Postprint: Multispectral Image Matching Algorithm Based on Multi-scale Support Region Descriptor

Authors: Zhao Enbo, Shi Zelin, Liu Yunpeng

Date: 2018-06-19T00:00:00+00:00

Abstract

To address the problem of insufficient robustness in existing multispectral image matching algorithms, a novel multispectral image matching algorithm based on multi-scale support region descriptors is proposed. The algorithm first extracts Harris corner points as feature points; then, edge orientation histograms within different scale neighborhoods of the feature points are respectively calculated and combined to construct the feature descriptor; Euclidean distance is employed as the similarity criterion, and the ratio method is used to obtain initial matching results; finally, an outlier removal algorithm based on the RANSAC algorithm is proposed. Experimental results demonstrate that the algorithm can effectively match multispectral images, exhibits stronger robustness compared with existing algorithms, and obtains more correct matching pairs.

Full Text

Preamble

Title: Multi-spectral Image Registration Algorithm Based on Multi-scale Support Region Descriptors

Authors: Zhao Enbo[1],[2],[3], , Shi Zelin[1],[3], , Liu Yunpeng[1],[3],
[1]Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China
The Key Laboratory of Image Understanding & Computer Vision, Shenyang 110016, China

Abstract: To improve the robustness of existing multi-spectral image registration algorithms, this paper proposes a novel algorithm based on multi-scale support region descriptors. The algorithm first extracts Harris corner points as feature points. It then constructs the descriptor by combining edge direction histograms calculated respectively in support regions of different sizes around each feature point. Using Euclidean distance as the similarity criterion, the ratio method is employed to obtain initial matches. Finally, an outlier removal algorithm based on RANSAC is proposed. Experimental results demonstrate that the proposed algorithm can effectively match multi-spectral images, exhibits stronger robustness compared to existing algorithms, and obtains more correct matching pairs.

Keywords: multi-spectral image registration; feature point descriptor; eliminating outliers

0 Introduction

Multi-spectral images are widely applied in remote sensing [1], video surveillance [2], and military applications [3]. Compared to single-spectral images, multi-spectral images provide diverse, reliable, and complementary information. For instance, visible light images offer rich detail but are significantly affected by environmental conditions, whereas infrared sensors can operate around the clock with minimal weather impact, albeit at lower resolution. By registering and fusing these two image types, their respective advantages can be combined to produce richer, more comprehensive images containing greater detail, which proves more beneficial for practical problem-solving such as real-time temperature monitoring of power equipment [4] and addressing limited nighttime driving visibility [5].

To obtain complementary information from multi-spectral images, image registration is required, which necessitates establishing feature correspondences between images. The primary challenge lies in overcoming the non-linear relationship in pixel intensity between visible and infrared images. Typically, infrared sensors operate in the 0.75–15 m spectral range, while visible light sensors operate at approximately 0.4–0.7 m. Consequently, pixel intensity variations in visible images relate to object color and reflected light, whereas infrared image pixel intensity varies with object temperature, resulting in a non-linear relationship between the two modalities. This non-linearity severely limits the effectiveness of classical algorithms such as SIFT [6] and SURF [7] when matching infrared and visible images.

Despite differences in pixel intensity, edges at boundaries between different materials tend to be consistent across modalities. This occurs because objects of different materials typically possess distinct appearance characteristics (e.g., color, shape) that determine which portions of light are reflected and how, causing pixel intensity variations at material boundaries in visible images. Different materials also exhibit different radiation characteristics, which primarily affect pixel intensity in infrared images. In essence, different materials have varying reflection and radiation properties, resulting in different pixel intensities in both visible and infrared images. Therefore, boundaries between different materials appear at corresponding locations in multi-spectral images, motivating the utilization of edge information to achieve effective registration.

1.1 Feature Point Selection

As previously discussed, edge information in multi-spectral images is more similar and consistent than grayscale information. Therefore, theoretically, feature point detection algorithms that leverage edge information should yield more consistent feature points. The earliest proposed feature points were corners, which utilize intensity information from neighboring pixels to distinguish corners from irrelevant pixels. Subsequently, the autocorrelation matrix and eigenvalue criteria were employed to compute corner maps [19]. Corners can essentially be viewed as intersections of two or more edges. Later research shifted toward algorithms capable of estimating feature point scale characteristics to accommodate scale transformations in matching scenarios. The widely applied SIFT algorithm represents a typical example, extracting feature points and estimating their scale characteristics using scale-space theory. The extracted feature points are also referred to as "blobs" because they reflect the similarity between the gray-level distribution in the feature point neighborhood and a Gaussian function. Later, to improve computational efficiency for real-time applications, the FAST feature point detector [20] and its improved versions were proposed, though these algorithms require machine learning to enhance the quality of detected feature points. Consequently, using corners as feature points for multi-spectral image matching represents a reasonable choice.

To demonstrate that corners are more suitable for multi-spectral images, this paper designed a comparative experiment evaluating the performance of the SIFT algorithm and the representative corner extraction algorithm—Harris algorithm. The evaluation metric is repeatability, defined as the ratio of corresponding feature points in location to the total number of feature points. Unlike experiments where the same algorithm applies identical parameters to both infrared and visible images, this experiment extracts the same number of feature points. This approach eliminates the need to adjust each parameter of the algorithm according to multi-spectral image differences and ensures fairness for both methods. Here, the top 1,000 feature points are extracted. Since images are acquired from different sensors, a tolerance of 4 pixels is allowed for corresponding feature points in location, which applies to all experiments mentioned in this paper. The experiment uses a visible-long-wave infrared image dataset [21] as test data. The results are shown in [Figure 1: see original paper]. It is evident that in most cases, Harris corners exhibit higher repeatability, thus Harris corners are selected as the feature points to be extracted.

1.2 Edge Pixel Gradient Direction Calculation

When calculating the gradient direction of edge pixels, the EOH algorithm uses a 5-direction Sobel operator, yielding relatively coarse directions. To enhance the representativeness of the descriptor, an 8-direction Sobel operator [22] is employed to calculate the direction of edge pixels. Increasing the number of edge directions can increase the dimensionality of the descriptor and enhance the distinctiveness between descriptors. The 8-direction Sobel filter templates are shown in [Figure 2: see original paper].

The direction of an edge pixel is calculated using Equation (1):

$$\mathrm{Ori}(x, y) = \arg\max_i \left(F_i \cdot \mathrm{Patch}(x, y)\right), \quad i = 1, 2, 3, \ldots, 8$$

where (x, y) represents the coordinates of the current edge pixel, $\mathrm{Patch}(x, y)$ denotes the neighborhood pixel block of the edge pixel in the grayscale image of the original image with the same size as the convolution template, F_i represents the Sobel operator, $\cdot$ denotes convolution operation, and $\mathrm{Ori}(x, y)$ corresponds to the direction represented by the maximum value among the convolution results obtained from the 8 Sobel operators.

1.3 Descriptor Construction

Descriptors calculated from multi-scale neighborhoods can fully utilize edge information in multi-spectral images. Artificial environments often contain multiple objects with similar shapes. As shown in the red rectangular box in Figure 3: see original paper, the shape and size of windows are very similar, and from Figure 3: see original paper it can be observed that the edges of these windows are essentially identical. In such cases, descriptors constructed from single neighborhoods face certain limitations. When the neighborhood is small, the descriptor contains only local information about the feature point. If the edges within the neighborhoods of two feature points are similar, the similarity between descriptors will be high, degrading matching performance. In this case, it is necessary to expand the neighborhood to incorporate different edges and thereby increase descriptor distinctiveness. However, when the neighborhood is large, if two feature points are close to each other, the overlap ratio between their support regions increases, resulting in high similarity between the constructed descriptors and leading to mismatches. In such situations, differences can be increased by including small-scale support regions containing local information. Therefore, this paper proposes using multiple support regions of different scales to compute descriptors, meaning the descriptor consists of several sub-descriptors calculated from multiple neighborhoods of different scales around the feature point. This method is called Multi-scale Support Region Edge Orientation Histogram (EOHMSR). The goal is to improve descriptor robustness for multi-spectral images of different scenes, thus geometric transformations such as rotation, scaling, and affine transformation are not considered.

The descriptor calculation process is illustrated in [Figure 4: see original paper]. The edge image $\mathrm{Edge}(x, y)$ is obtained from the original image $I(x, y)$ using the Canny algorithm, with thresholds automatically set according to the maximum gradient magnitude and parameter $\sigma = 2$. The direction of each edge pixel in $\mathrm{Edge}(x, y)$ is obtained by $\mathrm{EdgeOri}(x, y)$. For a feature point, different scale support regions $\mathrm{SR} = \{\mathrm{SR}_1, \mathrm{SR}_2, \dots, \mathrm{SR}_n\}$ are obtained. Each support region is divided into $4 \times 4 = 16$ sub-regions, and the edge direction histogram H_i of each sub-region is calculated, yielding a sub-descriptor with 128 dimensions ($4 \times 4 \times 8 = 128$). Finally, the H_i are combined to obtain the final descriptor.

2 Feature Point Matching

Feature point matching involves comparing feature point descriptors to find corresponding feature point pairs. During descriptor calculation, feature points with too little edge information in their neighborhoods are removed, as descriptors calculated from insufficient edge information have weak representativeness. Euclidean distance is used as the similarity measure, and the ratio criterion [6] is employed to determine correspondences. Two feature points are considered matched only when the ratio between the shortest distance and the second shortest distance is below a given threshold.

The initial matching result is shown in Figure 5: see original paper. Through observation, mismatched point pairs can be divided into two types: single-point correspondence type and one-point-to-multiple-points type. The main reasons for the second error type include numerous similar regions in the image and relatively close positions of some feature points. Due to the low inlier rate (26/208), directly using the RANSAC algorithm [23] yields poor results (Figure 5: see original paper), where the calculated transformation model only reflects the transformation relationship of the second outlier type, indicating that the proportion of the second mismatch type is relatively high.

Inspired by this observation, this paper proposes a practical outlier removal algorithm. The key is to divide the initial matching results into two parts based on the number of points matching the same point, then process each part separately. The algorithm flowchart is shown in [Figure 6: see original paper]. If the number is 2, it indicates one-to-one matching; if greater than 2, it indicates one-to-multiple matching. Accordingly, the initial matching results D_{original} are divided into D_{single} and D_{multiple}. Since the proportion of the second mismatch type is high, the RANSAC algorithm is first applied to the one-to-one matching part D_{single} to obtain a rough but basically correct model M_{single}, while obtaining most inliers $\mathrm{InL}_{\mathrm{single}}$.

However, in some cases, D_{multiple} still contains correct matching pairs. To make the obtained transformation model suitable for all correct matches, M_{single} is applied to D_{multiple} to extract potentially correct matching pairs $D_{\mathrm{candidate}}$. First, feature points in the infrared image belonging to D_{multiple}, denoted as $\mathrm{PIr}_{\mathrm{multiple}}$, are transformed to the visible light image using Equation (2), with the transformation result denoted as $\mathrm{PIr}^{T}_{\mathrm{multiple}}$.

$$\mathrm{PIr}^{T}_{\mathrm{multiple}} = M_{\mathrm{single}}(\mathrm{PIr}_{\mathrm{multiple}})$$

The distance between each point in $\mathrm{PIr}^{T}_{\mathrm{multiple}}$ and each point in $\mathrm{PVs}_{\mathrm{multiple}}$ (feature points in the visible light image belonging to D_{multiple}) is then calculated using Equation (3):

$$d = \|\mathrm{PIr}^{T}_{\mathrm{multiple}} - \mathrm{PVs}_{\mathrm{multiple}}\|$$

If $d < 4$, the two points are considered correct matching candidates. Finally, $\mathrm{InL}_{\mathrm{single}}$ and $D_{\mathrm{candidate}}$ are combined to form D_{new}, and the RANSAC algorithm is applied to obtain the final transformation model M_{final} and all inliers $\mathrm{InL}_{\mathrm{final}}$.

As shown in [Figure 7: see original paper], compared with directly using the RANSAC algorithm, the proposed method can overcome the problem of low inlier rate and obtain correct matching models.

3.1 Parameter Selection

The proposed algorithm has several parameters that need to be determined experimentally, including the number of extracted feature points K, the number of support regions N, and the minimum size of support regions $D \times D$. To determine these parameters, experiments were conducted on a visible-long-wave infrared dataset [21] containing 44 image pairs, each consisting of a visible image and an infrared image with corrected perspective and identical resolution.

The variation of repeatability across the entire dataset with respect to the number of feature points K is shown in [Figure 8: see original paper]. Repeatability is defined as the ratio of corresponding feature points in location to the total number of feature points. It is evident that repeatability increases with K, as the likelihood of feature point repetition increases with more feature points. Considering computational efficiency, $K = 1000$ is set, as repeatability improvement becomes insignificant beyond this number.

Precision and recall are used as evaluation criteria for selecting the number of support regions N and the minimum support region size $D \times D$, calculated using Equations (4) and (5):

$$\text{Precision} = \frac{\#\text{correct matches}}{\#\text{matches}}$$

$$\text{Recall} = \frac{\#\text{correct matches}}{\#\text{correspondences}}$$

where #correct matches is the number of matching pairs after outlier removal using the method described in Section 2, #matches is the number of initial matching pairs including both inliers and outliers, and #correspondences is the number of feature points that should be correctly matched. Precision and recall are set to 0 when too few correct matches exist to obtain a transformation model or when the obtained model is incorrect.

The variation of average precision and average recall with N and $D \times D$ is shown in [Figure 9: see original paper]. D increases by the same interval of 20 as N increases. When N increases, average precision increases regardless of $D \times D$ value, but does not exceed 35%. Average recall follows the same trend and reaches its maximum (approximately 15%) when $N = 5$. Since each increment of N increases descriptor dimensionality by 128, considering descriptor dimension, the final setting is $N = 5$, $D = 30$.

3.2 Performance Comparison

The experimental environment is a 64-bit Windows system with an Intel Core i7 3.4 GHz CPU and 4 GB RAM, running MATLAB R2014b. The proposed method is evaluated against three classical algorithms on the visible-long-wave infrared image dataset [21] using precision and recall as criteria. The dataset contains 44 image pairs, each with a visible image and an infrared image corrected to identical perspective and resolution (639×43). [Figure 10: see original paper] shows several image pairs. Since all pairs are corrected, lines connecting correct matches should be horizontal, as shown in [Figure 11: see original paper]. The proposed algorithm yields more correct matching results than the other three algorithms.

[Figure 12: see original paper] shows the recall and precision for each image pair in the dataset. The proposed method clearly outperforms the other three methods in the recall curve and surpasses them in precision for most image pairs. The average results per image pair across the entire dataset are summarized in . The proposed algorithm significantly outperforms the other three methods in average recall, average precision, and average number of correct matches. In terms of matching time, the proposed algorithm is slightly better than Sym-SIFT but approximately 1.5 times slower than EOH. However, to achieve the same average number of correct matches (18) as the EOH algorithm, the proposed algorithm requires only 5.52 seconds, about one-third of EOH’ s time, making it superior from this perspective. This paper focuses on algorithm effectiveness and improving applicability across different scenes. Future work will improve algorithm real-time performance through software and hardware optimizations.

4 Conclusion

This paper proposes an effective multi-spectral image matching algorithm. First, Harris corners are selected as feature points to ensure high repeatability of extracted features. Then, descriptors are calculated using edge information from different scale support regions to enhance distinctiveness between descriptors. Finally, to address the low inlier rate in initial matching results, a novel outlier removal algorithm is developed based on RANSAC. Experimental results demonstrate that the proposed algorithm is more robust than traditional multispectral image algorithms and obtains more correct matching pairs. Future work will enhance the proposed method' s robustness to rotation, scaling, and affine transformations while improving real-time performance.

References

[1] Cheng Guohua, Wang Achuan, Chen Shuchang, et al. High accuracy-automatic registration method research on multi-source remote sensing image [J]. Chinese Journal of Liquid Crystals & Displays, 2016, 31 (6): 604-612.

[2] Li Yanmei, Chen Leiting, Rao Yunbo, et al. Infrared and visual image fusion enhancement by dual-transform [J]. Application Research of Computers, 2013, 30 (10): 3142-3145.

[3] Guo Shaojun, Liu Feng, Xi Xiaoliang. The registration of warship infrared and visible images [J]. Infrared Technology, 2016, 38 (5): 403-408.

[4] Xiao Erliang, Liu Wenwen. Research of multi-scale gradient domain visible and thermal image fusion method [J]. Application Research of Computers, 2015, 32 (10): 3160-3163, 3167.

[5] Guo Quanmin, Dong Liang, Li Daidi. Vehicles anti-halation system based on infrared and visible images fusion [J]. Infrared & Laser Engineering, 2017, 46 (8): 171-176.

[6] Lowe D. Distinctive image features from scale-invariant keypoints [C]// Proc of International Journal of Computer Vision. 2004: 91-110.

[7] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features [J]. Computer Vision & Image Understanding, 2008, 110 (3): 404-417.

[8] Brown L. A survey of image registration techniques [J]. Computing Surveys, 1992, 24: 325-376.

[9] Zitova B, Flusser J. Image registration methods: a survey [J]. Image and Vision Computing, 2003, 21: 977-1000.

[10] Xiong Zhen, Zhang Yun. A critical review of image registration methods [J]. International Journal of Image and Data Fusion, 2010, 1: 137-158.

[11] Maes F, Vandermeulen D, Suetens, P. Medical image registration using mutual information [J]. Proceedings of the IEEE, 2003, 91: 1699-1622.

[12] Coiras E, Santamaria J, Miravet C. Segment-based registration technique for visual-infrared images [J]. Optical Engineering, 2000, 39: 282-289.

[13] Han J, Pauwels E, Zeeuw P. Visible and infrared image registration employing line-based geometric analysis [C]// Proc of International Conference on Computational Intelligence for Multimedia Understanding. Berlin: Springer, 2011: 114-125.

[14] Wei Lijun. Line matching algorithm on multispectral images [D]. Beijing: Beijing University of Posts and Telecommunications, 2017.

[15] Chen Jian, Tian Jie. Real-time multi-modal rigid registration based on a novel symmetric-sift descriptor [J]. Progress in Natural Science-Materials International, 2009, 19: 643-651.

[16] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// Proc of IEEE Computer Society Conference on Computer Vision & Pattern Recognition. [S. l.] : IEEE Computer Society, 2005: 886-893.

[17] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24: 509-522.

[18] Aguilera C, Barrera F, Lumbreras F, et al. Multispectral image feature points [J]. Sensors, 2012, 12: 12661-12672.

[19] Harris C. A combined corner and edge detector [J]. Proc Alvey Vision Conf, 1988, 1988 (3): 147-151.

[20] Rosten E, Drummond T. Machine learning for high-speed corner detection [C]// Proc of European Conference on Computer Vision. Berlin: Springer, 2006: 430-443.

[21] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]// Proc of IEEE Computer Society Conference on Computer Vision & Pattern Recognition. [S. l.] : IEEE Computer Society, 2005: 886-893.

[22] Zheng Yingjuan, Zhang Huiyou, Wang Zhiwei, et al. Edge detection algorithm based on the eight directions Sobel operator [J]. Computer Science, 2013, 40 (11A): 354-356.

[23] Fischler M, Bolles R. Random sample consensus: a paradigm for model-fitting with applications image-analysis and automated cartography [J]. Communications of the ACM, 1981, 24: 381-395.

Note: Figure translations are in progress. See original paper for figures.