A NOVEL TECHNIQUE TO MODEL THE VARIATION OF THE INTRINSIC PARAMETERS OF AN AUTOMATIC ZOOM CAMERA USING ADAPTIVE DELAUNAY MESHES OVER MOVING LEAST-SQUARES SURFACES

Michel Sarkis, Christian T. Senft and Klaus Diepold

Technische Universität München, Institute for Data Processing, Arcisstr. 21, 80290, Munich, Germany

Emails: michel@tum.de, lausbub@mytum.de, kldi@tum.de

ABSTRACT

The accuracy of computer vision systems is highly dependent on the correct estimates of the camera intrinsic parameters. This accuracy is important in numerous applications like telepresence and robot navigation. In this work, a novel technique is proposed to model the variation of the camera's intrinsic parameters as a function of the focus and the zoom. The proposed method computes the complete surfaces of the intrinsic parameters from a predefined number of focus/zoom measurements using a moving least-squares (MLS) regression technique. Then, it approximates the generated MLS surfaces by employing adaptive Delaunay meshes. Compared to a previous technique using bivariate polynomial functions, the new method results in a 94% enhancement of the mean estimation error. In addition, the new method leads to the same accuracy of the results as compared to a previous version of the MLS technique while requiring a less amount of computations.

Index Terms**—** Machine vision, lenses, modeling, optical distortion

1. INTRODUCTION

The usage of automatic motorized zoom lenses in camera systems has become significantly important in the field of computer vision. This is due to their flexibility and controllability, as compared to mono-focal lenses. By varying the focal length and the aperture values, a zoom camera system can be adjusted to different fields of view, depth of fields and lighting conditions. This flexibility is the main reason why zoom lenses are increasingly being adopted in applications like 3D scene depth reconstruction, visual tracking, telepresence and robot navigation [1–3].

The challenge of the employment of zoom lens camera systems lies in modeling the process of image formation as the lens parameters focus, zoom and aperture are varied in a continuous manner. This image-formation process describes the relationship between an existing point $P_w = (x_w, y_w, z_w)^T$ in the real world coordinate system $\mathbb{R}^3$ and its corresponding image point $P_i = (u_i, v_i)^T$ in the image coordinate system $\mathbb{R}^2$. Due to the imperfectness of a camera lens system, errors are introduced. These error are known as the radial and tangential distortion [4, 5]. To attain the flexibility required in computer vision applications, a zoom lens camera system has to be calibrated at a number of lens settings and the values are then saved in a look-up table. Nevertheless, this process is difficult to accomplish due to the numerous measurements that have to be done.

This research is sponsored by the German Research Foundation (DFG) as a part of the SFB 453 project, High-Fidelity Telepresence and Teleaction.

One way to deal with this limitation by using auto-calibration algorithms. These methods are either based on finding the image of the absolute conic, which is used to compute a transformation from the projective frame to the (calibrated) metric frame, or on solving the Kruppa equations which represent an algebraic representation of the correspondence of the epipolar lines tangent to the absolute conic [2]. However, these techniques are not very reliable since they are highly susceptible to errors. This is due to the inaccuracy in the obtained results especially the ones involved in finding the absolute conic in the projective form [2]. Another problem is the fact the a lot of constraints have to be imposed on the camera system which makes its application in scenarios like telepresence limited.

Another way is to treat the zoom lens camera system as an input/output function. The inputs in this case are the focus and the zoom while the outputs are the camera's intrinsic parameters. The main advantage over the auto-calibration techniques is the fact that there is no constraints imposed on the used camera system. Such an analysis has previously been conducted by [6] using bivariate polynomial functions; however, the results are not satisfactory due to the introduced errors. Recently, a new algorithm based on the MLS technique has been proposed in [7]. This algorithm operates by generating a local function at each focus/zoom setting point and leads to accurate results. The drawback of this method is its excessive computation if it has to be used in an application that requires low delay response, e.g. telepresence, when compared to [6].

In this work, a new technique based on the moving least-squares (MLS) method of [7] is implemented to model the intrinsic parameters of a zoom lens camera system. MLS is used to generate the surfaces of the camera's intrinsic parameters as a function of focus and zoom. Then, an adaptive Delaunay mesh is applied to approximate these surfaces. The obtained meshes can then be stored and used to estimate the intrinsic parameters in any application that requires a variable focal length. As will be seen in the results, the proposed method retains the accuracy of MLS while keeping the amount of computation low.

Section 2 defines the assumptions made in this work. Section 3 illustrates a review of some standard techniques of modeling the camera intrinsic parameters. Section 4 presents the proposed technique. Section 5 shows an analysis and comparison of the proposed method with previous techniques. Finally, conclusions are drawn in Section 6.

2. ASSUMPTIONS MADE

What is studied in this type of analysis is the function that describes the relation between the focus and the zoom settings of a zoom lens

1-4244-1437-7/07/$20.00 ©2007 IEEE

camera and the corresponding variation of the intrinsic and the distortion parameters. The intrinsic parameters are expressed by the focal length, the coordinates of the camera center point and the skew parameter while the distortion terms are described by the coefficients of their Taylor series expansion [2, 6, 8]. To facilitate the modeling operation, some assumptions have to be made.

- **Aperture Setting:** In the current camera's technology, the aperture setting is supposed to have a negligible influence on the intrinsic parameters [8]. Therefore, the lens parameter aperture is not considered in this analysis.
- **Skew Parameter:** In modern camera systems, the pixel angle can always be expected to be 90° [2]. Thus, the skew parameter can be safely neglected.
- **Lens Distortion Coefficients:** The imperfectness of lens systems compared to a pinhole camera model leads to non-linear distortion. In [4], an expression that corrects the distortion of mono-focal lens systems was proposed as:

$$\begin{aligned} u_i &= u_d + \bar{u}\left(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \cdots\right) \\ &\quad + \left[\tau_1\left(r^2 + 2\bar{x}^2\right) + 2\tau_2 \bar{x}\bar{y}\right]\left(1 + \tau_3 r^2 + \cdots\right), \quad (1) \\ v_i &= v_d + \bar{v}\left(\kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \cdots\right) \\ &\quad + \left[2\tau_1 \bar{x}\bar{y} + \tau_2\left(r^2 + 2\bar{y}^2\right)\right]\left(1 + \tau_3 r^2 + \cdots\right), \end{aligned}$$

where

$$\begin{aligned} \bar{u} &= u_d - u_0, \\ \bar{v} &= v_d - v_0, \\ r &= \sqrt{\bar{u}^2 + \bar{v}^2}. \end{aligned} \quad (2)$$

u_d and v_d represent the measured (lens distorted) pixel coordinates; u_i and v_i are the corrected (ideal) pixel coordinates of the world point P_w; u_0 and v_0 describe the location of the principal point P_0; r is the radial distance between $P_d = (u_d, v_d)^T$ and $P_0 = (u_0, v_0)^T$; κ_1, κ_2, ... are the coefficients of the radial distortion and τ_1, τ_2, ... are the coefficients of the tangential distortion.

The radial distortion is mostly dominated by the leading term of the power series expansion $\kappa_1 r^2$, whereas higher order terms $\kappa_2 r^4$, $\kappa_3 r^6$, ... are barely significant in today's commercial lenses. It can also be shown that the higher order terms r^4, r^6, ... result in a numerical instability in (1) [8]. Furthermore, the insignificance of the tangential distortion component in today's lens systems is mentioned in [4]. Hence, only the first radial distortion coefficient κ is considered.

3. EXISTING TECHNIQUES

Several techniques are available in the literature which address the problem of modeling the camera's intrinsic parameters. The simplest method is to measure the intrinsic parameters using a standard offline calibration technique and then store the data in a look-up table [9]. Unfortunately, this leads to a waste in the memory resources. This is justified since for a single focus/zoom setting it is necessary to store at least 7 values: the focus, the zoom, the focal length in u- and v-direction of the pixel axes, the two coordinates of the principal point and one radial lens distortion coefficient. By taking the motorized zoom lens system used in this work, it is possible to choose between 700 different settings for the zoom and 600 different settings for the focus. This results in 700×600×7 = 2.940.000 values that have to be stored.

In [6], the intrinsic parameters are computed by finding a bivariate polynomial function, empirically estimated, that fits the data points by the minimization of an error function. The advantage of this method is its very efficient memory consumption since only the coefficients of the polynomial functions need to be stored, instead of requiring every single value of the intrinsic parameters for each lens setting. The general equation that defines a bivariate polynomial function of degree m is given by:

$$f(s, o) = \sum_{i=0}^{m} \sum_{j=0}^{m-i} a_{ij} s^i o^j, \quad (3)$$

where s, o represent a focus and a zoom setting and k reflects the number of coefficients defined as

$$k = \frac{(m+1)(m+2)}{2}. \quad (4)$$

To approximate the intrinsic parameters of a zoom lens camera system in [6], a bivariate polynomial function of degree $m = 2$ was chosen for the coefficient of the radial distortion κ. The components of the focal length (f_u, f_v) and the coordinates of the principal point (u_0, v_0) were approximated with a function of degree $m = 5$.

4. THE PROPOSED TECHNIQUE

Given is a set of measured focus/zoom inputs and the corresponding intrinsic parameters at these points. Let $\mathbf{x}_i = \{s_i, o_i\}_{i \in I}$ be the set of these distinct data points in $\mathbb{R}^3$, and let $\{f(\mathbf{x}_i)\}_{i \in I}$ be the intrinsic parameter values at these points, i.e. $f(s_i, o_i) = z_i$. The moving least-squares approximation of degree m at $\mathbf{x} \in \mathbb{R}^3$, as described in [7], is the value $\tilde{p}(\mathbf{x})$ where $\tilde{p} \in \Pi_m^3$ minimizes, among all $p \in \Pi_m^3$, the weighted least-squares error function defined by

$$\sum_{i \in I} \theta\left(d_i\right)\left[p(\mathbf{x}_i) - f(\mathbf{x}_i)\right]^2. \quad (5)$$

θ is a non-negative weight function for each neighboring point $\mathbf{x}_i$ of $\mathbf{x}$, $d = \left\|\mathbf{x} - \mathbf{x}_i\right\|$ is the corresponding Euclidean distance in $\mathbb{R}^3$ and $p \in \Pi_m^3$ is the space of polynomials of degree m in $\mathbb{R}^3$. The approximation of the data introduced by (5) is defined to be *local*, if the weight function $\theta(d_i)$ is rapidly decreasing as $d_i \to \infty$, or is of finite support. Similarly, (5) is defined to be *interpolable* if $\lim_{d_i \to \infty} \theta(d_i) = 0$ [10]. A weighting function that satisfies these conditions is

$$\theta(d_i) = \exp\left(\frac{-{d_i}^2}{h^2}\right), \quad (6)$$

where h is a prescribed real constant.

The MLS approach is an appropriate technique to the problem of modeling the intrinsic parameters since it uses a local regression scheme. At each point setting $\mathbf{x}$, i.e. at each combination of focus and zoom, it only uses data values within a sphere of radius h around $\mathbf{x}$. In addition, the weighting function in (6) forces the weights of the points that are outside of this sphere to vanish. Thus, it is possible to compute a regression surface by only using the set of points within a distance around $\mathbf{x}$ less than h.

MLS results in a good approximation of the intrinsic parameters since the obtained model of a specific intrinsic parameter is the concatenation of several local models. For more information about the MLS zoom lens camera modeling technique, refer to [7].

The main drawback of the MLS method is its requirement of a large amount of computations. To estimate the intrinsic parameters at each focus/zoom setting point $\mathbf{x}$, the k nearest neighbors have to be determined to calculate the Euclidean distances. Then, the computed distances are weighted using (6), followed by a minimization of (5). These extra computations introduce a noticeable delay which cannot be tolerated if MLS is used in an application where the delay is critical, e.g. in telepresence [3].

To overcome this deficiency, the MLS generated surfaces will be approximated with a Delaunay triangulation. The motivation behind this idea is that MLS has the ability to interpolate new points from the measured focus/zoom settings. Therefore, if the number of generated points is large enough, i.e. the sampling rate is high, a mesh approximation can be performed without a significant loss in the accuracy [11]. To compensate for the discrete type of the generated surfaces, one can take advantage of the constructed meshes to compute the intrinsic parameters at the focus/zoom setting points which are not generated with MLS.

To estimate the intrinsic values in such cases, one has to search for the triangle that encloses a focus/zoom setting $\mathbf{x}$ and use the vertices of the triangle to interpolate the new value. Suppose that $\mathbf{v}_i$, where $i = 1 \ldots 3$, are the vertices of the triangle T that encloses $\mathbf{x}$. The intrinsic value $\hat{z}$ of $\mathbf{x}$ is computed as the weighted sum of the intrinsic values of $\mathbf{v}_i$

$$\hat{z} = \sum_{i=1}^{3} \omega_i \cdot z(\mathbf{v}_i), \tag{7}$$

where $z(\mathbf{v}_i)$ is the intrinsic value of the vertex $\mathbf{v}_i$ and ω_i is the weight assigned to $\mathbf{v}_i$. The weights ω_i in this case are computed by determining the barycentric coordinates of $\mathbf{x}$ in T [11].

The proposed scheme used to estimate the intrinsic parameters' surfaces from a small set of measured focus/zoom setting points is shown in Table 1. The offline version is used to generate the surfaces of the intrinsic parameters, i.e. to interpolate new points. The online version is the one to be employed in applications that require a camera with a variable focal length.

Table 1. The Proposed Algorithm

For every focus/zoom setting $\mathbf{x}$:

Offline:

1- Calculate the Euclidean distances between a chosen setting $\mathbf{x}$ and all measured data points $\mathbf{x}_i$.

2- Determine at least k nearest neighbors of $\mathbf{x}$ where k is obtained using Equation (4).

3- Weight the selected points with Equation (6).

4- Calculate the MLS regression surfaces with these weighted points by minimizing Equation (5).

5- Approximate the obtained surfaces using Delaunay triangulation and save the meshes.

Online:

1- Search for the position of the setting $\mathbf{x}$ in the mesh.

2- **If** $\mathbf{x}$ exists, fetch the intrinsic parameters from the MLS surfaces.
Else, locate the triangle T that encloses $\mathbf{x}$ and estimate the intrinsic parameters using Equation (7).

5. EXPERIMENTAL RESULTS

The lens camera system used in the experiments is a PROSILICA (EC 1280C) CCD camera with a motorized PENTAX zoom lens (C6Z1218M3). The pixel size, as specified in the manual, is 6.7 μm in both directions of the pixel. The skew, defined as the angle of the pixel axes, has been estimated to $90^\circ \pm 0.01$. Thus, for an angle close to 90°, the skew factor is set to 0. As discussed in Section 2, the aperture setting is set to a constant value of $F = 5.6$. The lens system's zoom range is between 12.5 mm - 75 mm and that of the focus between ∞ - $1.2m$ respectively.

In this work, the zoom setting in motor units was varied between $0 - 700$ in steps of 100. This is equivalent to a change in focal length between 12.5 mm - 65 mm. For the focus, a range of 150 - 600 motor units with a step size of 50 was used, which roughly corresponds to the focused distance of the lens system. To avoid a hysteresis problem due to backlash, as mentioned in [6], the lens motor was driven to the desired setting by starting with a smaller value for both, focus and zoom.

The measured set of data points, consisting of a total of 80 focus/zoom settings, is used to generate the models. For each lens setting, 40 images of a calibration grid with 64 points for calibration on it were taken from different fields of view. Due to the wide range of focal length, 6 checkerboards, different in size, were used for the calibration procedure. The calibration is done separately for each focus/zoom setting by treating each one as a single mono-focal lens. To determine the intrinsic lens parameters, the Camera Calibration Toolbox for MATLAB was used [12]. The computed parameters are then considered to as the ground truth data of the modeling process.

As a measure of the estimated model's accuracy, the Undistorted Image Plane Error (UIPE) defined in [6] is used to represent the reprojection error. The UIPE is given as

$$\text{UIPE} = \sqrt{(u_d - u_i)^2 + (v_d - v_i)^2}, \tag{8}$$

where (u_d, u_d) are the measured coordinates of the given world point P_w mapped to the image point P_d; (u_i, v_i) are the estimated coordinates of an image point P_i, determined by the camera's intrinsic parameter model. A value UIPE $= 0$ means that the camera model has totally corrected the reprojection error.

To make the proposed measure invariant to the number of the data points, it is also suggested to use the Mean UIPE (M_UIPE)

$$\text{M\_UIPE} = \frac{1}{n} \sum_{i,d=1}^{n} \text{UIPE}, \tag{9}$$

where n is the number of calibration points in all calibration grids for one single focus/zoom setting.

In these tests, the proposed algorithm is compared to the MLS method in [7] and the bivariate polynomial approach with the global regression scheme in [6].

These techniques are applied to the measured ground truth data to determine the intrinsic parameter models. The performance of the algorithms is measured by computing the M_UIPE, as in (9). The obtained results are illustrated in Fig. 1. As can directly be noticed, the proposed method is able to model the variation of the intrinsic parameters in a very good manner. The error in the pixels' reprojection error is of the same order as compared to the look-up table (ground truth) and the MLS technique in [7].

The values represented by these plots are better reflected in Table 2, where the MM_UIPE is the mean of M_UIPE over all ground truth points, MAX_UIPE and STD_UIPE are the maximum and the standard deviation of the UIPE. The RMSE value represents the Root Mean Squared Error of the reprojection defined as

$$\text{RMSE} = \sqrt{\frac{1}{l \times n} \sum_{i,d=1}^{l \times n} (u_d - u_i)^2 + (v_d - v_i)^2}, \tag{10}$$

where l is the total number of focus/zoom settings.

Comparing the proposed technique to Willson's method, i.e. [6], it can be seen that the pixel RMSE has decreased tremendously. The RMSE decreased from 23.31 to only 0.14 which gives an improvement of 94%.

The main reason for this result is that the proposed technique has a better ability of modeling the surfaces as previously mentioned due to the employment of MLS. This can be seen by looking at Figures 2 and 3a which present the generated models of the coordinate of the camera center u_0. Due to the scattered nature of the camera center, the global regression scheme cannot compute a good model of the intrinsic parameter. This can be also checked by the results in [6, 7].

The proposed method leads to the same accuracy as the original MLS method in [7] as shown in Table 2. Nevertheless, it can be noticed from Fig. 2 that the interpolated surface of the proposed technique is not as smooth as the original method due to the mesh approximation. However, by increasing the sampling rate at which the MLS curves are estimated, it is possible to attain smoother surfaces at the cost of storing larger meshes.

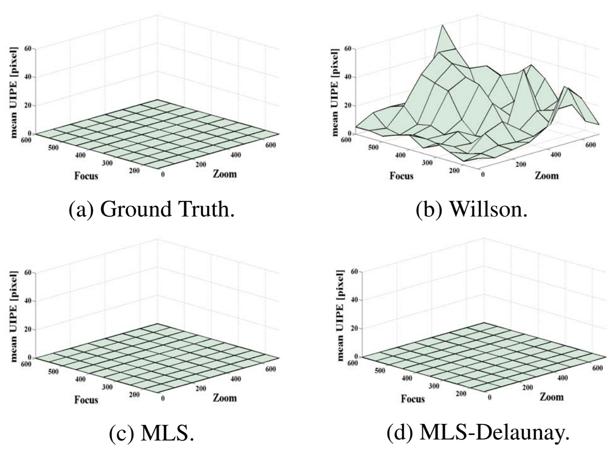
Finally, the hold-out cross validation test is applied to the three techniques by taking the RMSE defined in (10) as a criterion while varying the number of samples from 10 to 79. The results illustrated in Figure 3b clearly show that the mesh approximation do not deteriorate the accuracy of the MLS interpolation. Compared to Willson's technique, the improvement is easily noticed.

Table 2. Comparison of the reprojection error of the different intrinsic zoom lens parameter modeling techniques.

	RMSE	MM_UIPE	MAX_UIPE	STD_UIPE
LOOK-UP TABLE	0.14	0.08	7.91	0.12
MLS	0.14	0.08	7.91	0.12
MLS-DELAUNAY	0.14	0.08	7.91	0.12
WILLSON	23.31	19.4	75.03	14.00

6. CONCLUSION

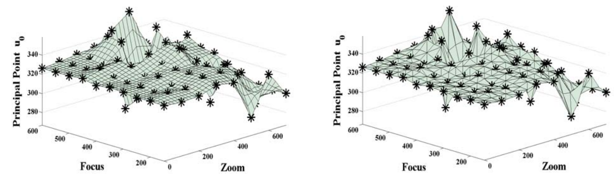
A new technique is proposed based on MLS, to model the variation of the camera's internal parameters as a function of the focus and the zoom. Compared to the previous version of MLS, the proposed approach is able to retain the accuracy of the intrinsic parameters' estimates of a zoom lens camera while minimizing the amount of computations. This is achieved since the proposed method approximates the MLS generated surfaces with a Delaunay mesh which makes the estimation of the intrinsic parameters simple.



(a) Ground Truth. (b) Willson.

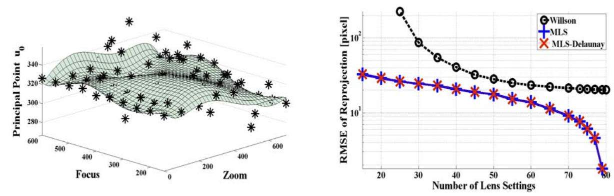
(c) MLS. (d) MLS-Delaunay.

Fig. 1. Plot of the M_UIPE of the different techniques. In (a), the ground truth data. In (b), the method of [6]. In (c), the original MLS method of [7]. In (d), the proposed method.



(a) MLS. (b) MLS-Delaunay.

Fig. 2. Plot of the model for the coordinate of the camera center u_0 obtained from the original MLS method of [7] in (a) and the proposed method in (b) along with the measured settings.



(a) Willson. (b) Cross Validation.

Fig. 3. (a): Plot of the model for the coordinate of the camera center u_0 obtained from the method of [6] along with the measured settings. (b): Cross validation test of all the methods. At each lens setting, the test is repeated 100 times and the average is plotted.

7. REFERENCES

[1] E. Hayman, *The use of zoom within active vision*, Ph.D. thesis, Robotics Research Group, Univ. of Oxford, Sep. 2000.

[2] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge Univ. Press, 2nd edition, Mar. 2004.

[3] J. Mulligan, X. Zabulis, N. Kelshikar, and K. Daniilidis, "Stereo-based environment scanning for immersive telepresence," *IEEE T. Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 304–320, Mar. 2004.

[4] D.C. Brown, "Close-range camera calibration," *Photogram. Eng.*, vol. 37, no. 8, pp. 855 – 866, 1971.

[5] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE T. Pat. Anal. and Mach. Intell.*, vol. 14, no. 10, pp. 965 – 980, Oct. 1992.

[6] R.G. Willson, *Modeling and calibration of automated zoom lenses*, Ph.D. thesis, The Robotics Institute, Carnegie Mellon Univ., Jan. 1994.

[7] M. Sarkis, C. T. Senft, and K. Diepold, "Modeling the variation of the intrinsic parameters of an automatic zoom camera system using moving least-squares," in *IEEE Conf. Automation Science and Engineering*, Sept. 2007.

[8] M. Li and J.-M. Lavest, "Some aspects of zoom lens camera calibration," *IEEE T. Pat. Anal. and Mach. Intell.*, vol. 18, no. 11, pp. 1105 – 1110, Nov. 1996.

[9] K. Tarabanis, R.Y. Tsai, and D.S. Goodman, "Modeling of a computer-controlled zoom lens," *IEEE Int. Conf. Robotics and Automation*, vol. 2, pp. 1545 – 1551, May 1992.

[10] D. Levin, "The approximation power of moving least-squares," *Math. of Computation*, vol. 67, no. 224, pp. 1517–1531, 1998.

[11] F. Nielsen, *Visual Computing: Geometry, Graphics, and Vision*, Charles River Media, 1st edition, 2005.

[12] J.-Y. Bouguet, "Camera calibration toolbox for MATLAB," http://www.vision.caltech.edu/bouguetj/calib_doc/index.html.