Image Stitching Based on Binocular Vision

Bo Zhang¹,²,³, Yue Ma¹,³, Min Xu¹,³
¹Shenyang Institute of Automation Chinese Academy of Sciences¹, Shenyang 110016, China
²University of the Chinese Academy of Sciences², Beijing 100049, China
³Institute of Robotics and Intelligent Manufacturing³, Chinese Academy of Sciences,
Shenyang 110016, China
zhangbo@sia.cn

Abstract. In the field of photography, image stitching is used to generate panorama. In most cases, the position and intrinsic parameters of the camera can not be obtained, so the transform matrix can only be calculated using the information of the image. In industrial machine vision detection system, the information of the camera can be obtained by calibration. It is feasible to improve the speed of image stitching algorithm by using the information of camera calibration. An image stitching algorithm based on binocular vision was proposed. Compared with the traditional stitching algorithm based on image information only, it greatly improved the running speed of the algorithm. The effectiveness of the algorithm was verified by experiments.

1. Introduction
In modern machine vision detection system, it is vital work to get a high-resolution picture of the object. Once an industrial camera was chosen, the only thing that can be done to improve the resolution of the picture is shortening the distance between the camera and the object. But for many objects, this kind of operation may cause that only part of the object can be captured by the camera instead of the whole object. To solve the problem, several industrial cameras are fixed in different positions to take pictures of the object from different views and stitch all the pictures into a panorama. In general, two-camera detection system is more common.

Many theories and algorithms about image stitching were put forward, which have similar algorithm flow. Before feature-based image stitching algorithm was proposed, many works had been done about image stitching mainly based on the idea of directly minimizing pixel-to-pixel dissimilarities. Then a different class of algorithms works by extracting a sparse set of features and then matching these to each other was presented and now becomes the mainstream and represents the trend of the image stitching, because it has the advantage of being more robust against scene movement and is potentially faster, if implemented the right way[1]. In sausage package detection line system, it is important to get the high-resolution picture of the sausage to detect the package defects. So there are two cameras parallelly fixed over the conveyor belt, which capture the left part and the right part of the sausage respectively, then stitch the picture of the left camera and the picture of the right camera together.
2. **Dual camera joint calibration and stereo rectification**

The physical model of the binocular camera is shown in Figure 1. The left and right cameras are selected as the same type and placed in parallel. In this ideal case, the imaging point coordinates of the point $P$ in the world coordinate system should only differ by a horizontal offset $T$ between the image planes of the two cameras, as shown in Figure 2. But in practice, it is impossible to fix two cameras completely parallel. Because of the radial distortion and tangential distortion in camera imaging, it is necessary to rectify the distortion of each camera and make the imaging plane of left and right cameras coplanar through the joint calibration of binocular cameras. Finally, stereo rectification of binocular cameras is made to make the pixel plane of left and right cameras row-aligned. The stereo-rectified image is shown in Figure 3. It can be seen that when a point in the three-dimensional world was projected onto the left and right pixel planes these two projection point have the same $y$-coordinate[2-3].
3. Image stitching

3.1. SURF descriptors

Unlike the Gauss difference image is used in SIFT(Scale Invariant Feature Transform)[4-5] operator, SURF (Speed-up Robust Feature)[6] use Hessian matrix determinant approximation image. Hessian matrix is a square matrix composed of second-order partial derivatives of real-valued functions with vectors as independent variables. The Hessian matrix of a pixel in the image is as follows.

\[
H(f(x, y)) = \begin{bmatrix}
\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\
\frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2}
\end{bmatrix}
\]  

(1)

However, because the feature points need to be scale independent, the image needs to be filtered by Gauss filter before constructing Hessian matrix. In order to improve the performance, SURF adopts block filter instead of second-order Gauss filter, and uses integral graph to accelerate convolution speed and reduce computational complexity. When constructing scale pyramids, SURF does not sample downwards and repeat doing Gauss filtering like SIFT, but directly enlarges the size of the box to form image pyramids of different scales, thus improving the processing speed of the algorithm. To ensure the rotation invariance of feature points, the principal direction of the feature points must be chosen and then the SURF descriptor can be constructed.

3.2. Feature points matching

After getting the SURF feature vector, the next step is finding the matching point pairs using the SURF feature vector[7-9]. The similarity measurement based on Euclidean distance is often used in feature points matching. Traditionally, the nearest neighbor algorithm is used to find matching points. A feature point is selected in the source image, and the nearest feature point and the second nearest neighbor point are found in the target image. The ratio of the nearest neighbor distance to the second nearest neighbor distance is calculated. When the ratio is less than a threshold, a pair of matching points is considered to be found. K-d tree fast search algorithm are often used to find the nearest neighbor point quickly and BBF(Best Bin Fast) which is the improved version of K-d tree algorithm is more often used nowadays.

But all these matching algorithms don’t use the information of the camera. A new matching algorithm was proposed based on the camera information and BBF. The steps of the new algorithm are as flows.

1. Set a threshold of the x-coordinate of the SURF feature points, remove the feature points beyond the threshold which means these points are out of the overlapping area.
2. for every feature point in the left image, define a searching rectangular area in the right image. The x-coordinate of the center of the rectangular is the x-coordinate of the feature point in the left image plus T which is the transplating distance between the two cameras. The y-coordinate of the center of the rectangular is equal to the y-coordinate of the feature point in the left image.
3. The height of the rectangular can be set to 2 pixels and the width can be set from 20 to 30 pixels which can be decided after test.
4. for every feature point in the left image, use BBF to search matching points within the defined area. If there are no feature point can be matched within the rectangular area then abandon the feature point in the left image and begin to match the next feature point.

The new matching algorithm is based the fact that only in the defined area is it possible to find a feature point that can match the feature point in the left image. Theoretically, the matched point in the right image has the same y-coordinate as the matched point in the left image has and the x-coordinate equal to the x-coordinate of the feature point in the left image plus T. But because of the perspective error, the matched point coordinates may vary within a small interval, a searching rectangular area is
defined. So it can speed up the searching process and improve the matching precision by using BBF only in a small-size area. A large number of mismatched points in rough matching results will have a serious impact on the transformation matrix, so it is necessary to eliminate mismatched points. RASAC (Random Sample Consensus)[10] was used to refine matched point pairs. The matching result is showed in Fig. 4.

![Fig. 4 Matching result by using new matching algorithm proposed above](image)

3.3. Image blending
Because of the illumination difference between the two images and other factors, there will be a obvious seam in the stitched image generated by directly transforming the target image into the coordinate system of the original image through the transformation matrix. So, image blending is to eliminate the seam[11]. Image blending algorithms include direct average blending, weighted average blending and multi-resolution blending, etc. Weighted average blending algorithm is used here based on gradual-in-gradual-out which has a good blending result, as shown in the following figure. The basic idea of the weighted average blending algorithm is shown in the following formula.

\[
 f(x, y) = \begin{cases} 
 I_1(x, y) & (x, y) \in I_1(x, y) \\
 d_1(x, y) I_1(x, y) + d_2(x, y) I_2(x, y) & (x, y) \in I_1(x, y) \cap I_2(x, y) \\
 I_2(x, y) & (x, y) \in I_2(x, y) 
\end{cases}
\]

(2)

In the overlapping regions of images to be blended, the pixel values of mosaic images are weighted by two source images. The weighted values are \(d_1(x, y) = 1 - \frac{x - x_0}{\text{width}}\), \(d_2 = 1 - d_1\), \(x_i, x_0\) represent the horizontal coordinates of the pixels to be computed and the horizontal coordinates of the left boundary of the blended region respectively, and \(\text{width}\) represents the width of the blended region. From left to right, \(d_1\) decreases linearly from 1 to 0, and \(d_2\) increases linearly from 0 to 1. The weight changing sketch is shown in Fig. 5.

![Fig.5 The weight changing sketch of blending function](image)

4. Experiment results
In the experiment, using three different matching algorithms based on SURF descriptors to stitch two images taken by two cameras which have the same type and are fixed parallelly. The first matching algorithm is k-d tree, the second is BBF and the third is the algorithm proposed in this paper. Table. 1
shows the calculation time of these three algorithms. It can be seen that the algorithm proposed in this paper can substantially reduce the matching time compared with the other two. The stitched image is showed in Fig. 6.

| matching algorithms | Time (ms) |
|---------------------|-----------|
| k-d tree            | 1628.25   |
| BBF                 | 1204.67   |
| new algorithm       | 500.01    |

5. Conclusion

Based on the results and discussions presented above, the conclusions are obtained as below:

1. It is shown that the image stitching algorithm based on binocular vision can work well in sausage package detection line system.

2. The improved matching algorithm of BBF based on SURF descriptors can significantly speed up the feature points matching process which can meet the real-time requirements in sausage package detection line system.

References

[1] Brown, M. and Lowe, D. (2003). Recognizing panoramas. In Ninth International Conference on Computer Vision (ICCV’03), pages 1218–1225, Nice, France.

[2] Zhang, Z. “Flexible camera calibration by viewing a plane from unknown orientations,” Proceedings of the 7th International Conference on Computer Vision(pp. 666–673), Corfu, September 1999.

[3] Zhang, Z. “A flexible new technique for camera calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000): 1330–1334.

[4] Bouguet, J.-Y. “Pyramidal implementation of the Lucas Kanade featuretracker description of the algorithm,” http://robots.stanford.edu/cs223b04/algo_tracking.pdf.

[5] Lowe D G . Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[6] Bay, H. , Tuytelaraars, T. , and Gool, L. V. (2006). Surf: Speeded up robust features. In Leonards, A, Bischof, H. , and Pinz, A. , editors, Computer Vision – ECCV 2006, pages 404–417, Springer.

[7] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features.
International Journal of Computer Vision, 74(1): 59–73, 2007.
[8] R. Hartley and A. Zisserman. Multiview Geometry in Computer Vision. Cambridge University Press, 2004.
[9] R. Szeliski. Image alignment and stitching: a tutorial. Found. Trends. Comput. Graph. Vis., 2:1–104, 2006.
[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24: 381–395, June 1981.
[11] Z A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photo-montage. ACM Transactions on Graphics(SIGGRAPH), 23(3): 294–302, 2004.