ISSN: 2664-9640

Global Constraint with Scale-invariant Feature Matching for Visual Hull Computation

Feiyue Hu[1, a] and Li Shi[2, b, *]

[1]Logistics Engineering College, Shanghai Maritime University, Shanghai 200135, China;

[2]Logistics Engineering College, Shanghai Maritime University, Shanghai 200135, China.

[a]201730210145@stu.shmtu.edu.cn, [b]Corresponding author Email:shili@shmtu.edu.cn

Abstract

For changes in the Angle of the three-dimensional reconstruction of image sequence and high repetition structure of stereo matching error matching rate problem, we put forward a kind of constraint based on global scale invariant features matching, the algorithm based on SIFT points to extract the image feature and match, points to the main direction for the polar axis, using histogram statistics and other matching point distribution using sieve distribution similarity in addition to miss match point, reduce the error matching rate. The experimental results show that, compared with the polar constraint, it can increase the correct matching point and reduce the wrong matching point at the same time, reducing the time consumption by two orders of magnitude, improving the real-time performance, and being able to reconstruct the model with richer details.

Keywords

Constraint, Global scale invariant, mismatching points, real-time.

1. Introduction and Related Work

Professor Marr proposed has the most perfect computer visual information processing system framework, the framework's main goal is to extract the two-dimensional image information to restore 3D scene, to extract the 3D object pose, accurate estimates of the 3D model, realize the computer aided geometric design (CAGD) and computer graphics (CG), computer animation, computer vision, medical image processing, scientific computing and virtual reality, digital media creation application scenario.

There are two key tasks in 3D reconstruction: one is to calibrate the internal and external parameters of the camera; the other is to reconstruct the 3D geometric structure of the scene. Even though the theory of view geometry for 3D reconstruction has become mature, there is still much room for improvement in optimality, robustness and efficiency. To improve the robustness of stereo matching algorithm to improve the accuracy of internal and external parameters of the camera, the selection of efficient feature extraction and stereo matching algorithm can improve the real-time performance of 3D reconstruction. Generally, the stereo matching algorithm that can generate dense point clouds has poor real-time performance. The real-time improvement of 3D reconstruction technology can make the technology apply to more scenes.

After decades of research at home and abroad, 3D reconstruction technology based on image point features has received a lot of research results in theory and application.

Detection of image point features: Lowe[1] proposed the difference of DOG(difference of Gaussian) algorithm, which used the difference detection algorithm of Gaussian filter with two scales to detect corner points, filtered out the points and regions that did not affect the structure,

ISSN: 2664-9640

and took the extreme point as corner points. Harris-affine, hessian-affine and other detection algorithms for corner position obtained by convolution of gradient covariance matrix and image [2]; SIFT[1](scale-independent feature transform) based on the DOG algorithm USES a 5-layer image pyramid to perform DOG filtering with scale invariance and rotation invariance, which is one of the feature point detection methods with the best performance at present.

Description of image feature points: Mikolajczyk and Schmid[3] compared the performance of a variety of feature descriptors, SIFT descriptors increase the image brightness feature to make it invariant; Ke and Sukthankar[4] proposed PCA-SIFT based on SIFT and used principal component analysis to reduce the dimension of gradient vector. Bay by approximate method to calculate the gradient and the method of integral optimization of the SIFT algorithm, proposed the SURF (speeded up robust feature).

Feature points matching: the easiest way to make use of Euclidean distance directly performance characteristics of the degree of acquaintance, Hua and Brown [5] setting the threshold value using machine learning method is put forward, in the camera model simulation shooting multiple sets of samples as the training data, learned different characteristics of the threshold, the nearest neighbor in the feature space descriptor selected the most similar match. Kiana [6] proposed an index structure using multidimensional search tree, based on the priority search strategy of hierarchical k-means tree, and adopted geometric verification to further reduce mismatches.

Camera calibration: the most widely used ilf planar checkerboard used camera calibration method has high accuracy, but a more ideal application scenario is don't need calibration method of self-calibration Heyden [7-8] within three image parameters, such as unchanged under the premise of Kruppa equation and absolute conic and its dual imaging invariance to achieve the calibration of camera external parameters.

Multi-view stereo: Seitz [9] proposed an algorithm based on voxel coloring to calculate a cost function to extract object surface model from 3d volume. Yang [10] proposed a surface evolution cost function through iteration, which gradually shrinks inward from a large initial volume and makes surface reconstruction denser. Liu [11] proposed a method to create a 3d scene based on a series of depth map information fusion with consistency applied. Furukawa [12] proposed an algorithm to establish a quasi-dense three-dimensional point cloud directly through stereo matching and directly reconstruct the three-dimensional surface, which has the best overall performance at present.

For stereo matching Angle change is bigger, and repeat architecture matching error rate higher problem, this paper proposes a constraint based on global scale invariant features matching, the algorithm based on SIFT points to extract the image feature and match, points to the main direction for the polar axis, using histogram statistics and other matching point distribution using sieve distribution similarity in addition to miss match point, reduce the error matching rate. The experimental results show that compared with the pole-line constraint, it can increase the correct matching point and reduce the wrong matching point at the same time, reducing the time consumption by two orders of magnitude and improving the real-time performance.

2. Problem Formulation

2.1. Feature Detection and Description

Scale-invariant feature transform (SIFT) algorithm is a description of image features. It looks for key points in the image and extracts the location, Scale and rotation invariants. It is an image local feature descriptor based on key points. This algorithm has scale invariance, that is, it maintains invariance for rotation, scale scaling and brightness change, and maintains a certain degree of stability for perspective change, affine transformation and noise [13].

ISSN: 2664-9640

(1) Construct scale space, make difference to generate gaussian difference scale (DOG) space, and look for key points. The gaussian kernel is convoluted with the original image to obtain image sequences with different degrees of blur. The bottom layer of the sequence is the resolution of the original image, and the second layer is half of the size of the original image to form an image pyramid. Search for the extremum point in the image pyramid, and take 9 points from top to bottom of the image pyramid for three consecutive layers. If the middle point is the maximum or minimum, this point is considered as the key point.

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{1}$$

$$D(x,y,\sigma)=\left[G(x,y,k\sigma)-G(x,y,\sigma)\right]*I(x,y) \tag{2}$$

σ is scale.

(2) Select key points, remove low contrast points, edge points, improve stability.

$$\begin{aligned} &D(\hat{x})=D+\frac{1}{2}\frac{\partial D^T}{\partial x}\hat{x} \\ &\left|D(\hat{x})\right|\geq T \end{aligned} \tag{3}$$

(3) The characteristic direction of key points is extracted to ensure the invariance of rotation and scale. A vector representing the feature of key points is obtained for the corresponding position of key points of each layer of images, and the main feature vector of a key point is obtained by statistics.

$$\begin{aligned} &m(x,y)=sqrt\left(\left[L(x+1,y)-L(x-1,y)\right]^2+\left[L(x,y+1)-L(x,y-1)\right]^2\right) \\ &\theta(x,y)=\tan^{-1}\left[\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right] \end{aligned} \tag{4}$$

(4) Feature descriptors are generated, and a 4×4 square is depicted in the key points. The edge of the square is parallel to the main direction of the key points. The gradient direction and amplitude of each pixel are counted in each square.

2.2. Feature Point Matching

Feature point matching, find the Euclidean distance between one feature point description of an image and all feature points description of another image, and generate a set of matching points when the Euclidean distance takes the minimum value. Then verify one by one to get the sequence of matching points.

ISSN: 2664-9640

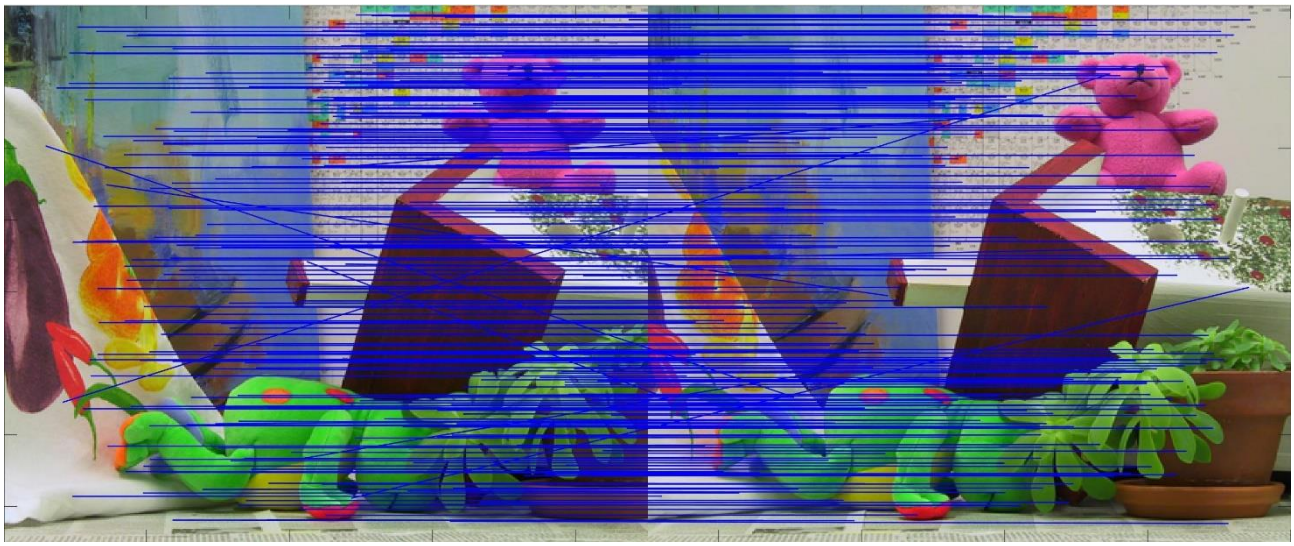

Fig 1. Stereo matching of Middlebury-Teddy with SIFT

The smaller the Euclidean distance allowed by the matching points, the higher the robustness of point matching. The smaller the number of corresponding points that are successfully matched is not enough to generate the parallax graph, and the lower the matching conditions, the more dense stereo matching can be generated. The matching results of the non-polar constraint or the global scale-invariant feature matching constraint are shown in figure 1. There are many repetitive structures in the left and right images, and there are obvious mismatching points.

2.3. Feature Point Matching

The camera imaging model USES the perspective camera model, and the field attractions in the camera coordinate system are projected on the imaging plane using the perspective projection principle. The model involves the transformation of multiple coordinate systems. Firstly, through rigid body transformation, the homogeneous coordinates of the field and scenic spot in the world coordinate system are transformed to the camera coordinate system. The transformation matrix in the middle is the external parameter, which determines the camera pose. Secondly, the similarity triangle principle is used to transform the three points into the points on the two-dimensional image. Finally, the point coordinates on the two-dimensional image are converted into pixel coordinates in pixels. The relationship between three-dimensional points $\mathbf{U}=(X,Y,Z,1)$ and image pixel $\mathbf{u}=(u,v,1)^{\top}$ coordinates:

$$u \simeq \mathbf{K}[\mathbf{R}\,|\,-\mathbf{R}\mathbf{t}]\mathbf{U} \tag{5}$$

$\mathbf{K}$ are internal parameter matrices. Generally, $\mathbf{K}=diag(f,f,1)$, f is the focal length of the camera. $\mathbf{R}$ and $\mathbf{t}$ are rotation and translation transformation matrices, collectively referred to as external camera parameters.

2.4. Multi-view Stereo

In the absence of noise, the line of sight of the same 3d point meets at a point in different images, and the actual noise and external point will inevitably affect the triangular modeling of the calibration view. Therefore, 3d reconstruction needs a global optimal multi-view triangulation method to obtain the estimation of 3d model. For the non-convex hull, the image information is obscured and all surface details cannot be obtained. Virtual engraving technology is adopted to

ISSN: 2664-9640

squeeze and cut the initial volume element through constraints, which can obtain more surface details.

Computing 3D bounding box:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{K} * \begin{pmatrix} P_{11}^i & P_{12}^i & P_{13}^i & P_{14}^i \\ P_{21}^i & P_{22}^i & P_{23}^i & P_{24}^i \\ P_{31}^i & P_{32}^i & P_{33}^i & P_{34}^i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{6}$$

P^i is the projection matrix for picture i.

$$K = P_{31}^i * x + P_{32}^i * y + P_{33}^i * z + P_{34}^i \tag{7}$$

The 2D border of the image wl_i , wh_i , hl_i , hh_i satisfies the following inequalities:

$$\begin{aligned} wl_i \le u = \frac{P_{11}^i * x + P_{12}^i * y + P_{13}^i * z + P_{14}^i}{P_{31}^i * x + P_{32}^i * y + P_{33}^i * z + P_{34}^i} \le wh_i \\ hl_i \le u = \frac{P_{21}^i * x + P_{22}^i * y + P_{23}^i * z + P_{24}^i}{P_{31}^i * x + P_{32}^i * y + P_{33}^i * z + P_{34}^i} \le hh_i \end{aligned} \tag{8}$$

4N linear inequalities, Each individual is a 3D dot (x,y,z). If you compute the maximum and minimum of x, the objective function are $f_{obj} = x$ and $f_{obj} = -x$.Select individuals according to fitness functions, If there are individuals that do not satisfy 4N inequalities, they are discarded, and those that satisfy the conditions are saved as optimized ones, and the octree of Visual Hull is established. For a 3D point v, the isosurface function:

$$f_{iso}(v) = \max_i D_i(P_i * v), i = 1, 2, \cdots, \tag{9}$$

D_i is the chamfer distance transformation of the contour, Minus inside, plus outside. If all 8 vertices are inside Visual Hull, that is, the voxels projected on all images are within the contour, the output type is in, otherwise, it is on. If part of the pixel targets are on the contour, then the polygon intersects the contour.

3. Global Constraint with Scale-invariant Feature Matching

Pole-line constraints, within the known left and right camera parameters, Al, Ar and two camera structure parameters R and T, The polar line constraint can be used to verify the matching point homography. The matching points in the right graph should correspond to the polar line of the point in the left graph, and the matching points in the left graph should correspond to the polar line of the point in the right graph. The result of pole line constraint filtering mismatching points is shown in figure 2.

ISSN: 2664-9640

$$\mathbf{F} = \mathbf{A}_r^{-\top}\mathbf{S}\mathbf{R}\mathbf{A}_l^{-1}$$
$$\mathbf{p}_r^{\top}\mathbf{F}\mathbf{p}_l = 0$$
$$\mathbf{S} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \tag{10}$$

$\mathbf{t} = \left(t_x, t_y, t_z\right)^{\top}$ is the translation vector.

Fig 1. Eliminate mismatches with polar constraints

With the constraint of feature matching with constant global scale, the polar coordinate system is established in the main direction of its key points, and the coordinate system is divided into 8 equal regions with 45° as the interval. The location distribution of all other matching points is investigated. If the number of point matches in each region is similar, the matching of corresponding points is the correct match. Calculate the coordinates of all other key points in this polar coordinate system. The polar Angle is used as the basis for the division.

The constraint theory of global scale invariant feature matching is based on the fact that there are supported point matches around the correct point matches and the wrong point matches are isolated. The difference between scale-invariant feature transformation and global scale-invariant feature matching constraint is that the former is a local feature descriptor, while the latter is a constraint on the global distribution invariance of feature points in binocular vision system.

$$\theta = -\arcsin\left[\frac{\mathbf{ab}}{|\mathbf{a}||\mathbf{b}|}\right] - \arcsin\left[\frac{x_1 y_2 - x_2 y_1}{|\mathbf{a}||\mathbf{b}|}\right] \tag{11}$$

$\left(x_1, y_1\right)$ is the coordinates of key points as poles, $\left(x_2, y_2\right)$ is the coordinates of any other key point, $\mathbf{a}$ is the principal direction vector of the key point as the pole, $\mathbf{b}$ is the vector that is different from the coordinates of any other key point.

ISSN: 2664-9640

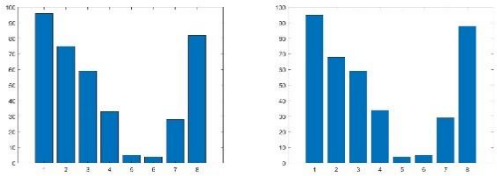

Fig 3. Histogram of robust matching

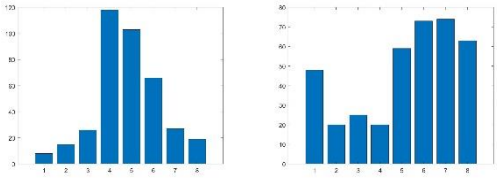

Fig 4. Histogram of mismatching

SIFT features were extracted and matched, and the Euclidean distance allowed by SIFT feature matching descriptor pairing was gradually increased. As the matching points became increasingly dense, the number of mismatching points gradually increased. The polar constraint and the global scale-invariant feature matching constraint were used to screen out the mismatching points.

Using polar constraint screen out false matching points, matching point right matching point distance left on the right of the corresponding polar distance less than the threshold of a tiny, right and left to match point distance matching point in the left corresponding polar distance is less than a minimum threshold, then the matching points for the correct matching points, otherwise identified as false matching points, to screen out it.

4. Experiment

Use global scale invariant feature matching constraint screen out false matching points, the correct matching points along the main direction of polar axis in polar coordinates, similar to other matching point distribution, the distribution of the histogram is slightly offset, collect two images is the cause of the deviation Angle is different, even large Angle changes, histogram remained similar distribution, about two small image histogram Euclidean distance. The histogram of the left and right of the mismatched points is completely irrelevant, and the histogram of the key points distribution of the left and right images can be screened out by the large Euclidean distance.

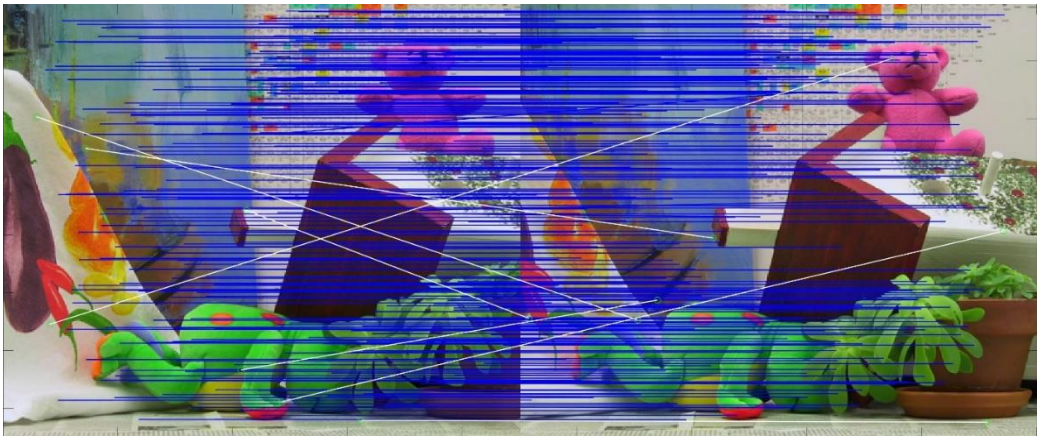

Fig 5. Performance of global constraint for scale-invariant feature matching

ISSN: 2664-9640

Table 1. Performance comparison of eliminating mismatches

	SIFT+Polar constraints	SIFT+ Global Constraint with Scale-invariant Feature Matching
Verify 100 matching points time /(s)	0.1943	0.0098
Screen out the number of mismatches	6	6
Whether reverse validation is required	Yes	No

As shown in table 1, the image matching algorithm is used to obtain a number of corresponding point matches, and the pole-line constraint is used to screen out the mismatches. The time to verify each 100 points is 20 times longer than the time to use the global scale invariant constraint. Screening effect is basically the same.

Visual Hull was calculated using the marching cubes algorithm to extract the surface.

The 3d reconstruction results of the data set of The Digital Michelangelo Project are shown in figure 6:

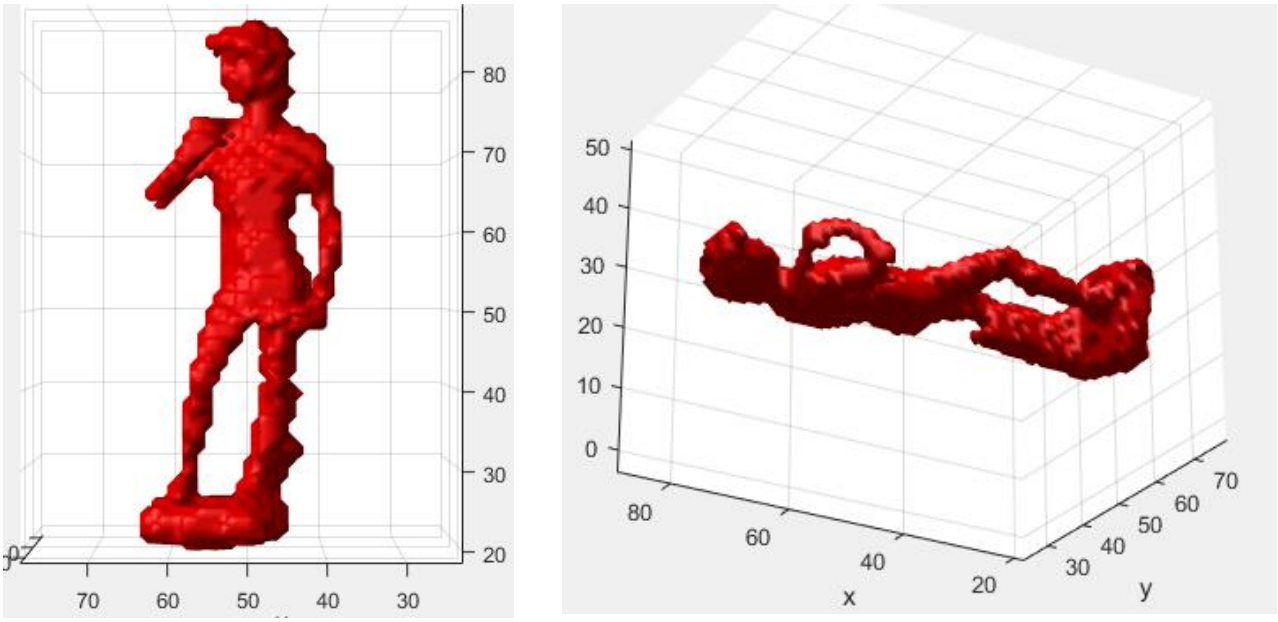

Fig 6. 3d reconstruction results of the data set of The Digital Michelangelo Project

Fig 7. Shows the 3d reconstruction results of Middlebury multi-view stereo:

ISSN: 2664-9640

Fig 7. Results of Middlebury multi-view stereo 3d reconstruction

5. Summary

This paper studies the algorithm of image matching based on global scale invariant feature constraint to screen out mismatched points and calculates Visual Hull. Firstly, the image matching algorithm is used to extract the original corresponding point matching of two images, and the stereoscopic matching results are screened by using the pole-line constraint and the feature matching constraint based on the global scale invariant. Through the method in this paper, image matching results with more repetitive structures in Middlebury data set were optimized, and common polar line constraints were compared to filter out the same false matching points and at the same time the speed was an order of magnitude faster than the polar line constraint, so that the reconstruction details were richer.

References

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

[2] T. Tuytelaars, K. Mikolajczyk: Local Invariant Feature Detectors: A Survey [J]. Foundations and Trends® in Computer Graphics and Vision, 2007, 3(3):177-280.

[3] K. Mikolajczyk, C. Schmid: A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10):1615-1630.

[4] N.Y. Ke, R. Sukthankar: PCA-SIFT: a more distinctive representation for local image descriptors[C].Proceedings of Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2004.

[5] E.A. Gang: Discriminant Embedding for Local Image Descriptors[C].Proceedings of IEEE International Conference on Computer Vision. IEEE, 2007.

[6] K. Hajebi, Y. Abbasi-Yadkori and H. Shahbazi: Fast Approximate Nearest-Neighbor Search with k-Nearest Neighbor Graph[C].Proceedings of IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011. DBLP, 2011.

ISSN: 2664-9640

[7] A. Heyden, K. Astrom: Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point.[C].Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR),1997:438-443.

[8] A. Heyden, K. Astrom: Flexible Calibration: Minimal Cases for Auto-calibration[C].Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition(ICCV),1999:350-355.

[9] A. Treuille, A. Hertzmann and S.M. Seitz: Example-Based Stereo with General BRDFs [J].Lecture Notes in Computer Science, 2004, Vol II:457--469.

[10] R. Yang, M. Pollefeys and G. Welch: Dealing with Textureless Regions and Specular Highlights-A Progressive Space Carving Scheme Using a Novel Photo-consistency Measure[C].Proceedings of IEEE International Conference on Computer Vision. IEEE, 2008.

[11] Y. Liu, X. Cao, Q. Dai, et al. Continuous depth estimation for multi-view stereo[C].Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)Workshops, 2009:2121-2128.

[12] Y. Furukawa, J. Ponce: Accurate, Dense, and Robust Multiview Stereopsis [J].Proceedings of IEEE Transactions on Software Engineering, 2010, 32(8):1362-1376.