An efficent method of binocular data reconstruction

YunBo Rao[1], Xianshu Ding[1], Bojiang Fan[1]
[1] School of Information and Software Engineering,University of Electronic Science and Technology of China, Chengdu, 610054, P.R China
[e-mail: uestc2008@126.com]
*Corresponding author: Yunbo Rao

Received April 12, 2015; accepted June 21, 2015; published September 30, 2015

Abstract

3D reconstruction based on binocular data is significant to machine vision. In our method, we propose a new and high efficiency 3D reconstruction approach by using a consumer camera aiming to: 1) address the configuration problem of dual camera in the binocular reconstruction system; 2) address stereo matching can hardly be done well problem in both time computing and precision. The kernel feature is firstly proposed in calibration stage to rectify the epipolar. Then, we segment the objects in the camera into background and foreground, for which system obtains the disparity by different method: local window matching and kernel feature-based matching. Extensive experiments demonstrate our proposed algorithm represents accurate 3D model.

Keywords: 3D reconstruction, binocular data, stereo matching, local window matching, kernel feature

A preliminary version of this paper appeared in IEEE ICC 2009, June 14-18, Dresden, Germany. This version includes a concrete analysis and supporting implementation results on MICAz sensor nodes. This research was supported by the National Natural Science Foundation of China (Grant No.61300092), the Fundamental Research Funds for the Central Universities of China, (Grant No. ZYGX2013J068).

http://dx.doi.org/10.3837/tiis.2015.09.024 **ISSN : 1976-7277**

1. Introduction

3D reconstruction is an important research topic in many fields, like computer vision and computer graphic[1]. 3D model has been playing a vital role in many applications, such as medical image processing, object recognition, segmentation and mapping[2-5]. The traditional approaches apply the geometric modeling technology, obtaining the surface depth of object directly by 3D scanning equipment which are so expensive that can not be generalized. However, reconstructing based on vision(images) is more popular and more likely to be accepted by different kinds of users. The main theory of vision-based reconstruction is to extract depth from the vision information. For example, the 2D features like shading, contours, voxel, give some cue of depth information. The widely researched technology can be classified by the number of input images. Some of them use only one image while others employ two and even more. Binocular vision reconstruction which makes a trade-off between accuracy and time efficiency has been the focus of vision-based 3D reconstruction.

Binocular stereo vision attracts much attention due to it simulates our human eyes which can fuse two images into a stereo mapping. Hence, the system uses two industrial CCD cameras to get two images of an object under the same optical condition in different viewpoints, and then extracts the depth information according to parallax of 2D images and the pixel disparity [10-12]. The main processing flow is as follow: camera-self calibration, feature-based stereo matching, depth solving from disparity, and 3D representation. Camera calibration aims to be capable of delivering lens distortion compensated images[6]. So most of systems try to employ high-precision cameras to avoid large lens distortion that would cause large mistake to intrinsic and external parameters. Among all algorithms, depth is computed directly by projecting points for matching into the image views. Stereo matching which is the fatal step in the processing can be achieved by local window matching or feature-based matching. The former can compute disparities of all pixels by matching, while the later just extract some feature points, and others would be computed by interpolation method.

As mentioned, most exiting approaches have various limitations in binocular data reconstruction. Our work is an extension of our previous work presented in [13]. Besides more concise representation, the major contributions are:

(1)an efficent 3D reconstruction approach by using a consumer camera is presented aiming to: 1) address the configuration problem of dual camera in the binocular reconstruction system; 2)address stereo matching can hardly be done well problem in both time computing and precision;

(2)we use ORB feature instead Kernel feature in the feature extraction stage, and also give comparion using our method and related works, such as [14];

(3)image enhancement technology is applied before the segmentation step for the consideration of reducing the light environment variation in two views;

(4)to improve the accuracy of disparity map, we improve Random Sample Consensus (RANSAC) method to delimit errant matched points, and even apply median filtering method to denoising the disparity map;

The remainder of the paper is organized as follow: the related work of binocular data reconstruction is introduced in Section 2. The proposed method is descripted in Section 3. The

final experimental settings and results are demonstrated in Section 4. Finally, we make a conclusion objectively in Section 5.

2. Related Work

The computer vision communities has developed many techniques for 3D mapping using monocular cameras, unsorted collections of photos, and even range scans[15,16], such as Kinect from Microsoft [17]. Most of 3D techniques require these following: feature extraction and matching, spatial alignment. However, in the vision-based 3D mapping, techniques about feature become the only relevant one.

Robotics vision using consumer scans like Kinect, utilizes joint optimization algorithm combining visual features and shape-based alignment, and visual and depth information are also combined for view-based loop-closure detection to achieve globally consistent maps. Nevertheless, this RGB-D mapping can only be applied to indoor environments due the camera limits. Hsiao et al [11] propose a approach to obtain the high-resolution disparity map from a low-resolution binocular image pair, using region-based fusion and refinement techniques. Bradley et al [12] present their algorithm for multi-view reconstruction using adaptive point-based filtering of the merged point clouds, and efficient, high-quality mesh generation. All geometry processing algorithms work only on local neighborhoods, and have the time complexity of O(k log k), where k is the number of points processed by the respective algorithm. Since the binocular stereo part is linear in the number of images, the complete algorithm remains highly scalable despite the high quality reconstructions it produces, which demonstrate one of the fastest techniques.

As one of the most popular techniques in vision-based 3D reconstruction, binocular data 3D reconstruction require these: camera-self calibration, stereo matching, depth extraction, and 3D representation. The traditional features, such as Scale Invariant Feature Transform (SIFT [18]), Principal Component Analysis (PCA)-SIFT and Speeded Up Robust Features (SURF [19]), are the most accepted ones in computer vision communities. SURF and PCA-SIFT are both the variants of SIFT, however, they has different weakness. Juan et al.[20] summarizes the three robust feature detection methods: SIFT presents its stability in most situations but it's slow; SURF is the fastest one with worse performance than SIFT; PCA-SIFT show its advantages in rotation and illumination changes, but sensitive to other variants. In recent years, some new features are proposed instead. Bo et al [21] propose several depth kernel descriptors that capture different recognition cues including size, shape and edges (depth discontinuities). Experiments demonstrates that matched patches in kernel view are proved highly accurate for object recognition. However, kernel computing is too expensive to general memory setting, so it is too computationally intensive for use in real-time applications of any complexity. Rosten et al.[22] propose a faster feature which is based on non-maximal suppression. But the proposed feature also suffers from a number of disadvantages: it is not robust to high level noise.It is dependent on a threshold, and it can respond to 1 pixel wide lines at certain angles. Rublee et al.[14] propose a very fast binary descriptor based on BRIEF [23], called ORB, which is rotation invariant and resistant to noise.

In this paper, we employ image enhancement technique to reduce the light variants in two views. About enhancement techniques, the existing techniques of image enhancement can be classified into two categories: self-enhancement and context-based fusion enhancement [24], in which fusion method use light information from other images. self-enhancement cannot resolve the problem of low quality. The reason is that in the dark environment, some areas are

so dark that all the information is already lost in those regions. No matter how much illumination enhancement you apply, it will not be able to bring back lost information. Context-based fusion enhancement refers to high quality environment, which fuse illumination information in different time image. The approach is that it is by extracting high quality background information to embed low quality image/video. So in our algorithm, we use pairs from two views to help each other to equalize the light environment, more detail description, please see previous work in [24,25].

3. The proposed method

In this paper, we propose a new and high-efficiency processing algorithm, which uses two consumer digital camera to avoid the large cost of mechanical settings. Most of the advanced machine vision cameras contribute best to make the two CCD on the same epipolar line. Consumer camera settings may make epipolar line much more warped, nevertheless, we revise the epipolar line by kernel feature calibration. About the stereo matching, in order to reduce the time consumption, we match the foreground objects by window to acquire confident disparity, while we use kernel features to label background to avoid largely computing.

3.1 Camera-self calibration and epipolar rectification

In the binocular vision simulation system, two cameras that have the same intrinsic parameters are arranged in a line, their focal length is identical and the two retinal planes are the same. This parallel binocular reconstruction principle model is show in **Fig.1**. The pixel of point P is projected on the left and right images, $(x1, y1)$ and $(x2, y2)$, respectively. Under ideal condition, the two images will be in the same plane, namely $(x1 = x2)$. According to triangular geometry principle, the distance Z can be computed by the following equation:

$$Z = \frac{Tf}{x1 - x2} \tag{1}$$

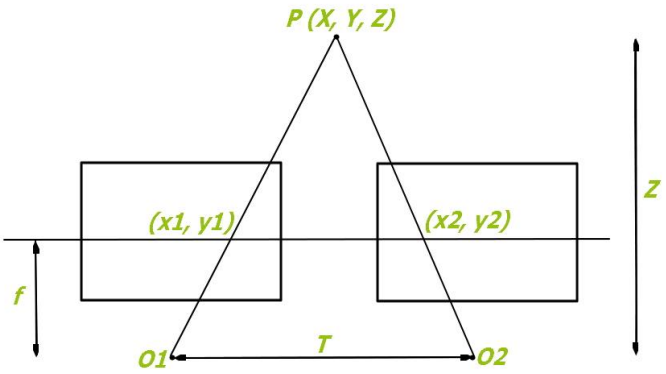

Fig. 1. Binocular stereo imaging principle diagram.

So in the camera-self calibration stage, obtaining other parameters, like camera focal length f, baseline distance T, is also an important task. In this research, we employ Zhang Zhengyou's calibration method [7], which has been the most popular these years.

The computing is under ideal condition, however, we can not guarantee $(y1 = y2)$, especially in the consumer camera systems. In order to correct epipolar, we propose gradient kernel feature to match images. Kernel function is equivalent to measuring the similarity of image patches using a liner kernel in the feature map in the Kernel space:

$$K(P,Q) = \sum_{x \in P} \sum_{x^1} (\delta(x))^T \delta(x^1) \tag{2}$$

where P and Q are patches in two matched images. $(\delta(x))^T \delta(x^1)$ are the inner product of two vectors which are positive definite kernels. The most helpful point of kernel view is that it can map the definite feature into higher dimensions in which these features are linearly separable ones. In this study, we use gradient kernel descriptors listed as follow:

$$K_{grad}(P,Q) = \sum_{x \in P} \sum_{x^1 \in Q} m(x) m(x^1) k_0(\theta(x), \theta(x^1)) k_p(x, x^1) \tag{3}$$

where $k_p(x, x^1)$ is a Gaussian position kernel with x denoting the 2D position of a pixel in an image patch. k_0 is another Gaussian kernel over orientations. Three kernel feature points in the left image are $Al(x11, yl1)$, $Bl(x12, yl2)$, $Cl(x13, yl3)$, and in the right image are $Ar(xr1, xr1)$, $Br(xr2, xr2)$, $Cr(xr3, xr3)$. Then we have the following affine transformation:

$$\begin{bmatrix} xl1 & yl1 & 1 \\ xl2 & yl2 & 1 \\ xl3 & yl3 & 1 \end{bmatrix} = H \begin{bmatrix} xr1 & xr2 & xr3 \\ yr1 & yr2 & yr3 \\ 1 & 1 & 1 \end{bmatrix} \tag{4}$$

where H is the affine matrix, which can be appled to other pixels to correct y coordinate. As shown in **Fig.2**. the left is the original image (left), the middle is right image from the same consumer image, and the right is the affine revised image.

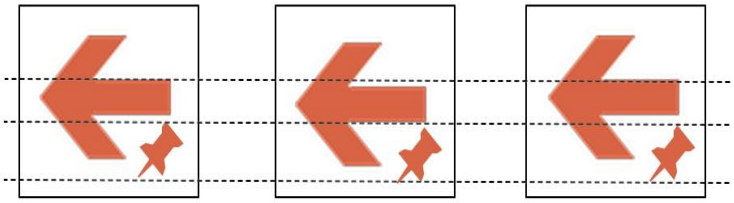

Fig. 2. Epipolar line rectification.

However, the kernel computing is too expensive to general memory setting, to be used in real-time applications. So we have to find a better feature detector instead. ORB (Oriented Fast and Rotated Brief) which builds on well-known FAST keypoint detector and the well-developed BRIEF descriptor is attractive because of its good performance and low cost. It runs much faster than SIFT, while performing as well in many real-time situations. ORB uses a simple measure of corner orientation-intensity centroid in which corner's intensity is offset from its center, and this vector imply an orientation. The moments of a patch can be described as follow:

$$\mathrm{M}_{ij} = \sum_{m,n} m^i n^j \mathrm{I}(m,n) \tag{5}$$

and the centroid can be defined with thes moments:

$$\mathrm{C} = (\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{11}}) \tag{6}$$

Then the orientation of the patch is :

$$\alpha = ar\tan 2(M_{01}, M_{10}) \tag{7}$$

The orientation is an important information of the feature point, and sometime the dark or light environment my be considered in different conditions which would be processed by different flow. In our proposed application: binocular vision reconstruction, due to the two cameras are in the nearly same X line, so the corner is consistent, dark and light environment consideration can be ignored. **Fig. 3** shows the ORB feature points matching, neither foreground nor background objects are both extracted well in the same line.

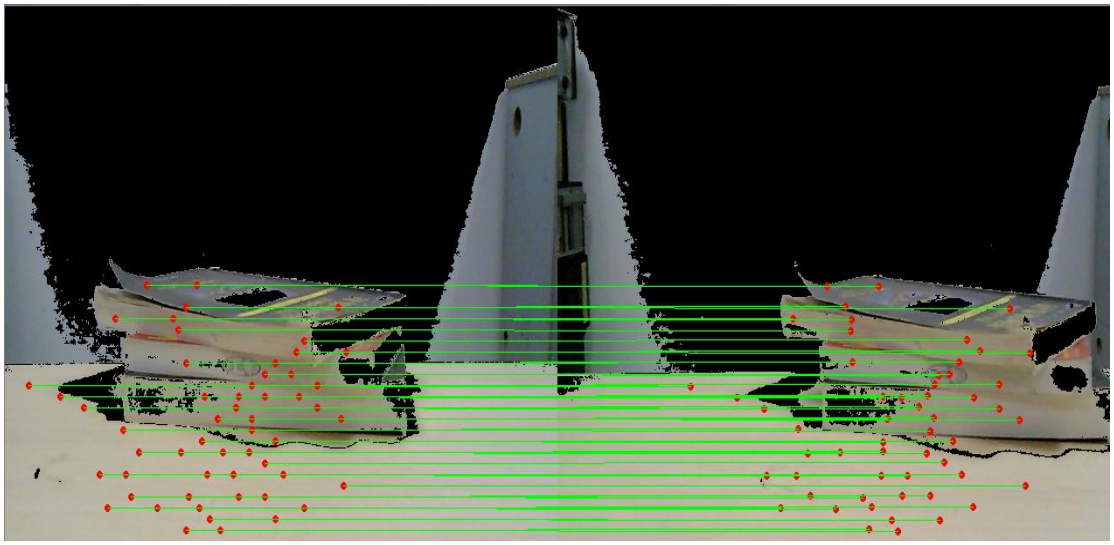

Fig. 3. Epipolar rectification with ORB feature matched points.

3.2 Stereo matching

The imaging coordinates of the same pixel should be found in different images of stereo matching stage. According to triangular geometry principle, we can compute the depth data from the extracted disparity. Local matching needs much more time computing but gets better results, while feature-based matching consumes less time with not very good depth result. In our work, objects segmentation from the background is done firstly. Due to the foreground objects are relatively more valuable to be reconstructed, we use local slipping window to compute disparity of each pixel directly. To background which would occupy a larger part of an area, we use feature-based method to reduce the time consuming.

Because the two cameras are in different views, the light environment is not the same, which will more or less negatively affect the feature extraction and points matching. So before segmentation step, we firstly process some enhancement operations to equalize the light environment. Histogram equalization is one of the most commonly used methods for contrast enhancement. It attempts to alter the spatial histogram of an image to closely match a uniform

distribution. The main objective of this method is to achieve a uniform distributed histogram by using the cumulative density function of the input image. In our proposed algorithm, we employ self-enhanced technique in the fusion view, and proposed local area histogram equalization. We use the nearly same areas to equalize the light environment, like shown in **Fig. 4**. And the basic processing steps are follow:

Step1: ORB feature extraction and matching;

Step2: Select one feature point as centroid and program a local area with fixed number of feature points:

Step3: Histogram equalization in selected area in pair to make them light equalization.

In our proposed method, we find that the smaller the area is, the better the results are, but the time is becomes costly, as shown in **Fig. 5**. For the trade-off, we set the number of the feature points as five.

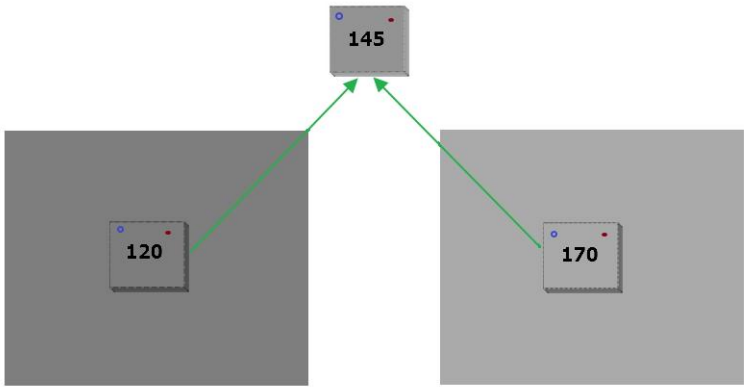

Fig. 4. Local area histogram equalization.

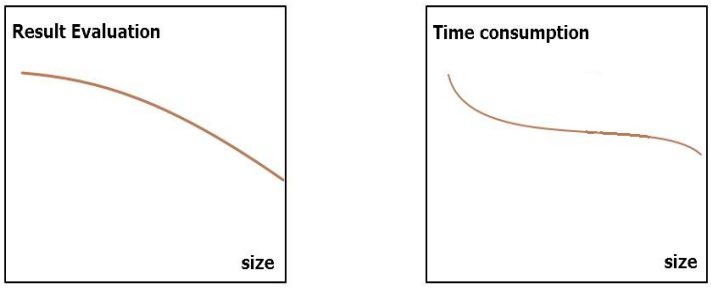

Fig. 5. Result evaluation vs size of local area (left); Time consumption vs size of local area(right).

Image segmentation has been the pre-defined processing for many research applications, especially for medical imaging[8] and 3D reconstruction. In our study, we choose K_means for the consideration of its simplicity. K_means is an unsupervised classification algorithm based on clustering, and the basic theory is described in equation (8)-(9):

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} \left\| x_i - u_k \right\|^2 \gamma_{ik} \tag{8}$$

$$u_k = \frac{\sum_i \gamma_{ik} x_i}{\sum_i \gamma_{ik}} \tag{9}$$

Assuming that there are N data points in total, they are divided into K cluster. The K_means algorithm is to minimize J, γ_{ik} is 1 or 0. When we get the smallest J, u_k should be meet the equation (8). The segmentation sample is shown in **Fig. 6**.

(a) (b) (c)

Fig. 6. K_means segmentation.(a) Original Image, (b) Background, (c) Foreground.

Because most pixels of the background image have the same depth in 3D space, so we do not need to compute each of them directly. The proposed algorithm extracts disparity information in background as follow:

Step1: Match using gradient kernel feature extracted in epipolar rectification;
Step2: Compute disparity of matched feature points;
Step3: Compute disparity of other pixels using interpolation method.

The sample of matching result is shown in **Fig. 7**.

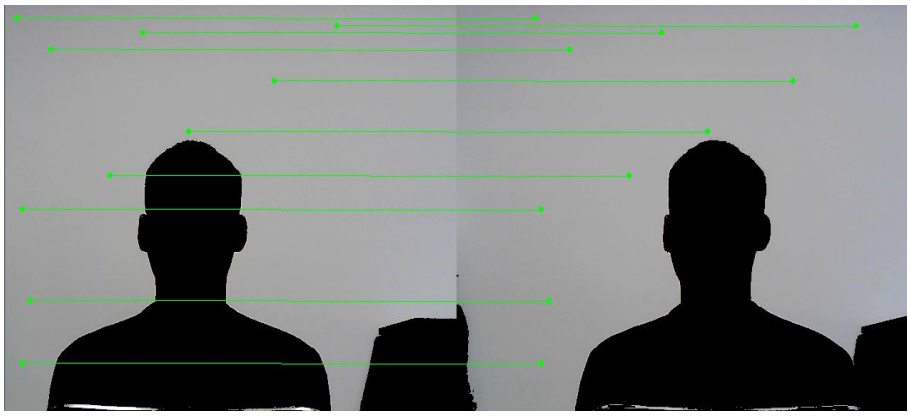

Fig. 7. Background matching.

If we use the ORB feature instead, the processing flow will be changed as follow:

Step1: Match using ORB feature extracted in epipolar rectification;
Step2: Compute disparity of matched feature points;
Step3: Compute disparity of other pixels using interpolation method.

The sample of matching result with Kernel points is shown in **Fig. 7**. Compared to **Fig. 3**. which is ORB feature points matching, we find that the number of ORB feature points is higher than of Kernel points.

For foreground objects, we design method with higher accuracy. Sliding window is applied to compare feature difference in local area between left and right view. The BRIEF feature (Binary Robust Independent Elementary Feature) easily outperforms other fast descriptors such as SURF in terms of speed and recognition rate. The simple principle is shown in **Fig. 8**. We adopt the algorithm for approximate matching of binary features based on priority search of multiple hierarchical clustering tree[9]. The size of the sliding window can be changed by different texture. The main processing steps for foreground objects described as follow:

Step1: Compute binary feature for all pixels in foreground.

Step2: Compare NCC (normalized cross correlation) distance between two pixels in both view sliding from $Xi-M$ to Xi in the right image.

Step3: Select the minimal distance in M compared results.

Step4: Compute X coordinate disparity between the two matched pixels.

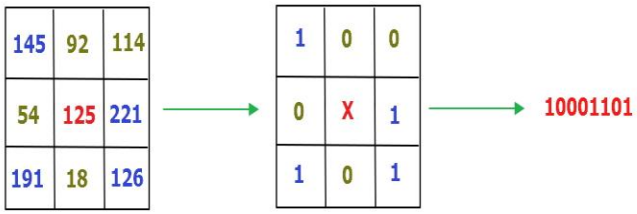

Fig. 8. Simple principle of all binary feature descriptor.

The NCC on square pixel regions as a metric for the best match:

$$NCC(v_0, v_1) = \frac{\sum_{i=1}^{N^2}(v_0(i)-\bar{v}_0)(v_1(i)-\bar{v}_1)}{\sqrt{\sum_{i=1}^{N^2}(v_0(i)-\bar{v}_0)^2\sum_{i=1}^{N^2}(v_1(i)-\bar{v}_1)^2}} \tag{10}$$

where v_0 and v_1 are local neighborhoods of size $N*N$ in both view images. $\bar{v}_0$ and $\bar{v}_1$ represent the averages over the same neighborhoods. M is the maximum of all possible pixel, which is set to reduce the number of sliding windows.

The RANSAC algorithm is a general parameter estimation approach designed to cope with a large proportion of outliers in the input data. Unlike many of the common robust estimation techniques such as M-estimators and least-median squares that have been adopted by the computer vision community from the statistics literature, RANSAC was developed from within the computer vision community. RANSAC is a resampling technique that generates candidate solutions by using the minimum number observations (data points) required to estimate the underlying model parameters. In the proposed processing algorithm, we employ RANSAC to reduce the number of outliers in the matched pairs to improve the matching accuracy. The matched image is shown in **Fig. 9**, in which we just select some pixels of them. Then we use Median Filtering to processing the map result, and fuse the background and foreground disparity map into the whole map shown in **Fig. 10**.

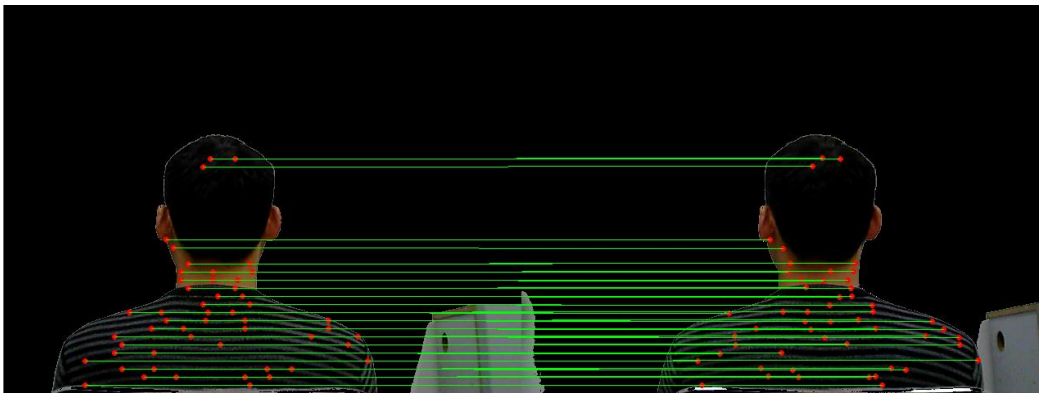

Fig. 9. Foreground objects matching of back.

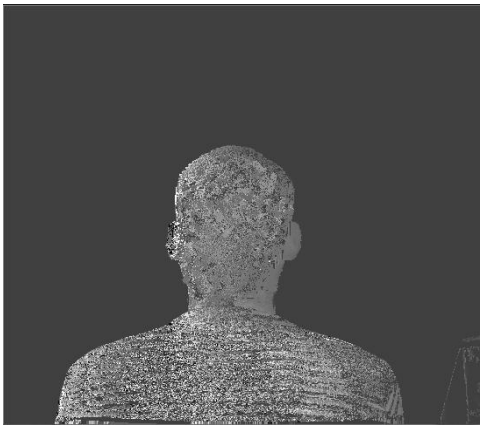

Fig.10. Disparity map of back.

4. Experimented results

In this paper, the experimental hardware parts includes two cameras and PC. Cameras parameters: Dostyle CA101, resolution: 640*480, Canon EOS 700D, resolution: 3456*2304 and PC: Inter(R) Core(TM) i5-4460 CPU@3.20GHz, 8GRAM, and all the specific algorithm carried out under VS2013, OpenCV 2.4.8, OpenGL software environment. All experiment database is collected from our lab.

Dostyle CA101 is a consumer camera which has been widely applied on PC or Laptop. Its lens distortion might be more serious than advanced EOS cameras. In our experiments, we firstly calibrated it using Zhang Zhengyou method with the 20 pieces of chessboard images, shown in **Fig. 11**. As a comparison, we use Canon EOS 700D in our experiment. The calibration results are shown in **Fig. 12**, in which the left is calibrated Dostyle CA101 imaging with intrinsic matrix (11), the right is from Canon EOS 700D with intrinsic matrix (12). From the comparison, it is obvious that consumer camera has larger distortion rate.

$$\begin{bmatrix} 4.18045801e+003 & 0 & 2.88928223e+002 \\ 0 & 2.82097119e+003 & 2.63142944e+002 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

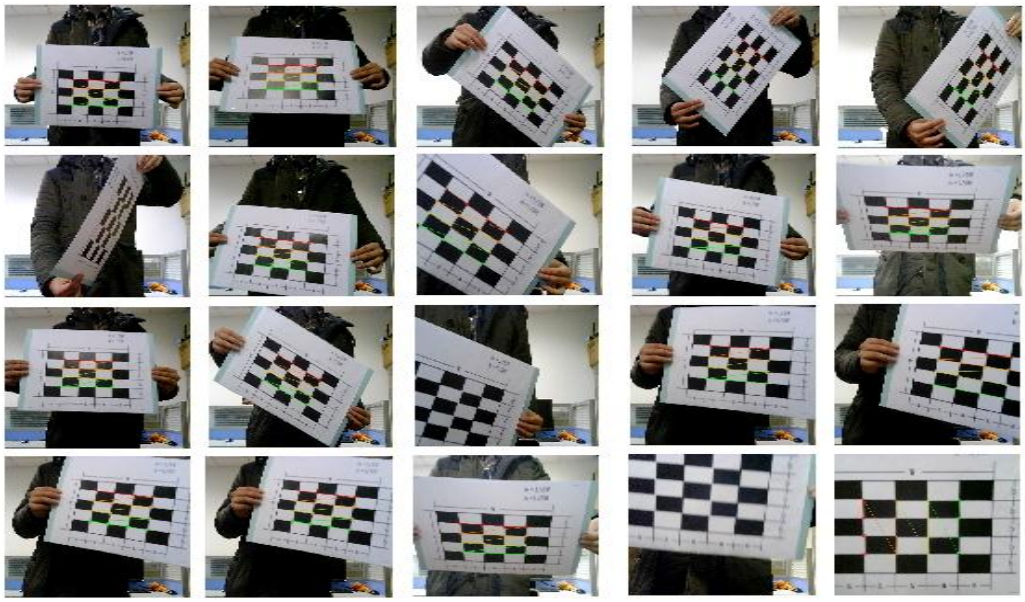

Fig. 11. Calibration using chessboards.

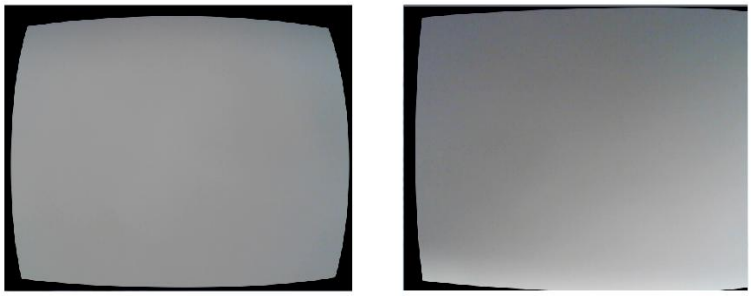

Fig. 12. Calibration results. Left: common consumer camera; Right: advanced EOS camera.

$$\begin{bmatrix} 4.35354980e+003 & 0 & 3.01830750e+002 \\ 0 & 3.54884521e+003 & 2.18060471e+002 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

In our experiments, we capture some common objects in our daily life as input in our system, such as Cup, Book. More importantly, some body parts, Knee, Hand, Brain, Back are chosen as experimental objects due to they are often reconstructed in medical imaging which has helpful influence on our life. The 3D reconstruction results represented by OpenGL are shown in **Fig. 13**. From **Fig. 13**, the left column is the matched bin-view images, in which the green matched feature points is background points while the red are foreground ones; the middle column is the disparity map of the these scene; the right column is the reconstructed results represented by OpenGL. As the error matching was inevitable in the too close or too far field of the scene, there are some discrete noise points caused by the error matching or not matching. However, the wrongly matched points occupy a small part of all the pixels, and would affect the reconstruction negatively, and our proposed processing algorithm has obtained highly accurate reconstruction results.

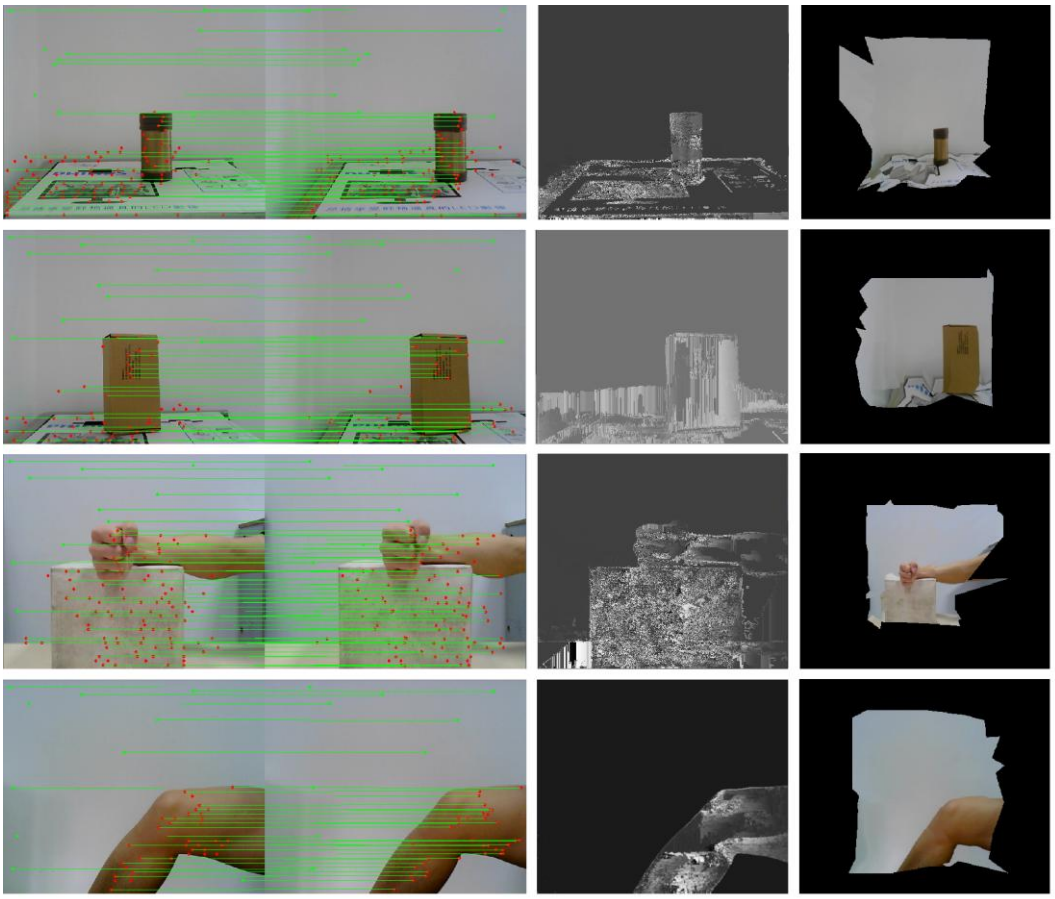

Fig. 13. Results show. the left: the match bin-view images; the middle: the disparity map; the right: 3D representation using OpenGL.

Table 1. Time consumption of different kernel feature computing (unit: ms)

Feature	Gradient Kernel	Shape Kernel	PCA Kernel	Edge Kernel	ORB
Time	107.21	98.47	104.59	89.75	11.18

Table 2. Sub-processing time (unit: ms)

Sub-processing	Time
Epipolar Rectification	19
Image Enhancement	12
Image Segmentation	9
Foreground matching	32
Background matching	26
Depth computing	18
3D Representation	27
Total	143

Time consumption is an important evaluation item for the new version. As mentioned above, ORB feature computing is much faster than Kernel feature does. **Table 1** shows the

comparison using 640*480 images. From the table, we can find that ORB feature computing takes just about 10 ms, while the Kernel features need about 100 ms. **Table 2** shows all the sub-processing time in different stages. From the time table, image segmentation cost less, and foreground matching cost the most share, the three processing steps: Foreground matching, background matching and 3D representation are the main time consuming. Actually, foreground matching and background matching are two independent steps that we can process them parallelly, then the total time will be reduced from 26ms to 117ms.

Compared to other state-of-the-art techniques [11-12] and our previous conference work, the proposed method reduces the time consumption dramatically, as shown in **Table 3**. Ref [11] is super-resolution reconstruction that costs most due to its complex processing and computing. Ref [12] costs more because it pays much time to surface reconstruction and meshing. Our previous conference work has already reduced the time computing, from 452ms to 367ms, moreover, the new version reduces the time consumption again, from 367ms to 143ms (117ms), due much to ORB kernel feature.

Table 3. In terms of PSNR (Unit: DB), performance comparison of the different methods.

Method	Ref [11]	Ref[12]	Ref [13]	Proposed method
PSNR	27.48	29.48	29.17	30.47

In this study, the peak signal to ratio (PSNR) and the bad pixel rate (BPR) are employed as two objective performance measures. PSNR method, which is most commonly used as a measure of quality of reconstruction in image compression and images/videos enhancement, is adopted to evaluate objectively the enhanced images. PSNR can be computed by equation:

$$PSNR = 10 * \log_{10} \frac{(2^b - 1)^2}{mn \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} | I(x,y) - I_e(x,y) |^2} \tag{13}$$

where b is the bytes of signal (8 for 8-bits images), m and n are width and height of image, $I(x,y)$ and $I_e(x,y)$ are the pixel value of original dark image and enhanced image. The comparison is that the greater PSNR the value, the better the enhanced result. BPR denotes the percentage of "bad" pixels in an image, as:

$$\mathrm{BPR} = \frac{1}{M} \sum_{i=1}^{M} (\left| X_i - X_i^{'} \right| > Th) \tag{14}$$

where Th denotes the disparity error threshold (set to 15 in this experiments), X_i is the value in the ground truth disparity map, $X_i^{'}$ is the corresponding value in the reconstructed disparity map, and M is the total number of pixels in the original image. Here a lower BPR indicates that the reconstructed disparity map has higher accuracy. We use the Leg object as comparison image material, the Leg Reconstruction is shown in **Fig. 14**. **Table 4** and **Table 5** show the comparison results, in which we can find that our new proposed method outstrip other methods, even the previous conference work, obtaining the nearly accurate reconstruction results. And we test the two evaluation items in other image materials, such as Book, Cup, we can also get the same level results.

Table 4. Total time comparison among different method (unit: ms).

Method	Ref [11]	Ref [12]	Proposed with Kernel	Proposed with ORB
Time	452	1005	367	143 (117)

Table 5. In terms of bad pixel rate (BPR, %), performance comparisons of the different methods.

Method	Ref [11]	Ref [12]	Ref [13]	Proposed method
BPR	9.68	10.87	11.24	10.84

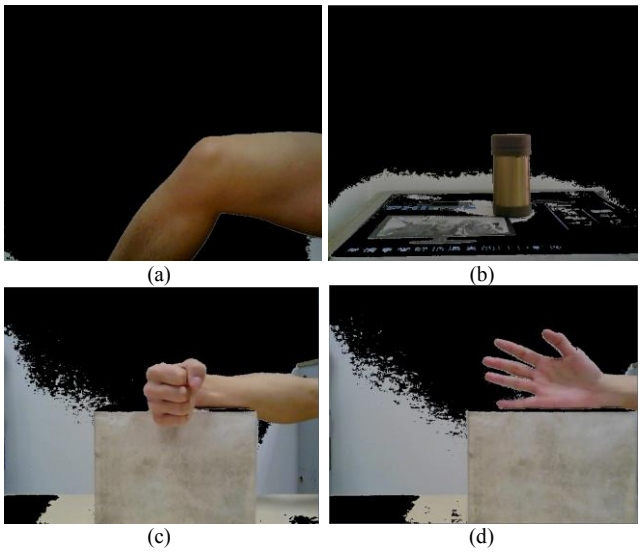

Fig. 14. Reconstruction results using our new proposed method.

5. Conclusion

In this paper, we propose a new and high-efficiency processing algorithm that we just need two consumer digital camera which avoid the large cost of mechanical settings. Actually, the proposed method has some weakness, explained as follow. Firstly, we assume that the background is simple tone, and the image materials are all belong to the kind of unchanged background. This setting would greatly help the segmentation and background matching, however, making our proposed method does not apply to complex background. In the futue, how to resolve complex background problem is our point. In a addition, we employ OpenGL to represent reconstruction results directly, and we pay no attention to meshing and surface. So the reconstruction show is not as perfect as others. We also further consider the meshing and surface problems of reconstruction processing in our work.

In our work, the main processing flow has changed as follow: (1)camera calibration, (2)ORB feature epiolar rectification, (3)local area histogram equalization, (4)segmentation, (5)background and foreground matching using ORB feature points, (6)RANSAC and median filtering, (7)disparity computing and 3D representation. Experiments demonstrate that epipolar line rectification can positively affect the feature points searching and matching, making the consumer camera binocular vision reconstruction be possible, and reconstruction results shown that the proposed method can get an accurate reconstruction map. Because using ORB instead, our method reduces the whole time consumption dramatically, from 367ms to 143ms. And even, the proposed method outstrips other state-of-the-art techniques in the terms

of PNSR and BRP evaluation items.

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments. This work is partly supported by the National Natural Science Foundation of China (Grant No.61300092), the Fundamental Research Funds for the Central Universities of China, (Grant No. ZYGX2013J068).

Reference

[1]F Nex, F Remondino,"UAV for 3D mapping applications: a review," *Applied Geomatics*, vol. 6, on.1, pp.1-15, 2014. Article (CrossRef Link)

[2]C.Y Ren, V Prisacariu, D Murray, and I Reid, "STAR3D:simultaneous tracking and reconstruction of 3D objects using RGB-D data," in *Proc. of International Conference on Computer Vision (ICCV)*, pp.561-1568, 2013. Article (CrossRef Link)

[3]Q.S Zhang, X Song, X.W Shao, H.J Zhao, and R Shibasaki, "When 3D reconstruction meets ubiquitous RGB-D images," in *Proc. of International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 700-707, 2014. Article (CrossRef Link).

[4]J. Gallego, J Salvador, J.R Casas, and M Pardas, "Joint multi-view foreground segmentation and 3D reconstruction with tolerance loop," in *Proc. of International Conference on Image Processing(ICIP)*, pp. 997-1000, 2011. Article (CrossRef Link)

[5]R Sagawa, R Furukawa , and H Kawasaki, "Dense 3D reconstruction from high frame-rate video using a static grid pattern," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 36, on. 9, pp. 1733-1747, 2014. Article (CrossRef Link)

[6] N Michael,C David. Slaughter, and G Chris, "Vision-based 3D peach tree reconstruction for automated blossom thinning," *IEEE Transactions on Industrial Informatics*, vol. 8, no.1, pp.188-196, 2012. Article (CrossRef Link)

[7] Z.Y Zhang, "A flexible new technique for camera calibration," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, on. 11, pp. 1330-1334, 2000. Article (CrossRef Link)

[8]A.B Tosun, and C Gunduz-Demir, "Graph run-length matrices for histopathological image segmentation," *IEEE Transactions on Medical Imaging*, vol. 30, in. 3, pp. 721-732, 2011. Article (CrossRef Link)

[9]M Muja, and D.G Lowe, "Fast matching of binary features," in *Proc. of The 9th Conference on Computer and Robot Vision*, pp.404-410, 2012. Article (CrossRef Link)

[10]M Sizintsev, and R.P Wildes, "Spacetime stereo and 3D flow via binocular spatiotemporal orientation analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 36, on. 11, pp. 2241-2254, 2014. Article (CrossRef Link)

[11]W.-T Hsiao, J.-J Leou, and H.-H Hsiao, "Super-resolution reconstruction for binocular 3D data," in *Proc. of International Conference On Pattern Recognition (ICPR)*, pp. 4206-4211, 2014. Article (CrossRef Link)

[12]D.Bradley, T. Boubekeur, and W.Heidrich , "Accurate multi-view reconstruction using robust binocular stereo and surface meshing," in *Proc. of International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-8, 2008. Article (CrossRef Link)

[13]Rao Y.B, Fan B.J, Ding X.S, "Object-based binocular data reconstruction using consumer camera," *Application of Image and Graphics Technology (IGTA2015)*, June 19-20, 2015. Article (CrossRef Link)

[14]E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proc. of International Conference on Computer Vision (ICCV)*, pp. 2564-2571, 2011. Article (CrossRef Link)

[15]S. Anderson and T. D. Tarfoot, "RANSAC for motion-distorted 3D visual sensors," in *Proc. of International Conference on Intelligent Robots and Systems (IROS)*, pp. 2093-2099, 2013. Article (CrossRef Link)

[16]G. R. Arce, "Nonlinear Signal Processing: A Statistical Approach," *Wiley: New Jersey*, USA, 2005. Article (CrossRef Link)

[17]Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren and Dieter Fox, "RGB-D Mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments," *The International Jouranl of Robotics Research*, vol. 31, no. 5, pp. 647-663, 2012. Article (CrossRef Link)

[18]D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004. Article (CrossRef Link)

[19] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Computer Vision Image Understanding*, vol.110, no. 3, pp.346-359, 2008. Article (CrossRef Link)

[20] L. Juan and O. Gwun, "A comparison of SIFT, PCA_SIFT and SURF," *International Journal of Image Processing(IJIP)*, vol. 3, no. 4. pp. 143-152, 2009. Article (CrossRef Link)

[21] L.F. Bo, X.F. Ren, and D. Fox, "Depth kernel descriptors for object recognition," in *Proc. of International Conference on Intelligent Robots and Systems*, pp. 821-826, San Francisco, CA, USA, 2011 Article (CrossRef Link).

[22] E. Rosten, R. Porter and T. Drummond, "Faster and Better: A machine learning approach to corner detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, on. 1, pp. 105-109, 2010. Article (CrossRef Link)

[23] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. of Europe. Conf. on Computer Vision (ECCV)*, pp.778-792, 2010. Article (CrossRef Link)

[24] Y.B. Rao and L.T. Chen, "A survey of video enhancement techniques," *Int. J. Electr. Eng. Inform*, vol. 3, no. 1, pp. 71-99,2012. Article (CrossRef Link)

[25] Y.B Rao, and L.T Chen, "Illumination-based nighttime video contrast enhancement using genetic algorithm," *Multimedia Tools and Applications*, vol.70, no. 3, pp.2235-2254, 2014. Article (CrossRef Link)

Yunbo Rao received the B.S. and M.S. degrees from the Sichuan Normal University and the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2003 and 2006, respectively, both in School of Computer Science and Engineering (SCSE), and the PhD degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2012. He has been as a visiting scholar of Electrical Engineering of the University of Washington from Oct 2009 to Oct 2011, Seattle, USA. Since 2012, he has been an associate professor at the School of Information and Software Engineering, University of Electronic Science and Technology of China. His research interests include video enhancement, computer vision, and crowd animation. More information: http://yunborao.drivehq.com/
E-mail: uestc2008@126.com School of Information and Software Engineering, University of Electronic Science and Technology of China. Chengdu, Sichuan, P.R China. 610054 Address： No.4, Section 2, North Jianshe Road, Chengdu, Sichuan, P.R China. 610054

Xianshu Ding received the Bachelor's degree in computer science and technique from Chongqing Three Gorges University, Chongqing, China in 2012. Currently, he is pursuing his Master degree in computer science from University of Electronic Science and Technology of China (UESTC), Chengdu, China, His research interests include image segmentation, image representation, sparse coding, image understanding, 3D mapping etc.

`

Bojiang Fan received the Bachelor's degree from School of Information and Software Engineering of University of Electronic Science and Technology of China(UESTC), Sichuan, China in 2015. His research interests include image processing, 3D reconstrction etc.