

## Research on Large Space Building Fire Positioning Technology Using Video Surveillance Image

Ying Lu

Information and Control Engineering School ,  
Xi'an University of Architecture and Technology,  
Xi'an, China 710055  
applepeas@126.com

Hui-qin Wang

Information and Control Engineering School ,  
Xi'an University of Architecture and Technology,  
Xi'an, China 710055  
hqwang@xauat.edu.cn

Zhen-Bo Bi

Architecture School , Xi'an University of Architecture  
and Technology, Xi'an, China 710055  
Department of Computer Science, Zhejiang Ocean  
University, Zhoushan, China 316000  
bzb136@sina.com

Fei XU

Information and Control Engineering School ,  
Xi'an University of Architecture and Technology,  
Xi'an, China 710055  
326032513@qq.com

**Abstract**—The large space building fire source locating technology is an important link to automatic fire detection and extinguishing. This paper introduces a method that determines fire source location using three-dimensional reconstruction techniques based on accurate fire detection. The process includes: feature extraction, feature matching, and 3D spatial locating. The feature extraction process improves the scale invariant feature transform (SIFT) algorithm. It reduces feature points' 128-dimensional vector description to 16-dimension. The improved SIFT algorithm accelerates feature extraction while maintaining the good invariance of rotation, scale, and lamination factors from the original algorithm. To improve the positioning accuracy of the feature point matching process, this article adopts bi-directional matching method to achieve the exact match of image features according to the principle of unique match of feature points. In order to further improve the positioning accuracy, this article calculates the depth of the fire source from the scene by improving the binocular vision method and the use of the camera-cruise in determining different locations of a fire in the image. And finally it corrects the fire depth information by applying the confidence value.

**Keywords**- SIFT algorithm; fire vide; parallel binocular vision

### I. INTRODUCTION

Fires are often sudden and destructive, and pose serious threats to the safety of human life and property and the natural environment. Fire video, as the basis of image-type fire detection technology, possesses the advantages of real-timeliness and convenience, compared to traditional temperature/smoke sensing fire detection and fighting equipments. Some large buildings, once on fire, due to their interior height and spaciousness, the flames and smoke are often not detected in time before they can reach the traditional low-sensitivity alarm device. In these cases, fire video surveillance becomes particularly important. A fire detection and quenching system that based on video

surveillance can effectively detect a fire in its early stage, preventing further loss. In the early stage of video-based fire detection research, fire locating technology is crucial to the whole undertake of fire suppression. When a fire breaks out, finding the exact spatial location of the fire, and drive the high-pressure water guns for fire suppression, is of significant research and application value.

Research of positioning technology focuses on finding the target position in the background environment. Target position is selected by identifying consistent spatial point in two or multiple images generated by CCD cameras. With machine vision and digital image processing, nowadays technology can successfully restore three-dimensional fire spatial location, which is then used to drive the fire extinguishing water cannons. Distributed video fire detection products are already available on market at present, such ZANB-CFSD by Zhi An Bang Technology Co., an image-based fire detector that applicable to indoor large space and civil constructions. However, in order for the associated fire extinguishing system to put out the fire, it is imperative to accurately locate the fire after it is detected [1].

Binocular vision technology is inspired by how human eyes see objects [2][3]. Each human eye observes the three-dimensional world from its own angle and projects it onto a two-dimensional image plane. Image points scatter on different positions in two eyes. This difference in position is known as parallax, which reflects the depth of the object within the scene. Binocular vision technology is built upon parallax theory. Two cameras are used to sense the image parallax of one spatial object, and this parallax is then applied in the calculation to obtain the spatial depth of the observed object.

### II. MECHANISM OF BINOCULAR VISION

The most common binocular vision system is the one that is parallel and equal height. It consists of two identical cameras parallel placed, whose fronts are aligned

at the same height.

A binocular imaging model can be seen as a combination of the two monocular camera imaging model, which obtains two different images in the same space scenarios through two parallel cameras. Figure 1 illustrates the mechanism of binocular imaging model. The line connecting the centers of the two cameras are referred to as the baseline for the imaging system, indicated by letter B.

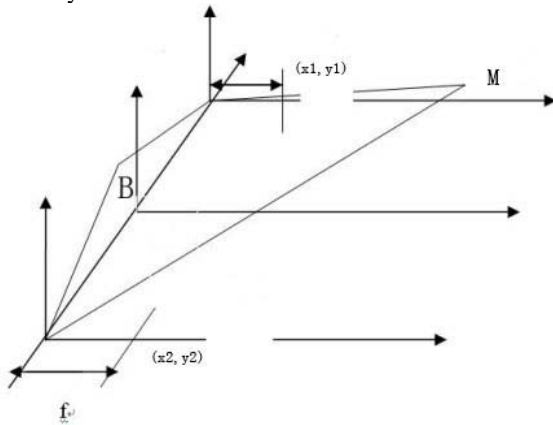


Figure 1. Binocular vision principle

By the mechanism of binocular imaging model [2][3], we can determine a spatial point's coordinates on each camera image as, for example,  $(x_1, y_1)$  and  $(x_2, y_2)$ . If the camera coordinate system and the world coordinate system coincide, then the image plane parallels the world coordinate system's XY plane, and the Z coordinate of object M should be the same in two camera coordinate systems as that in the world coordinate system.

Due to the fact that parallel binocular system uses two identical cameras, while the origins of each camera's coordinate system are different, the corresponding axes of which (i.e., optical axis) are parallel to each other. Therefore, we can analyze by simplify the binocular vision mechanism as shown in Fig 2. Fig 2 shows an illustration of the two camera connection plane.

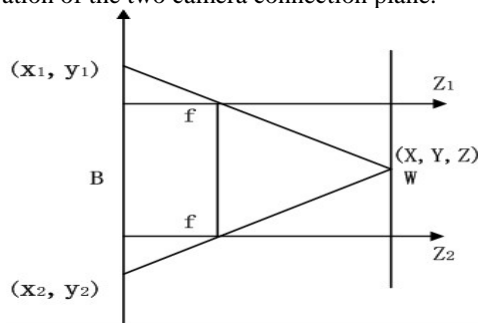


Figure 2. Parallax principle in binocular vision

As illustrated, B is baseline length and W's Z coordinate has the same value within two camera coordinate system. Placing two cameras into the world coordinate system where their origin point overlap, we can get their respective coordinate  $(X_1, Y_1, Z_1)$  and  $(X_2, Y_2, Z_2)$ . Since baseline length is B and Z value is identical in on both cameras' coordinate system, we can

draw the following equations:

$$\begin{aligned} X_2 &= X_1 + B \\ Z_2 &= Z_1 + Z \end{aligned} \quad (1)$$

Known from the camera imaging formation model that,

$$X_1 = \frac{x_1}{f}(f - Z) \quad (2)$$

$$X_1 + B = \frac{x_2}{f}(f - Z)$$

if we integrate these two equations, Z can be solved as:

$$Z = f - \frac{f \times B}{x_2 - x_1} \quad (3)$$

In this equation, Z is the distance between the object the image plane, i.e. the depth of the object in the three-dimensional scene, and  $(x_2 - x_1)$  stands for the parallax. Not hard to be seen that the depth information relates to the parallax as well as the baseline length. The accuracy of the match determines the accuracy of the parallax. Thus increasing baseline length can improve the calculation accuracy of the depth information, but will also increase the difficulty of matching, and reduce the accuracy of the parallax. We can reorganize the set of equations and obtain the spatial point's coordinates by binocular parallel visual formula set shown below, where X, Y, Z are the spatial coordinates of the observed spatial point;  $x_1, y_1$  and  $x_2, y_2$  are the coordinates of projection points of point M on respective cameras; f stands for the focal length of the observing camera; and finally B is the baseline length.

$$\begin{cases} Z = f - \frac{f \times B}{x_2 - x_1} \\ X = \frac{x_1}{f}(f - Z) \\ Y = \frac{y_1}{f}(f - Z) \end{cases} \quad (4)$$

### III. FIRE POSITIONING IN LARGE SPACE BUILDINGS

Figure 3 illustrates the process of locating fire source in large space buildings. The main difficulty lies in the extraction of feature points and identification of spatial depth. When a computer receives two or multiple images of a scene from different angles, it first records parameters of each camera when the image was taken, such as its position within the three-dimensional scene, etc. Then it captures and matches feature points from collected images, which is to build a corresponding relationship between two projections or among multiple projections of a spatial point. Finally, it identifies the spatial depth of the feature points using the relationship map we built in the last step, and calculates the actual location of the fire source.

#### A. Camera Calibration in Large Space Fire Scene

Camera calibration in large space fire detection and positioning scenarios is usually complicated, mainly due

to the abrupt changes of fire flames and uncertain backgrounds. With these characteristics, traditional camera calibration methods are destined to be inappropriate. Because of the industrial need for detection integration of the fire extinguishing system, it is required that these systems can adapt to different scenarios. Active Vision based calibration method is more suitable for the use of a monocular camera, but is not applicable in large space fire scenes due to the uncertainty of the scene

background sudden bursts of fire flames. Beside, using monocular camera is less accurate in positioning and more complex in calculation compared to those of dual camera calibration calculation. Therefore, in the large space fire source positioning technology, self-calibration method is most commonly used which is comparatively cheap and easy to experiment and therefore can meet with the real world requirements of the large space fire detection and positioning.

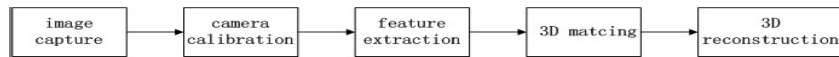


Figure 3. Flow chart of large space fire positioning with video images

### B. Fire Feature Point Extraction from Large Space Surveillance Video

Scale Invariant Feature Transform algorithm has strong ability to keep invariant of the rotation, scaling, and brightness change, and also maintains certain degree stability to perspective changes, affine transformation, and noise [4]. It carries rich information which can be used to extracted abundant feature points vector. Even scenes absent of large number of objects can produce considerable amount of SIFT feature vectors, this feature is essential for the reconstruction of three-dimensional information.

SIFT algorithm uses a 128-dimensional vector to represent each feature points in the image. However, it has limitations, such a data redundancy [5], which becomes especially obvious in live fire surveillance in large spaces where a longer processing time usually results in a reduction in images' instantaneity. Experiments have proved that the fourth step in SIFT algorithm, which employs a 128 feature point description, can take up considerable system time, gravely jeopardizing the real-timeliness of the algorithm. Therefore, this article has made some improvements to the traditional SIFT algorithm.

Specific improvements steps are as follows:

(1) Modify the direction distribution threshold value of key points.

After formation of multi-scale spatial feature points, during the detection of the key point direction, we lower the threshold value of the auxiliary direction so that when each key point histogram is generated, that pixel may produce multiple directions (a main direction and one or more auxiliary directions). The more directions of each pixel, the more directional characteristics it contains. Using such multi-directional pixels in describing key points can help maintaining more characteristics and the uniqueness of them, and this therefore enhances the robustness and stability of the key point matching [6].

(2) Reduce the vector number of feature points.

1) We use a 4\*4 window feature vector that centered on key point to represent a key point feature description. Shown in Figure 4, each grid increases auxiliary direction

of the pixel. It diminishes feature vector number on the key point, it nonetheless maintains matching stability.

2) Then on the 4 × 4 grid, as shown in Figure 5, we add in weighted information and get the statistics on 8 gradient directions of the key point.

3) In order to ensure illumination invariance, we normalize eight gradient directions. Assume we have a 8-dimensional feature point vector  $\bar{a} = [1,2,3,4,5,6,7,8]$ , we plug these values into formula (5) to obtain the normalized value.

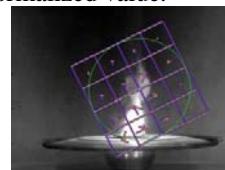


Figure 4. Eigenvector generated from feature points

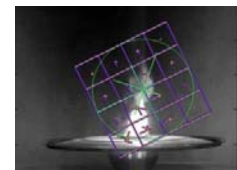


Figure 5. 8-dimensional vector feature of key points

$$\bar{a} = \frac{a}{\sqrt{\sum_{i=1}^8 a_i^2}} = (\bar{a}_1, \bar{a}_2, \bar{a}_3, \bar{a}_4, \bar{a}_5, \bar{a}_6, \bar{a}_7, \bar{a}_8) \quad (5)$$

4) Find the maximum gradient direction. After obtaining the 8-dimensional vector for the key point, we apply formula (6) to this vector which moves the largest element to the front and keeps the rest of the order remain still. If we input vector  $\bar{a} = [1, 2, 3, 4, 5, 6, 7, 8]$ , the output will be a new eigenvectors of the key points of  $\bar{a} = [8, 1, 2, 3, 4, 5, 6, 7]$ . This is to ..., which helps eliminating false matching in the subsequent step. Meanwhile, the use of the 8-dimensional vector matching and the removal of the seed point, can drastically promote matching velocity therefore shorten the time in extracting feature points.

$$\bar{a}_i = \max\{\bar{a}_i | \bar{a}_i \in \bar{a}\} \quad (6)$$

These improvements to SIFT algorithm are made specific to the characteristics of fire images. These improvements make the algorithm faster and more real-time than the original SIFT algorithm having about the same amount of feature points to be matched. While boosting performance, it also maintains a satisfying accuracy as well as rotation, scaling, and illumination invariance characteristics from the original SIFT algorithm.

### C. Large Space Fire Video Footage Feature Point Matching

This paper uses Euclidean distance (7) as primary theoretic basis of the similarity metric for key points within two images, where  $(x_1, y_1)$  and  $(x_2, y_2)$  each represents a spatial object's projection onto a camera. The  $d$ , i.e. the Euclidean distance, here means the similarity. The closer the key points, the more similar they are.

$$d = \text{sqrt}((x_1 - x_2)^2 + (y_1 - y_2)^2) \quad (7)$$

Euclidean distance calculation method is: First, take a feature key point from one image, and then find a key point in another image and accepts it as a matching point under the condition that their Euclidean distance is shortest. The main advantages of the using Euclidean distance to judge key point similarity are: easy calculation, low complexity, high real-timeliness, and high matching efficiency.

Because the feature point descriptor vector has reduced its original 128-dimension to a lower 8-dimensional, it loosens up the matching condition which enlarges chance of error. Therefore, in order to ensure the accuracy of point matching, i.e. to maintain the one-to-one mapping. This paper suggests bi-directional matching. The unique matching constraint condition means the one-to-one mapping of points between two images.

$$\{p \rightarrow p'\} = \{p' \rightarrow p\} \quad (8)$$

In our improved SIFT algorithm, we added in bi-directional Euclidean distance matching to offset accuracy loss by reducing descriptor dimension. Binocular positioning normally use two images, thus two Euclidean matching processes. If the two matching processes render the same pair of match, then the matching completes. Monocular video image sequence usually uses 6 frames per second. In the case of monocular video image, we divide 6 images into small group of the size of 2, and apply bi-directional matching within each.

### D. Large Space Fire Positioning

#### (1) Parallel binocular vision positioning

According to the national standard GB15631-2008 requirements for special fire detectors, with high-pressure water cannon auto fire-extinguishing systems, detected position of the fire source and its actual location should be no larger than 1.2 m. Large-space surveillance cameras have an effective monitoring range of 30-100 meters. Depending on the distance of protection, the camera perspective ranges between 20 ° -60 °. If the high-pressure water cannon has a protection distance of 100 meters, then the camera's viewing angle has to be 20 degrees horizontally and 15 degrees vertically to cover the fire spot. To improve the accuracy of the fire source positioning, the camera is usually fixed on the top of the high-pressure water cannons, and rotates along when the water cannons moves. When there is a displacement in position and angle, another group of value will be generated with the same method.

Through the above analyses, this paper designed fire source localization method with the use of two cameras

which comprise a binocular vision system. These cameras will scan through the fire site and record three images in which fire is located where it was first detected, on the left-most, and right-most, respectively. Then the computer will process these images and retrieve a spatial location of fire source out of each image. Finally, we add confidence value into three groups of data and correct the coordinates. Specific steps are as follows:

1) After a fire is detected, the system calculates the first set of three-dimensional spatial coordinates of the fire as  $(x_1, y_1, z_1)$  by binocular vision positioning.

2) The camera keeps collecting fire site images and rotating with the water cannon until the fire center reaches the left rim of the image. We mark down the coordinates of the conflagration area center as  $(i_1, j_1)$ , and also calculate the second set of fire location  $(x_2, y_2, z_2)$ .

3) Likewise, we record  $(i_2, j_2)$  and  $(x_3, y_3, z_3)$  when the camera rotates to an angle where within the captured image, fire reaches the right rim.

4) By adding confidence values  $a_1, a_2, a_3$  to three groups of spatial coordinates, where  $a_1, a_2, a_3$  satisfying  $a_1+a_2+a_3 = 1$ , we can have corrected spatial coordinates  $(x, y, z)$  of the fire.

5) In the final step, the system transmits the calculated fire source coordinate to the water cannon which is then used to drive the water hose in the fire suppression task.

#### (2) Determine spatial depth of a fire

A fire is not simply a pixel in a video footage. Fire pattern is formed by a plurality of coordinates or pixels. Therefore, when we try to represent a fire, we normally use the center point of the conflagration area. This center point's coordinates are averaged by matching feature points. However, this center point doesn't always correctly represent the true hot spot and causes errors in retrieving the fire's spatial depth.

In the previous sector, we have introduced our improved SIFT feature extraction algorithm which can be used to effectively capture a large amount of valid feature points. As discussed in the second chapter, we can calculate one  $z$  coordinate out of one matched feature point. Therefore if we obtain  $n$  valid points from feature matching process, it is convenient to get the average depth by plugging all  $z$  coordinates into formula (9) and calculate their mean value:

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i \quad (9)$$

## IV. RESULTS

We set up our experiment environment on a computer of CPU model E5300 2.9GHz and 2GB of memory. The simulation is carried out in Matlab 7.8 installed on Microsoft Windows XP SP3 operating system.

(1) Simulate improved SIFT algorithm in matching feature points using different types of fire images. Figure 6 reveals that the improved SIFT algorithm and the bi-directional matching algorithm perform better in matching accuracy.

(2) This experiment used parallel binocular visual

system and chose pictures from fire video library as subjects in the simulation.

To fully verify the validity of the algorithm and to find a convincing control group. We adopted experiment conditions used in referenced literature [7], in which baseline length was 100 mm, focal length 388mm, and acquired images in the size of  $768 \times 576$  pixels. We tested in different distance range. Results are shown in Table 1.

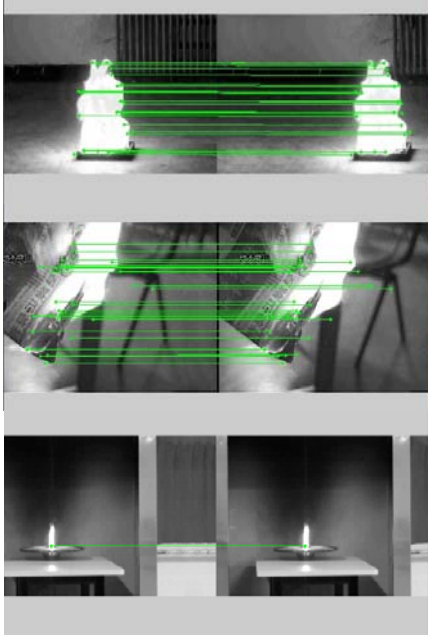


Figure 6. Feature Matching of Fire images

TABLE I COMPARISON OF MEASURE RESULTS (UNIT: MM)

No.	Actual distance	Results by [7]	Our results
1	800	831.4	819
2	1000	1023.3	1012
3	1200	1241.6	1231
4	1400	1432.6	1417

As can be seen from the data in Table 1, the improved spatial positioning algorithm, to a certain degree, enhances the recovery accuracy of the three-dimensional information of large spaces.

## V. CONCLUSION

This paper applies to video fire surveillance systems in the large spaces buildings. Such systems calculate the spatial locations of fires through diagnoses of videos and images taken by surveillance cameras. This paper improves the SIFT algorithm. The improved SIFT algorithm boosts its real-timeliness by reduces feature point extraction time while maintains the accuracy of the original algorithm as well as the invariance of rotate, scale, illumination. Feature point matching is based on Euclidean distance. It uses bi-directional matching which ensures matching accuracy. Recovery of three-dimensional depth information bases on parallel binocular vision mechanism. By calculating three groups of the fire source spatial data, and adding confidence value to the correction step, this improved algorithm also enhances

calculation accuracy of the depth information.

## ACKNOWLEDGMENT

This work was financially supported by Shaanxi Science Technology Research and Development Program Project(2011K17-04-01), Xi'an Industrial application technology research and development projects (CX1122(2)), Construction Science and Technology Projects for Xi'an City Urban and Rural Construction Committee (2011008) and Beilin District, Xi'an City Science and Technology Project(GX1104).

## REFERENCE

- [1] Jiang, B. Research and development on intelligent water spray fire extinguisher [D]. Nanjing University of Science and Technology, 2004.
- [2] Liu, C. Image understanding and computer vision [M]. Xiamen University Press, 9.2009
- [3] Duan, F., Wang, Y., Lei, X. etc. Review of machine vision technology and its application[J]. Automation Browse. 19(3): 59-61.2002
- [4] Guo, L., Liao, Y., Chen, W., etc. Research on improved SIFT based super resolution video reconstruction and fast calibration [J]. Journal of Hubei Institute for Nationalities, 28(02): 177-183, 2008
- [5] Cao, J., Li, X., Lin, W., etc. Research on SIFT Feature matching algorithm improvements [J]. System simulation journal, 22(11): 2760-2763, 2010
- [6] David G. Lowe. Distinctive Image Features from Scale-Invariant Key-points [D]. Computer Science Department University of British Columbia Vancouver, B.C., Canada.2004
- [7] Meng, H., Cheng, K., SIFT feature points based binocular vision positioning [J].Journal of Harbin Engineering University, 30(06): 649-652, 2009
- [8] Plinio Moreno , Alexandre Bernardino. Improving the SIFT descriptor with smooth derivative filters[D]. Pattern Recognition Letters.2008
- [9] Canlin Li , Lizhuang Ma. A new framework for feature descriptor based on SIFT[J].Pattern Recognition Letters.2009. Volume 30, Issue 5 , 544-557
- [10] David G. Lowe. Local Feature View Clustering for 3D Object Recognition [D]. Computer Science Department University of British Columbia Vancouver, B.C., Canada.2001
- [11] David G. Lowe. Object Recognition from Local Scale-Invariant Features [D]. Computer Science Department University of British Columbia Vancouver, B.C., Canada.1999.