Stereo Image Matching Using Adaptive Morphological Correlation

Victor H. Diaz-Ramirez 1,*, Martin Gonzalez-Ruiz 1, Vitaly Kober 2,3 and Rigoberto Juarez-Salazar 4

1 Instituto Politécnico Nacional-CITEDI, Instituto Politécnico Nacional 1310, Tijuana 22310, BC, Mexico
2 Department of Computer Science, CICESE, Ensenada 22860, BC, Mexico
3 Department of Mathematics, Chelyabinsk State University, 454001 Chelyabinsk, Russia
4 CONACYT-Instituto Politécnico Nacional, CITEDI, Instituto Politécnico Nacional 1310, Tijuana 22310, BC, Mexico
* Correspondence: vdiazr@ipn.mx; Tel.: +52-664-6231344

Abstract: A stereo matching method based on adaptive morphological correlation is presented. The point correspondences of an input pair of stereo images are determined by matching locally adaptive image windows using the suggested morphological correlation that is optimal with respect to an introduced binary dissimilarity-to-matching ratio criterion. The proposed method is capable of determining the point correspondences in homogeneous image regions and at the edges of scene objects of input stereo images with high accuracy. Furthermore, unknown correspondences of occluded and not matched points in the scene can be successfully recovered using a simple proposed post-processing. The performance of the proposed method is exhaustively tested for stereo matching in terms of objective measures using known database images. In addition, the obtained results are discussed and compared with those of two similar state-of-the-art methods.

Keywords: stereo vision; disparity estimation; morphological correlation; locally adaptive image processing

1. Introduction

Stereo vision recovers three-dimensional (3-D) information about the observed scene by processing at least two images of the scene captured from different viewpoints. Stereo vision is widely used in high-impact technologies, such as robot navigation, autonomous vehicles, augmented reality and medical diagnosis, among others [1,2]. Stereo vision has many advantages over other existing 3-D technologies; for instance, simplicity and flexibility, high-rate performance, large field of view, and low cost. A fundamental task in stereo vision is disparity estimation. This task, also known as stereo matching, consists of determining the correspondence of all points in a pair of stereo images. The 3-D distribution of the scene can be retrieved from the disparity by triangulation [3].

Over the years, several approaches for stereo matching have been proposed. These approaches can be classified as local, global or hybrid [4,5]. In many applications, the local approach is preferable over the global and hybrid approaches because it is suitable for high-rate performance. In general, local methods estimate the disparity of each point of the scene by matching local windows centered at given corresponding points in each stereo image. Local methods usually perform the following steps: matching-cost computation, cost aggregation, disparity computation, post-processing and refinement [5–7]. The matching cost quantifies the similarity of two corresponding image points for a given disparity value. Commonly, the matching cost is computed by comparing the intensity values of two given image points. The cost aggregation reduces the uncertainty in the association of matching points. This step is usually carried out by matching adaptive windows [8,9] or adaptive weight support functions [10]. Disparity computation is performed by selecting the best
aggregation cost value for each corresponding point. Post-processing recovers the disparity of occluded image points. Finally, the refinement reduces estimation errors [11,12].

Within the state-of-the-art, several methods for matching-cost computation have been suggested [13]. The absolute difference (AD), squared difference (SD) and normalized cross-correlation (NCC) are widely known intensity-based matching measures [4]. Stereo matching based on the AD, SD or NCC is computationally efficient and possesses good tolerance to image noise. However, it tends to produce incorrect disparity estimates in image regions of low texture, nonstationary intensity or that are partially occluded [6]. Alternatively, matching-cost measures based on the relative order of pixel intensities have been considered [14]. The census transform (CT) is a non-parametric technique based on the local spatial structure [7,15]. The CT maps a given image point to a binary string. Each element of this string is true if the intensity of a given point is higher than that of a prespecified reference point; otherwise, it is false. Usually, the cost aggregation in CT-based methods is computed with the Hamming distance of two resultant binary strings. The CT is more accurate than intensity-based matching methods [6]. However, it is more sensitive to image noise [16].

Several variants of the CT have been suggested to improve the stereo matching accuracy and noise robustness. A simple approach consists of replacing the intensity value of the central element of the matching window (reference point) with the mean intensity value of their neighbor elements when computing the binary string [17]. Another approach is to compute the binary string from different pairs of image points within the matching window, excluding the central point [15,16]. Recently, the use of a weighting mask in the CT matching-cost computation has been suggested [18]. In addition, a trade-off between intensity-based AD and CT has been considered [19]. This approach, known as AD-Census, has good tolerance to image noise and accuracy of disparity estimation.

Although existing local methods for stereo matching have had great success, new alternatives still need to be explored to improve their performance. For instance, in the matching-cost and cost aggregation steps, it is desirable to obtain a low cost for image points with high similarity to those belonging to the object formed at the origin of the reference window and a high cost for the remaining points. To do this, we propose a robust method for stereo matching based on adaptive morphological correlation optimized with respect to a new criterion called binary dissimilarity-to-matching ratio (BDMR). First, locally adaptive windows constructed for a reference point and a potential corresponding point in the stereo image pair are preprocessed using binary threshold decomposition. Next, the morphological correlation is computed between the two preprocessed adaptive windows for different disparity values. Finally, a disparity estimate is obtained by finding the corresponding point coordinate of the maximum correlation. In addition, we propose a simple post-processing method to recover the disparity in occluded and not matched stereo image points.

The main contributions of this research are as follows. A binary dissimilarity-to-matching ratio (BDMR) is introduced. By minimizing the BDMR, a matching-cost measure based on adaptive morphological correlation is derived. A locally adaptive cost aggregation method for stereo matching based on morphological correlation is proposed. An efficient post-processing method for recovering the disparity in occluded points is proposed. This paper is organized as follows. Section 2 presents the proposed method for stereo image matching. Section 3 presents the results obtained with the proposed stereo matching method using images from the Middlebury stereo dataset [20–22]. These results are discussed and compared with those obtained with two recent existing similar methods. Finally, Section 4 presents our conclusions.

2. Stereo Matching with Adaptive Morphological Correlation

This section provides details of the proposed approach for stereo matching. First, we review the preliminaries of stereo vision. Secondly, we present the proposed method for image matching based on adaptive morphological correlation. Finally, we introduce the suggested approach for disparity post-processing.
2.1. Stereo Vision

Consider the stereo imaging system depicted in Figure 1. A pair of cameras project a point \( P \) in their corresponding image planes as the points \( p_1 \) and \( p_2 \), respectively. This setup assumes that the cameras are horizontally aligned, and the captured images \( I_1(x, y) \) and \( I_2(x, y) \) are rectified \([23,24]\). Thus, the points \( p_1 \) and \( p_2 \) can be located along the horizontal epipolar line, as shown in Figure 1. The location of the points \( p_1 \) and \( p_2 \) with coordinates \((x_1, y_1)\) and \((x_2, y_1)\), respectively, allows us to compute the disparity as

\[
\delta = |x_1 - x_2|.
\]  

(1)

The depth \( D \) to point \( P \) from the stereo baseline can be obtained as

\[
D = \frac{fB}{\delta},
\]

(2)

where \( f \) is the focal length of the camera lens and \( B \) is the distance between the optical camera centers. It should be noted that the parameters \( f \) and \( B \) are obtained by camera calibration, and the disparity \( \delta \) is determined by stereo matching.

![Figure 1. Description of a stereo imaging setup.](image)

2.2. Proposed Method for Stereo Matching

The block diagram of the proposed method is shown in Figure 2. The first step is the estimation of the disparity map from the input pair of rectified stereo images \( I_1(x, y) \) and \( I_2(x, y) \). Let \( w_1(x, y) \) and \( w_2(x, y) \) be two image windows, both of size \( N_w \times N_w \), obtained from \( I_1(x, y) \) and \( I_2(x, y) \) at the coordinates \((x_0, y_0)\), respectively. According to the theory of morphological image processing, the image window \( w_i(x, y) \) can be represented by the binary threshold decomposition in a given range as \([25–27]\)

\[
\tilde{w}_i(x, y) = \sum_{q=q_0}^{q_N} b_{iq}(x, y),
\]

(3)

where \( q_0 = \min\{w_i(x, y)\} \), \( q_N = \max\{w_i(x, y)\} \), and
is a binary image of $w_i(x, y)$ for the $q$-th intensity value. Note that if $q_0 = 1$ then $w_i(x, y) = \tilde{w}_i(x, y)$.

Figure 2. Block diagram of the proposed method for stereo image matching.

Now, assuming the horizontal epipolar constraint, we introduce the binary dissimilarity-to-matching ratio (BDMR) as follows:

$$BDMR(\tau) = \frac{D(\tau)}{M(\tau)} = \frac{\sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} [b_{1,q}(x, y) - b_{2,q}(x - \tau, y)]}{\sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} [b_{1,q}(x, y) + b_{2,q}(x - \tau, y) - 1]}$$

Evaluating Equation (5) can be rewritten as

$$BDMR(\tau) = \frac{\sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} [b_{1,q}(x, y) + b_{2,q}(x - \tau, y) - 2\text{MIN}\{b_{1,q}(x, y), b_{2,q}(x - \tau, y)\}]}{\sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} [1 + b_{1,q}(x, y) + b_{2,q}(x - \tau, y) - 2\text{MIN}\{b_{1,q}(x, y) + b_{2,q}(x - \tau, y), 1\}]}.$$ (6)

Note that the summation terms in Equation (6) can be calculated as

$$\mu_{b_1} = \frac{1}{QN^2} \sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} b_{1,q}(x, y),$$

$$\mu_{b_2}(\tau) = \frac{1}{QN^2} \sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} b_{2,q}(x - \tau, y),$$ (7)

where $Q = (q_N - q_0)$ is the number of quantization levels in the binary threshold decomposition. Moreover, by considering that

$$\text{MIN}\{b_{1,q}(x, y) + b_{2,q}(x, y), 1\} = \text{MAX}\{b_{1,q}(x, y), b_{2,q}(x, y)\},$$ (8)

Equation (6) can be rewritten as
\[ BDMR(\tau) = \frac{\mu_{b_1} + \mu_{b_2}(\tau) - \frac{2}{QN_w} \sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MIN}\{ b_{1,q}(x,y), b_{2,q}(x-\tau,y) \}}{1 + \mu_{b_1} + \mu_{b_2}(\tau) - \frac{2}{QN_w} \sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MAX}\{ b_{1,q}(x,y), b_{2,q}(x-\tau,y) \}}}. \] (9)

The minimum value of Equation (9) is obtained by maximizing

\[ C(\tau) = \frac{\sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MIN}\{ b_{1,q}(x,y), b_{2,q}(x-\tau,y) \}}{\sum_{q=q_0}^{q_N} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MAX}\{ b_{1,q}(x,y), b_{2,q}(x-\tau,y) \}}, \] (10)

where the term \(1/QN_w^2\) is added to the denominator to avoid singularities. Now, by interchanging the order of summations and considering [25,27]

\[ \sum_{q=q_0}^{q_N} \text{MIN}\{ b_{1,q}(x,y), b_{2,q}(x,y) \} = \text{MIN}\{ w_1(x,y), w_2(x,y) \}, \]
\[ \sum_{q=q_0}^{q_N} \text{MAX}\{ b_{1,q}(x,y), b_{2,q}(x,y) \} = \text{MAX}\{ w_1(x,y), w_2(x,y) \}. \] (11)

Equation (10) can be rewritten as

\[ C(\tau) = \frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MIN}\left\{ \sum_{q=q_0}^{q_N} b_{1,q}(x,y), \sum_{q=q_0}^{q_N} b_{2,q}(x-\tau,y) \right\}}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MAX}\left\{ \sum_{q=q_0}^{q_N} b_{1,q}(x,y), \sum_{q=q_0}^{q_N} b_{2,q}(x-\tau,y) \right\}}. \] (12)

Equation (12) is a nonlinear correlation that minimizes the BDMR when the maximum correlation value is reached. For the problem of stereo matching, the maximum value of Equation (12) should occur in the coordinate \(\tau = \delta\); that is, at the location where the sliding window \(w_2(x-\tau,y)\) matches the reference window \(w_1(x,y)\). To improve the accuracy and robustness of stereo matching using Equation (12), the values \(\{q_0, q_N\}\) can be chosen to properly describe the implicit object formed at the origin \((x_0, y_0)\) of the window \(w_i(x,y)\), identified as the target. Thus, the values \(\{q_0, q_N\}\) can be specified as

\[ q_0 = w_i(x_0, y_0) - \epsilon_v \sigma_{w_i}, \]
\[ q_N = w_i(x_0, y_0) + \epsilon_v \sigma_{w_i}, \] (13)

where \(\sigma_{w_i}\) is the standard deviation of \(w_i(x,y)\) with respect to \(w_i(x_0, y_0)\) and \(\epsilon_v\) is a dispersion parameter. Thus, Equation (12) can be adapted to each point of the pair of stereo images as

\[ C(\tau) = \frac{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MIN}\{ \tilde{w}_1(x,y), \tilde{w}_2(x-\tau,y) \}}{\sum_{x=1}^{N_x} \sum_{y=1}^{N_y} \text{MAX}\{ \tilde{w}_1(x,y), \tilde{w}_2(x-\tau,y) \}}, \] (14)

where
\[ \tilde{w}_1(x, y) = \sum_{q=1}^{Q} b_1(q_0 + q\Delta) (x, y), \]
\[ \tilde{w}_2(x - \tau, y) = \sum_{q=1}^{Q} b_2(q_0 + q\Delta) (x - \tau, y), \]

are preprocessed image windows of \( I_1(x, y) \) and \( I_2(x, y) \), respectively, using adaptive binary threshold decomposition, with a quantization step as
\[ \Delta_q = \frac{2\epsilon_v\sigma_{w1}}{Q}. \] (16)

To perform stereo matching using Equations (14)–(16), consider a reference point \( p_1 \) with coordinates \((x_0, y_0)\) in the image \( I_1(x, y) \). The corresponding point \( p_2 \) in image \( I_2(x, y) \) can be detected and located as depicted in the block diagram shown in Figure 3. First, a reference window \( w_1(x, y) \) with origin at the point \( p_1 \) and size of \( N_w \times N_w \) is constructed from \( I_1(x, y) \), where \( N_w = 2s + 1 \) and \( s \) are computed adaptively as
\[ s = (s_0 - 1) \exp \left( -\beta \frac{I_1(x_0, y_0)}{\sigma_{s0}^2} \right), \] (17)
where \( \beta \) is a scalar, \( s_0 \) is a prespecified parameter defining the maximum allowable window size and \( \sigma_{s0}^2 \) is the standard deviation of the intensity values of the points within the reference window with a maximum size of \((2s_0 + 1) \times (2s_0 + 1)\) with respect to \( p_1 \). Then, a sliding window \( w_2(x - \tau, y) : \tau \in [0, \delta_{\text{max}}] \), with a size of \( N_w \times N_w \) is constructed from \( I_2(x, y) \). Note that \( w_2(x - \tau, y) \) is shifted along the horizontal epipolar line of \( I_2(x, y) \). Afterward, \( w_1(x, y) \) and \( w_2(x - \tau, y) \) are preprocessed by binary threshold decomposition as described in Equation (15). Next, the adaptive morphological correlation is given in Equations (14)–(16) is computed for all values of \( \tau \). Finally, a disparity estimate is obtained as
\[ \delta(x_0, y_0) = \arg \max_{\tau} \{ C(\tau) \}. \] (18)

The disparity maps \( \delta_1(x, y) \) and \( \delta_2(x, y) \) can be obtained by applying the proposed method to all points of the stereo images \( I_1(x, y) \) and \( I_2(x, y) \).

**Figure 3.** Block diagram of the proposed method for stereo matching based on adaptive morphological correlation.
2.3. Disparity Post-Processing

The estimated disparity maps \( \delta_1(x, y) \) and \( \delta_2(x, y) \), can be verified as

\[
m_k(x, y) = \begin{cases} 
1, & \text{if } |\delta_k(x, y) - \delta_l(x - \delta_k(x, y), y)| \leq \epsilon_\delta, \\
0, & \text{otherwise},
\end{cases}
\]

(19)

where \( \epsilon_\delta \) is a tolerance parameter, \( k = \{1, 2\} \) and \( l = \begin{cases} 
k + 1, & \text{if } k = 1 \\
k - 1, & \text{if } k = 2.
\end{cases} \)

Note that a value of \( m_k(x, y) = 1 \) in Equation (19) indicates a verified estimated disparity, whereas a value of \( m_k(x, y) = 0 \) denotes an incorrectly estimated disparity caused by an occlusion or any other perturbation. Let \( T = \{(x_T, y_T) : m_i(x_T, y_T) = 1\} \) be the set of coordinates of all verified estimated disparities and \( F = \{(x_F, y_F) : m_i(x_F, y_F) = 0\} \) be the set of coordinates of all incorrectly estimated disparities. A desirable post-processing method requires replacing the incorrectly estimated disparity value \( \delta_i(x_F, y_F) \) with verified disparity values from the set \( \{\delta_i(x_T, y_T)\} \). In this context, we consider the prior probability that a verified estimated disparity at arbitrary coordinates \( (x, y) \) can replace the incorrect disparity at the coordinates \( (x_F, y_F) \), which is given by

\[
P(x, y) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left[ -\frac{(x - x_F)^2 + (y - y_F)^2}{2\sigma_1^2} \right] : (x, y) \in T,
\]

(20)

where a normal distribution with variance \( \sigma_1^2 \) is assumed. Furthermore, the probability density function that an image point with intensity value \( I_i(x, y) \) has a similar disparity as that expected at the coordinates \( (x_F, y_F) \), can be given by

\[
P(I_i(x, y)|\delta_i(x_F, y_F)) = \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left[ -\frac{(I_i(x, y) - I_i(x_F, y_F))^2}{2\sigma_2^2} \right] : (x, y) \in T,
\]

(21)

where \( \sigma_2^2 \) is the variance of the target’s intensity values. According to Bayesian theory, the posterior probability that an image point with disparity \( \delta(x, y) \) and intensity \( I_i(x, y) \) can replace unknown disparity \( \delta_i(x_F, y_F) \) given that \( I_i(x_F, y_F) \) is the intensity of \( I_i(x, y) \) at the coordinates \( (x_F, y_F) \) is given as

\[
P(\delta_i(x, y)|I_i(x_F, y_F)) = \frac{P(I_i(x, y)|\delta_i(x_F, y_F))P(x, y)}{P(I_i(x, y))},
\]

(22)

where \( P(I_i(x, y)) \) is the prior probability density function of the intensity of \( I_i(x, y) \). As a result, the coordinates \( (x_T, y_T) \in T \) of the disparity \( \delta(x_T, y_T) \) with the highest probability corresponds to the unknown disparity \( \delta(x_F, y_F) \), and can be obtained as

\[
(\hat{x}_T, \hat{y}_T) = \arg\max_{(x, y) \in T} \{P(\delta_i(x, y)|I_i(x_F, y_F))\}.
\]

(23)

By substituting Equations (20) and (21) into Equation (23), and by applying the logarithm function, we get

\[
(\hat{x}_T, \hat{y}_T) = \arg\min_{(x, y) \in T} \left\{ \frac{(x - x_F)^2 + (y - y_F)^2}{2\sigma_1^2} + \frac{(I_i(x, y) - I_i(x_F, y_F))^2}{2\sigma_2^2} \right\},
\]

(24)

where \( \delta_i(\hat{x}_T, \hat{y}_T) \) is an estimate of the incorrect disparity \( \delta_i(x_F, y_F) \). Thus, by applying the estimator given in Equation (24) to all elements of the set \( F \), one can obtain the improved post-processed disparity maps \( \delta_1(x, y) \) and \( \delta_2(x, y) \).
3. Results

This section presents the results obtained with the proposed approach for stereo matching using images from the Middlebury stereo dataset [20–22]. The results are discussed and compared with those obtained with two recent variants of the CT, namely, the improved weighted census transform (IWCT) [18] and the improved AD-Census (AD-C) algorithm [19]. The accuracy of disparity estimation by the proposed, IWCT and AD-C methods is quantified in terms of the bad-matched pixels (BMP) and root mean squared (RMS) error between estimated and ground truth disparities. For the BMP measure, we set the tolerance of $\epsilon_d = 2$. First, we quantify the performance of the proposed and considered methods for disparity estimation of non-occluded regions in input stereo images. Next, we evaluate the performance of the suggested disparity post-processing method. Additionally, we show refined disparity maps obtained with the proposed approach using a generic refinement method. Finally, we present the statistical performance results of the proposed and considered methods for stereo matching with twenty-five images from the Middlebury stereo dataset.

The proposed method, IWCT and AD-C were implemented using the Python 3.10.7 language on a personal computer with an Intel Core I5 2.4 GHz processor, 16 GB of RAM and Linux Ubuntu 20.04 operating system. Figure 4a shows the right image from eight different stereo image pairs from the dataset. The window size for all tested methods is $N_w \times N_w$, where $N_w = 2s_0 + 1$ and $s_0 = 6$. For the proposed method we set $Q = 31$, $\epsilon_v = 1.5$ and $\beta = 2.5s_0$. Figure 4b shows the ground truth disparities of non-occluded regions of the input images shown in Figure 4a. The non-occluded regions are obtained by applying the verification method given in Equation (19) to the ground truth disparities provided by the dataset. The estimated disparity maps obtained with the IWCT, AD-C and proposed method are presented in Figure 4c–e, respectively. Notice that the proposed method produces the lowest values of BMP and RMS measures in all cases compared to those obtained with the IWCT and AD-C methods. The proposed method is able to estimate the disparity in homogeneous regions with high accuracy. This feature is obtained when the specified number $Q$ of quantification values for the binary threshold decomposition is sufficiently large ($Q > 8$). Furthermore, it can be seen that the proposed method is also able to correctly estimate the disparity at the edges of the objects in the scene. This feature is due to the dynamic adaptation of the sliding windows employed for point matching given in Equation (17). On the other hand, the IWCT method produces the worst results of all tested methods. This approach yields many incorrectly estimated disparity values in homogeneous image regions. Note that the test images shown in Figure 4a present several challenges for stereo matching, such as image regions with little texture, partial occlusions, nonstationary intensity changes, objects with sharp edges and abrupt disparity variations. According to the obtained results shown in Figure 4c–e, the proposed method adapts better to challenging situations than the other tested methods. However, the lack of texture in image regions larger than the search space of the algorithm causes the matching method to be unable to determine the point correspondences. For instance, this can be seen in the central area of the Recycle image. The AD-C algorithm yields good results in the majority of the performed tests. This algorithm can estimate the disparity values at the edges of the scene objects very well. However, its performance is lower than that of the proposed method.
Afterward, we evaluated the performance of the post-processing method described in Section 2.3. We applied the suggested post-processing to the estimated disparity maps obtained with the IWCT, AD-C and proposed method, see Figure 4c–e. The resultant post-processed disparity maps are presented in Figure 5b–d. It can be seen that the suggested post-processing is successful in retrieving the unknown disparity values in occluded regions of the input stereo images. Furthermore, it can also retrieve several incorrectly estimated disparity values in homogeneous image regions, which were not verified by Equation (19). Additionally, Figure 5b–d presents the BMP and RMS values of all tested methods between the estimated and ground truth disparities shown in Figure 5a. The IWCT and AD-C methods yield higher BMP and RMS values in comparison with those obtained with the proposed method. The AD-C algorithm yields slightly better performance than the IWCT. However, the post-processed disparity maps obtained with the proposed method yielding the best results of all the tested methods. It is worth mentioning that the large occluded regions on the right side of the estimated disparity maps shown in Figure 4c–e were correctly recovered by the suggested post-processing in all tested methods. However, note that the post-processing was unable to recover the disparity values in the wood knot shown in the Wood2 image. This is because the verified disparity values in the vicinity of this region are associated with image points with intensity values that are significantly different from those of the wood knot.

Figure 4. Disparity estimation results for non-occluded image regions. (a) Right input stereo image. (b) Ground truth disparity map of non-occluded image regions. Estimated disparity maps of non-occluded image regions, obtained with: (c) IWCT, (d) AD-C and (e) Proposed method.
Figure 5. Results of disparity post-processing using the suggested method. (a) Ground truth disparity maps. Post-processed disparity maps with the suggested method, obtained with: (b) IWCT, (c) AD-C and (d) Proposed method.

Now, the post-processed disparity maps shown in Figure 5b–d, were refined by applying the well-known weighted least-squares filter [28]. The refined disparity maps are shown in Figure 6. Note that the refinement significantly reduces anomalous disparity errors for all tested methods. The refined disparity maps using the proposed adaptive morphological correlation approach produce the best results of all tested methods in terms of the BMP and RMS measures. Furthermore, we see that the refined disparity maps using the IWCT and AD-C methods of images, such as Adirondack, Recycle and Rocks1, contain very noticeable artifacts, while the refined disparity maps using the proposed approach contain fewer artifacts. This result is expected because any refinement method performs better when the input disparity map contains fewer incorrect disparity estimates, such as those obtained with the proposed approach.

Finally, we compare the statistical performance of the proposed method, IWCT and AD-C, in terms of both BMP and RMS measures. In this experiment, we estimated the disparity map of twenty-five different stereo images from the Middlebury stereo dataset using each of the considered stereo matching methods. The mean value and standard deviation of the BMP and RMS measurements were computed for each tested method. The results are presented in Figure 7 and Table 1. Figure 7a shows the statistical results for all tested stereo matching methods considering only the non-occluded regions of the input stereo images. Note that the proposed approach yields the best results of all tested methods. In contrast, the IWCT produces the worst results. This low performance is because the IWCT produces many wrong disparity estimates in homogeneous image regions, as shown in Figure 4c. The AD-C algorithm yields good statistical results in general terms. The AD-C approach produces fewer incorrect disparity estimates in homogeneous image regions than the IWCT. Additionally, it correctly estimates the disparity values at the edges of the objects present in the scene. However, the performance of both AD-C and IWCT methods is lower than that of the proposed approach.
Figure 6. Refined disparity maps after applying the suggested post-processing method, obtained with: (a) IWCT, (b) AD-C and (c) Proposed method.

Figure 7b presents the statistical results of the post-processed disparity maps using the suggested approach. It should be noted that the reference for computing the BMP and RMS measures consists of the ground truth disparities provided by the dataset, see Figure 5a. Note that the proposed approach yields the best results, whereas IWCT yields the worst results. The AD-C algorithm produces acceptable results in general terms. The results shown in Figure 7a and Table 1 confirm that the proposed method based on adaptive morphological correlation is effective and robust for stereo image matching. Additionally, the results presented in Figure 7b and Table 1 indicate that the suggested post-processing method is successful in retrieving the disparity values in occluded image regions.

Figure 7. Statistical performance results in terms of BMP and RMS measurements in twenty-five stereo images of the evaluated methods. (a) Disparity estimation results in non-occluded image regions. (b) Statistical performance of the suggested method of disparity post-processing.
Table 1. Statistical results in terms of BMP and RMS for stereo matching of non-occluded points and evaluation of the proposed post-processing method.

| Stereo Matching Non-Occluded Points | Proposed Post-Processing |
|-------------------------------------|--------------------------|
| Method | BMP | RMS | BMP | RMS | BMP | RMS |
| IWCT | 10.69 | 4.41 | 8.44 | 2.51 | 14.46 | 6.07 | 10.99 | 3.57 |
| AD-C | 6.65 | 2.33 | 6.10 | 1.69 | 10.11 | 3.28 | 8.77 | 2.45 |
| Proposed | 3.42 | 2.20 | 4.11 | 1.62 | 6.15 | 3.28 | 7.29 | 2.87 |

4. Conclusions

An accurate and robust method for stereo image matching based on adaptive morphological correlation was presented. The correspondence of non-occluded points in a pair of rectified stereo images was accurately determined by matching locally adaptive image windows using the suggested morphological correlation operation, which is optimal with respect to the new, introduced criterion called binary-to-dissimilarity ratio. In addition, a simple disparity post-processing method for recovering point correspondences of occluded points was suggested. The performance of the proposed method for stereo matching was exhaustively tested in terms of the mean absolute error and peak signal-to-noise ratio objective measures using images of the well-known Middlebury stereo dataset. The obtained results were discussed and compared with two recent state-of-the-art methods based on the census transform. According to the performed experiments and obtained results, the proposed method for stereo matching outperformed the existing tested methods in terms of the considered performance measures. Additionally, the obtained results confirmed that the suggested post-processing method allowed the disparity values of partially occluded image points to be successfully recovered.

Author Contributions: V.H.D.-R.: Conceptualization, Methodology, Software, Visualization, Writing—Original draft preparation, Funding acquisition. M.G.-R.: Investigation, Data Curation, Writing—Reviewing and Editing. V.K.: Formal analysis, Investigation, Writing—Reviewing and Editing. R.J.-S.: Formal analysis, Writing—Reviewing and Editing, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Consejo Nacional de Ciencia y Tecnología (CONACYT) (Basic Science and/or Frontier Science 320890, Basic Science A1-S-28112, Cátedras CONACYT 880), and by Instituto Politécnico Nacional (SIP-20221288).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: vision.middlebury.edu (accessed on 14 October 2022).

Acknowledgments: V. Kober thanks the Russian Science Foundation, grant No. 22-19-20071.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Fan, D.; Liu, Y.; Chen, X.; Meng, F.; Liu, X.; Ullah, Z.; Cheng, W.; Liu, Y.; Huang, Q. Eye Gaze Based 3D Triangulation for Robotic Bionic Eyes. *Sensors* 2020, 20, 5271. [CrossRef]
2. Brown, N.E.; Rojas, J.F.; Goberville, N.A.; Alzubi, H.; AlRousan, Q.; Wang, C.; Huff, S.; Rios-Torres, J.; Ekti, A.R.; LaClair, T.J.; et al. Development of an energy efficient and cost effective autonomous vehicle research platform. *Sensors* 2022, 22, 5999. [CrossRef] [PubMed]
3. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2nd ed.; Cambridge University Press: Cambridge, MA, USA, 2003.
4. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* 2002, 47, 7–42. [CrossRef]
Sensors 2022, 22, 9050

5. Hamzah, R.A.; Ibrahim, H. Literature survey on stereo vision disparity map algorithms. J. Sensors 2016, 2016, 8742920. [CrossRef]
6. Hirschmuller, H.; Scharstein, D. Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans. Pattern Anal. Mach. Intell. 2009, 31, 1582–1599. [CrossRef]
7. Banks, J.; Corke, P. Quantitative evaluation of matching methods and validity measures for stereo vision. Int. J. Robot. Res. 2001, 20, 512–532. [CrossRef]
8. Adhyapak, S.; Kehtarnavaz, N.; Nadin, M. Stereo matching via selective multiple windows. J. Electron. Imaging 2007, 16, 013012. [CrossRef]
9. Fusiello, A.; Roberto, V.; Trucco, E. Symmetric stereo with multiple windowing. Int. J. Pattern Recognit. Artif. Intell. 2001, 14, 1053–1066. [CrossRef]
10. Yoon, K.J.; Kweon, I.S. Adaptive support-weight approach for correspondence search. IEEE Trans. Pattern Anal. Mach. Intell. 2006, 28, 650–666. [CrossRef]
11. Zhan, Y.; Gu, Y.; Huang, K.; Zhang, C.; Hu, K. Accurate Image-Guided Stereo Matching with Efficient Matching Cost and Disparity Refinement. IEEE Trans. Circuits Syst. Video Technol. 2016, 26, 1632–1645. [CrossRef]
12. Jiao, J.; Wang, R.; Wang, W.; Dong, S.; Wang, Z.; Gao, W. Local stereo matching with improved matching cost and disparity refinement. IEEE Multimed. 2014, 21, 16–27. [CrossRef]
13. Ma, J.; Jiang, X.; Fan, A.; Jiang, J.; Yan, J. Image matching from handcrafted to deep features: A survey. Int. J. Comput. Vis. 2021, 129, 23–79. [CrossRef]
14. Zabih, R.; Woodfill, J. Non-Parametric Local Transforms for Computing Visual Correspondence. In Proceedings of the Third European Conference-Volume II on Computer Vision-Volume II; Springer: Berlin/Heidelberg, Germany, 1994; pp. 151–158.
15. Fife, W.S.; Archibald, J.K. Improved census transforms for resource-optimized stereo vision. IEEE Trans. Circuits Syst. Video Technol. 2000, 12, 16–22. [CrossRef]
16. Lee, J.; Jun, D.; Eem, C.; Hong, H. Improved census transform for noise robust stereo matching. Opt. Eng. 2016, 55. [CrossRef]
17. Chen, S.; Wu, Y.; Zhang, Y. S-census transform algorithm with variable cost. Comput. Eng. Des. 2018, 39, 414–419.
18. Hou, Y.; Liu, C.; An, B.; Liu, Y. Stereo matching algorithm based on improved census transform and texture filtering. Optik 2022, 249, 168186. [CrossRef]
19. Wang, Y.; Gu, M.; Zhu, Y.; Chen, G.; Xu, Z.; Guo, Y. Improvement of AD-Census Algorithm Based on Stereo Vision. Sensors 2022, 22, 6933. [CrossRef]
20. Scharstein, D.; Pal, C. Learning conditional random fields for stereo. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
21. Hirschmuller, H.; Scharstein, D. Evaluation of cost functions for stereo matching. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
22. Scharstein, D.; Hirschmüller, H.; Kitajima, Y.; Krathwohl, G.; Nešić, N.; Wang, X.; Westling, P. High-resolution stereo datasets with subpixel-accurate ground truth. In Proceedings of the German Conference on Pattern Recognition; Springer: Berlin/Heidelberg, Germany, 2014; pp. 31–42.
23. Juarez-Salazar, R.; Rios-Orellana, O.I.; Diaz-Ramirez, V.H. Stereo-phase rectification for metric profilometry with two calibrated cameras and one uncalibrated projector. Appl. Opt. 2022, 61, 6097–6109. [CrossRef]
24. Fusiello, A.; Trucco, E.; Verri, A. A compact algorithm for rectification of stereo pairs. Mach. Vis. Appl. 2000, 12, 16–22. [CrossRef]
25. Maragos, P. Optimal morphological approaches to image matching and object detection. In Proceedings of the 1988 Second International Conference on Computer Vision, Computer Society, Tampa, FL, USA, 5–8 December 1988; pp. 695–696.
26. Martinez-Diaz, S.; Kober, V.I. Nonlinear synthetic discriminant function filters for illumination-invariant pattern recognition. Opt. Eng. 2008, 47, 067201. [CrossRef]
27. Garcia-Martinez, P.; Ferreira, C.; Garcia, J.;Arsenault, H.H. Nonlinear rotation-invariant pattern recognition by use of the optical morphological correlation. Appl. Opt. 2000, 39, 776–781. [CrossRef] [PubMed]
28. Min, D.; Choi, S.; Lu, J.; Ham, B.; Sohn, K.; Do, M.N. Fast global image smoothing based on weighted least squares. IEEE Trans. Image Process. 2014, 23, 5638–5653. [CrossRef] [PubMed]