
# A Novel Cross-Resolution Image Alignment for Multi-camera System

Kuo Chen[1(✉)], Tianqi Zheng[1], Chenxing He[1], and Yeru Wang[2]

[1] Chongqing University of Posts and Telecommunications, Chongqing 400065, China
chenkuo@cqupt.edu.cn
[2] Hangzhou Dianzi University, Hangzhou 310018, ZJ, China

**Abstract.** In complicated computer vision tasks, the multi-camera system is more effective than a single camera owing to the image fusion of multiple cameras, in which the image alignment is the essential first step. Especially, the cross-resolution image alignment caused by focal length difference has been extensively studied. A usual solution is using pyramid based local feature matching to create the mapping between input images with high and low-resolution. However, this kind of algorithm has high time and space complexity and is not applicable to front-end equipment such as the multi-camera system. Therefore, this paper proposes a fast and novel cross-resolution image alignment method based on the approximate focal length difference, including the coarse feature matching in low-resolution and the matching model estimation in high-resolution. At last, two types of test experiments are carried out using the industrial camera and SLR camera respectively. And the experimental results show that the proposed method performs well for cross-resolution image alignment, which can be widely used for multi-camera system.

**Keywords:** Local feature matching · Image alignment · Multi-camera imaging

## 1 Introduction

Powered by computer vision and artificial intelligence technology, more visual information and higher-quality images can be obtained by multi-camera system [1], which has been widely used in medicine, industry, remote sensing and other fields. By reasonably fusing the image information between different cameras, higher-quality composite images can be obtained, and panoramic imaging, high dynamic range imaging, extend depth of field imaging, night vision and so on can be realized with multi-camera system [2, 3].

In the process of image fusion between different cameras, the image alignment is the essential first step, which is limited by the possible differences between cameras, such as attitude difference, focal length difference and spectral difference [4]. Currently, the general solutions are pixel based, patch based and local feature based methods [5], among which the most flexible one is local feature based matching. It has a robust alignment effect.


Y. Chenggang et al. (Eds.): MobiMedia 2022, LNICST 451, pp. 3–14, 2022.
https://doi.org/10.1007/978-3-031-23902-1_1

This paper focuses on cross-resolution image alignment, which is image alignment between the long-focus camera and the short-focus camera. For this problem, benefiting from the image pyramid technology, accurate feature points can be extracted on the cross-resolution images by feature detection algorithms such as Scale-Invariant Feature Transform [6] (SIFT), Speeded-Up Robust Features [7] (SURF) and Oriented Fast and Rotated Brief [8] (ORB). Then the final homography matrix can be estimated by using inlier filtering algorithms such as Random Sample Consensus [9] (RANSAC).

However, in order to achieve scale invariance, the existing technologies have designed a complex feature description method, which leads to high time and space complexity. Therefore, this paper studies a novel cross-resolution image alignment method.

## 2 Related Work

To realize cross-resolution image alignment, local feature detection and outlier removal technologies are involved. This paper briefly discusses the related works of these two aspects.

For local feature detection, many excellent algorithms have been developed. For example, SIFT detects the sub-pixel feature point using difference of Gaussians on pyramids with different scales, and generates the robust feature descriptor by gradient histogram. But its time complexity and space complexity are extremely high. Therefore, SURF uses box filtering, wavelet transform, lower dimensional vector and other techniques to reduce the complexity of SIFT. But there are still some shortcomings in real-time performance. While ORB effectively combines the detection of local extremum feature and binary descriptor and realizes real-time local feature matching. However, to deal with the cross-resolution problem, a complicated pyramid structure is still required.

In the aspect of outlier removal, a global transformation is usually used to judge inliers or outliers, such as RANSAC, in which similar transformation, affine transformation, projective transformation and so on can be used. There are also a series of later improvements of RANSAC [10, 11]. Or the probability of sampling inliers is increased, the speed of iteration is accelerated, and the threshold of inlier discrimination is optimized, etc. In addition, with the development of deep learning technology, there are many excellent feature detection and outlier removal algorithms based on deep learning, such as LF-Net [12], Superpoint [13] and SuperGlue [14], which can only achieve high real-time performance on GPU at present, and cannot meet the application requirements of multi-camera system which needs edge computing.

Inspired by the related works, this paper focuses on the image alignment of long-focus camera and short-focus camera. According to the approximate ratio of spatial resolution between multi-camera images, the high-resolution images are downsampled to the low-resolution scale to realize coarse matching. Then the feature points are filtered and matched iteratively on the high-resolution scale to achieve final accurate alignment.

## 3 The Proposed Method

As shown in Fig. 1, a narrow field-of-view image with high-resolution is captured by the long-focus camera, defined by $I_s$, and a wide field-of-view image with low-resolution is

captured by the short-focus camera, defined by $I_r$. Considering the difference in image resolution caused by focal length, this paper designs a method consisting of matching feature point on the low-resolution scale and estimating the alignment model on the high-resolution scale. It is worth mentioning that in the practical application of multi-camera imaging, the non-overlapping areas of images $I_s$ and $I_r$ can be marked in advance to reduce their influence on the final alignment. The approximate ratio of spatial resolution between images caused by focal length can be estimated by geometric measurement and defined by integer $k$.

In this paper, the image $I_s$ is downsampled to the image $I_{s\_d}$, and the feature points are extracted without complex pyramid structure, as shown in Fig. 1. In order to realize more accurate alignment result, the feature points of the image $I_{s\_d}$ are restored to a high-resolution scale after the coarse feature matching, and then the alignment model is estimated in high-resolution.

Normally, restoring feature points from low-resolution scale to high-resolution scale involves two aspects: image texture and spatial coordinate. For the former, it still requires reconstructing the missing image information under high-resolution, which consumes a lot of time and space. Therefore, this paper only restores the spatial coordinates of feature points to high-resolution, which greatly reduces the time and space complexity. And then finds the best alignment in the alignment model estimation stage. For specific implementation methods and steps, please refer to Sect. 3.1 and Sect. 3.2 respectively.

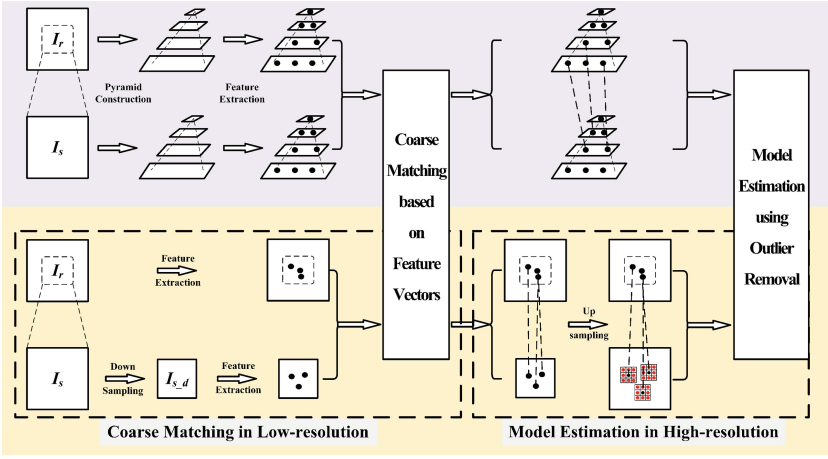


**Fig. 1.** The flow chart of proposed cross-resolution image alignment. The upper half is the pyramid-based solution, and the lower half is the proposed solution.

### 3.1 Coarse Matching in Low-Resolution

Firstly, the high-resolution image $I_s$ is downsampled by $k$ to obtain the low-resolution image $I_{s\_d}$, where $k$ is an integer. And the downsampling process does not need interpolation and has high time efficiency. Then, on the low-resolution scale, the feature points

on image $I_{s\_d}$ and image $I_r$ can be detected using any one fast detection operators without scale invariance. And then the feature point sets $P_{s\_d}$ and $P_r$ are obtained respectively. Although the true resolution ratio of images $I_s$ and $I_r$ is not equal to integer $k$ absolutely, which may cause mismatch in sets $P_{s\_d}$ and $P_r$, it can be compensated in Sect. 3.2.

According to the practical application of multi-camera system, descriptor without scale can be used flexibly to get the feature vectors $V_{s\_d}$ and $V_r$ corresponding to the feature point set $P_{s\_d}$ and $P_r$. Following the general steps of feature matching, the matching quality is measured by the ratio of maximum and submaximum of feature vector distance, and the search process is accelerated by kd-tree, and then the matched point pair $\left\{P'_{s\_d}, P'_r\right\}$ of image $I_{s\_d}$ and image $I_r$ is obtained. It only represents the correspondence of input images on the low-resolution scale. So, in this paper, the feature point $P'_{s\_d}$ of image $I_{s\_d}$ is upsampled $k$ times on the spatial coordinate to get $P'_s$, as shown in Fig. 2, which represents the coarse matching of images on the high-resolution scale.

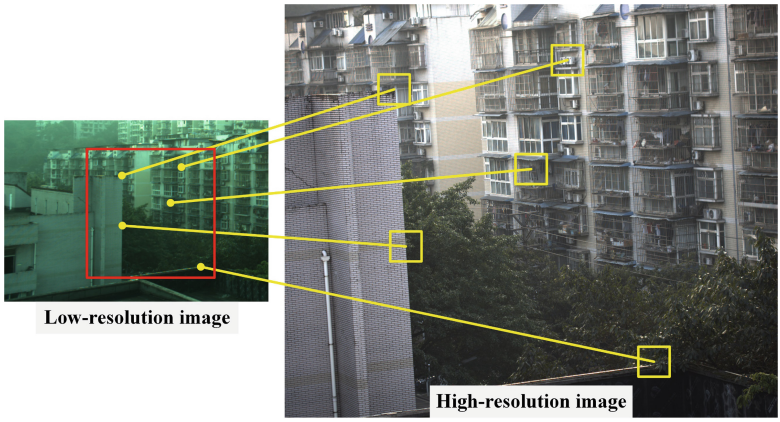


**Fig. 2.** Coarse matching result. A feature point on low-resolution image is corresponding to a patch on high-resolution image.

### 3.2 Model Estimation in High-Resolution

As shown in Fig. 2, due to the difference of spatial resolution, coarse matching in low-resolution establishes the correspondence between a feature point of $I_r$ and a patch of $I_s$. In order to estimate the accurate alignment model, it is necessary to find the pixel-wise correspondence between high-resolution image $I_s$ and low-resolution image $I_r$. Therefore, this paper designs the following iterative algorithm to estimate the alignment model.

**Algorithm1**

**Input:**
  $I_r$: Wide field-of-view image with low-resolution
  $I_s$: Narrow field-of-view image with high-resolution
  $k$: Approximate ratio of spatial resolution
**Function:**
  Downsample high-resolution image
  Feature detection using FAST and others
  Feature descriptor using BRIEF and others
  Get the initial matched point pair $\{P_{s\_d}^{'}, P_r^{'}\}$
  Upsample feature point $P_{s\_d}^{'}$ to $P_s^{'}$
  **for** $i$=1 to $T$ **do**
    Select randomly $m$ feature points in $P_r^{'}$, and the corresponding $m$ patches in $P_s^{'}$
    Select one feature point for each patch
    Estimate the alignment model $H_i$
    Find all inlier sets $P_{in}$ and projection error $E_i$
    **If** the number of $P_{in}$ is large enough and the value $E_{in}$ is small enough
      Update current alignment model $H_i$
      Update current projection error $E_i$
      Update $T$
**Output:**
  $H$: Optimal cross-resolution alignment model

In order to solve the optimum solution, this paper designs an iterative approach to estimate the alignment model inspired by RANSAC. Specifically, firstly randomly select $m$ matched points on the low-resolution image $I_r$ and corresponding $m$ patches on the high-resolution image $I_s$. And then estimate an alignment model $H_i$ according to these $m$ matching relationships. Decide whether all feature points meet the current mapping relationship of model $H_i$, except the selected m points, and obtain the inlier set $P_{in}$ and the projection error $E_{in}$ based on a threshold. Because a feature point in the low-resolution image $I_r$ is corresponding to a patch in the high-resolution image $I_s$, this paper selects the smallest projection error point in a patch as the inlier point. After getting the set of inliers, decide whether to update the alignment model according to the inliers number and projection error until the end of the iteration process, and output the optimal cross-resolution alignment model $H$.

## 4 Experiments

In order to verify the validity of the proposed method, this paper designs three test experiments. In test experiment No. 1, a multi-camera system is built, including a 5-megapixel Daheng industrial camera with an 8 mm focal length lens, and a 12-megapixel

Daheng industrial camera with a 16 mm focal length lens. Five groups of wide field-of-view images with low-resolution and narrow field-of-view images with high-resolution images are captured, with resolution ratio 7.4, named as H1, H2, H3, H4, and H5. In test experiment No. 2, a Nikon D7100 SLR camera with an 18–105 mm zoom lens is used. Two groups of cross-resolution test images with resolution ratio 3 are captured, named as D1 and D2. Two groups of cross-resolution test images with resolution ratio 3.89 and 4.17 are captured, named as E1 and E2. Two groups of cross-resolution test images with resolution ratio 5.11 are captured, named as F1 and F2. Two groups of cross-resolution test images with resolution ratio 5.83 are captured, named as G1 and G2. In test experiment No. 3, we used the SLR camera in Experiment No. 2 to capture images of the same resolution with slight displacement, rotation, etc., and then obtain 9 sets of images with resolution differences of 1–9 times by downsampling, named as A.

### 4.1 Implementation

In the specific implementation process, this paper selects the simple FAST feature extraction operator [15] and BRIEF feature descriptor [16]. And other two cross-resolution image alignment algorithms are compared in this paper. The specific descriptions are as follows. To verify the effect of model estimation in Sect. 3.2, Algorithm 1 only extracts feature points and estimates the alignment model on low-resolution scale. Algorithm 2 is the pyramid based ORB algorithm. Algorithm 3 is the proposed method. Considering the comparability of experimental results, the above three algorithms keep the same parameters in the processes of feature extraction, coarse feature matching, outlier removal and iteration.

In addition, although the real resolution ratio of high-resolution and low-resolution images are mostly decimals, the integer $k$ is always used for upsampling and downsampling in our experiment. This can further save time and space cost in particular, and the parameter $k$ has a certain tolerance and has no significant impact on the final alignment result. As shown in Fig. 3, the resolution ratio of the test image is 5.83 times. When the parameter $k$ is set as 4, 5, 6, 7 and 8, the alignments can still be successful. And the closer the integer $k$ is to the real resolution ratio, the better the alignment effect is, which can be seen from the fused image and the distribution of inlier points.

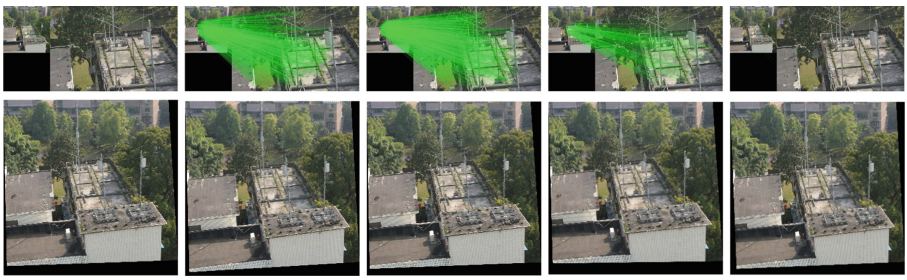


**Fig. 3.** Influence of approximate ratio $k$ on alignment result

### 4.2 Results

The results of test experiment No. 1 are shown in Fig. 4 and Table 1. Considering the limited space for this article, Fig. 4 only shows the image alignment results of groups H1 and H5, and Table 1 shows the final projection errors of all five groups.

It can be seen from Fig. 4 that Algorithm 2 and our method both have a better effect in detail than Algorithm 1, such as the outdoor region of the air conditioner, in which Algorithm 1 has obvious artifacts. We know that Algorithm 1 directly performs feature extraction and matching on low-resolution scale, that results in inadequate accuracy of feature points. While the proposed model estimation in high-resolution can fix this problem.

Furthermore, the projection error of Algorithm 1 is much bigger than Algorithm 2 and our method, and our method perform better than Algorithm 2 slightly as shown in Table1. Since our method does not utilize the pyramid structure to extract feature points, it can reduce time and space consumption in feature extraction phase compared with Algorithm 2.

The results of test experiment No. 2 are shown in Fig. 5 and Table 2. Figure 5 only shows the image alignment results of groups D2, F2 and G1, and Table 2 shows the final projection errors of all eight groups.

As shown in Table 2, the projection error of Algorithm 1 increases along with the increase of the resolution ratio, while algorithm 2 and proposed method are not influenced. Again, the proposed method performs best. With the increase of the resolution ratio, the measuring accuracy of feature points extracted by Algorithm 1 becomes more and more insufficient, so the corresponding projection error increases. However, Algorithm 2 handles this problem using complex pyramid structure. And the proposed method estimates the alignment model in high-resolution iteratively, so it succeed in cross-resolution image alignment with different resolution ratio.

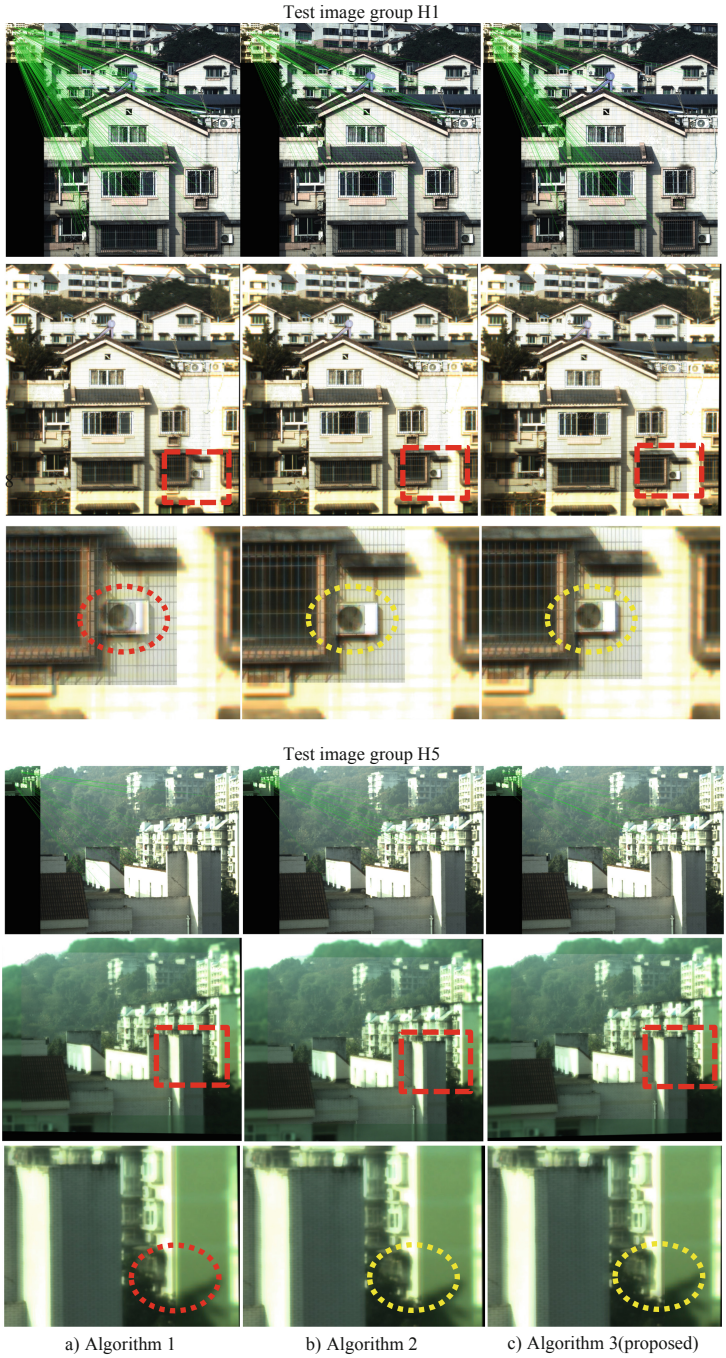


**Fig. 4.** Results of test experiment No. 1

**Table 1.** Projection errors of test experiment No. 1

| Resolution ratio | Test group | Algorithm comparison | | |
|---|---|---|---|---|
| | | Algorithm 1 | Algorithm 2 | Algorithm 3 (proposed) |
| 7.4 | H1 | 5.97 | 2.96 | **2.35** |
| | H2 | 8.49 | **3.33** | 3.88 |
| | H3 | 4.44 | 2.66 | **2.05** |
| | H4 | 7.12 | 3.41 | **3.40** |
| | H5 | 7.45 | 2.71 | **1.97** |

**Table 2.** Projection errors of test experiment No. 2

| Resolution ratio | Test group | Algorithm comparison | | |
|---|---|---|---|---|
| | | Algorithm 1 | Algorithm 2 | Algorithm 3 (proposed) |
| 3.00 | D1 | 2.20 | 2.37 | **1.85** |
| 3.00 | D2 | 4.52 | 1.82 | **1.75** |
| 3.89 | E1 | 2.82 | 2.20 | **1.70** |
| 4.17 | E2 | 4.64 | 2.64 | **2.18** |
| 5.11 | F1 | 4.52 | 3.09 | **2.24** |
| 5.11 | F2 | 5.21 | 2.67 | **1.23** |
| 5.83 | G1 | 6.18 | 3.05 | **1.30** |
| 5.83 | G2 | 6.11 | **2.55** | 3.37 |

The results of test experiment No. 3 are shown in Fig. 6. In order to better explore the relationship between the projection error and the resolution magnification, simulation experiments are conducted. Figure 6 shows the registration error results of 9 groups of experiments with the resolution multiplier from 1 to 9 under image A.

As shown in Fig. 6, the abscissa of the figure represents different resolution magnification, and the ordinate represents projection error. It can be clearly seen in the line chart that the projection error of Algorithm 1 presents an upward trend with the improvement of resolution. However, the projection error of Algorithm 2 and the proposed method is not affected by the resolution multiplier. Meanwhile, the projection error line of the proposed method is below the projection error line of Algorithm 2. This is exactly consistent with the results of Experiment No. 2, thus verifying the correctness and effectiveness of our method.

Although the test images have a large difference in spatial resolution, the experimental results show that the proposed method can always obtain a good alignment effect without the complex pyramid structure. Even in some tests, as shown in Fig. 5 and Fig. 6,

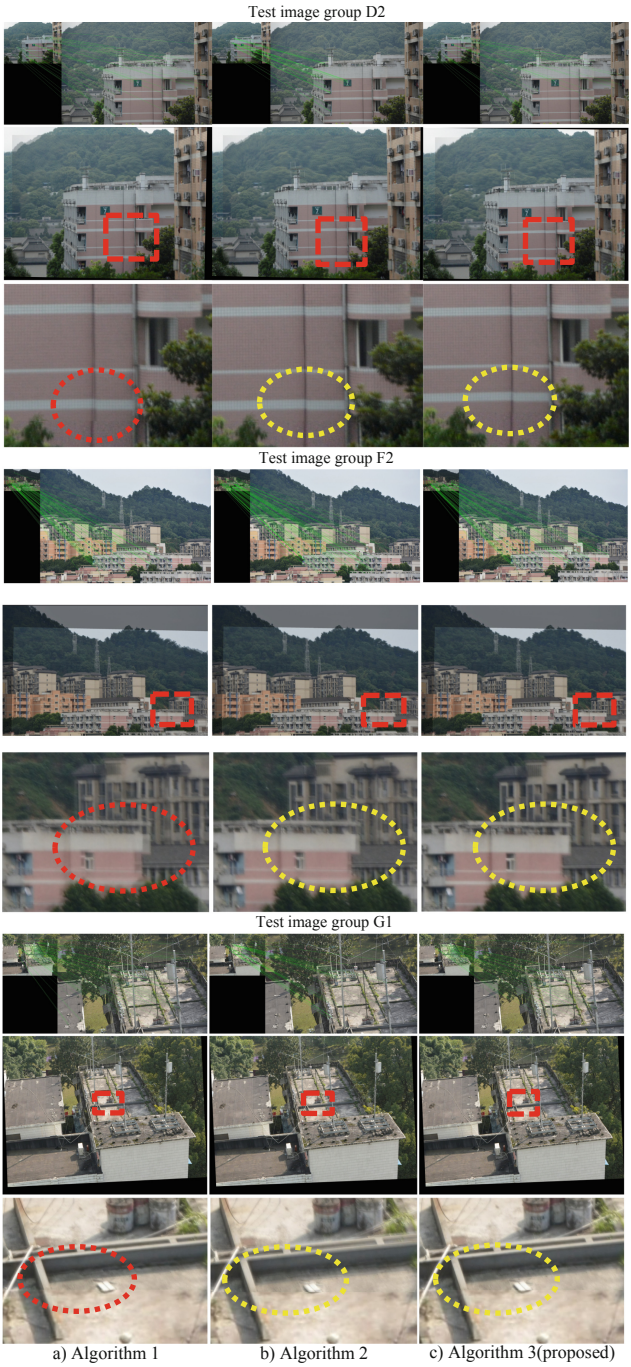


**Fig. 5.** Results of test experiment No. 2

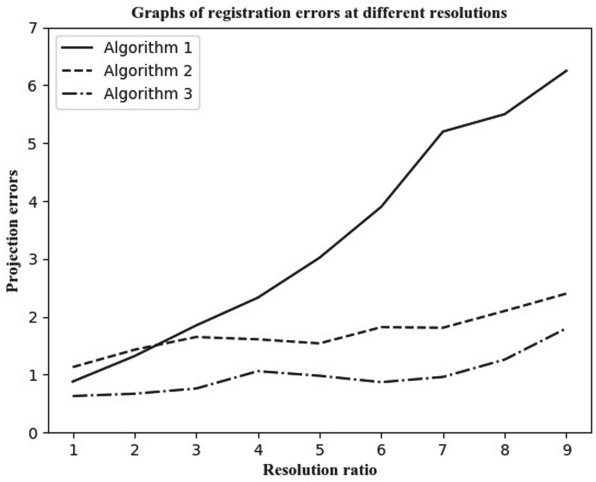


**Fig. 6.** Results of test experiment No. 3

the proposed method performs better than the pyramid based ORB method. The precondition is getting the approximate resolution ratio of high-resolution and low-resolution image, which is very easy to measure in multi-camera system applications.

## 5 Conclusion

According to the problem that the current pyramid structure has high time and space complexity and is not applicable to the multi-camera system, this paper proposed a novel cross-resolution image alignment method, including the coarse feature matching in low-resolution and the matching model estimation in high-resolution. Based on our method, combined with artificial intelligence technology, the interesting computational imaging on the multi-camera system like high-quality continuous zooming, wide field-of-view imaging with high-resolution and so on can be realized. However, when the feature points are restored from low-resolution scale to high-resolution, the influence of image texture information can be deeply considered in the future. In addition, for the iterative solution of the optimal alignment model, the perspective of selecting points from high-resolution patch for further speeding up this method can be taken into consideration in the future.

**Acknowledgement.** The authors thank the financial support from National Natural Science Foundation of China (Grant No. 61905033), and Chongqing Basic and Frontier Research Project (Grant cstc2018jcyjAX0314).

## References

1. Yuan, X., Fang, L., Dai, Q., et al.: Multiscale gigapixel video: a cross resolution image matching and warping approach. In: 2017 IEEE International Conference on Computational Photography (ICCP), pp. 1–9. IEEE (2017)

2. Chen, Y., Jiang, G., Yu, M., et al.: Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters. IEEE Trans. Comput. Imaging **6**, 1044–1058 (2020)
3. Milgrom, B., Avrahamy, R., David, T., et al.: Extended depth-of-field imaging employing integrated binary phase pupil mask and principal component analysis image fusion. Opt. Express **28**(16), 23862–23873 (2020)
4. Cui, J., Zhang, S., Jiang, Z., et al.: Approach of spectral information-based image registration similarity. J. Appl. Remote Sens. **14**(2), 026520 (2020)
5. Ma, J., Jiang, X., Fan, A., et al.: Image matching from handcrafted to deep features: a survey. Int. J. Comput. Vis. **129**(1), 23–79 (2021)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
7. Bay, H., Ess, A., Tuytelaars, T., et al.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
8. Rublee, E., Rabaud, V., Konolige, K., et al.: ORB: an efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision, pp. 2564–2571. IEEE (2011)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
10. Barath, D., Matas, J., Noskova, J.: MAGSAC: marginalizing sample consensus. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10197–10205 (2019)
11. Chum, O., Matas, J.: Matching with PROSAC-progressive sample consensus. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 220–226. IEEE (2005)
12. Ono, Y., Trulls, E., Fua, P., et al.: LF-Net: learning local features from images. Adv. Neural Inf. Process. Syst. **31**, 6237–6247 (2018)
13. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 224–236 (2018)
14. Sarlin, P.E., DeTone, D., Malisiewicz, T., et al.: Superglue: learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4938–4947 (2020)
15. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) Computer Vision – ECCV 2006, vol. 3951, pp. 430–443. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_34
16. Calonder, M., Lepetit, V., Strecha, C., et al.: Brief: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) Computer Vision, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_56