

## Research Article

# A Mean-Shift-Based Feature Descriptor for Wide Baseline Stereo Matching

Yiwen Dou, Kuangrong Hao, Yongsheng Ding, and Min Mao

*Engineering Research Center of Digitized Textile & Apparel Technology, Ministry of Education, College of Information Science and Technology, Donghua University, Shanghai 201620, China*

Correspondence should be addressed to Kuangrong Hao; [krhao@dhu.edu.cn](mailto:krhao@dhu.edu.cn)

Received 30 July 2015; Accepted 25 October 2015

Academic Editor: Erik Cuevas

Copyright © 2015 Yiwen Dou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a novel Mean-Shift-based building approach in wide baseline. Initially, scale-invariance feature transform (SIFT) approach is used to extract relatively stable feature points. As to each matching SIFT feature point, it needs a reasonable neighborhood range so as to choose feature points set. Subsequently, in view of selecting repeatable and high robust feature points, Mean-Shift controls corresponding feature scale. At last, our approach is employed to depth image acquirement in wide baseline and Graph Cut algorithm optimizes disparity information. Compared with the existing methods such as SIFT, speeded up robust feature (SURF), and normalized cross-correlation (NCC), the presented approach has the advantages of higher robustness and accuracy rate. Experimental results on low resolution image and weak feature description in wide baseline confirm the validity of our approach.

## 1. Introduction

Stereo matching [1, 2] is an analyzing and representing approach of which different objects in real scene should be in different relative depth relationships. The key technology is disparity map acquired from two images taken at different angles of the same scene. There are two realization approaches so far, local matching [3, 4] and global optimization matching [5] algorithm. The former initially sets a fixed size window in each pixel and then searches the corresponding pixel in another image by pixels' gray values. The main advantage of this approach is simple while it has a few of drawbacks. It will give rise to big errors when the gray value of image window center is much closer to its surrounding pixels. Oversized image window is destined to increase matching search time and the error in low resolution. In contrast, the depth image will be distortion because there is no matching feature point in small size window. For the sake of avoiding the problem of window size, many researchers focus on global optimized approach which chooses the least error assignment scheme in whole image. Nevertheless, the above algorithms are only limited to narrow baseline. The essential reason is that the corresponding relation between reference image

point and the searched image center point is not at the same horizontal line. So it increases point searching range and easily leads to several-for-one relationship which will destroy the uniqueness of point matching. On another facet, considering the image complexity in real scene, only using pixel-based algorithm will bring large error (which is attributed to 1D searching in narrow baseline and 2D searching in wide baseline [6–8]) when searching area expands to a certain degree. In summary, the two above approaches can generate depth image in narrow baseline and do nothing about it in wide baseline.

Generally speaking there are two ways of depth image acquisitions including active type and passive type. In recent years, active depth image sensors have been promoted to a certain degree. In [9], Shieh and Hsieh used structured light to project 3D scene and then a complex algorithm of triangle measuring with geometry relation was utilized to compute depth image. But there is no denying that its applications in engineering are hindered by expensive cost and stringent calibration condition. On the contrary, passive depth image acquisition is the basis of the construction of 3D stereo vision and some relatively perfect theories promote its developments. Typically, depth image is computed from



FIGURE 1: Texture information around the same name image point under different angles.

photograph taken by binocular camera in narrow baseline. It does not need to consider point matching problem because the captured photograph is on the same horizontal line. Thus, the depth of points in weak texture image can be computed by using different disparity scales [10, 11] and continuity hypothesis. But in wide baseline, it cannot work out disparity map just utilizing horizontal disparity because the matching point is not on the same horizontal line. Hence, the matching approach between different images feature points always uses high repeatability and robustness feature detection operator such as SIFT [12, 13], SURF [14], and Harris [15]. However, those operators usually only collect those feature points which owned high gray change rate. The acquirement of depth map needs all image points' depth estimation. As a result, above algorithms do not compute depth map on wide baseline because traditional feature extraction and matching algorithms cannot deal with those weak texture points in image.

In fact, many researchers describe feature by local information surrounding feature point instead of focusing on feature point itself. Binary affine invariant descriptor (BAND) which is defined by local affine invariant area is the best case [16] and it results in good matching results. In this paper, a feature descriptor based on Mean-Shift [17, 18] is represented to overcome the above-mentioned problem which is especially adapted to real situation.

The main contributions of this paper are summarized as follows. (1) A novel feature description approach based on Mean-Shift is proposed. It designs feature point description operator of weak texture region which not only has effect on weak texture feature point but also fits other strong texture feature points. (2) The Bhattacharyya coefficient is utilized in feature point matching processing and a fast matching search scheme is proposed. Bhattacharyya coefficient plays an important role for the similarity estimation between two multidimension vectors. Moreover, in order to decrease matching search time in wide baseline, we give searching constrain condition. Searching area of each point is

constrained on corresponding epipolar line according to basic matrix computed by eight-point approach. (3) In non-calibration wide baseline situation, we embed our descriptor in stereo matching algorithm. Experimental verification shows that our algorithm owned high matching accuracy and low complexity outperforms many state-of-the-art stereo matching algorithms and is especially suited for engineering applications. In order to facilitate understanding, we collect important abbreviations mentioned in this paper and show them in Nomenclature.

The rest of this paper is organized as follows. In Section 2, we present a novel stereo matching algorithm fitting in wide baseline and give the detailed deduction processing. In Section 3 we describe how to compute depth image based on Graph Cut and the method of our descriptor embedded. In Section 4, some experimental results are achieved to test the effectiveness and efficiency of the proposed algorithm. Section 5 concludes the whole paper and gives the further direction in the future.

## 2. Mean-Shift-Based Feature Point Description Operator

*2.1. Texture-Based Feature Point Description.* In order to get accurate matching point pairs, it is better to collect texture pattern near the point and then describe corresponding features. However, the images captured at wide baseline situation are harshly affected by light changing, visual angle and scale changing, and so on. Therefore, it does not result in texture robustness between different image pairs. As shown in Figure 1, it is not the same texture (circle area) around the same name image point under different angles. Obviously, it cannot directly describe feature points.

In contrast, the classic feature point detection algorithm can get the stable matching feature points at wide baseline situation. Therefore, feature points obtained by this approach can be used for feature description. So we use classical SIFT approach to get the corresponding robustness feature points.

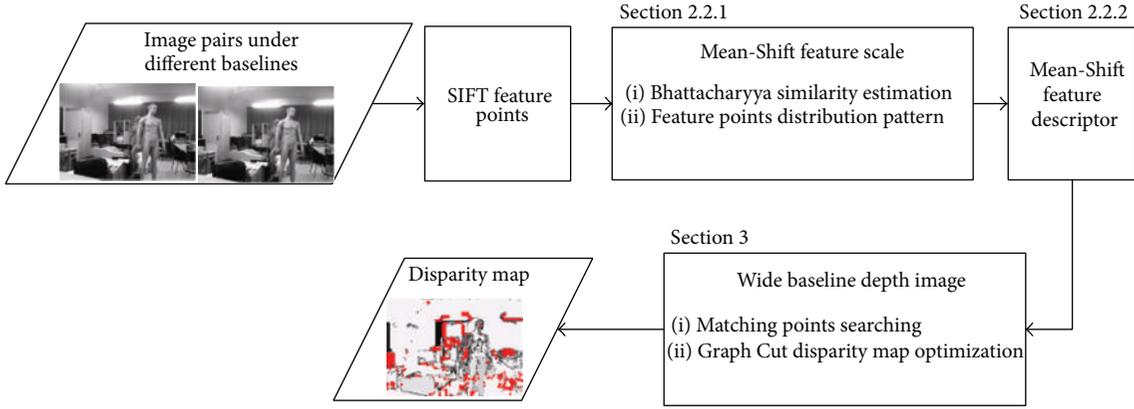


FIGURE 2: Flowchart of the proposed approach and its application.

On the other hand, in view of acquirement of the image depth information under wide baseline, each of image points should be matched. SIFT will get large number of feature points and if all feature points are involved to describe corresponding features, it surely brings a great many pieces of information that will decrease compute efficiency [19, 20]. So a compromise approach is to select partial feature points to build each feature description vector, in sure about each image point can correspond to feature points set; we use those feature points in its circle area as members of feature points set and compute the scale of this circle area with Mean-Shift approach. At last, the statistical information of feature points set will be used to describe each point's feature. Generally, a useful building process of feature descriptor is shown in Figure 2.

First of all, image pairs can be captured from different wide baseline and different angle of view. Stable SIFT feature points are set and will be, respectively, extracted by SIFT method from the two images. In view of improving efficiency, we present the method of deleting redundant SIFT feature points and improving feature descriptor coverage. Then, it randomly generated circular feature template which includes many SIFT feature points. The Bhattacharyya coefficient between two feature points will be computed in order to check their similarity and the size of circular area must be changed when the similarity exceeds threshold. The method of computing scale is Mean-Shift. At the end of feature extraction, the corresponding feature descriptor will be obtained by regular arranging of the feature points set within feature pattern. At last, we apply the above methods to two captured images and use Graph Cut algorithm for global optimization and then get disparity map finally.

**2.2. The Building Approach for Mean-Shift-Based Feature Point Descriptor.** Through careful analysis on Figure 1, it is indicated that the corresponding feature points set should be found around each point and then built its feature descriptor. Thus, we use Mean-Shift algorithm to compute the corresponding feature scale and determine the feature points set.

**2.2.1. Feature Scale Based on Mean-Shift.** Mean-Shift algorithm can be used to detect and track moving objects. It determines moving objects' locations by sample density distribution maximum value [21]. If a set  $\{\mathbf{x}_i\}_{i=1,\dots,n}$  belongs to  $R^d$  space, the corresponding density estimation function  $\hat{f}(\mathbf{x})$  will be computed by density kernel  $K(\mathbf{X})$  and window radius  $h$ :

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right). \quad (1)$$

Due to practical problem, density kernel  $K(\mathbf{X})$  has many forms. The mostly useful expression is

$$K_N(\mathbf{x}) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} \|\mathbf{x}\|^2\right). \quad (2)$$

According to the above two equations, density estimation function will be

$$\hat{f}(\mathbf{x}) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right). \quad (3)$$

If  $g(\mathbf{x})$  is equal to  $-k'(\mathbf{x})$ , kernel  $G(\mathbf{x})$  can be defined as

$$G(\mathbf{x}) = Cg(\|\mathbf{x}\|^2), \quad (4)$$

where  $C$  is a normal constant number. Now, we can deduce the gradient of density function

$$\begin{aligned} \hat{\nabla} f_{\kappa}(\mathbf{x}) &\equiv \nabla \hat{f}_{\kappa}(\mathbf{x}) = \frac{2}{nh^{d+2}} \sum_{i=1}^n (\mathbf{x} - \mathbf{x}_i) k' \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) \\ &= \frac{2}{nh^{d+2}} \sum_{i=1}^n (\mathbf{x} - \mathbf{x}_i) g \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) = \frac{2}{nh^{d+2}} \left[ \sum_{i=1}^n g \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right) \left( \frac{\sum_{i=1}^n \mathbf{x}_i g \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right)}{\sum_{i=1}^n g \left( \left\| \frac{\mathbf{x} - \mathbf{x}_i}{h} \right\|^2 \right)} - \mathbf{x} \right) \right]. \end{aligned} \quad (5)$$

If  $M_{h,G}(\mathbf{x})$  is equal to  $\sum_{i=1}^n \mathbf{x}_i g(\|\mathbf{x} - \mathbf{x}_i\|/h) / \sum_{i=1}^n g(\|\mathbf{x} - \mathbf{x}_i\|/h) - \mathbf{x}$ , we can get another form as follows:

$$M_{h,G}(\mathbf{x}) = \frac{h^2}{2/C} \frac{\hat{\nabla} f_{\kappa}(\mathbf{x})}{\hat{f}_G(\mathbf{x})}. \quad (6)$$

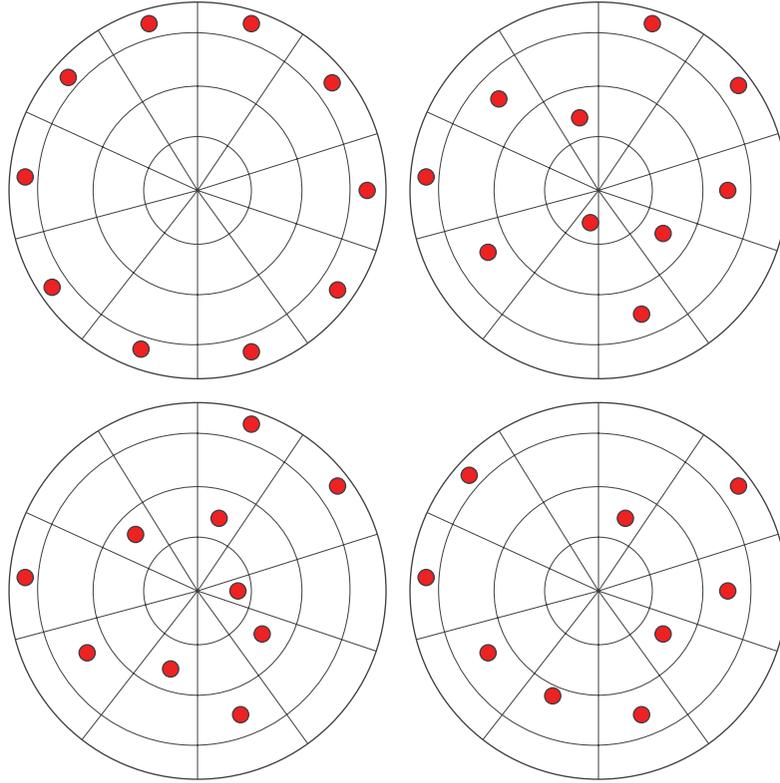


FIGURE 3: Feature area patterns.

This expression indicates that the gradient of density estimation function can be denoted as kernel function. Furthermore, the corresponding coordinate of the maximum point of density estimation function will be obtained through iterative computing.

In real application, it always needs to translate corresponding problem into mathematical model as mentioned before. Bhattacharyya coefficient [22] is usually used to estimate corresponding position coordinates. Suppose an object feature is  $z$  and its density function is  $q_z$ . Meanwhile, the center position of candidate object is  $\mathbf{y}$  and the corresponding density estimation function is  $p_z(\mathbf{y})$ . For the sake of finding the center position of candidate object which is most close to original object, we use a kind of Bhattacharyya coefficient presented as follows to estimate the distance between  $q_z$  and  $p_z(\mathbf{y})$ :

$$d(\mathbf{y}) = \sqrt{1 - \rho[\hat{p}(\mathbf{y}), \hat{q}]}, \quad (7)$$

where

$$\hat{\mathbf{q}} = \{\hat{q}_u\}_{u=1, \dots, m} \left( \sum_{u=1}^m \hat{q}_u = 1 \right),$$

$$\hat{\mathbf{p}}(\mathbf{y}) = \{\hat{p}_u(\mathbf{y})\}_{u=1, \dots, m} \left( \sum_{u=1}^m \hat{p}_u(\mathbf{y}) = 1 \right), \quad (8)$$

$$\hat{\rho}(\mathbf{y}) \equiv \rho[\hat{p}(\mathbf{y}), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}) \hat{q}_u}.$$

If  $\hat{\mathbf{p}}(\mathbf{y})$  and  $\hat{\mathbf{q}}$  are all  $m$  dimensional vectors,  $\rho[\hat{p}(\mathbf{y}), \hat{q}]$  will be regarded as the cosine function between them. So (7) can denote the vector similarity.

Here, we need to estimate a reasonable area for every point in image so that it can involve enough SIFT feature points to describe the point's features. According to Mean-Shift approach, we need to give an object's feature initially. A key point should be paid attention on our concern: the involved problem is not tracking problem. Thus, it results in being unable to find corresponding pattern in image. On the other hand, we believe that it can get the only feature descriptor if SIFT feature around the point has uniform distribution (the detailed descriptor structure approach will be given on the next section). Figure 3 gives several feature area pattern types.

We use Mean-Shift approach to estimate the area and covered angle. The corresponding coordinate of covered area, unlike tracking algorithms, is denoted as polar coordinate form  $\mathbf{x} = (r, \theta)$ , where  $r$  is radius and  $\theta$  is angle variable. Supposing  $\{\mathbf{x}_i\}_{i=1, \dots, n}$  is denoted as the polar coordinate position of feature points set in pattern, we define function  $b: \mathbb{R}^2 \rightarrow \{1, \dots, m\}$  as corresponding number index function, where  $m$  is the area number of circle area and  $b(\mathbf{x}_i)$  is the sequence number of  $\mathbf{x}_i$ . Then, the density estimation function [23, 24] can be written as

$$\hat{q}_u = C \sum_{i=1}^n k(\|\mathbf{x}_i\|^2) \delta(b(\mathbf{x}_i) - u), \quad (9)$$

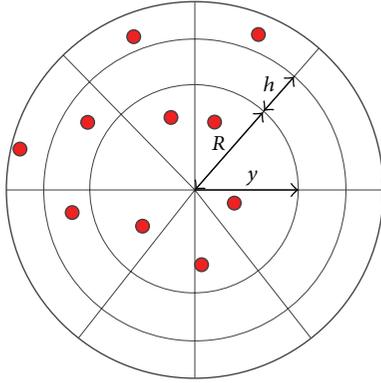


FIGURE 4: Mean-Shift tracking parts of object.

where  $\delta(\cdot)$  is impulse response function and  $\hat{q}_u$  should satisfy  $\sum_{u=1}^m \hat{q}_u = 1$ . The normalization constant  $C$  can be denoted as follows:

$$C = \frac{1}{\sum_{i=1}^n k(\|\mathbf{x}_i\|^2)}. \quad (10)$$

As shown in Figure 4, in view of estimating area radius, there is an assumption  $\mathbf{y} = (R, 0)$ , where  $R$  is estimation radius. If  $\{\mathbf{x}_i\}_{i=1, \dots, n_h}$  is the polar coordinate of real feature points set and  $h$  is the width of every little area, then the density estimation function of corresponding candidate radius will be

$$\hat{p}_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\frac{\|\mathbf{x}_i - \mathbf{y}\|^2}{h}\right) \delta(b(\mathbf{x}_i) - u), \quad (11)$$

where  $\sum_{u=1}^m \hat{p}_u = 1$ . The normal constant  $C_h$  can be denoted as

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k(\|\mathbf{x}_i - \mathbf{y}\|^2/h)}. \quad (12)$$

The following iterative procedures can be used to compute Bhattacharyya coefficient.

*Step 1.* Initialize  $R_0$  and compute corresponding Bhattacharyya coefficient:

$$\rho[\hat{\mathbf{p}}(\mathbf{y}_0), \hat{\mathbf{q}}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}) \cdot \hat{q}_u}. \quad (13)$$

*Step 2.* Compute weight  $\{w_i\}_{i=1, \dots, n_h}$ , where  $w_i = \sum_{u=1}^m \delta(b(\mathbf{x}_i) - u) \sqrt{\hat{q}_u / \hat{p}_u(\mathbf{y}_0)}$ .

*Step 3.* Compute new radius by Mean-Shift and the corresponding Bhattacharyya coefficient:

$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i g(\|\hat{\mathbf{y}}_0 - \mathbf{x}_i\|^2/h)}{\sum_{i=1}^{n_h} w_i g(\|\hat{\mathbf{y}}_0 - \mathbf{x}_i\|^2/h)}, \quad (14)$$

$$\rho[\hat{\mathbf{p}}(\mathbf{y}_1), \hat{\mathbf{q}}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\mathbf{y}_1) \cdot \hat{q}_u}.$$

*Step 4.* When  $\rho[\hat{\mathbf{p}}(\mathbf{y}_1), \hat{\mathbf{q}}] < \rho[\hat{\mathbf{p}}(\mathbf{y}_0), \hat{\mathbf{q}}]$ , old radius value will be updated as  $\hat{\mathbf{y}}_1 \leftarrow (1/2)(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)$ .

*Step 5.* The algorithm will be halted when  $\|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\| < \varepsilon$ . Otherwise, it sets  $\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1$  and returns to Step 1.

According to the above algorithm, we can get the estimation of radius  $R$  and final radius  $R'$  should be summation of two parts; that is,

$$R' = R + hc, \quad (15)$$

where  $c$  is the layer of the area and  $h$  is the width of every layer. Figure 5 shows the results of Mean-Shift algorithm and different radius depended on different patterns.

*2.2.2. Acquisition of Feature Point Descriptor.* With regard to real position, there are some points whose texture features are not available in their neighbor areas. We hope to use Mean-Shift approach to get the SIFT feature points set around this kind of points. At the same time, SIFT approach can collect those repeatable and robust feature points [25]. So those points' set can describe the feature of the point. Here, we build the corresponding feature descriptor of the point by feature points' space and pixel information. For the sake of feature matching, it has two conditions for building feature points descriptor.

- (1) The feature descriptor should be invariant for rotating, scaling, and lighting [26].
- (2) The feature descriptor of each point should be strong differentiation [27].

After given feature distribution pattern, each area can be cut into several fan areas. And then, according to texture feature distribution position of every fan area, we can define the corresponding descriptor as follows.

*Step 1.* According to feature points distribution pattern, we use Mean-Shift algorithm to search corresponding area and then find the number of corresponding SIFT texture feature points according to each fan area.

*Step 2.* According to the number difference of feature point distribution in each fan area, we define a line from detecting point to center of feature point (this point is the average position of feature points in the fan area) as shown in Figure 6.

*Step 3.* According to gray value of the end of the line and gradient information, the corresponding feature description vector is denoted as

$$d_{ij} = \mu(t_i) - \mu_{\text{line}}(t_i), \quad (16)$$

where  $\mu_{\text{line}}(t_i) = \mu(0) + t_i(\mu(1) - \mu(0))$ ,  $t_i = i/(N + 1)$ ,  $t_i \in [0, 1]$ .  $\mu(\cdot)$  is the characteristics on a point position (such as pixel value, gradient value, and other feature values). Here, we cut several same sizes ( $N$ ) in this line.  $\mu_{\text{line}}(t_i)$  is the pixel characteristics of  $t_i$  position using linear interpolation. So  $d_{ij}$

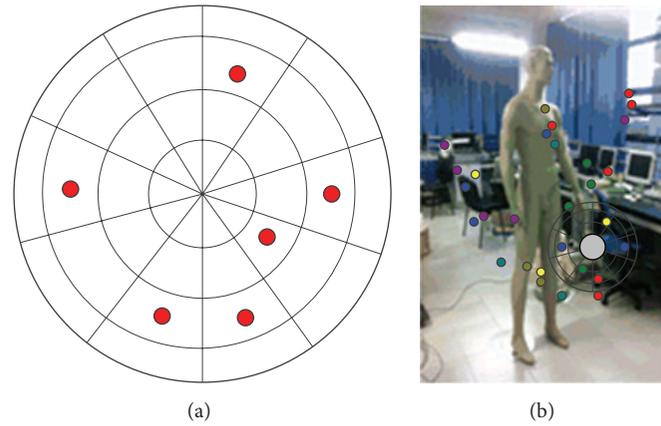


FIGURE 5: As the feature area pattern gives on (a), Mean-Shift will provide corresponding feature points on (b).

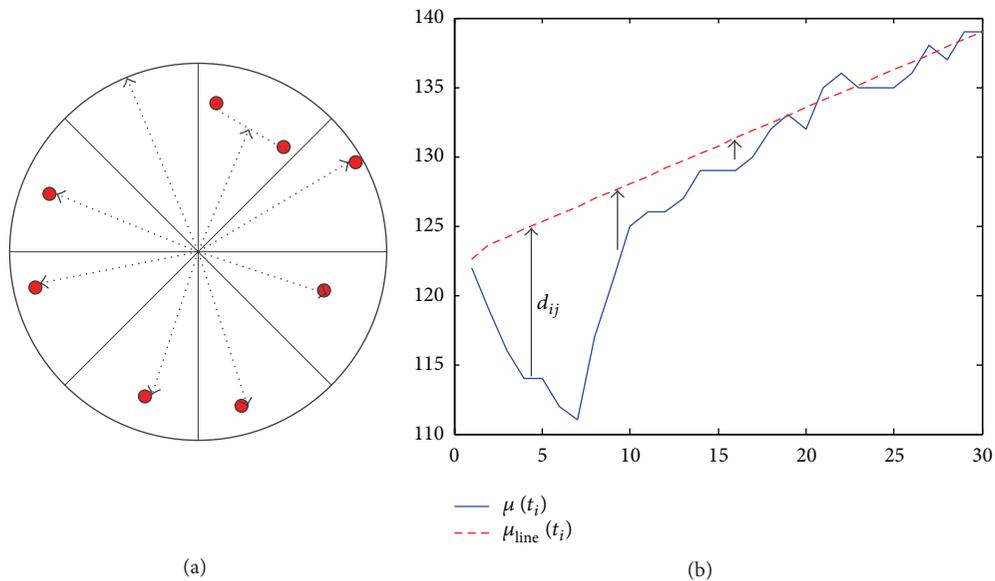


FIGURE 6: (a) Dotted lines with arrow denote the paths which are coming from the feature points to the center of the fan area. And the center is decided by those SIFT feature points' positions in the fan area. (b) The differences (image gradient feature is used to present each point's characteristics) between real feature and linear interpolation of each point.

is the difference between real feature value and linear interpolated value and  $j$  is the corresponding fan area. Supposing  $M$  is the number of fan area, the length of corresponding feature description vector is  $M \times N$  dimension.

*Step 4.* Taking  $\max \mu_j(1)$  corresponding fan area as initial area, the corresponding feature vector can be counter-clockwise ranked.

After the above steps, each point can obtain  $M \times N$  dimension feature vector. Significantly, we do not choose single feature point as feature description fundamental element. Within self-adaptive Mean-Shift feature pattern area, our descriptor is represented by texture distribution structure around the feature point. As a result, the structure is not changed in different scale. That is, the operator is scale

invariance. Besides, in view of image rotation, it cannot apply in image matching if the feature vectors are out-of-order. In other words, feature vectors of the same point will vary with rotation. Ingeniously, we define the initial position and its rank order in Step 4. So it is obviously known that those vectors produced by this algorithm are rotation variance.

### 3. Wide Baseline Depth Image Based on Graph Cut Algorithm

In traditional narrow baseline disparity map, it is always supposed that those captured images only have horizontal deviation. Figure 7(a) shows that the disparity value is horizontal deviation  $d$  of the same name point in image pairs. Dot point denotes the position of the same point in

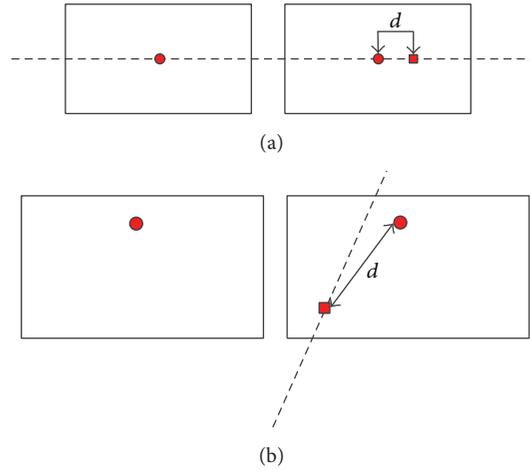


FIGURE 7: (a) Disparity map in narrow baseline; (b) disparity map in wide baseline.

reference image and square point shows the real position of disparity map. Because of the above supposition, the point's disparity value can be easily obtained by searching the position of the same name point along 1D horizontal level. This approach is easily achieved and has high computing efficiency. However, real captured images are almost not satisfied to horizontal deviation supposition. Thus, the stereo matching algorithm needs to rectify captured image and the rectification processing will surely bring a certain error. This will largely restrict its application. On the other hand, the algorithm complexity must be increased when directly dealing with disparity map in wide baseline situation. The key problem of the same name point searching way will not be limited on a horizontal line. It needs to search a whole image space.

Taking the above two problems into account, we give a disparity map building algorithm to overcome them. Figure 7(b) shows the disparity sketch map in wide baseline. Here, the definition of dot point and square point is the same as shown in Figure 7(a). Their distance is the disparity value. Then, the corresponding same point needs to tackle the two critical problems described as follows:

- (1) How to search the same name point in disparity map.
- (2) How to adjust noninteger disparity value to integer form.

As to the first problem, we use eight-point approach and feature descriptor by weak feature description (WFD) to search the position of the same name point in disparity map. The dotted line in Figure 7(b) is the epipolar line constraint based on eight-point approach to find the corresponding same name point. In view of accuracy, we use WFD descriptor to match it.

As to the second problem, we should know the problem which results from traditional disparity map processing. The disparity value is denoted by discrete integer. But, we will get noninteger form of disparity value by our above definitions. Thus, our approach needs to adjust them to integer form

so as to subsequent optimization process. The distance from the reference point to the extreme line of each point is segmented by the given value of the disparity and then the obtained distance values can be output as the integer form.

Chen and Jhang [28] used improved Graph Cut algorithm to transform stereoscopic image sequence to depth map sequence. Bleyer and Gelautz [29] used their Graph Cut-based algorithm to overcome three classical problems in stereo methods. We simply review Graph Cut algorithm in order to get better understanding of disparity map generating process. In stereo matching, the disparity is usually considered as a label of pixels. Suppose there are two images to be matched. The set of all pixels in the first image is represented by  $X$ , and the matching target is assigned a label  $f(x) \in \{l_1, l_2, \dots, l_n\}$  to each pixel  $x$  in  $X$ , where the significance of the label is denoted as a possible disparity of pixel  $x$  and  $L$  represents the set of all the possible labels. Under the above assumption of narrow baseline, the disparity will only be changed depending one component like horizontal component. At this time, the elements in the set  $L$  are one-dimensional vector. However, as for the case of wide baseline, label set  $L$  is a two-dimensional vector set. Thus, matching problem will be changed so as to find an appropriate label  $f(x)$  and make pixel  $x$  in first image matching pixel  $x + f(x)$  in second image.

In order to get the proper labels of pixels, an energy function of label is constructed by two matching rules. First, the pixels' data of matched pixels cannot have too much differences. Second, adjacent pixels have similar disparity values. The former is called the data constraint and the latter is called smooth constraint. The energy function is shown as follows:

$$E(f) = E_D(f) + E_S(f) = \sum_{x \in X} D_x(f(x)) + \sum_{\{x,y\} \in N} V_{x,y}(f(x), f(y)), \quad (17)$$

where  $N$  is denoted as a set of adjacent pixel pairs.



FIGURE 8: Depth image in different baselines.

As to minimize  $E(f)$ , an undirected network  $G$  is constructed, which makes the label function set and the Graph Cut of the network  $G$  to form a bidirectional mapping, and the energy  $E(f)$  is the corresponding cutting capacity of each label  $f$ . Thus, the minimization cutting solving problem of the undirected network  $G$  is the minimization problem of the energy function  $E(f)$  and the minimum cut problem is attributed to the computation of the maximum flow. We use the Potts model [30] which is described as follows:

$$E(f) = \sum_{\{x,y\} \in N} u_{\{x,y\}} \cdot T(f_x \neq f_y) + \sum_{x \in p} D_x(f_x), \quad (18)$$

where  $f_x$  and  $f_y$  are the corresponding index of image point  $(x, y)$ , respectively. Data supporting item  $D_x(f_x)$  denotes support degree of real data when the index of point  $x$  is  $f_x$ . Generally, the support degree is computed by points' similarity of two images. Experimental conclusion shows that the corresponding support degree will be smaller when the features of two points are more similar and the energy

function will be less subsequently. Here, we define support degree of index  $f_x$  as follows:

$$C_{\text{fwd}}(\mathbf{x}, \mathbf{y}, f) = \min_{|\Delta f| \leq 1/2} |D_{\mathbf{x}} - D_{\mathbf{y}+\Delta f}|, \quad (19)$$

where  $D_{\mathbf{x}}$  is the corresponding WFD feature vector and the setting of region of neighborhood can make the accuracy reach subpixel level.  $C_{\text{fwd}}$  is the error minimum between all points coming from points set  $\{\mathbf{y}_i\}$  whose distance holds  $f \pm 1/2$  with position  $\mathbf{x}$  and other feature vectors.

Similarly, it needs to rectify position  $x$  in reference map. That is,

$$C_{\text{rev}}(\mathbf{x}, \mathbf{y}, f) = \min_{|\Delta f| \leq 1/2} |I_{\mathbf{x}+\Delta f} - I'_{\mathbf{y}+f}|. \quad (20)$$

Comprehensively considering (18) and (19), we can obtain final form of data support degree as follows:

$$\begin{aligned} C(\mathbf{x}, \mathbf{y}, f) \\ = (\min \{C_{\text{fwd}}(\mathbf{x}, \mathbf{y}, f), C_{\text{rev}}(\mathbf{x}, \mathbf{y}, f), \text{const}\})^2. \end{aligned} \quad (21)$$

Because of 2D searching area, the accuracy 0.5 is denoted as four neighborhoods of this point. In addition, we need to

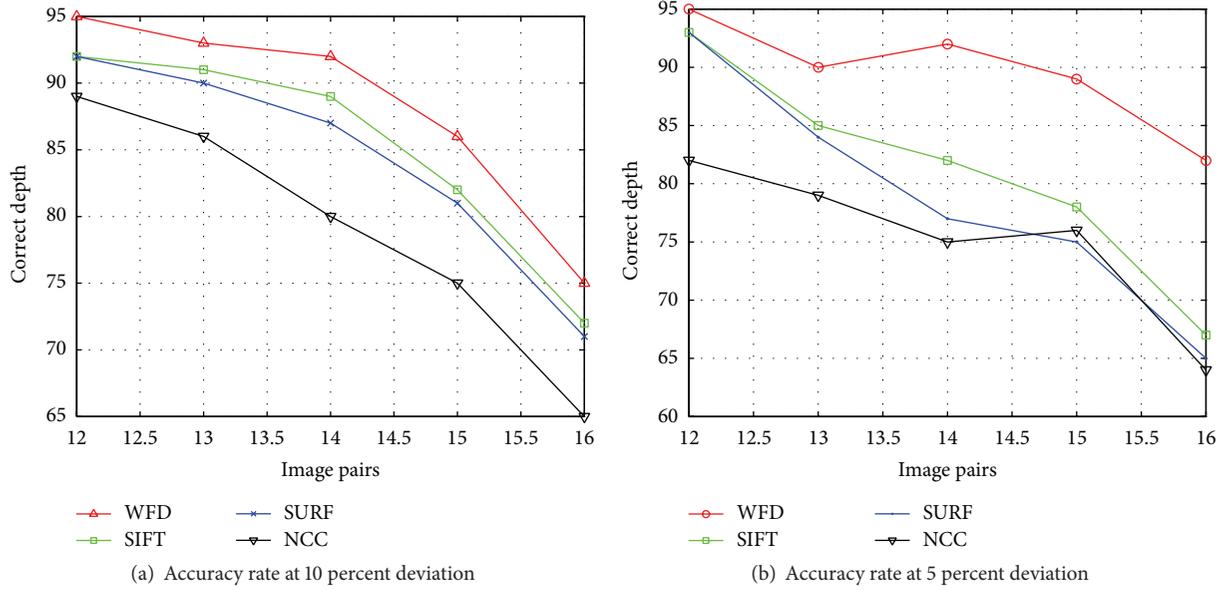


FIGURE 9: Accuracy rate of feature descriptor in wide baseline changing.

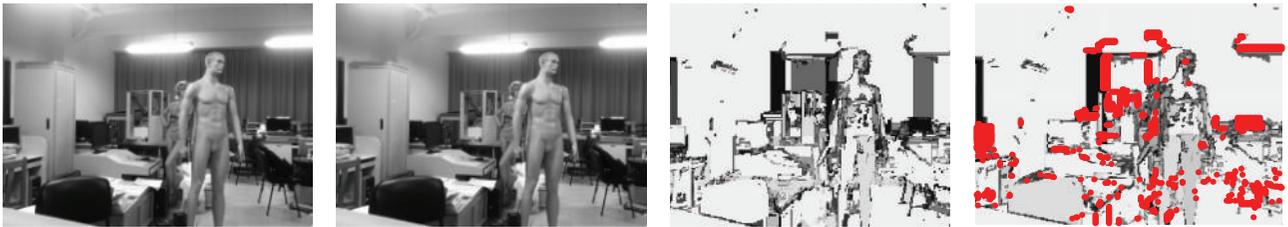


FIGURE 10: Depth image based on WFD feature.

assign noninteger index  $f$  to integer index sets  $\{f'\}$ . So (20) will simplify as follows:

$$C'(x, y, f') = \min_{f \in f'} C(x, y, f). \quad (22)$$

In summary, we give the corresponding definitions of Graph Cut parameters. Subsequently, some stereo matching effects will be represented in detail in Experimental Results.

#### 4. Experimental Results

In this section, we show the efficiency detection results of our feature point descriptor. We compare it with traditional scale-invariance feature transform (SIFT), speeded up robust feature (SURF), and normalized cross-correlation (NCC) and estimate the change ranges of corresponding baseline in available accuracy.

*4.1. Comparisons of Different Descriptors.* Figure 8 shows the efficiency comparisons of other descriptors including SIFT, SURF, and NCC. The upper-left image is the reference image

and the corresponding depth image can be computed by using image pairs which combined the reference and each image at the right side. In spite of using the same Graph Cut algorithm in wide baseline, their differences only rest with different feature descriptors. On the other hand, matching range of each point is the corresponding epipolar of another image in order to improve algorithm efficiency and the corresponding baseline increases in turn from left to right. Meanwhile, we provide the corresponding depth image in narrow baseline as standard image. The second row image is the depth image in narrow baseline and the third to the sixth row images are the depth of weak feature description (WFD), SIFT, SURF, and NCC. The dot point is denoted as the difference between each depth image and standard image.

Figure 9 shows the curve of depth error which includes the error rate between the standard depth image and the depth image obtained by different descriptors at baseline changing situation. Here, we, respectively, give curve changing maps in five and ten percent of depth value deviation. For the sake of detecting algorithm efficiency of WFD, we chose many weak texture images as testing map. It obviously surpasses other descriptors from comparisons results.

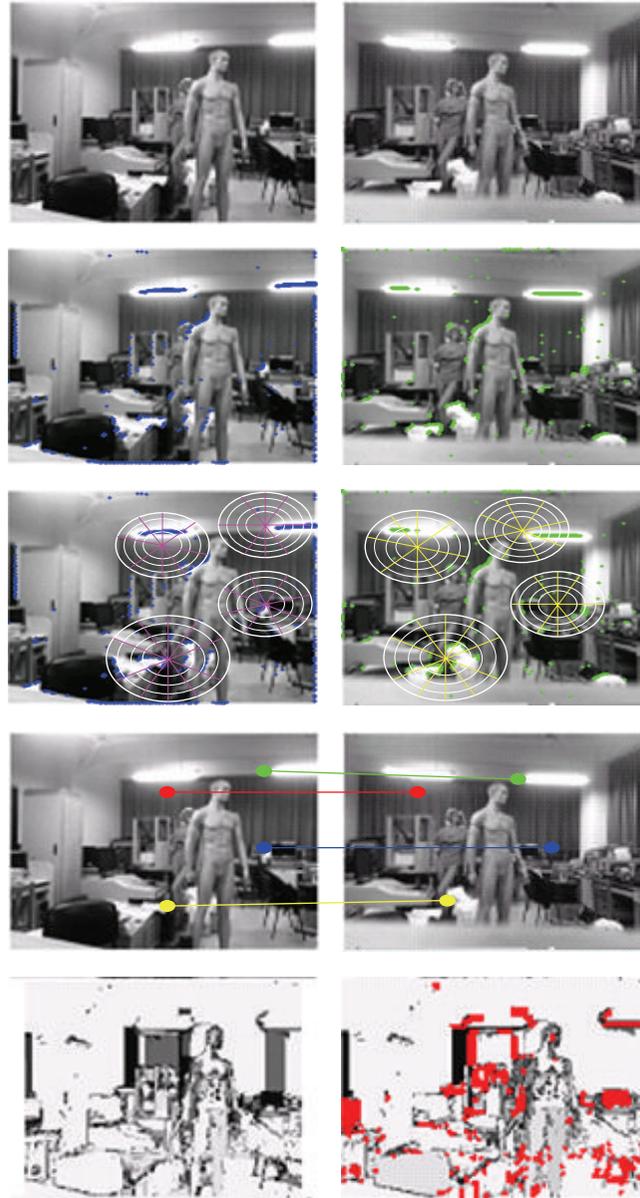


FIGURE 11: An example of the complete process.

In addition, when the error set of WFD is 5 percent, the accuracy rate is higher and their results are very close when the error set of WFD and SIFT is 10 percent.

On the other hand, we use different colors to present the right and wrong matching points. Figure 10 shows an analysis example selected from a set of image pairs. There, the depth matching error threshold is 5 percent.

In view of comprehensive understanding of our algorithm, we give an example of the complete process using a pair of images which is shown in Figure 11. The left column in Figure 11 shows the ground truth image processing and the right column displays the processing of image captured in another baseline. The first line gives original images of ground truth and another situation. The second line represents the feature points obtained after SIFT processing and the third

line shows the feature descriptor using our proposed algorithm. The fourth line is to find the depth of the two images through matching with our feature descriptor. The left side of the fifth line is the depth image computed by narrow baseline algorithm, while the right is a combination method between Mean-Shift and Graph Cut global optimization algorithm.

*4.2. Depth Image in Different Baselines.* Figure 12 shows depth images on different image pairs using WFD operator and the depth image on diagonal line is standard image in narrow baseline. Table 1 gives the accuracy rate changing of depth image in different baselines. On the other hand, we use six visual angles to capture images and each visual angle changes 10 degree in horizontal shown in Figure 12.



FIGURE 12: The depth image based on WFD feature descriptor in wide baseline.

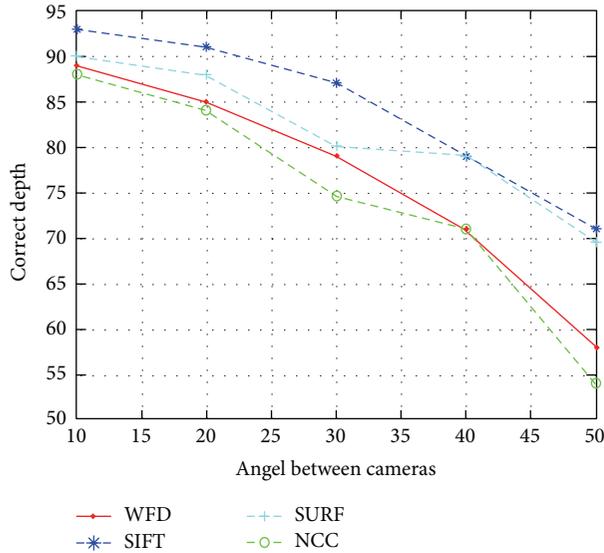


FIGURE 13: Accuracy rate changing curve of WFD in different capture angles.

TABLE 1: Depth image accuracy estimation percentage.

	1	2	3	4	5	6
1	—	89.7	47.2	44.8	69.6	42.1
2	84.2	—	90.1	80.1	72.1	66.7
3	81.9	83.2	—	92.9	74.6	83.5
4	80.8	73.9	87.6	—	78.3	86.4
5	77.2	55.2	87.1	90.4	—	87.6
6	70.9	53.9	71.7	79.0	88.9	—

Figure 13 describes the average accuracy rate changing curve of WFD in different baselines. Here, the depth image deviation threshold is 10 percent. That is, the right depth matching point is less than 10 percent deviation between depth image and referent image. It is not difficult to see that the accuracy is higher when the capture angle changes between plus and minus 20 percent.

**4.3. Robustness of WFD in Different Image Changing.** From Figure 8, we can find that our descriptor has strong robustness on image scale, resolution, and luminance according to the error deviation of depth image in different baseline. It also spelled out strong robustness from the error rate curve changing in Figure 9. Otherwise, in order to test real effect in real scene, we get some depth images of several image pairs (see Figure 14) by our algorithm. Because of matching difficulty of weak texture area in a large range, it generally uses high resolution image (like  $3027 * 2048$ ) to do stereo matching. Here, we use  $684 * 512$  resolution image and also get well corresponding stereo matching image.

## 5. Conclusions and Future Work

This paper presents an algorithm using feature point extraction and description to solve the depth image acquirement

problem in wide baseline. Initially, we introduce a Mean-Shift-based feature descriptor which can be used to build corresponding feature description vector according to its texture feature point distribution situation. The approach makes it describe the point in weak area and realize right location in wide baseline. On the other hand, we expand disparity map of narrow baseline to wide baseline situation. Utilizing Graph Cut algorithm to optimize disparity information, the depth for arbitrary image pairs can be extracted without calibration. Experimental data shows that WFD descriptor can get good depth image in low resolution image which means adapting the algorithm to video stream. Otherwise, we apply Mean-Shift algorithm to WFD and get texture feature of each point. So the corresponding feature descriptor ensures strong robustness for image changing and it allows defining arbitrary pattern according to situations and results in strong flexibility.

This research can be extended in two directions. Firstly, because of various conditions restriction, we do not have access to high resolution images and it does not result in perfect performance of our approach on a large range of weak texture area. We hope that the subsequent work can be carried out on the basis of the high resolution image to expand our stereo matching strategy. Secondly, it needs to dynamically adjust the window scale of Mean-Shift to obtain a stable feature descriptor and this is directly responsible for a huge amount of algorithm run time. In the latter stage, it will be more suitable for the practical engineering application if it can decrease the frequency of dynamic adjustment and appropriately fix the window scale in low dynamic changing.

## Nomenclature

SIFT: Scale-invariance feature transform  
 SURF: Speeded up robust feature  
 NCC: Normalized cross-correlation  
 WFD: Weak feature description  
 Harris: Harris corner detector  
 BAND: Binary affine invariant descriptor.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was supported in part by the Key Project of the National Nature Science Foundation of China (no. 61134009), the National Nature Science Foundation of China (nos. 61473077 and 61473078), Program for Changjiang Scholars from the Ministry of Education, Specialized Research Fund for Shanghai Leading Talents, Project of the Shanghai Committee of Science and Technology (no. 13JC1407500), and Innovation Program of Shanghai Municipal Education Commission (no. 14ZZ067).



FIGURE 14: The stereo matching effect map using WFD for wide baseline and low resolution image.

## References

- [1] F. Cheng, H. Zhang, M. Sun, and D. Yuan, "Cross-trees, edge and superpixel priors-based cost aggregation for stereo matching," *Pattern Recognition*, vol. 48, no. 7, pp. 2269–2278, 2015.
- [2] X. Tan, C. Sun, X. Sirault, R. Furbank, and T. D. Pham, "Feature matching in stereo images encouraging uniform spatial distribution," *Pattern Recognition*, vol. 48, no. 8, pp. 2530–2542, 2015.
- [3] C. Stentoumis, L. Grammatikopoulos, I. Kalisperakis, and G. Karras, "On accurate dense stereo-matching using a local adaptive multi-cost approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 91, pp. 29–49, 2014.
- [4] A. Hosni, M. Bleyer, and M. Gelautz, "Secrets of adaptive support weight techniques for local stereo matching," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 620–632, 2013.
- [5] F. Cheng, H. Zhang, D. Yuan, and M. Sun, "Stereo matching by using the global edge constraint," *Neurocomputing*, vol. 131, pp. 217–226, 2014.
- [6] D. Shin and T. Tjahjadi, "Clique descriptor of affine invariant regions for robust wide baseline image matching," *Pattern Recognition*, vol. 43, no. 10, pp. 3261–3272, 2010.
- [7] D. Zhang, Y. Wang, W. Tao, and C. Xiong, "Epipolar geometry estimation for wide baseline stereo by clustering pairing consensus," *Pattern Recognition Letters*, vol. 36, no. 1, pp. 1–9, 2014.
- [8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.
- [9] M.-Y. Shieh and T.-M. Hsieh, "Fast facial detection by depth map analysis," *Mathematical Problems in Engineering*, vol. 2013, Article ID 694321, 10 pages, 2013.
- [10] Y. Jiang, Y. Xu, and Y. Liu, "Performance evaluation of feature detection and matching in stereo visual odometry," *Neurocomputing*, vol. 120, pp. 380–390, 2013.
- [11] G. A. Kordelas, D. S. Alexiadis, P. Daras, and E. Izquierdo, "Enhanced disparity estimation in stereo images," *Image and Vision Computing*, vol. 35, pp. 31–49, 2015.
- [12] K. Liao, G. Liu, and Y. Hui, "An improvement to the SIFT descriptor for image representation and matching," *Pattern Recognition Letters*, vol. 34, no. 11, pp. 1211–1220, 2013.
- [13] Y. Pang, X. Cao, L. Zhang, and A. Hussein, "Special issue on image feature detection and description," *Neurocomputing*, vol. 120, pp. 1–3, 2013.
- [14] P. F. Alcantarilla, L. M. Bergasa, and A. J. Davison, "Gauge-SURF descriptors," *Image and Vision Computing*, vol. 31, no. 1, pp. 103–116, 2013.
- [15] Q. Zhang, T. Rui, H. Fang, and J. Zhang, "Particle filter object tracking based on harris-SIFT feature matching," *Procedia Engineering*, vol. 29, pp. 924–929, 2012.
- [16] X. Qu, F. Zhao, M. Zhou, and H. Huo, "A novel fast and robust binary affine invariant descriptor for image matching," *Mathematical Problems in Engineering*, vol. 2014, Article ID 129230, 7 pages, 2014.
- [17] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," *Pattern Recognition Letters*, vol. 49, pp. 250–258, 2014.
- [18] D. Ming, J. Li, J. Wang, and M. Zhang, "Scale parameter selection by spatial statistics for GeOBIA: using mean-shift based multi-scale segmentation as an example," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 106, pp. 28–41, 2015.
- [19] S. Zhong, J. Wang, L. Yan, L. Kang, and Z. Cao, "A real-time embedded architecture for SIFT," *Journal of Systems Architecture*, vol. 59, no. 1, pp. 16–29, 2013.
- [20] C.-C. Chen and S.-L. Hsieh, "Using binarization and hashing for efficient SIFT matching," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 86–93, 2015.
- [21] Y.-R. Lai, K.-L. Chung, C.-H. Chen, G.-Y. Lin, and C.-H. Wang, "Novel mean-shift based histogram equalization using textured

- regions,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 2750–2758, 2012.
- [22] B. K. Patra, R. Launonen, V. Ollikainen, and S. Nandi, “A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data,” *Knowledge-Based Systems*, vol. 82, pp. 163–177, 2015.
- [23] Y. A. Ghassabeh, T. Linder, and G. Takahara, “On some convergence properties of the subspace constrained mean shift,” *Pattern Recognition*, vol. 46, no. 11, pp. 3140–3147, 2013.
- [24] Y. Aliyari Ghassabeh, “A sufficient condition for the convergence of the mean shift algorithm with Gaussian kernel,” *Journal of Multivariate Analysis*, vol. 135, pp. 1–10, 2015.
- [25] F. Bellavia, D. Tegolo, and C. Valenti, “Keypoint descriptor matching with context-based orientation estimation,” *Image and Vision Computing*, vol. 32, no. 9, pp. 559–567, 2014.
- [26] S. Bianco, D. Mazzini, D. P. Pau, and R. Schettini, “Local detectors and compact descriptors for visual search: a quantitative comparison,” *Digital Signal Processing*, vol. 44, pp. 1–13, 2015.
- [27] W.-J. Choi and T.-S. Choi, “Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 37–54, 2014.
- [28] W.-M. Chen and S.-H. Jhang, “Improving Graph Cuts algorithm to transform sequence of stereo image to depth map,” *The Journal of Systems and Software*, vol. 86, no. 1, pp. 198–210, 2013.
- [29] M. Bleyer and M. Gelautz, “Graph-cut-based stereo matching using image segmentation with symmetrical treatment of occlusions,” *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 127–143, 2007.
- [30] W. E. Frazier, G. S. Rohrer, and A. D. Rollett, “Abnormal grain growth in the Potts model incorporating grain boundary complexion transitions that increase the mobility of individual boundaries,” *Acta Materialia*, vol. 96, pp. 390–398, 2015.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

