A Maximum Disparity Estimation Algorithm based on CLG-TV Optical Flow Method

Zhenwei Zhang, Yingyun Yang

Beijing, China
E-mail: 490259250@qq.com.

***Abstract* - Disparity range estimation is a very important part of stereo vision system in machine vision. In order to estimate the disparity range effectively, a new algorithm based on CLG-TV optical flow method is proposed. For the computation of the maximum disparity, the algorithm improves data item and smooth item of optical flow method. The position estimation of the maximum disparity is obtained by using the coarse-to-fine method at low resolution, then the image is cut in the position to carry out the accurate optical flow calculation, and finally the optical flow field is analysed to obtain the maximum disparity. Experimental results show that the proposed algorithm has a high accuracy in real images and synthetic sequences, and the maximum disparity can be effectively estimated in a short time.**

Keywords - stereo vision, maximum disparity, optical flow method

I. INTRODUCTION

Disparity is the difference between the object's positions got in the two observation points that are equally spaced from the same object. The stereo vision system uses a binocular camera to imitate human eyes to obtain external information and use the left and right image to obtain the disparity map. The disparity map can be used for three-dimensional reconstruction, image-based rendering, augmented reality and other fields, and it has great research value. Maximum disparity is very important for disparity range estimation of stereo matching algorithm. The current disparity range estimation method can be divided into: 1) variogram method; 2) features statistics; 3) dense disparity statistics etc. [1-3]These methods look for the corresponding relationship between the projection points of different viewpoint images, then the disparity map of the pixel points in the image of the actual scene is obtained, and the maximum value is obtained in the disparity map. In the practical application, the image of the same scene in the left and right view has a great difference, which makes the advantage of the present method is not obvious because of the influence of illumination change, depth discontinuity, weak texture and noise interference etc.

Optical flow is used to describe the two dimensional motion of objects in the observed images. The results of the optical flow motion estimation are the motion vectors of the target in the image. Optical flow method is one of the most effective methods to estimate the motion of sequential image and video. Ideally, the horizontal motion vector of the pixels in the image can be obtained by calculating the optical flow field of the binocular image. The horizontal motion vector is equal to the disparity of each pixel by the stereo matching algorithm, so the motion vector by optical flow method contains the disparity information, which can be used to obtain the maximum disparity of the binocular images. The basic assumption of optical flow method is a combination of the brightness constraint and the smoothness constraint, which not only ensures the accuracy of the calculation of the motion vector, but also makes the motion vector in each area can be transited smoothly. The actual image is complex and volatile, and the same scene can be imaged differently in the left and right view, and it will be difficult to obtain accurate disparity value. The advantage of optical flow method for solving the maximum disparity is that optical flow method has a high accuracy to calculate the maximum disparity, and it is more effectively compared with traditional methods to complex scenes which are difficult to handle.

Aiming at the problem of the current method, this paper presents a new algorithm of maximum disparity estimation based on CLG-TV optical flow method[4]. In this method, the data item and smooth item in optical flow algorithm are improved. Optical flow is estimated by the coarse-to-fine multi-scale method. At first, the position estimation of the maximum disparity is obtained by computing the optical flow at the lower resolution, then a part of the original image is cut in the position to carry on the high resolution accurate optical flow calculation, and finally the optical flow field is analyzed to obtain the maximum disparity. The algorithm in this paper has strong adaptability to large displacement, and overcomes the problem of large amount of computation due to extracting of feature points and matching of dense stereo, which can effectively enhance the accuracy and robustness of optical flow computation.

II. OPTICAL FLOW ESTIMATION MODEL

HS method[5] first proposed optical flow constraint equations and optical flow estimation model of variational method, and it is a global optimization method. Firstly, the optical flow constraint equation is obtained as the data item by the first order Taylor expansion with gray consistency hypothesis, and then the optical flow model is established by adding smooth item which are obtained by the smoothness of each pixel's motion. The limitation of HS method is that it can only be used for small displacement calculation, and the error of large motion is assignable.

In LK[6] method, the local optical flow field is calculated by the least square method. LK method is a local optimization method; it is assumed that a small area with single pixel has the same optical flow, so the advantage of the LK method is that it has accurate calculation for larger motion vector.

CLG method [7] combines LK and HS with the aim of enhancing the accuracy of large motion estimation to improve the robustness of the algorithm.

The energy calculation of HS and CLG methods are based on the L2 type norm of the variational energy functional, and the L2 type norm has the problem of high noise sensitivity, over expansion, poor selectivity, and no retention of moving boundary etc.

In order to overcome the shortcomings [8] of L2 model, [8] proposes a method of L1 norm which can preserve the edge information of the image . Its energy function is:

$$E = \int_{\Omega}\left[\left(\sum_{region} W(x)\left(I_x u + I_y v + I_t\right)^2\right) + \lambda\left(\|\nabla u\| + \|\nabla v\|\right)\right] \qquad 2\text{-}1$$

For the above equation, i.e. data item, the local computation of the LK form including more date items is used to improve the robustness, for smooth item, the L^1 norm is used to enhance the accuracy. The above equation still uses square calculation for reducing the amount and time of calculation.

A. Data Item Weight

The bilateral filter[8] combines Gauss's spatial filtering and distance filtering, which can not only remove the noise but also overcome the problem caused by the discontinuity of the edge. The calculation is as follows:

$$BFW(\boldsymbol{x}) = \exp\left\{-\frac{|\boldsymbol{x} - \boldsymbol{x}_0|^2}{2\sigma_1^2} - \frac{|I_x - I_{x_0}|^2}{2\sigma_2^2}\right\} \qquad 2\text{-}2$$

$\boldsymbol{x}$ refers to the coordinates of x point, I_x is the gray level of color space at x point. Using the results of the bilateral filtering as data item indicates that when the distance between pixels in a region and center pixel is larger, the weight is smaller; when the gray level difference between pixels in a region and center pixel is larger, the weight is smaller.

B. Smooth Item Weight

Smooth item of (2-1) for the smoothness of motion vectors are same in all directions, in fact, the image brightness information in the image can be used to obtain a diffusion coefficient[9], the diffusion coefficient indicates the continuity of objects in the image, When the degree of continuity is high, the diffusion coefficient is large and the continuous degree is low, the diffusion coefficient is small. Smooth item uses it to be weighted so that an textureless image region get high smooth weight, while the corner region of this non-continuous get low smooth weight, this anisotropic smooth item weight effectively suppresses the blind diffusion, to improve the calculation accuracy. At the same time, the global smoothing information should be considered, so the global and local smoothness is combined, and get the smooth item weight formula as follows:

$$D\left(\|\nabla I\|\right) = \alpha_g + \alpha_l e^{-\|\nabla I\|^{\beta}} \qquad 2\text{-}3$$

Finally, the error estimates of the two coefficients are obtained:

$$E = \int_{\Omega}\left[\lambda\left(\sum_{region} BFW(x)\rho\left(I_x u + I_y v + I_t\right)\right) + \left(\rho\left(D\cdot\nabla u\right) + \rho\left(D\cdot\nabla v\right)\right)\right] \qquad 2\text{-}4$$

III. OPTICAL FLOW SOLUTION

A. Optical Flow Solution

Using method in[4] (2-4) is divided into three parts:

$$E_{CLG-TV-1} = \int_{\Omega}\left[\left(\sum_{region} BFW(x)\left(I_x\hat{u} + I_y\hat{v} + I_t\right)^2\right) + \frac{1}{2\cdot\theta}\left(u-\hat{u}\right)^2 + \frac{1}{2\cdot\theta}\left(v-\hat{v}\right)^2\right] \qquad 3\text{-}1$$

$$E_{TV-u} = \int_{\Omega}\left[\frac{1}{2\cdot\theta}\left(u-\hat{u}\right)^2 + \|D\cdot\nabla u\|\right] \qquad 3\text{-}2$$

$$E_{TV-v} = \int_{\Omega} \left[\frac{1}{2\cdot\theta} (v - \hat{v})^2 + \|D \cdot \nabla v\| \right] \quad 3\text{-}3$$

The minimal value solution is obtained by using the alternative optimization scheme:

1) u and v are considered fixed, $\hat{u},\hat{v}$ are calculated to obtain the minimum value of (3-1).
Solution of (3-1) is a convex function optimization problem, which can be directly used as the optimal solution. In (3-1) put $\hat{u},\hat{v}$ as unknown variables, compute partial derivative in the first formula respectively to $\hat{u},\hat{v}$ and order them to zero, after simplification and consolidation can get equations:

$$\begin{pmatrix} 1+2\theta\sum BFW\cdot I_x^2 & 2\theta\sum BFW\cdot I_x I_y \\ 2\theta\sum BFW\cdot I_x I_y & 1+2\theta\sum BFW\cdot I_y^2 \end{pmatrix}\cdot$$

$$\begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} = \begin{pmatrix} u - 2\theta\sum BFW\cdot I_x I_t \\ v - 2\theta\sum BFW\cdot I_y I_t \end{pmatrix} \quad 3\text{-}4$$

Solution $\hat{u},\hat{v}$ of (3-1) can be obtained by solving linear equations above.

2) $\hat{u}$ and $\hat{v}$ are considered fixed, u,v are calculated to obtain the minimum value of (3-2) and (3-3).
Euler - Lagrange equation is rewritten as:

$$u = \lambda\cdot div\left(D\cdot\overrightarrow{p_u}\right) + \hat{u} \quad 3\text{-}5$$

The dual algorithm is used and the dual vector $\overrightarrow{p_u}$ is defined as:

$$\overrightarrow{p_u} = \frac{\nabla u}{\|\nabla u\|} \quad 3\text{-}6$$

$$\overrightarrow{p_u}^{n+1} = \frac{\overrightarrow{p_u}^{n} + \tau\cdot\nabla\left(div\left(D\cdot\overrightarrow{p_u}\right) + \frac{\hat{u}}{\lambda}\right)}{1+\tau\cdot\left\|\nabla\left(\lambda\cdot div\left(D\cdot\overrightarrow{p_u}\right)+\hat{u}\right)\right\|} \quad 3\text{-}7$$

(3-4) is used to compute $\hat{u},\hat{v}$, (3-5) and (3-7) are used to update u,v and $\overrightarrow{p_u},\overrightarrow{p_v}$. Iterations should be repeated until precision is met. The result is the optic flow field got by the above pyramid calculation.

B. Coarse-to-fine calculation

When the pixel displacement is larger than the size of the filter window, using the data item with linearization to compute optical flow directly will significantly increase the error of the optical flow. In order to calculate large motion vectors, one solution is to use Coarse-to-Fine[10] Estimation, a larger flow in low resolution scale is small enough to meet the optical flow constraint equation. By looking for the appropriate low resolution pyramid layer and pyramid layer, the optical flow is computed and mapped to the original image layer, so the computation of the large motion vector is finished. As shown in Figure 3.1, the original image is the bottom image, high level image is obtained by the particular Gaussian filtering and under sampling to its low layer image, where selected images of each layer of length and width is than its low layer image. When calculating the top (the lowest resolution) layer of optical flow initial vector $(u_{L\max}, v_{L\max}) = 0$, after calculating flow field in a layer, the flow is amplified by using bilinear interpolation and be used as the initial vector of the next layer. Such a repeat iteration is calculated to the required number of layers.

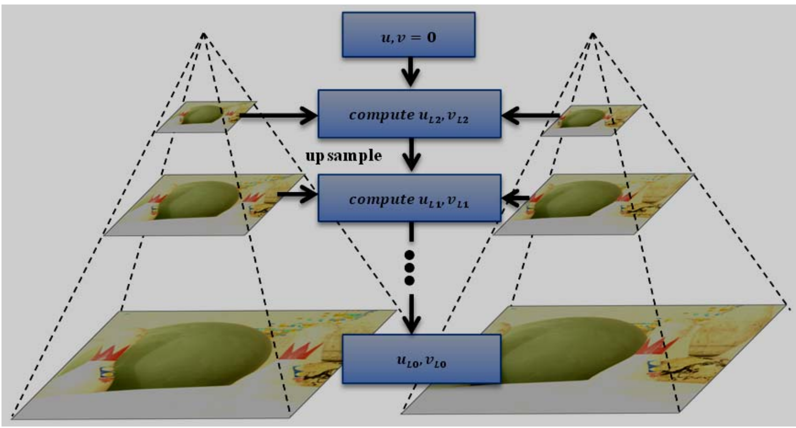

Figure 3.1 coarse-to-fine multi-scale estimation

C. Warping iteration method

In the actual calculation, in order to further strengthen the control of the linearization of the data, the motion vector can be obtained by repeated warping. Thought of this idea is that the movement is decomposed into many steps, in each step flow are used to warp another image, then the motion vector is computed after several iterations. In this case, the small motion vector point is made a slight adjustment for many times, and the large motion vector point is calculated through a lot of large precision adjustment, which makes the result of point where the motion vector is large is more accurate.

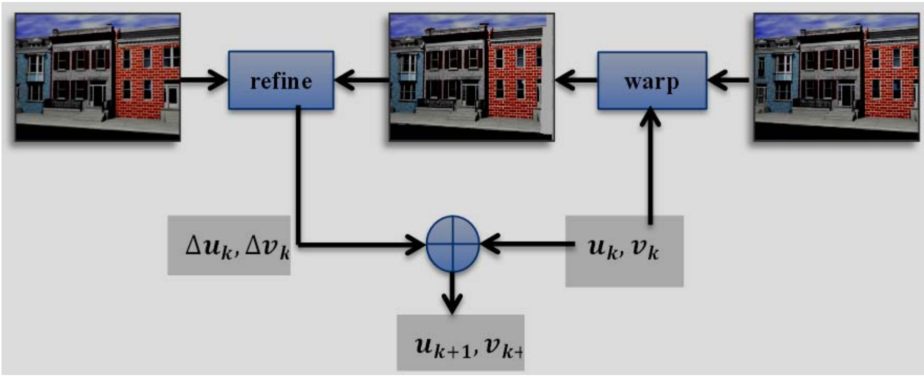

Figure 3.2 warping iteration

As shown in figure 3.2, left figure is the first frame; right figure is the second frame. The calculation results are used to update right figure by warping with bilinear interpolation, and then the results of left figure and the warped right figure are used to update last result.

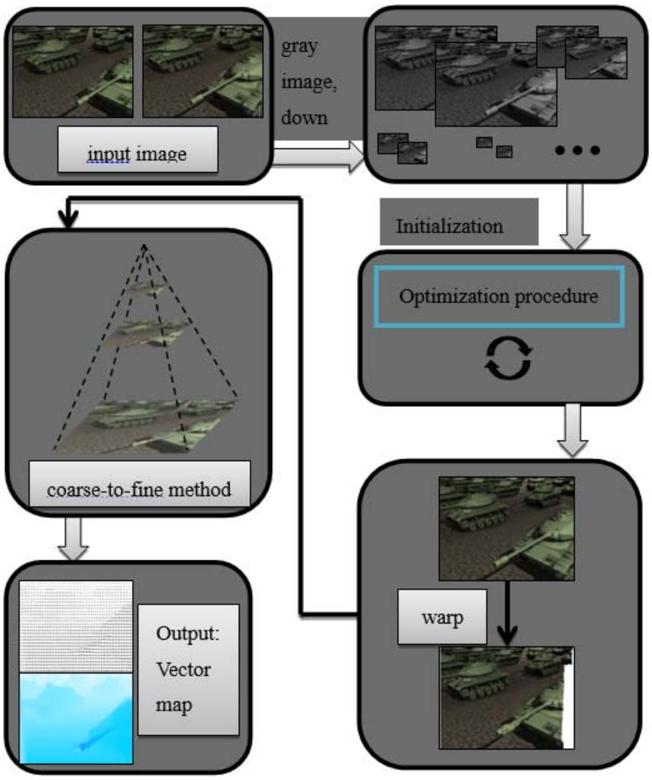

Figure 3.3 the calculation process of CLG-TV optical flow method

IV. MAXIMUM DISPARITY ESTIMATION METHOD

Ideally, the horizontal motion vector of the pixels in the image can be obtained by calculating the optical flow field of the binocular image, and the horizontal motion vector are equal to the disparity of each pixel. The calculation amount of optical flow field in the inputted original image is large, and there is a redundant part. After the analysis of the ideal disparity of the binocular image, a fast algorithm for calculating the maximum disparity based on CLG-TV optical flow is proposed, which reduces the amount of computation and maintains the accuracy of maximum disparity estimation. First, the input image is preprocessed to reduce the noise and adjust the image size, and then using the coarse-to-fine multiresolution calculation, a preliminary estimation of the optical flow field is obtained by solving the full amplitude image at a lower resolution. On the basis of the preliminary estimation, the results of the preliminary estimation are used to extract the image data and then make accurate estimates, and the maximum disparity value is obtained by calculating the horizontal motion vector. The whole calculation process for maximum disparity is shown in figure 4.1:

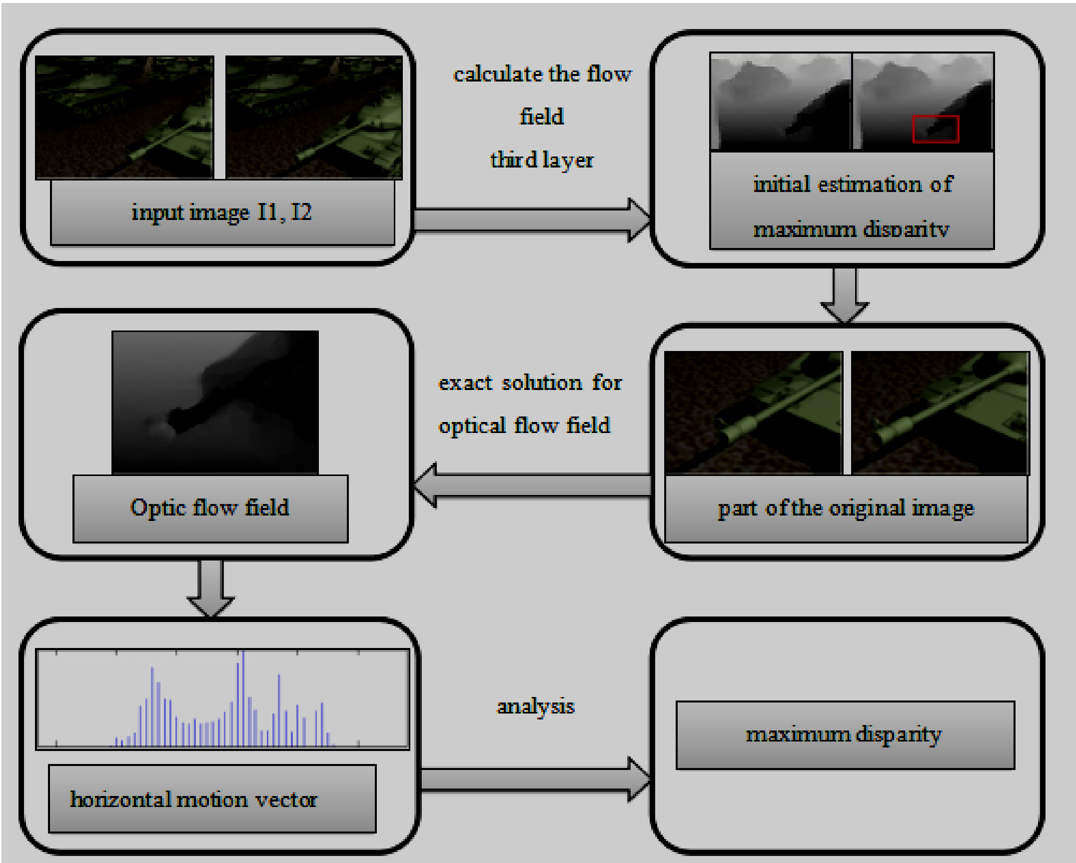

Figure 4.1 maximum disparity calculation process

A. Initial estimation of maximum disparity

The binocular image is acquired by using the left and right cameras, and the original image resolution is not uniform and there is noise. If the resolution of the original image is relatively large, the image will be reduced to the VGA level to reduce the amount of computation. Then, the color image is converted into gray image and the image is smoothed by Gauss filter, which eliminates the random noise generated during the process of camera acquisition.

Direct calculation of the original image optical flow field will get the entire pixel motion vector map, while we only concern the maximum motion vector, so the solving process is divided into two parts: a global motion vector estimation at low resolution and local motion vector estimation at high resolution. The preliminary estimate of the maximum disparity is made using a coarse-to-fine calculation for the flow field, from the highest level to the third level and the sampling coefficient is 0.25as shown in Figure 4.2, (the darker in the gray image, the larger the disparity is):

DOI 10.5013/IJSSST.a.17.26.24

ISSN: 1473-804x online, 1473-8031 print

a) left image in first frame of the Tanks sequence. b) initial estimation at lower resolution. c) the position of the maximum disparity
Figure 4.2 initial estimation at lower resolution

The low resolution estimation accuracy is common, but it can be capable of finding the position of the maximum disparity at the lower resolution.

B. Accurate estimation of maximum disparity

The position of the maximum disparity in the lower resolution is obtained, and the position of maximum disparity in original image can be obtained according to the zoom ratio. At this point, the other position of the part of the original image is obviously not the location of the maximum disparity. So they are not calculated, only the preliminary estimate of the location near the image is calculated.

The preliminary estimation of the optical flow field is processed to find the maximum value of the optical flow field. According to the zoom ratio, the maximum value u_{max} corresponding to the original image and the location $\left(i_{umax}, j_{umax}\right)$ of the maximum value is obtained. This position in I1 is used as the midpoint, and a small part is obtained by cutting off the length and width of each part of the original image $1/\beta$, then the location of the maximum value of the I2$\left(i_{umax}+u_{max}/scale, j_{umax}\right)$ in the mobile is obtained, and a small part is obtained in the same way, and scale shows reciprocal of the sampling coefficient. In grey-scale map, the darker the color is, the larger the disparity is as shown in figure 4.3. Then the two images are taken as the input image, and the optical flow field is obtained. The sampling coefficient is α, and from the highest level to the zero level, the accuracy of the calculation is higher.

a) part of the left image b) part of the right image c)disparity map of partial image
Figure 4.3 estimation of partial image

C. Selection of maximum disparity

In order to enhance the robustness of the output results, the maximum value is not directly to be find but the following way is adopted, and eventually the maximum disparity is obtained.

The horizontal motion vector of the optical flow field in 4.2 is obtained. They are ordered from small to large, and then the first order differential of this sequence is calculated when looking for maximum value. First order differential sequence began to be observed from the smallest value, when the first order difference absolute value is greater than a threshold η, all of the motion vectors below this threshold will be taken out. Their variances are computed and then make judgments based on the variance:

$$u_{\psi max} = \begin{cases} median\left(u_\psi\right), D\left(u_\psi\right) < \lambda \\ \max\left(u_\psi\right), otherwise \end{cases} \quad 4\text{-}1$$

The formula indicates that if the horizontal motion vector changes more evenly, the intermediate value is taken. If the change in the horizontal motion vector is random, the maximum value is taken.

The arithmetic sum is calculated from maximum value $u_{\psi max}$ and maximum value u_{max} in 4.2. As a result, the output of the whole system is the maximum disparity.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, five sets of standard disparity test sequences and test images are selected, which have the synthetic images, and also have the real scene images. In the experiment, the sampling coefficient α and the intercept coefficient β of 4.2 are adjusted, and the results of three different complexities are obtained. The following table lists the accuracy and computational time of the results. (Computer hardware: Intel i7-4700MQ 2.40GHz processor, memory 8GB).

TABLE I. RESULTS OF IMAGE TEST

Test image	**Real value of the maximum disparity / pixel**	$(\alpha=0.25,\beta=4)$		$(\alpha=0.5,\beta=3)$		$(\alpha=0.75,\beta=2)$	
		maximum disparity / pixel	time / s	maximum disparity / pixel	time / s	maximum disparity / pixel	time / s
aloe	70	60	0.158	65	0.318	68	0.748
Baby1	45	44	0.147	45	0.296	45	0.685
Bowling1	77	72	0.151	72	0.308	75	0.716
Cloth1	57	54	0.148	54	0.302	56	0.688
flowerpots	60	64	0.160	60	0.342	59	0.758
Lampshade1	65	64	0.150	64	0.310	62	0.697
Midd1	69	62	0.211	75	0.419	68	1.070
Monopoly	53	48	0.153	56	0.306	53	0.720
Plastic	65	64	0.147	65	0.297	64	0.683
Rocks1	57	52	0.144	52	0.291	56	0.674
Wood1	72	76	0.174	70	0.344	72	0.840
average error / pixel		4.18		2.73		1.09	
average time/s		0.158		0.321		0.753	

Table I lists the maximum disparity estimation results and the computation time of three different complexities. The images are collected from the real scene images, from the Middlebury data set. The error is defined as the absolute value of the difference between the calculated maximum and the real maximum value. From table I it can be concluded that with the decrease of the average computation time, the average error increases, the average error of the maximum disparity and the real maximum disparity of the maximum complexity is 1.09 pixels, and the average computation time is 0.753s. The table II shows the 5 synthetic sequences, each of which has 100 frames. The estimated value is u_{est} and the true value is u_{gt}, the results is believed to be accurate when $\left|u_{est} - u_{gt}\right| < 2$. From table II, it can be concluded that the accuracy of the video sequence calculation is more than 90%. The two parts of the experiment results show that the method can effectively calculate the maximum disparity.

TABLE II. 2 RESULTS OF SEQUENCE TEST

Test sequence	$(\alpha=0.25,\beta=4)$		$(\alpha=0.5,\beta=3)$		$(\alpha=0.75,\beta=2)$	
	accuracy /%	average time/s	accuracy /%	average time/s	accuracy /%	average time/s
Street	42	0.119	76	0.232	99	0.683
Tunnel	37	0.120	64	0.241	91	0.701
Temple	73	0.113	89	0.227	95	0.652
Tanks	80	0.112	77	0.223	98	0.650
Book	85	0.109	100	0.220	100	0.648

VI. CONCLUSION

The maximum disparity is an important parameter of the stereo vision system in machine vision, and it has a great significance for the estimation of disparity range. In order to estimate the maximum disparity of binocular images, a new maximum disparity algorithm based on CLG-TV optical flow method is proposed, which uses the CLG-TV optical flow method to initially and accurately estimate the binocular images and finally gets the maximum disparity. Experimental results show that the proposed algorithm has better robustness to image noise and complex scenes, and it is more accurate than the traditional methods, which can be applied to the estimation of disparity range.

ACKNOWLEDGEMENTS

This paper is sponsored by the No.2012BAH39F02 project of Communication University of China.

REFERENCES

[1] B. Cyganek, and J. Borgosz, "Maximum disparity threshold estimation for stereo imaging systems via variogram analysis," *Computational Science—ICCS 2003*, pp. 591-600: Springer, 2003.
[2] J. Kostkova, and R. Sara, “Automatic disparity search range estimation for stereo pairs of unknown scenes,” *Proc. CVWW*, pp. 1-10, 2004.
[3] O. Veksler, "Reducing Search Space for Stereo Correspondence with Graph Cuts." pp. 709-718.
[4] M. Drulea, and S. Nedevschi, "Total variation regularization of local-global optical flow." pp. 318-323.
[5] B. K. Horn, and B. G. Schunck, "Determining optical flow." pp. 319-331.
[6] B. D. Lucas, and T. Kanade, "An iterative image registration technique with an application to stereo vision." pp. 674-679.
[7] A. Bruhn, J. Weickert, and C. Schnörr, “Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods,” *International Journal of Computer Vision,* vol. 61, no. 3, pp. 211-231, 2005.
[8] C. Tomasi, and R. Manduchi, "Bilateral filtering for gray and color images." pp. 839-846.
[9] P. Perona, and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 12, no. 7, pp. 629-639, 1990.
[10] J. Weber, and J. Malik, “Robust computation of optical flow in a multi-scale differential framework,” *International Journal of Computer Vision,* vol. 14, no. 1, pp. 67-81, 1995.