マルチモーダル全方位カメラのキャリブレーション方法

Calibration Methodology for Multimodal Omnidirectional Camera

張　宇鵬* 伊　紅* 宮　衛涛* 于　海華* 王　煒*
Yupeng ZHANG Hong YI Weitao GONG Haihua YU Wei WANG

要　旨

本論文では，RGB-IRデュアル・バンド全方向カメラのキャリブレーション方法を紹介する．この方法は，高次多項式による非線形マッピングを使用し，撮影したIR画像をRGB画像にワープさせる．シミュレーションでは，ワープしたIR画像がRGB画像に完全にマッチすることを示した．さらに，ホモグラフィ変換によって得られた結果よりも優れていることを示した．

ABSTRACT

We propose a method to calibrate RGB and IR images captured by an RGB-IR dual band omnidirectional camera. The proposed method utilizes a non-linear transformation that warps the captured IR image to the RGB image by using higher order polynomials. Simulation results showed that the warped IR image perfectly matches the RGB image, which is better than the results obtained by homography transformation.

* リコーソフトウェア研究所（北京）有限公司
Ricoh Software Research Center (Beijing) Co., Ltd.

1. Introduction

An omnidirectional camera is useful when wide coverage of a scene is required. Such cameras can capture the surrounding areas with a 360° field of view (FOV) in a single shot without losing any of the target, which makes them invaluable for a variety of security applications.

However, most omnidirectional cameras use RGB sensors such as CCD or CMOS sensors, which work fine in the daytime and for indoor applications with good illumination but falter under some low light conditions, outdoor scenes, and special cases (e.g., nighttime, foggy, rainy, and snowy days, people wearing camouflage, objects behind leaves/bushes, etc.), which greatly limits their application.

Therefore, we aim to develop a camera that not only captures a 360° FOV but also operates both day and night for indoor and outdoor scenes and for some special cases. We previously designed an RGB-IR dual band omnidirectional camera (called a multimodal omnidirectional camera, Fig. 1) that was originally intended for surveillance purposes but can be adapted to other applications. In the surveillance scenario, human detection and tracking is an important task. It is very useful to use multimodal cameras (RGB and IR) and to acquire RGB and IR images simultaneously so that if one cue fails to track a human, the other will do it. In order to achieve this, the RGB image is calibrated to match the IR image such that the positions in both images overlap.

For perspective cameras, image matching (also known as image registration) between two cameras can be realized

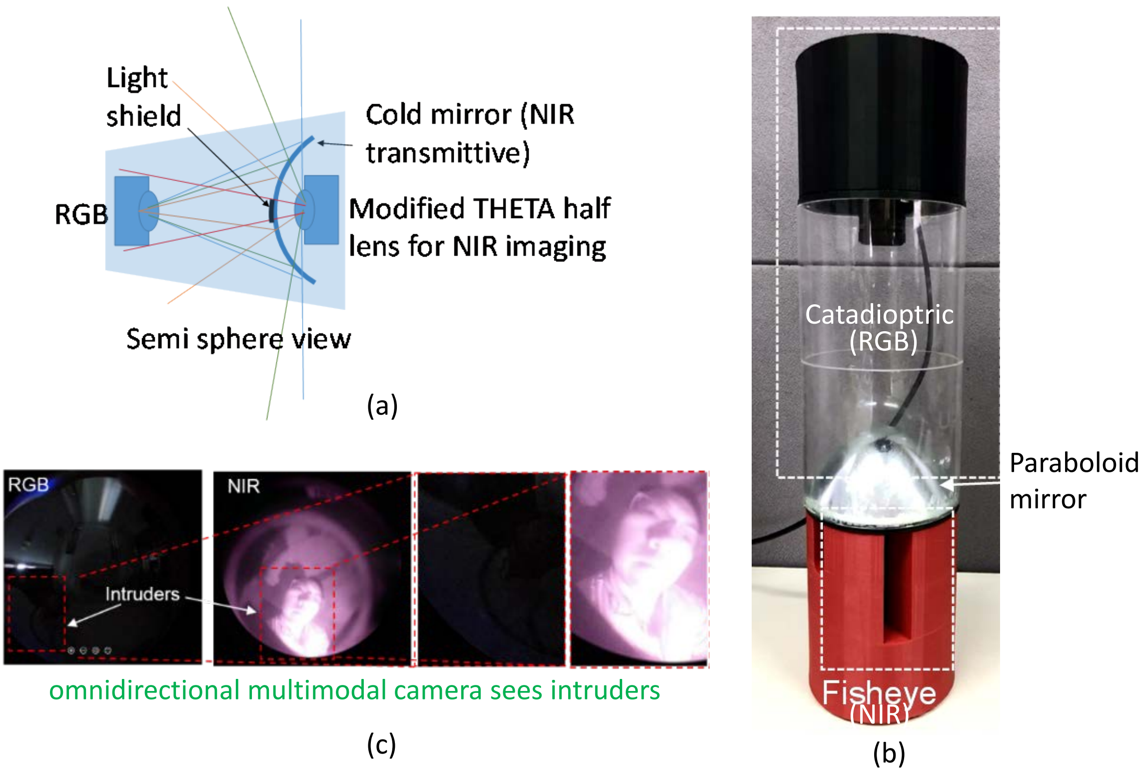


Fig. 1 Multimodal omnidirectional camera[1]. The camera consists of a fisheye near-infrared (NIR) camera modified from RICOH THETA, a catadioptric RGB camera, and a parabolic mirror sandwiched in between.
(a) System configuration of the multimodal omnidirectional camera. The lines represent light rays penetrating into the fisheye NIR camera from different angles and bouncing back towards the catadioptric RGB camera at the same time. (b) A prototype camera. (c) Comparison of RGB and NIR images captured by the prototype camera. As shown, the intruders cannot be visualized in the RGB image but can be clearly visualized in the NIR image.

by affine transformation, which is a linear mapping method. However, for omnidirectional cameras, the captured images are heavily distorted around the image boundaries[2]. In this case, linear mapping is not effective for matching. Non-linear mapping using higher order polynomials is an effective way to obtain a better matching result.

In this paper, we introduce a non-linear mapping method that can be used to calibrate two omnidirectional cameras. We describe the theory and simulation results and also perform comparisons with linear mapping methods such as homography transformation.

2. Property of co-axis and same FOV of the fisheye and catadioptric camera

One important property of the proposed camera system in Fig. 1 is that the fisheye and catadioptric sides share same field of view. This can be proved by simulation as shown in Fig. 2. As shown in the optical system in Fig. 2 below, the fisheye and catadioptric cameras are essentially co-axes. So, there is essentially no disparity between the two cameras. The right column in Fig. 2 shows image simulation results. It can be seen that the fisheye and catadioptric image includes all nine vertical and horizontal lines of the original image, and they share same center of the grid image.

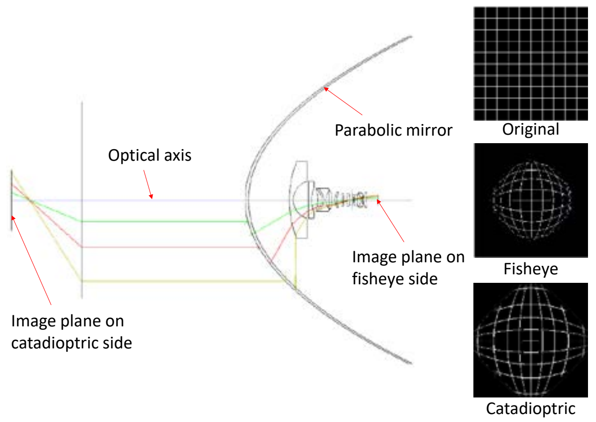


Fig. 2 Optical simulation showing co-axis and same FOV shared by fisheye and catadioptric camera.

3. Non-linear transformation

The motivation for our non-linear mapping method is taken from the concept outlined in a work by Zhang and Ueda[3]. Suppose we have a target image to be matched: namely, we want to warp the IR image $I_{IR}(x,y)$ to the RGB image $I_{RGB}(x',y')$ by a non-linear function f, where x,y and x',y' represent the index of pixels in the IR and RGB image, respectively. The radial distance in both images can be expressed as $r=\sqrt{x^2+y^2}$ and $r'=\sqrt{x'^2+y'^2}$, where r and r' represent radial distance calculated from the image center of the IR and RGB images, respectively.

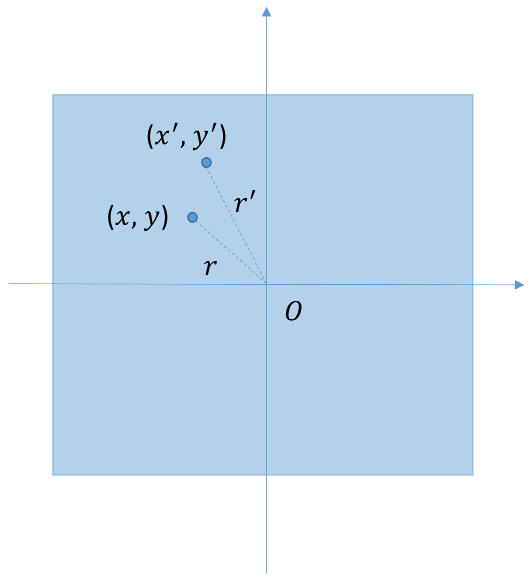


Fig. 3 Two pixels in the IR and RGB image, and their radial distance from the image center.

The proposed method uses a back-forward mapping method[3] that maps the index of the RGB image to that of the IR image by

$$f:\ r = ar'^n + br'^{n-1} + \cdots,$$

where a, b…are coefficients of each polynomial term in the non-linear function f. f can be obtained by curve fitting. The warped points have a decimal pixel index. The pixel values can be obtained by interpolation from the surrounding pixels.

4. Results and discussion

In this section, we present a simulation in which a checkerboard is "captured" by the multimodal omnidirectional camera optical model. The distortion in the captured IR and RGB images are different due to different optical aberrations. Our objective is to warp the checkerboard from the IR image to the RGB image by the proposed calibration method.

4-1 Optical model

The details of the optical model are depicted in Figs. 4 and 5, which show a fisheye and a catadioptric camera, respectively. The fisheye camera is used to capture the IR image and the catadioptric camera is used to capture the RGB image.

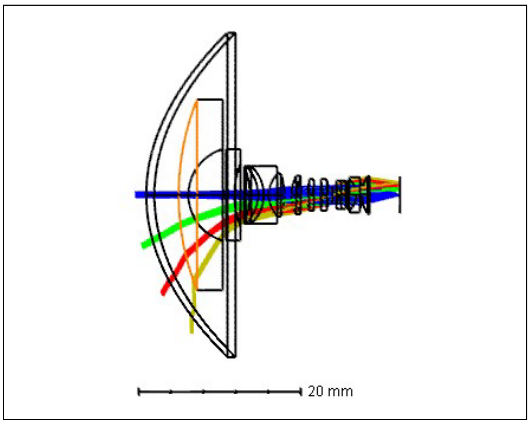


Fig. 4 Optical model of the multimodal omnidirectional camera the fisheye side created by Zemax OpticStudio.

The optical parameters of the fisheye lens are listed in Table 1. The lens is designed to capture a 180° FOV. The blue, green, red, and yellow light rays (the rays from top to bottom) represents light rays at 0°, 30°, 60°, and 90° semi-FOV, respectively.

The catadioptric camera consists of a parabolic mirror and an RGB camera, as shown in Fig. 5. Table 2 lists the optical parameters.

Table 1 Optical parameters of the fisheye lens.

Parameter	Value
Focal length	1.33 mm
Entrance pupil diameter	0.652 mm
F/#	2.1
FOV	180°

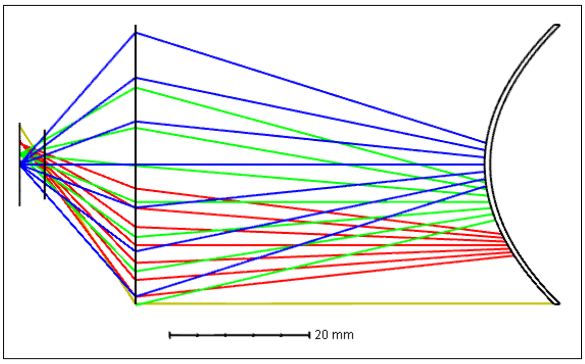


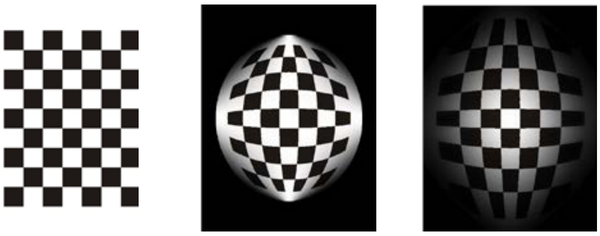
Fig. 5 Optical model of the multimodal omnidirectional camera (the catadioptric side) created by Zemax OpticStudio.

Table 2 Optical parameters of the catadioptric side.

Parameter	Value
Focal length	13 mm
Entrance pupil diameter	40 mm
F/#	0.325
FOV	180°

The catadioptric camera reflects light rays by the parabolic mirror. This is a single effective viewpoint camera using the combination of a parabolic mirror and an orthographic lens[4]. All the reflected chief rays are parallel to the optical axis of the camera. The blue, green, red, and yellow rays (the bunch of rays from top to bottom) represent reflected light rays at 0°, 30°, 60°, and 90° semi-FOV, respectively.

The simulation results of the checkerboard images "captured" by the fisheye (IR) and catadioptric (RGB) camera are shown in Fig. 6. It is clear that the boundaries of the captured images are heavily distorted.



(a) (b) (c)

Fig. 6 Image simulation results. (a) Original checkerboard image. Image "captured" by (b) fisheye side and (c) catadioptric side.

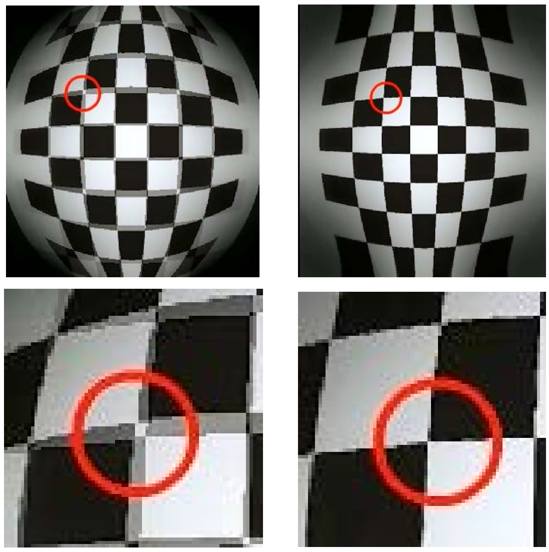
4-2 Calibration

The basic flow of the calibration is as follows. 1) Capture checkerboard images and find corners in both IR and RGB checkerboard images. 2) Compute radial distance of each corner point (the radial distance is the distance from the image center to the corner) in the fisheye and catadioptric images. 3) Use curve fitting to find the higher order polynomial equation between the radial distances in the fisheye image and those in the catadioptric image. 4) Use bilinear interpolation to calculate pixel values of the warped image.

In order to determine the effectiveness of the proposed method, we compare its results with those using homography transformation. The visual results, in which the warped IR image overlaps with the RGB image, are shown in Fig. 7.

It is clear that the warped IR image is perfectly matched to the RGB image by using the non-linear transformation method while it is not by using the homography transformation. This is because the latter utilizes linear equations that lead to mis-matching between the two images. For the non-linear mapping, we selected 12 corner points along the radial direction in both images and tried to fit the data using higher order polynomials. We found that a 3rd order polynomial equation is good enough. The fitted equation is $r = -2.016\mathrm{e}{-}06 * r'^3 - 0.0005955 * r'^2 + 1.041 * r' - 0.5823$, where r' is the radial distance in the catadioptric image and r is the radial distance in the fisheye image. Note that both images are 480 * 640 pixels in horizontal and vertical direction, respectively.

To evaluate the proposed method quantitatively, we calculated the Euclidean distance between corner points on the warped image and those on the catadioptric (RGB) image. We used all 48 corner points for evaluation and computed the average distance for both homography and non-linear transformation. The results (Table 3) demonstrate that the proposed non-linear transformation is significantly better than homography transformation.



(a) (b)

Fig. 7 Camera calibration results. Calibration by (a) homography transformation and (b) proposed non-linear transformation. A close-up is provided at the bottom.

Table 3 Comparison of homography transformation with non-linear transformation by Euclidean distance.

	Homography	Non-linear
Mean (Euclidean)	7.5811	1.3248

Note that the proposed method can calibrate not only fisheye and catadioptric cameras but also any two omnidirectional camera images, e.g., images captured by two fisheye cameras or by two catadioptric cameras.

5. Conclusion

In this paper, we proposed a method to calibrate two omnidirectional images captured by a multimodal omnidirectional camera. The proposed method utilizes a non-linear and backward mapping using higher order polynomial equations to warp one image to the other. This method is useful for a variety of applications, particularly those used for surveillance purposes.

References

1) Y. Zhang et al.: Multimodal omnidirectional camera, Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP), OSA Technical Digest (online), paper ITh3E.4 (2017).
2) V. K. Asari: Non-linear spatial warping of endoscopic images: An architectural perspective for real time applications, *Microprocessors and Microsystems*, Vol. 26, No. 4, pp. 161-171 (2002).
3) Y. Zhang, T. Ueda: Field-Dependent Distortion Coefficient and Backward Mapping for Distortion Correction of Singlet Lens Cameras, *IEEJ Trans on Electrical and Electronic Engineering*, Vol. 5, No. 2, pp. 203-210 (2010).
4) D. Scaramuzza: OCamCalib: Omnidirectional Camera Calibration Toolbox for Matlab, https://sites.google.com/site/scarabotix/ocamcalib-toolbox (2010).