Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 2, pp. 225–245, 2020

SWAP-MOVE WITH LONGITUDINAL NEIGHBORING OPTIMIZATION AND MAXIMUM A POSTERIOR ESTIMATE FOR VISUAL CORRESPONDENCE*

Qing Tian, Yinchu Wang, Xinguo Wei, Yuan Zhang, Wei Li and Li Fang

Abstract: The stereo vision is based on accurate image calibration. In practice, the low-cost stereo vision systems have limited calibration accuracy. Therefore, the matching points will have disparities not only in the latitudinal neighborhood, but also in the longitudinal neighborhood, which will seriously affect the performance of existing algorithms. In this paper, we redesign the swap-move strategy with both longitudinal and latitudinal disparities, which expand the optimization method from one-dimensional to two-dimensional in the max-flow optimization procedure. We prove that it can be locally optimal from the perspective of maximum a posteriori (MAP) estimation. Finally, the effectiveness of this algorithm is verified by real data experiments.

Key words: *stereo matching, alpha-beta swap, longitudinal disparity, MAP, max-flow*

Mathematics Subject Classification: *68U10, 94C15*

1 Introduction

Stereo matching uses two or more images of the same scene from different view angles to calculate the pixel disparities that can be further used to calculate the depth, i.e., distance of scene objects from the camera, by trigonometry. There already exist some technologies to obtain the depth, such as infrared laser camera and radar. Compared with infrared laser camera, the stereo matching based on binocular camera has the advantage of low cost and easy implementation. For radar technology, the active electromagnetic waves are often useful for strong backscattering points. Taking into account the cost and coverage factors, we use binocular stereo matching to obtain depth information.

According to the optimization theory, the stereo matching algorithm can be divided into local stereo matching [2, 3, 32] and global stereo matching [1, 5, 7]. The global stereo matching algorithm can integrate a variety of prior knowledge into the global energy function, so its performance is better than the local stereo matching algorithm. As a global optimization algorithm, graph-cuts has been studied and developed by the scholars due to its good performance [8]. Moreover, its idea of minimizing the energy function of the constructed edge weighted graph has been widely used to solve many computer vision problems, such

*This research was funded by National Key R&D Program of China, grant number 2017YFC0806005; National Natural Science Foundation of China (Grant No.61806008, No.61801007); Beijing Natural Science Foundation (4194075).

as image segmentation [14,19,36], image completion [33], medical imaging [6,21], 3D reconstruction [20,25], stereo matching [18,34] and machine learning [15]. The stereo matching methods usually adopt the graph-cuts by transforming the disparity map into a multi-label update problem [22,30]. Namely, different disparities are treated as different labels. With the maximum posteriori probability (MAP) estimation of label distribution [24,29] or minimizing the energy function [12,28], the disparities can be calculated. Under the assumption of Markov Random Field, the maximum posterior probability and the minimum energy are equivalent [26].

Graph-cuts, as an optimization method, is originally proposed by Ford and Fulkerson to solve the combinatorial optimization problems [10]. Max-flow and min-cut theorem implied by their algorithm is an important core theorem in network flow and graph theory. Nevertheless, graph-cuts had not been applied to the field of computer vision until 1989. Greig et al. first applied graph-cuts in binary image segmentation and overcomes the defects of local optimal solution and slow convergence rate [11]. After that, graph-cuts was limited to binary image segmentation until 1998. Roy et al. first used graph-cuts to solve multi-value optimization problems and make graph-cuts have a wider applicability [27]. But finding the minimum value of energy function objectively is a NP-hard problem. To solve this problem, Boykov et al. adopted the idea of multi-label to solve multi-value problems and proposed two swap-move algorithms, i.e., alpha-beta swap and alpha-expansion algorithm. These algorithms can achieve good performance [4]. In terms of the optimization performance of alpha-expansion algorithm, it has been proved that the difference between the local and the global optimization is only a fixed factor while alpha-beta swap algorithm isn't proved to have similar optimization performance. Nashihatkon et al. did a meticulous research on label update strategy and pointed out that alpha-beta swap algorithm has a better performance than alpha-expansion algorithm in the condition of a large number of labels [23]. Nevertheless, the optimization performance of alpha-beta swap algorithm is still unknown.

Previous researches mainly focus on the multi-label problem and are able to produce very good results with ideal calibration in vertical direction. However, in practice, there is always calibration errors in the vertical direction for binocular camera, which causes the longitudinal disparities. For example, the optical axis of the cameras is not exactly parallel, or the two cameras are not set at the same height. Moreover, there also could be some factors affecting the accuracy of calibration, such as the non-standard calibration plate, the insufficient number of feature points, the accuracy of feature points extraction method and the shooting angles [35]. Meanwhile, the fully automatic calibration method of camera is still existed errors [16,31]. However, the previous algorithms only consider the latitudinal disparity, and will fail when both longitudinal and latitudinal (LL) disparities are present. To solve this problem, we propose a LL-swap-move algorithm with the consideration of the longitudinal disparities. Firstly, we reconstruct the model with the help of MAP. Then, we redesign the graph, i.e., weights, source and sink points. After that, max-flow algorithm is executed to optimize. Next, the label of pixel is changed based on update strategy. Afterwards iteration strategy is applied to obtain the final result. Furthermore, we prove the reason why the LL-swap-move algorithm can achieve good optimization performance. Finally, we verify the effectiveness of this algorithm by real experimental data.

The rest of the paper is organized as follows. Section 2 describes the problems of binocular camera system in practice. Section 3 introduces the LL-swap-move algorithm as well as corresponding proving process. The actual test results of LL-swap-move algorithm are given in Section 4. Also, comparison between different algorithms and analysis on the results are made. At last, the conclusion of this paper is given in Section 5.

2 Existing System Problem

2.1 Binocular Stereo Matching Theory

In classical stereo vision problems, there are two cameras to observe a static scene and the relative coordinate system assumption is known for both cameras. Under this condition, stereo matching is to determine the position of two corresponding points, $P_L(X_L, Y_L)$ in the left image and $P_R(X_R, Y_R)$ in the right image, for the same point $P(X_C, Y_C, Z_C)$ in the scene. Then, the positional differences, between P_L and P_R, can be regard as the distance from P to cameras. In order to simplify the problem of stereo matching when finding corresponding points, two cameras are usually placed on the same line as much as possible. Thus the two optical axes of cameras can be placed as parallel as possible and there is only a latitudinal difference between P_L and P_R. Their abscissa difference is the latitudinal disparity d. From this, the depth of the scene and the coordinates of the points in the scene mapped from pixels can be calculated by combining disparities with the triangular geometry principle. Depth is the distance from the point in the scene to the camera. On condition that the calculated disparity map is accurate, disparity map can be used for 3D reconstruction, target detection, non-contact measurement, and so on. Therefore, accurate disparity map is the focus of stereo matching.

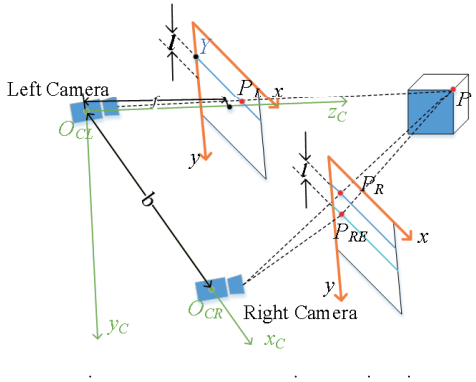

Figure 1: Sketch map of binocular imaging.

When a binocular camera is used to collect images for stereo matching, O_{CL} and O_{CR} represent two horizontal placed cameras as shown in Figure 1. O_{CL} is selected as the origin and the coordinate system is established as $O_{CL}x_cy_cz_c$, where the x_c axis is parallel to the x axis of imaging plane coordinate, the y_c axis is parallel to the y axis of the imaging plane coordinate system, and the z_c axis is the optical axis of the left camera. If the camera calibration is very accurate, the coordinates of P can be calculated according to the coordinates of P_L and P_R, the focal distance f of the camera, and the baseline distance b.

From $\Delta PP_LP_R \sim \Delta PO_{CL}O_{CR}$ and triangular geometry principle we have that

$$X_L = f\frac{X_C}{Z_C}, \tag{2.1}$$

$$X_R = f\frac{X_C - b}{Z_C}, \tag{2.2}$$

$$Y = f\frac{Y_C}{Z_C}, \tag{2.3}$$

where $Y = Y_L = Y_R$. If d is known, we can substitute $X_L = X_R + d$ into equation (2.1), (2.2) and (2.3), giving

$$X_C = \frac{b \cdot X_L}{d}, \tag{2.4}$$

$$Y_C = \frac{b \cdot Y}{d}, \tag{2.5}$$

$$Z_C = \frac{b \cdot f}{d}. \tag{2.6}$$

Then, the depth of P can also be calculated by the geometric relationship

$$D = \frac{b}{d}\sqrt{X_L^2 + Y^2 + f^2}. \tag{2.7}$$

Different pixel positions in the image and disparities reflect the different distance from points in the scene to the cameras. The position of the pixel in the image is relatively fixed. Therefore, accurate disparity is the focus of the stereo matching. The generation of accurate and reliable disparity map is of great significance for the subsequent use for 3D reconstruction, target detection and so on.

2.2 System Problems

The 3D coordinate accuracy (e.g. accuracy of calibration plate printed and measured), the number of feature points, the extraction accuracy of feature points, the quantity of calibration picture, the angles of taking pictures and other factors affect the accuracy of calibration. But the current calibration methods are still existing calibration errors [9, 13, 16, 31]. Meanwhile, the position of two cameras affects whether the calibration image pairs can achieve sufficient line alignment. If the line alignment is not sufficient, the disparities calculated by the stereo matching algorithms, may be inadequate and inaccurate, e.g. alpha-beta swap algorithm [17]. In Figure 1, P_{RE} is the point after inaccurate camera calibration. In this case, the corresponding points not only have differences in latitudinal direction, that is latitudinal disparity d, but also have certain differences in longitudinal direction, that is longitudinal disparity l. However, when the network graph of the max-flow algorithm is constructed by the alpha-beta swap, the source and sink points and weights are designed only considering disparities along the latitudinal neighborhood direction.

Take the label set F with α and β as an example, that is, $F=\{\alpha, \beta\}$. The alpha-beta swap is first given a set of initial labels f, $P=\{P_m | m \in F\}$, where $P_m=\{p \in P | f_p = m\}$ represents a pixel set whose label is assigned as m. The energy function to be minimized by the algorithm is

$$E(f) = E_{smooth}(f) + E_{data}(f), \tag{2.8}$$

where E_{smooth} measures the piecewise smoothness of f, while E_{data} measures the difference between f and the observed data. The form of E_{data} is typically

$$E_{data}(f) = \sum_{p \in P} D_p(f_p), \tag{2.9}$$

where D_p measures how well label f_p fits the observed data at pixel p.

The first image is represented by I and the second image is represented by I'. I_p represents the intensity of p in the first image. D_p's expression is

$$D_p(f_p) = min(\left| I_p - I'_{p+f_p} \right|, U), \tag{2.10}$$

where U is a constant and usually set as 20. The form of E_{smooth} is typically

$$E_{smooth}(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q), \tag{2.11}$$

where N represents a set of adjacent pixel pairs on the image, $V_{p,q}$ represents the smoothness between adjacent pixel pairs p and q, and its expression is

$$V_{p,q}(f_p, f_q) = min(\left| I'_{p+f_p} - I'_{q+f_q} \right|, U). \tag{2.12}$$

For clarity, Figure 2 shows 1D network graph for a 1D image.

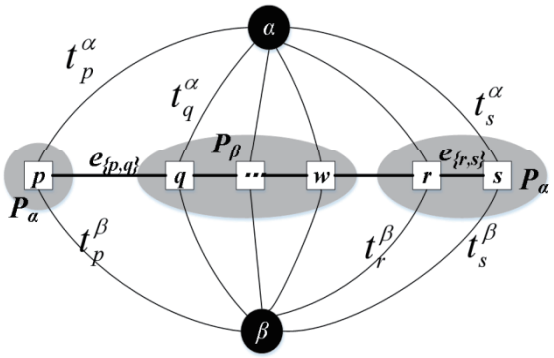

Figure 2: Alpha-beta swap algorithm composition.

In Figure 2, α is the source point and β is the sink point, $p \in P_{\alpha\beta}$, where $P_{\alpha\beta} = P_\alpha \bigcup P_\beta$ and $f_p \in \{\alpha, \beta\}$. The edges connecting with α and β are t-link edges, that is t_p^α and t_p^β. A pixel pair of $\{p, q\} \subset P_{\alpha\beta}$ and $\{p, q\} \in N$, $\{p, q\}$ are connected by n-link edges, that is $e_{\{p,q\}}$. The weights of the edges are shown in Table 1.

Table 1: Weight design of alpha-beta swap algorithm

edge	weight	for
t_p^α	$D_p(\alpha) + \sum_{\substack{q \in N_p \\ q \notin P_{\alpha\beta}}} V_{p,q}(\alpha, f_q)$	$p \in P_{\alpha\beta}$
t_p^β	$D_p(\beta) + \sum_{\substack{q \in N_p \\ q \notin P_{\alpha\beta}}} V_{p,q}(\beta, f_q)$	$p \in P_{\alpha\beta}$
$e_{\{p,q\}}$	$V_{p,q}(\alpha, \beta)$	$\substack{\{p,q\} \in N \\ p,q \in P_{\alpha\beta}}$

Only latitudinal disparities are considered in the alpha-beta swap algorithm as shown in Table 1. Once the image pair has both the latitudinal and longitudinal disparity, alpha-beta swap algorithm is ineffective. In this paper, LL-swap-move algorithm is proposed for image pairs with both latitudinal and longitudinal disparity. With the consideration of longitudinal disparity, the optimal model is constructed from MAP, and the iterative optimization is performed with the help of the max-flow algorithm. Finally, our algorithm has a better performance.

3 Swap-move with longitudinal neighborhood optimization

3.1 Optimization Function Model

For each pixel $p \in P$, f_p represents the label of p in the image $f = (f_1, f_2, ..., f_m)$ and the label set is $F = u_1, u_2, ..., u_{s\times t}$. $y = (y_1, y_2, ..., y_m)$ means the value of the actual label. $y_1, y_2, ..., y_m$ are conditionally independent of the given f, each of which has a known conditional density function $f(y_p|f_p)$, dependent on f only through f_p.

A Bayesian formulation specifies a priori distribution $p(x)$ over all allowable images. The likelihood $l(y|f)$ of any image f is combined with $p(x)$, in accordance with Bayes's theorem, to form an a posteriori distribution

$$p(f|y) \propto l(y|f)p(f). \tag{3.1}$$

The actual disparity map is calculated when $p(f|y)$ is maximized by MAP. So $p(f|y)$ is defined as the optimization function model which should be maximized.

The expression of log-likelihood ratio is

$$\ln l(y|f) = \sum_{p\in P}\sum_{f_p\in F}\left(\prod_{u_q\in F}\prod_{\substack{u_k\in F\\ u_k\neq u_q}}\frac{f_p-u_k}{u_q-u_k}\right)\ln f(y_p|f_p). \tag{3.2}$$

The prior distribution model $p(x)$ can be expressed as

$$p(f) \propto exp(\frac{1}{2}\sum_{\substack{p\in P\\ q\in P}}\beta_{pq}G_{pq}), \tag{3.3}$$

wherein

$$\beta_{pq} = \begin{cases} V_{pq}(f_p, f_q) & \{p,q\}\in N \\ 0 & others \end{cases}, \tag{3.4}$$

$$G_{pq} = \begin{cases} 1 & f_p = f_q \\ 0 & f_p \neq f_q \end{cases}. \tag{3.5}$$

Thus, apart from an additive constant, $\ln p(f|y)$ can be written as

$$L(f) = \xi(f) + \eta(f), \tag{3.6}$$

wherein

$$\xi(f) = \sum_{p\in P}\sum_{f_p\in F}\left(\prod_{u_q\in F}\prod_{\substack{u_k\in F\\ u_k\neq u_q}}\frac{f_p-u_k}{u_q-u_k}\right)\ln f(y_p|f_p), \tag{3.7}$$

$$\eta(f) = \frac{1}{2}\sum_{\substack{p\in P\\ q\in P}}\beta_{pq}G_{pq}. \tag{3.8}$$

The MAP estimation is the image f which maximizes L. α and β are chosen to construct the optimization function $L_{\alpha\beta}$, and its expression is

$$L_{\alpha\beta}(f) = \xi_{\alpha\beta}(f) + \eta_{\alpha\beta}(f), \tag{3.9}$$

wherein

$$\xi_{\alpha\beta}(f) = \sum_{p \in P_{\alpha\beta}} \lambda_p \frac{f_p - \beta}{\alpha - \beta}, \tag{3.10}$$

$$\eta_{\alpha\beta}(f) = \frac{1}{2} \sum_{\substack{p \in P_{\alpha\beta} \\ q \in P_{\alpha\beta}}} \beta_{pq} + \sum_{\substack{p \in P_{\alpha\beta} \\ q \notin P_{\alpha\beta}}} \beta_{pq}, \tag{3.11}$$

Where $\lambda_p = \ln\{f(y_p|\alpha)/f(y_p|\beta)\}$. Consider a capacity network comprising $n+2$ nodes, including a source s, a sink t and n pixels. If $\lambda_p > 0$, the capacity of a directed edge (s, p) from s to pixel p is designed as

$$c(s, p) = \lambda_p + \sum_{q \notin P_{\alpha\beta}}^{p \in P_{\alpha\beta}} \beta_{pq}; \tag{3.12}$$

Otherwise, there is a directed edge (p, t) from p to t with capacity

$$c(p, t) = -\lambda_p + \sum_{q \notin P_{\alpha\beta}}^{p \in P_{\alpha\beta}} \beta_{pq}. \tag{3.13}$$

There is an undirected edge (p, q) between two internal nodes(pixels) p and q with capacity

$$c(p, q) = \begin{cases} \beta_{pq} & p \in P_{\alpha\beta}, q \in P_{\alpha\beta} \\ 0 & otherwise \end{cases}. \tag{3.14}$$

For any image f, let $S = \{s\} \cup \{p : f_p = \alpha\}$ and $T = \{t\} \cup \{p : f_p = \beta\}$ define a two-set partition of the network nodes and put

$$C(f) = \sum_{k \in S} \sum_{l \in T} c_{kl}. \tag{3.15}$$

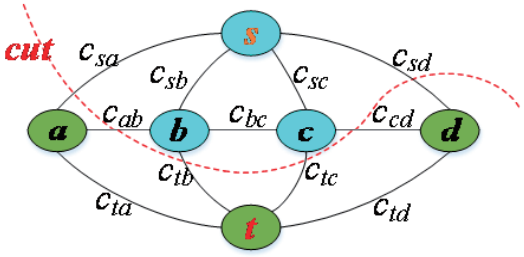

Figure 3: A cut of a network graph.

As shown in Figure 3, the set of edges with a node in S and a node in T is called a cut and $C(f)$ is called the capacity of the cut. The min-cut is the cut with minimum capacity and its capacity is equal to the max-flow.

Theorem 3.1. *If α and β are chosen to be source and sink to construct the network graph $G_{\alpha\beta}$, $L_{\alpha\beta}(f)$ will be minimized to $L_{\alpha\beta}(f^C)$ after update labels by max-flow algorithm. Meanwhile, $L(f)$ will be local maximized to $L(f^C)$.*

Proof. The cost of a cut in $G_{\alpha\beta}$ is obtained

$$C(f) = \left\{ \sum_{p\in P_{\alpha\beta}} \frac{\alpha - f_p}{\alpha - \beta} \left[max(0, \lambda_p) + \sum_{q\notin P_{\alpha\beta}} \beta_{pq} \right] \right\} + \left\{ \sum_{p\in P_{\alpha\beta}} \frac{f_p - \beta}{\alpha - \beta} \left[max(0, -\lambda_p) + \sum_{q\notin P_{\alpha\beta}} \beta_{pq} \right] \right\} + \frac{1}{2} \sum_{\substack{p\in P_{\alpha\beta} \\ q\in P_{\alpha\beta}}} \beta_{pq} \left(\frac{f_p - f_q}{\alpha - \beta} \right)^2, \tag{3.16}$$

which differs from $-L_{\alpha\beta}(f)$ by a term which does not depend on f and this term does not affect that the minimization of $C(f)$ is to maximize $L_{\alpha\beta}(f)$. The difference between $L(f)$ and $L_{\alpha\beta}(f)$ is

$$L(f) - L_{\alpha\beta} = \sum_{p\notin P_{\alpha\beta}} \sum_{f_p\in F} \left[\left(\prod_{u_p\in F} \prod_{\substack{u_k\in F \\ u_k\neq u_q}} \frac{f_p - u_k}{u_q - u_k} \right) \ln f(y_p|f_p) \right] + \frac{1}{2} \sum_{\substack{p\in P_{\alpha\beta} \\ q\in P_{\alpha\beta}}} \beta_{pq} \left(\frac{f_p - f_q}{\alpha - \beta} \right)^2. \tag{3.17}$$

No matter which cut is used to form the new label set f^C, the labels of $(P - P_{\alpha\beta})$ region will not update, so $L(f^C) - L_{\alpha\beta}(f^C)$ is a constant for all cuts C. Thus $L(f)$ will be local maximized to $L(f^C)$ when min-cut is acquired.

Theorem 3.2. *f^* is the label which maximize $L(f)$. Different label pairs are selected to construct network graph to execute max-flow algorithm each swap, the relationship between the current label f^C and f^* is $L(f^*) \geq L(f^C) \geq sL(f^*)$, where $s \in (0, 1]$.*

Proof. Let

$$c = min_{p,q\in N} \left(\frac{min_{\alpha\neq\beta\in F} V_{p,q}(\alpha, \beta)}{max_{\alpha\neq\beta\in F} V_{p,q}(\alpha, \beta)} \right), \tag{3.18}$$

be the smallest ratio of the smallest nonzero value of V to the largest nonzero value of V. After one time swap, the labels are updated to

$$f_p^{\alpha\beta} = \begin{cases} \alpha & p \in P_\alpha \\ \beta & p \in P_\beta \\ f_p^C & otherwise \end{cases}. \tag{3.19}$$

When swap moves are allowed, i.e., when the max-flow is still decreasing, we can get

$$L(f^C) \geq L(f^{\alpha\beta}). \tag{3.20}$$

Let Ω be a set consisting of pixels in P and adjacent of pixels in N. We define $L(f|\Omega)$ to be a restriction of the contribution of labeling f to the set Ω

$$L(f|\Omega) = -\sum_{p\in\Omega} D_p(f_p) + \sum_{\{p,q\}\in\Omega} \beta_{pq}(1 - G_{pq}). \tag{3.21}$$

Let $H_{\alpha\beta}$ be the set of pixels and pairs of neighboring pixels contained in $P_{\alpha\beta}$. Also, let $I_{\alpha\beta}$ be the set of pairs of neighboring pixels on the boundary of $P_{\alpha\beta}$ and $J_{\alpha\beta}$ be the set of pixels and pairs of neighboring pixels contained outside of $P_{\alpha\beta}$. That is

$$H_{\alpha\beta} = P_{\alpha\beta} \cup \{\{p, q\} \in N : p \in P_{\alpha\beta}, q \in P_{\alpha\beta}\}, \tag{3.22}$$

$$I_{\alpha\beta} = \{\{p,q\} \in N : p \in P_{\alpha\beta}, q \in (P - P_{\alpha\beta})\}, \tag{3.23}$$

$$J_{\alpha\beta} = (P - P_{\alpha\beta}) \cup \{\{p,q\} \in N : p \notin P_{\alpha\beta}, q \notin P_{\alpha\beta}\}. \tag{3.24}$$

The following facts hold

$$L(f^{\alpha\beta}|H^{\alpha\beta}) = L(f^*|H^{\alpha\beta}), \tag{3.25}$$

$$L(f^{\alpha\beta}|I^{\alpha\beta}) \geq cL(f^*|I^{\alpha\beta}), \tag{3.26}$$

$$L(f^{\alpha\beta}|J^{\alpha\beta}) = L(f^C|J^{\alpha\beta}). \tag{3.27}$$

The left and right of equations (3.25), (3.26), and (3.27), are respectively added to each other so that

$$L(f^{\alpha\beta}|H^{\alpha\beta}) + L(f^{\alpha\beta}|I^{\alpha\beta}) + L(f^{\alpha\beta}|J^{\alpha\beta}) \geq L(f^*|H^{\alpha\beta}) + cL(f^*|I^{\alpha\beta}) + L(f^C|J^{\alpha\beta}). \tag{3.28}$$

When max-flow falls after swap moves, the inequality relation is as follows

$$L(f^C|H^{\alpha\beta}) + L(f^C|I^{\alpha\beta}) + L(f^C|J^{\alpha\beta}) \geq L(f^{\alpha\beta}|H^{\alpha\beta}) + L(f^{\alpha\beta}|I^{\alpha\beta}) + L(f^{\alpha\beta}|J^{\alpha\beta}). \tag{3.29}$$

Using (3.22), (3.23), and (3.24), we get from the equation above

$$L(f^C|H^{\alpha\beta}) + L(f^C|I^{\alpha\beta}) \geq L(f^*|H^{\alpha\beta}) + cL(f^*|I^{\alpha\beta}), \tag{3.30}$$

The total $L(f)$ is obtained by using all label pairs. Therefore, we need to sum (3.30) over all $\alpha \neq \beta \in F$

$$\sum_{\alpha\in F}\sum_{\substack{\beta\neq\alpha\\ \beta\in F}} [L(f^C|H^{\alpha\beta}) + L(f^C|I^{\alpha\beta})] \geq \sum_{\alpha\in F}\sum_{\substack{\beta\neq\alpha\\ \beta\in F}} [L(f^*|H^{\alpha\beta}) + cL(f^*|I^{\alpha\beta})]. \tag{3.31}$$

Let different label pairs have w species, w >2. And let $I = \cup_{\alpha,\beta\in F} I^{\alpha\beta}$. On the left side of (3.28), for every pixel p, D_p appears w-1 times. Meanwhile, for $\{p,q\} \in N$, the term $V(f_p, f_q)$ appears once for $f_p = \alpha$, $f_q = \beta$ in $L(f^C|H^{\alpha\beta})$, w-2 times for $f_p = \alpha$, $f_q \neq \beta$ in $L(f^C|I^{\alpha f_q})$, and w-2 times for $f_p = \beta$, $f_q \neq \alpha$ in $L(f^C|I^{f_p\beta})$. Similarly, corresponding number of times also be known on the right side of (3.31). Thus, (3.31) can be rewritten to get the bound

$$L(f^C) + \frac{w-2}{w-1}L(f^C|I) \geq L(f^*) + \left(\frac{(2w-3)c}{w-1} - 1\right)L(f^*|I). \tag{3.32}$$

Then, we get from the equation above

$$\begin{aligned} L(f^*) &\leq L(f^C) + \frac{w-2}{w-1}L(f^C|I) - \left[\frac{(2w-3)c}{w-1} - 1\right]L(f^*|I) \\ &\leq L(f^C) + \frac{(2w-3)(1-c)}{w-1}L(f^C|I) \\ &\leq \left[\frac{(2w-3)(1-c)}{w-1} + 1\right]L(f^C). \end{aligned} \tag{3.33}$$

It proves that each new label distribution f^C has both upper and lower bound

$$sL(f^*) \leq L(f^C) \leq L(f^*), \tag{3.34}$$

where

$$s = \left(\frac{(2w-3)(1-c)}{w-1} + 1 \right)^{-1}. \tag{3.35}$$

So we can use max-flow algorithm for different label pairs to approximates $L(f^C)$ to the local maximum $L(f^*)$. Finally, their difference is within a fixed factor. Moreover, this factor, can be as large as 1 which depends on c.

3.2 The Process and Weight Design

Two labels u_1 and u_2 are taken as example to construct the network graph as shown in Figure 4. u_1 represents the longitudinal disparity l_1 and latitudinal disparity d_1. Similarly, u_2 represents the longitudinal disparity l_2 and latitudinal disparity d_2.

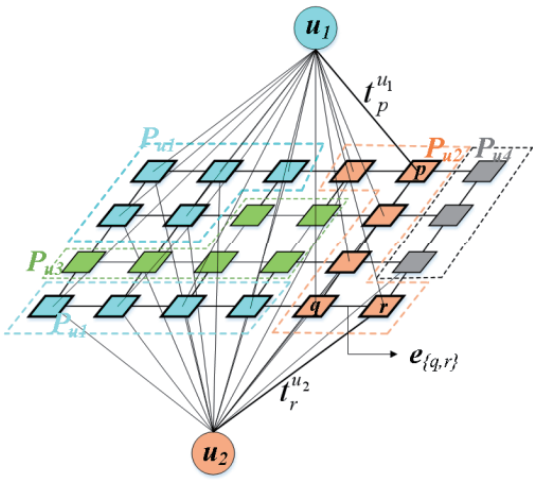

Figure 4: A network graph with label pairs u_1 and u_2.

Considering the longitudinal neighborhood, we still use D_p to measure how well the label f_p fits pixel p given the observed data and use $V_{p,q}$ to represent the smoothness between adjacent pixel pairs p and q. Thus we have

$$D_p(l_p, d_p) = min(\left| I_p - I'_{p_x+l_p, p_y+d_p} \right|, U), \tag{3.36}$$

$$V_{p,q}(l_p, l_q, d_p, d_q) = min(\left| I'_{p_x+l_p, p_y+d_p} - I'_{q_x+l_q, q_y+d_q} \right|, U), \tag{3.37}$$

where p_x is the abscissa of p and p_y is the ordinate of p. The edge weights are set as shown in Table 2.

Then the max-flow can be got by max-flow algorithm after the weight is set. Before the max-flow converges, pixels will be divided into different cut sets and their labels will also be updated. When the max-flow converges, the optimal disparity map can be generated. The updated rules are expressed as follows

$$f_p = \begin{cases} u_1 & p \in T \\ u_2 & p \in S \\ f_p & others \end{cases}, \tag{3.38}$$

Table 2: Weight design of LL-swap-move algorithm

edge	weight	for
$t_p^{u_1}$	$D_p(l_1,d_1)+\sum_{\substack{q\in N_p\\ q\notin P_{u_1u_2}}} V_{p,q}(l_1,l_q,d_1,d_q)$	$p\in P_{u_1u_2}$
$t_p^{u_2}$	$D_p(l_2,d_2)+\sum_{\substack{q\in N_p\\ q\notin P_{u_1u_2}}} V_{p,q}(l_2,l_q,d_2,d_q)$	$p\in P_{u_1u_2}$
$e_{\{p,q\}}$	$V_{p,q}(l_1,l_2,d_1,d_2)$	$\substack{\{p,q\}\in N\\ p,q\in P_{u_1u_2}}$

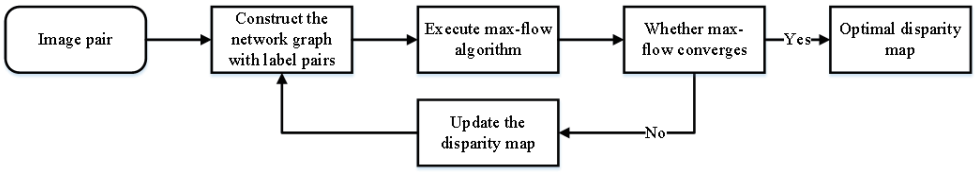

Figure 5: The flow chart of LL-swap-move algorithm.

The update of the label corresponds to the calculation result of longitudinal disparity and latitudinal disparity. For instance, if f_p is updated to u_1, the longitudinal disparity of p is updated to l_1 and the latitudinal disparity of p is updated to d_1. The complete flow chart of proposed algorithm is shown in Figure 5.

4 Results and Analysis of Experiments

Since the experimental results are not tested on the classic data set which is fully calibrated, our environmental scenario is shown in Figure 6. Some images are taken from laboratory environment, where ZED binocular cameras are chosen and the yellow shadow is the intersecting area of two cameras. All images used in the experiments are obtained after the flexible camera calibration by viewing a plane.

Figure 6: The scene of laboratory.

4.1 The Calibrator Test

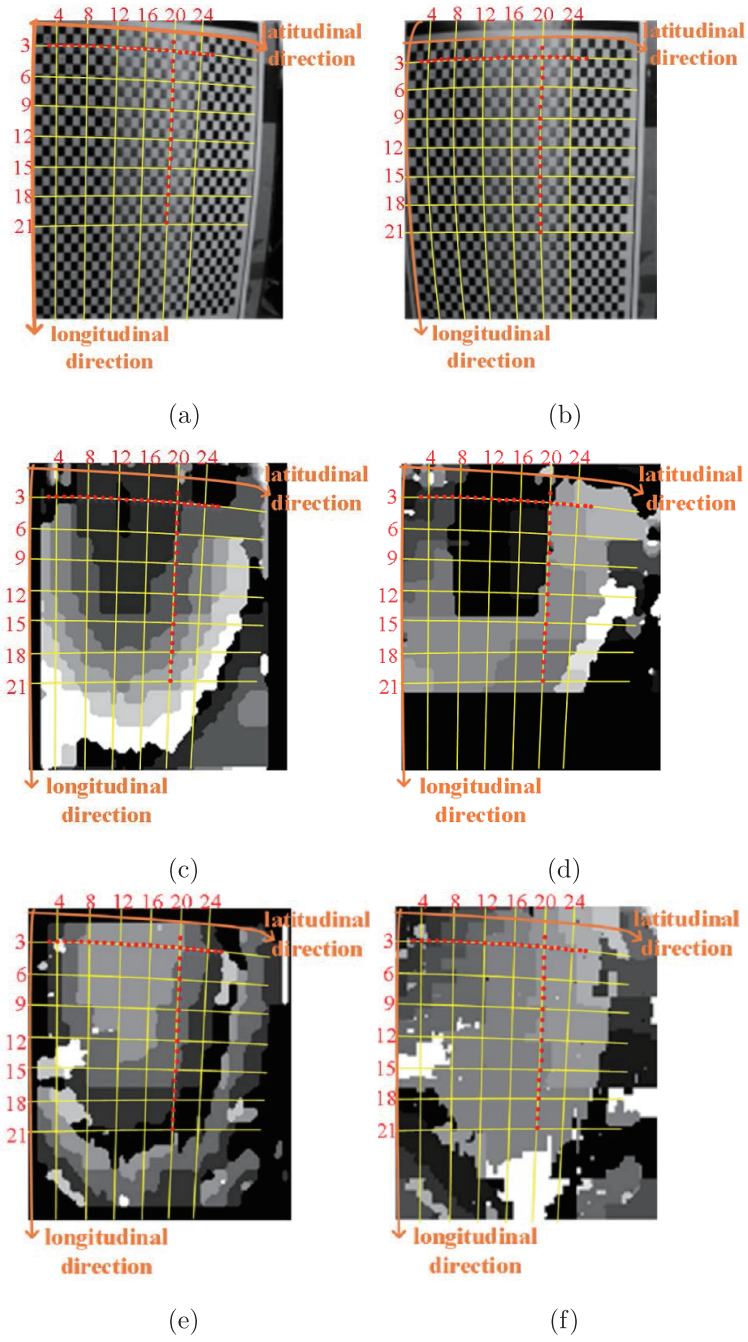

Figure 7: (a) Left image (b) Right image (c) Latitudinal disparity map of alpha-beta swap algorithm (d) Longitudinal disparity map of alpha-beta swap algorithm (e) Latitudinal disparity map of LL-swap-move algorithm (f) Longitudinal disparity map of LL-swap-move algorithm.

In order to verify the optimization performance, we use a calibrator to perform the matching test as shown in Figure 7. Some corners on the calibrator are selected for disparity measurement, which are indicated in Figure 7 by red points, i.e., 24 corners in the 3rd row and 20 corners in the 20th column. After that, we select the image in Figure 7(b) as the reference image and perform alpha-beta swap algorithm and LL-swap-move algorithm on the image pair. Figure 7(c)(d) show the results of alpha-beta swap algorithm and the LL-swap-move algorithm's results are shown in Figure 7(e)(f).

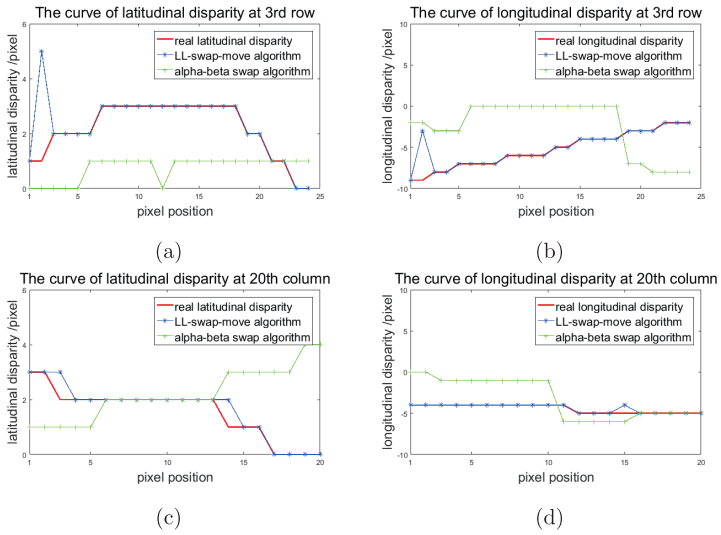

Figure 8: (a) The curve of latitudinal disparity at 3rd row (b) The curve of longitudinal disparity at 3rd row (c) The curve of latitudinal disparity at 20th column (d) The curve of longitudinal disparity at 20th column.

As shown in Figure 8, there is only one obvious error of the calculated longitudinal and latitudinal disparity at the second corner in the 3rd row. Most of the other corners' calculated results are accurate. Moreover, we also make the mean square error analysis among the results of the proposed algorithm, alpha-beta swap algorithm and actual disparity. For the latitudinal and longitudinal disparity, the mean square errors of LL-swap-move algorithm are 0.41px and 0.84px while those of alpha-beta swap algorithm are 3.30px and 19.23px.

We also analyze the corner M as shown in Figure 9(a) which has obvious errors. After processed by LL-swap-move algorithm, the corresponding corner R is in the 4th row and the 5th column in the right image as shown in Figure 9(b). The corresponding corner N which is got by measuring is also shown in Figure 9(b). According to the measurements, this error is an apparent error of position. In section 3, we use the four-neighbor model and the grayscale information to construct the network graph. So, the gray values of the four-neighbor pixels of corner M, N, R are extracted as shown in Figure 9(c∼e). In accordance with the construction method of the network graph in LL-swap-move algorithm, the local network graph of the result of measurements and this algorithm can be obtained as shown in Figure 9(f)(g). In order to make it clearer, we only show the grayscale difference without normalizing. The accuracy of this corner is concerned in the measured result while other pixels in the neighborhood are not considered completely. But because the positional information of the relationship between the pixels is considered in the measurements, the accuracy of the result is ensured. On the other hand, in order to obtain the smooth effect, LL-swap-move algorithm

is performed with considering of the grayscale difference in the four-neighbor. However, the positional relationship is not taken into account. Thus, after this algorithm, the grayscale difference in the neighborhood is smaller than the actual measurements, but the positional error is obvious.

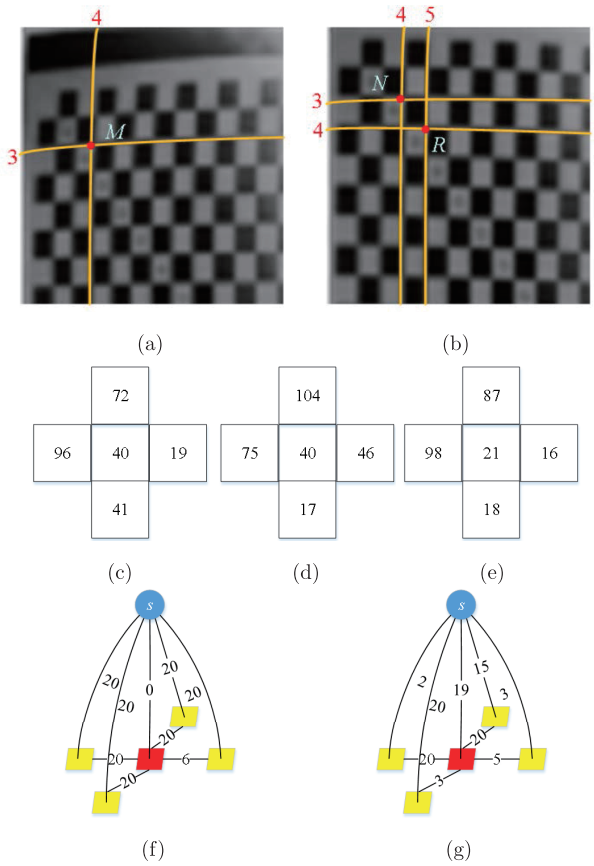

Figure 9: (a) The error corner in the left image (b) The corresponding corners of the measured and the calculation of LL-swap-move algorithm (c) The gray value of M's four-neighbor (d) The gray value of N's four-neighbor (e) The gray value of R's four-neighbor (f) The local network graph of measured result (g) The local network graph of LL-swap-move algorithm's result.

4.2 Single-target Image Test

In order to further verify the effectiveness of LL-swap-move algorithm on more complex targets, we make a test on the pedestrian images. The four consecutive frames extracted from the pedestrian video are selected as the reference images as shown in Figure 10(a), and the corresponding images are shown in Figure 10(b). Shoulder and leg areas are selected to show the local effect which are indicated in Figure 10(a) by red and blue circles. The results of alpha-beta swap algorithm are shown in Figure 10(c). Figure 10(d) shows LL-swap-move algorithm's results. The local results are shown in Figure 10(e∼h).

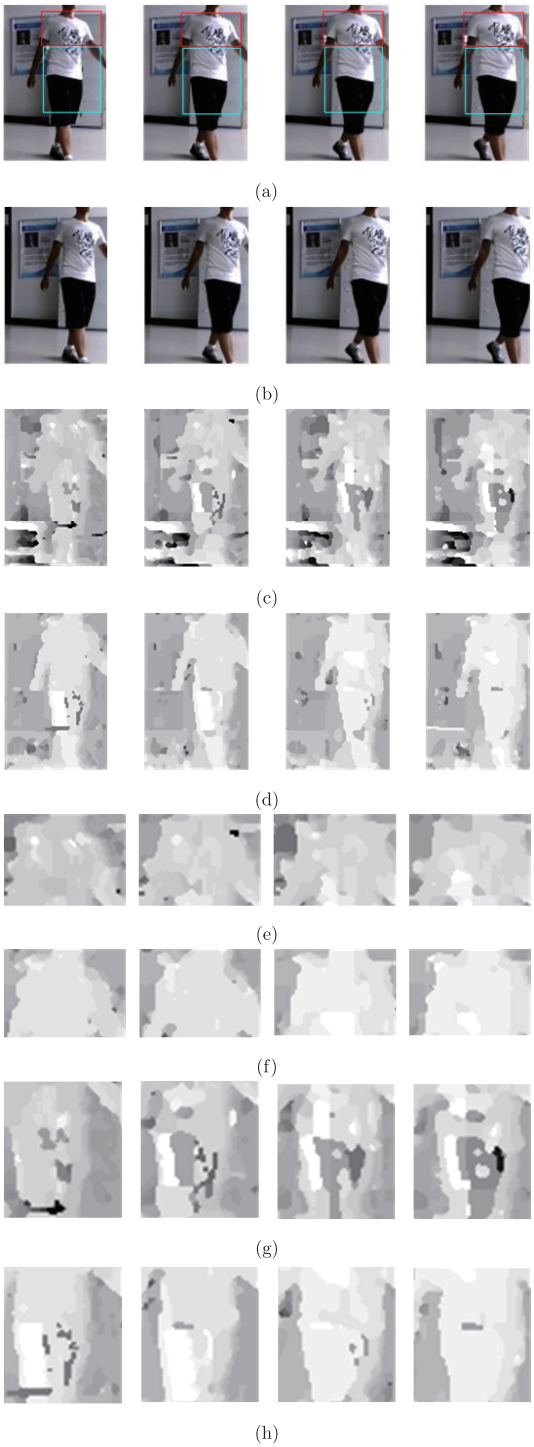

Figure 10: (a) The reference images (frame 1 to 4) (b) The correspondence images (c) Alpha-beta swap algorithm (d) LL-swap-move algorithm (e) The shoulder results of alpha-beta swap algorithm (f) The shoulder results of LL-swap-move algorithm (g) The leg results of alpha-beta swap algorithm (h) The leg results of LL-swap-move algorithm.

From the test results of the collected images, it can be found that alpha-beta swap algorithm performs poorer than LL-swap-move algorithm because of the systematic calibration error and the existence of the longitudinal disparity. In addition, the results of alpha-beta swap algorithm as shown in Figure 10(e)(g) show that the contour information cannot be well performed at the strong texture area and the performance of smoothness is poor at the weak texture area as well. With the consideration of longitudinal disparity, LL-swap-move algorithm have a better optimization performance as shown in Figure 10(f)(h). Because of the consideration of the smoothness of the labels, the results are relatively smooth.

4.3 Multi-target Image Test

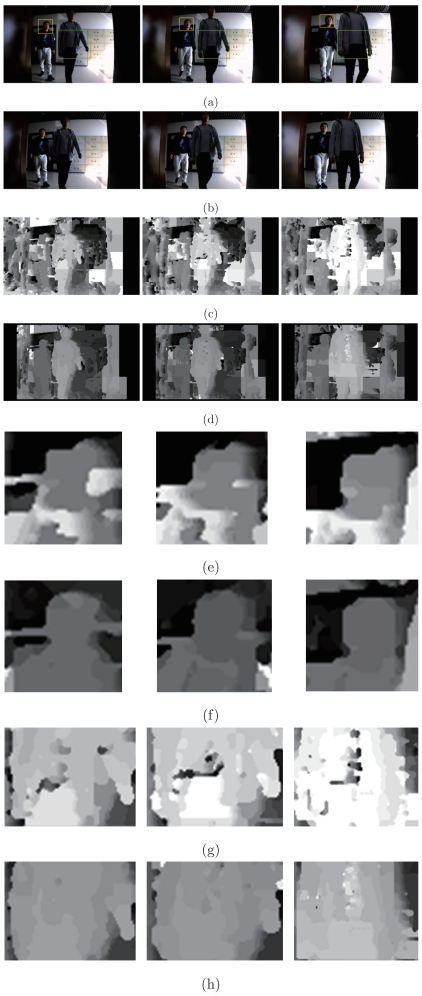

Figure 11: (a) The reference images (b) The correspondence images (c) Alpha-beta swap algorithm (d) LL-swap-move algorithm (e) The head results of alpha-beta swap algorithm (f) The head results of LL-swap-move algorithm (g) The body results of alpha-beta swap algorithm (h) The body results of LL-swap-move algorithm.

We also make an experiment on images with multiple pedestrians. Figure 11(a)(b) show the reference images and the correspondence images. Moreover, head and body areas are selected to show the local effect which are indicated in Figure 11(a) by yellow and green circles. Because of the systematic calibration error and the existence of the longitudinal disparity, the results generated by alpha-beta swap algorithm are not satisfactory as shown in Figure 11(c). LL-swap-move algorithm has a better performance as shown in Figure 11(d). As shown in Figure 11(e), the effects of the results generated by alpha-beta swap algorithm are rough in the strong texture region, and are not smooth in the weak texture region as shown in Figure 11(g). But LL-swap-move algorithm has the better effect as shown in Figure 11(f)(h) even there are systematic calibration errors.

In order to describe the convergence performance of LL-swap-move algorithm more clearly, we choose the first image pair in Figure 11(a)(b) to visually illustrate the convergence process as shown in Figure 12(a∼f). As the iterations progress, the disparities are updated step by step. Finally, the optimal disparity map can be obtained. Moreover, we show the different convergence performance between alpha-beta swap algorithm and LL-swap-move algorithm in Figure 13. Because the consideration of longitudinal disparity is adding into the optimization function, LL-swap-move algorithm has a lower initial value. In terms of convergence speed, alpha-beta swap algorithm is a little faster than LL-swap-move algorithm. Alpha-beta swap algorithm converges at the 16th iteration while LL-swap-move algorithm converges at the 45th iteration. But in terms of convergence results, LL-swap-move algorithm has a lower convergence value. Alpha-beta swap converges at 1044 while LL-swap-move algorithm converges at 469. Thus, LL-swap-move algorithm has better convergence performance.

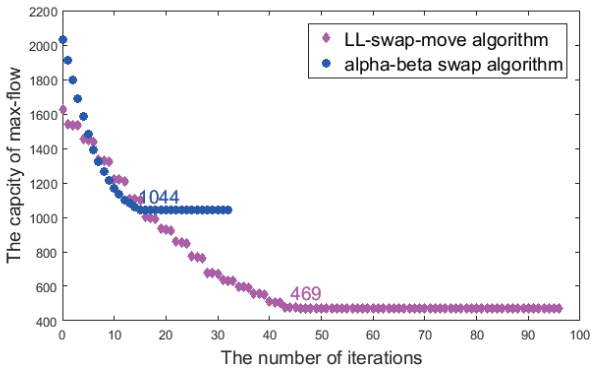

Figure 12: The comparison of different convergence processes.

5 Conclusion

In the field of stereo matching, the inaccurate calibration and non-ideal cameras position will result in image pairs disparities not only in latitudinal neighborhood but also in longitudinal neighborhood. With the consideration of longitudinal neighborhood, we propose the LL-swap-move algorithm and construct the optimization model. Then, we execute the max-flow procedure to generate the dense disparity map. The optimization performance of LL-swap-move algorithm are verified by real data experiments, which shows that the proposed algorithm can ensure the retention of disparity information and the accuracy of disparity

calculation when dealing with stereo matching image data with longitudinal deviation. It can be applied to the image matching with two-dimensional disparity, which effectively solves the problem that the alpha-beta swap algorithm cannot perform well with the inaccuracy of the camera calibration. Also, it can provide accurate disparity reference value for further image detection and 3D reconstruction. Moreover, this two-dimensional matching points optimization strategy can be applied to other stereo matching algorithms when camera calibration is not ideal.

References

[1] A. Ahmadzadeh, H. Madani, K. Jafari and et al., Fast and adaptive BP-based multi-core implementation for stereo matching, *Formal Methods and Models for Code-sign* 42 (2013) 135–138.

[2] A.F. Bobick and S.S. Intille, Large occlusion stereo, *International Journal of Computer Vision* 33 (1999) 181–200.

[3] Y. Boykov, O. Veksler and R. Zabih, A variable window approach to early vision, *IEEE Trans. On PAMI* 20 (1998) 1283–1294.

[4] Y. Boykov, O. Veksler and R. Zabih, Fast approximate energy minimization via graph cuts, *IEEE Trans. Anal. Mach. Intell* 23 (2001) 1222–1239.

[5] J. Cai, Integration of optical flow and dynamic programming for stereo matching, *IET Image Processing* 6 (2012) 205–212.

[6] J. Chen, Z.W. Shen and W.W. Xi, Super-resolution reconstruction for 4-dimensional computed tomography of graph cuts, *Journal of Southern Medical University*, 36 (2016) 1260–1264.

[7] I.J. Cox, S.L. Hingorani, S.B. Rao and et al., A maximum likelihood stereo algorithm, *Computer Vision and Image Understanding* 63 (1996) 542–567.

[8] S. Daniel and S. Richard, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *International Journal of Computer Vision* 47 (2002) 7–42.

[9] X.F. Feng and D.F. Pan, A camera calibration method based on plane mirror and vanishing point constraint, *OPTIK* 154 (2018) 558–565.

[10] L. Ford and D. Fulkerson, *Flows in network*, Princeton University Press, New Jersey, 1962.

[11] D.M. Greig, B.T. Porteous and A.H. Seheult, Exact maximum a posterior estimate for binary images, *Journal of the Royal Statistical Society. Series B (Methodological)* 51 (1989) 271–279.

[12] S. Gould, F. Amat and D. Koller, Alphabet soup: a framework for approximate energy minimization, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 903–910.

[13] B. Guan, Y. Shang and Q.F. Yu, Planar self-calibration for stereo cameras with radial distortion, *Appl Opt* 56 (2017) 9257–9267.

[14] N. Habiba and D. Ali, Split Bregman's algorithm for three-dimensional mesh segmentation, *Journal of Electronic Imaging* 25 (2016) 033011.

[15] Q.Y. Huang, X.H. Zhang and S. Huang, Sparse graph-based inductive learning with its application to image classification, *Journal of Electronic Imaging* 25 (2016) 050502.

[16] F. Jin and X.Y. Wang, An autonomous camera calibration system based on the theory of minimum convex hull, in: *International Conference on Instrumentation Measurement Computer Communication and Control(IMCCC)*, 2015, pp. 857–860.

[17] E.K. Jung, C. Kim and S.Y. Park, Vertical disparity correction of stereoscopic video using fast feature window matching, in: *2012 IEEE International Conference on Consumer Electronics(ICCE)*, 2012, pp. 463–464.

[18] A.L. Kaczmarek, Stereo vision with equal baseline multiple camera set (EBMCS) for obtaining depth maps of plants, *Computers And Electronics in Agriculture* 135 (2017) 23–37.

[19] J. Lee and S.I. Yoo, Defect detection on images using multiple reference images: solving a binary labeling problem using graph-cuts algorithm, *Journal of Electronic Imaging* 21 (2012).

[20] M. Li, W.L. Zhang and D.Y. Fan, Automatic texture optimization for 3D urban reconstruction, *Acta Geodetica et Cartographica Sinica* 46 (2017) 338–345.

[21] Y.H. Li, F.C. Jia and J. Qin, Brain tumor segmentation from multimodal magnetic resonance images via sparse representation, *Artificial Intelligence in Medicine*, 73 (2016) 1–13.

[22] Y. Lim, K. Jung and P. Kohli, Efficient energy minimization for enforcing label statistics, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (2014) 1893–1899.

[23] B. Nashihatkon and R. Hartley, Move-Based algorithms for the optimization of an isotropic gradient MRF model, in: *International Conference on Digital Image Computing Techniques & Application*, 2012, pp. 1–8.

[24] C. Nieuwenhuis and D. Cremers, Spatially varying color distributions for interactive multilabel segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013) 1234–1247.

[25] R.J. Pan and G. Taubin, Automatic segmentation of point clouds from multi-view reconstruction using graph-cut, *Visual Computer* 32 (2016) 601–609.

[26] S. Paul, S. Alexander and K.J. Hendrik, Partial optimality by pruning for MAP-Inference with general graphical models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (2014) 1170–1177.

[27] S. Roy and I. Cox, A maximum-flow formulation of the n-camera stereo correspondence problem, in: *International Conference on Computer Vision*, 1998, pp. 492–499.

[28] N. Papadakis and A. Bugeau, Tracking with occlusions via graph cuts, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (2011) 144–157.

[29] M.J. Shafiee, A. Wong and P. Fieguth, Deep randomly-connected conditional random fields for image segmentation, *IEEE Access* 5 (2017) 366–378.

[30] R. Szeliski, R. Zabih, D. Scharstein and et al., A comparative study of energy minimization methods for Markov random fields with smoothness-based priors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2008) 1068–1080.

[31] F. Vasconcelos, J.P. Barreto and E. Boyer, Automatic camera calibration using multiple sets of pairwise correspondences, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (2018) 791–803.

[32] O. Veksler, Stereo correspondence with compact windows via minimum ratio cycle, *IEEE Trans. on PAMI* 24 (2002) 1654–1660.

[33] M. Xiao, G.Y. Li,L. Xie and et al., Contour-guided image completion using a sample image, *Journal of Electronic Imaging* 24 (2015) 023029.

[34] Q.S. Wang, Z. Yu and C. Rasmussen, Stereo vision-based depth of field rendering on a mobile device, *Journal of Electronic Imaging* 23 (2014) 1709–1717.

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

[36] J. Zhao, C.M. Zhang and X.Y. Zhang, Image segmentation algorithm based on optimal region boundary map, *Application Research of Computers* 33 (2016) 307–310.

Manuscript received 24 April 2018
revised 20 September 2018
accepted for publication 20 November 2018

Qing Tian
School of Information Science and Technology
North China University of Technology, Beijing, 100144, P.R. China
E-mail address: tianqing@ncut.edu.cn

Yinchu Wang
15 BeiSanhuan East Road, ChaoYang District, Beijing, 100029, P.R. China
E-mail address: wangyinchuncut@outlook.com

Xinguo Wei School of Instrumentation Science and Opto-electronics Engineering
Beihang University No. 37 Xueyuan Road, Haidian District, Beijing, 100191, P.R. China
E-mail address: wxg@buaa.edu.cn

Yuan Zhang
Address: School of Information Science and Technology
North China University of Technology, Beijing, 100144, P.R. China
E-mail address: zhangyuan@ncut.edu.cn

Wei Li
School of Information Science and Technology
North China University of Technology, Beijing, 100144, P.R. China
E-mail address: lwsar@ncut.edu.cn

Li Fang
Quanzhou Institute of Equipment Manufacturing, Haixi Institutes
Chinese Academy of Science, Quanzhou, Fujian, 362200, P.R. China
E-mail address: fangli@fjirsm.ac.cn