An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions

A. Luo and H. Burkhardt
Technische Informatik I
Technische Universität Hamburg-Harburg
Postfach 90 10 52, 21071 Hamburg, Germany

Abstract

This paper presents a new intensity-based stereo algorithm using cooperative bidirectional matching with a hierarchical multilevel structure. Based on a new model of piecewise smooth depth fields and the consistency constraint, the algorithm is able to estimate the 3-D structure and detect its discontinuities and the occlusion reliably with low computational costs. In order to find the global optimal estimates, we utilize a multiresolution two-stage algorithm minimizing nonconvex cost functions, which is equivalent to the MAP estimation. This basic framework computing the 3-D structure from binocular stereo images has been extended to the trinocular stereo vision for a further improvement of the performance. A few examples for the binocular and trinocular stereo problems are given to illustrate the performance of the new algorithms.
Keywords: binocular and trinocular stereo, cooperative bidirectional matching, discontinuity and occlusion, MAP estimation and multiresolution.

1 Introduction

Stereo vision is a fundamental technique to obtain the 3-D structure of a scene from 2-D images, which has been extensively studied in the past by many researchers. The methods for stereo matching can be grouped into three major categories: area-based [21], feature-based [20] [12] and intensity-based approaches [15] [10]. Usually the feature-based methods are preferably used in many publications, because sparse features can be reliably located and their matching is relatively simple. In contrast the intensity-based methods have long been overlooked as a potential method to cope with different intensities of stereo image pairs. These methods, however, have some advantages over feature-based methods: one directly can get a dense depth field estimate by matching algorithms and the performance of the algorithms does not rely on more or less reliable features. Extracting robust features in natural scenes can sometimes be very difficult and time consuming. Moreover, the algorithms can use the whole information of the images without loss through image preprocessing. The main causes for different intensity images that lead to difficult matching in intensity-based methods include photometric effects, occlusions, sensor and discretization noise. The problem with occlusion exists commonly in all three matching methods and will be discussed later. The last cause (mainly noise) can be partially overcome by statistical optimal estimations, for example, MAP methods [8] [24] [4] which are well suited to solve in general ill-posed problems in low level vision. Photometric effects can partly be eliminated by some adequate techniques, such as using a spatial coherent multiplier in the matching process [10]. Thus the intensity-based methods are also efficient and useful compared with other methods. Recently some works integrating both intensity-based and feature-based methods have been reported [6] [27], which, however, have a high computational complexity.

Although the stereo techniques have achieved a great progress, some problems have not yet been satisfactorily solved. One of the most important reasons is that discontinuities and occlusions are often not explicitly treated in many matching algorithms. Often stereo algorithms utilize the constraints of uniqueness, smoothness and ordering to simplify the matching process which, however, are invalid assumptions in occluded regions. If occluded areas are not detected explicitly, they may incorrectly match adjacent parts and interfere with the correct matching in the neighborhood of occluded regions. Thus a matching process taking into account occlusions is necessary to accurately recover the 3-D structure from 2-D stereo images. Recently several authors [27] [14] [6] [5] [19] proposed some new computational frameworks for stereo matching incorporating occlusion information. A common characteristic among these methods is that two matching processes (L to R and R to L) run separately and benefit little from each other.

A way for further improvement of stereo vision is to decrease the ambiguities of stereo matching by an additional triangulation geometry constraint with trinocular stereo (recently [13] [23]). It turns out that trinocular stereo vision overcomes many of the problems in binocular stereo accompanied with greater computational complexity. Besides eliminating the ambiguous matching, trinocular stereo techniques also provide a better performance in occluded regions and their adjacent regions.

Our contribution in this paper is to put forward a new intensity-based computational framework for stereo vision, which uses the mechanism of cooperative bidirectional stereo matching to estimate the depth fields and to detect the occluded regions on the stereo pairs reliably. Based on well known models, i.e. a deterministic structure model from imagery triangulation and a statistic model for image acquisition and a new Markov random field model for depth fields and occlusions, the a posteriori probability distribution or cost function of piecewise smooth depth fields is introduced in a similar way to general intensity-based regularization methods ([15] [10] etc.). But in contrast to existing solutions the discontinuities and occlusions are explicitly taken into account in our new cost function. Instead of simulated annealing methods or the graduated non-convexity algorithm for minimizing this nonconvex function, we simply use a deterministic relaxation algorithm of two stages handling the discontinuities of depth fields. Implementing it in a hierarchical multilevel structure, we get with high probability the global optimal depth estimates and simultaneously the occlusion maps for both stereo images with a rapid convergence. This algorithm which computes dense depth fields and occlusions in a unified framework is different from feature-based or correlation-based methods where the matching and interpolation (although in the later step it is also possible to preserve discontinuities) must separately be performed. In [9] it is tried to model occlusions and combine it into a Bayesian approach for stereo matching. But due to the high computational complexity only lateral spatial coherence is considered in their algorithm. Our framework is rather different and can easily be implemented.

We have directly extended this method to the case of trinocular stereo to reduce even further the remaining ambiguities.

Compared with the strategy of estimating a single depth field and computing occlusions from it, our algorithm using two dependent maps has a higher complexity. But the cooperative bidirectional matching can help each other for overcoming some wrong relaxation and increases the reliability and stability of the depth estimation, which is important for intensity-based methods to estimate a dense depth map. The detection and utilization of a reliable occlusion information is then simplified as well.

This paper is organized as following: in the next section we analyze occlusions in detail, give various models for MAP estimation and derive an objective cost function. In section 3 and 4 an algorithm minimizing the objective cost function for the depth estimation and the related hierarchical multilevel implementation are put forward. Section 5 provides some experimental results of this binocular stereo matching method by a few examples. In section

6 the algorithm is extended into the trinocular stereo matching, and the improvement of the depth estimation is shown with a real example, and then a conclusion follows.

2 The Bayesian model and occlusions

As shown in Fig. 1, we assume a general epipolar camera model throughout the context of the whole paper, where two coplanar images $g_l(x,y)$ and $g_r(x,y)$ are formed by the perspective projection with the focal length F, and with parallel optical axes Z_l and Z_r separated by the baseline length B. Non-parallel axis stereo images can easily be reprojected into parallel axis stereo images by rectification [13], which can be treated according to the simpler epipolar constraint. Rather than estimating a disparity field as usually, we estimate the depth field directly. In trinocular stereo it has the advantage that there are identical estimates for the horizontal and vertical matching.

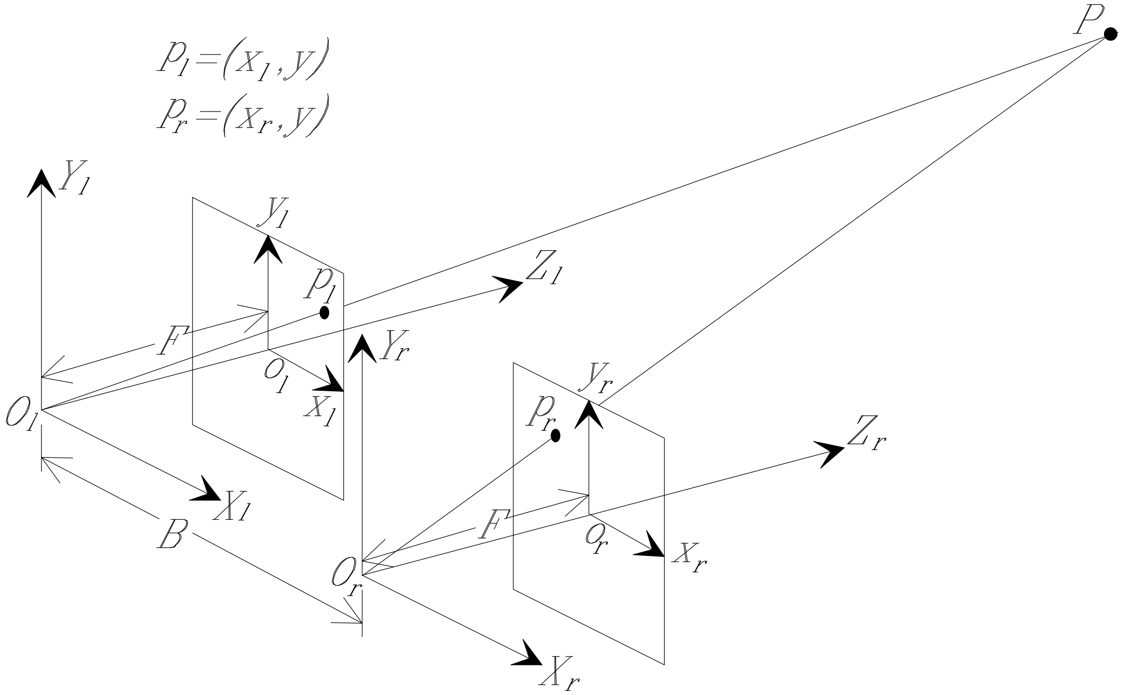


Figure 1: The camera geometry of the binocular stereo vision

Structure and observation models from the triangulation According to the triangulation geometry, we can know the corresponding points of the stereo images, and derive the relationship between their intensity values $g_l(x,y)$ and $g_r(x,y)$ with the knowledge of depth values $Z_l(x,y)$ or $Z_r(x,y)$ in image coordinate systems, i.e.

$$g_l(x,y) - g_r(x_r,y) = n_1(x,y) \tag{1}$$

or

$$g_r(x,y) - g_l(x_l,y) = n_2(x,y) \tag{2}$$

with the abbreviations $x_r = x - \frac{BF}{Z_l(x,y)}$ and $x_l = x + \frac{BF}{Z_r(x,y)}$, where $n_1(x,y)$ and $n_2(x,y)$ are random terms due to sensor and discretization noise. To emphasize our main goal, we neglect photometric effects here. If necessary, however, a spatial coherent multiplier due to photometric effects can easily be introduced into the above equations (1) and (2) (see [10]).

The above equations are valid only when neither of the corresponding points on both images is occluded by others. Under this condition both the depth maps in different image coordinate systems have the relationship as following, which means that both maps represent a consistent 3-D physical structure:

$$Z_l(x, y) = Z_r(x_r, y) \tag{3}$$

or

$$Z_l(x_l, y) = Z_r(x, y) \tag{4}$$

with the same abbreviations x_r and x_l as above.

Occlusions The problem with occlusions plays an important role in stereo matching, because some basic assumptions, for example uniqueness and smoothness, and the fundamental triangulation geometry are invalid in occluded regions. Occlusion occurs when a part of the scene that is visible in one of the views is hidden in the other view or is beyond the other view. The past stereo approaches without taking into account this problem usually have some troubles in the neighborhood of occluded regions, which often lead to some wrong matches.

A constraint from the occlusion geometry is then necessary for further significant improvements. On the other hand, the matching processes using the information about occluded regions can help to determine the associated depth discontinuities disambiguously. For binocular stereo matching two kinds of occlusions must separately be considered: left occluded regions and right occluded regions. The left occluded regions are visible only in the left stereo image and have no corresponding points in the right image and vice versa. Therefore any matching in these occluded regions must be false.

From a simple geometric analysis, the occluded regions in one stereo image must correspond to discontinuities of the depth map in another stereo image coordinate system. The location of occluded regions in both stereo images can be represented by two binary occlusion maps $O_l(x, y)$ and $O_r(x, y)$:

$$O_s(x, y) = \begin{cases} 1 & : \quad (x, y) \notin \text{occluded regions of } s\text{−image} \\ 0 & : \quad (x, y) \in \text{occluded regions of } s\text{−image} \end{cases} \tag{5}$$

where s stands for l(eft) or r(ight) (or b(ase), u(pper) in the whole paper).

The occlusion map $O_l(x, y)$ can directly be derived from the accurate depth map $Z_r(x, y)$ in the other system (but not easily from $Z_l(x, y)$), and similarly $O_r(x, y)$ from $Z_l(x, y)$:

$$\{(x, y)|O_l(x, y) = 1, (x, y) \in \mathcal{B}\} = \{(x_l, y)|x_l = x + \frac{BF}{Z_r(x, y)}, (x, y) \in \mathcal{B}\} \tag{6}$$

$$\{(x, y)|O_r(x, y) = 1, (x, y) \in \mathcal{B}\} = \{(x_r, y)|x_r = x - \frac{BF}{Z_l(x, y)}, (x, y) \in \mathcal{B}\} \tag{7}$$

where $\mathcal{B}$ is the whole region of an image. In the discrete case, we can calculate the occlusion map $O_r(x, y)$ as follows, when $Z_l(x, y)$ has already been obtained (Similarly for the dual process):

The right occlusion map $O_r(x, y)$ is initialized to zero. By the triangulation geometry, every pixel (x,y) of the left image is reprojected into a subpixel-point (x_r, y) of the right image with the a priori knowledge of $Z_l(x, y)$, such that the occlusion map $O_r(\lfloor x_r \rfloor, y)$ and $O_r(\lfloor x_r \rfloor + 1, y)$ at the two nearest pixels of (x_r, y) are accumulated with the factors, which are equal to the distance from the subpixel to another nearest pixel, where $x_r = (x - \frac{BF}{Z_l(x,y)})$. After applying all the above operations through the whole left image this occlusion map is filtered by a 3 × 3 lowpass average filter and then thresholded with a value, e.g. 0.65, so as to remove some small occluded regions and noise. We represent this process as a transformation $\mathcal{OT}$:

$$O_r(x,y) = \mathcal{OT}_{LR}(Z_l(x,y)) \tag{8}$$

$$O_l(x,y) = \mathcal{OT}_{RL}(Z_r(x,y)) \tag{9}$$

When the depth maps are accurately estimated, then the occlusion maps are also accurately obtained.

A priori model of the depth field The probability distribution of a Markov random field (MRF) process depends only on a finite neighborhood. Consequently it can be used to model the depth field that has certain local spatial coherent properties, e.g. piecewise smoothness. With the help of the equivalence between the MRF and Gibbs distribution, we can obtain the a priori probability distribution of the depth field in an explicit form (see [8]):

$$P(\mathbf{Z}_s) = \frac{1}{C_s} exp(-\frac{U(\mathbf{Z}_s)}{T}) \tag{10}$$

where T is a constant (temperature of the model), C_s is a normalization constant (partition function) and does not depend on $\mathbf{Z}_s$. The energy function $U(\mathbf{Z}_s)$ can be written as a sum of local potential functions:

$$U(\mathbf{Z}_s) = \sum_{(x,y)\in\mathcal{B}} V_{x,y}(\mathbf{Z}_s) \tag{11}$$

where $\mathcal{B}$ is the set of all pixels (x,y) within the image and each potential function $V_{x,y}(\mathbf{Z}_s)$ depends only on a certain neighborhood.

The choice of the potential function $V_c(\mathbf{Z}_s)$ can directly reflect how to model the local properties of the 3-D surface. For the assumption about the global smoothness of 3-D surfaces, we can choose a simple quadratic function of neighboring difference on the 4-connected neighborhood of each pixel as potential function (the standard a-priori smooth model):

$$V_c(\mathbf{Z}_s) = (Z_s(x+1,y) - Z_s(x,y))^2 + (Z_s(x,y+1) - Z_s(x,y))^2 \tag{12}$$

This assumption about the global smoothness of 3-D surfaces, however, often contradicts real scenes. The assumption of a piecewise smoothness is much better suited to reflect the conditions of real scenes. Thus various models are put forward to describe the piecewise smoothness [8] [26] [2], where an additional line variable field represents the discontinuity process. However, the line field, which plays an important part in these models, complicates the associated algorithms of the optimal estimation.

In this paper we use a simple form of potential function modeling the local properties of piecewise smooth 3-D surfaces, which takes into account discontinuities implicitly. Therefore, a probability distribution or energy function of the depth field only (without the line field for discontinuities) should be defined.

It is well known, that the median filtering is a nonlinear local operation on neighborhoods. A piecewise smooth field is continuous almost everywhere. It means that all neighbors of each pixel on the field are strongly correlated, unless there are discontinuities between them. Even the pixels near discontinuities are still strongly correlated with most of their respective neighbors. Like a low-pass filtering, a median filtering smoothes an image in flat regions and, therefore, is useful to preserve the smoothness and to reduce noise. Unlike a low-pass filtering, a median filtering can preserve discontinuities on a piecewise smooth image without explicitly detecting them and, moreover, removes impulsive or salt-and-pepper noise without affecting other pixels. These properties are interesting and essential for preserving both the strong correlation and discontinuities of a piecewise smooth depth field. Based on this, we can define the following local potential function of weak smooth depth fields to model their local interaction, which involves only local operations:

$$V_c(\mathbf{Z}_s) = (Z_s(x,y) - Z_s^*(x,y))^2 \tag{13}$$

where $Z_s^*(x,y) = \text{median}_{(x',y')\in W_{x,y}}(Z_s(x',y'))$ and the window $W_{x,y}$ is a relative large neighborhood of the pixel (x,y), e.g. a 5×5 window.

For example, suppose a piecewise smooth depth field without noise. The median filtering would tend to preserve edges and smooth regions without degradation, such that the above local potential function at each pixel of the whole field tends to be low or zero. If this field is not in a state of piecewise smoothness and corrupted by noise, the median filter usually removes this noise, so that the filtered field differs from the original field at these noisy pixels and there the local potential functions have high values. Therefore, only piecewise smooth fields have a high probability according to this local potential function, and the other possible fields have a low probability. This is only a simple explanation about the validity of Eq. (13) as the potential function of a weak smooth field.

It is difficult to prove theoretically the properties of this model, but the empirical results indicate that this model reflects the local properties of a piecewise smooth surface satisfactorily.

Bayesian model – MAP estimation Assuming that both the stereo images $g_l(x,y)$ and $g_r(x,y)$ are given, we need to estimate the optimal depth field $Z_s(x,y)$. The optimal depth field $Z_s(x,y)$ should have the maximal a posteriori probability over all possibilities:

$$max_{\mathbf{Z}_s} \quad P(\mathbf{Z}_s|\mathbf{g}_l, \mathbf{g}_r) \tag{14}$$

From Eq. (1), we can derive the conditional probability of $\mathbf{g}_r$ given $\mathbf{g}_l$ and $\mathbf{Z}_l$ and $\mathbf{O}_l$ as

$$P(\mathbf{g}_r|\mathbf{g}_l, \mathbf{Z}_l) = \frac{1}{C} \quad exp(-\frac{1}{2\sigma^2}\sum O_l(x,y)(g_l(x,y) - g_r(x_r,y))^2) \tag{15}$$

with a normalization constant C under the assumption of Gaussian noise.

According to the Bayesian rule, we can derive the a posteriori probability of $Z_l(x,y)$ from Eqs. (10) and (15) as follows:

$$P(\mathbf{Z}_l|\mathbf{g}_l, \mathbf{g}_r) = \frac{P(\mathbf{g}_r|\mathbf{Z}_l, \mathbf{g}_l)P(\mathbf{Z}_l)}{P(\mathbf{g}_r|\mathbf{g}_l)} = \frac{1}{C_{pl}} exp(-\frac{U_{pl}(\mathbf{Z}_l)}{T}) \tag{16}$$

where C_{pl} is a normalization constant and the a posteriori energy function is given by:

$$U_{pl}(\mathbf{Z}_l) = \lambda \sum_{(x,y)\in\mathcal{B}} O_l(x,y)(g_l(x,y) - g_r(x_r,y))^2 + \sum_{(x,y)\in\mathcal{B}} V_{x,y}(\mathbf{Z}_l), \tag{17}$$

with $\lambda = T/2\sigma^2$. Eq. (16) holds because the depth field and only a single image are statistically independent. Similarly, one can derive the other a posteriori energy function $U_{pr}(\mathbf{Z}_r)$:

$$U_{pr}(\mathbf{Z}_r) = \lambda \sum_{(x,y)\in\mathcal{B}} O_r(x,y)(g_l(x_l,y) - g_r(x,y))^2 + \sum_{(x,y)\in\mathcal{B}} V_{x,y}(\mathbf{Z}_r) \tag{18}$$

Obviously one can convert the original MAP estimation alternatively into the problem of minimizing these (a posteriori) cost energy functions (s=l and r):

$$min_{\mathbf{Z}_s} U_{ps}(\mathbf{Z}_s) \tag{19}$$

One can see that these two MAP estimation processes (19) (s=l and r) are dependent through the consistent constraints (3) (4) and the transformations (8) (9). Thus in order to estimate the optimal depth fields simultaneously taking into account occlusions, we need to solve the following problem:

$$min_{\mathbf{Z}_l,\mathbf{Z}_r} \left(U_{pl}(\mathbf{Z}_l) + U_{pr}(\mathbf{Z}_r)\right) \tag{20}$$

with the constraints (3), (4), (8) and (9).

So far we have established a complete mathematical model of the depth estimation. In the next section we give a solution.

3 The new algorithm

Solving the problem described in the previous section is not simple, because the cost energy functions are usually nonconvex and there exist many complicated constraints. In order to solve the constraint optimization problem, we utilize an iterative method, where within each step the depth fields are relaxed to optimize the cost energy function (20) incorporating the constraints of consistency (3) and (4), and in cascade the occlusion fields are refreshed according to Eq. (8) and (9).

At first we consider the methods to optimize the cost energy functions. The cost energy function described by (20) under the assumption of a piecewise smooth surface may have many local minimal solutions, i.e. it may be nonconvex. This means that the simple gradient descent will fall into a local minimum rather than into the global minimum, depending strongly on the chosen initial point.

There exists a commonly known statistical method to find the global optimum of a nonconvex cost function, which calls simulated annealing (see [8]). Despite its great success to nonconvex problems, however, the simulated annealing method has the following disadvantages in realistic applications. The algorithm converges very slowly, especially in our problem which has a huge number of state variables. Secondly the standard algorithm handles only a discrete problem and is difficult to get subpixel accuracy.

An alternative method, namely the graduated non-convexity algorithm, is provided to minimize a nonconvex function for the visual reconstruction in a special case [2]. The kernel of the algorithm includes two main steps: The first is to construct a convex approximation to the non-convex function and then to find its minimum. The second step is to construct a sequence of functions, controlled by a parameter and ending with the true nonconvex function, and to use a deterministic relaxation method with the searching history descending on them in cascade. And each process provides a nearly global optimum for the next. The authors claim that the algorithm is very effective for problems with weak continuities and converges in a few steps, although there is no theoretical proof.

Inspired by the above work [2] that is not appropriate for our model, we construct a framework of two-stage algorithm, where only a simple function approximating the true nonconvex function is used. At first we optimize the approximate function with a deterministic relaxation method. The optimal result of minimizing this function provides a good initial guess for further optimizing the original nonconvex objective function, which lies in a convex neighborhood of the global minimum. Beginning with this initial point, we continue to minimize the true nonconvex function also with the deterministic relaxation method in order to find the global optimum. Here due to the absence of a smooth transition between the two functions, it is not guaranteed that the initial point lies near enough to the global optimal point. To avoid possible local minima near the global minimum, we allow some small fluctuations during the descent process without strictly restricting the criteria to a monotonic decrease.

In order to find the global optimal estimate under the assumption of the piecewise smoothness tied with (13), we first minimize an approximate objective function under the assumption of the global smoothness tied with (12). Although this reconstructed global smooth field differs greatly from the true field near discontinuities, these regions, namely edges of the objects, are relatively sparse in the whole field. Otherwhere, both models are qualitatively consistent and this global smooth estimate should be rather accurate. Therefore this smooth estimate can provide a good initial point near the global minimum for optimizing the true objective function (20) with (13). If the stereo images are smooth enough, we can take the approximate function tied with (12) to be convex, because the images with low frequencies are almost linear in a large enough neighborhood. It means that we can obtain a single optimal solution for this cost function at a coarse resolution, which gives us an approximation of the true global solution.

The deterministic relaxation process can be derived from the discrete approximation of the Euler equation in the continuous case, or the discrete version of the objective function is directly minimized by differentiation:

$$Z_l^{k+1}(x,y) = \bar{Z}_l^k(x,y) + \lambda O_l^k(x,y)(g_l(x,y) - g_r(x_r,y))g_{rx}(x_r,y)\frac{BF}{(Z_l^k(x,y))^2} \tag{21}$$

$$Z_r^{k+1}(x,y) = \bar{Z}_r^k(x,y) + \lambda O_r^k(x,y)(g_l(x_l,y) - g_r(x,y))g_{lx}(x_l,y)\frac{BF}{(Z_r^k(x,y))^2} \tag{22}$$

with the abbreviations $x_r = x - \frac{BF}{Z_l^k(x,y)}$, $x_l = x + \frac{BF}{Z_r^k(x,y)}$ and $g_{sx}(x,y) = \partial g_s(x,y)/\partial x$. $\bar{Z}_s^k(x,y)$ is the filtered field of $Z_s^k(x,y)$ which is divided into two cases: if global smooth surfaces with (12) are assumed, $\bar{Z}_s^k(x,y)$ is the local average of $Z_s^k(x,y)$, e.g. on the 4-connected neighborhood we get:

$$\bar{Z}_s^k(x,y) = \frac{1}{4}(Z_s^k(x-1,y) + Z_s^k(x,y-1) + Z_s^k(x+1,y) + Z_s^k(x,y+1)) \tag{23}$$

if piecewise smooth surfaces with (13) are assumed, $\bar{Z}_s^k(x,y)$ is the local median value of $Z_s^k(x,y)$ in a neighborhood, e.g. a 5×5 window:

$$\bar{Z}_s^k(x,y) = \text{median}_{(x',y')\in W_{x,y}}(Z_s(x',y')) \tag{24}$$

where a simple approximation is assumed, i.e. (not in a strict mathematical sense):

$$\frac{\partial}{\partial Z_s(x',y')}(\text{median}_{(x',y')\in W_{x,y}} Z_s(x',y')) \approx 0, \tag{25}$$

because an impulse corrupting a single variable has almost no influence on the median value. Of course we must utilize a simplified fast implementation for such a 5×5 median filter in our stereo vision algorithm.

In order to improve the stability of the algorithm, we can rewrite the update equations (21) and (22) into the following form:

$$\begin{aligned} Z_l^{k+1}(x,y) &= \bar{Z}_l^{k+}(x,y) + \lambda O_l^k(x,y)(g_l(x,y) - g_r(x_r^+,y))g_{rx}(x_r^+,y)\frac{BF}{(\bar{Z}_l^{k+}(x,y))^2} \quad (26) \\ Z_r^{k+1}(x,y) &= \bar{Z}_r^{k+}(x,y) + \lambda O_r^k(x,y)(g_l(x_l^+,y) - g_r(x,y))g_{lx}(x_l^+,y)\frac{BF}{(\bar{Z}_r^{k+}(x,y))^2} \quad (27) \end{aligned}$$

with the abbreviations $x_r^+ = x - \frac{BF}{\bar{Z}_l^{k+}(x,y)}$ and $x_l^+ = x + \frac{BF}{\bar{Z}_r^{k+}(x,y)}$, where

$$\begin{aligned} O_l^k(x,y) &= \mathcal{OT}_{RL}(\bar{Z}_r^k(x,y)) \quad (28) \\ O_r^k(x,y) &= \mathcal{OT}_{LR}(\bar{Z}_l^k(x,y)) \quad (29) \\ \bar{Z}_l^{k+}(x,y) &= \frac{\bar{Z}_l^k(x,y) + \bar{Z}_r^k(x_r,y)O_r^k(x_r,y)}{1 + O_r^k(x_r,y)} \quad (30) \\ \bar{Z}_r^{k+}(x,y) &= \frac{\bar{Z}_r^k(x,y) + \bar{Z}_l^k(x_l,y)O_l^k(x_l,y)}{1 + O_l^k(x_l,y)} \quad (31) \end{aligned}$$

with the abbreviations $x_r = x - \frac{BF}{\bar{Z}_l^k(x,y)}$ and $x_l = x + \frac{BF}{\bar{Z}_r^k(x,y)}$, where the Eqs. (30) and (31) enforce both depth fields $Z_r(x,y)$ and $Z_l(x,y)$ representing a consistent 3-D structure with a weighted averaging according to the constraints (3) and (4).

The equations (26) to (31) provide a complete iterative step of the relaxation algorithm estimating the optimal depth fields with simultaneous detection of the occlusion maps. Noticing the two-stage process, we begin to calculate the expected filtered depth fields $\bar{Z}_s^k(x,y)$ using Eq. (23) in the iterative steps at first until the objective cost function converges very slowly or fluctuates, then switch to the model of piecewise smooth surfaces and utilize Eq. (24) to compute $\bar{Z}_s^k(x,y)$ until no further improvement is expected in further steps.

In this section we have given the relaxation algorithm for an intensity-based stereo vision. If one has good initial approximating estimates for the depth fields and occlusion maps, then better estimates for them would be achieved by the relaxation steps described above. Once the estimates don't fall into local minima, the estimation will continue to be improved until approaching the global minimum nearly. Because the energy function of stereo vision is more complicate than that of the surface interpolation in [2] and [8], our energy function even with the global smooth model (12) is often still nonconvex too. Although this two-stage algorithm can greatly help avoiding many local minima of the energy function, it isn't guaranteed to obtain the global optimum.

The remaining problems are how we get a better initial estimate that can lead to the global optimum easily, and accelerate the convergence of the algorithm. This can be solved satisfactorily by a hierarchical multilevel structure in the implementation that is described in the next section.

4 The multilevel implementation

The relaxation algorithm given in the previous section needs a good initial estimate, especially if large disparities in the stereo image pairs exist. An inadequate initial point would cause the changes of the estimates into a wrong direction due to the local properties of intensity variations, and lock them in wrong matches. To reduce the effect of the local properties, it is necessary to smooth and blur the images by removing high frequency components and intensify the global structure in a coarse level using only low frequency components. It motivates us to represent the data structure and implement the algorithm by the hierarchical multilevel technique [22]. Besides its most obvious advantages of greatly reducing the computational cost of various algorithms, the multilevel structure provides a useful tool of converting global properties into local properties.

Some biological researchers [20] show that there exist spatially tuned binocular channels in the human vision system, and the interaction across channels eliminates the ambiguities of matching, where the perception in low frequency channels provides a better initial estimate for searching in high frequency channels. Hierarchical multilevel structures have extensively been used in various problems of low-level computer vision [25] [11] [7] [1] [10] [27] [4] and have proved to be a very efficient in saving computational costs and overcome errors due to local properties. The fast convergence of a multilevel relaxation algorithm is based on the mechanism that high-frequency errors are quickly smoothed and the low-frequency errors resist to a decrease in contrast. The low-frequency components in a finer level remain in the relatively higher frequencies in a coarser level. The whole error from low to high frequencies can progressively be eliminated quickly from the coarsest to the finest level. Thus we also utilize the multilevel structure to implement our algorithm. In the coarse levels the algorithm can more globally find rough matchings between the smoothed images. Then the rough estimates are propagated into the fine levels, which provide a good initial value for the relaxation there. As discussed in the previous section, our approximate energy function tied with the global smooth model (12) is believed to be convex at a coarse enough level, so that only a unique estimate for the rough optimal matching can be found there, which is used for further global optimization.

The hierarchical multilevel structure for implementing our algorithm is shown in Fig. 2. At first we produce the Gaussian pyramids for both stereo images, in which the image in the coarse level (t) is the subsample of a low-pass filtered image in the next finer level ($t-1$). Let the finest level (0) have the original resolution and the coarsest level be M (here M=2, only the 3 levels are shown). Then for $1 \leq t \leq M$ we have (see [3]):

$$g_{s,t}(x,y) = \sum\sum_{m,n=-2}^{+2} w(m,n) g_{s,t-1}(2x+m, 2y+n) \tag{32}$$

where w() is a Gaussian kernel.

Let's initialize the whole depth fields $Z_{s,M}(x,y)$ with a reasonable constant, e.g. in the range from half to ten times as large as an approximate true depth, and the whole occlusion maps $O_{s,M}(x,y)$ as value 1 (no occluded regions are assumed) in the coarsest level M. Then we carry out the relaxation algorithm from the coarsest level M to the finest level 0. The optimal estimates of the depth fields $Z_{s,t}(x,y)$ and the occlusion maps $O_{s,t}(x,y)$ at the level t are propagated into the next finer level ($t-1$) as good initial estimates through an interpolation process:

$$\begin{aligned} Z_{s,t}(x,y) &= 4\sum\sum_{m,n=-2}^{+2} w(m,n) Z_{s,t+1}\left(\frac{x+m}{2}, \frac{y+n}{2}\right) && (33) \\ O_{s,t}(x,y) &= bfunc(4\sum\sum_{m,n=-2}^{+2} w(m,n) O_{s,t+1}\left(\frac{x+m}{2}, \frac{y+n}{2}\right)) && (34) \end{aligned}$$

where $bfunc(x) = u(x-0.5)$ is a thresholding function with threshold 0.5 ($u(x)$ is the unit step function), and $\frac{x+m}{2}$ and $\frac{y+n}{2}$ are integer division operations.

One relaxation step in each level utilizes the equations from (26) to (31) with (23) or (24) to get better occlusion maps and depth fields, in which two parallel relaxation processes for $Z_{l,t}(x,y)$ and $O_{l,t}(x,y)$, and $Z_{r,t}(x,y)$ and $O_{r,t}(x,y)$ receive the results of another process

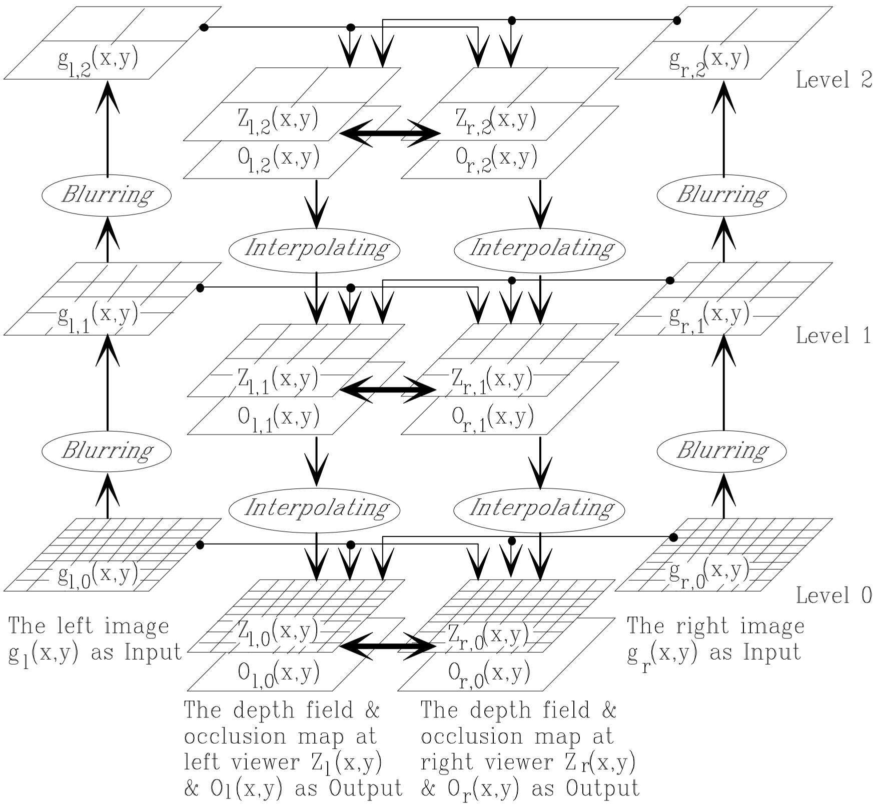


Figure 2: The multilevel implementation of the binocular stereo algorithm

and are dependently carried out. This framework has some advantages. At first, it provides the consistent results for depth fields and occlusion maps in the left and right different image coordinate systems. Secondly, the algorithm converges faster and more stable than a simple matching in one direction, because both relaxation processes use more information about the better refreshed depth fields and occlusion maps in due time from each other and overcome some wrong relaxation. The two-stage algorithm at each level guarantees the accurate detection of discontinuities and occlusions.

Whereas for a single resolution the convergence rate is strongly influenced by the initial estimates and their error characteristics and is very slow for low-frequency components, the multigrid methods have a rather good convergence rate, which is less dependent on the initial values. Normally, this multigrid algorithm converges in tens of relaxation steps all together. The complexity of one relaxation step is only proportional to the number of pixels of the image. Suppose the quantity “**WU**” (working unit) as the computational cost of one relaxation step at the finest level. Then one “**WU**” hat $\mathcal{O}(N)$ scalar operations, where N is the size of the image at the finest level. Therefore, the total cost of the algorithm is less than $\frac{4}{3}l\mathbf{WU}$, where l is the number of relaxation steps at each level and equal to about 10 or little more.

5 Experimental results

From the above discussion in the previous section one can see that our cooperative bidirectional stereo matching with multilevel structure shown in Fig. 2 can provide accurate and stable results with moderate computational costs. In order to further illustrate the performance of the algorithm, a few examples estimating the depth fields and detecting the occlusions from synthetic and real stereo image pairs are given in the following:

Synthetic stereo image pairs:

In the top-right of Fig. 3 (a) and (b) two images (left and right) $g_l(x,y)$ and $g_l(x,y)$ of a synthetic stereo pair are separately shown. Assuming a focal length of $F = 225$ pixels and a baseline length of $B = 2cm$, the images are digitally simulated, where the scene contains a square planar object of size $20cm \times 20cm$ size with a distance of $75cm$ and a uniform background with the distance of $150cm$ from the viewer. The estimated depth fields $Z_l(x,y)$ and $Z_r(x,y)$ of the original resolution in the left and right image coordinate systems are separately shown in the bottom-left of Fig. 3 (a) and (b) (the brightest stand for 255 cm, and the darkest for 0 cm). Similarly the detected occlusion maps $O_l(x,y)$ and $O_r(x,y)$ of the original resolution are individually shown in the bottom-right of Fig. 3 (a) and (b) (the darkest represent occluded regions). The intermediate results in the coarsest and the middle levels are shown in the top-left of Fig. 3 (a) and (b) similarly only in a smaller size. Our algorithm has obtained very good results. The average relative error of the whole estimated depth field outside the occluded regions is only 0.283%. The occluded regions are almost exactly detected, although there exist ragged edges due to some discretization effect and small errors.

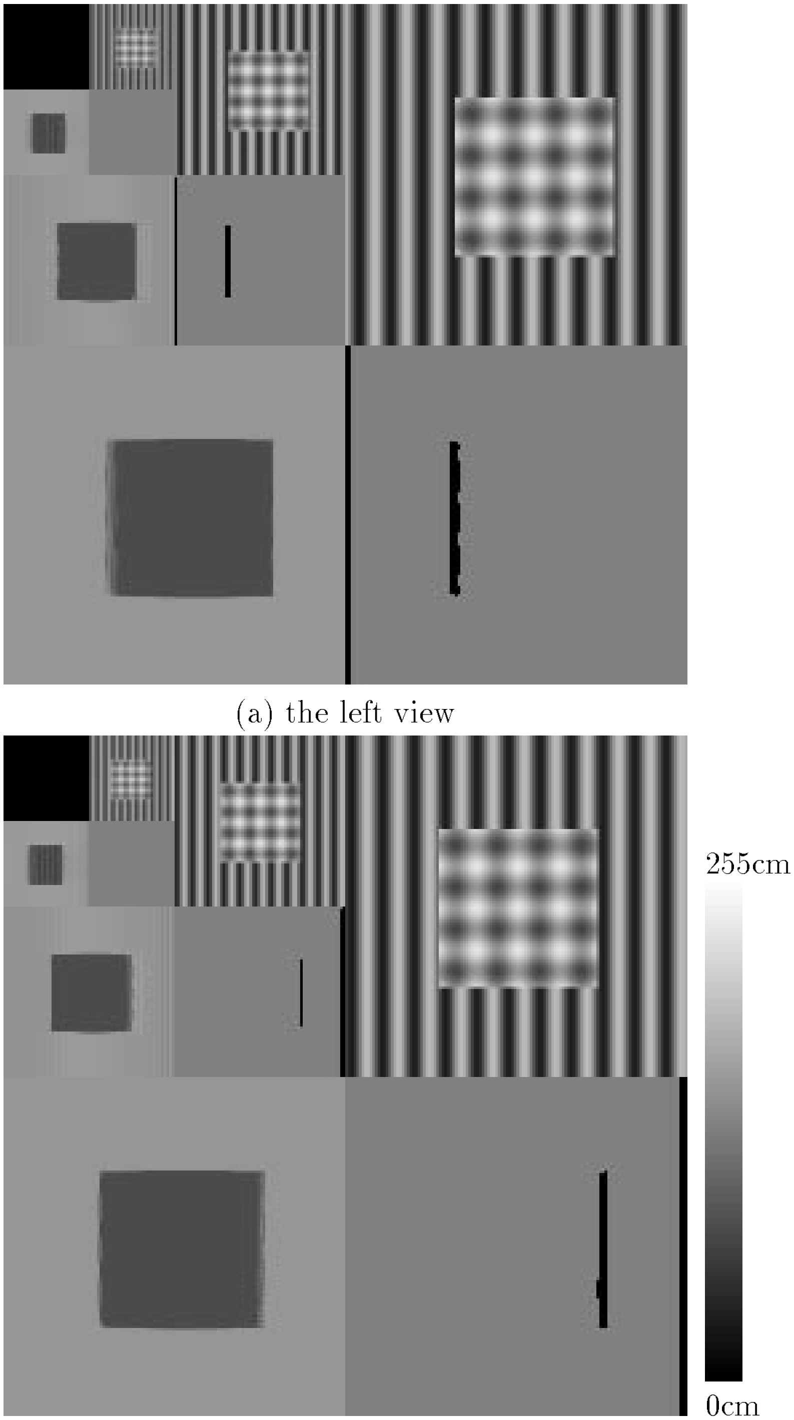
Four subpictures in Fig. 4 represent the results $Z_l(x,y)$ (top-left), $O_l(x,y)$ (top-right), $Z_r(x,y)$ (bottom-left) and $O_r(x,y)$ (bottom-right) of the original resolution, when a Gaussian noise is added to the original synthetic images with a signal to noise ratio of $20dB$. The algorithm shows a good robustness against noise. The average relative error of the whole estimated depth field outside the occluded regions is only 1.05%. The occluded regions are still satisfactorily detected.

Real stereo image pair:

In the same way as in Fig. 3, the figures from Fig. 5 to 9 show the estimated results individually from five pairs of real stereo images, which are grabbed with CCD cameras. The original image format was $512 * 512$ pixels which then was averaged and subsampled into a format of $128 * 128$ pixels. The internal parameters of the stereo cameras in all cases are the same as in the synthetic example, i.e. the focal length $F = 8\text{mm} = 225\text{pixels}$. In the first three examples (Fig. 5 to 7) the external parameters of the stereo cameras are also the same as in the synthetic example, i.e. the baseline length $B = 2cm$, whereas the baseline length $B = 4cm$ (double) and $B = 1cm$ (half) are respectively used in the fourth example (Fig. 8) and in the fifth example (Fig. 9).

The scene in the first pair of stereo images (Fig. 5) contains a poster as background, a Mickey mouse and a cup as objects, which are respectively 121cm, 50cm and 82cm far away from the cameras. The results of this stereo pair with our algorithm are also shown in Fig. 5, which are satisfactory. The occluded regions are qualitatively correct detected. In most of the area, the depth fields are well estimated. Only in the background at the lower-left corner and between the Mickey mouse and the cup there exist large errors of the depth fields because of the absence of horizontal intensity variations. The average relative error of the depth field within the objects and the background is 4.6%.

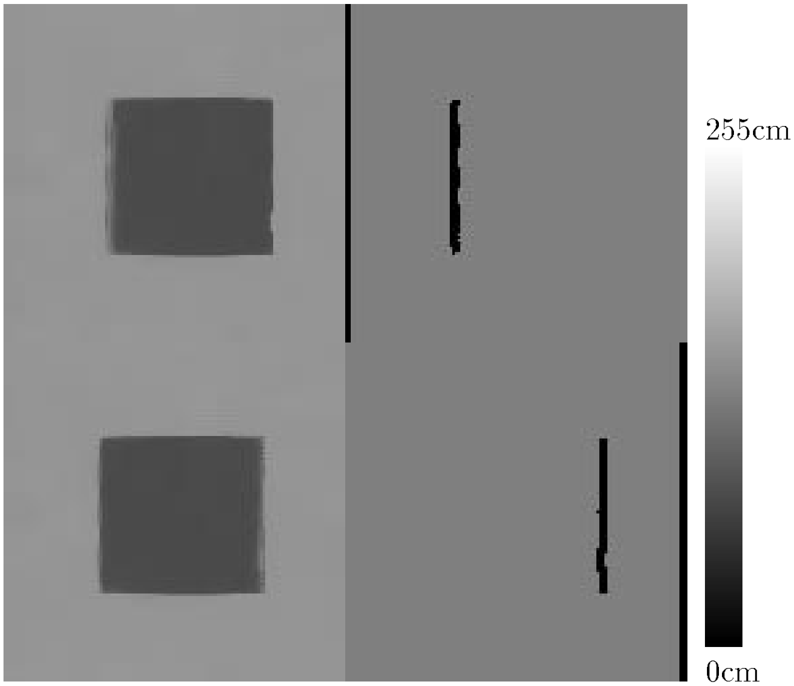
The scene in the second pair of the stereo images (Fig. 6) contains the wall as background, a monitor and an oscillometer etc. as objects. The wall is 220cm far away from the cameras, and the monitor and the oscillometer respectively 70cm and 130cm. The results of the depth and occlusion estimation are shown in Fig. 6. All the occluded regions have been detected. The depth fields and their discontinuities are also well estimated except at the bottom of



(a) the left view

(b) the right view
upper right: the original image
lower left: the depth field
lower right: the occlusion map
upper left: the results with reduced resolutions

Figure 3: The results for a noise-free synthetic stereo image pair



upper left: the left depth field
upper right: the left occlusion map
lower left: the right depth field
lower right: the right occlusion map

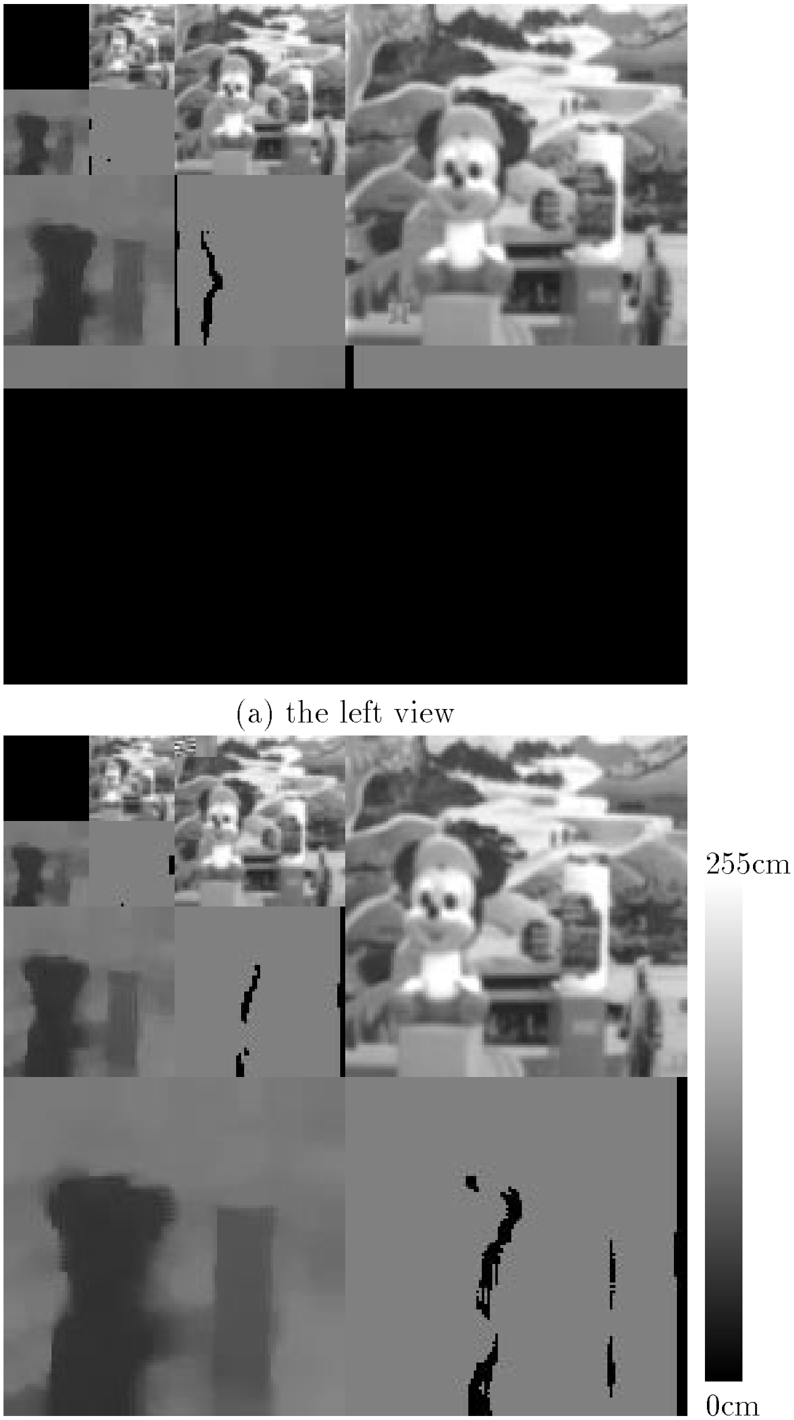
Figure 4: The results for a synthetic stereo image pair with additive noise

and above the monitor. There is the same reason for errors, i.e. absence of horizontal intensity variations in these both regions.

The scene in the third pair of stereo images (Fig. 7) contains the wall and a number of books on the table as background, a toy on the table as object, which are respectively 175cm, 130cm and 82cm far away from the cameras. From the results in Fig. 7 one can see that our algorithm is efficient and robust. The depth fields of almost the whole scene are satisfactorily estimated. Their discontinuities and the corresponding occluded regions of the real scene are qualitatively well detected.

In the scene of the forth stereo image pair (Fig. 8) the wall, posters and a book-shelf etc. are considered as background, which have the distance of $360 - 390$cm to the cameras. A person was sitting in the chair on the ground and his chest is 170cm far away from the cameras. Fig. 8 shows the estimated results of the depth fields and occlusions. All the occluded regions including the occlusions caused by the person are satisfactorily detected. The depth fields of the whole scene except within and above the head are well estimated. The error on the background above the head was caused by the smoothing effect because of no horizontal intensity variations. The small motion of the head leads to some error on it.

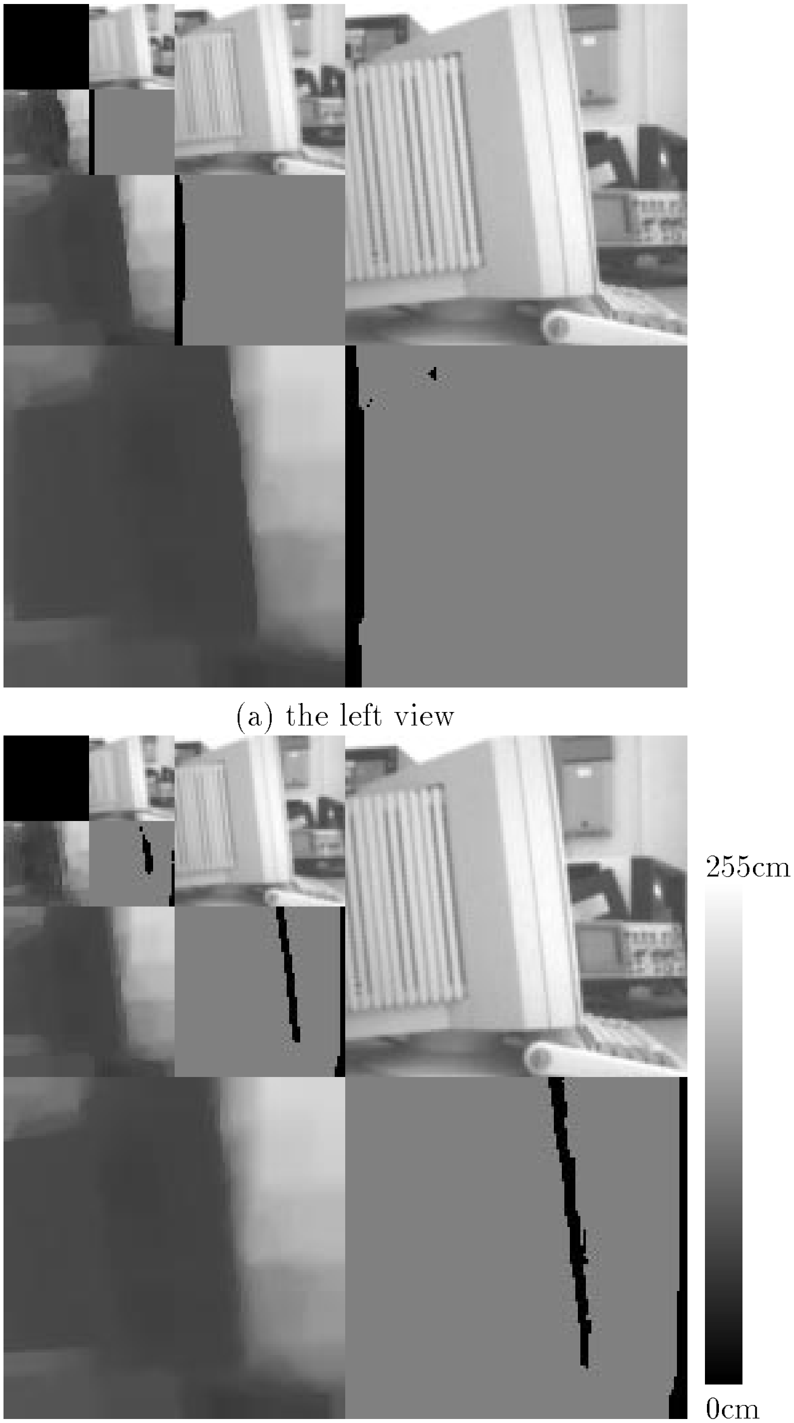
The scene in the fifth pair of stereo images (Fig. 9) is composed by two toy-houses standing on a table in front of a background-poster. These two houses and the poster are respectively 24cm, 50cm and 77cm far away from the cameras. With the new algorithm we obtain the depth estimates and the detected occlusion maps shown in Fig. 9. Most occluded regions of the scene are found, and the whole depth fields except a small background-region



(a) the left view

(b) the right view
upper right: the original image
lower left: the depth field
lower right: the occlusion map
upper left: the results with reduced resolutions

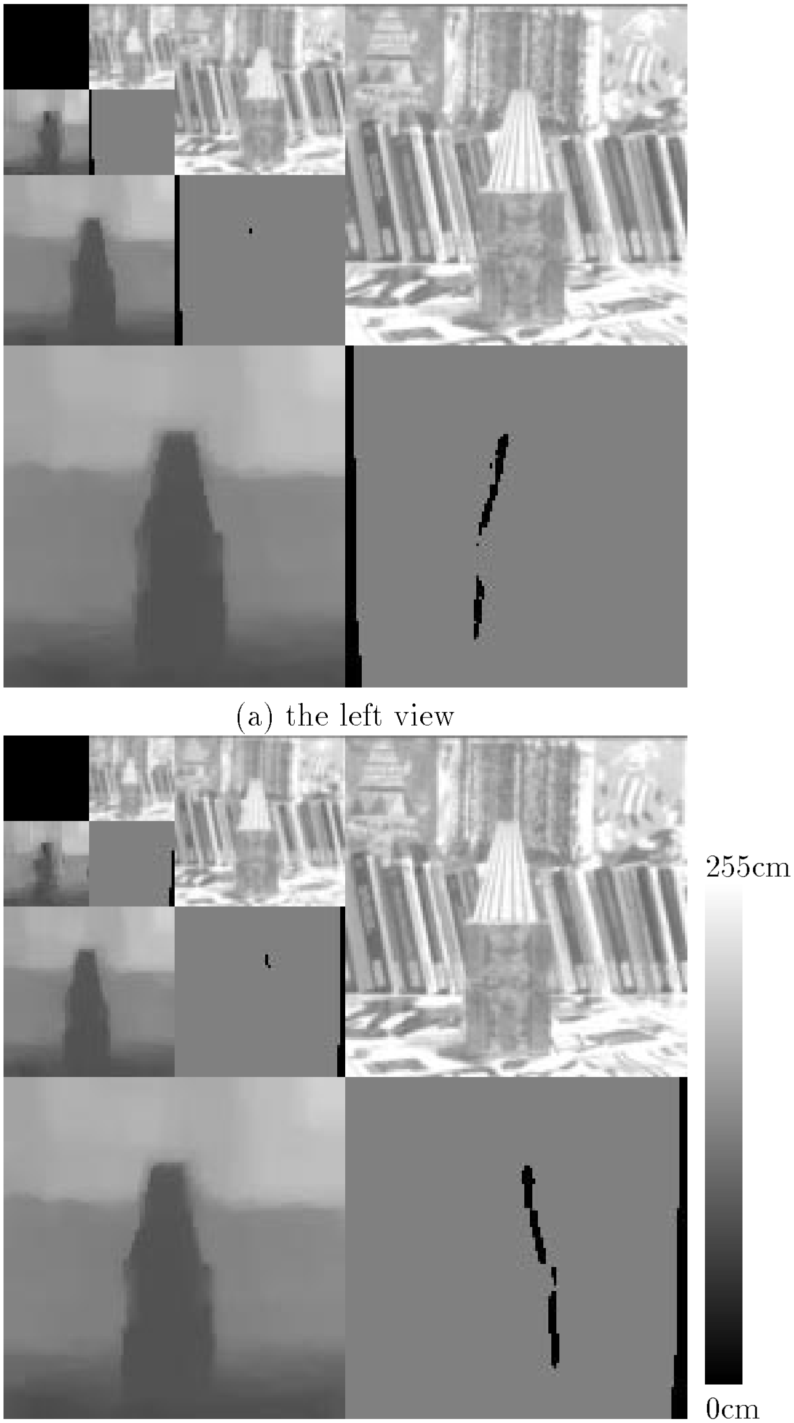
Figure 5: The results for the real stereo image pair 1 with the binocular approach



(a) the left view

(b) the right view
upper right: the original image
lower left: the depth field
lower right: the occlusion map
upper left: the results with reduced resolutions

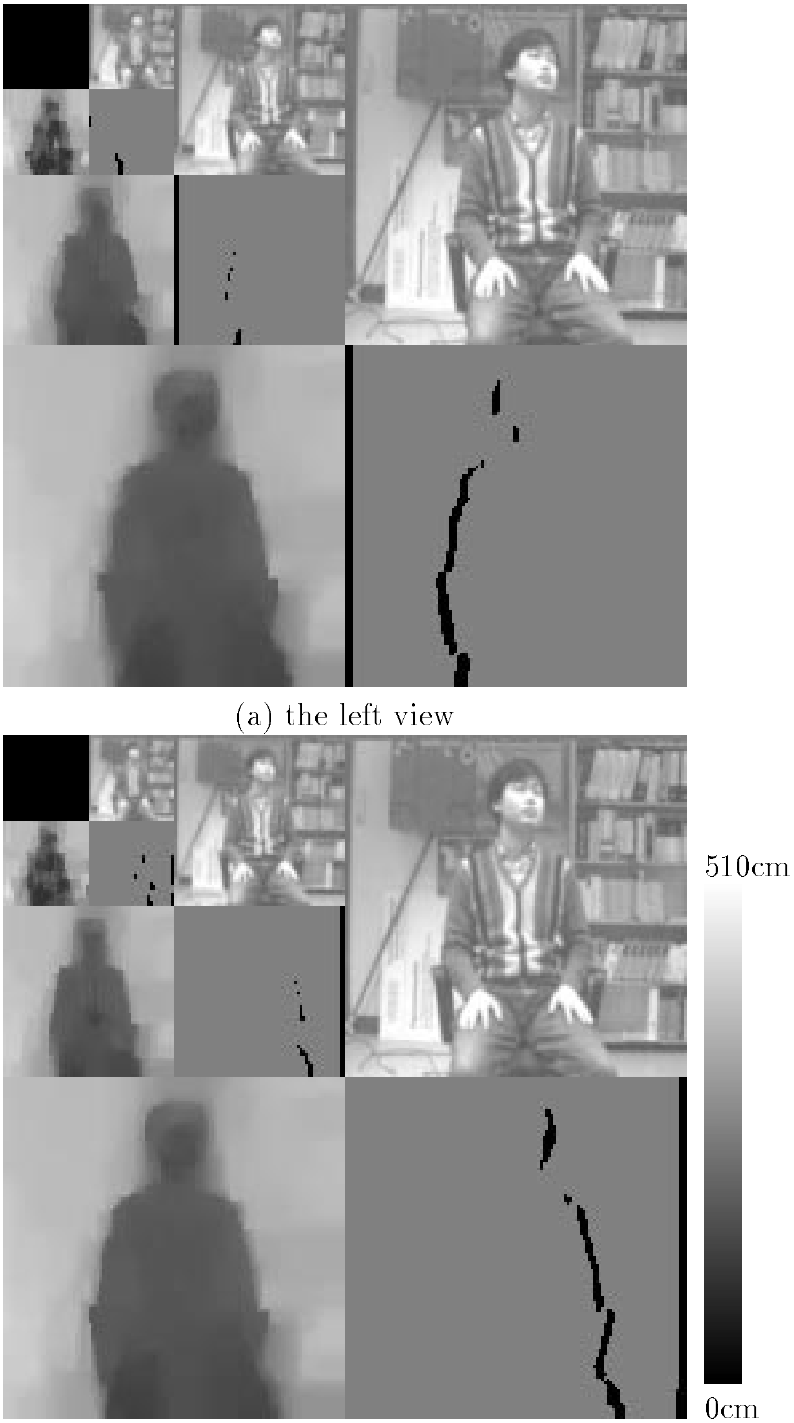
Figure 6: The results for the real stereo image pair 2 with the binocular approach



(a) the left view

(b) the right view
upper right: the original image
lower left: the depth field
lower right: the occlusion map
upper left: the results with reduced resolutions

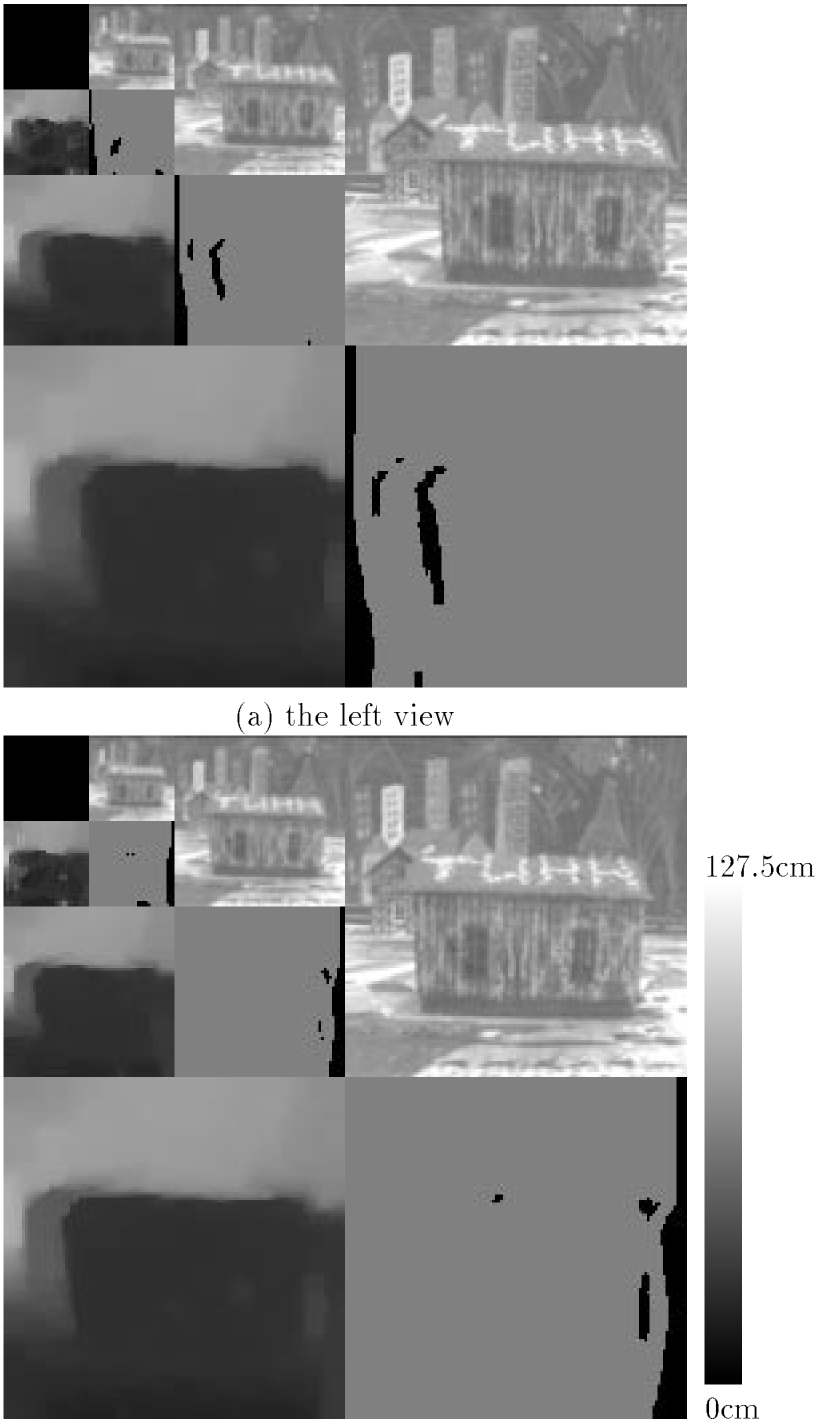
Figure 7: The results for the real stereo image pair 3 with the binocular approach



(a) the left view

(b) the right view
upper right: the original image
lower left: the depth field
lower right: the occlusion map
upper left: the results with reduced resolutions

Figure 8: The results for the real stereo image pair 4 with the binocular approach



(a) the left view

(b) the right view
upper right: the original image
lower left: the depth field
lower right: the occlusion map
upper left: the results with reduced resolutions

Figure 9: The results for the real stereo image pair 5 with the binocular approach

right to the foreground-house are satisfactorily estimated. The depth discontinuities are kept, although some of them produce no occlusions.

Experiments show that almost all of the discontinuities of depth fields have been blurred and no occlusions have been found from these synthetic and real stereo image pairs, if only the standard a priori smooth model tied with (12) is used instead of our new a priori piecewise smooth model and the two-stage algorithm. Thus, the second stage using the new a priori model with the median filter is very important to preserve discontinuities and to detect occlusions. For saving the space, these blurred estimates are not given here.

All of the above five real examples demonstrate the efficiency and robustness of our algorithm estimating depth fields and detecting discontinuities and occlusions from binocular stereo image pairs of real scenes. The occluded regions of these five scenes are effectively detected. The average relative errors of the depth fields within the objects and background outside the occluded regions are below 5%. In most areas the depth fields are satisfactorily estimated. Actually, the performance of this algorithm handling discontinuities depends on the horizontal intensity variations of the areas near these discontinuities and occluded regions. If there are such obvious intensity variations, the discontinuities, whose intervals are not smaller than the size of the median filter window, can be handled. But due to discretization errors, some occluded regions that are not wider than two pixels can be neglected, although the discontinuities of estimated depth fields exist. In the absence of horizontal intensity variations, the triangulation geometry of binocular stereo vision can not detect the accurate disparities on such regions, and the relaxation algorithm can only assign approximate values of their neighborhoods to them, so that the discontinuities may easily be neglected. A further improvement to this problem is discussed in the next section.

6 The trinocular stereo vision: a further improvement

Sometimes there exist large ambiguities for the binocular matching between both stereo images due to the absence of intensity variations on the epipolar lines, e.g. the examples of the real scenes in the previous section. In such a situation, it is difficult to estimate the real depth field from the spatial coherence only by the binocular stereo vision, unless concrete models for objects are assumed. One potential alternative to improve the optimal depth estimation is trinocular stereo techniques using an additional third camera. The basic advantage of trinocular techniques is that the third camera provides an extra epipolar geometry constraint to the stereo matching, such that the ambiguities during the local binocular matching are largely eliminated.

Recently trinocular stereo techniques have been extensively studied [13] [23]. But almost current trinocular approaches are feature-based, and only the features that are sparsely extracted can be considered for matching, where a dense depth map must be obtained later in a following process of the surface reconstruction. However, the new intensity-based binocular approach that was discussed in the previous sections can easily be extended to a intensity-based trinocular approach, where the dense depth map can directly be estimated within a relaxation process. The camera geometry, which is shown in Fig. 10, involves a base camera and two other cameras: right and upper camera displacing the horizontal and vertical direction respectively, and all having axes parallel to each other.

The trinocular approach shows many advantages over the binocular approach: At first, the algorithm does not depend only on the intensity variations on one epipolar line direction and overcomes the difficulty of the matching in the regions having intensity changes only along other directions. Secondly, the algorithm has a better performance against noise, because the relaxation is carried out to match in two directions. Thirdly, the algorithm can eliminate some ambiguities of matching due to some nearly periodical intensity texture. Finally, it is also important to eliminate some difficulties to match in partly occluded regions. The area of occluded regions of the base image with respect to both other images are greatly decreased, and the algorithm can provide an accurate depth estimate in the regions that are occluded in only one image by matching the base image with another image. The new trinocular algorithm and its implementation are given as follows.

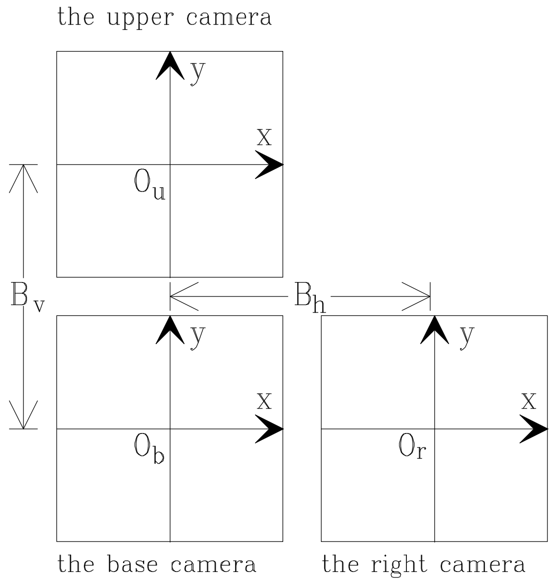


Figure 10: The camera geometry of the trinocular stereo vision

At first we give some notations used below: $g_b(x,y)$, $g_r(x,y)$ and $g_u(x,y)$ stand for the base, right and upper images of a stereo pair respectively, and $Z_b(x,y)$, $Z_r(x,y)$ and $Z_u(x,y)$ for the depth fields in the different image coordinate systems. $O_{br}(x,y)$ and $O_{bu}(x,y)$ stand for the occlusion maps of the base system respectively to the right and upper image, and $O_r(x,y)$ and $O_u(x,y)$ for the occlusion maps of the right or upper system to the base image. For convenience, it is assumed that the vertical and horizontal baseline lengths are equal, i.e. $B_h = B_v = B$. We match the base image simultaneously to the right and upper images to estimate $Z_b(x,y)$, whereas the matchings of the two others back to the base image produce the two other depth fields.

According to the consistent constraint and the piecewise smooth model of depth fields, we can derive the relaxation algorithm of the cooperative bidirectional matching between the three trinocular stereo images, which is similar to the derivation in the binocular case:

$$\begin{aligned} Z_b^{k+1}(x,y) &= \bar{Z}_b^{k+}(x,y) + \lambda O_{br}^k(x,y)(g_b(x,y) - g_r(x_r^+,y))g_{rx}(x_r^+,y)\frac{BF}{(\bar{Z}_b^{k+}(x,y))^2} + \\ & \quad \lambda O_{bu}^k(x,y)(g_b(x,y) - g_u(x,y_u^+))g_{uy}(x,y_u^+)\frac{BF}{(\bar{Z}_b^{k+}(x,y))^2} \qquad (35) \end{aligned}$$

$$Z_r^{k+1}(x,y) = \bar{Z}_r^{k+}(x,y) + \lambda O_r^k(x,y)(g_b(x_b^+,y) - g_r(x,y))g_{bx}(x_b^+,y)\frac{BF}{(\bar{Z}_r^{k+}(x,y))^2} \qquad (36)$$

$$Z_u^{k+1}(x,y) = \bar{Z}_u^{k+}(x,y) + \lambda O_u^k(x,y)(g_b(x,y_b^+) - g_u(x,y))g_{by}(x,y_b^+)\frac{BF}{(\bar{Z}_u^{k+}(x,y))^2} \qquad (37)$$

with the abbreviations $x_r^+ = x - \frac{BF}{\bar{Z}_b^{k+}(x,y)}$, $y_u^+ = y - \frac{BF}{\bar{Z}_b^{k+}(x,y)}$, $x_b^+ = x + \frac{BF}{\bar{Z}_r^{k+}(x,y)}$ and $y_b^+ = y + \frac{BF}{\bar{Z}_u^{k+}(x,y)}$, where

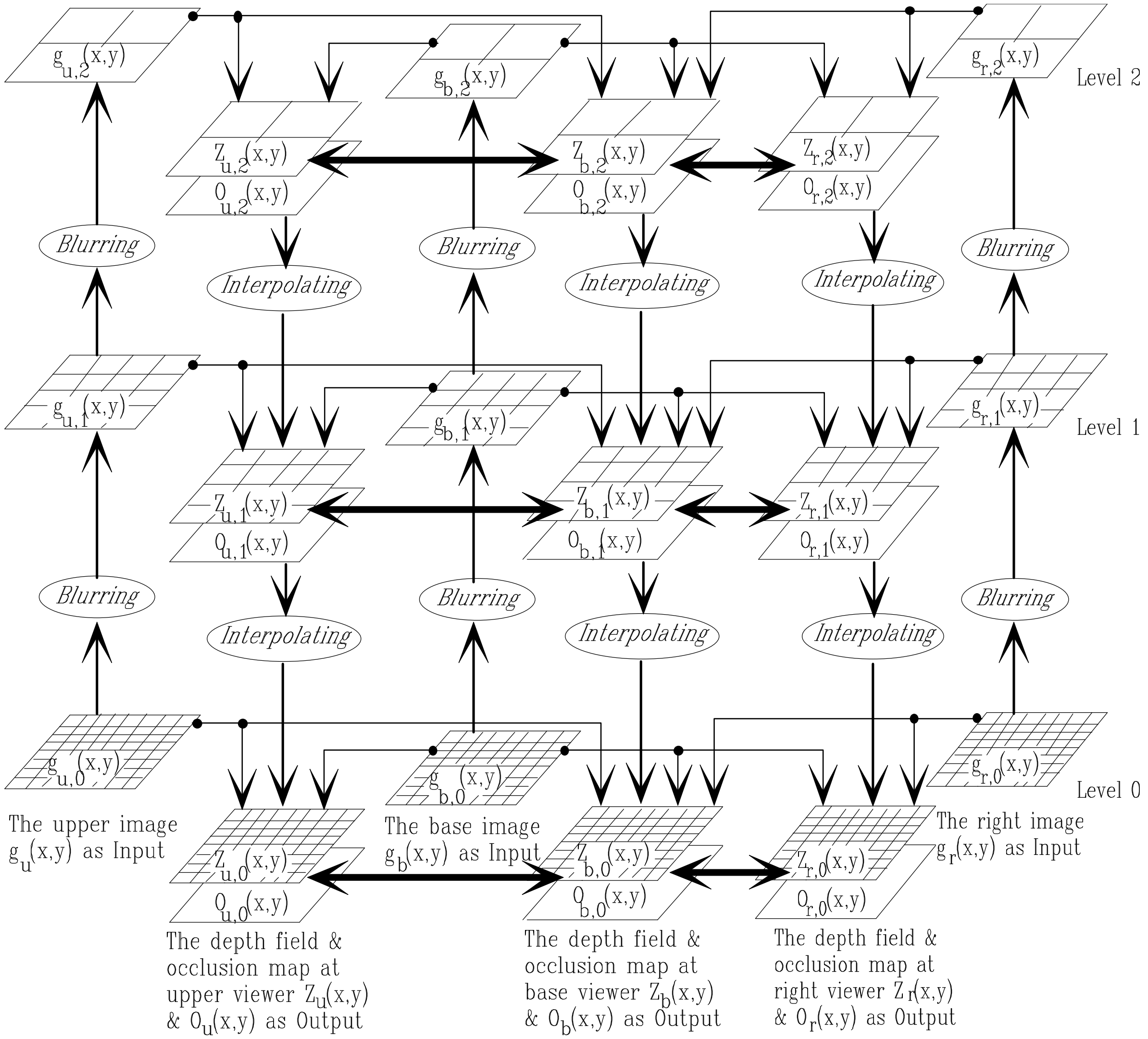


Figure 11: The multilevel implementation of the trinocular stereo algorithm

$$O_{br}^k(x,y) = \mathcal{OT}_{RB}(\bar{Z}_r^k(x,y)) \tag{38}$$

$$O_{bu}^k(x,y) = \mathcal{OT}_{UB}(\bar{Z}_u^k(x,y)) \tag{39}$$

$$O_r^k(x,y) = \mathcal{OT}_{BR}(\bar{Z}_b^k(x,y)) \tag{40}$$

$$O_u^k(x,y) = \mathcal{OT}_{BU}(\bar{Z}_b^k(x,y)) \tag{41}$$

$$\bar{Z}_b^{k+}(x,y) = \frac{\bar{Z}_b^k(x,y) + \bar{Z}_r^k(x_r,y)O_r^k(x_r,y) + \bar{Z}_u^k(x,y_u)O_u^k(x,y_u)}{1 + O_r^k(x_r,y) + O_u^k(x,y_u)} \tag{42}$$

$$\bar{Z}_r^{k+}(x,y) = \frac{\bar{Z}_r^k(x,y) + \bar{Z}_b^k(x_b,y)O_{br}^k(x_b,y)}{1 + O_{br}^k(x_b,y)} \tag{43}$$

$$\bar{Z}_u^{k+}(x,y) = \frac{\bar{Z}_u^k(x,y) + \bar{Z}_b^k(x,y_b)O_{bu}^k(x,y_b)}{1 + O_{bu}^k(x,y_b)} \tag{44}$$

with the abbreviations $x_r = x - \frac{BF}{\bar{Z}_b^k(x,y)}$, $y_u = y - \frac{BF}{\bar{Z}_b^k(x,y)}$, $x_b = x + \frac{BF}{\bar{Z}_r^k(x,y)}$ and $y_b = y + \frac{BF}{\bar{Z}_u^k(x,y)}$, where $\bar{Z}_s^k(x,y)$ are computed according to (23) at the first steps or according to (24) later.

In the algorithm three relaxation processes are cooperatively and dependently carried out. Its implementation with the hierarchical multilevel structure is shown in Fig. 11. The Gaussian pyramids of three stereo images are progressively produced by Eq. (32). The relaxation processes are also progressively carried out from the coarsest level to the finest level, and both the global smooth model (12) and the new piecewise smooth model (13) are successively used at each level. The optimal results in a coarser level provide the good initial estimates for the relaxation in the next finer level by the interpolation processes with Eq. (33) and (34). Initializing the whole depth field with a reasonable constant, e.g. in the range from half to ten times as large as an approximate true depth, and occlusion maps with "1", the algorithm can converge fast to the nearly global optimal estimates of the depth fields and occlusion maps for original stereo scenes in tens of relaxation steps altogether.

Experimental results with the trinocular algorithm

Here we give a real example to illustrate the performance of the new trinocular stereo algorithm. In the image acquisition, the focal length of the stereo cameras with the baseline length of $B = B_h = B_v = 2$cm is the same as in the real examples of section 5, i.e. $f = 8$mm $= 225$pixels. In Fig. 12 (a), (c) and (d) the three stereo images $g_u(x,y)$, $g_b(x,y)$ and $g_r(x,y)$ of the original resolution are shown respectively. The scene contains a poster as background, a Mickey mouse and a book as objects, which are respectively 125cm, 50cm and 95cm far away from the cameras. The corresponding results of the depth fields $Z_u(x,y)$, $Z_b(x,y)$ and $Z_r(x,y)$ with our trinocular algorithm are given in Fig. 12 (e), (g) and (h) individually, and the occlusion maps $O_u(x,y)$, $O_r(x,y)$, $O_{bu}(x,y)$ and $O_{br}(x,y)$ respectively in Fig. 12 (b), (f), (i) and (j) (in the original resolution of the finest level). We can see that the estimates of the depth fields are almost completely consistent with the true 3-D structure, and the discontinuities and occlusions of the 3-D structure are very satisfactorily detected. The average relative error of this estimated depth field within the background and objects is only 3.6%, whereas the error of the estimate that is obtained from the two of the same stereo images with the binocular algorithm is 4.9% (the dense representation of this binocular estimate is omitted here). In Fig. 13 two profiles of depth fields $Z_b(x,y^*)$ and $Z_l(x,y^*)$ with a certain $y = y^*$ in $\frac{2}{5}$ height of the field are given, which are respectively obtained with the trinocular and binocular algorithm. The examples show that the new trinocular stereo algorithm gives great improvements over the binocular stereo algorithm to estimate the 3-D structure of real scenes, but of course, at the expense of higher (but still reasonable) computational costs.

7 Conclusion

In this paper we first discuss the existing problems in stereo vision due to discontinuities and occlusions, and modify the probabilistic model of piecewise smooth depth fields. Then we put forward a two-stage relaxation method to minimize our nonconvex a posteriori energy function derived from the triangulation geometry, the observation model and the new depth field model, which is equivalent to the MAP estimation. Incorporating the consistent constraint of depth fields in different viewer systems, we solve binocular stereo vision problems by a new cooperative bidirectional matching algorithm, which is implemented with a hierarchical multilevel structure. The examples of synthetic and real scenes show that the depth field can reliably be estimated with simultaneous detection of discontinuities and occlusion by our algorithm with low computational costs. To improve the estimation of 3-D structures further, we extend the new algorithm to the trinocular approach, and satisfactory results

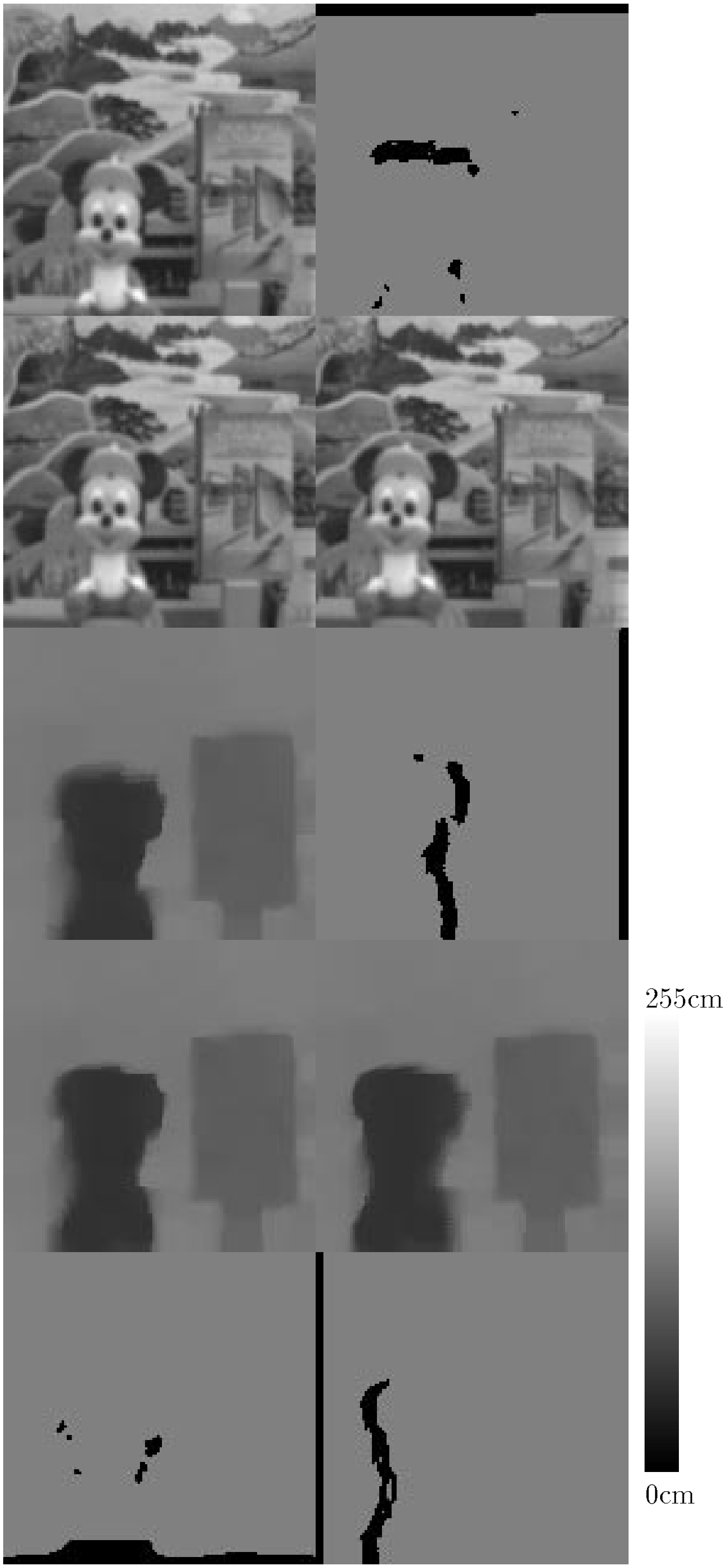


Figure 12: The results of a real trinocular stereo image pair with the trinocular approach (a)-(j) (from left to right then from top to bottom): (a) the original upper image, (b) the occluded regions of the upper image to the base, (c) the base image, (d) the right image, (e) the upper depth field, (f) the right occlusion to the base, (g) the base depth field, (h) the right depth field, (i) the base occlusion to the upper, (j) the base occlusion to the right.

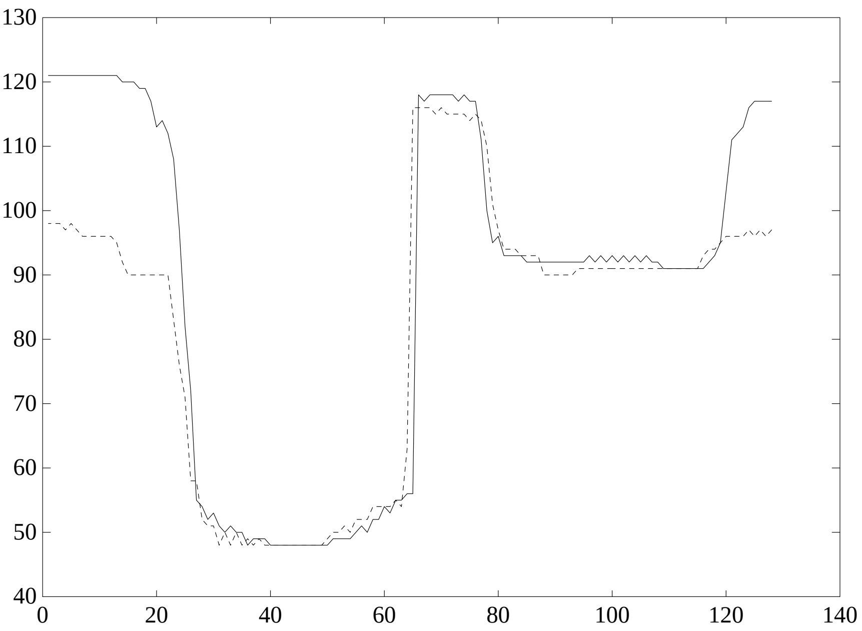


Figure 13: Profiles of both the estimated depth fields $Z_b(x,y)$ and $Z_l(x,y)$, which are obtained respectively with the trinocular (real line) and binocular (dotted line) algorithms, in $\frac{2}{5}$ height of the field (Fig. 12 (g))

are obtained. This computational framework proved to be efficient and robust in stereo problems, and could easily be extended to other problems in low-level vision [18] e.g. optical flow computation [16] and shape from shading [17].

Acknowledgment We would like to thank the reviewers for the valuable comments which helped us to improve the paper.

References

[1] S.T.Barnard, Stochastic stereo matching over scale, Int. Journal of Computer Vision, Vol. 2, 1989, pp. 17-32

[2] A.Blake and A.Zisserman, Visual Reconstruction, MIT Press, Cambridge, MA, 1987

[3] P.J.Burt, The pyramid as a structure for efficient computation, in the book [22], 1984, pp. 6-35

[4] C.Chang and S.Chatterjee, Multiresolution stereo - a Bayesian approach, Proc. of 10th Int. Conf. on Pattern Recognition, Atlantic City, NJ, 1990, pp. 908-912

[5] C.Chang, S.Chatterjee and P.R.Kube, On an analysis of static occlusion in stereo vision, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii, 1991, pp. 722-723

[6] R.Chung and R.Nevatia, Use of monocular groupings and occlusion analysis in a hierarchical stereo system, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Hawaii, 1991, pp. 50-56

[7] W.Enkelmann, Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences, Computer vision, graphics and image processing, Vol. 43, 1988, pp.150-177

[8] S.Geman and D.Geman, Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, IEEE Trans Pattern Anal. Mach. Intelligence, Vol.6, 1984, pp. 721-741

[9] D.Geiger, B.Ladendorf and A Yuille, Occlusions and binocular stereo, Proc. of 2nd Europe Conf. on Computer Vision, Italy, 1992, pp. 425-433

[10] M.A.Gennert, Brightness-based stereo matching, Proc. of 2nd Int. Conf. on Computer Vision, Tampa, FL, 1988, pp. 139-143

[11] F.Glazer, Multilevel relaxation in low-level computer vision, in the book [22], 1984, pp. 312-330

[12] W.E.L.Grimson, Computational experiments with a feature based stereo algorithm, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol.7, 1985, pp. 17-34

[13] C.Hansen, N.Ayache and F.Lustman, Toward real-time trinocular stereo, Proc. of 2nd Int. Conf. on Computer Vision, Tampa, FL, 1988, pp. 129-133

[14] W.Hoff and N.Ahuja, Surface from stereo: Integrating feature matching, disparity estimation and contour detection, IEEE Trans. Pattern Anal. Machine Intell., Vol. 11, 1989 pp. 121-136

[15] B.K.P.Horn, Robot Vision, MIT Press, Cambridge, MA, 1986

[16] A. Luo and H. Burkhardt, Bestimmung des optischen Flusses aus Bildfolgen unter Berücksichtigung von Diskontinuität und Okklusion, In S.J. Pöppl und H. Handels (eds.), *15. DAGM - Symposium "Mustererkennung"*, pp. 59–66, Lübeck, 1993. Reihe Informatik aktuell, Springer.

[17] A. Luo and H. Burkhardt, Shape from Shading (SFS) und Integration von SFS und Stereo unter perspektivischer Projektion, In S.J. Pöppl und H. Handels (eds.), *15. DAGM - Symposium "Mustererkennung"*, pp. 584–591, Lübeck, 1993. Reihe Informatik aktuell, Springer.

[18] A. Luo, Helligkeitsbasiertes Rechnersehen zur direkten Ermittlung räumlicher Eigenschaften, Ph.D. Dissertation, Technische Universität Hamburg-Harburg, December 1993.

[19] D.G.Jones and J.Malik, A computational framework for determining stereo correspondence from a set of linear spatial filters, Proc. of 2nd Europe Conf. on Computer Vision, Italy, 1992, pp. 395-410

[20] D.Marr and T.Poggio, A computational theory of human stereo vision, Proc. Royal Soc. London B. Vol. 204, 1979, pp. 301-328

[21] H.P.Moravec, Robot Rover visual navigation, UMI Research Press, Ann Arbor, MI, 1981

[22] A. Rosenfeld (eds.), Multiresolution Image Processing and Analysis, Springer-Verlage 1984

[23] C.V.Stewart and C.R.Dyer, The trinocular general support algorithm: a three-camera stereo algorithm for overcoming binocular matching errors, Proc. of 2nd Int. Conf. on Computer Vision, Tampa, FL, 1988, pp. 134-138

[24] R.Szeliski, Bayesian modeling of uncertainty in low-level vision, Int. Journal of Computer Vision, Vol.5, 1990 pp. 271-301

[25] D.Terzopoulos, Multilevel reconstruction of visual surface: variational principles and finite-element representation, in the book [22], 1984, pp. 237-310

[26] D.Terzopoulos, Regularization of inverse visual problems involving discontinuities, IEEE Trans. Pattern Anal. Mach. Intelligence, Vol.8, 1986, pp. 413-424

[27] J.Weng, N.Ahuja and T.S.Huang, Two-view matching, Proc. of 2nd Int. Conf. on Computer Vision, Tampa, FL, 1988, pp. 64-73