Stereo Vision Using Cost-Relaxation with 3D Support Regions

Roland Brockers, Marcus Hund, and Bärbel Mertsching

Faculty of Electrical Engineering, Informatics and Mathematics,
University of Paderborn, Germany

Email: {brockers, hund, mertsching}@get.upb.de

Abstract

In this paper a new stereo algorithm is presented for computing dense disparity maps from stereo image pairs using a cost relaxation approach, where the disparity map is the momentary state of a dynamic system. Following biologically motivated cooperative approaches a disparity space is defined with cooperating probability variables. In a first step a correlation based similarity measure is performed to initialize the relaxation process. Due to a very simple mathematical formulation, the relaxation itself could be realized as an optimization of a global cost function taking into account both the stereoscopic continuity constraint and considerations of the pixel similarity. The continuity constraint is implemented using a 3D Gaussian-weighted local support area of coupling probability variables which interact during the relaxation process. A special construction of the global cost function guarantees the existence of a unique global minimum of the cost function, which can be easily found with the help of a standard numerical procedure. In a post-processing step occlusions are detected and a sub-pixel precise disparity map is computed.

Keywords: stereo vision, correspondence problem, cost function, occlusion detection, sub-pixel precision

1 Introduction

In the last years a lot of promising algorithms have been developed to solve the correspondence problem in Stereo Vision [4], [14], [17]. In particular the application of graph based methods [2], [7] or segmentation based methods [1], [8] have improved the quality of the results. Unfortunately, the accuracy of the results in general can only be achieved with a higher computational effort, making the algorithms slower and unsuitable for real time applications like mobile robot stereo vision, where a fast 3D information acquisition with good accuracy is a basic requirement.

To consider these basic requirements, we developed a new stereo algorithm, which is based on a cost-relaxation. Relaxation approaches are a relatively old group of stereo vision algorithms, which were famously invented by the work of Marr and Poggio [9] in the seventies and are picked up again by Zitnick and Kanade in 2000 [17]. These algorithms calculate a disparity map using an iterative process, which improves an initial guess, computed from a specific similarity measurement. The improvement takes place with an iterative application of stereoscopic constraints to a set of probability variables that are organized inside a well-defined disparity space, each representing a certain disparity for a certain pixel in the reference frame.

Due to the dimensions of the disparity space the computational expense of such an approach is highly dependant on the velocity of convergence during the iteration process. For instance, a lot of algorithms are using competitive structures during the organization, decelerating the convergence of the mathematical system. At the other hand the main advantage of the iterative approach is the availability of a system state, representing the momentary best result for a disparity map at any time of the iteration process. So even after some iteration an improved disparity map can be computed in case of limited computational resources like in time-critical applications.

In our algorithm we fuse the advantage of the internal state, coding the momentary optimal disparity with a new mathematical formulation of the relaxation process that is based on minimizing a global cost function.

The approach was motivated by the work of Reimann, Haken [11] and Trapp [16] who formulated the correspondence problem as a biologically abutted self-organization process, based on Haken's synergetic pattern recognition equation [5].

To model the correspondence problem, all free system parameters are described via so-called order variables, containing the probability of the concerned disparities as explained above. For each pixel n we get a set of assigned variables containing the possibilities of different disparities in the later disparity map. The variables are organized in a single order parameter vector:

$$\boldsymbol{\xi} = (\xi_{(1,d_{\min})}, \ldots, \xi_{(n,d_{\min})}, \xi_{(1,d_{\min}+1)}, \ldots, \xi_{(n,d_{\max})})^T \quad (1)$$

The interesting aspect of this formulation is that the values of the variables can be interpreted as neural activities behaving like binocular neurons in visual cortex [9], [11].

Reimann's model uses the order parameter vector to state a system of coupled differential equations introducing a competition between all variables of a certain pixel **k**. Only one of the variables attached to pixel **k** can win the competition, claiming the final disparity value of **k**, all other variables are inhibited to zero activity.

Beside the attractiveness of the biologically equivalent, the solution via local competition has a major disadvantage. The numerical relaxation of the system takes place in a computationally expensive iteration process, making the approaches unsuitable for time-critical applications. For this reason we formulated a new approach, taking advantage of the principal formulation, but avoiding the local competition for acceleration. The algorithm picks up Reimann and Haken's attractive idea of coupling binocular neurons represented by order variables to implement a self-organization process with a modified differential equation system which is defined by a new cost approach.

Our algorithm works in two steps: According to former algorithms we use the initial guess of a similarity measurement which is computed in a first step. After this the organization process takes place as an optimization of a global cost function, formulated as a system of linear equations with a unique global minimum, which can easily be computed with the help of a fast standard numerical procedure. The cost function implements a local support of neighbouring order variables to realize the stereoscopic continuity constraint together with quality assumptions of the similarity measurement.

In the following we give a brief overview of the algorithm, concentrating on the question of forming the local support area to implement the continuity constraint. After describing the used similarity measurement and the cost-relaxation we explain the post-processing where an explicit occlusion detection and sub-pixel precise refining of the disparity map is implemented. Section 3 demonstrates the capability of the cost-relaxation algorithms showing the performance of the entire algorithm applied to standard stereo test images in competition with other algorithms.

2 Algorithm overview

2.1 Similarity measure

To compute an initial guess for the existing displacements in the two input images the first step of the algorithm is to compute a similarity measure as a metric for the probability that two pixels in the two input images form a correspondence pair, this means that they correspond to the two projections of a 3D scene point onto the camera planes.

Expecting the grey-level input images i_l and i_r to be rectified and undistorted, permits the application of a general correlation-based measurement. Several standard methods are possible, like SAD or SSD (cf. [12]). In our algorithm we use normalized cross-correlation, windowed and weighted by a function f_s to compute the similarity in a local neighbourhood area of the concerned pixels **x** and **x**+**d**:

$$s_0(\mathbf{x}, \mathbf{d}) = \frac{m_{\mathbf{x},\mathbf{x}+\mathbf{d}}}{s_{\mathbf{x}} s_{\mathbf{x}+\mathbf{d}}} \quad (2)$$

with

$$m_{\mathbf{x},\mathbf{x}+\mathbf{d}} = \sum_{\tilde{\mathbf{x}}} f_s^2(\mathbf{x}-\tilde{\mathbf{x}})[i_r(\mathbf{x})-\bar{i}_r(\mathbf{x})][i_l(\tilde{\mathbf{x}}+\mathbf{d})-\bar{i}_l(\mathbf{x}+\mathbf{d})] \quad (3)$$

$$s_{\mathbf{x}} = \sqrt{\sum_{\tilde{\mathbf{x}}} f_s^2(\mathbf{x}-\tilde{\mathbf{x}})[i_r(\tilde{\mathbf{x}})-\bar{i}_r(\mathbf{x})]^2} \quad (4)$$

$$s_{\mathbf{x}+\mathbf{d}} = \sqrt{\sum_{\tilde{\mathbf{x}}} f_s^2(\mathbf{x}-\tilde{\mathbf{x}})[i_l(\tilde{\mathbf{x}}+\mathbf{d})-\bar{i}_l(\mathbf{x}+\mathbf{d})]^2} \quad (5)$$

As the result we get a "traditional" similarity measure which computes an initial match value indicating the probability of d being the correct 1D disparity of the point **x**. Due to ambiguity occurring in the input images it is not sufficient to declare the pixel pair with the highest match as the resulting correspondence pair according to the pixel **x**. The resolution of this ambiguity occurs in the second step of the algorithm.

2.2 Global optimization via cost relaxation

The elimination of ambiguity is performed via a special relaxation procedure formulated in terms of a cost function, allowing to state stereoscopic constraints in single cost terms. The great advantage lies in the fact that with an appropriate definition of the cost terms, the search for the optimal solution is reduced to solving a system of linear equations, which can be done easily and, more important, rapidly, with the help of a standard numerical procedure.

According to Reimann's approach the disparity space *s(x, y, d)* is arranged as a vector $\boldsymbol{\xi}$ (see eq. 1) to follow the order parameter vector idea, which allows to achieve a valid disparity map at any time during the computation by simply choosing the highest variable. Additionally, the arrangement of the variables as a vector later on results in a more simple formulation of a gradient descent method to minimize the cost function.

Two assumptions are made about the cost function. First, costs arise if the elements $\xi_{(x,d)}$ of the disparity space differ from the initial values $\xi_{(x,d)_0}$ given by the similarity measure $s_0(x,y,d)$. Second, there will be costs if the stereoscopic continuity constraint [9] is not fulfilled. The first requirement leads to a cost term punishing the distance of the parameter vector $\boldsymbol{\xi}$ to $\boldsymbol{\xi}_0$:

$$P_1(\boldsymbol{\xi}) = c_1 \sum_{d=d_{\min}}^{d_{\max}} \sum_{i=1}^{n} (\xi_{(i,d)} - \xi_{(i,d)_0})^2 \tag{6}$$

The distance is added across all components of the disparity space and weighted by a positive constant c_1. According to the stereoscopic continuity constraint the disparity values of neighboured pixels in one image are expected to be piecewise smooth [9]. This leads to the formulation of a second cost term which generates costs, if the disparity of a pixel diverges from those of its neighbours:

$$P_2(\boldsymbol{\xi}) = c_2 \sum_{d=d_{\min}}^{d_{\max}} \sum_{i=1}^{n} \sum_{j \in U_i} w_{ji} (\xi_{(i,d)} - \xi_{(j,d)})^2 \tag{7}$$

c_2 again is a positive constant weight, while U_i represents the local support area to a given pixel *i*, defining the neighbouring pixels of *i*.

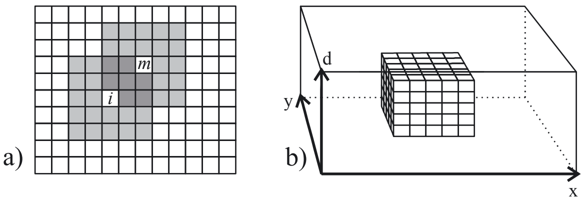

Figure 1: Example for local support areas *U*, (a) Fixed quadratic neighbourhood of two pixels *i* and *j*, (b) 3D cubic local support area

UU_i can be defined in different ways: In some cooperative algorithms, like the algorithms from Marr and Poggio [9], Reimann and Haken [11], or Trapp [16], the coupling surroundings are layer specific neighbourhoods within a constant disparity level (see fig. 1a). In other approaches like the one of Zitnick and Kanade [17] a three-dimensional support area inside the disparity space is used to define the coupling neighbourhood (fig. 1b).

In former versions of our algorithm [3] we used a two-dimensional support area inside a constant disparity level defined by a symmetric window function *f* which also defines the weighting factors w_{ij} to limit the influence of a variable depending on the distance to the centre pixel (see fig. 2).

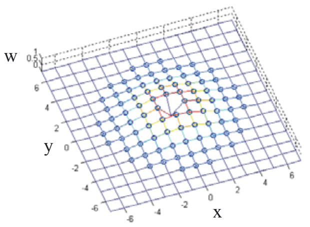

Figure 2: Circular window function *f* to describe the coupling 2D surrounding area *U* with Gaussian weights w_{ij}

To define a 3D support area for a point inside the disparity space we use an ellipsoidal neighbourhood region, which is also weighted by Gaussian weights (fig. 3).

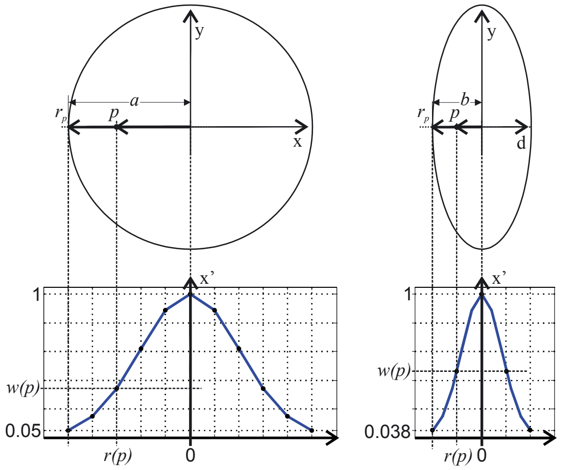

Figure 3: Definition of the ellipsoidal 3D local support area with Gaussian weights

The weights w_{ij} are defined inside an ellipsoid with the radiuses *a* and *b* in such a way that the minimal value on the x- and y-axis is 5% and on the d-axis is 3.8% of the maximum at the centre (fig. 3). The centre is not included in *U*, like in the 2D case (cp. fig. 2).

The complete cost function appears as the sum of the individual cost terms:

$$P(\boldsymbol{\xi}) = P_1(\boldsymbol{\xi}) + P_2(\boldsymbol{\xi}) \tag{8}$$

Due to the quadratic terms of the cost function there must exist a minimum which represents the optimal solution.

This minimum has to satisfy $\nabla P(\xi) = 0$, which leads to a system of linear equations:

$$\mathbf{A}\boldsymbol{\xi} - \boldsymbol{\xi}_0 = 0 \qquad (9)$$

Since the matrix **A** is symmetric and positive definite, there must exist an inverse $\mathbf{A}^{-1}$. Therefore we have a unique solution of equation (10).

In our approach the minimum of equation (8) is computed numerically with the gradient descent method. The fact of formulating $\boldsymbol{\xi}$ as a vector simplifies the resulting iteration rule to:

$$\boldsymbol{\xi}_{i+1} = \boldsymbol{\xi}_i - \lambda \nabla P(\boldsymbol{\xi}_i) \qquad (10)$$

with a positive fixed increment λ.

It can be shown that the iteration converges very rapidly to the global minimum of the cost function, which represents the optimal solution.

2.3 Disparity estimation, explicit occlusion detection and sub-pixel precision

Once the minimum of the cost function is found, the valid disparity *d(k)* of a pixel *k* can be retrieved via a maximum search across the parameters $\xi_{(k,j)}$ belonging to *k*.

The fact that the variable of the winning disparity of a pixel does not converge towards a fixed value allows to set up an explicit occlusion detection. Keeping in mind that $\xi_{(k,j)}$ still represents a measurement of the probability of a pixels disparity offers the possibility to select, in cases of ambiguity, the parameter with the highest value to obtain the most probable disparity.

A maximum search across all pixels *r* in one image corresponding with an identical pixel x_l in the other image (fig. 4a) eliminates correspondences of pixels originating from occluded areas corresponding with regular pixels in the second view (eq. 11), whereas a search for correspondence pairs that aim at scene points lying in succession from a viewpoint of a cyclopean camera (fig. 4b) eliminates correspondences of occlusion areas in both images (eq. 12) (see also [3]).

$$d(k) = \begin{cases} d(k), & \xi_{(k,d(k))} = \max\{\xi_{(r,d(r))}\}\Big|_{x_l = r + d(r)} \\ c, & \text{otherwise} \end{cases} \qquad (11)$$

$$d(k) = \begin{cases} d(k), & \xi_{(k,d(k))} = \max\{\xi_{(r,d(r))}\}\Big|_{x_c = r + \frac{d(r)}{2}} \\ c, & \text{otherwise} \end{cases} \qquad (12)$$

To label occluded areas, a constant value *c* outside of the disparity range is assigned to occluded pixels.

Finally, the same cost approach can be used in a modified form to compute sub-pixel precise disparity values for each pixel (eq. 13) which means in substance applying again the continuity constraint to the computed pixel-accurate disparity map. Before this can be done a filter operation eliminates detected one pixel wide occlusion areas which appear as a result of the pixel precise resolution. The detected pixels are filled with the average disparity value of their neighbours, now being able to participate in the following final relaxation process:

$$P(\tilde{\boldsymbol{\xi}}) = c_3 \sum_{i=1}^{n} (d(i) - d_0(i))^2 + c_4 \sum_{i=1}^{n} \sum_{j \in \tilde{U}_i} (d(i) - d(j))^2 \qquad (13)$$

Similar to $\boldsymbol{\xi}$, $\tilde{\boldsymbol{\xi}}$ is composed of the counted pixel disparity values:

$$\tilde{\boldsymbol{\xi}} = (d(1), \ldots, d(k))^T \qquad (14)$$

$\tilde{U}_i$ contains neighbouring pixels in a quadratic surrounding with $|d(i) - d(j)| < 1.3$ ensuring that disparity discontinuities are maintained.

3 Experimental Results

To evaluate the quality of our approach, the algorithm was tested with several test images that provide ground-truth data of the observed scene. Figure (5) shows the utilized images of the Tsukuba, Venus, Teddy and Cones scene that where introduced by Szeliski and Scharstein [12], [13], together with the according ground-truth maps.

To compare our results with those of other stereo algorithms the percentage of bad matching pixels is computed similarly to error functions found in literature [12] with an disparity tolerance threshold of $\delta_d = 1.0$ according to previously published studies [7], [12].

We compute two different errors. Because of the occlusion detection, the first measurement $B_{\overline{oo}}$ counts all bad pixels in regions that are not occluded in the ground-truth and the computed disparity map to consider exclusively false correspondences.

In comparison with other algorithms we also compute a second error measure $B_{\bar{o}}$ which is computed in the way of [12]. In this paper, Scharstein and Szeliski replace detected occlusions with the nearest background neighbouring disparity in the scan line and compute a bad pixel percentage for all pixels that are not occluded in the ground-truth. Doing so, the results of different algorithms can be compared to each other independent of the existence of occlusion detection.

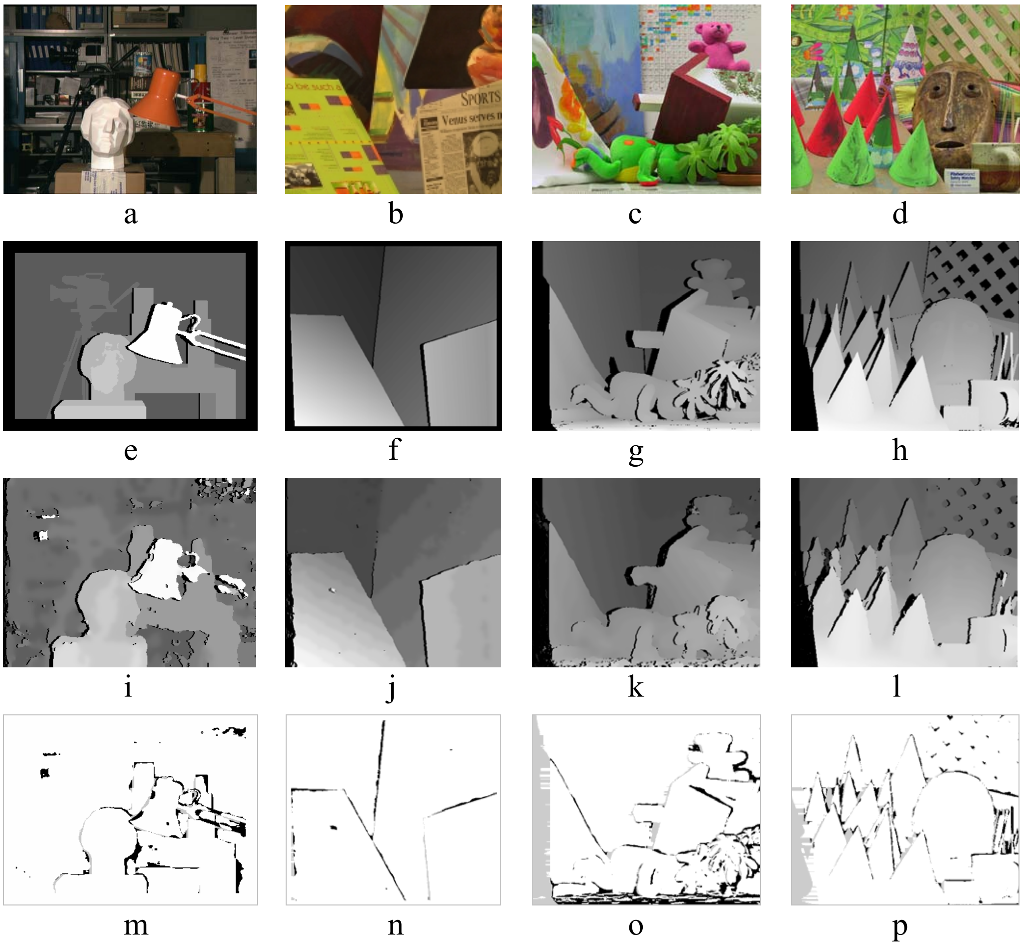

Figure 5: Tsukuba (a), Venus (b), Teddy (c) and Cones scene (d); a) – d) original images, left view; e) – h) ground-truth referring to the left view with black labelled occlusions, unconsidered borders are marked black as well; i) – l) computed sub-pixel precise disparity maps using ellipsoidal support regions (Cost relaxation II) (c_1=1; c_2=5.5; c_3=1; c_4=0.8; f_s=quadratic 3x3, U_i=ellipsoid (a=2, b=1), Gaussian weights; $\widetilde{U}_i$=max. 5x5); m) – p) error maps, bad matching pixels are marked black ($B_{\bar{o}}, \delta_d$=1), unconsidered occlusions are labelled grey, the border was added for visibility reasons

Table 1: Percentage of bad matching pixels in unoccluded regions, (* cp. [12] and: http://www.middlebury.edu/stereo [visited 9/2005])

	Tsukuba	Venus	Teddy	Cones
Graph Cut [2]	1.94*	1.79*	16.50*	7.70*
Dynamic Programming [1]	4.12*	10.10*	14.00*	10.50*
SymBP+occ [14]	0.97*	0.16*	6.47*	4.79*
Cost relaxation I [3], $B_{\overline{oo}}$ ($B_{\bar{o}}$)	5.77 (6.33)	1.44 (1.44)	7.97 (9.60)	3.65 (5.24)
Cost relaxation II, $B_{\overline{oo}}$ ($B_{\bar{o}}$)	4.46 (4.76)	1.35 (1.41)	7.81 (8.18)	3.52 (3.91)

Figure (5) indicates the computed disparity maps for the four scenes together with the used ground-truth maps. The similarity measure was computed with a minimum quadratic 3x3 correlation window because of the high amount of texture in every scene. Table (1) shows the quantitative results.

To compare the effect of using a 3D local support area we show the results for a relaxation with a circular, Gaussian weighted 2D support area (Cost Relaxation I) in contrast to the results achieved with the ellipsoidal 3D support region (Cost Relaxation II). The set of parameters was constant for all four scenes.

The comparison with some other selected stereo algorithms that were tested in [12] shows that algorithms like the Graph Cut algorithm introduced by Boykov, Veksler and Zabih [2] generally achieve better results while requiring much more computation time [12] whereas fast approaches like the Dynamic Programming algorithm cannot keep up with the quality of the results (tab. 1). In particular the performance for the Cones and Teddy scenes shows the capability of the new algorithm applied to well-textured real stereo images containing a deep disparity range.

To evaluate the advantage of the immanently coded disparity map, we computed the error changes during the iteration process.

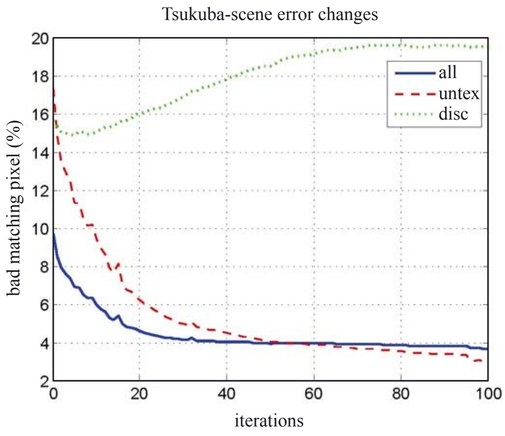

Figure 6: Error changes of $B_{\overline{oo}}$ during the relaxation applied to the Tsukuba scene (cf. fig. 5a). (all: $B_{\overline{oo}}$; untex: $B_{\overline{oo}}$ in untextured image regions; disc: $B_{\overline{oo}}$ near object borders; cf. [12]; Tsukuba optimized parameter set)

Figure (6) shows the changes of $B_{\overline{oo}}$ together with the error in untextured regions (untex) and image regions with disparity discontinuities (disc), computed for the Tsukuba scene (image regions according to [12]). The development shows a rapid descent of the over all error and the error in untextured regions during the first iterations making it possible to cut the iteration very early to achieve improved disparity maps in time-critical applications. The increase of the error rate near discontinuities is an immanent effect of the relaxation process, because of the simple formulation of the continuity constraint. At object borders an increase of sharpness could be achieved by limiting the local support area depending on additional information like contour information.

4 References

[1] A. F. Bobick, S. S. Intille, “Large occlusion stereo“, *IJCV*, 33(3), pp. 181-200, 1999.

[2] Y. Boykov, O. Veksler, R. Zabih, “Fast approximate energy minimization via graph cuts“, *IEEE TPAMI*, 23(11), pp. 1222-1239, 2001.

[3] R. Brockers, M. Hund, B. Mertsching, “Stereo matching with occlusion detection using cost relaxation”, *ICIP2005*, vol. III, pp. 389-392, 2005.

[4] S. Forstmann, J. Ohya, Y. Kanou, A. Schmitt, S. Thuering, “Real-time stereo by using dynamic programming“, *CVPRW2004*, vol. III, p. 29, 2004.

[5] H. Haken, *Synergetik*. Berlin: Springer Verlag, 1982.

[6] L. Hong, G. Chen, “Segment-based stereo matching using graph cuts”, *CVPR2004*, vol. I, pp. 74-81, 2004.

[7] V. Kolmogorov, R. Zabih, “Visual correspondence with occlusions using graph cuts“, *ICCV2001*, pp. 508-515, 2001.

[8] M. Lin, T. Tomasi, “Surfaces with occlusions from layered stereo“, *CVPR2003*, vol. I, pp. 710-717, 2003.

[9] D. Marr, T. Poggio, “Cooperative computation of stereo disparity“, *Science*, vol. 194, pp. 283-287, 1976.

[10] G. F. Poggio, T. Poggio, “The analysis of stereopsis“, *Annual Reviews of Neuroscience*, vol. 7, pp. 379-412, 1984.

[11] D. Reimann, H. Haken, “Stereovision by self-organisation“, *Biological Cybernetics*, vol. 71, pp. 17-26, 1994.

[12] D. Scharstein, R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms“, *IJCV*, 47(1/2/3), pp. 7-42, 2002.

[13] D. Scharstein, R. Szeliski, “High-accuracy stereo depth maps using structured light“, *CVPR2003*, vol. I, pp. 195-202, 2003.

[14] J. Sun, Y. Li, S. B. Kang, H. Y. Shum, “Symmetric Stereo Matching for Occlusion Handling”, *CVPR2005*, pp. 1075-1082, 2005

[15] J. Sun, H.-Y. Shum, N.-N. Zheng, “Stereo matching using belief propagation“, *ECCV2002*, vol. II, pp. 510-524, 2002.

[16] R. Trapp, S. Drüe, G. Hartmann, “Stereo matching with implicit detection of occlusions“, *ECCV1998*, pp. 17-33, 1998.

[17] C. L. Zitnick, T. Kanade, “A cooperative algorithm for stereo matching and occlusion detection”, *IEEE TPAMI*, vol. 22(7), pp. 675-684, 2000.