Genetic Algorithm for Stereo Correspondence with a Novel Fitness Function and Occlusion Handling

Alvaro Arranz[1], Manuel Alvar[1], Jaime Boal[1], Alvaro Sanchez-Miralles[1] and Arturo de la Escalera[2]

[1]*Insitute for Research in Technology (IIT), ICAI School of Engineering, C/ Alberto Aguilera 23, 28015 Madrid, Spain*

[2]*Intelligent Systems Lab, University Carlos III of Madrid, C/ Butarque 15, 28911 Leganes, Madrid, Spain*

Keywords: Stereo Reconstruction, Genetic Algorithm.

Abstract: This paper proposes a genetic algorithm for solving the stereo correspondence problem. Applied to stereo, genetic algorithms are flexible in the cost function and permit global reasoning. The main contribution of this paper is a new crossover and a mutation operator which accounts for occlusion management and a new fitness function which considers occluded pixels and photometric derivatives. Both left and right disparity images are analysed in order to classify occluded pixels correctly. The proposed fitness function is compared to the traditional energy function based in the framework of the Markov Random Fields. The results show that a 32% bad-pixel error reduction can be achieved on average using the proposed fitness function. The results have been uploaded to the Middlebury ranking webpage, as the first evolutionary algorithm evaluated.

1 INTRODUCTION

Passive stereo has received a huge amount of attention from the research community over the past two decades. The first algorithms that dealt with the stereo correspondence were sparse-feature based algorithms. Considering that some applications would find a per-pixel estimation of the disparity for the reference image more useful, dense stereo algorithms started to show their value. New algorithms such as global methods, dynamic programming, multi-resolution, or cooperative algorithms soon appeared to deal with dense disparity estimation. A taxonomy and evaluation of the most important dense stereo algorithms was proposed in (Alahari et al., 2010). This paper presented a methodology for comparing different stereo algorithms incorporating a ranking system (Middlebury,).

Results shown in (Alahari et al., 2010) suggest that global methods are the most accurate ones while local methods are ideal for real-time applications due to its simplicity and their parallelizable nature. However, new algorithms that outperform them, such as (Mei et al., 2011), have been published. Generally speaking, the best algorithms in the Middlebury ranking use some kind of optimization process for global reasoning followed by a refinement process for outliers and occluded pixels.

In this paper, a stereo algorithm using a genetic optimization approach is proposed. The main contributions are the occlusion handling procedure that is included as a part of the genetic algorithm and the proposed fitness function that is demonstrated to improve the number of bad pixels in the solution.

This document is organized as follows. In 2, previous genetic algorithms in stereo are briefly reviewed. In 3 detailed information about the genetic algorithm proposed is given. 4 puts forward the experiments carried out as well as a comparison between algorithms. Finally in 5 the main conclusions are drawn.

2 GENETIC ALGORITHMS IN STEREO

Genetic algorithms are a class of evolutionary algorithms that have been widely used as an heuristic search for optimization problems in a many different applications. In (Saito and Mori, 1995) a genetic algorithm is used to combine solutions of window-based methods with different window sizes while favouring photo-consistency and smoothness. In order to reduce the size of the problem, it proposed also to divide the solution into blocks and find optimal disparity maps block-by-block. The authors in (Han et al., 2001) use a region extraction algo-

Arranz A., Alvar M., Boal J., Sanchez-Miralles A. and de la Escalera A..
Genetic Algorithm for Stereo Correspondence with a Novel Fitness Function and Occlusion Handling.
DOI: 10.5220/0004291202940299
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 294-299
ISBN: 978-989-8565-48-8

rithm for dividing the image. Their fitness function is made of an intensity similarity and a smoothing term between regions. A multi-resolution approach is proposed in (Gong and Yang, 2001) and (Gong and Yang, 2002), where a quad-tree structure is used for representing each individual. A Markov Random Field (MRF) based fitness function for global reasoning is used. In (Wang et al., 2003) it is proposed to use the whole disparity map as a representation of genomes with no mutation operation. Recently, (Dai et al., 2008) use an adaptive crossover and mutation while their fitness function do not include any smoothing term. Finally, (Zhang et al., 2009) use a pyramidal propagation stratagem for solution representation and (Nie et al., 2009) implement a stereo correspondence genetic algorithm in GPU for performance enhancement. Genetic algorithms have also been used for matching sparse features, for instance in (Issa et al., 2002) a genetic algorithm is employed to match edges.

The utilization of genetic algorithms in stereo correspondence has some advantages over other traditional methods. Firstly, genetic algorithms may optimize an energy function for some global reasoning. In this sense it resembles to global methods based on MRF such as graph-cuts (Kolmogorov and Zabin, 2004). It has the advantage of its flexibility, given that practically any fitness function can be used, although getting close to the optimum is not guaranteed. Secondly, unlike most global or local methods, genetic algorithms can provide various local solutions during the optimization process.

Our approach uses some of the ideas found in the literature and proposes new crossover and mutation operators. One of the main contributions of this paper is that a method for occlusion handling is included natively in the genetic algorithm, not only as a refinement process. The disparity estimation is performed on both right and left images. This permits to calculate an occlusion map for both images, and treat occluded pixels and the common ones differently. The new crossover proposed permits to make a combination of a large number of pixels at the same time while favouring the child to inherit the best regions of both parents. The new mutation operator radically changes the disparity values of some regions to enable large jumps in the solution space. Finally, another contribution is the fitness function proposed, that really achieves to optimize correctly the number of bad pixels in the image considering occlusions.

As mentioned before, (Middlebury,) has become one of the main resources for evaluation and comparison of stereo correspondence algorithms. Any of the genetic stereo correspondence methods previously cited neither used the standard set of stereo images, nor compared their results with the state-of-the-art stereo algorithms. In this paper a comparison between several stereo methods is shown and results have been uploaded to the ranking system in (Middlebury,) for future comparison.

3 PROPOSED ALGORITHM

Generally, stereo correspondence has been demonstrated to be an ill-posed, NP-hard problem. Moreover, considering a common size image of 400 by 300 pixels and sixty different disparity labels, the number of different possible solutions is completely overwhelming. A naive implementation of a genetic algorithm, with highly random disparity assignments to each pixel, does not perform correctly due to the fact that almost any random disparity image does not even make sense as an image. Hence, due to the huge search space involved in stereo correspondence, properly guiding the genetic algorithm towards feasible disparity maps is fundamental to make the algorithm computationally tractable.

In the following subsections the genetic stereo correspondence algorithm proposed in this paper is explained in more detail.

3.1 Genome Representation

The most remarkable genome representations used in the literature are the quad-tree and the disparity map representation. Given that, in the method herein proposed, no multi-resolution is used and the disparity map representation makes it easier to compute the fitness function, the whole disparity map representation has been chosen. An important novelty is to include both left and right disparity images in the genome representation.

$$\bar{g}\begin{cases}\bar{g_L} = \{X_{1_L}, X_{2_L}, ..., X_{N_L}\} \\ \bar{g_R} = \{X_{1_R}, X_{2_R}, ..., X_{N_R}\}\end{cases}, \quad X_i \in \{L_1, L_2, ..., L_k\} \tag{1}$$

where g is the genome, g_L and g_R are the representation of the left and right disparity images respectively, X_{i_L} and X_{i_R} are the disparities estimated for pixel i on the left and right disparity images respectively, N the total number of pixels in each image and L_i the set of labels representing the set of disparities analysed.

3.2 Initialization

Some algorithms in the literature use random sampling for their initialization process. The probability of each pixel having a certain disparity value is based on a photo-consistency measurement. Others use a solution of other local window-based algorithm with a random window size. This approach is similar to the one proposed in (Saito and Mori, 1995) with the main difference that the disparity range is not restricted to the range obtained by the local methods.

For the initialization process we have used two different window-based algorithms with different window sizes, the adaptive support-weight approach (Yoon and Kweon, 2006) with random parameters and the census based with window-cost aggregation. For the census transform, a constant window size of 9x7 was used as suggested in (Mei et al., 2011). During the initialization process, each pixel is sampled with a probability proportional to the times it has appeared in the local window-based algorithms. It is important to notice that either algorithm alone performs well in discontinuities, occluded or untextured areas.

3.3 Fitness Function

As the stereo correspondence problem can be formulated as a MRF, the fitness function will be assigned the related energy value of the left disparity image. A classical formulation is given by the following equations

$$E_{classic}\left(\bar{g}\right) = E_{data-classic}\left(\bar{g_L}\right) + E_{smooth-classic}\left(\bar{g_L}\right) \tag{2}$$

$$E_{data-classic}\left(\bar{g_L}\right) = \sum_{i \in \bar{g_L}} \left|I_L(x_i, y_i) - I_R(x_i - X_i, y_i))\right| \tag{3}$$

$$E_{smooth-classic}\left(\bar{g_L}\right) = \sum_{\{p,q\} \in N} V_{\{p,q\}}(X_p, X_q) \tag{4}$$

where g is a certain individual, g_L is the left disparity image, I_l and I_r stand for the left and right stereo pair, x_i and y_i are the image coordinates of pixel i, and $V_{\{p,q\}}$ is a smoothing function favouring the neighbouring pixels having the same disparity.

This energy configuration has been widely used in the literature by global stereo algorithms. It is simple and it can be optimized by heuristic algorithms such as graph-cuts (Kolmogorov and Zabin, 2004). As stated in (Kolmogorov and Zabin, 2004), not any function can be used for this purpose.

The classic energy function might present some problems in the disparity estimation along the discontinuities and the occluded areas. Due to the flexible nature of the genetic algorithms, other more complicated functions can be chosen as fitness functions. Herein, an energy function that considers discontinuities and occlusions for the energy evaluation is proposed:

$$E\left(\bar{g}\right) = E_{data}\left(\bar{g_L}\right) + E_{smooth}\left(\bar{g_L}\right) \tag{5}$$

$$E_{data}\left(\bar{g_L}\right) = \begin{cases} \lambda_d & \text{if } i \text{ is occluded} \\ \sum_{i \in \bar{g_L}} \left|I_L(x_i, y_i) - I_R(x_i - X_i, y_i))\right| & \text{otherwise} \end{cases} \tag{6}$$

$$E_{smooth}\left(\bar{g_L}\right) = \sum_{\{p,q\} \in N} \frac{\beta_s}{\varphi_s} |X_p - X_q| \tag{7}$$

$$\beta_s = max(\lambda_s, \gamma_s - |I_L(p) - I_L(q)|) \tag{8}$$

where λ_s, γ_s and φ_s are constant parameters for every pixel.

The main modification in the E_{data} term is that occluded pixels, as they do not have a correspondent pixel on the right image, contribute to the energy with a constant value λ_d. This enables low energy configurations where the occluded pixels are matched correctly.

This new smooth function establishes a relation between the colour consistency of the neighbours and the associated weight of their disparity difference. For neighbouring pixels that are very different in colour, low weight is assigned to their disparity difference, while neighbouring pixels that are very similar are forced to have the same disparity.

At this point it is very important to emphasize that in any case the genetic algorithm, neither using the classic energy function nor the proposed one, is guaranteed to find the optimum energy configuration. Even more, how close we can get to the optimum will depend on the fitness function, the crossover and mutator operators, their parameters, the population size, etc. Probably, in any case the genetic algorithm will get close to the optimum, but the experiments carried out and shown in 4, suggest that the proposed energy function is more adequate than the classic one for both discontinuity and occlusion management.

3.4 Crossover

On first place, our method employed as a crossover algorithm that is very similar to the uniform crossover. For each crossover, a random block size is selected representing a region on each disparity image g_L and g_R. Then, a random assignation of each parent block to the children is performed.

While this stochastic approach to the crossover operation is inherent to the genetic algorithms, some tests with a deterministic crossover were also carried out. A new crossover was defined, instead of assigning the blocks to the sons randomly, first it evaluates the fitness function on each parent block and then put the blocks with the best fitness function on the same son. In this sense, this approach contradicts the stochastic nature of the genetic algorithm and might involve getting stuck in local minima. However, after testing both approaches, the deterministic crossover achieved a lot better fitness function than the stochastic one, so this one was used on our final tests. This is also suggested in (Wang et al., 2003).

3.5 Mutation

Three different mutation operations that may occur to each individual have been defined. Firstly, one possible mutation operation is to initialize again some pixels of one of the left or right images following the steps explained in 3.2. That is, the disparity of the pixel is changed stochastically with a probability proportional to those suggested by local methods. This mutation operation may happen with a probability P_{Ma}.

Secondly, a median filter operation with a random window size is also performed as a mutation function. It is not any novelty, but sometimes it is effective for managing some sparse outliers. This median filter operation is performed with a probability P_{Mb}.

Finally, an occlusion detection and handling is also included as a mutation with probability P_{Mc}. This process is a two step operation: an occlusion detection followed by an occlusion management. Given that both left and right disparity images are being estimated by our algorithm, we can use the right image disparities to estimate which pixels cannot have possible matches on the left one and vice-versa. The following operations are defined for calculating the left occlusion map:

$$O_L(p) = \begin{cases} 0 & \exists i / \begin{pmatrix} x(i) + \bar{g_R}(i) \\ y(i) \end{pmatrix} = \begin{pmatrix} x(p) \\ y(p) \end{pmatrix} \quad p, i \in P \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

being O_L the left occlusion map, $x(p)$ and $y(p)$ the x and y coordinates of point p respectively and P the set of disparity image points.

Similarly, an expression for the right occlusion map for the right image is:

$$O_R(p) = \begin{cases} 0 & \exists i / \begin{pmatrix} x(i) - \bar{g_L}(i) \\ y(i) \end{pmatrix} = \begin{pmatrix} x(p) \\ y(p) \end{pmatrix} \quad p, i \in P \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

being O_R the right occlusion map.

Once the occlusion maps are calculated for both images, a very simple occlusion management is performed. We follow an iterative process based on the neighbouring disparities of the occluded pixels. For the left image, each occluded pixel is assigned the disparity value of the most photo-consistent non-occluded neighbour from left to right and afterwards it is marked as non-occluded. If no non-occluded neighbours exist, it maintains its occluded status for the next iteration. Special status have the occluded pixels whose $x(p)$ coordinate is less than the number of disparities analysed. In this case the iteration is made from right to left and bottom-up. The iteration is finished when no occluded pixels are left on the left occluded map.

For the right image it is similarly done but vice versa (right to left for common pixels and left to right for pixels whose $x(p)$ is at a distance of the number of disparities analysed from the right image border). This fast an simple algorithm demonstrates to be effective in 4.

4 EXPERIMENTAL RESULTS

The genetic algorithm proposed has been applied to solve the Middlebury standard stereo dataset in (Middlebury,) that consists of four images. The parameters used related with the new energy function proposed are shown in 1, while the parameters related with the genetic algorithm are shown in 2. For local methods, window sizes between 3 and 45 have been used and random values for the adaptive-weight parameters. All the test-cases were run using the same parameters.

The resulting left disparity images with their bad pixels percentage image representation are shown in 1 and 2. Looking to the bad-pixel images, it is clear that Tsukuba and Venus obtain the best results. Although the algorithm performs quite well all along non-occluded and discontinuity regions, in other areas such as untextured regions it fails substantially. This can be attributed to the fact that the local algorithms used in the initialization process also fail in these untextured regions, so the genetic algorithm is unable to generate individuals with proper disparities on that region.

All four images were uploaded and evaluated using the Middlebury web-site. The algorithm achieved an average rank of 38.5 and an average percent of bad pixels of 5.81. This is an improvement over, for example, the adaptive-weight algorithm used for its initialization step which has an average rank of 61.4 and an average percent of bad pixels of 6.67. Moreover, the proposed genetic algorithm achieved the best rank in the discontinuity areas of the Tsukuba image. Com-

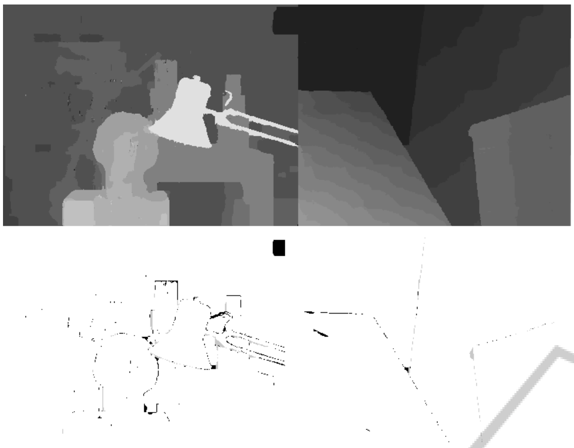

Figure 1: Tsukuba and Venus results. Disparity images (first row) and bad pixels (second row)

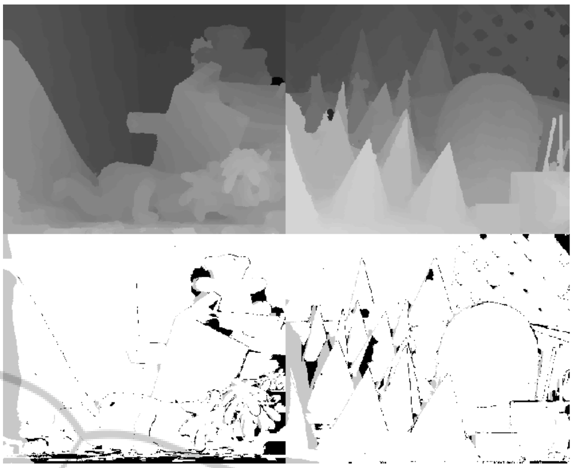

Figure 2: Teddy and Cones results. Disparity images (first row) and bad pixels (second row).

paring the proposed algorithm to the best reported one (Mei et al., 2011), our algorithm performs 1.84% worse. This can be attributed mainly to the untextured regions already explained where the local methods fail considerably.

Table 1: Parameters for the new energy function

λ_d	λ_s	γ_s	φ_s
10.0	50.0	2.0	10.0

Table 2: Parameters for the genetic algorithm.

Population	Generations	$P_c ross$	P_{Ma}	P_{Mb}	P_{Mc}
50.0	1000	0.9	0.1	0.1	0.5

In order to evaluate the performance of the energy functions described in 3.3, some tests were carried out using exactly the same genetic algorithm but applying the classic energy formulation as fitness function instead. The truncated linear function was used for the smoothing function with a cost of 1.0 and a truncation value of 10.0. The results were uploaded to the Middlebury stereo web-page, following the same steps as in the previous case. The average percent of bad pixels increases from 5.81 to 8.56, which is near 3 more bad pixel percentage error if the classic energy function is used.

The evolution of each energy function during the optimization process compared to the bad-pixels error measurement for Tsukuba stereo pair is shown in 3. The image on the first row shows the whole energy which is being minimized during the first 500 generations. Both algorithms follow a similar descendant curve. However, they cannot be compared in terms of the minimum energy achieved given that different functions and parameters are used.

The image on the second row shows the evolution of the bad-pixels measurement of the best individual for each population. These charts were selected because they show empirically how the real disparity error evolves when each energy function is minimized. 3 shows that a reduction of the classic energy function not always translates into an effective reduction of the bad-pixels. Actually, in this test-case it produces some kind of unstable behaviour. The experiments carried out with the rest of the stereo pairs show the same trend. Meanwhile, using the energy function proposed, a much more stable behaviour and a much better final error for all the tests carried out is obtained.

However, it is important to notice that it cannot be stated that the proposed energy function represents the real disparity images better, i.e. the true disparity images obtain lower energy values than others. Genetic algorithms work fine for finding good approximations to real optimum values only when all the genetic operators are well set. It cannot be guaranteed that the genetic algorithm will perform better using the proposed energy function for any genetic configuration. Neither can be guaranteed that the optimum in one case has less error than the optimum in the other case. However, this trend has appeared in every tests carried out.

5 CONCLUSIONS

A new genetic algorithm has been proposed for stereo correspondence. Applying genetic algorithms has some benefits such as global reasoning and unrestricted fitness function. The contributions of this paper, is twofold. Firstly, compared to other genetic al-

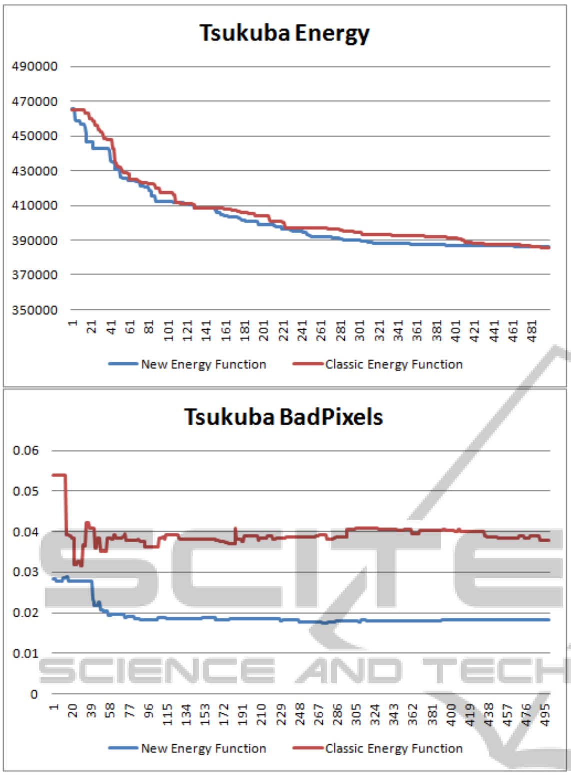

Figure 3: Evolution of the energy functions and the badpixels in Tsukuba.

gorithms previously proposed, it uses new crossover and mutation operators that account for occlusion handling. Both left and right disparity images are estimated in order to manage occlusions adequately. Secondly, it has been proposed and analysed a new energy function that includes occluded pixels handling in the formulation and enables depth discontinuities on pixels with high photometric derivatives.

The genetic algorithm has been evaluated using the standard Middlebury stereo dataset using both classic and proposed energy functions. Our implementation outperformed the classical one in 2.75 of bad pixels percentage on average, which represents a 32% error reduction using the new energy function. Moreover, an analysis of the evolution of the badpixels error measurement suggests that the new formulation is more adequate for representing real disparities. The algorithm proposed was rated with an average rank of 38.5 in the Middlebury ranking and as far as we know, is the first evolutionary algorithm included on this table.

REFERENCES

Alahari, K., Kohli, P., and Torr, P. H. S. (2010). Dynamic hybrid algorithms for map inference in discrete mrfs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(10):1846–1857.

Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 23(11):1222–1239.

Dai, C., Wu, X., and Liu, J. (2008). Stereo matching using adaptive genetic algorithm. In *Audio, Language and Image Processing, 2008. ICALIP 2008. International Conference on*, pages 1225–1228.

Gong, M. and Yang, Y.-H. (2001). Multi-resolution stereo matching using genetic algorithm. In *Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on*, pages 21–29.

Gong, M. and Yang, Y.-H. (2002). Genetic-based stereo algorithm and disparity map evaluation. *International Journal of Computer Vision*, 47(1):63–77.

Han, K.-P., Song, K.-W., Chung, E.-Y., Cho, S.-J., and Ha, Y.-H. (2001). Stereo matching using genetic algorithm with adaptive chromosomes. *Pattern Recognition*, 34(9):1729–1740.

Issa, H., Ruichek, Y., and Postaire, J. G. (2002). Stereo correspondence using a genetic scheme with a new solution encoding. In *Systems, Man and Cybernetics, 2002 IEEE International Conference on*, volume 6, page 5 pp. vol.6.

Kolmogorov, V. and Zabin, R. (2004). What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2):147–159.

Mei, X., Sun, X., Zhou, M., Jiao, S., Wang, H., and Zhang, X. (2011). On building an accurate stereo matching system on graphics hardware. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 467 –474.

Middlebury. http://vision.middlebury.edu/stereo/.

Nie, D.-H., Han, K.-P., and Lee, H.-S. (2009). Stereo matching algorithm using population-based incremental learning on gpu. In *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*, pages 1–4.

Saito, H. and Mori, M. (1995). Application of genetic algorithms to stereo matching of images. *Pattern Recognition Letters*, 16(8):815–821.

Wang, B., Chung, R., and Shen, C.-L. (2003). Genetic algorithm-based stereo vision with no block-partitioning of input images. In *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, volume 2, pages 830–836 vol.2.

Yoon, K. J. and Kweon, I. S. (2006). Adaptive support-weight approach for correspondence search. *Ieee Transactions On Pattern Analysis And Machine Intelligence*, 28(4):650–656.

Zhang, Z., Hou, C., and Yang, J. (2009). A stereo matching algorithm based on genetic algorithm with propagation stratagem. In *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on*, pages 1–4.