HAL
open science

Build your own hybrid thermal/EO camera for autonomous vehicle

Yigong Zhang, Yicheng Gao, Shuo Gu, Yubin Guo, Minghao Liu, Zezhou Sun, Zhixing Hou, Hang Yang, Ying Wang, Jian Yang, et al.

To cite this version:

Yigong Zhang, Yicheng Gao, Shuo Gu, Yubin Guo, Minghao Liu, et al.. Build your own hybrid thermal/EO camera for autonomous vehicle. ICRA 2019 - IEEE International Conference on Robotics and Automation, May 2019, Montreal, Canada. hal-02051880

HAL Id: hal-02051880

https://inria.hal.science/hal-02051880v1

Submitted on 28 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Build your own hybrid thermal/EO camera for autonomous vehicle

Yigong Zhang[1], Yicheng Gao[1], Shuo Gu[1], Yubin Guo[1], Minghao Liu[1], Zezhou Sun[1], Zhixing Hou[1], Hang Yang[1], Ying Wang[1], Jian Yang[1,*], Jean Ponce[2] and Hui Kong[1,3,**]

***Abstract*— In this work, we propose a novel paradigm to design a hybrid thermal/EO (Electro-Optical or visible-light) camera, whose thermal and RGB frames are pixel-wisely aligned and temporally synchronized. Compared with the existing schemes, we innovate in three ways in order to make it more compact in dimension, and thus more practical and extendable for real-world applications. The first is a redesign of the structure layout of the thermal and EO cameras. The second is on obtaining a pixel-wise spatial registration of the thermal and RGB frames by a coarse mechanical adjustment and a fine alignment through a constant homography warping. The third innovation is on extending one single hybrid camera to a hybrid camera array, through which we can obtain wide-view spatially aligned thermal, RGB and disparity images simultaneously. The experimental results show that the average error of spatial-alignment of two image modalities can be less than one pixel.**

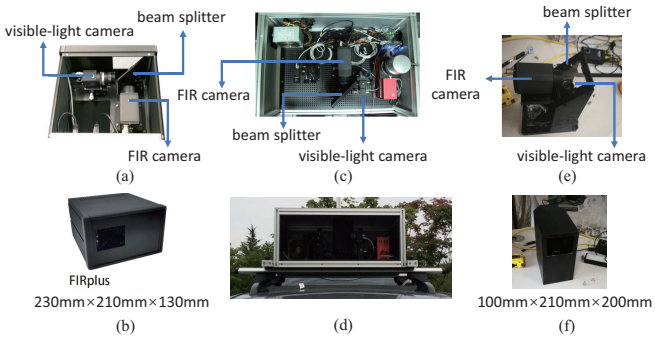


Fig. 1. A comparison of our hybrid camera with the other existing two. Left column: a camera designed by FIRplus[1]. Middle column: A camera designed by KAIST[4], [5], [9]. Right column: our design.

I. INTRODUCTION

Visual sensor fusion is an important research topic in academia and has found its tremendous applications in mobile-robot sensing and navigation, autonomous driving vehicles, and industrial inspection for product-quality control, etc. Among them, the fusion of visible-light camera with the non-visible-spectrum one is probably the most-widely adopted fusion strategy.

This work aims at introducing a new design of hybrid thermal/EO camera. The main driven force behind it is our nocturnal autonomous-driving vehicle project, which requires the vehicle to automatically drive at night with functions such as object detection and recognition, self-localization and obstacle avoidance, etc.

Our expectations on this hybrid camera include two aspects. First, the output thermal and RGB frames of the hybrid camera at any time instance should be pixel-wisely aligned and temporally synchronized as well. This expectation facilitates the tasks of object detection, recognition and pixel-wise semantic segmentation because it can save us tremendous labor in image labeling, training data preparation, and exploiting information fusion if we are seeking the deep-learning based techniques. Second, the dimension of one single hybrid camera should be as small as possible. Considering that a single thermal or EO camera might has a small field of view, we hope to combine more than one hybrid cameras to acquire a wide-FoV (Field of View) hybrid imagery. Thus, the designed hybrid camera should be compact enough to be extendable.

Almost all existing schemes of similar hybrid thermal/EO cameras exploit thermal/EO cameras and one beam splitter to make the output of the thermal/EO cameras spatially aligned. However, we find that none of them can completely satisfy our expectations and requirements. The biggest issue of existing similar hybrid thermal/EO cameras is their dimensions, and all of them are too large to be extendable easily. The major factor inducing their large sizes is that the existing hybrid cameras cannot be detached from their adjusters and dock, which are actually quite large in dimension. To deal with this issue, our camera is specifically designed in three innovative ways so that it is more compact in size, and thus more practical for real-world applications.

Among the three innovations, the first is a redesign of the structure of the hybrid camera, where we are able to minimize of the size of the beam splitter without sacrificing the hybrid-camera's range of sensing field. Meanwhile, we find out a way to detach the thermal/EO cameras from their adjusters. The second is how to obtain a pixel-wise spatial alignment of the thermal/RGB frames, where we first attach the hybrid camera to two tuners (one tuner per camera) to get a rough spatial alignment of the two image modalities. Thereafter, the hybrid camera is detached from the tuners after the thermal and EO cameras are firmly locked onto their dock. To get a fine alignment, the two image modalities can

[1] These authors are with the PCA Lab, Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and Jiangsu Key Lab of Image and Video Understanding for Social Security, School of Computer Science and Engineering, Nanjing University of Science and Technology, China.

[2] Jean Ponce is with the WILLOW team of INRIA/Ecole Normale Superieure, France.

[3] Hui Kong is also with Institute of Advanced Artifical Intelligence in Nanjing, Horizon Robotics, China. He is responsible for the correspondence of this paper. E-mail: konghui@njust.edu.cn, hui.kong@horizon.ai.

[*] The work of Jian Yang was supported by the National Science Fund of China under Grant No. U1713208, Program for Changjiang Scholars.

[**] The work of Hui Kong was supported by the National Natural Science Fund of China under Grant No. 61672287.

(a) (b) (c)

Fig. 2. (a) and (b) A combination of three hybrid cameras. (c) Two hybrid-camera arrays to obtain spatially aligned wide-view thermal, RGB and disparity images simultaneously.

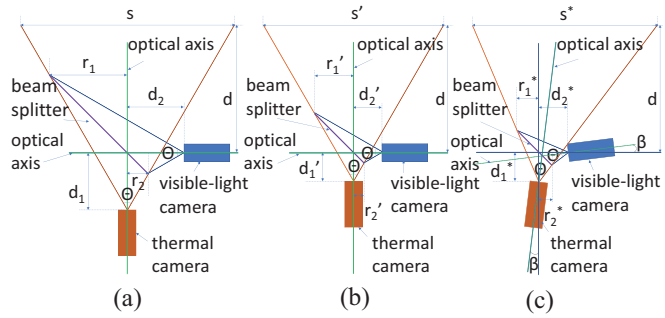


Fig. 3. Calculation of the minimal side length of the beam splitter given the FOV angles of both thermal and EO cameras for three different cases.

be viewed as captured by two cameras which are slightly subject to a rotation. Therefore, we can obtain an accurate pixel-wise spatial alignment of two image modalities by one constant homography transformation induced by the rotation. The third innovation is that we are able to obtain a wide-view output of thermal, RGB and depth images simultaneously, and these images are aligned pixel-wisely. Some demos are showed at `https://www.youtube.com/playlist?list=PLoxxUaKK06F3BVvei-o5plxilxWXQ6Lvi`.

II. RELATED WORKS

In the literature, there are a few works that have used co-axial thermal/EO cameras for their applications, but very few works give detailed introduction on how their hybrid cameras are built.

St-Laurent et al. proposed the method of combining visible and thermal cameras [2], [3]. They used the co-axial structure of visible and thermal cameras, where the axes of the two cameras are identical. The additional glass beam splitter with indium-tin-oxide (ITO) coating was used for the co-axial structure, which can reflect thermal energy while transmitting visible waves. However, the camera viewing angle of thermal and visible cameras is usually large in order to be used for outdoors surveillance systems, which inevitably makes the size of the glass beam splitter very large and consequently the dimension of the system also increases.

Hwang et al. proposed to use their dual-spectrum (thermal/EO) co-axial cameras for pedestrian detection [4], [5], [9]. Their hybrid camera is almost the same as the one in [2]. The size of their camera's splitter is about 14cm. Hwang et al.'s hybrid camera is inspired by the work of Bienkowski et al. [6] who propose the co-axial hybrid thermal/EO camera for enhancing active thermography for nondestructive testing.

Except for the aforementioned hybrid cameras with the co-axial structure, some researchers and manufacturers also designed the hybrid cameras without the co-axial structure, for example, Treible's work [7] and TELOPS Hyperspectral IR Cameras [8]. However, these cameras cannot capture piexel-wisely-aligned RGB/thermal images simultaneously. They are often used to capture the objects in the distance, which is not suitable for driving scenarios.

III. A SINGLE HYBRID CAMERA

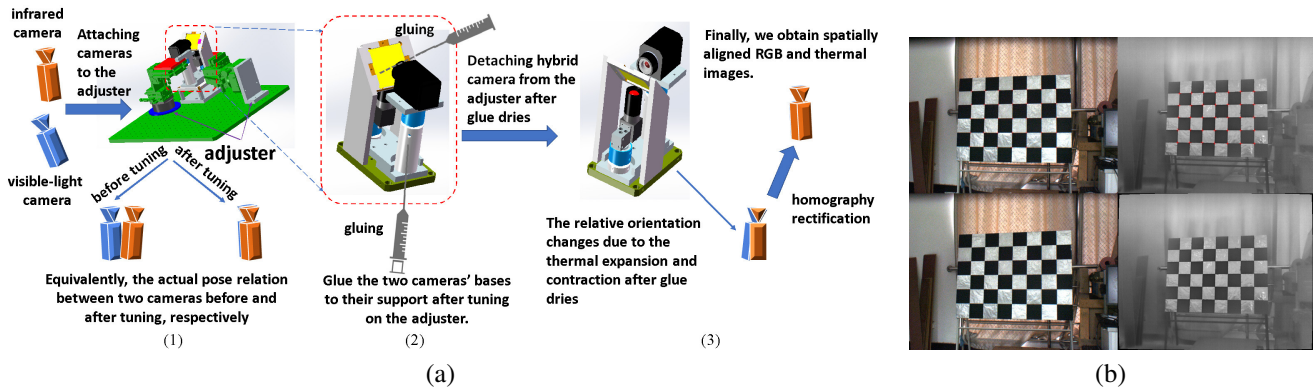
We starts with the design of the hybrid camera's structural layout. To make the hybrid camera smaller, we have changed the layout of the thermal and visible-light cameras. Contrary to the existing designs (e.g., Fig. 1(a) and (c)), we adopt an upright layout of the two component cameras (Fig. 1(e)). The new layout design can make the hybrid camera more compact. Because of its smaller dimension, it can also be beneficial for us to combine two or more hybrid cameras if we hope to obtain panoramic/wide-view dual-spectrum images (e.g., the combinations shown in Fig. 2).

A. How large is the beam splitter?

The thing that affects the dimension of the hybrid camera is the size of its components, i.e., the thermal camera, the EO camera, and the beam splitter. At this moment, we are not able to make the thermal and EO cameras by ourselves. Therefore, the size of beam splitter is the only thing that we can play with to make the hybrid camera as small as possible. Given the specifications of the thermal/EO cameras (e.g., focal length) and their lens (e.g., field-of-view angle), we calculate the expected size of the beam splitter as follows. As shown in Fig. 1 (e), it is straightforward to see that the dimension of the hybrid camera is proportional to the side length of the beam splitter (assuming that the beam splitter has a shape of square). Obviously, we prefer to use a small beam splitter.

Without losing generality, we assume that the thermal and EO cameras have equal FoV angles θ. Although, in our practice, the thermal camera has a larger FoV angle than the EO camera, we take the smaller of them as the common FoV and let θ be 40°. We analyze the three cases shown in Fig. 3, where we assume that the two cameras are equally close to the beam splitter, i.e., $d_1 = d_2$ in Fig. 3(a), $d_1' = d_2'$ in Fig. 3(b) and $d_1^* = d_2^*$ in Fig. 3(c).

If the optical axes of the two cameras are orthogonal to each other, as shown in Fig. 3(a), $\tan(\theta/2) = \frac{r_1}{d_1+r_1}$, so we get $r_1 = \frac{\tan(\theta/2)}{1-\tan(\theta/2)} d_1$. Likewise, $\tan(\theta/2) = \frac{r_2}{d_1-r_2}$, and $r_2 = \frac{\tan(\theta/2)}{1+\tan(\theta/2)} d_1$. Supposing that the splitter is a square, we can obtain its side length as $r = \sqrt{2}(r_1 + r_2) = \sqrt{2}(\frac{\tan(\theta/2)}{1-\tan(\theta/2)} + \frac{\tan(\theta/2)}{1+\tan(\theta/2)}) d_1$. Obviously, the side length of the beam splitter



(a) (b)

Fig. 4. (a) The flowchart of spatially aligning the thermal and visible-light frames of a hybrid camera. (b) Camera calibration and undistortion. Top row: the original RGB and thermal images of a hybrid camera, respectively, overlaid by detected corners. Bottom row: the undistorted RGB and thermal images, respectively..

is proportional to d_1 given the camera's FoV angle θ. If we move both the thermal and EO cameras closer to the beam splitter, as the case shown in Fig. 3(b), so that d_1' of Fig. 3(b) is equal to one half of the d_1 in Fig. 3(a), we can notice a significant reduction in the size of beam splitter with hardly sacrificing the camera's physical sensing field at a certain distance. The side-length of beam splitter in Fig. 3(b) is calculated as

$$r' = \sqrt{2}(r_1' + r_2') = \sqrt{2}\left(\frac{\tan(\theta/2)}{1-\tan(\theta/2)} + \frac{\tan(\theta/2)}{1+\tan(\theta/2)}\right)d_1' \quad (1)$$

Obviously, if $d_1' = \frac{1}{2}d_1$, the splitter's length of Fig. 3(b) is only a half of that of Fig. 3(a). Meanwhile, the width of sensing field at a distance d from the hybrid camera is calculated as $s = 2(d+d_1)\times\tan(\theta/2)$ in Fig. 3(a), and the counterpart $s' = 2(d+d_1')\times\tan(\theta/2)$ in Fig. 3(b). For real-world applications of automated vehicle, d is hundreds of times larger than d_1 and d_1', so the width of sensing field s in Fig. 3(a) is almost equal to s' of Fig. 3(b). Thus, it is an effective way to reduce the size of splitter without affecting the range of actual sensing field by moving both cameras closer to the splitter.

Another way is to change the relative pose between the thermal and EO cameras, shown in Fig. 3(c), where the optical axes of both cameras are not orthogonal. This can be done by expanding the angle between both camera's optical axes via rotation until one of them blocks the other's sight. Assuming that the rotation angle is denoted by β, we have $\tan(\theta/2-\beta) = \frac{r_1^*}{d_1^*+r_1^*}$, and $r_1^* = \frac{\tan(\theta/2-\beta)}{1-\tan(\theta/2-\beta)}d_1^*$. Likewise, we have $\tan(\theta/2+\beta) = \dfrac{r_2^*}{d_1^*-r_2^*}$, and $r_2^* = \frac{\tan(\theta/2+\beta)}{1+\tan(\theta/2+\beta)}d_1^*$. Thus, the beam splitter's side length is $r^* = \sqrt{2}(r_1^*+r_2^*)$,

$$r^* = \sqrt{2}\left(\frac{\tan(\theta/2-\beta)}{1-\tan(\theta/2-\beta)} + \frac{\tan(\theta/2+\beta)}{1+\tan(\theta/2+\beta)}\right)d_1^* \quad (2)$$

where β can be tuned only within a limited range because camera sight can be occluded beyond this range.

Note that $d_1^* = d_1'$, which means that, in comparison to Fig. 3(b), we only tune the directions of both cameras instead of changing the distances from the splitter. The value of the first item of Eq.2 monotonically increases with β, while the second item is a monotonically decreasing function of β. By summing them up, we can obtain a function of β which looks like a parabola, where we change β from 0° to 90°. Apparently, the setting of cameras in Fig. 3(c) is equal to that in Fig. 3(b) when β is 0°. When β is 45°, the beam splitter has the smallest side length in theory. However, the FOV of one camera is fully occluded by the other one. Considering that the largest rotation angle β that can avoid sight occlusion is about 10°, the smallest side length of the beam splitter that we can actually set based on the configuration in Fig. 3(c) is about 0.69 times of the splitter's side length in Fig. 3(b).

In Fig. 3(c), the range of sensing field s^* can be calculated as

$$s^* = (d_1^*+d)\left(tan\left(\frac{\theta}{2}-\beta\right)+tan\left(\frac{\theta}{2}+\beta\right)\right) \quad (3)$$

where d_1^* is ignorable compared with d. When β is 10° and θ is 40°, s^* is $0.75d$. In contrast, the s' in Fig. 3(b) is $0.73d$. Therefore, s^* is $0.02d$ smaller than s'. When d is 5m, the difference is about 0.1m.

By calculating the side length of beam splitter based on the three configurations of Fig. 3, we observe that we can have the smallest beam splitter based on the configuration in Fig. 3(c). However, the configuration in Fig. 3(c) has several disadvantages compared with that of Fig. 3(b). First, it is not easy to make the hybrid camera dock for the configuration in Fig. 3(c). In contrast, we can use some standard parts for Fig. 3(b). Second, the sight view of the hybrid camera in Fig. 3(c) is not symmetric. Third, the sensing range is smaller compared with Fig. 3(b). Based on these observations, we make our hybrid camera based on Fig. 3(b) to calculate the size of splitter.

B. How to align the thermal/EO frames pixelwisely?

We introduce the procedure of aligning the thermal/EO camera frames based on Fig. 4(a).

1) Coarse alignment by manual adjustment: First, the two cameras are attached to their docks, which are attached to their 5-DOF adjusters, respectively. Next, we estimate the

intrinsic parameters of the two cameras with a large calibration chessboard by using the Camera Calibrator Application in MATLAB. The chessboard is made by sticking two layers of materials onto a pure-flat board, where the bottom layer is a large piece of paper printed with chessboard pattern, and we cover the white-region of the bottom layer with foil paper. During camera calibration, we blow the completed chessboard with cold air, and we can see clear chessboard patterns in both thermal and RGB frames, as shown in Fig. 4(b). With the intrinsic parameters of individual camera, we remove the distortion in images. Because the EO camera has a higher resolution than the thermal camera, we resize either of them to make their image pixels have equivalent scales. We first tuning the adjusters to align the principal points (i.e. optical axes) of both cameras.

Before adjusting, the optical axes of the two cameras are generally not coincident. After tuning the adjusters, both optical axes of cameras should be aligned very well (as shown in the Fig. 4(a)(1)). Before we detach the cameras from the adjuster platform, we glue both camera docks to their support (Fig. 4(a)(2)). The hybrid camera is detached from the adjuster until the glue dries completely. Unexpectedly, we found that the thermal and RGB frames are not perfectly aligned due to the effect of thermal expansion and contraction after the glue dries, but the misalignment is not that serious at this moment. In the following, we give a warping step to refine the alignment.

2) Fine alignment by a constant homography warping: Let P be a 3D point. $p_1 = (u_1, v_1, 1)^T$ and $p_2 = (u_2, v_2, 1)^T$ are the projections of P onto the EO and thermal camera, respectively. $P_1 = (x_1, y_1, z_1)^T$ and $P_2 = (x_2, y_2, z_2)^T$ are the corresponding 3D coordinates in the EO and thermal camera coordinate systems for P. We have the following equations,

$$\begin{cases} z_1 \cdot p_1 = K_1 \cdot P_1 \\ z_2 \cdot p_2 = K_2 \cdot P_2 \end{cases} \tag{4}$$

where K_1 and K_2 are the intrinsic parameters of the EO and thermal cameras, respectively. If the two cameras are well aligned, $z_1 = z_2$ and $P_1 = P_2$, the (4) can be simplified as

$$p_1 = K_1 \cdot K_2^{-1} \cdot p_2 \tag{5}$$

This describes a transformation between thermal and EO cameras. We can warp the thermal image to the RGB one based on the constant transformation $K_1 \cdot K_2^{-1}$, or warp the RGB image to the thermal one based on $K_2 \cdot K_1^{-1}$.

As mentioned in the coarse alignment step, we believe that, equivalently, there exists a small transformation (a small rotation plus a very small translation) between the thermal and EO cameras after detachment in contrast to the moment before detachment. Next, we take pictures of our calibration chessboard with the hybrid camera (after detached from the adjuster platform). We can obtain the extrinsic parameters between the thermal and EO cameras, i.e., the small rotation R and translation t between the two cameras. Let the transformation between P_1 and P_2 be $P_1 = R \cdot P_2 + t$. The (4) can be rewritten as,

$$p_1 = z_1^{-1} \cdot z_2 \cdot K_1 \cdot R \cdot K_2^{-1} \cdot p_2 + z_1^{-1} \cdot K_1 \cdot t \tag{6}$$

In general, considering the fact that the shift (translation) between two cameras is very small, and the depth z_1 is much larger than t, we can assume that $z_1^{-1} \cdot z_2 \approx 1$ and $t \approx 0$. The (6) can be approximated as follows,

$$p_1 \approx K_1 \cdot R \cdot K_2^{-1} \cdot p_2 \tag{7}$$

Therefore, we only need to estimate this rotation transformation, and apply a constant homography transformation, which is derived from this rotation, to warp between the thermal and RGB frames to make their alignment accurate.

Fig. 5 shows examples of camera alignment, where we have selected three sets of hybrid cameras (shown in Fig. 2) to exhibit the effect of pixel-wise alignment. In the first row of Fig. 5, we show the result of each hybrid camera before homography warping, respectively. In the second row, we shows the corresponding alignment results after homography warping. More explanation is given in the figure's caption.

IV. Beyond a single hybrid camera

For applications in intelligent vehicle society, it is a rich supplement to the existing road-scene image datasets (e.g. KITTI) if we are able to create a wide-view thermal/EO image database that contains pixel-wisely registered image frames. Therefore, we aim at creating such a kind of dataset which includes thermal/RGB images of varieties of road scenes. To achieve it, we extend our one single hybrid camera to build an array of six hybrid cameras (the last image of Fig. 2). Specifically, we first build two sets of hybrid-camera arrays, where each array consists of three hybrid cameras, as shown in the first two images of Fig. 2, and one hybrid-camera array is stacked on top of the other. Via image stitching and synchronization, we are able to obtain wide-view pixelwisely aligned thermal and RGB frames from both the top and bottom camera arrays. Further, we are able to acquire wide-view depth images by stereo matching of the top and bottom stitched images. The flowchart of generating these images are shown in Fig. 6. Note that we do not directly stitch the output of the top hybrid camera array, and the output of the bottom hybrid camera array. Before stitching, we apply stereo rectification to the top-left and bottom-left hybrid cameras, the top-middle and bottom-middle hybrid cameras, and the top-right and bottom-right cameras, respectively. We stitch the rectified images (both thermal and RGB) into wide-view ones. Finally, we do the stereo matching to obtain the disparity map. All outputs are pixel-wisely aligned.

V. Experiments

A. Hardware systems and camera specs

Our single hybrid camera consists of one visible-light EO camera and one thermal camera (Fig. 1). The built-in cameras support the Ethernet interface, so these cameras can be controlled from eBUS protocol. The EO camera is the PointGrey Flea3 color camera (FL3-GE-13S2C-C), with the

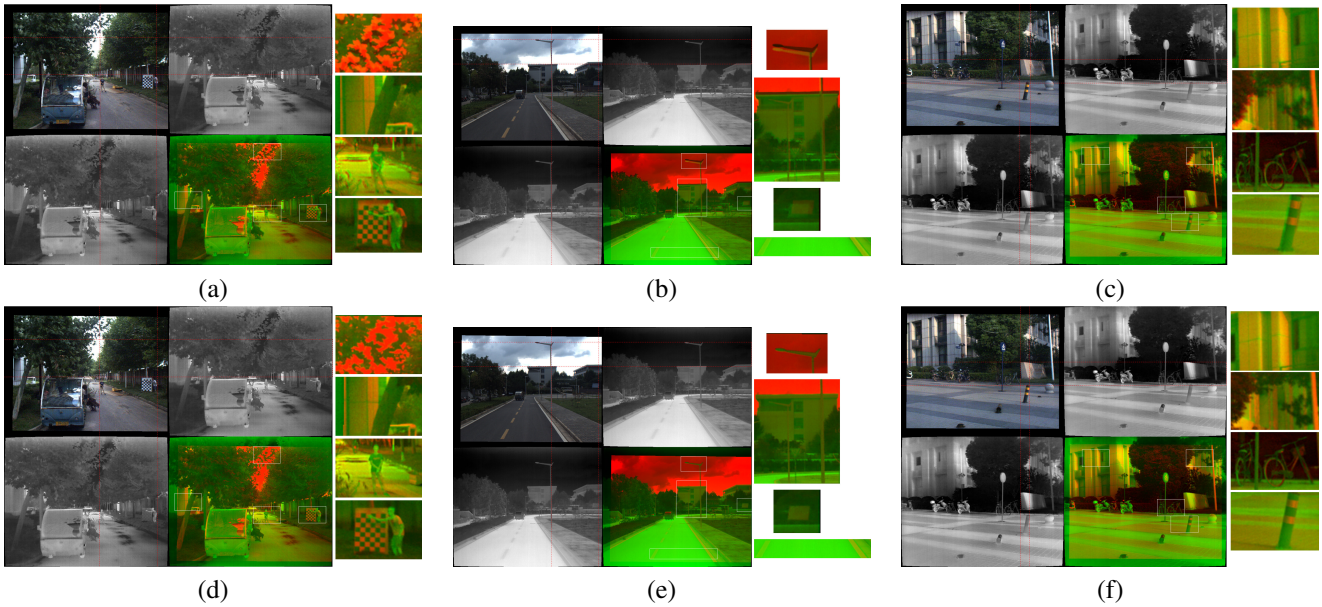


Fig. 5. Results of alignment between the thermal and RGB frames of three sets of hybrid cameras before and after homography warping, respectively. (a), (b) and (c) are the alignment results before the homography warping, respectively. In each sub-figure, the layout of images is arranged as follows. Top-left: the aligned RGB image. Top-middle and bottom-left: the same aligned thermal image. Bottom-middle: the fusion image. (d), (e) and (f) are the alignment results after the homography warping, respectively. Likewise, the layout of images in each sub-figure is the same as those of (a), (b) and (c). To show the effect of homography rectification, we have overlaid red dotted lines horizontally and vertically onto the each sub-figure. In addition, the right column of each sub-figure zooms in four selected image regions to help us to view the warping result.

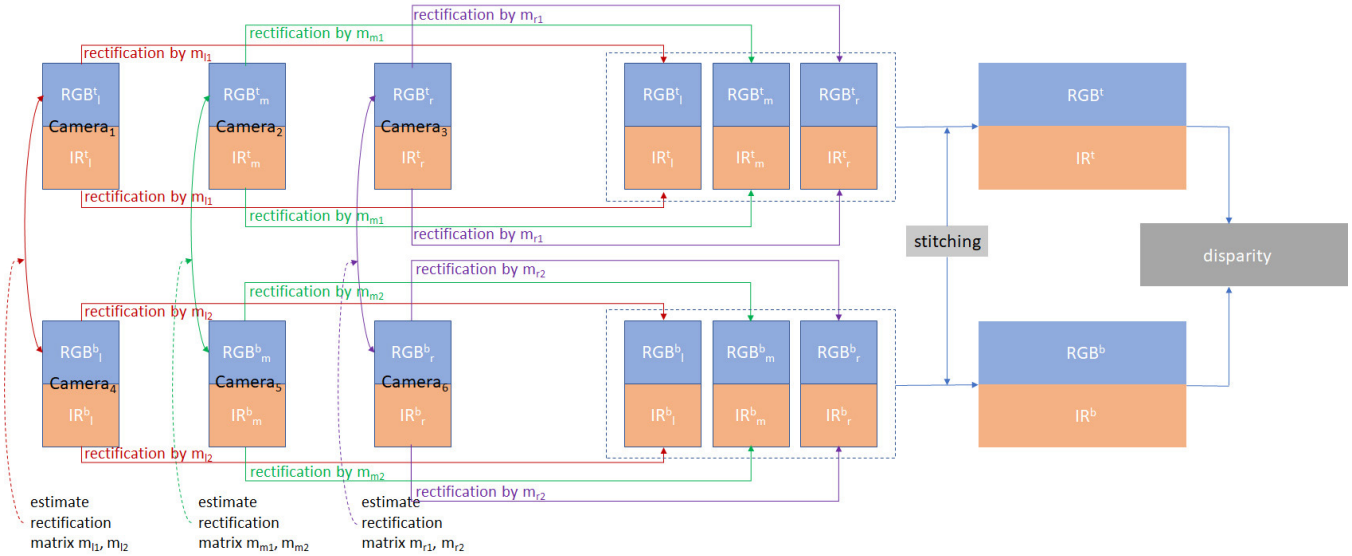


Fig. 6. The flowchart on how to obtain the pixel-wisely aligned wide-view RGB, thermal and disparity maps.

Fig. 7. The output pixel-wisely aligned wide-view RGB, thermal and disparity images at a certain time instance. Left column: the output of the top camera array. Middle column: the output of the bottom camera array. Right column: the top one is the disparity map obtained only based on the RGB image, and the bottom one is obtained based on the RGB and thermal frames.

TABLE I

EVALUATION FOR THE PIXEL-WISE ALIGNMENT BASED ON THE SYMMETRIC TRANSFER ERROR.

Hybrid camera	1	2	3	4	5	6
RGB to thermal (pixel)	0.77	0.61	0.60	0.80	0.50	0.77
thermal to RGB (pixel)	1.62	1.31	1.30	1.69	1.05	1.63

largest output image size of 1280×960 at the wavelength of 400∼700nm. The thermal camera is the FLIR A65 camera, with the output image size of 640×512 at the wavelength of 7.5∼13μm. The side length of the beam splitter of our camera is about 6.5cm. The FoV of the EO and thermal cameras are about 40° and 45°, respectively. The dimension of a single hybrid camera is 10cm×21cm×20cm. The two cameras are synchronized using an external sync interface. The frame rate of both cameras is set to 29.97fps due to the demand of the thermal cameras. The size of the stitched wide-view RGB and thermal images is 2881×993. The horizontal FOV of the camera array is about 100°. The baseline between the top and bottom array is about 30cm. The extend camera array consists of six hybrid cameras, and connected to a high-performance server via Ethernet. The total one-second storage is more than 1G bytes.

B. Accuracy estimation for pixel-wise alignment

We first evaluate the accuracy of the pixel-wise alignment for all the six hybrid cameras. We place three large calibration chessboards in different poses and use each hybrid camera to capture 20 images in various views. To ensure the accuracy of corner detection, all the three chessboards are fully displayed in the images in a sufficient size. To quantify the accuracy of the pixel-wise alignment, The symmetric transfer error is used: one is the average distance between the corners detected in the RGB images and the corners projected from the thermal images, and the other is the average distance between the corners detected in the thermal images and the corners projected from the RGB images. The experimental results are listed in Table I. It indicates that the accuracy of our system is high enough to ensure the pixel-wise alignment between the thermal and RGB images, which is suitable for various applications, such as detection, segmentation, recognition, etc. Because the size of RGB images are much larger than the one of thermal images, the symmetric transfer errors from thermal images to RGB images are also larger than the ones from RGB images to thermal images.

C. Running time and performance

In this subsection, we evaluate the running time for homography warping, stitching and disparity computation. We conduct all the experiments on a general PC with a 16G memory and a CPU of Intel Core-i7 8550U. We first evaluate the running time for the homography warping between thermal images and RGB images. It only spends 6ms to transform one thermal image to the space of RGB image. In practice, with a fixed assembly of our system, all the parameters used for undistortion, homography warping and rectification are constant, so that a look-up table can be generated to complete all the transformations mentioned above. It also spends about 6ms to complete all of these transformations for one image.

Next, we evaluate the running time for stitching all 6 RGB images and 6 thermal images in a frame. The average running time for stitching 6 RGB images and 6 thermal images is about 2.92s. Due to the pixel-wise alignment between RGB images and thermal images, the same seam lines can be shared, which saves a lot of time.

We also test the running time of the disparity computation. It costs about 1.8s to compute a disparity map from two stitched RGB images and two stitched thermal images. A sample result is showed in Fig. 7, the upper RGB and thermal images are listed in the first column. The lower RGB and thermal images are listed in the second column. The top image in the third column is the disparity map obtained by only the two RGB images, and the bottom one is the disparity map obtained by both the RGB and thermal images. It shows that it could recover the disparity information in the region with extreme illumination by using infrared images, such as the street lamp in the sky region in Fig. 7.

Note that our code is almost based on OpenCV 2.4.13 and LIBELAS[10] with little modifications in the multi-thread and pipeline techniques without using GPU. If the size of stitched images is resized to 30% of the original size, the same code can run in about 12fps on a server with an Intel Xeon E5 CPU of 20 cores and 40 threads (i.e. the server used for acquiring data).

VI. CONCLUSION

In this work, we propose a novel paradigm to design a hybrid thermal/EO camera, whose thermal and RGB frames are pixel-wisely aligned and temporally synchronized. We innovate in three ways to make it more compact in dimension, and more practical and extendable for real-world applications. The experimental results show that the average error of spatial-alignment of two image modalities can be less than one pixel.

REFERENCES

[1] FIRplus: http://www.viewplus.co.jp/product_viewplus_products/18_index_detail.html

[2] L. St-Laurent et al., Thermal imaging for enhanced foreground-background segmentation, International Conference on Quantitative Infrared Thermography, 2006.

[3] L. St-Laurent et al., Combination of colour and thermal sensors for enhanced object detection, International Conference on Information Fusion, 2007.

[4] S. Hwang, et al., Low-Cost Synchronization for Multispectral Cameras, International Conference on Ubiquitous Robots and Ambient Intelligence, 2015.

[5] S. Hwang, et al., Multispectral Pedestrian Detection: Benchmark Dataset and Baseline, CVPR, 2015.

[6] L. Bienkowski, et al., Hybrid camera and real-view thermography for nondestructive evaluation, Quantitativ InfraRed Thermography, 2012.

[7] W. Treible, et al., CATS: A Color and Thermal Stereo Benchmark, CVPR, 2017.

[8] TELOPS Hyperspectral IR Cameras: http://telops.com/products/hyperspectral-cameras

[9] Y. Choi, et al., KAIST multi-spectral day/night data set for autonomous and assisted driving. IEEE Transactions on Intelligent Transportation Systems, 19(3), 2018: 934-948.

[10] A. Geiger, et al., Efficient Large-Scale Stereo Matching, ACCV, 2010.