3-D Point Cloud Reconstruction of Infrared Images Based on Improved Structure from Motion

YINGYING KONG, BOWEN ZHANG,
YINGYING CHEN
CEIE
Nanjing University of Aeronautics and Astronautics
Nanjing
CHINA
yayako_zy@nuaa.edu.cn
http://ceie.nuaa.edu.cn

HENRY LEUNG, SHIYU XI
ECE
University of Calgary
Calgary, AB
CANADA
Leungh@ucalgary.ca
https://www.ucalgary.ca

Abstract: Structure from Motion (SfM) has been proved an efficient algorithm of 3-D point cloud reconstruction derived from optical images. This paper extends it to infrared images taken by thermal cameras. To solve the absence of distinctive features and presence of thermal reflections with low contrast, this paper proposed a new TAC-RANSAC model to eliminate the mismatches using a feature detection algorithm suitable for infrared images. The experiment shows that the proposed method reduced the number of mismatches and obtain an ideal result of reconstruction.

Key-Words: 3-D Reconstruction, MVG, infrared images, SfM, RANSAC, mismatch eliminating

Received: September 15, 2018. Revised: November 3, 2018. Accepted: December 28, 2018.
Published: January 23, 2019

1 Introduction

3-D reconstruction has always been an active topic in the field of computer vision. We human understand the real world through our eyes and form a 3-D model in our brain. Computers are expected to do the same thing to make the real world comprehensive to themselves. Some efforts have been done in the field of 3-D reconstruction with optical images, such like Shape from Shading [1], Structure from Motion (SfM) [2], Multi-view Stereo [3] and some deep learning methods [4, 5] in recent years. Compared to other methods, SfM is fast, little-memory needed, resulted relatively well, thus cost-effective.

Unlike optical light, infrared light is invisible to the human eye and requires a thermal camera to generate images. Infrared images show the temperature distribution the object. As a result, they have low resolution, poor contrast, blurred visual effect and various forms of noise compared to optical images. However, thermal imaging is not affected by environmental factors such as weather, light, and smoke, as optical imaging always be.

Therefore, infrared images are often used for the detection and maintenance of power system. However, infrared images can only provide 2-D information. Due to the lack of depth information in 2-D images, it is meaningful to reconstruct 3-D models with rich temperature information as well as stereo information, which can show both the surface information and 3-D structure of the monitoring target more realistically. Besides, 3-D thermal models are also contributory to some medical applications [6] and human face recognition [7] due to these desirable properties.

Some efforts have been done in the field of 3-D infrared reconstruction. Zeng et al. [8] proposed an IR-SFS algorithm, which improved the traditional SFS algorithm by both considering external light source and self-radiation of infrared objects. Jia et al. [9] proposed an improved belief propagation algorithm to enhance the stereo match of thermal images. A convolutional auto-encoder based method is applied to image matching and 3-D thermal reconstruction [10], which is capable of extracting features from low or nontextured objects, and perform robust patch matching from multiview stereo infrared imagery. Yet, there are still a trade-off between running speed, needed memory and accuracy of the result of reconstruction.

In the conference version of this paper [17], we improve the SfM algorithm so that make it w ork with infrared images. To overcome the absence of distinctive features and the presence of low contrast, we summarize a feature detection algorithms suitable for infrared images, and proposed a new TAC-RANSAC model to eliminate the mismatches. The proposed TAC-RANSAC model utilizes the feature of temperature in infrared images.

This journal paper extends our earlier work through further analysis and more results. Alternative environments, conditions and implementation details better cover the proposed method. The experiment shows that the proposed method reduces mismatches and obtain an ideal result of reconstruction from infrared images, seen in section 3.

2 Improved SfM

Structure from Motion (SfM) is a wide-used algorithm in the field of optical 3-D reconstruction. In this chapter, we briefly introduced the typical SfM algorithm with its post-processing, and then describe our improvement to extend it to the application of infrared images.

2.1 Structure from Motion

SfM is a mature algorithm in the field of 3-D reconstruction of images taken by optical sensors such like personal digital cameras. The workflow of SfM is shown in Fig. 1.

The feature detection part extracts key points from all images, and describes them mathematically by descriptors such as SIFT [14], SURF [15] and other algorithms. The matching part matches the keypoints in corresponding images with the descriptors. The obtained keypoint correspondence and RANSAC algorithm are utilized to calculate the camera poses. Finally, the epipolar constraint is used to verify all correspondences and remove mismatches.

When correct pairs of matching keypoints are obtained, the fundamental matrix $\boldsymbol{F}$ and essential matrix $\boldsymbol{E}$ can be estimated through the epipolar constraint, which indicates that each pair of matching keypoints $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ satisfies

$$\tilde{\mathbf{x}}^{\mathrm{T}} \boldsymbol{E}\, \tilde{\mathbf{x}}' \sim 0 \qquad (1)$$

where $\boldsymbol{E}$ is determined only by the rotation matrix $\boldsymbol{R}$ and translation matrix $\boldsymbol{T}$ and thus the camera matrix $\boldsymbol{P}$ is obtained.

Then points in the real world can be generated by triangulation. The back-projection radial lines are not usually intersected due to the reprojecting errors, as shown in Fig. 2. The Golden Rule [16] is usually adopted to minimize the error sum of squares of the distance between the generated real-world points reprojecting to the image $\hat{\mathrm{u}}_i(P_i, \mathrm{X})$ and the real-image points u_i, which is

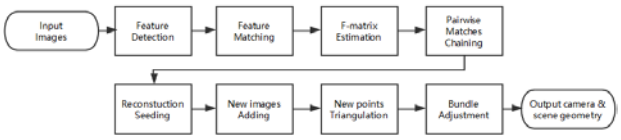


Fig.1 The Workflow of SfM.

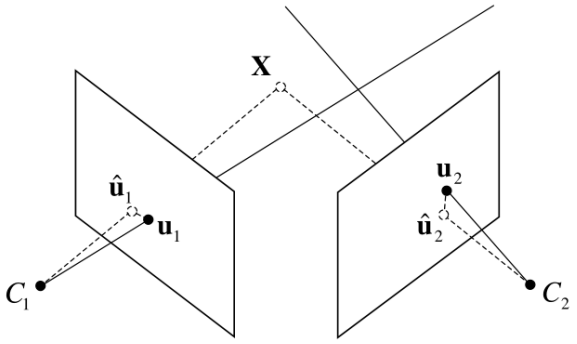


Fig.2 The generated real-world points reproject to the image with error.

$$X = \arg\min_{X} \sum_{i} \| \mathrm{u}_i - \hat{\mathrm{u}}_i(P_i, \mathrm{X}) \|^2 \qquad (2)$$

which gives the maximum likelihood solution of X.

It is obviously that the feature detection part and the matching part are the basic parts of SfM, which almost determine the quality of the reconstruction result. Common methods to eliminate mismatches include cross-filtering, KNN (K-Nearest Neighbor), RANSAC and AC-RANSAC (A Contrario Random Sample Consensus) [11, 12].

2.2 TAC-RANSAC

TAC-RANSAC is proposed on the basis of AC-RANSAC, adding the unique feature of thermal images, the temperature. In the following content, first we briefly introduce the basic RANSAC and AC-RANSAC and then follows the detailed idea of the proposed TAC-RANSAC.

2.2.1 RANSAC and AC-RANSAC

RANSAC is the most common-used algorithm to eliminate mismatches. Its core idea is quiet simple and can be separated to 3 steps:

1) Randomly sample N pairs of data to estimate the model;

2) Compute matches constrained by the threshold;

3) End the iteration if the number of iterarion is less than n and the number of matches is larger than m. Otherwise restart the algorithm from 1).

The key of RANSAC is the hyperparameters set by the users of the algorithm. RACSAC may not able to find a reasonable model if the threshold of

difference between matches and mismatches, σ, is set too low, and the model resulted may be not optimized if the number of iterations is set too small. Consequently, a problem occurs which the hyperparameters depend on specific data and models. To solve this problem, AC-RANSAC is proposed which does not need a threshold. Instead, the algorithm finds a balance between σ and the number of matches by controlling NFA (Number of False Alarms). The common formula of NFA is

$$NFA = N_{out}(n - N_{sample})\binom{n}{k}\binom{k}{N_{sample}}\left(\varepsilon_k^d \alpha_0\right)^{k-N_{sample}} \tag{3}$$

where N_{sample} denotes the minimum number of needed samples to estimate one model. N_{out} denotes the number of models that can be inferred from N_{sample} samples, which equals to 1 in this case. Assume the threshold between matches and mismatches is ε_k, which represents the kth residual of n samples. α_0 denotes the probability of the maximum error of randomly selected samples equaling one pixel. d refers to the dimension which equals 1 for point-to-line distance and 2 for point-to-point, respectively. The probability of each point pair becoming a match is $\varepsilon_d^k \alpha_0$ since all samples are independent in the background model.

2.2.2 TAC-RANSAC

TAC-RANSAC uses the temperature information as another constraint to distinguish matches from mismatches. The thermal camera used in experiments, FLIR® Tau 2, can take 9 pictures per second. Thus, it can be assumed that the temperature of a certain point remains almost the same in the second. In view of this, the temperature information is introduced into NFA as another error term, which is the proposed NFAT (Number of False Alarms with Temperature)

$$NFAT(M,k) = \chi N_{out}(n - N_{sample})\binom{n}{k}\binom{k}{N_{sample}} \cdot\left(\sigma_1(M)^d \alpha_0 + \sigma_2(M)\beta_0\right)^{k-N_{sample}} \tag{4}$$

where χ is a normalized value. $\sigma_1(M) = \xi_k$ represents the first minimum distance error in all corresponding relations. $\sigma_2(M) = \varepsilon_k$ denotes the first lowest temperature error in all corresponding relations under model M. β_0 is the probability of a random corresponding pair having at most 0.01 temperature difference, which is

$$\beta_0 = \frac{0.01}{\overline{T}} \tag{5}$$

where $\overline{T}$ is the average of the difference of the highest and lowest temperature of two matched images, I_i and I_j, which is

$$\overline{T} = \frac{(T^i_{\max} - T^i_{\min}) + (T^j_{\max} - T^j_{\min})}{2} \tag{6}$$

where $T^i_{\max}$, $T^i_{\min}$, $T^j_{\max}$ and $T^j_{\min}$ represent the highest and lowest temperature in the image I_i and image I_j, respectively. The temperature and the pixel value of each point in the infrared image are proportional, which is

$$T = I_p \times 0.04 - 273.15 \tag{7}$$

$\sigma_2(M)\beta_0$ is the probability of a random correspondence pair having at most ε_k temperature error for under the assumed background model distribution. Thus the last factor of *NFAT* denotes probability of $k - N_{sample}$ pairs of matches having at most ε_k and ξ_k errors. In other words, this is the expectation that model M has a false alarm of k matches in the hypothetical background model. If M is considered reasonable, then

$$NFAT(M) = \min_{k=N_{sample+1\ldots n}} NFAT(M,k) \le \varepsilon \tag{8}$$

where ε is usually set to be 1 (the lower, the more accuracy). The whole algorithm is listed in Table 1.

We use TAC-RANSAC to estimate the homography matrix to eliminate mismatches. In that case, $N_{sample} = 4$, $N_{out} = 1$, $d = 2$, $\alpha_0 = \pi / (w * h)$. Then *NFAT* becomes

$$NFAT(M,k) = \chi(n-4)\binom{n}{k}\binom{k}{4} \cdot\left(\xi_k^2 \frac{\pi}{w*h} + \varepsilon_k \frac{0.01}{\overline{T}}\right)^{k-4} \tag{9}$$

where ξ_k is the d or max(d, d') in Fig. 3.

The experiment results are listed in section 3, demonstrating the effectiveness and practicability of TAC-RANSAC in 3.1, and the resulting improvement of reconstruction in 3.2, respectively.

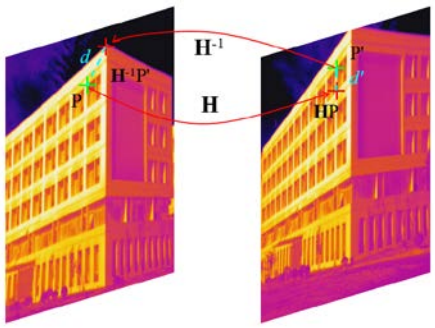


Fig.3 The residual from the homography matrix.

Table1. The TAC-RANSAC Algorithm.

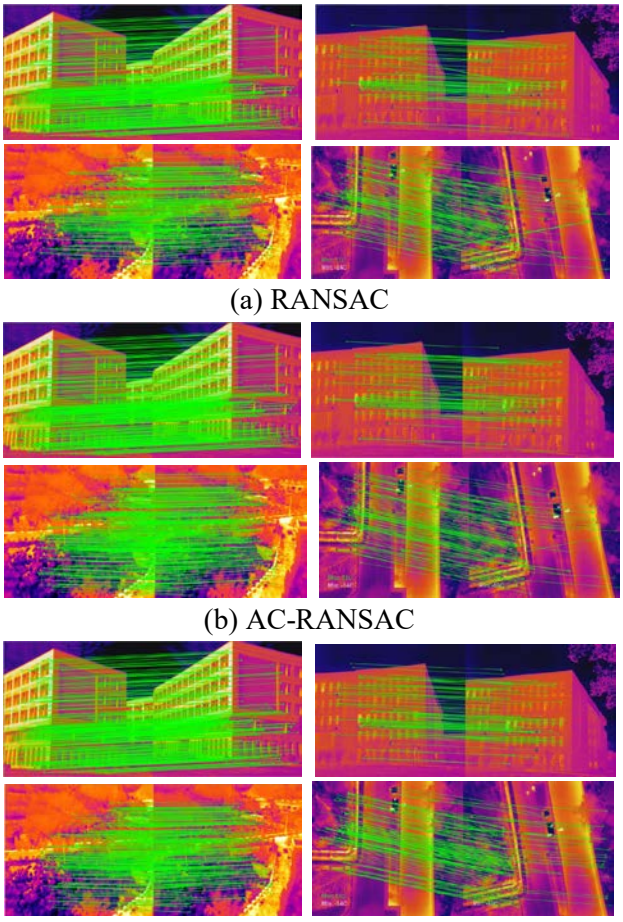
Input:	N_{sample} : the number of samples needed to estimate model parameters.
	ε: the *NFAT* threshold.
	nIter : iterations.
Process:	
	1: Randomly select N_{sample} samples.
	2: Estimate model parameter P_{aram} .
	3: Calculate the residual of $n - N_{sample}$ and sort in ascending order.
	4: Put the remain in matches in order and calculate *NFAT*, save the P_{aram} with lowest *NFAT*.
	5: Decide whether P_{aram} is reasonable and discard it if not.
	6: Go to step 1 if *nIter* has not been reached.
Output:	P_{aram}: the parameter of the model.

3 Experiment

In this section, we use infrared image pairs with different environments and color table to test our method. The direct result and the resulting improvement of reconstruction are listed in 3.1 a nd 3.2, respectively.

3.1 The Direct result from TAC-RANSAC

In this paper, cross-filtering, KNN and TAC-RANSAC algorithms are combined to produce an accurate and efficient model to eliminate mismatches. Cross-filtering can significantly remove the obvious mismatches. Both the time complexity and space complexity of the algorithm are low, but so is the accuracy. So cross-filtering is utilized as the first and a primary step to eliminate mismatches. KNN algorithm is better than cross-filtering in terms of accuracy, thus can be used as t he consequent step (In this paper, k = 2). Running KNN after cross-filtering is faster and more efficient than running KNN directly since cross-filtering filtering out obvious mismatches. The searching range of the KNN algorithm can be narrowed, thereby shortening the matching time and improving the efficiency. After this two steps of filtering, a quantity of mismatches with large error has been removed. The influence of large error has been eliminated so that the TAC-RANSAC algorithm is able to estimate an accurate geometric constraint model. The combination of cross-filtering, KNN and TAC-RANSAC to eliminate mismatches is theoretically reasonable and effective.



(a) RANSAC

(b) AC-RANSAC

(c) TAC-RANSAC

Fig.4 Results of eliminating mismatches with 3 methods presented above.

4 pairs of infrared images are used to estimate the result of RANSAC, AC-RANSAC, and proposed TAC-RANSAC. The matching result is shown in Fig. 3 a nd the performance is listed in Table 2, 4 rows for 4 pairs of images, respectively. The result shows that the TAC-RANSAC improves the result in terms of the number of matches, error sum, and matching speed.

3.2 The resulting improvement of reconstruction

After the matching step, 60 and 72 infrared images, respectively, are used to build the 3-D point clouds based on SfM using the result of AC-RANSAC and TAC-RANSAC. Then PMVS method [13] is applied to generate the dense point clouds. The results of reconstruction are performed in Fig. 4, which show that the generated 3-D point cloud based on TAC-RANSAC matching model has more matching points. The point cloud is also fuller, denser and more complete. The result is shown quantitatively in Table 3 in terms of the number of points, mean angular error, running time, mean position error and RMSE, 2 rows for 2 reconstruction results, respectively.

Table2. Performance of 3 models

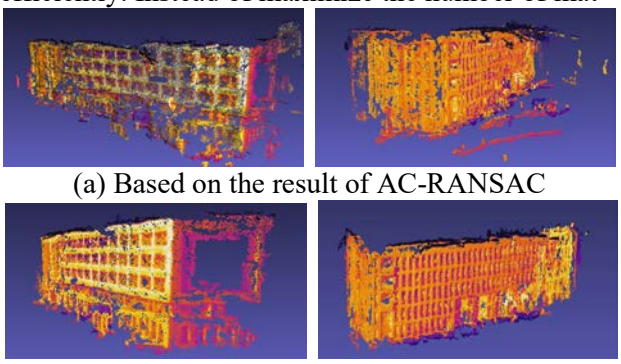
	RANSAC	AC-RANSAC	TAC-RANSAC
Error sum	1.772	1.615	**0.975**
	2.238	1.934	**0.985**
	4.449	3.903	**1.676**
	2.901	1.823	**0.725**
Matching time	0.478s	0.521s	**0.404s**
	0.307s	0.378s	**0.247s**
	0.563s	0.611s	0.599s
	0.353s	0.368s	**0.304s**
The number of matches	348	**356**	350
	146	176	**191**
	366	**388**	369
	224	229	**234**

Table3. Performance of 3-D point cloud based on AC-RANSAC and TAC-RANSAC

Models	AC-RANSAC	TAC-RANSAC
The number of points	23700	**30067**
	51023	**60047**
Mean angular error	0.273°	**0.097°**
	0.55	**0.12**
Running time	333s	**301s**
	603s	**555s**
Mean position error	8.71cm	**3.26cm**
	2.56cm	**1.80cm**
RMSE	3.98cm	**3.23cm**
	4.76cm	**3.59cm**

4 Conclusion

SfM is a mature algorithm in the field of 3-D reconstruction for optical images. This paper extends it to infrared images took by thermal cameras, and propose a new TAC-RANSAC model based on AC-RANSAC by adding a temperature error term to eliminate mismatching points more efficiently. Instead of maximize the number of matches, the goal of TAC-RANSAC is to minimize *NFAT*. Compared to AC-RANSAC, the added error term is beneficial to the accuracy of keypoint matching and reduce the needed time of convergence. Once found a reasonable model, TAC-RANSAC algorithm greatly saves time by either reducing the number of random samples or directly sampling in the matches to achieve the purpose of reduce the needed time of convergence. The result of 3-D reconstruction shows that the proposed model can improve the matching accuracy, reduce the mismatching rate, and generate a better 3-D point cloud with lower position error, angle error and RMSE. The further experiment reveals the potential 3-D monitoring abilities for different kinds of man-made constructions.



(a) Based on the result of AC-RANSAC

(b) Based on the result of TAC-RANSAC

Fig.5 The generated 3-D point clouds.

References:

[1] Zhang, Ruo, et al. "Shape from Shading: A Survey." IEEE Transactions on Pattern Analysis & Machine Intelligence 21.8(2002):690-706.

[2] Havlena, Michal, and K. Schindler. VocMatch: Efficient Multiview Correspondence for Structure from Motion. Computer Vision – ECCV 2014. Springer International Publishing, 2014:46-60.

[3] Goesele, Michael, B. Curless, and S. M. Seitz. "Multi-View Stereo Revisited." 2.2(2016):2402-2409.

[4] Rezende, Danilo Jimenez, et al. "Unsupervised Learning of 3D Structure from Images." (2016).

[5] Fan, Haoqiang, H. Su, and L. Guibas. "A Point Set Generation Network for 3D Object Reconstruction from a Single Image." (2016):2463-2471.

[6] Abreu, de Souza M, et al. "Generation of 3D thermal models for dentistry applications." Engineering in Medicine and Biology Society IEEE, 2016:1397.

[7] Reiter, Michael, et al. "3D and Infrared Face Reconstruction from RGB data using Canonical Correlation Analysis." International Conference on Pattern Recognition IEEE, 2006:425-428.

[8] Zeng, Wang, Liu, et al. “3D reconstruction of space target IR image based on IR-SFS algorithm.” Chinese Optics 03(2014):376-388.

[9] Jia, Tong, et al. "3D temperature distribution model based on vision method." IEEE International Conference on Robotics and Biomimetics IEEE, 2017:852-855.

[10] Knyaz, Vladimir A., et al. "Deep Learning of Convolutional Auto-Encoder for Image Matching and 3D Object Reconstruction in the Infrared Range." IEEE International Conference on Computer Vision Workshop IEEE Computer Society, 2017:2155-2164.

[11] Rabin, Julien, et al. "MAC-RANSAC: a robust algorithm for the recognition of multiple objects." Proceedings OFDPTV (2009):051.

[12] Moulon, Pierre, P. Monasse, and R. Marlet. "Adaptive structure from motion with a contrario model estimation." Asian Conference on Computer Vision Springer Berlin Heidelberg, 2012:257-270.

[13] Furukawa, Yasutaka, and J. Ponce. "Accurate, Dense, and Robust Multiview Stereopsis." IEEE Transactions on P attern Analysis & Machine Intelligence 32.8(2010):1362-1376.

[14] Bay, Herbert, T. Tuytelaars, and L. V. Gool. "SURF: speeded up robust features." European Conference on Computer Vision Springer-Verlag,2006:404-417.

[15] Lowe, David G. "Distinctive Image Features from Scale-Invariant Keypoints." International Journal of Computer Vision 60.2(2004):91-110.

[16] Hartley, Richard, and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.

[17] Kong, Zhang, Chen, et al. "3-D Point Cloud Reconstruction of Infrared Images Based on Improved Structure from Motion" EECS2018-123.