15-16

15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Estimation of Radial Distortion Using Local Spectra of Planar Textures

Benjamin Spitschan and Jörn Ostermann
Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover
Appelstr. 9A, 30167 Hannover, Germany
spitschan@tnt.uni-hannover.de

Abstract

A novel self-calibration method for estimation of radial lens distortion is proposed. It requires only a single image of a textured plane that may have arbitrary orientation with respect to the camera. A frequency-based approach is used to estimate the perspective and non-linear lens distortions that planar textures are subject to when projected to a camera image plane. The texture is only required to be homogeneous and may exhibit a high amount of stochastic content. For this purpose, we derive the relationship between the local spatial frequencies of the texture and those of the image. In a joint optimization, both the rotation matrix and the radial distortion are subsequently estimated. Results show that with appropriate textures, a mean reprojection error of $9.76 \cdot 10^{-5}$ relative to the picture width is achieved. In addition, the method is robust to image corruption by noise.

1 Introduction

Camera calibration is of great importance in many computer and machine vision applications. Along with photometric calibration, geometric calibration is a key component in the identification of a camera system's imaging properties, and involves the estimation of intrinsic and extrinsic camera parameters. Extrinsic parameters define the 3-D coordinate system transformations from world coordinates to camera coordinates, i. e. the relative pose of an object with respect to the camera. Intrinsic parameters determine the perspective mapping of 3-D points given in camera coordinates to the image plane, i. e. the camera sensor. Commonly, a pinhole camera model is assumed for the mapping, resulting in a linear projective transformation.

Real-world cameras exhibit several additional non-linearities including non-planar sensor geometry or lens distortion. In particular, wide-angle lenses usually lead to strong distortions. They can only partly be mitigated by increasing the expenditure of lens production. For many industrial machine vision tasks, however, geometrically correct, i. e. linear projective portrayal of the scene is required. Undistorting an image corrupted by lens distortion is therefore a common pre-processing step. The usual approach is to measure lens distortion at a single point of time before camera deployment using specialized calibration targets. In some applications, however, calibration prior to operation is not feasible, or only the resulting image is available. In these cases, self-calibration methods have to be employed. They enforce constraints on the scene, the camera motion or the intrinsic parameters, e. g. by incorporating a-priori knowledge of the image content such as the presence of straight lines. In contrast, we present a method for the joint estimation of extrinsic and intrinsic camera parameters by using planar textures in the scene. It is based on the distortion of texture patches caused by perspective projection and lens distortion (Fig. 1). This projection is analyzed in the frequency domain.

Planar textures occur in many industrial applications such as optical quality inspection of man-made surfaces. In order to accommodate for flexible imaging environments, a versatile self-calibration mechanism is needed. For instance, images of machined—i. e. ground and milled—metal surfaces are well-suited for the proposed method.
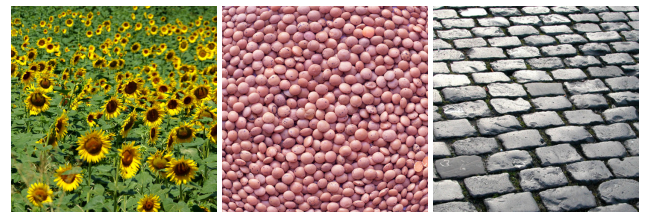

Figure 1. Examples of textured images exhibiting perspective distortion (left panel), radial lens distortion (center panel), and combined distortion (right panel).

Our paper is organized as follows. In Section 2 an overview over previous work and its limitations is given. In Section 3 we recapitulate the principles of perspective and non-linear projection and present our frequency-based approach to estimate it. The method is subsequently evaluated in Section 4. We conclude with a discussion of the results.

2 Related Work

Geometric calibration is of ongoing interest in both the computer vision and the photogrammetry communities. Surveys can be found in [1–3]. A recent assessment of several state-of-the art methods was carried out in [4]. Target-based algorithms require checkerboards [5–7] or other markers. Sinusoidal patterns displayed on computer screens were proposed in [8]. The use of dedicated calibration targets, however, is not possible in many configurations. Also, it does not allow for a dynamic re-calibration to account for changing camera parameters such as the change of focus or focal length in a zoom lens. Self-calibration techniques do not rely on specialized targets, but instead exploit the image content itself. The "plumbline" approach attempts to detect and rectify lines in the image [9, 10]. The assumption of straight lines within the image is not always fulfilled, though. The possibility to utilize textures for calibration was explored in [11], making a low-rank assumption on texture patches and requiring multiple occurrences of those patches. Highly stochastic textures therefore fail due to their high entropy and full rank.

The non-linear projection of planar surfaces on the one hand and the linear projection of arbitrarily shaped surfaces on the other hand represent to a limited extent a dual problem. Therefore, the proposed method is strongly connected to the shape-from-texture (SfT) problem. A recent general survey on SfT methods is given in [12]. In [13], frequency-based shape recovery was performed, based on earlier work [14,

15]. However, shape recovery poses integrability constraints which are different from those introduced by perspective and radial distortion.

3 Radial Distortion from Local Spectra

We propose a method that exploits the spectral content of projected textures for estimating radial lens distortion. It requires only a single picture of a texture (Fig. 1). We assume a textured plane to be freely rotated in space. We then find an expression for the backprojection of non-linearly projected coordinates of that plane imaged by a projective camera with lens distortion. If the planar texture is homogeneous, its spectral content is constant over the location in texture coordinates. Following the parlance of [16], the "instantaneous" 2-D frequency of an image intensity signal will be called *local spatial frequency* (LSF). After the projection, the texture and therefore its LSFs are unequally distorted depending on the image coordinate. Using the backprojected coordinates and the relation between frequency and phase of a signal, we are able to give the relationship between the LSFs of the projected image and those of the original texture. The location-dependent spectral content of the image is estimated using a Gabor filter band. Using these spectra as observation vector and the analytically-derived backprojected spectra as objective function, the process of estimating the extrinsic and non-linear intrinsic camera parameters is expressed as a non-linear least-squares minimization problem.

Notation Independent of letter case and font style, variable names "p" indicate spatial coordinates of different kinds and variable names "I" indicate intensity signals. Vectors and matrices are printed in boldface. For clarity, spatial quantities such as coordinates, frequencies, and phases with superscript index $(\cdot)^{\mathrm{I}}$ denote quantities of the *projective image*, whereas quantities with index $(\cdot)^{\mathrm{Id}}$ are quantities of the *lens-distorted image* and quantities with index $(\cdot)^{\mathrm{T}}$ are *texture* quantities. A lower-case sans-serif superscript $(\cdot)^{\mathrm{t}}$ marks the transpose of a vector or matrix.

3.1 Geometric camera model

In the pinhole camera model [17], a projective camera is assumed with rotation $\boldsymbol{R} \equiv (r_{ij}) \in \mathbb{R}^{3\times 3}$ and translation $\boldsymbol{t} \equiv (t_x, t_y, t_z)^{\mathrm{t}}$ (extrinsic parameters), and camera calibration matrix $\boldsymbol{K} \in \mathbb{R}^{3\times 3}$ (intrinsic parameters). We assume a simplified calibration matrix with focal length f, vanishing skew coefficient, and a principal point centered at the origin, i.e.

$$\boldsymbol{K} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{1}$$

Let the rotation matrix

$$\boldsymbol{R} \equiv \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \boldsymbol{R}_\gamma \boldsymbol{R}_\beta \boldsymbol{R}_\alpha \tag{2}$$

be constructed of concatenated rotations $\boldsymbol{R}_\gamma$, $\boldsymbol{R}_\beta$, and $\boldsymbol{R}_\alpha$ around the unit vectors by α, β, and γ.

In general, a scene point $\boldsymbol{P} \equiv (P_x, P_y, P_z)^{\mathrm{t}}$ in 3-D world coordinates is projected to a homogeneous coordinate

$$\boldsymbol{p} \equiv (p_x, p_y, p_z)^{\mathrm{t}} = \boldsymbol{K}\boldsymbol{R}(\boldsymbol{P} - \boldsymbol{t}), \tag{3}$$

obtaining the resulting image coordinate $\boldsymbol{p}^{\mathrm{I}} \equiv \left(p_x^{\mathrm{I}}, p_y^{\mathrm{I}}\right)^{\mathrm{t}}$ through normalization by the z component:

$$\boldsymbol{p}^{\mathrm{I}} = \begin{pmatrix} \dfrac{p_x}{p_z} \\ \dfrac{p_y}{p_z} \end{pmatrix} = f \cdot \begin{pmatrix} \dfrac{r_{11}(P_x - t_x) + r_{12}(P_y - t_y) + r_{13}(P_z - t_z)}{r_{31}(P_x - t_x) + r_{32}(P_y - t_y) + r_{33}(P_z - t_z)} \\ \dfrac{r_{21}(P_x - t_x) + r_{22}(P_y - t_y) + r_{23}(P_z - t_z)}{r_{31}(P_x - t_x) + r_{32}(P_y - t_y) + r_{33}(P_z - t_z)} \end{pmatrix}. \tag{4}$$

3.2 Lens distortion model

The geometric lens distortion introduced by lens imperfections is a non-linear function $L: \mathbb{R}^2 \to \mathbb{R}^2$ and results in a location-dependent displacement $\delta\left(\boldsymbol{p}^{\mathrm{I}}\right) \in \mathbb{R}^2$ such that the resulting distorted image coordinate $\boldsymbol{p}^{\mathrm{Id}} \equiv \left(p_x^{\mathrm{Id}}, p_y^{\mathrm{Id}}\right)$ is

$$\boldsymbol{p}^{\mathrm{Id}} = L\left(\boldsymbol{p}^{\mathrm{I}}\right) \equiv \begin{pmatrix} L_x\left(\boldsymbol{p}^{\mathrm{I}}\right) \\ L_y\left(\boldsymbol{p}^{\mathrm{I}}\right) \end{pmatrix} = \boldsymbol{p}^{\mathrm{I}} + \delta\left(\boldsymbol{p}^{\mathrm{I}}\right). \tag{5}$$

Commonly, the Brown-Conrady approach [18, 19] is used to parametrize L, modeling the distortion by radial and tangential displacement components which are approximated by a Taylor series expansion. It was shown in [5] that the tangential components are negligible for many applications. Thus, only the radial components of the Taylor series are considered, accounting for barrel and pincushion distortion:

$$\boldsymbol{p}^{\mathrm{Id}} = L\left(\boldsymbol{p}^{\mathrm{I}}\right) = \boldsymbol{p}^{\mathrm{I}}\left(1 + \sum_{i=1}^{\infty} c_i\, r^{\mathrm{I}\,2i}\right), \tag{6}$$

with $r^{\mathrm{I}} \equiv \left\|\boldsymbol{p}^{\mathrm{I}}\right\|_2 = \sqrt{p_x^{\mathrm{I}\,2} + p_y^{\mathrm{I}\,2}}$ denoting the distance from the image center, and coefficients $c_i \in \mathbb{R}$.

In order to rectify the image, non-linear distortions have to be reversed. However, Eq. (6) cannot be inverted to a closed-form solution. In accordance with [9], the inverse transformation L^{-1} is modeled by a Taylor expansion as well, which is stopped after the first radial displacement component determined by the coefficient $\kappa \in \mathbb{R}$:

$$\boldsymbol{p}^{\mathrm{I}} = L^{-1}\left(\boldsymbol{p}^{\mathrm{Id}}\right) \equiv \begin{pmatrix} L_x^{-1}\left(\boldsymbol{p}^{\mathrm{Id}}\right) \\ L_y^{-1}\left(\boldsymbol{p}^{\mathrm{Id}}\right) \end{pmatrix} = \boldsymbol{p}^{\mathrm{Id}}\left(1 + \kappa\, r^{\mathrm{Id}\,2}\right), \tag{7}$$

with distance $r^{\mathrm{Id}} \equiv \left\|\boldsymbol{p}^{\mathrm{Id}}\right\|_2 = \sqrt{p_x^{\mathrm{Id}\,2} + p_y^{\mathrm{Id}\,2}}$ from the center of the distorted image.

3.3 Textures

We presume that larger, not necessarily connected regions of an image are a projection of a texture. We further require that this texture fulfills the homogeneity constraint, i.e. it is location-independent [20, 21]. Furthermore, it may contain strong stochastic signal components.

Our method exploits the assumption that the perspective and non-linear distortions of texture patches are locally affine. Instead of allowing arbitrary orientations of the textured plane, we keep it fixed and rotate the camera instead. We therefore assume a texture intensity signal $I\left(\boldsymbol{p}^{\mathrm{T}}\right)$ at texture coordinate $\boldsymbol{p}^{\mathrm{T}} \equiv \left(p_x^{\mathrm{T}}, p_y^{\mathrm{T}}\right)^{\mathrm{t}}$ located in the x-y plane of the 3-D world coordinate system, corresponding to

$$\boldsymbol{P} = \begin{pmatrix} p_x^{\mathrm{T}} \\ p_y^{\mathrm{T}} \\ 0 \end{pmatrix}, \tag{8}$$

and

$$t = \begin{pmatrix} 0 \\ 0 \\ t_z \end{pmatrix}. \qquad (9)$$

In this case, Eq. (4) simplifies to

$$p^{\mathrm{I}} = f \cdot \begin{pmatrix} \dfrac{r_{11}p_x^{\mathrm{T}} + r_{12}p_y^{\mathrm{T}} - r_{13}t_z}{r_{31}p_x^{\mathrm{T}} + r_{32}p_y^{\mathrm{T}} - r_{33}t_z} \\[2mm] \dfrac{r_{21}p_x^{\mathrm{T}} + r_{22}p_y^{\mathrm{T}} - r_{23}t_z}{r_{31}p_x^{\mathrm{T}} + r_{32}p_y^{\mathrm{T}} - r_{33}t_z} \end{pmatrix}. \qquad (10)$$

Due to the planarity constraint stated in Eq. (8), Eq. (10) can be inverted up to a constant scale factor. The backprojected homogeneous coordinate $p' \equiv \left(p'_x, p'_y, p'_z \right)^{\mathrm{t}}$ is

$$p' = R^{-1}K^{-1} \begin{pmatrix} p_x^{\mathrm{I}} \\ p_y^{\mathrm{I}} \\ 1 \end{pmatrix} = R^{-1}K^{-1} \begin{pmatrix} L_x^{-1}\left(p^{\mathrm{I_d}}\right) \\ L_y^{-1}\left(p^{\mathrm{I_d}}\right) \\ 1 \end{pmatrix} \qquad (11)$$

and the backprojected coordinate $p^{\mathrm{T}'} \equiv \left(p_x^{\mathrm{T}'}, p_y^{\mathrm{T}'} \right)^{\mathrm{t}}$ in texture coordinates is

$$p^{\mathrm{T}'} = \begin{pmatrix} p'_x / p'_z \\ p'_y / p'_z \end{pmatrix} = -\frac{p^{\mathrm{T}}}{t_z}. \qquad (12)$$

The scale factor $-1/t_z$ is introduced by the ambiguity with respect to the distance of the camera.

Without loss of generality, we consider a 2-D sinusoidal texture signal

$$I\left(p^{\mathrm{T}}\right) \equiv \tfrac{1}{2} + \tfrac{1}{2} \cos \omega_0^{\mathrm{T}\,\mathrm{t}} p^{\mathrm{T}} \qquad (13)$$
$$= \tfrac{1}{2} + \tfrac{1}{2} \cos \left(\omega_{0,x}^{\mathrm{T}} p_x^{\mathrm{T}} + \omega_{0,y}^{\mathrm{T}} p_y^{\mathrm{T}} \right)$$

with 2-D spatial frequency $\omega_0^{\mathrm{T}} \equiv \left(\omega_{0,x}^{\mathrm{T}}, \omega_{0,y}^{\mathrm{T}} \right)^{\mathrm{t}}$. Let $I\left(p^{\mathrm{I_d}}\right)$ denote the intensity signal of the distorted image. Since $p^{\mathrm{I_d}}$ corresponds to p^{T}, the intensity values in the projected image at distorted coordinates $p^{\mathrm{I_d}}$ are the corresponding intensity values in the texture image at original coordinates p^{T}:

$$I\left(p^{\mathrm{I_d}}\right) = I\left(p^{\mathrm{T}}\right). \qquad (14)$$

Here, we neglect distance-depending haze effects and low-pass filtering introduced by the point-spread function of the lens.

Subsequently, the local phases in the projected image and the texture image are equal:

$$\varphi^{\mathrm{I}}\left(p^{\mathrm{I_d}}\right) = \varphi^{\mathrm{T}}\left(p^{\mathrm{T}}\right). \qquad (15)$$

3.4 Estimation of lens distortion

In general, the 2-D LSF $\omega(p) \equiv \left(\omega_x(p), \omega_y(p) \right)^{\mathrm{t}}$ is the gradient of the local phase:

$$\omega(p) = \nabla\varphi(p). \qquad (16)$$

Applying the gradient operator on both sides of Eq. (15) and substituting its backprojection $-t_z p^{\mathrm{T}'}$ for p^{T} according to Eq. (11), we find

$$\omega^{\mathrm{I}}\left(p^{\mathrm{I_d}}\right) = \nabla\varphi^{\mathrm{T}}\left(-t_z p^{\mathrm{T}'}\right). \qquad (17)$$

Hence,

$$\omega^{\mathrm{I}}\left(p^{\mathrm{I_d}}\right) = -t_z \nabla\left(\omega_0^{\mathrm{T}\,\mathrm{t}} p^{\mathrm{T}'} \right). \qquad (18)$$

We find the LSFs

$$\omega^{\mathrm{I}}\left(p^{\mathrm{I_d}}\right) = -t_z \begin{pmatrix} \dfrac{\partial}{\partial p_x^{\mathrm{I_d}}} \\[2mm] \dfrac{\partial}{\partial p_y^{\mathrm{I_d}}} \end{pmatrix} \left(\omega_{0,x}^{\mathrm{T}} \dfrac{p'_x}{p'_z} + \omega_{0,y}^{\mathrm{T}} \dfrac{p'_y}{p'_z} \right) \qquad (19)$$

in the distorted image as 2-D function

$$\omega^{\mathrm{I}}\left(p^{\mathrm{I_d}}\right) = F\left(p^{\mathrm{I_d}}, \kappa, K, R\right) \qquad (20)$$

of $p^{\mathrm{I_d}}$ and κ, K, and R.

After inserting Eq. (7) and Eq. (11) into Eq. (19), F is derived analytically by a symbolic solver to a closed-form solution, which is not given here for reasons of brevity.

The estimation of the camera parameters can then be formulated as the non-linear least-squares minimization problem

$$\arg\min_{\kappa, K, R} \left\| \widehat{\omega^{\mathrm{I}}}\left(p^{\mathrm{I_d}}\right) - F\left(p^{\mathrm{I_d}}, \kappa, K, R\right) \right\|_2^2, \qquad (21)$$

with observations $\widehat{\omega^{\mathrm{I}}}\left(p^{\mathrm{I_d}}\right)$ at observation points $p^{\mathrm{I_d}}$.

We therefore need to find estimates $\widehat{\omega^{\mathrm{I}}}\left(p^{\mathrm{I_d}}\right)$ of the 2-D LSFs in the projected image $I\left(p^{\mathrm{I_d}}\right)$.

Unlike previous methods such as [8] for calibration with sinusoidal patterns, or [13] for SfT techniques, we do not assume a single dominant frequency and drop the confinement to sinusoidal textures (13) in the following. Instead, the entire local spectrum of a texture patch at image coordinate $p^{\mathrm{I_d}}$ is utilized. Thus, the ill-posed problems of 3-D integration and 2-D phase unwrapping are avoided.

For spectral estimation, a 2-D k-band Gabor filter bank with impulse responses

$$g_n\left(p, \omega^{\mathrm{I}}\right) = \mathrm{e}^{-\frac{1}{2}\left(\frac{p_x^2}{\sigma_x^2} + \frac{p_y^2}{\sigma_y^2}\right)} \cdot \mathrm{e}^{-\mathrm{j}\left(\omega_{n,x} p_x + \omega_{n,y} p_y\right)}, \qquad (22)$$

$n = 1 \ldots k$, is used. The center frequencies are linearly spaced within the frequency range. Its lower and upper boundaries are determined in a pre-processing step by identifying the range of the significant frequencies in the magnitude of the 2-D Fourier spectrum of the projected image.

Applying the Gabor filter bank on the textured regions of the distorted image $I\left(p^{\mathrm{I_d}}\right)$, the 2-D spectral estimate

$$\widehat{S}\left(p^{\mathrm{I_d}}\right) = \underbrace{\begin{pmatrix} g_{1,x} & \cdots & g_{k,x} \\ g_{1,y} & \cdots & g_{k,y} \end{pmatrix}}_{\equiv\, G(p^{\mathrm{I_d}})} * I\left(p^{\mathrm{I_d}}\right) \qquad (23)$$

of size $2 \times k$ with the elements

$$g_{n,x} = g\left(p^{\mathrm{I_d}}, \begin{pmatrix} \omega_{n,x} \\ 0 \end{pmatrix}\right), \quad g_{n,y} = g\left(p^{\mathrm{I_d}}, \begin{pmatrix} 0 \\ \omega_{n,y} \end{pmatrix}\right) \qquad (24)$$

of $G\left(p^{\mathrm{I_d}}\right)$ is obtained for each image coordinate. Subsequently, the spectrum magnitude $\left| \widehat{S}\left(p^{\mathrm{I_d}}\right) \right|$ is calculated.

To accommodate for low-pass filtering due to the point-spread function of the optical system, the spectral estimate is

normalized to its L^2 norm. Since the homogeneity constraint is fulfilled, we can assume constant spectra throughout the original texture image.

The full optimization problem can now be written as

$$\arg\min_{\kappa,\boldsymbol{K},\boldsymbol{R}} \left\| \left| \widehat{\boldsymbol{S}}\left(\boldsymbol{p}^{\mathrm{Id}}\right) \right| - \left| \mathcal{F}\left\{\boldsymbol{G}\left(\boldsymbol{p}^{\mathrm{Id}}\right)\right\} \right|_{\boldsymbol{\omega}=F\left(\boldsymbol{p}^{\mathrm{Id}},\kappa,\boldsymbol{K},\boldsymbol{R}\right)} \right\|_{\mathrm{Frob}}^2 , \tag{25}$$

with $\mathcal{F}\{(\cdot)\}$ denoting the Fourier transform. The objective function is the magnitude of the frequency-domain filter bank response evaluated at the projected LSFs. The Frobenius norm is used as matrix distance to be minimized. It can be found from Eqs. (1) and (2) that a total of five parameters (κ, f, α, β, and γ) have to be estimated.

4 Evaluation

4.1 Implementation

The number of bands of the Gabor filter bank was set to $k = 16$. For solving Eq. (25), the Levenberg-Marquardt algorithm [22, 23] was chosen due to its robust convergence characteristics. The Jacobian of the objective function was derived analytically by a symbolic engine and provided as input. $\kappa = 0$, $f = 1$, and $\alpha = \beta = \gamma = 0°$ were used as initial parameters for all experiments.

4.2 Test data

For evaluation, three natural grayscale textures (D22, D77, and D92) from the BRODATZ dataset [24] (Fig. 2, top row) and three high-resolution textures (blanket1, lentils1, wall1) from the KYLBERG dataset [25] were selected. Textures with strong repetitive structures (e.g. D22 and D77) or strong stochastic content (e.g. D92) were chosen. All images exhibit negligible perspective and lens distortion and fulfill the homogeneity and isotropy requirements. In addition, two textures (stones, coffee, (Fig. 2, bottom left and center) were evaluated that were modified to allow for seamless perfect repetition. As reference, a superimposition of two orthogonal 2-D sinusoidals was used (Fig. 2, bottom right).
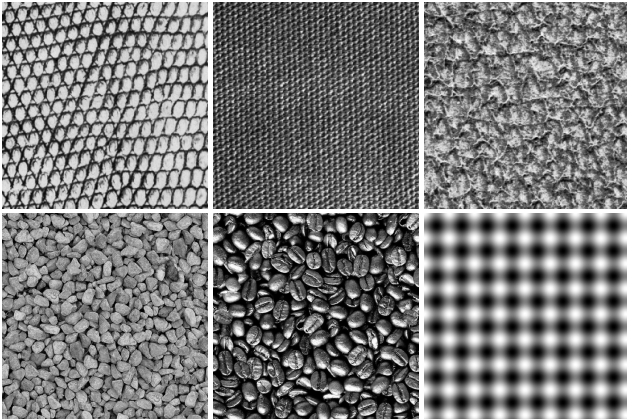

Figure 2. Top, from left to right: Details of BRO-DATZ textures D22, D77, and D92 (taken from [24]). Bottom, from left to right: Seamless textures stones and coffee, and superimposition of orthogonal sinusoidals (reference image).
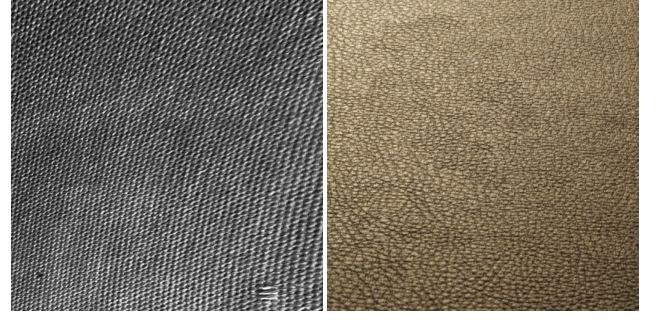

Figure 3. Left: Example of non-linearly projected texture D77. Right: Example of faux leather specimen.

4.3 Results

Non-linear projection was simulated by rotating the texture images around the x, y, and z axes by different angles in the range of $\pm 25°$ and subsequently applying different radial distortions with values of κ in the range of ± 0.5. Image coordinates were converted to floating-point values and normalized to ± 0.5 prior to processing; all results relate to a normalized image width (see Fig. 3, left panel, for a simulated projection).

Simulated radial and perspective distortion The RMS error of κ depending on κ and the angle between the projection axis and the z axis is depicted in Fig. 4, left panel. The results were averaged across all eight test images excluding the sinusoidal reference pattern. It can be seen that the estimation performs worse if the texture is less slanted and tilted. For the reference pattern, the RMS error of κ is almost constant over κ and the axis angle and amounts to $4.17 \cdot 10^{-4}$.

Simulated radial distortion only We assumed a projection axis perpendicular to the texture plane and only estimated κ, i.e. we fixed $\alpha = \beta = \gamma = 0$. κ was varied in 20 steps between -0.5 and 0.5. The RMS error of κ is depicted in Fig. 4, right panel. It can be noticed that our algorithm performs worse when the range of LSFs is low within the distorted image. This is similar to the findings in the combined radial/perspective distortion case. For the reference pattern, the RMS error of κ is $8.63 \cdot 10^{-5}$.
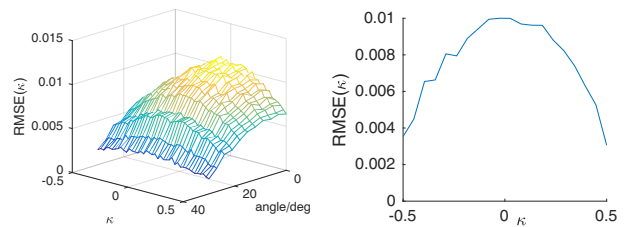

Figure 4. Left: Estimation of perspective and radial distortion. Right: Estimation of radial distortion only.

Physical projection Additional experiments were carried out with a faux leather specimen that was photographed in different orientations. The image (Fig. 3, right panel) exhibits a narrow depth of field and inhomogeneous illumination. The mean reprojection error of 32 manually

Table 1. Impact of additive Gaussian noise on estimation accuracy.

σ_n	0	1	2	5
$e_{\mathrm{repr}}/10^{-4}$	1.41	1.65	2.87	7.31
$\mathrm{RMSE}(\kappa)/10^{-3}$	4.74	6.81	10.40	18.29

marked points distributed equally across the image is $e_{\mathrm{repr}} = 9.76 \cdot 10^{-5}$. We compared our method with the results of the Caltech toolbox [26] implementation of [6, 7]. Evaluating three images of different poses of a 9×9 checkerboard target, a mean reprojection error of $e_{\mathrm{repr}} = 3.12 \cdot 10^{-5}$ is obtained. Hence, we are able to achieve very similar accuracy by evaluating just a single image instead of three.

Influence of noise To investigate the susceptibility to noise, we rotated the D77 texture by $15°$ around the x and y axes, respectively, distorted the image with $\kappa = 0.25$, and added Gaussian noise of different standard deviations σ_n. The values of σ_n relate to an intensity value range of 0 to 255. Each experiment was conducted 100 times. Average results for the mean reprojection error e_{repr} and the RMS error of κ are shown in Table 1. For a noise standard deviation of $\sigma_n < 2$, there is no significant impact on estimation accuracy.

5 Conclusions

In this paper, we presented a novel method for estimation of lens distortion. It requires only a single image of homogeneous textures. We derived the transformation of local spatial frequencies under perspective and radial distortion and formulated the estimation of camera parameters as a non-linear least-squares optimization problem. The experimental results show that with suitable textures, our algorithm performs similar to the state of the art in terms of the orders of magnitude of the reprojection error and the estimation error of the radial distortion coefficient. However, a direct comparison is not possible since previous approaches enforce other constraints on the image content and focus on different applications. Limitations of the method were identified in the estimation of pure radial distortion with no perspective projection involved.

References

[1] Joaquim Salvi, Xavier Armangué, and Joan Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognition*, vol. 35, no. 7, pp. 1617–1635, Jul. 2002. DOI: 10.1016/s0031-3203(01)00126-1.

[2] Zhengyou Zhang, "Camera calibration," in: *Emerging Topics in Computer Vision*, Gerard Medioni and Sing Bing Kang, Eds. Upper Saddle River, NJ: Prentice Hall, 2004, ISBN: 978-0-13-101366-7.

[3] Fabio Remondino and Clive Fraser, "Digital camera calibration methods: Considerations and comparisons," in: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, ISPRS, Ed., vol. 36, Dresden, Germany, 2006, pp. 266–272.

[4] Elijs Dima, Mårten Sjostrom, and Roger Olsson, "Assessment of multi-camera calibration algorithms for two-dimensional camera arrays relative to ground truth position and direction," in: *2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, Institute of Electrical and Electronics Engineers (IEEE), Jul. 2016. DOI: 10.1109/3dtv.2016.7548887.

[5] Roger Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, Aug. 1987. DOI: 10.1109/jra.1987.1087109.

[6] Zhengyou Zhang, "A flexible new technique for camera calibration," Microsoft Research, Technical Report MSR-TR-98-71, Dec. 2, 1998.

[7] Zhengyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. DOI: 10.1109/34.888718.

[8] Tobias Elbrandt and Jörn Ostermann, "Enabling accurate measurement of camera distortions using dynamic continuous-tone patterns," *Integrated Computer-Aided Engineering*, vol. 18, no. 1, pp. 3–14, 2011. DOI: 10.3233/ICA-2011-0355.

[9] Frédéric Devernay and Olivier Faugeras, "Straight lines have to be straight," *Machine Vision and Applications*, vol. 13, no. 1, pp. 14–24, Aug. 2001. DOI: 10.1007/pl00013269.

[10] Thorsten Thormählen and Hellward Broszio, "Automatic line-based estimation of radial lens distortion," *Integrated Computer-Aided Engineering*, vol. 12, no. 2, pp. 177–190, 2005.

[11] Zhengdong Zhang, Yasuyuki Matsushita, and Yi Ma, "Camera calibration with lens distortion from low-rank textures," in: *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, Jun. 2011. DOI: 10.1109/cvpr.2011.5995548.

[12] Toby Collins, Jean-Denis Durou, Pierre Gurdjos, *et al.*, "Singleview perspective shape-from-texture with focal length estimation: A piecewise affine approach," *3D data processing visualization and transmission (3DPVT10)*, 2010.

[13] Fabio Galasso and Joan Lasenby, "Fourier analysis and gabor filtering for texture analysis and local reconstruction of general shapes," in: *Proceedings of the 22nd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. DOI: 10.1109/CVPRW.2009.5206591.

[14] Eraldo Ribeiro and Edwin Robert Hancock, "Analysis of curved textured surfaces using local spectral distortion," in: *ICAPR*, 2001, pp. 407–416. DOI: 10.1007/3-540-44732-6_42.

[15] Boaz J. Super and Alan Conrad Bovik, "Shape from texture using local spectral moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 4, pp. 333–343, 1995. DOI: 10.1109/34.385983.

[16] Boaz J. Super and Alan Conrad Bovik, "Planar surface orientation from texture spatial frequencies," *Pattern Recognition*, vol. 28, no. 5, pp. 729–743, May 1995. DOI: 10.1016/0031-3203(94)00140-h.

[17] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[18] Alexander Eugen Conrady, "Decentred lens-systems," *Monthly notices of the royal astronomical society*, vol. 79, no. 5, pp. 384–390, 1919.

[19] Duane C. Brown, "Decentering distortion of lenses," *Photogrammetric Engineering*, vol. 32, no. 3, pp. 444–462, 1966.

[20] Fabio Galasso and Joan Lasenby, "Shape from texture: Fast estimation of planar surface orientation via fourier analysis," in: *British Machine Vision Conference*, 2007. DOI: `10.5244/C.21.71`.

[21] Eraldo Ribeiro and Edwin Robert Hancock, "Estimating the 3d orientation of texture planes using local spectral analysis," *Image and Vision Computing*, vol. 18, no. 8, pp. 619–631, 2000. DOI: `10.1016/S0262-8856(99)00064-5`.

[22] Kenneth Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[23] Donald W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[24] Phil Brodatz, *Textures: A photographic album for artists and designers*. Mineola, NY: Dover Publications, 1966.

[25] Gustaf Kylberg, "The Kylberg Texture Dataset v. 1.0," Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, Uppsala, External report (Blue series) 35, Sep. 2011. [Online]. Available: `http://www.cb.uu.se/~gustaf/texture/`.

[26] Jean-Yves Bouguet. (2015). Camera Calibration Toolbox for Matlab, [Online]. Available: `http://www.vision.caltech.edu/bouguetj/calib_doc` (visited on 03/03/2017).