Confidence-based camera calibration with modified census transform

Qicong Dong, Lei Wang & Jieqing Feng

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501
Volume 79
Combined 31-32

Multimed Tools Appl (2020)
79:23093-23109
DOI 10.1007/s11042-020-09023-0

Springer

Multimedia Tools and Applications (2020) 79:23093–23109
https://doi.org/10.1007/s11042-020-09023-0

Confidence-based camera calibration with modified census transform

Qicong Dong[1] · Lei Wang[1] · Jieqing Feng[1]

Received: 12 June 2019 / Revised: 17 March 2020 / Accepted: 1 May 2020 /
Published online: 6 June 2020

Abstract
Most camera calibration methods assume that the camera is nearly in-focus for precisely estimating the intrinsic and extrinsic camera parameters. However, the camera is generally out-of-focus in real calibration applications. This paper presents a method that is capable of accurately calibrating an out-of-focus camera. A confidence-based camera calibration with modified census transform (MCT) is proposed for a checkerboard calibration pattern. No additional devices or special calibration targets are required. First, the MCT utilizes the intensity information and local image structure to construct a descriptor. Then, the dissimilarity between the control point on the real image and the one on the sharp image is evaluated using the Hamming distance. This similarity metric is treated as the confidence of each control point. Finally, the confidence of each control point is added to the calibration energy minimization procedure to enhance the calibration performance. The experimental results on real images demonstrate that the proposed method achieves a more accurate calibration result than conventional methods.

Keywords Camera calibration · Stereo matching · Modified census transform

1 Introduction

Camera calibration is a fundamental problem in computer vision. It is used to estimate the intrinsic and extrinsic parameters of a camera. The intrinsic parameters describe the internal geometric and optical characteristics of the camera system, whereas the extrinsic parameters are defined as the rotation and translation transformations between the world coordinate system and the camera coordinate system [33]. Because camera calibration determines the

✉ Jieqing Feng
jqfeng@cad.zju.edu.cn

Qicong Dong
qicongdong@zju.edu.cn

Lei Wang
wanglei96@zju.edu.cn

[1] State Key Lab of CAD, CG, Zhejiang University, Hangzhou, 310058, China

mapping relation between the 3D points in the real world and the observed 2D projections on the image plane, its accuracy has a significant impact on the overall performance of computer vision applications. Therefore, improving the camera calibration accuracy could significantly benefit subsequent tasks in the computer vision workflow, especially for triangulation-based applications, such as binocular stereo systems [6] and structured light systems [17].

As one of the most fundamental problems in the field of computer vision, camera calibration has been extensively studied for decades [34]. Various camera calibration methods have been proposed, and they can be roughly classified into three categories: three-dimensional reference-object-based calibration [15, 23, 31], two-dimensional plane-based calibration [30, 33], and self-calibration [22, 26].

In three-dimensional reference-object-based calibration, a precisely constructed calibration object with known 3D geometry [9, 29] is required to establish the correspondences between the 3D points on the object and the associated 2D image points projected on the image plane. The transformation from the 3D real world to the 2D image plane is approximated by some optimization algorithms. However, the fabrication of accurate 3D targets is typically difficult. Two-dimensional plane based calibration requires the acquisition of a few well-planned independent views of a planar pattern. Self-calibration does not use any specific calibration target. Two-dimensional plane-based calibration has attracted considerable attention owing to its flexibility, robustness, and ease of construction. Tsai [31] presents a calibration method using 2D calibration targets. The calibration procedure is simplified because the 2D calibration targets are easier to obtain. Zhang [33] further simplifies the calibration procedure by adding some control points (checkerboard or circular point patterns) to enable arbitrary poses and orientations of a 2D calibration target. Zhang's method and its extensions have been integrated into many software applications, such as the Camera Calibration Toolbox for MATLAB [5] and OpenCV [8], because of its flexibility. The proposed method also utilize the frame of Zhang's method. However, we assign each control point on the checkerboard a confidence in the camera calibration procedure. As shown in Fig. 1, the confidence of each control point is different, depending on the proposed MCT descriptor. These confidences guide the calibration procedure in a more proper way. As a result, the calibration performance is enhanced by the confidence-based method with MCT. Moreover, no additional devices or special calibration targets are required in the whole calibration procedure.

The remainder of this paper is organized as follows. Closely related work is reviewed in Section 2. The concept of a defocused image and the definition of the traditional census transform are described in Section 3. The MCT is explained in detail in Section 4. The confidence-based calibration procedure is presented in Section 5. The experimental results are presented in Section 6. Finally, conclusions are drawn and future work directions are discussed in Section 7.

2 Related work

In this section, we review the closely related literature since a large volume of studies on calibration have been published.

The rapid development of digital imaging devices has led to widespread use of computer-vision-based methods in measurement applications, resulting in a growing demand for high accuracy camera calibration techniques [1]. To achieve accurate camera calibration results, it is important to acquire a set of exact correspondences between the 3D control points in

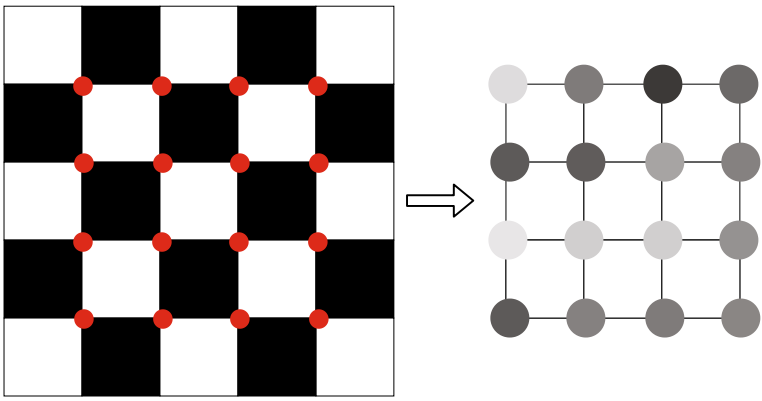


Fig. 1 Control points are assigned with different confidences (different colors) according to the proposed MCT descriptor

the real world and the associated 2D control points projected on the image plane. Under the assumption that the camera calibration target is known with sufficient precision, it is worth noting that the localization precision of the control points on the image plane is one of the main factors affecting the accuracy and stability of camera calibration, and it has been attracting increasing interest in recent years [14, 27]. Accurate detection of control points on the 2D image plane is often difficult in real-world scenarios, especially for defocused images.

In general, conventional methods for extracting control points require well-focused input images to achieve localization of the control points with high accuracy. However, few studies have investigated the out-of-focus blurring effect, which commonly exists in real lens systems. In practice, input images for camera calibration commonly exhibit varying degrees of out-of-focus blur, especially for camera systems having a very limited depth of field (DoF) [3, 7]. Lens focusing for such camera systems is often tedious even if the camera is equipped with an auto-focus lens. In general, a calibration target is required to be placed at a number of positions and orientations such that it can span around the 3D space, and it will be processed in the subsequent vision tasks. Thus, the out-of-focus effect of input images for camera calibration is practically unavoidable even after fine-tuning the acquisition setup. The effect of defocus blur tends to be increasingly severe in current high-resolution cameras with high-density optical sensors, i.e., each pixel has a much smaller physical size in the optical sensor. Thus, conventional methods for extracting the control points are no longer suitable for directly application to the out-of-focus areas of the input images.

Baba et al. [2] propose a unified camera calibration method using a thin-lens-based camera model that adds the reprojected blur width error of the control points to the traditional cost function for camera parameter estimation. Kannala et al. [24] propose a generic camera model that is applicable for conventional cameras, wide-angle lens cameras and fish-eye lens cameras. This generic model uses a radially symmetric projection assumption that is more flexible for various camera lenses. Geiger et al. [16] find that for each control point, the image gradient at a neighboring pixel should be approximately orthogonal to the direction from the neighboring pixel to the control point. The position of the control point is estimated according to that rule. The method proposed in [28] finds the saddle point of the polynomial surface that is estimated around the candidate control point, and it considers the position of this saddle point to be the accurate location of the control point. Datta et al. [14]

use the parameters computed by traditional calibration methods for initialization, and then they refine the positions of the control points in an iterative manner. The reprojection errors of [14] are considerably lower than those of the traditional calibration methods. Ha et al. [18] propose a camera calibration method for defocused images using five complementary binary patterns consecutively displayed on a smartphone. They model the defocus blur as a convolution of a sharp image with a Gaussian point spread function on the binary patterns, which simplifies the 2D Gaussian deconvolution problem to a 1D Gaussian deconvolution problem. Bell et al. [4] propose a calibration method that encodes control points into the carrier phase with the fringe patterns obtained by a digital display. Consequently, the accuracy of the localization of control points is not affected by the amount of defocusing. Chuang et al. [13] proposes a new calibration method that consider the outliers in the computation and the focal length in the calibration procedure that enhances the correctness and robustness. Lopez et al. [25] predict extrinsic and intrinsic parameters from a single image to handle radial distortion problem, which introduces a new parameterization of radial distortion tailored for the learning-based method.

Accurate localization of the control points in defocused images is important, and many studies have focused on this task. However, few studies have investigated the final calibration energy minimization procedure. The proposed method assigns a confidence to each control point in the calibration procedure, and this confidence is able to guide the calibration procedure in a more proper manner. We present a modified census transform (MCT) to describe each control point on the checkerboard pattern and the dissimilarity between the MCT of the control point on the real image and the one on the sharp image is evaluated using the Hamming distance. The similarity metric is treated as the confidence of each control point in the calibration procedure to enhance the calibration performance.

Feature extraction [19, 20] is important for camera calibration. Census transform [32] is widely used as a similarity metric in stereo matching. It uses one binary string as a descriptor. The census transform encodes local image structures with the relative magnitude of the pixel intensities. Consequently, the census transform is capable of describing the relationship between a single pixel and its neighboring pixels within the image patch. Given the special shape of the checkerboard pattern, we modify the traditional census transform by not only considering the relative magnitude of pixels but also utilizing the pixel's intensity information itself. We divide the image patch centered at each control point into two parts: the cross region and the other regions, which will be described in Section 4. In the cross region, we compute the intensity similarity between the real image and the sharp image. In the other regions, we utilize the census transform to consider the structure information, and we compute the structure similarity between the real image and the sharp image. Experiments demonstrate that the MCT achieves superior results compared to the traditional census transform in the camera calibration procedure.

3 Preliminaries

In this section, we describe the phenomenon of defocused images and the traditional census transform in stereo matching.

3.1 Defocused image

The camera calibration method is generally based on the pinhole camera model. As shown in Fig. 2, if one point is placed at the focal distance in front of the lens, then all the light rays

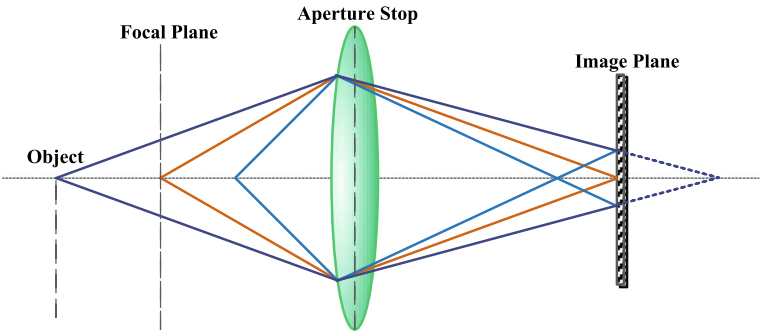


Fig. 2 Thin-lens camera model

from the point will ideally converge to a single point on the image plane, which generates a sharp image point. However, when a point is placed at a distance that is closer or farther than the focal plane, the light rays from that point will converge before or after reaching the image plane. Consequently, the rays spread over a small area on the image plane, and the point will be defocused. As shown in Fig. 3, the images that are closer or farther than the focal plane are blurred. Thus, the positions of these control points on the image are blurred and unreliable, thereby influencing the calibration precision. The proposed method assigns each control point a confidence to resolve this defect. Section 4 presents the details of defining the confidence.

3.2 Census transform

The census transform is widely used in stereo matching for describing the neighboring structure of a single pixel. Fig. 4 illustrates the traditional census transform. The census transform is a nonparametric measure, and it encodes local image structures with the relative magnitudes of the pixel intensities. For each pixel p, the relative magnitudes of intensities between p and its neighboring pixels are compared. For each pixel in the neighborhood, if its intensity is lower than that of the center pixel p, the value of the pixel is defined as zero. Otherwise, the value of the pixel is defined as one:

$$c(p,q) = \begin{cases} 0 & if\ I(q) < I(p) \\ 1 & otherwise \end{cases} \tag{1}$$

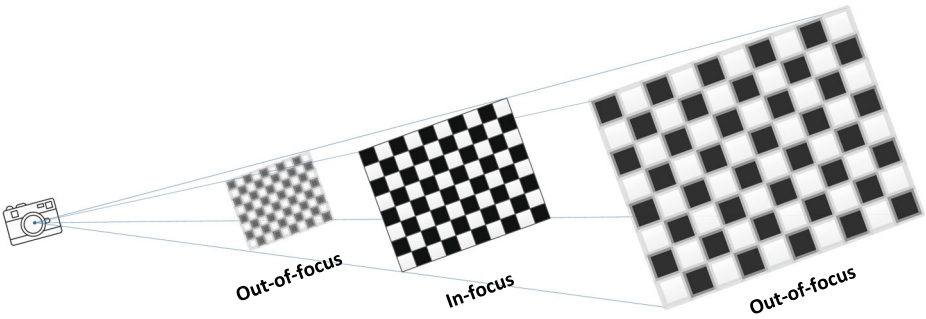


Fig. 3 In-focus and defocused images

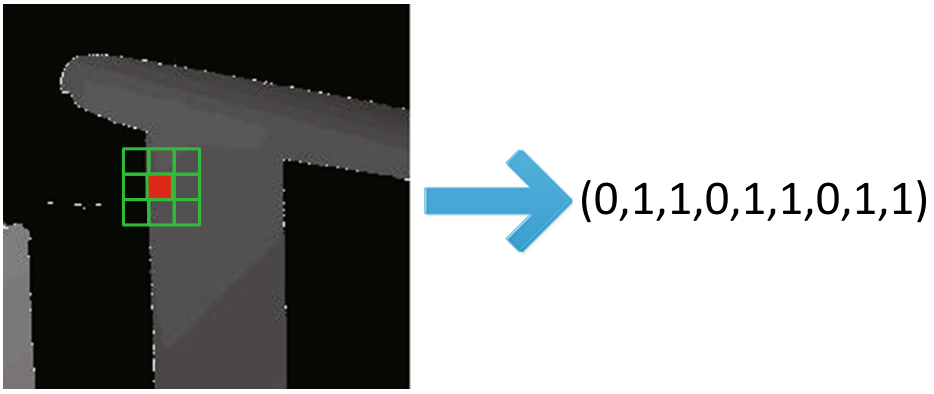


Fig. 4 An illustration of the traditional census transform. The intensities of the center pixel (red) and the neighboring pixels are compared. The result of the census transform is a binary string

where q is a pixel in p's neighborhood, and $I(q)$ denotes the intensity of pixel q. Then, the values in the neighborhood are concatenated as a binary string as the descriptor of pixel p:

$$cen(p) = \bigotimes_{q \subseteq W_p} c(p, q) \tag{2}$$

where W_p is the neighborhood of pixel p, and $\bigotimes$ is the concatenate operator.

In stereo matching, the census transform has an advantage over describing the structure information. For corresponding pixels in the left and right images, the Hamming distance between the two pixels' census transform descriptors is defined as a similarity metric. The proposed method is inspired by this similarity metric definition. We utilize this similarity metric to compute the dissimilarity between the control point in the real image and the one in the sharp image. The dissimilarity reflects the confidence of one control point in the real image, and this confidence can be utilized in the final calibration procedure. The experimental results demonstrate the effectiveness of this similarity metric.

4 Modified census transform for checkerboard pattern

The flowchart of the proposed method is illustrated in Fig. 5. The proposed method utilizes the MCT to compute the difference of control point between the sharp image and the real image. The difference is formulated as the confidence of the control points, which is integrated into the final camera calibration step.

First, for each input defocused image, initial control point detection of the checkerboard pattern is performed using the Harris corner detector [21] as implemented in OpenCV [8]. The control points are refined with subpixel accuracy with an iterative gradient-based searching algorithm.

After the initial control points are detected, we assign each control point a confidence by utilizing the MCT. In (2), the conventional census transform uses a binary string as a pixel's descriptor. However, this is a definition for general images. During calibration, the target in the images is special, such as the checkerboard pattern, and we modify the traditional census transform. First, we define the MCT of the control point in the sharp image. The

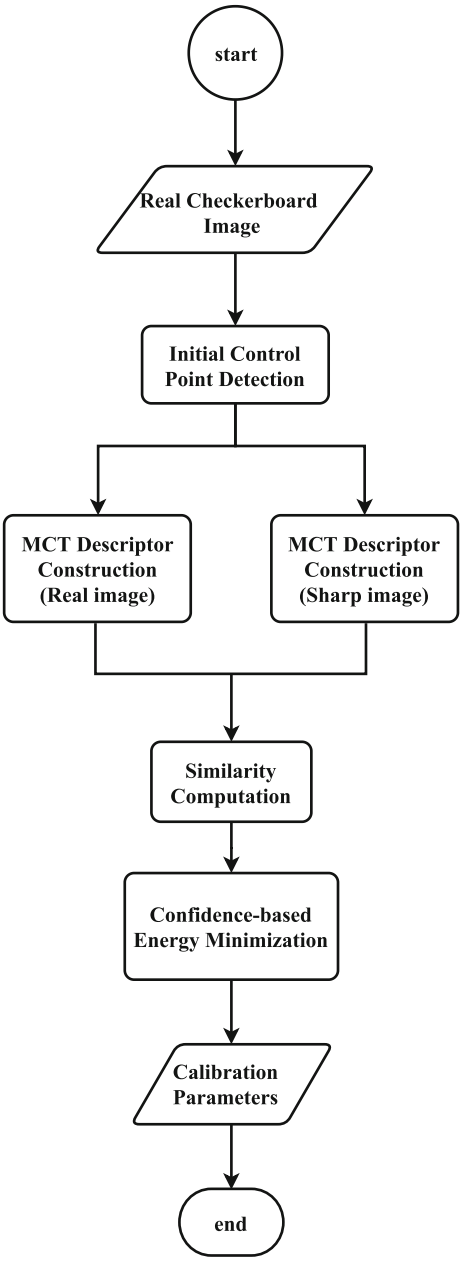
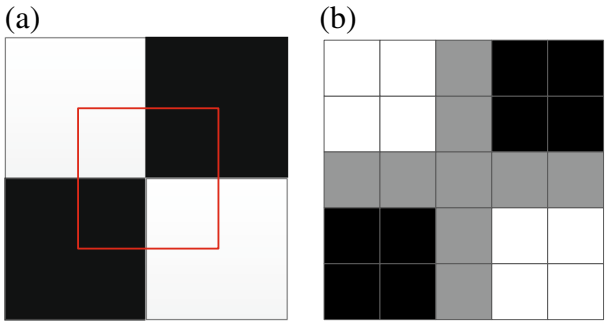


Fig. 5 Flowchart of the proposed method:

sharp image has sharp edges that have no degree of blurring, as shown in Fig. 6a. The control point and the edge points at the four adjacent sharp edges define the cross region, as shown in Fig. 6b (the gray pixels). Note that in practice, the direction of the sharp edge is defined as the line connecting the detected control points. In the checkerboard pattern, the positions of the pixels in the cross region (sharp edges) should have subpixel accuracy, and



(c)

1	1	1	0	0
1	1	1	0	0
1	1	1	1	1
0	0	1	1	1
0	0	1	1	1

(d)

1	1	I_S	0	0
1	1	I_S	0	0
I_S	I_S	I_S	I_S	I_S
0	0	I_S	1	1
0	0	I_S	1	1

Fig. 6 The modified census transform. **a** For each control point, the neighborhood of the control point is defined (the red square). **b** The magnified pixels in the red square region (5×5, for instance), where each small square denotes a pixel. **c** The traditional census transform result of (**b**). **d** The definition of MCT in the sharp image

we define the intensities of the pixels in the cross region as the average of the neighboring pixels' intensities. The definition of the conventional census transform of the control point in the sharp image is illustrated in Fig. 6c. In the cross region in Fig. 6c (the gray pixels in Fig. 6b), the values are all ones according to the census transform definition (the intensities of the pixels in the cross region and the center pixel are equal). Consequently, the census transform of the center pixel explores the same information in the cross region as in the white regions, which have higher intensity than the center pixel. This approach clearly has a negative effect on describing the control points in the checkerboard patterns. To distinguish the information between the cross region and the white regions, we modify the traditional census transform. For each pixel in the cross region, the relative intensity magnitude is not compared with the center pixel. Rather, we retain the intensity values for the pixels in the cross region. Figure 6d illustrates the descriptor of the MCT. In the sharp image, the intensity I_s of the cross region is defined as the average of the intensities of the neighboring pixels. The size of the window is defined as 11×11 here. Then, the value is divided by 255 to be controlled in the range of zero to one. For the real image, the definition of the MCT is similar to that for the sharp image. The only difference is that the pixel intensity in the real image is the real sampled pixel intensity.

Then, the computation of similarity between the control points in the real image and those in the sharp image can be divided into two parts: the dissimilarity in the cross region, and the dissimilarity in other regions. For the pixels in the cross region, we compute

the sum of absolute intensity difference between the real image and the sharp image as follows:

$$D_{cross}(p_r, p_s) = \frac{\sum_{r \subseteq R_{pr}, s \subseteq R_{ps}} |I_r - I_s|}{I_{max}} \tag{3}$$

where p_r is the control point in the real image; p_s is the control point with the same position in the sharp image; R_{p_r} and R_{p_s} are the cross region of p_r and p_s, respectively; I_r denotes the intensity of real image pixels in the cross region, I_s denotes the intensity of sharp image pixel in the cross region; and I_{max} denotes the maximum intensity value, i.e., 255.

For pixels in other regions, the census transform values are concatenated as a binary string, and the Hamming distance between the descriptors of the real image and the sharp image is computed as the dissimilarity of the two points in other regions, as follows:

$$D_{other}(p_r, p_s) = cen(p_r) \bigoplus cen(p_s) \tag{4}$$

where $cen(p_r)$ denotes the census transform values of the control point in the real image, $cen(p_s)$ denotes the census transform values of the control point in the sharp image, and $\bigoplus$ denotes the Hamming distance operator.

Finally, the dissimilarity between p_r and p_s is defined as the sum of the two dissimilarities in (3) and (4), as follows:

$$D(p_r, p_s) = D_{cross}(p_r, p_s) + D_{other}(p_r, p_s) \tag{5}$$

5 Confidence-based calibration procedure

First, we introduce the definition of the Mean Reprojection Errors (MRE):

$$MRE = \sum_{i=1}^{M} \sum_{j=1}^{N} \left\| P_{ij} - \hat{P}_{ij} \right\| \tag{6}$$

where M is the number of images and N is the number of control points in each image, P_{ij} is the projection of the control point in the 3D world coordinate system and $\hat{P}_{ij}$ is the corresponding 2D image point.

After the similarity metric is defined, we can obtain the dissimilarity between the control points in the real image and the ones in the sharp image. The dissimilarity measures the confidence of the detected control point localization. Intuitively, the localization error of the control point increases as the dissimilarity increases. The modified census transform is adopted to define the weight for each control point, which depicts the confidence of the detected control point localization. In this way, the out-of-focus problem in camera calibration is alleviated by adding the weights for each control point. For the jth control point in the ith image, the dissimilarity from the corresponding control point in the sharp image is defined as D_{ij}, where $D = D(p_r, p_s)$ for short. The confidence c_{ij} in the calibration procedure is defined as follows:

$$c_{ij} = e^{-q \cdot D_{ij}} \tag{7}$$

where q is a constant value and will be discussed in Section 6.5. The confidence reflects the importance of each control point, and this confidence is normalized as follows:

$$C_{ij} = c_{ij} / \overline{c} \tag{8}$$

and

$$\overline{c} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} c_{ij} \tag{9}$$

The normalized confidence of each control point reflects its confidence in the defocused image, which determines the relative importance of the control point when minimizing the energy function of the camera calibration.

Finally, the energy function is defined as follows:

$$E = \sum_{i=1}^{M} \sum_{j=1}^{N} C_{ij} \left\| P_{ij} - \hat{P}_{ij} \right\| \tag{10}$$

6 Experimental results

The experiments are all conducted on a desktop equipped with an Intel(R) Core(TM) i7-3770K CPU at 3.50 GHz and 16 GB of RAM. Real images are used to evaluate the performance of the proposed method. The camera is a Canon EOS 600D DSLR with a Canon EF 50 mm f/1.8 STM lens. The resolution of the captured images is 5184 × 3456 pixels. The camera system is focused at a distance of approximately 1.5 m. For each dataset, 30 input images are taken from different positions and orientations. Three other test images for each dataset are used for the calibration accuracy measurement in Section 6.2. Note that the three test images are well placed in the focal plane. For each dataset, 21 images were randomly selected from the 30 input images as the training input images under various blur conditions. A 24 × 14 checkerboard pattern is used. The size of the neighborhood of MCT is defined as 11 × 11. The mean reprojection error is used as the criterion for evaluating the calibration performance.

The proposed method is evaluated from various perspectives, including the improvement over traditional calibration methods in Section 6.1, the evaluation of the accuracy of camera parameters in Section 6.2, the effectiveness of the MCT in Section 6.3, the evaluation of the neighborhood definition in Section 6.4, the evaluation of sharp image construction in Section 6.5, and the parameter and time analysis in Section 6.6.

6.1 Improvement over traditional calibration methods

First, the improvements from applying the confidence-based calibration method over the traditional calibration methods are evaluated. The control points are extracted using three traditional localization methods: the control point refinement procedure in OpenCV [8], the algorithm presented by Geiger et al. [16], and ROCHADE [28]. The mean reprojection errors are shown in Table 1. The percentage denotes the improvement in the calibration accuracy. As shown in Table 1, the calibration accuracy of the proposed method is higher than those of the conventional methods. Compared to the method in [16], there is a considerable improvement in the calibration performance. The confidences of the control points guide the calibration procedure in a more proper way.

6.2 Evaluation of the accuracy of camera parameters

In this subsection, we further evaluate the accuracy of the camera parameters by using three additional test images captured around the focal plane that were not included in the

Table 1 Mean reprojection errors of the camera calibration with and without confidences

	Without confidences		
	OpenCV	Geiger	ROCHADE
Data1	0.56080	1.00017	0.56012
Data2	0.58262	1.21417	0.58006
Data3	0.56579	1.01078	0.56466
Data4	0.56272	1.00281	0.56341
Data5	0.56302	1.00673	0.56320
	With confidences		
Data1	0.46757 (16.62%)	0.50203 (49.81%)	0.50524 (9.80%)
Data2	0.50043 (14.11%)	0.50527 (58.39%)	0.47577 (17.98%)
Data3	0.47956 (15.24%)	0.49964 (50.57%)	0.51537 (8.73%)
Data4	0.50122 (10.93%)	0.52747 (47.40%)	0.48697 (13.57%)
Data5	0.51928 (7.77%)	0.50834 (49.51%)	0.48536 (13.82%)

calibration process. The OpenCV [8] control point extraction method is used in this experiment. The performances of the calibration with and without the confidences are compared. As shown in Table 2, the mean reprojection errors are clearly reduced by the proposed confidence-based camera calibration method. The trained parameters are robust and perform well in the calibration procedure of test images.

6.3 Effectiveness of the MCT

We also compare the similarity metric performances of the conventional census transform and the proposed MCT. The result is shown in Fig. 7, where the mean reprojection error is lower when using the MCT as the similarity metric. The conventional census transform cannot explore the information in the cross region well, as stated in Section 4, which has a negative effect on defining the confidences of the control points. However, the MCT processes the information of the cross region more effectively, and the confidences are defined more reasonably.

6.4 Evaluation of the sampling strategy

The MCT uses binary strings as the descriptor of neighboring sampling pixels, and the sampling strategy of the proposed method is as follows: for each location pair, we choose the center pixel p and the pixel q in p's neighborhood (except for the cross region). In

Table 2 Quantitative analysis on test images using OpenCV [8] control point extraction

	Without confidences	With confidences
Dataset1	0.482997	0.458526 (5.07%)
Dataset2	0.547746	0.505966 (7.63%)
Dataset3	0.604864	0.566243 (6.38%)
Dataset4	0.668804	0.606188 (9.36%)
Dataset5	0.502463	0.454530 (9.54%)

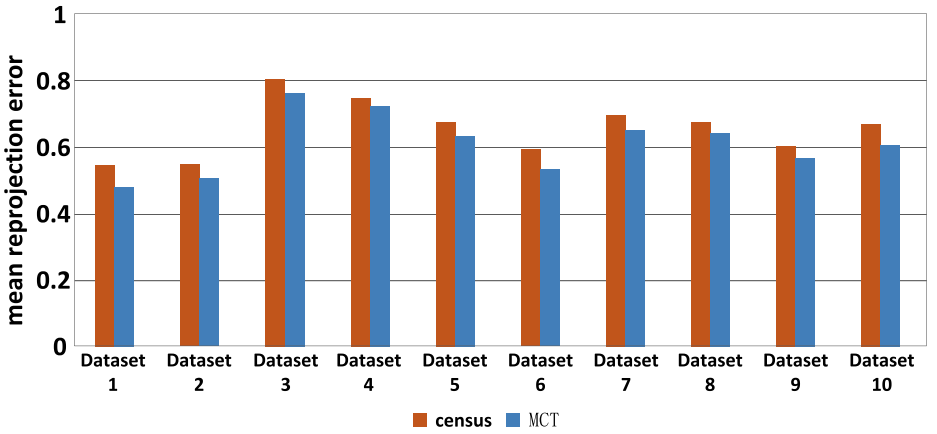


Fig. 7 The comparison of the performances between the traditional census transform and the proposed MCT

other fields, there is also a binary string descriptor called binary robust independent elementary features (BRIEF) [10]. In [10], there are five sampling strategies for generating the binary string. These five sampling strategies are random sampling methods with different probability distributions. In other words, for each location pair, the positions of p and q are randomly chosen according to some probability distributions. Details of five sampling strategies can refer to the reference [10]. We also evaluate the performances when applying these five sampling strategies to generate the binary strings in our confidence-based camera calibration procedure. Table 3 shows the mean reprojection errors when choosing different sampling strategies. For most datasets, the sampling strategy of the proposed MCT performs the best. The sampling strategy of the MCT is more suitable for the checkerboard pattern since the information of the center pixel (control point) is fully considered.

6.5 Evaluation of sharp image construction

In Section 4, we demonstrate the method of sharp image construction. For each control point, the direction of the sharp edge is defined as the line connecting the neighboring

Table 3 Comparison of the five sampling strategies in [10] and the sampling strategy in the proposed method, where the bold number in each row indicates the minimal mean reprojection error among different sampling strategies

	Origin	Sampling1	Sampling2	Sampling3	Sampling4	Sampling5	Ours
Dataset1	0.548	0.482	0.543	0.510	0.494	0.484	**0.480**
Dataset2	0.805	0.763	0.781	0.763	0.763	0.785	**0.763**
Dataset3	0.747	**0.721**	0.724	0.727	0.723	0.747	0.723
Dataset4	0.593	0.539	0.543	0.539	0.543	0.593	**0.532**
Dataset5	0.674	0.654	0.663	0.654	0.663	0.654	**0.643**
Dataset6	0.668	0.641	0.668	0.665	0.606	0.652	**0.606**
Dataset7	0.619	0.615	0.598	0.585	0.594	0.595	**0.583**
Dataset8	0.609	0.574	0.574	0.597	0.570	0.574	**0.570**

detected control points. However, the control point detection is not guaranteed to be accurate and as a result, the direction of sharp edges may not be accurate. The dissimilarity computation is influenced by this fact. To evaluate the degree of this influence, for a cross region in the sharp image, we add a small random value to the positions of the four adjacent control points and compute the dissimilarity accordingly. The test is repeated 300 times and Fig. 8 illustrates the result. kh and kv denote the slope of the horizontal line and the vertical line, respectively. The variance of these dissimilarities is 0.0017, which is considerably small. The dissimilarity computation is insensitive to the position turbulence of the control points.

6.6 Parameter and time analysis

In this subsection, the choice of the parameter q in (7) is evaluated. We evaluate the mean percent improvement in camera calibration on all the datasets when choosing q from 1 to 12, and the result is shown in Fig. 9. When q is small, the mean improvement for all the datasets increases with q since the control points with a smaller dissimilarity to the sharp image control points play a more important role in the calibration procedure, and the error from the control points that indicate a larger control point localization error are suppressed. However, the mean percent improvement decreases after it reaches the maximum value since the confidences of the control points with inaccurate positions are too small. Consequently, the number of control points that constrain the calibration procedure is reduced, and the calibration procedure becomes unstable. Considering the efficiency and stability of the calibration, we chose $q = 7.2$ in our experiment.

The window size of the MCT is also evaluated. The neighbor size is chosen from 3 to 19, and the corresponding weight is computed and illustrated in Fig. 10. It shows that the weight is not sensitive with the range from 11 to 17.

Regarding runtime, each dataset can be completed in 6 min with an Intel(R) Core(TM) i7-3770K CPU at 3.50 GHz and 16 GB of RAM. Because camera calibration is generally an offline task, the runtime of the proposed method is acceptable in practice. The running time is also compared with the state-of-the-art methods [11, 12] shown in Table 4.

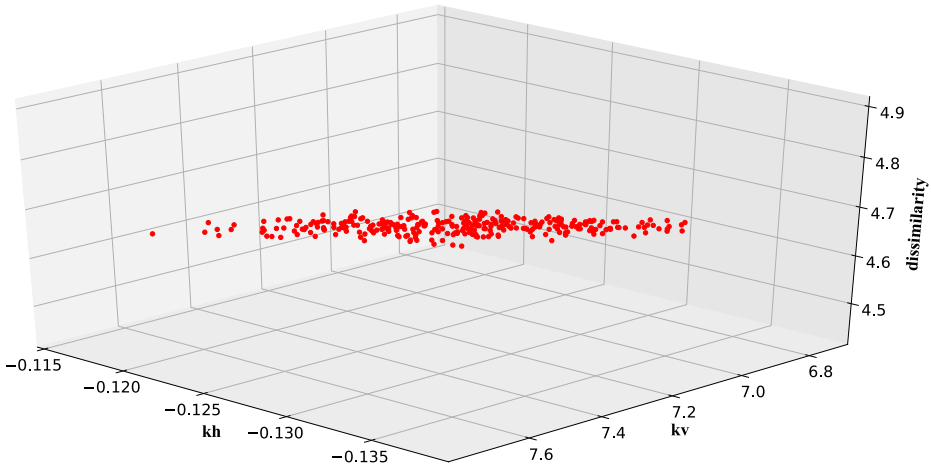


Fig. 8 Dissimilarities with different control point positions

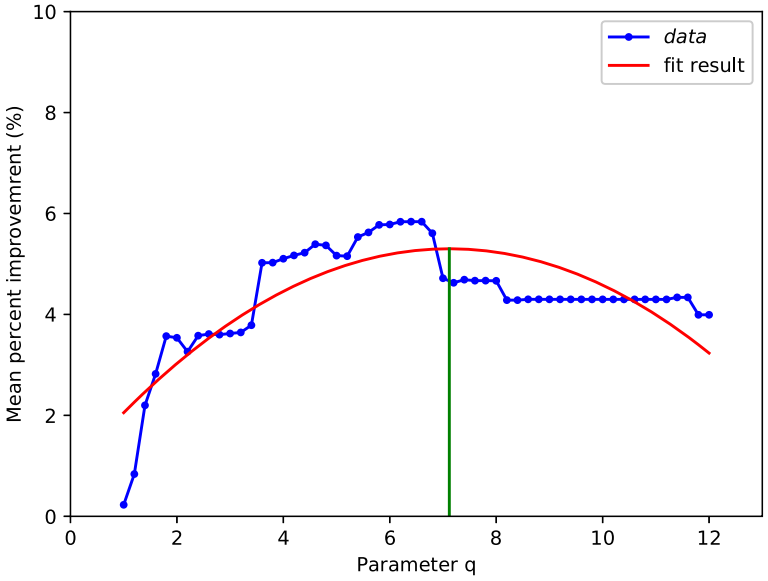


Fig. 9 Mean percent improvement on the test images according to parameter q in (7)

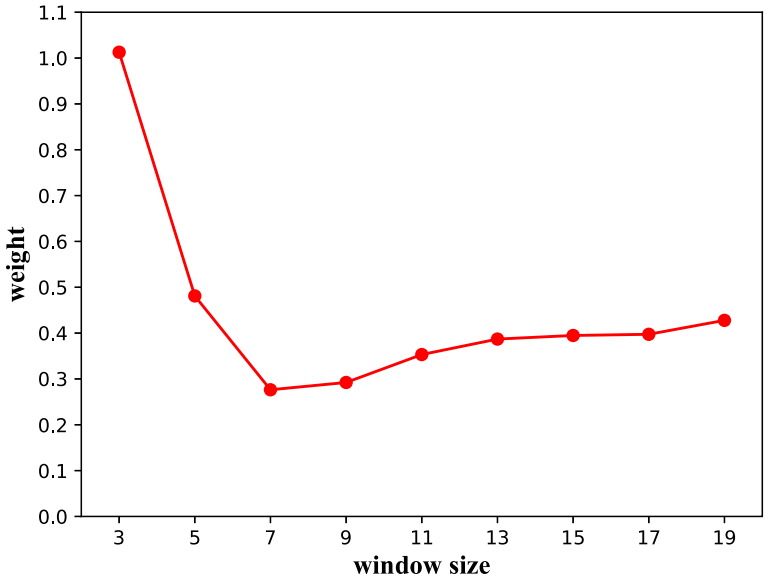


Fig. 10 The weight of MCT from different window sizes

Table 4 Comparison of the calibration running time

	Method in [11]	Method in [12]	The proposed method
Time (second per mega-pixel)	0.9766	0.3255	0.2791

7 Conclusion

We propose a confidence-based camera calibration with MCT. The MCT assigns each control point a reasonable confidence and guides the calibration procedure to achieve superior performance. Moreover, the proposed method can easily be applied in various fields, such as out-of-focus detection, control point confidence description, and camera calibration for stereo or multi-view stereo. The proposed method does not require additional devices or a special calibration target. The experimental results demonstrate the effectiveness and robustness of the proposed method. How to define the confidence for each control point in different calibration patterns might be a future work. The definition of the confidence of the control points is also an open issue to study with. The information of the blur on the control points may be considered in the subsequent steps such as the rectification in stereo matching, or the feature extraction in multi-view geometry, to enhance the performance of the 3D reconstruction system.

Acknowledgments This work was jointly supported by the National Natural Science Foundation of China under Grant Nos. 61732015 and 61472349 and by Key Research and Development Program of Zhejiang Province under Grant No 2018C01090.

References

1. Albarelli A, Rodolà E, Torsello A (2010) Robust camera calibration using inaccurate targets. In: British machine vision conference, pp 16.1–16.10
2. Baba M, Mukunoki M, Asada N (2006) A unified camera calibration using geometry and blur of feature points. In: 18th international conference on pattern recognition (ICPR). vol 1, pp 816–819
3. Beeler T, Bickel B, Beardsley P, Sumner B, Gross M (2010) High-quality single-shot capture of facial geometry. ACM Trans Graph 29(4):40
4. Bell T, Xu J, Zhang S (2016) Method for out-of-focus camera calibration. Appl Opt 55(9):2346–2352
5. Bouguet JY (2013) Camera calibration toolbox for matlab. https://www.vision.caltech.edu/bouguetj
6. Bradley D, Boubekeur T, Heidrich W (2008) Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In: IEEE conference on computer vision and pattern recognition, pp 1–8
7. Bradley D, Heidrich W, Popa T, Sheffer A (2010) High resolution passive facial performance capture. ACM Trans Graph 29(4):41
8. Bradski G (2000) The opencv library. Doctor Dobbs Journal 25(11):120–126
9. Brown DC (1971) Close-range camera calibration. Photogramm Eng 37(8):855–866
10. Calonder M, Lepetit V, Strecha C, Fua P (2010) Brief: binary robust independent elementary features. In: European conference on computer vision, pp 778–792
11. Chen J, Little JJ (2017) Where should cameras look at soccer games: Improving smoothness using the overlapped hidden markov model. Comput Vis Image Underst 159:59–73
12. Chen J, Zhu F, Little JJ (2018) A two-point method for ptz camera calibration in sports. In: 2018 IEEE winter conference on applications of computer vision, pp 287–295
13. Chuang JH, Ho CH, Umam A, Chen H, Lu MT, Hwang JN, Chen TA (2019) A new technique of camera calibration: A geometric approach based on principal lines. arXiv:1908.06539
14. Datta A, Kim JS, Kanade T (2009) Accurate camera calibration using iterative refinement of control points. In: IEEE 12th international conference on computer vision workshops, pp 1201–1208
15. Faugueras OD, Toscani G (1989) The calibration problem for stereoscopic vision. Sensor devices and systems for robotics, pp 195–213

16. Geiger A, Moosmann F, Car Ö, Schuster B (2012) Automatic camera and range sensor calibration using a single shot. In: 2012 IEEE international conference on robotics and automation, pp 3936–3943
17. Geng J (2011) Structured-light 3d surface imaging: a tutorial. Adv Opt Photon 3(2):128–160
18. Ha H, Bok Y, Joo K, Jung J, So Kweon I (2015) Accurate camera calibration robust to defocus using a smartphone. In: IEEE international conference on computer vision, pp 828–836
19. Han J, Farin D, de With PH (2008) Broadcast court-net sports video analysis using fast 3-d camera modeling. IEEE Trans Circ Syst Video Technol 18(11):1628–1638
20. Han J, Pauwels EJ, de Zeeuw P (2013) Visible and infrared image registration in man-made environments employing hybrid visual features. Pattern Recogn Lett 34(1):42–51
21. Harris C, Stephens M (1988) A combined corner and edge detector. In: Alvey vision conference, pp 147–151
22. Hartley RI (1994) An algorithm for self calibration from several views. IEEE Comp Soc Conf Comp Vis Pattern Recogn 94:908–912
23. Heikkila J (2000) Geometric camera calibration using circular control points. IEEE Trans Pattern Anal Mach Intell 22(10):1066–1077
24. Kannala J, Brandt SS (2006) A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. IEEE Trans Pattern Anal Mach Intell 28(8):1335–1340
25. Lopez M, Mari R, Gargallo P, Kuang Y, Gonzalez-Jimenez J, Haro G (2019) Deep single image camera calibration with radial distortion. In: IEEE conference on computer vision and pattern recognition, pp 11,817–11,825
26. Luong QT, Faugeras OD (1997) Self-calibration of a moving camera from point correspondences and fundamental matrices. Int J Comput Vis 22(3):261–289
27. Nakano K, Okutomi M, Hasegawa Y (2002) Camera calibration with precise extraction of feature points using projective transformation. In: 2002 IEEE international conference on robotics and automation. vol 3, 2532–2538
28. Placht S, Fürsattel P, Mengue EA, Hofmann H, Schaller C, Balda M, Angelopoulou E (2014) Rochade: Robust checkerboard advanced detection for camera calibration. In: European conference on computer vision, pp 766–779
29. Sobel I (1973) On calibrating computer controlled cameras for perceiving 3-d scenes. Artif Intell 5(2):185–198
30. Sturm P, Maybank S (1999) On plane-based camera calibration: a general algorithm. IEEE Comp Soc Conf Comp Vis and Pattern Recogn 1:432–437
31. Tsai RY (2003) A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. IEEE J Robot Autom 3(4):323–344
32. Zabih R, Woodfill J (1994) Non-parametric local transforms for computing visual correspondence. In: European conference on computer vision, pp 151–158
33. Zhang Z (2000) A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 22(11):1330–1334
34. Zhang Z, Matsushita Y, Yi M (2011) Camera calibration with lens distortion from low-rank textures. In: IEEE conference on computer vision and pattern recognition, pp 2321–2328

Qicong Dong is now a Ph.D student in State Key Lab of CAD&CG, Zhejiang University of China. His research interests include stereo matching and multi-view stereo.

Lei Wang is now a master student in State Key Lab of CAD&CG, Zhejiang University of China. Her research interests include camera calibration and multi-view stereo.

Jieqing Feng is a professor in the State Key Lab of CAD&CG, Zhejiang University, Peoples Republic of China. He received his BSc in applied mathematics from the National University of Defense Technology in 1992 and his Ph.D in computer graphics from Zhejiang University in 1997 respectively. His research interests include digital geometry modeling, stereo vision, and thermal solar energy simulation and modeling.