AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 55(1) (2023) 125-130
DOI: 10.22060/eej.2022.21687.5489

Dense Stereo Matching Based on the Directional Local Binary Pattern

P. Bagheri, A. M. Fotouhi *

Department of Electrical Engineering, Tafresh University, Tafresh, Iran

ABSTRACT: New applications such as 3D graphics, 3D displays, and image-based modeling have made stereo vision an active research area in recent years. In dense disparity map estimation, which is a basic problem in stereo vision, using two left and right images taken from a scene from two different positions, the disparity of each pixel of the reference image is determined (meaning determining each pixel with how displacement is appeared in the other image). Based on the disparity value, the depth of each pixel in scene is simply determined. For dense disparity map estimation, local stereo matching methods are simpler and faster than global methods, and therefore suitable for real time applications. In these methods, defining proper window which aggregate intensity pattern as well as keeping disparity consistency in all the window area, is an important challenge. To overcome this challenge, the idea of directional multiple window has been proposed in the previous researches. On the other hand, local binary patterns have considerable success in pattern recognition applications, while computationally simple. Therefore, the idea of using local binary pattern in a directional multiple window arrangement is proposed for dense stereo matching in this paper. Experimental results on standard stereo images show the better performance of the proposed method with respect to other proposed binary descriptors.

Review History:

Received: Aug. 24, 2022
Revised: Dec. 16, 2022
Accepted: Dec. 17, 2022
Available Online: Mar. 01, 2023

Keywords:

Stereo vision
Stereo matching
Dense disparity map
Local binary pattern

1- Introduction

Estimating depth or disparity using two rectified left and right images, taken from a scene from two different positions (called stereo vision), is one of the most important issues in machine vision. New applications such as 3D graphics, 3D displays, and image-based modeling have made stereo vision an active research area in recent years. The main problem in stereo vision is to find the corresponding pixels of the two images or stereo matching. In feature-based stereo, the matching of features, such as edges or corners, is desired and a sparse disparity map is generated. In the dense stereo, which is considered in this paper, the corresponding search process is performed for all pixels of the reference image and a dense disparity map is obtained.

The simplest way to solve the stereo matching problem is considering minimum-maximum disparity and epipolar constraints. Based on the intensity or color consistency constraint, each pixel of a row of the left image is matched with a pixel in the corresponding row of the right image, which have the most color or intensity similarity. This idea is logical if it is assumed that each point of the environment has a unique color. On the other hand, it appears with a unique and identical color in two images. However, the limitation of the number of brightness levels, the difference in brightness of a point of the scene in two images due to viewing from two different positions, the presence of shadows, different camera specifications, noise, occlusion, etc., have made stereo matching a complicated problem.

In [1], dense stereo matching algorithms are categorized and compared. According to [1], these algorithms generally consist of four main steps: 1) calculating the cost of pixel-to-pixel matching, 2) cost aggregation, 3) selection of disparity for each pixel and 4) refinement or improvement of dense disparity map. Depending on the strategy used in the third step, the stereo matching algorithms are divided into two categories: local methods and global methods.

In global methods, an aggregative cost function is defined by components based on certain matching constraints, such as disparity continuity and intensity consistency across a row or all area of the image. Afterwards, disparity for each pixel is selected by minimizing this cost function. These algorithms have a high computational complexity, and in some cases, their convergence or reaching the global minimum depends on the initial values of the parameters, the quality of the initial estimation of the disparity values, and the accurate segmentation of the image. In local methods, a window around each pixel of the reference image is considered, and by sliding a window with the same shape in the other image on the matching candidate pixels, the cost of the matches is calculated and the corresponding pixel is simply selected with the minimum matching cost. Local methods have less computational complexity than global methods, but in these

*Corresponding author’s email: fotouhi@tafreshu.ac.ir

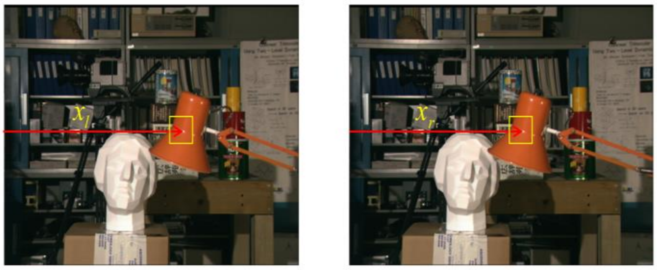

Fig. 1. Matching pixels in left and right images

methods, it is important to determine the exact matching cost to obtain a smooth and detailed disparity map due to the simplicity of the selection process.

In this paper, a dense stereo local matching algorithm is presented, which describes each pixel of the left and right images using a directional local binary pattern, and by comparing the feature vector of each pixel of the reference image with the matching candidate pixels in the other image, selects the pixel with the minimum matching cost as the corresponding pixel.

The rest of the paper is organized as follows: in Section 2, an overview of local stereo matching methods is given. In Section 3, the proposed algorithm is considered. Section 4 presents the experimental results and Section 5 is dedicated to the summary and conclusions.

2- An Overview of Local Dense Stereo Matching Methods

Using the depth continuity assumption, the neighborhood information of each pixel can be used to reduce the ambiguity in the matching, and thus reduce the incorrect matches. This is the main idea of the basic local methods, which can also be called the windowing method.

In the windowing method, as shown in Fig. (1), a square window with specified dimensions around the desired pixel in the reference image is considered. Based on a specific difference criterion function, the matching cost between this window and the similar shaped window in each matching candidate position in the second image is calculated [1].

Matching candidate positions for the desired pixel in other image are limited by considering the epipolar constraint and the minimum-maximum disparity constraint. The position in which the cost of matching is minimized determines the pixel of the second image that corresponds to the desired pixel of the first image. In other words, the value of disparity is determined by a simple selection of the best among candidates or WTA (Winner Takes All) in this method and the disparity d is determined for each pixel from Eq. (1):

$$d=x_l-x_r \quad (1)$$

In Eq. (1), x_l and x_r are the positions of the matching pixels in the left and right images, respectively. In local methods, the selection of disparity value for each pixel is made regardless of the calculated disparity values for other pixels. In addition to being conceptually simple, the basic windowing method can be implemented with low computational cost and is therefore suitable for real-time applications.

In areas without texture, little texture, repetitive texture, and without changing the depth, the size of the window should be chosen large enough so that the window contains more pixels to obtain the higher signal to noise ratio, more discrimination power, and thus more reliable matchings. On the other hand, in the vicinity of depth discontinuities, the dimensions of the window should be small enough to avoid changing the disparity of the pixels inside the window because the matching cost for the left and right windows is calculated with the assumption of disparity consistency of window pixels called DCWP assumption [2]. Therefore, it is a good idea to choose the size of the window adaptively in different pixels of the image so that the size of the window is large enough and on the other hand, the disparity of the pixels inside the window remains constant. However, disparity is the output of the problem, so different methods try to maintain the DCWP assumption inside the window with different ideas.

In the adaptive window method [3, 4], the idea has been proposed to obtain the optimal window size and shape per pixel by initial estimation of disparity and an iterative algorithm. In the multiple-window method, unlike the adaptive window method, which is often based on iterative algorithms or complex and time-consuming optimizations, a specific window group is used in all image pixels [5-10]. In methods presented in [11] and [12], in a specific window space, appropriate weights are assigned to the pixels inside the window. The basic idea of these methods is to assign a larger weight to the pixels of the window that are more likely to have a disparity equal to the center pixel. In [11], the weight of each pixel is determined based on the color similarity of that pixel to the central pixel and the geometric distance. The greater the color similarity of a pixel to a central pixel and the smaller the geometric distance, the more likely they are to be on the same surface and therefore have the same spatial disparity, so it should weigh more in the window.

In recent researches, the use of binary descriptors has become popular in various applications of machine vision, including texture analysis and pattern recognition, due to their resistance to changes in brightness and noise as well as simple calculations [13]. A various range of binary operators has been proposed, the main difference being the in-window sampling pattern for local intensity comparisons to produce the binary string descriptor in each pixel [13]. Binary descriptors are also used in stereo matching, and while they are simple to calculate, the simple Hamming distance function is usually used to check their matching in two images [14, 15]. These patterns can be divided into two general categories including central pixel and non-central pixel based. In central pixel-based descriptors, unlike non-central pixel based descriptors, the central pixel of the window must be present in all comparisons to produce a bit zero or one.

Fig. (2) shows the sampling patterns for a 7x7 window for the five most important feature descriptors, including

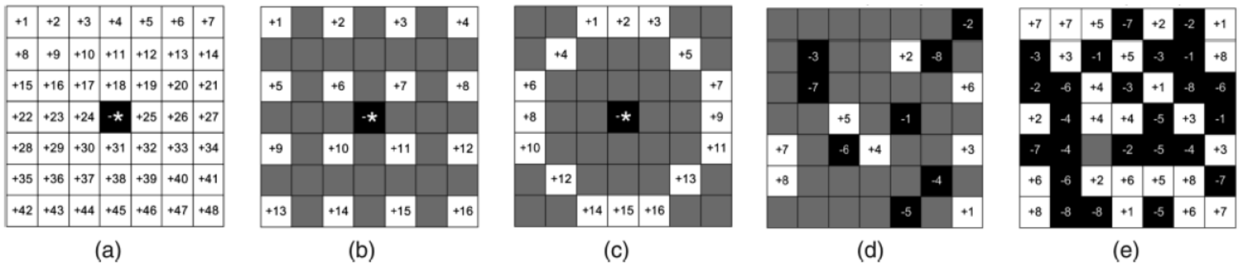

Fig. 2. Sampling pattern in binary descriptors, (a) CENSUS-dense, (b) CENSUS-sparse, (c) LBP, (d) BRIEF, (e) STABLE [15]

CENSUS-dense, CENSUS-sparse, LBP, BRIEF and STABLE (Stochastic Binary Local Descriptor). The first three descriptors are central pixel-based and the other two descriptors are non-central pixel-based descriptors. These descriptors produce a string of 48-, 16-, 16-, 8-, and 8-bit length, respectively. The *i*'th bit in the K-bit string of the central pixel-based descriptor for the central pixel *p* of the window denoted by *b(p.i)* is calculated as:

$$b(p.i)=S(f(+i)-f(p)) \tag{2}$$

and for the non-central pixel-based descriptor is calculated as:

$$b(p.i)=S(\sum^{J}_{(j=1)}(f(+i_j)-f(-i_j))) \tag{3}$$

In Eq. (3), *J* is the number of pixel pairs participating to form *i*'th bit of the descriptor. The S function in Eq. (3) is also defined as Eq. (4):

$$S(x)=\begin{cases}1 & if\ x\geq 0\\ 0 & if\ x<0\end{cases} \tag{4}$$

Finally, the value of the descriptor *des* in pixel *p* is obtained as:

$$des(p)=\sum^{K}_{i=1} b(p.i)*2^{i-1} \tag{5}$$

The characteristics of BRIEF and STABLE descriptors is that their size is independent of the size of the window.

In the BRIEF method, the bit of the descriptor is obtained with one pair of pixels generated by applying an isotropic Gaussian distribution function to the window, and therefore in Eq. (3), J = 1. In [16], it is shown that the BRIEF descriptor on standard databases with lower computational cost has comparable recognition accuracy compared to SURF and SIFT descriptors. In [14], it is mentioned that the use of BRIEF descriptor alone in the vicinity of depth discontinuities causes the phenomenon of fattening (progress of closer surface disparity at the farther surface) in the disparity map and therefore combines it with a binary mask based on color similarity of pixels inside the window.

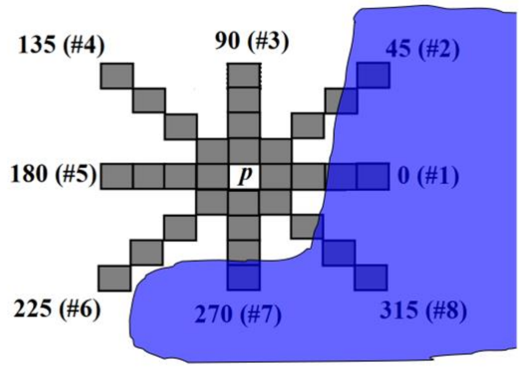

Fig. 3. Radial window behavior near a depth discontinuity

In the STABLE descriptor proposed in [15], which is actually an extension of the BRIEF descriptor, more than one pixel pair contributes to a single-descriptor bit, and therefore in Eq. (3), J > 1. For example, in the descriptor shown in Fig. 2(e), three pairs of pixels are used to calculate each bit of the descriptor, and therefore in Eq. (3), J = 3. In [15], the use of STABLE in stereo matching leads to better results than CENSUS and LBP, and comparable or better results than BRIEF. In this paper, a new method for dense stereo matching based on the idea of radial multiple window and directional binary descriptor is proposed.

3- Proposed Method

The use of a multiple-window method in [2] based on eight radial windows with the arrangement shown in Fig. (3) is proposed as a preprocessing to reduce the disparity space. The basic idea of this method is that in the vicinity of depth discontinuities, some radial windows do not overlap with depth discontinuities. For example, in Fig. (3), radial windows number 3, 4, 5, and 6 do not intersect with depth discontinuities and can therefore most likely lead to the correct disparity response due to maintaining DCWP assumption. Using the weighted window method [2, 17] and dynamic programming method [18] in this space, while leading to better results, the speed of extracting the dense disparity map has also been significantly increased. Therefore, it can be concluded from the results of previous researches that using the radial window pattern can increase the probability of achieving the correct disparity response for a pixel. In [2], to calculate the

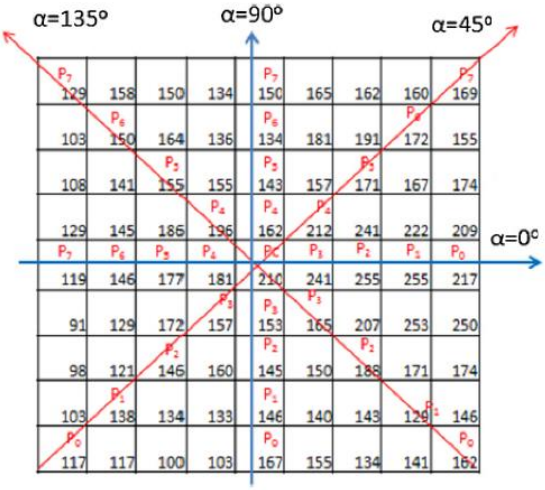

Fig. 4. Sampling pattern in window for dLBP_α for α=0, α=45, α=90 and α=135 [19]

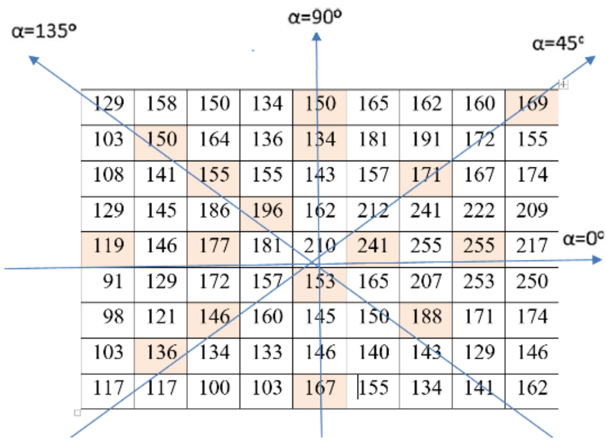

Fig. 6. Sampling pattern in window for rdLBP_α for α=0, α=45, α=90 and α=135

matching cost function along the radial windows, the sum of the absolute difference of intensity and the difference of the intensity gradient have been used.

Due to the success of radial windows in stereo matching and the benefits of binary stereo matching discussed, based on the combining of these two ideas, we propose using radial binary descriptor in dense stereo matching.

In [19], this idea has been proposed that certain texture patterns can exist along directional lines passing through a pixel, and that these types of texture patterns may not be present in a circular search as in the base LBP according to the Fig. (2). Therefore, a new descriptor based on called directional LBP in direction α ($dLBP_\alpha$) has been proposed. The sampling pattern in a 9x9 window for the values $\alpha = 0$, $\alpha = 45$, $\alpha = 90$, and $\alpha = 135$ are shown in the Fig. (4).

Bit i of $dLBP_\alpha$ descriptor is obtained from Eq. (6):

$$dLBP_\alpha(p.i)=S(f(p_c)-f(p_i)) \tag{6}$$

Step1: Compute $\mathrm{dLBP}_{\propto}(p_l)$

$for \ \propto = 0.45.90.135$

Step2: $\mathrm{dLBP}(p_l) = \mathrm{dLBP}_{135}(p_l)\mathrm{dLBP}_{90}(p_l)\mathrm{dLBP}_{45}(p_l)\mathrm{dLBP}_{0}(p_l)$

Step3: Compute $\mathrm{dLBP}_{\propto}(p_{r.d})$

$for \ \propto = 0.45.90.135 \ and \ d = d_{min}.\dots.d_{max}$

Step4: $\mathrm{dLBP}(p_{r.d}) =$

$\mathrm{dLBP}_{135}(p_{r.d})\mathrm{dLBP}_{90}(p_{r.d})\mathrm{dLBP}_{45}(p_{r.d})\mathrm{dLBP}_{0}(p_{r.d})$

$for \ \ d = d_{min}.\dots.d_{max}$

Step5: $d(p_l) = \min_d(Hamming(\mathrm{dLBP}(p_{r.d}).\mathrm{dLBP}(p_l)))$

Step6: *Applying* Median *Filtering*

Fig. 5. Pseudo code for proposed algorithm

Where the function in Eq. (6) is defined in Eq. (4) and the decimal value $dLBP_\alpha$ is obtained from Eq. (5).

For window in Fig. (4) according to Eq. (6) and the intensity values of pixel inside the window, binary string $dLBP_0 = 00001111$, $dLBP_{45} = 11110111$, $dLBP_{90} = 11111111$ and $dLBP_{135} = 11111111$ are obtained, and so decimal corresponding values are $dLBP_0 = 15$, $dLBP_{45} = 247$, $dLBP_{90} = 255$ and $dLBP_{135} = 255$. For example, with the data in Fig. (4) for , is calculated as follows:

$dLBP_0=S(P_7-PC)S(P_6-PC)S(P_5-PC)S(P_4-PC)S(P_3-PC)S(P_2-PC)S(P_1-PC)S(P_0-PC)=$

$=S(119-210)S(146-210)S(177-210)S(181-210)S(241-210)S(255-210)S(255-210)S(217-210)=00001111=15$

The big difference between $dLBP_0$ with other values $dLBP_{45}$. $dLBP_{90}$ and $dLBP_{135}$ shows that different texture patterns can exist in different directions.

Using dLBP descriptor on standard data bases Brodatz-1, Brodatz-2, Butterfly and Kylberg, lead to acceptable results.

Considering that radial multiple window method can lead to creating windows with disparity consistency within the window. On the other hand, the dLBP method leads to creating proper radial binary descriptors, using this descriptor is proposed for dense stereo matching. Pseudo code of the proposed algorithm is shown in Fig. (5).

Another proposed binary pattern to decrease the length of the final binary feature and make it independent of the window size is random directional Local Binary Pattern (rdLBP). In this method, instead of using all available pixels along each direction, only n pixels that are randomly selected in that direction are used.

For example, for the 9x9 window shown in Fig. (4), in-window sampling pattern for $n = 4$ has been shown in Fig. (6). For the window in Fig. (6), according to the Eq. (6) and the intensity values of the selected pixels inside the window, the binary string is obtained as $dLBP_0$=1100 ,$dLBP_{45}$=1111, $dLBP_{90}$=1111 and, $dLBP_{135}$= 1111 and therefore their decimal corresponding values are equal to $dLBP_0$=12, $dLBP_{45}$=15, $dLBP_{90}$=15 and $dLBP_{135}$=15.

Fig. 7. Left images and true disparity maps of Middlebury standard stereo images: (a) Tsukuba, (b) Teddy, (c) Venus, (d) Cones

Table 1. Comparison of error of proposed algorithm with other local binary algorithms

	Tsukuba	Teddy	Venus	Cones
LBP	10.64	27.4	13.44	32.28
BRIEF	9.08	22.94	11.45	25.23
CENSUS	7.17	23.93	8.4	27.33
dLBP	6.56	21.13	5.43	24.69

4- Experimental Results

In this paper, the proposed method is evaluated on Middlebury standard stereo images including four Tsukuba, Teddy, Venus, and Cones images, which are often used to compare different methods. The left images of this dataset is shown in Fig. (7). Each of these images has areas without texture, with repeated texture, ramps in different directions, many depth discontinuities and occlusions, which are considered challenging areas in the dense stereo matching problem.

To evaluate the results of the algorithms, assuming a true disparity map, a quantitative criterion of the percentage of false matched pixels based on Eq. (7) is used:

$$E_{\delta} = 100 \times \frac{1}{H \times W} \sum_{(x,y)} (|d(p_l) - d_{True}(p_l)| > \delta) \tag{7}$$

In Eq. (7), $d(p_l)$ is the disparity calculated in pixel p_l of the image and $d_{True}(p_l)$ is the value of the true spatial disparity according to the true disparity map, H is the length, and W is the width of the image. δ is the limit of error in pixels and is usually assumed to be one.

Table 2. Percentage improvement of the proposed algorithm compared to other local binary algorithms

	Tsukuba	Teddy	Venus	Cones
LBP	38.34	22.88	59.59	23.51
BRIEF	27.75	7.89	52.57	2.14
CENSUS	8.5	11.7	35.35	9.65

Table 1 compares the error of the proposed algorithm with other binary algorithms for the four standard images mentioned in non-occluded region and all of the image.

Considering the experimental results show that the proposed method is more effective than other local binary algorithms and reduces the error by an average of 20%. Table 2 shows the percentage of error reduction of the proposed algorithm compared to other algorithms.

Fig. (8) shows the final dense disparity map of the proposed algorithm on the for standard test images Tsukuba, Teddy, Venus and Cones.

5- Conclusion

In this paper, an advanced local binary pattern is used to extract the dense disparity map. This descriptor is based on the ideas of multiple window, radial or directional window, and local binary patterns. Experimental results show the performance of the proposed scheme. Since the proposed algorithm is based on binary and integer computations, it also

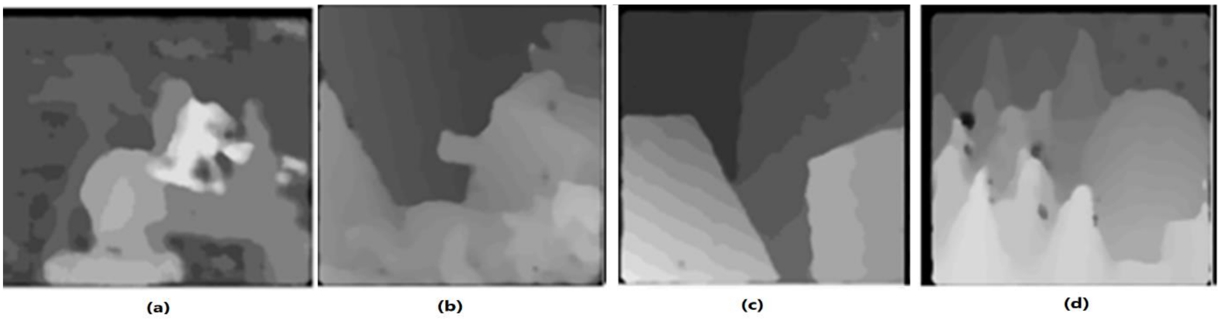

Fig. 8. Dense disparity maps obtained from proposed algorithm for images (a) Tsukuba, (b) Teddy, (c) Venus, (d) Cones

has a high computational efficiency. For further research we plan to integrate the proposed method in other complicated dense stereo matching with post processing to handle occlusions.

References

[1] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International journal of computer vision, 47(1) (2002) 7-42.

[2] A.M. Fotouhi, R. Abolghasem, Optimizing Disparity Candidates Space in Dense Stereo Matching, aut journal of modeling and simulation, 43(2) (2011) 43-51.

[3] T. Kanade, M. Okutomi, A stereo matching algorithm with an adaptive window: Theory and experiment, IEEE transactions on pattern analysis and machine intelligence, 16(9) (1994) 920-932.

[4] O. Veksler, Stereo correspondence with compact windows via minimum ratio cycle, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(12) (2002) 1654-1660.

[5] A. Fusiello, V. Roberto, E. Trucco, Efficient stereo with multiple windowing, in: Proceedings of IEEE Computer Society conference on computer vision and pattern recognition, IEEE, 1997, pp. 858-863.

[6] Y. Xu, D. Wang, T. Feng, H.-Y. Shum, Stereo computation using radial adaptive windows, in: 2002 International Conference on Pattern Recognition, IEEE, 2002, pp. 595-598.

[7] A.F. Bobick, S.S. Intille, Large occlusion stereo, International Journal of Computer Vision, 33(3) (1999) 181-200.

[8] S.B. Kang, R. Szeliski, J. Chai, Handling occlusions in dense multi-view stereo, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, IEEE, 2001, pp. I-I.

[9] H. Hirschmüller, P.R. Innocent, J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors, International Journal of Computer Vision, 47(1) (2002) 229-246.

[10] O. Veksler, Fast variable window for stereo correspondence using integral images, in: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., IEEE, 2003, pp. I-I.

[11] K.-J. Yoon, I.S. Kweon, Adaptive support-weight approach for correspondence search, IEEE transactions on pattern analysis and machine intelligence, 28(4) (2006) 650-656.

[12] A. Hosni, M. Bleyer, M. Gelautz, C. Rhemann, Local stereo matching using geodesic support weights, in: 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 2093-2096.

[13] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6) (2011) 765-781.

HOW TO CITE THIS ARTICLE

P.Bagheri, A. M. Fotouhi, Dense Stereo Matching Based on the Directional Local Binary Pattern, AUT J. Elec. Eng., 55(1) (2023) 125-130.

DOI: 10.22060/eej.2022.21687.5489