Detection, Rectification and Segmentation of Coplanar Repeated Patterns

James Pritts    Ondřej Chum    Jiří Matas
Center for Machine Perception
Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Cybernetics
[prittjam,chum,matas]@cmp.felk.cvut.cz

Abstract

This paper presents a novel and general method for the detection, rectification and segmentation of imaged coplanar repeated patterns. The only assumption made of the scene geometry is that repeated scene elements are mapped to each other by planar Euclidean transformations. The class of patterns covered is broad and includes nearly all commonly seen, planar, man-made repeated patterns. In addition, novel linear constraints are used to reduce geometric ambiguity between the rectified imaged pattern and the scene pattern. Rectification to within a similarity of the scene plane is achieved from one rotated repeat, or to within a similarity with a scale ambiguity along the axis of symmetry from one reflected repeat. A stratum of constraints is derived that gives the necessary configuration of repeats for each successive level of rectification. A generative model for the imaged pattern is inferred and used to segment the pattern with pixel accuracy. Qualitative results are shown on a broad range of image types on which state-of-the-art methods fail.

1. Introduction

This paper presents a method for the detection, segmentation and rectification of a broad class of coplanar repeated patterns. A pattern is defined to be a coplanar arrangement of a motif (see Fig. 3). This definition is very general: no restrictive constraints are imposed on motif placement, e.g., constrained to a lattice, a circle, or a row. The method does not require the presence of parallel scene lines to work. The proposed detection and rectification method applies to nearly all man-made structures with repetitions or symmetries, which include building facades, mosaics, decorative prints and text.

The problem of rectification is closely coupled with the detection of a pattern and its motif in a chicken-and-egg fashion: rectification is easier if the motif is known, and estimation of the motif is easier from a rectified image. The proposed approach begins by establishing tentative intra-image correspondences of local affine frames (LAFs) [15] that could belong to a pattern. Rectification is estimated exclusively from the ratio of scales and lengths of extents of the imaged repeated scene structure.

Three types of repeated patterns are considered with the goal being to bring each imaged pattern type as geometrically close as possible to its scene counterpart to improve scene understanding. The three pattern types differ by the class of transformations between motif instances in the scene plane. Each type provides different geometric constraints and allows for different levels of rectification. A set of linear single-view geometric constraints are used that, depending on the pattern arrangement, enable estimation of rectification to within a similarity of the scene pattern. A generative model for the imaging process of the repeated pattern and its corresponding non-linear estimator is given. An explicit model of the pattern is constructed and projected.

1James Pritts and Ondřej Chum were supported by the GACR P103/12/2310 and MSMT LL1303 ERC-CZ grants, and Jiří Matas by the Czech Science Foundation Project GACR P103/12/G084.
Vanishing points
Proposed

1. Related work

Leung and Malik [10] proposed an early solution that grows a 2-D lattice from initial patches of high texture in a manner analogous to SSD tracking. Schaffalitzky and Zisserman [17] also used local patches to grow a grid, and parameterized imaged planar repetitions as conjugate translations and rotations of local patches. Tuytelaars et al. [19] grouped fixed points and lines from symmetries consistent with homologies estimated by Hough transform. Liu et al. [12] modeled the lattice structure of textures using a minimal set of tiles. Park et al. [16] modeled lattice topology with a Markov random field and allowed for some non-planar deformation. Doubek et al. [4] developed a shift-invariant descriptor to detect patterns with translated motifs. Hong et al. [8] developed constraints based on several classes of symmetries to estimate camera pose and pattern structure, but assumed that the symmetry type is given. Wu et al. [20, 21] introduced a repetition constraint and a corresponding optimization method to densely reconstruct repetitive structures, but they assumed the presence of parallel scene lines and that repeats occur on a grid. Rectification of repeated structures by rank minimization of the intensity matrix was proposed in [22]. The method is sensitive to background clutter and occlusions, and the region of interest must be manually annotated in problem cases. Liebowitz and Zisserman [11] estimated affine rectification from vanishing lines, then, similar to our approach, upgraded the rectification to within a similarity by enforcing the congruence of the length extents of repeated scene elements. But only rotated repetitions were considered. Recently, Aiger et al. [1] also proposed estimating rectification by maximizing congruence of the length extents of repeats. However, optimization was performed jointly for affine and metric rectification, which has no convergence guarantee since congruence is not preserved by either projective or affine transforms.

Köser et al. [9] proposed the estimation of conjugate rotation from a planar image correspondence, which they use as a local first-order Taylor approximation to the interimage homography. Using the same linearization, an algebraic constraint on the scale change of a planar scene feature transformed by a homography was introduced by Chum et al. in [3]. The method attempts to find a rectifying homography \( H_{\infty} \) so that the matching patches of the repeated structure cover identical areas. The proposed method extends the method in [3] by further reducing the ambiguity between imaged pattern and scene pattern for certain types of repeated patterns and significantly increases the precision of rectification by a non-linear optimization that explicitly reconstruct the pattern and accounts for radial lens distortion.

2. Problem formulation

The method searches the input image for a coplanar repeated pattern that can be generated by the assumed model illustrated in Fig. 4. The motif is repeatedly stamped on the scene plane (a priori unknown) at locations specified by Eu-
Figure 4. Generative model of the image of a repetitive pattern: (a) a motif is 'stamped' by transformations $T_j$ onto (b) a plane in 3D. The image of the plane (c) is captured by a projective camera with lens distortion (d). Only the final image (d) is the input of the method; the previous stages are inferred from intra-image correspondences. The only assumptions are made on the type of transformations $T_j$.

Euclidean transformations $T_j$, which constructs the pattern in a rectified space. The construction is imaged by a radially distorted projective camera. Since the construction lies on a scene plane, the imaging process is modeled with a homography. The projective part of the homography is captured by the image of the line at infinity. Finally, radial lens distortion is applied to the image plane creating the input image. Note that the image is the sole input to the algorithm, and all parameters of the generative model are estimated. An assumption of the generative model is that the positioning transformations $T_j$ are one of three types typically encountered: pure translation, composition of an axial symmetry with translation, or a general Euclidean (rigid) transformation (Fig. 3). It is of little interest to assume that $T_j$ is as general as a homography because (i) such repeated elements rarely appear on real-world scene planes, and (ii) it is impossible to decouple the projective placement of the scene element from the projective effect of the camera. There are no other assumptions about the positioning transformations: repeated elements do not have to create a lattice or assume any other form.

The motif is a spatially localized collection of features that form the repeated element. Due to noise or occlusions, we assume that each instance of the motif in the pattern consists of a subset of spatially consistent motif features; i.e., not all motif features have to appear in all instances. To contribute to a solution, an instance of a motif must contain at least two features. For a solution to be found, at least two motif instances must be detected.

3. The Method

This section describes the step-by-step estimation of the parameters of the generative model from a single image of a repeated pattern. An overview of the approach is given in Alg. 1.

Feature appearance matching. To obtain sets of features potentially arising from different instances of the repeated motif, an intra-image matching procedure is performed. Affine invariant local features are first extracted in the image. We use MSER features [14] with the local affine frame (LAF) extension [15] since it provides a local coordinate system for each feature in the form of a triple of points. SIFT descriptors [13] are extracted from the distinguished regions defined by the LAFs. Features used for construction of the reflected instances are detected in a reflected image. Spectral descriptor matching is employed to robustly extract clusters of features with similar appearance. Examples of matching LAFs can be found in Figs. 5 and 6.

Projective distortion removal. The line at infinity is estimated from the change of scale of matching local features following [3]. In a projective image, patches that are farther from the camera appear smaller. A rectifying homography is estimated so that all matching features detected in different instances of the motif have equal area after rectification. The constraints on the homography are derived from the scale change at a point, or, equivalently, at an infinitesimally small patch around the point. This scale change is given by the determinant of an affine transformation locally approximating (first order Taylor expansion) the homography at the point. The infinitesimal scale changes are estimated from the areas of local, affine covariant features. This approach gives a set of linear equations

$$x_i h_7 + y_i h_8 - s_i^{1/3} \alpha_j = -1,$$

where $(x_i, y_i)$ are the coordinates of $i$-th feature point, $s_i$ is the local scale of that feature, $(h_7, h_8, 1)$ are the coordinates...
of the line at infinity, and $\alpha_j$ is related to the relative scale of different feature sets (indexed by $j$, if more than one set of matching features are used to estimate the line at infinity) after rectification. Three corresponding features in general position or three sets of two matching features are sufficient to estimate the line at infinity. For more details see [3].

**Motif construction.** Each SIFT cluster represents a set of intra-pattern tentative correspondences between motif elements that were detected as MSERs and represented as local affine frames (LAFs). The clustering implicitly assumes that local neighborhoods covered by LAFs have a large overlap with the motif and contain large subsets of the pattern. SIFTs of the largest clusters are used to define sets of overlapping neighborhoods in the pattern. Using the SIFTs’ corresponding LAFs, which define local coordinate systems at the SIFT’s origin, SIFT neighborhoods are transformed to the standard orthonormal basis, where surrounding features are aggregated in a common space. In the standard basis frame, spatially overlapping features are identified as repeated elements and added to the motif representation. This process is repeated for a fixed number of the top SIFT clusters. Mismatches are discarded by spatial verification in the canonical frame. Instances of a motif are shown in Fig. 6(c).

**Affine distortion correction.** The method of [3] provides planar rectification up to an unknown affine transformation. If the motif instances are reflected or rotated (which is easily detected even under an affine ambiguity), then it is possible to further reduce the ambiguity. The methods use constraints from the lengths of corresponding line segments and lead to systems of linear equations. The contribution is detailed in 3.1.

**Non-linear optimization.** The parameters of the generative model for the imaged pattern (Fig. 4) are initialized from the linear estimates of rectification and from the motif construction in rectified space. Subsequently, re-projection error in the image is minimized by a non-linear least squares optimizer (Levenberg–Marquardt). This step is detailed in 3.2.

**Robustness.** Since the intra-image matching is prone to false matches, estimation of rectification needs to be robust to outliers. RANSAC [6] is used to estimate the line at infinity and to estimate an further affine distortion reduction if reflected or rotated repeats are detected. The motif construction stage filters out sets of features that do not have sufficient support in their neighborhood, which is similar to the approach of [18].
The derived constraints are linear, so they are simple to implement and efficient to use in a RANSAC-like estimator. Fast estimation can be achieved from minimal sets (indexed by \( i \)) of corresponding rotated line segments of length \( r_j \), each set having at least two elements (indexed by \( i \)). The length constraint can be written as

\[
\hat{x}_{ij}^\top \hat{x}_{ij} = \hat{r}_j^2.
\]

(6)

The situation is depicted in Fig. 7(b). Corresponding vectors are color coded and Equation (6) is depicted as circles with diameter \( r_j \). The length constraint (6) in the image transformed by \( \tilde{A}^{-1} \) according to (2) changes to

\[
\Sigma x_{ij} = \tilde{r}_j,
\]

where \( \Sigma = \tilde{A}^\top \tilde{A} \).

(7)

In equation 7, \( \Sigma \) represents an ellipse (visualized in Fig. 7c), where

\[
\Sigma = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.
\]

(8)
Equation (7) can be rewritten as
\[ (x^2_{ij} - 1)(a \ b \ c) = 0, \] (9)
which gives a system of homogeneous linear equations. There are three unknowns for \( \Sigma \), and each set of matching line segments adds one unknown \( r_j \). Each participating line segment in general position (rotation) adds one constraint. For two pairs of line segments, there are \( 3 + 2 = 5 \) unknowns and four linear equations, yielding a one-dimensional linear space of solutions. An alternative minimal solution is given by one triplet of reflected line segments, having \( 3 + 1 = 4 \) unknowns and 3 linear equations.

The affine transformation can be derived from the solution of the system of linear equations (9) up to a scale factor and a rotation. The unknown scale comes from the homogeneous nature of the system–both \( \Sigma \) and \( r_j^2 \) can be multiplied by a positive scalar. The rotational ambiguity comes from the ambiguity of the decomposition \( \Sigma = \hat{H}^T \hat{A} \).

A rotation by 180 degrees (or an integer multiple) creates a special case: if the pattern is only rotated by integer multiplications of 180 degrees, then the matching vectors lie on parallel lines. Since affine transformations affect the lengths of vectors on parallel lines equally, the situation is similar to the pure translation case with full affine ambiguity.

3.2. Non-linear estimation

The parameters of the generative model depicted in (Fig. 4) are estimated by a non-linear minimization of pattern re-projection error. Scene plane rectification, position transformations and localization of detected features are jointly refined, and the camera model is extended to include radial distortion. Radial distortion is modeled by the division model [7]. The minimization is accomplished by Levenberg–Marquardt. The parameters of the optimization include the position of the feature points in the motif \( \hat{v}_i \), the parameters of the rectifying homography \( \hat{H} \), which is a composition of the non-projective upgrade (if there is one) and \( \hat{H}_{\infty} \), the parameters of the positioning transformations \( \hat{T}_j \) for each motif instance, and the radial distortion parameter \( \hat{r} \). A geometric re-projection of the reconstructed pattern with the measured feature points in the original image counterpart is given by the following cost function
\[ \min_{\hat{H}, \hat{r}, \hat{v}_i} \sum_{i=1}^{N} \sum_{j=\exists u_{ij}} d(\mathcal{L}_{\hat{A}}(\hat{H}^{-1} \hat{T}_j(\hat{v}_i)), u_{ij}). \] (10)
The outer sum is over all \( N \) points of the motif, the inner sum is over instances \( j \) where the \( i \)-th motif point was observed, \( d(x, y) \) is the Euclidean distance between the images of \( x \) and \( y \), and \( \mathcal{L}_{\hat{A}} \) implements the radial distortion.

The rectifying homography \( \hat{H} \) is minimally parameterized based on the geometric constraints available from the pattern arrangement (see 3.1).

4. Experiments

In this section, we present experimental results that validate the broad applicability of the proposed method. The ambiguity of image registration is resolved by registering the rectified image to the original image using a transformation that corresponds to the rectification type: affine for an affine ambiguity, similarity for similarity ambiguity, etc. This is done solely for the visualization purposes so that the rectified images are not significantly distorted.

4.1. Pure Translation

The case where the motif is only translated in the scene plane offers rectification only up to an arbitrary affine transformation of the scene plane: thus, it is the least interesting. Examples of repeated patterns with translated motifs are shown in Figs. 8 and 9.

4.2. Reflection

The presence of a reflected motif(s) allows for the detection of two perpendicular vanishing points, one of which is aligned with the axis of symmetry. Fig. 10 shows an example of a symmetric repeated pattern: the windows. The figure also provides a comparison to an affine rectification gotten by the method given in [3]. More examples are shown in Figs. 5, 1-top, Figs. 11. The motif in the FEET example in Fig. 11 is not self-symmetric.

4.3. Rotation

Fig. 2 compares the rectification results of different methods on the target image from [3], which contains a rotated repeated pattern. A rectification of an image of the famous Escher painting "Tessellation 85" is shown in Fig. 12. This experiment demonstrates the robustness of the proposed algorithm to noise. Since the painting is hand-drawn, each of the lizards, bats and fish are, in fact, unique.
4.4. Improving the descriptor accuracy

The non-linear minimization not only improves the estimate of the rectification transformation, but, more importantly, it estimates the noise free feature points. In order to quantify the gain, the following statistics were computed both before the rectification (i.e. over the detected features), and after the rectification (over the geometrically corrected features). For each feature in the motif, an average SIFT descriptor was computed from its instances. The average distance of feature SIFT descriptors to the corresponding average SIFT descriptor was computed. The rectification reduces the intra-cluster SIFT distance by 35.2% on average for the images in Figs. 2, 8, 9. This is an important result for applications such as image retrieval.

4.5. Multiple coplanar patterns

It is natural to consider multiple motifs for modeling images that contain multiple patterns. Fig. 8 shows two different groups of posters cosegmented from multiple motif detections. In this case, the instances are on the same plane, and the constraints on the line at infinity and the radial distortion are shared. If the instances of the motifs live on a different planes, then new parameters for the line at infinity for each plane are introduced, but the radial distortion parameter is the same for all the planes.

5. Conclusions

A novel and fully automated method for the detection, rectification and pixel accurate segmentation of coplanar repeated patterns is introduced. Besides coplanarity, the algorithm makes virtually no assumptions about the arrange-
ment of the motif. A set of linear single-view geometry constraints are introduced to estimate rectification up to a similarity of the scene plane from a minimal set of intra-image correspondences—affine distortion can be removed from just one rotated or reflected repeated scene element. These constraints are used in a fast RANSAC framework to provide robust intra-pattern feature matching and pattern rectification.

The imaged pattern is modeled generatively. Model optimization concurrently refines rectification, pattern geometry and lens distortion. The model-based approach enables the reprojection of the constructed pattern, which is used to segment the pattern with pixel-level accuracy.

Experiments verify that the method works when the pattern is a small part of the image, of low texture, with an arbitrarily arranged motif and with a significant perspective warp. Thus the proposed method is applicable to a broad class of patterns, including man-made, planar structures with repetition or symmetry, such as building facades, mosaics, decorative prints and text.

References