Continuous Zoom Calibration by Tracking Salient Points in Endoscopic Video*

Miguel Lourenço[1], João P. Barreto[1,2], Fernando Fonseca[3], Hélder Ferreira[4], Rui M. Duarte[4] and Jorge Correia-Pinto[4]

[1] Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal
[2] Perceive 3D, Coimbra, Portugal
[3] Coimbra Hospital and Universitary Centre, Faculty of Medicine, Coimbra, Portugal
[4] Life and Health Sciences Research Institute, University of Minho, Braga, Portugal

Abstract. Many image-based systems for aiding the surgeon during minimally invasive surgery require the endoscopic camera to be calibrated at all times. This article proposes a method for accomplishing this goal whenever the camera has optical zoom and the focal length changes during the procedure. Our solution for online calibration builds on recent developments in tracking salient points using differential image alignment, is well suited for continuous operation, and makes no assumptions about the camera motion or scene rigidity. Experimental validation using both a phantom model and *in vivo* data shows that the method enables accurate estimation of focal length when the zoom varies, avoiding the need to explicitly recalibrate during surgery. To the best of our knowledge this the first work proposing a practical solution for online zoom calibration in the operation room.

1 Introduction

Minimally Invasive Surgery has a number of well documented benefits for the patient, such as faster recovery time, and less trauma to surrounding tissues. However, since the surgeon has limited access to the anatomical cavity and the visualisation is carried indirectly through the video acquired by an endoscopic camera, the execution of MIS is more difficult than the (equivalent) open-surgery. In this context, systems for CAS that process the endoscopic video can be very helpful in assisting the doctor during the procedure, either by improving the visualisation [8], or by recovering the camera motion [2].

Most image-based CAS systems that use the endoscopic video as primary sensory input require the intrinsic camera calibration to be known at all times during the procedure [2,8]. Endoscopic camera calibration in the context of CAS is challenging for three reasons [8, 12]: (i) since the optics are exchangeable and

* Miguel Lourenço and João Barreto want to thank *QREN-Mais Centro* by generous funding through *Novas Tecnologias para apoio à Saúde e Qualidade de Vida, Projecto A- Cirurgia e Diagnóstico Assistido por Computador Usando Imagem* and the Portuguese Science Foundation by funding through grant SFRH/BD/63118/2009.

the camera cannot be pre-calibrated, the calibration procedure must be carried in the operation room (OR) by a non-expert user [8], (ii) in the case of oblique-viewing endoscopes the surgeon often rotates the lens scope with respect to the camera head, which changes the calibration parameters [12], and (iii) high-end endoscopy systems provide optical zoom, which means that camera focal length changes during the intervention. Melo *et al.* [8] describe effective solutions for overcoming challenges (i) and (ii). They improve usability by proposing a fully automatic calibration method that uses as input a single image of a planar checkerboard pattern and, in the case of oblique viewing endoscopes, they show that it is possible to estimate the lens rotation and update the initial calibration by tracking the image boundary contour. This paper addresses challenge (iii) meaning that it is shown that under varying zoom the only parameter that changes significantly is the focal length, and that it is possible to update the initial calibration information without the need of re-calibrate the camera.

Zoom calibration is closely related to the problem of unknown/variable focal length estimation [10, 11]. Stoyanov *et al.* [11] propose a solution for stereo endoscopy where the focal lengths are directly estimated from the fundamental matrix [4]. Given the offline extrinsic stereo calibration, the focal lengths can be determined using only two point matches across the stereo pair. Unfortunately, the solution only generalizes for monocular endoscopy if the camera motion is known. Stewenius *et al.* [10] propose a solution for computing the relative camera pose and unknown focal length from 6 correspondences that is used within a sample consensus framework. The method assumes a rigid scene and requires in practice a considerable baseline between images, which makes its use problematic in continuous video. Related to this article is the work of Lee *et al.* [5] that does online estimation of focal length based on the image of the boundary contour of the endoscope. This approach has the disadvantage of requiring explicit camera calibration for multiple zoom positions and, more importantly, it does not work whenever the boundary contour is not visible in the image.

This article reports a solution for efficient and accurate focal length estimation in endoscopic video. We built on recent advances in tracking image features between frames with radial distortion [6] to show that it is possible to recover the focal length variation in sequences with zoom variation. Since we built on tracking theory, our approach is well suited for processing continuous monocular endoscopic video, does not make assumptions about camera motion [11] or scene rigidity [10], and does not require the boundary contour of the lens to be visible [5]. Quantitative and qualitative validation in synthetic and *in vivo* scenarios show that the proposed method enables accurate, online estimation of the focal length when the camera zoom changes.

2 Methods

This section details the proposed method for online focal length calibration. We start by introducing the adopted camera model before moving to the method description.

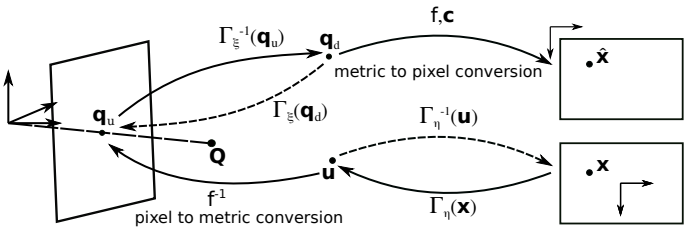


Fig. 1. Illustration of endoscopic camera modeling in the presence of radial distortion.

2.1 Endoscopic Camera Modeling

We assume that the radial distortion present in endoscopic cameras can be conveniently described using the so-called division model [3,8]. Let $\mathbf{q}_d$ be a generic 2D point with distortion. This point can be mapped in its undistorted counterpart $\mathbf{q}_u$ by the function $\mathbf{\Gamma}_\xi(\cdot)$

$$\mathbf{q}_u = \mathbf{\Gamma}_\xi(\mathbf{q}_d) = \left(1 + \xi \mathbf{q}_d{}^\mathsf{T} \mathbf{q}_d\right)^{-1} \cdot \mathbf{q}_d, \tag{1}$$

with ξ quantifying the amount of distortion.

Direct Projection Model and Single Image Calibration: Let $\mathbf{q}_u$ be the perspective projection of a 3D point $\mathbf{Q}$ in the canonical projective plane (see Fig.1). In the presence of distortion, and assuming the camera to be skewless and having unitary aspect ratio, point $\mathbf{q}_u$ is mapped into the point $\hat{\mathbf{x}}$ in the image plane by

$$\hat{\mathbf{x}} = f\, \mathbf{\Gamma}_\xi^{-1}(\mathbf{q}_u) \,+\, \mathbf{c}, \tag{2}$$

with $\mathbf{\Gamma}^{-1}(\cdot)$ being the inverse of Eq. 1 that maps $\mathbf{q}_u$ in its distorted counterpart

$$\mathbf{q}_d = \mathbf{\Gamma}_\xi^{-1}(\mathbf{q}_u) = 2\left(1 + \sqrt{1 - 4\xi \mathbf{q}_u{}^\mathsf{T} \mathbf{q}_u}\right)^{-1} \cdot \mathbf{q}_u, \tag{3}$$

f is the camera focal length that converts metric units into pixel units, and $\mathbf{c} = (c_x, c_y)$ is the principal point in pixels. With the single image calibration of [8] we can easily estimate ξ, f and $\mathbf{c}$ at an initial reference zoom position. Remark that ξ is the amount of distortion in metric units that is a characteristic of the lens and therefore independent of the zoom variation.

Modeling Radial Distortion in the Image Plane: An alternative way of modelling the projection is to consider that the radial distortion acts in the image plane as opposed to act in the metric projective plane. From the inversion of Eq. 2 it comes in a straightforward manner that

$$\mathbf{q}_u \,=\, \mathbf{\Gamma}_\xi(f^{-1}(\hat{\mathbf{x}} \,-\, \mathbf{c})). \tag{4}$$

For simplicity, let's assume that $\mathbf{x} = \hat{\mathbf{x}} - \mathbf{c}$, which means that image points are expressed in a coordinate frame centred in the principal point. Replacing $\mathbf{\Gamma}_\xi$ by the expression of Eq. 1 it comes that

$$f \cdot \mathbf{q}_u = (1 + \frac{\xi}{f^2}\mathbf{x}^\mathsf{T}\mathbf{x})^{-1} \cdot \mathbf{x}. \tag{5}$$

Let $\mathbf{u} = f \cdot \mathbf{q}_u$ be the undistorted image point in pixel units. From the equation above it follows that $\mathbf{u}$ is related with its distorted version $\mathbf{x}$ by $\mathbf{u} = \mathbf{\Gamma}_\eta(\mathbf{x})$ with

$$\eta = \xi \cdot f^{-2} \tag{6}$$

being the parameter that quantifies the distortion in pixel units. We conclude that, if the radial distortion is expressed in metric units, i.e. before the intrinsics, the corresponding parameter ξ does not depend of the camera focal length. However, if we quantify this same distortion in pixel units using η, then there is a dependence on the focal length which means that the distortion parameter varies with the zoom. We will use the relation of Eq. 6 for recovering the focal length f at each frame by combining offline calibration of the constant parameter ξ using [8] with online estimation of η using the tracking framework of [6].

2.2 Estimating Image Distortion at every frame using uRD-KLT

Lourenço and Barreto show in [6] that it is possible to estimate the radial distortion in the image plane by tracking feature points between adjacent frames. Their uncalibrated KLT algorithm for images with radial distortion (uRD-KLT) starts by extracting reference templates $\mathsf{T}(\mathbf{x})$ around a set of salient points $\mathbf{x}$ that are detected based on image derivatives [6]. Given an incoming image $\mathsf{I}(\mathbf{x})$, the goal is to align the templates $\mathsf{T}(\mathbf{x})$ with the corresponding image regions subject to the squared intensity difference and under the assumption of a 2D deformation model $\mathbf{v}(\mathbf{x};\mathbf{p})$, with parameters $\mathbf{p}$, that accounts for both the local motion $\mathbf{w}$ and the global effect of distortion. The deformation model is given by:

$$\mathbf{v}(\mathbf{x};\mathbf{p}) = \Big(\mathbf{\Gamma}^{-1} \circ \mathbf{w} \circ \mathbf{\Gamma}\Big)(\mathbf{x};\mathbf{p}), \tag{7}$$

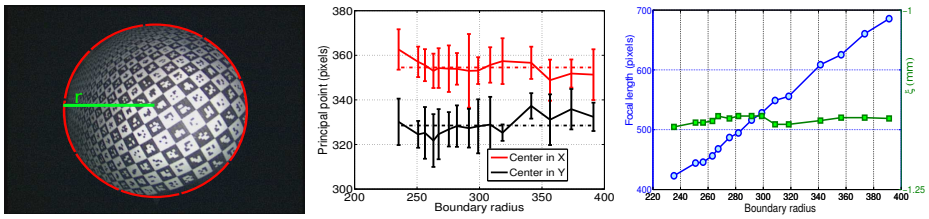
with $\mathbf{p} = (\mathbf{m}, \eta)$ where $\mathbf{m}$ is the vector of motion parameters that describes the local deformation undergone by each image patch in the absence of distortion [1], and η is the global distortion parameter that is common to all image regions.

Given an initial estimate of $\mathbf{p}$ the goal is to iteratively compute the updates $\delta\mathbf{p}$ of the warp parameters by minimizing the following cost function

$$\epsilon = \sum_{\mathbf{x} \in \mathcal{N}} \Big[\mathtt{I}(\mathbf{v}(\mathbf{x};\mathbf{p})) - \mathtt{T}(\mathbf{v}(\mathbf{x};\delta\mathbf{p}))\Big]^2 \tag{8}$$

This error function can linearised with respect to $\mathbf{p}$ by computing the first order Taylor expansion, and the final updates $\delta\mathbf{p}$ can be computed in closed-form as:

$$\delta\mathbf{p} = \mathcal{H}^{-1} \sum_{\mathbf{x}\in\mathcal{N}} \Big[\nabla\mathtt{T}\frac{\partial\mathbf{v}(\mathbf{x};\mathbf{0})}{\partial\mathbf{p}}\Big]^\mathsf{T} \Big(\mathtt{I}(\mathbf{v}(\mathbf{x};\mathbf{p})) - \mathtt{T}(\mathbf{x})\Big), \tag{9}$$



(a) Calibration image (b) Center of Coordinates (c) Focal length

Fig. 2. Intrinsic parameters for different zoom positions. Fig. 2(a) shows a calibration image where the radius of the boundary is used to index the current zoom position. Fig. 2(b) shows the variation of the center of coordinates and Fig. 2(c) shows the variation of the focal length (blue) and distortion in metric units (green) for increasing zoom. The experiment confirms that the focal length increases, while the principal point $\mathbf{c}$ and the distortion parameter ξ are virtually constant ($-\xi = 1.1515 \pm 0.007$).

with $\mathcal{H}$ being a 1st order approximation of the Hessian matrix, and $\partial \mathbf{v}(\mathbf{x}; \mathbf{0})/\partial \mathbf{p}$ being the Jacobian of the warp evaluated at the identity warp [1,6]. Since the η is a global parameter common to every image point, the corresponding distortion updates are computed using all tracked features, while the feature local motion $\mathbf{m}$ is computed for each feature separately [6].

2.3 Calibrating zoom by image alignment

So far, we have derived the relation between distortion parameters in metric and pixel units and showed that the distortion in pixels can be estimated at each time instant using the uRD-KLT. While in [6] it is assumed that the camera calibration is not known and that the principal point $\mathbf{c}$ is coincident with the image center, we use the single image calibration [8] at an initial zoom position to obtain the principal point $\mathbf{c}$ and the lens distortion ξ in metric units. The uRD-KLT is applied during operation to continuously estimate the image distortion parameter η and the focal length is estimated at each frame time instant using the relation of Eq. 6. The approach works as far as $\mathbf{c}$ and ξ remain constant.

3 Results

In this section we evaluate the proposed solution for recovering the focal length in continuous video. We start by conducting a set of experiments with ground truth to validate the assumptions made for the derivation of our solution. Afterwards, the method is validated in both a synthetic environment and in a *in vivo* sequence acquired in a porcine uterus.

3.1 Variation of Intrinsic Camera Parameters with Zoom Changes

In this experiment we used a Storz H3-Z endoscopy system with a Dyonics' arthroscopic lens with 4mm diameter. We placed the camera zoom in 15 distinct

(a) Phantom model (b) Focal length estimation

Fig. 3. Simulation experiment with a zoom only sequence. Fig. 3(a) shows the two phantom images with the corresponding boundary radius at different zoom positions. Fig. 3(b) shows that the focal length estimation of the uRD-KLT is accurate.

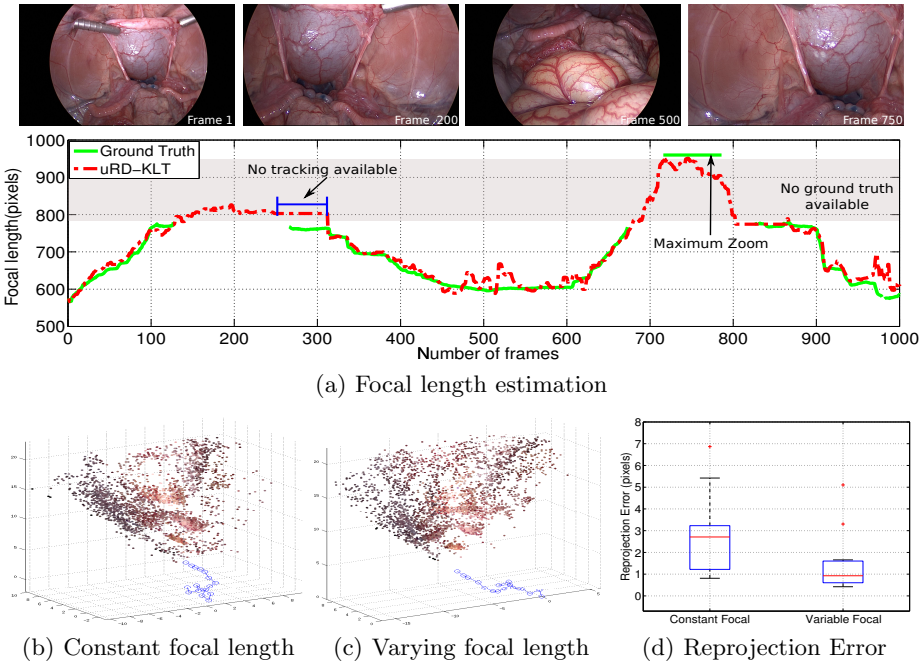
positions and, for each position, we collected 5 images of a checkerboard pattern that were used to obtain 5 independent intrinsic calibrations using the method described in [8]. Figure 2(b) shows the principal point estimation for successive zoom positions that are referenced using the radius of the boundary contour. Fig. 2(c) does the same for the focal length f and the lens distortion parameters ξ. The assumption that $\mathbf{c}$ and ξ are kept constant while the zoom varies holds in practice.

3.2 Validation with a phantom model

This experiment uses the camera setup of section 3.1 for acquiring a video sequence of a phantom model of the knee. The endoscope is kept stationary, while the zoom is increased. The focal length is estimated at each frame time instant by using the uRD-KLT to track 20 automatically detected points. The focal length estimates are related with the calibration results of section 3.1 using the radius of the boundary contour like in [5]. Figure 3 compares the on-line estimation results with the calibration ground truth. Please note that high-zoom values have no ground truth because the boundary contour is not visible and there is no manner of relating the f estimates with calibration results. Nevertheless, the estimation seems to be plausible and consistent with the calibration obtained for the end zoom position. The maximum relative estimation error was 2.5% for the maximum zoom position when the image distortion η reaches its minimum.

3.3 Validation in *in-vivo* data

The data used in this experiment was recorded in a *in-vivo* porcine uterus during a robotic assisted procedure. The sequence of 1000 frames with resolution 1920×1080 was acquired at 30Hz with a Storz H3-Z camera system equipped with a laparoscopic lens of 10 mm from Dyonics. We used the procedure of section 3.1 to obtain calibration ground truth. The surgeon was asked to vary the zoom against the direction of motion of the endoscope in an attempt to keep the size of the image structures constant and evaluate the robustness to changes



(a) Focal length estimation

(b) Constant focal length (c) Varying focal length (d) Reprojection Error

Fig. 4. Results in *in-vivo* data. (a) shows some sample video frames, and the focal length estimation. (b) and (c) show a visual odometry experiment without and with focal length compensation, respectively. Compensating the focal length bring clear benefits for visual odometry, as it can be seen in Fig. 4(d) where the reprojection error of the reconstructed 3D points decreases from ≈ 3 pixels to less than 1 pixel.

in scale. Figures 4(a) shows the online estimation results for the focal length by performing uRD-KLT tracking in the *in-vivo* sequence. These results were obtained with a straightforward Matlab implementation that ran at 2fps on a single core of an Intel i7-3630QM CPU @ 2.40GHz processor. It can be observed that the uRD-KLT-based estimation is quite accurate with an average relative error of $2.1927 \pm 2.3959\,\%$ when compared with the calibration ground truth. Please note that there are sequence segments for which there are no salient points (frames 250 to 300) or the accuracy of the estimation decreases due to temporary poor tracking (frames 475-525). However, and since the focal length measurement is carried in a frame-by-frame basis, these errors do not accumulate. Finally Fig. 4(b) to 4(d) show comparative visual odometry results for a sub-sequence of 17 frames where the camera moves forward while the zoom decreases. Since most of the scene is rigid the camera motion is computed by applying the five-point algorithm [9] using image correspondences obtained with sRD-SIFT [7]. Fig. 4(b) and 4(d) depict the motion estimation results when the focal length is kept constant and when the focal length is updated.

4 Discussion and Conclusion

This article presents a practical solution for keeping the camera calibrated when the camera zoom changes during operation. The method builds on recent developments in image alignment for tracking keypoints in video with radial distortion and, since there are no distortion free endoscopic cameras, it can be virtually used in any MIS. The approach was validated in both synthetic and *in-vivo* data, showing that is possible to keep the camera calibrated under zoom variations without the need to re-calibrate the camera during operation. To the best of our knowledge, this is the first work proposing an effective solution for the zoom calibration in continuous medical endoscopic video.

References

1. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. International Journal of Computer Vision 56(3), 221–255 (2004)
2. Burschka, D., Li, M., Taylor, R., Hager, G.: Scale-Invariant Registratiou of Monocular Endoscopic Images to CT-Scans for Sinus Surgery. In: Barillot, C., Haynor, D., Hellier, P. (eds.) MICCAI, LNCS, vol. 3217, pp. 413–421. Springer Berlin Heidelberg (2004)
3. Fitzgibbon, A.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: IEEE Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 125–132 (2001)
4. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edn. (2004)
5. Lee, T.Y., *et al.*: Automatic distortion correction of endoscopic images captured with wide-angle zoom lens. IEEE Transactions on Biomedical Engineering 60(9), 2603–2613 (Sept 2013)
6. Lourenço, M., Barreto, J.a.P.: Tracking Feature Points in Uncalibrated Images with Radial Distortion. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV, LNCS, vol. 7575, pp. 1–14. Springer Berlin Heidelberg (2012)
7. Lourenco, M., Barreto, J., Vasconcelos, F.: sRD-SIFT: Keypoint Detection and Matching in Images With Radial Distortion. IEEE Transactions on Robotics 28(3), 752–760 (June 2012)
8. Melo, R., Barreto, J., Falcao, G.: A New Solution for Camera Calibration and Real-Time Image Distortion Correction in Medical Endoscopy-Initial Technical Evaluation. IEEE Transactions on Biomedical Engineering 59(3), 634–644 (2012)
9. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(6), 756–770 (June 2004)
10. Stewenius, H., Nister, D., Kahl, F., Schaffalitzky, F.: A minimal solution for relative pose with unknown focal length. In: IEEE Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 789–794 (June 2005)
11. Stoyanov, D., Darzi, A., Yang, G.Z.: Laparoscope Self-calibration for Robotic Assisted Minimally Invasive Surgery. In: Duncan, J., Gerig, G. (eds.) MICCAI, LNCS, vol. 3750, pp. 114–121. Springer Berlin Heidelberg (2005)
12. Yamaguchi, T., *et al.*: Camera Model and Calibration Procedure for Oblique-Viewing Endoscope. In: Ellis, R., Peters, T. (eds.) MICCAI, LNCS, vol. 2879, pp. 373–381. Springer Berlin Heidelberg (2003)