# Fast and Robust Semi-Local Stereo Matching Using Possibility Distributions

Haythem Ghazouani, Moncef Tagina, René Zapata

▶ **To cite this version:**

Haythem Ghazouani, Moncef Tagina, René Zapata. Fast and Robust Semi-Local Stereo Matching Using Possibility Distributions. International Journal of Computational Vision and Robotics, 2011, 2 (3), pp.237-253. 10.1504/IJCVR.2011.042841 . lirmm-00738216

**HAL Id: lirmm-00738216**

**https://hal-lirmm.ccsd.cnrs.fr/lirmm-00738216**

Submitted on 3 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Fast and robust semi-local stereo matching using possibility distributions

Haythem Ghazouani* and Moncef Tagina

National School for Computer Studies,
SOIE Laboratory,
Campus Universitaire de La Manouba,
La Manouba 2010, Tunisia
E-mail: haythemghz@yahoo.fr
E-mail: moncef.tagina.ensi.rnu.tn
*Corresponding author

René Zapata

LIRMM Laboratory,
University of Montpellier II,
4-6-7, 161 rue Ada, 34392 Montpellier Cedex 5, France
E-mail: zapata@lirmm.fr

**Abstract:** Global stereo matching methods aim to reduce the sensibility of stereo correspondence to ambiguities caused by occlusions, poor local texture or fluctuation of illumination. However, when facing the problem of real-time stereo matching, as in robotic vision, local algorithms are known to be the best. In this paper, we propose a semi-local stereo matching algorithm (SLSM algorithm); an area-based method that embodies global matching constraints in the matching score. Our approach uses a fuzzy formularisation of the similarity assumption in order to define a *matching possibility distribution*. An *unmatching possibility distribution* is defined by applying global constraints to the matching possibility distribution. The final matching cost is computed using the two possibility distributions. Experimental results and comparison with other existing algorithms are presented to demonstrate the performance and effectiveness of our approach.

**Keywords:** semi-local stereo matching; possibility distribution; fuzzy logic; cost function; stereoscopic constraints.

**Reference** to this paper should be made as follows: Ghazouani, H., Tagina, M. and Zapata, R. (xxxx) 'Fast and robust semi-local stereo matching using possibility distributions', *Int. J. Computational Vision and Robotics*, Vol. x, No. x, pp.xxx–xxx.

**Biographical notes:** Haythem Ghazouani received his Masters degree in Computer Science from the National School for Computer Studies of Tunis, Tunisia in 2006. His research interest includes robotics, computer vision and artificial intelligence. Currently, he is a PhD student in the University of Montpellier II, France and the National School for Computer Studies of Tunis, Tunisia. His current research interest includes robot navigation using stereo vision, robot cooperation and map building.

Copyright © 200x Inderscience Enterprises Ltd.

Moncef Tagina is a Professor of Computer Science at the the National School for Computer Studies of Tunis, Tunisia. He received his PhD in Industrial Computer Science from Central School of Lille, France in 1995. He heads research activities at LI3 Laboratory in Tunisia (Laboratoire d'Ingénierie Informatique Intelligente) on metaheuristics, diagnostic, production, scheduling and robotics.

René Zapata is a Professor at the University of Montpellier II, France. He heads research activities at LIRMM Laboratory (Laboratoire d'Informatique, de Robotique et de Micro Electronique de Montpellier) on humanoïd robots, robot planification, vision, map building and localisation.

## 1 Introduction

Stereo vision aims to recover 3D information given incomplete and possibly noisy information of the scene (Horn, 1986; Marr, 1982). Depth (or shape) provided by Stereo vision systems is useful for robot controlling. Stereo correspondence search is the most important and costing task of stereo vision. The stereo matching task can be extremely challenging in presence of disturbance factors which typically affect images. A common source of disturbances can be related to photometric distortions between the images under comparison. These can be ascribed to the camera sensors employed in the image acquisition process (due to dynamic variations of camera parameters such as auto-exposure and auto-gain, or to the use of different cameras), or can be induced by external factors such as changes of the amount of light emitted by the sources or viewing of non-lambertian surfaces at different angles. All of these factors tend to produce brightness changes in corresponding pixels of the two images that can not be neglected in real applications implying stereo correspondence between images acquired from different spatial points (e.g., stereo vision) and/or different time instants (e.g., pattern matching, change detection). In addition to photometric distortions, differences between corresponding pixels can also be due to the noise introduced by camera sensors. Finally, the acquisition of images from different spatial points or different time instants can also induce occlusions. Classical stereo matching methods suffer from weakness in presence of these disturbances. Stereoscopic constraints are used in a refinement step by area-based methods to have higher quality results. Constraints used in stereo matching can be classified into two categories: local constraints, which rely only on a pixel and on some pixels in its surrounding, and global constraints, which must be verified by the whole pixels of a line or of the image. The local methods aim to find a matching for a given pixel without taking into account neighbour pixels correspondences. Global methods try to define a global model of the observed scene and to minimise a global cost function. They try to find the correspondences once for all pixels in one line or for all pixels in the image. This work presents an area-based method that uses global constraints which we classified as semi-local method. Similarity measures used in this approach are based on fuzzy logic to have more robustness against disturbance factors.

The paper unfolds as follows. Section 2 presents a brief review state of the art about stereo matching, focused on the local and global methods. Section 3 defines and explains two possibility distributions that will be used to compute the matching cost. Section 4 presents the main algorithm, and Section 5 presents results of the proposed

algorithm and comparison with existing methods, and Section 6 closes with the main contributions, drawbacks and possible extensions of this work.

## 2 Stereo matching: state of the art

The stereo vision aims at retrieving depth and shape of objects on a scene perceived by two or more images acquired at the same moment from different points of view. The stereo reconstruction is based on the aptitude to find in each image the projection of the same object in the scene. The recovery of depth information of an object is related from one side to the disparity which is the difference between projections of the same object in both images, and on the other side, to the relative position of both cameras (baseline) and to the image resolution. Thus, two different problems raise from the stereoscopic reconstruction. The first is the disparity calculation, which is attached to the problem of stereo correspondence. The second problem is the ability to inverse the projective geometry problem. In other words, the 3D reconstruction, or how to exploit disparity knowledge and relative position of the two sensors to find tridimentional information. The works of Faugeras (1993), on the projective geometry have established a solid basis for the tridimentional reconstruction problem. For the matching problem, there is no method sufficiently reliable, robust and effective that allows a simple use of stereo vision as a sensor of depth measurement. Binocular stereo vision uses two images acquired by two cameras. A preliminary phase of calibration is needed to estimate the different parameters of a stereo rig: the parameters of the projection model for each camera (pinhole geometric model) and the spatial relationship between the two cameras. This knowledge allows us to calculate the 3D coordinates of a point from its projections in the two images by a simple triangulation. Stereo matching is one of the most studied topics in computer vision since more than half a century (Julesz, 1962). A detailed taxonomy of stereo correspondence algorithms is proposed in Scharstein and Szeliski (2002). The authors classify stereo matching methods with respect to four criteria: the local matching cost, the aggregation area while computing the local cost, the optimisation method, and the method performed to refine matching results.

### *2.1 Local stereo matching methods*

In local stereo matching methods, best match of each pixel is searched separately starting from one image (the reference image) without taking into account the matches of other ones. The matching cost between two pixels is based on similarity measurements of the local intensity function. Intuitively, the projections of the same physical point will naturally have similar intensities in the two images. In fact, the Lambertian model (Horn, 1986) assumes that the object surface reflects uniformly the light in all directions. Using this model, we can suppose that the corresponding pixels in both images are similar, and indeed, their neighbours are also similar, assuming that view fields between the two cameras are very close (no or small occlusions). A correlation measurement can calculate a degree of similarity between two point sets. Local methods try to find a match $p_2$ in the right image for a point $p_1$ in the left one. The correlation measurement uses information given by $p_1$, $p_2$ and their neighbour pixels. The pixel $p_1$ and its neighbours form a first point set, and the point $p_2$ and its neighbours constitute the other point set. A correlation score evaluates the similarity between these datasets. Many

local approaches use comparison windows centred on the considered pixel. Among the most known measurements, we can find: sum of squared differences (SSD) (Cox et al., 1996), sum of absolute differences (SAD) (Hirschmüller, 2001), zero-mean normalised cross-correlation (ZNCC) (Chen and Medioni, 1999; Sára, 2002), etc.

There are several local stereo matching algorithms, generally classified into two categories: area matching and/or feature (element) matching (Horn, 1986). Area matching algorithms are characterised by comparing features distributed over regions. Feature matching uses local features, edges and borders for instance, with which it is possible to perform the matching. Area-based algorithms are usually slower than feature-based ones, but they generate full disparity maps and error estimates. Area-based algorithms usually employ correlation estimates between image pairs for generating the match. Such estimates are obtained using discrete convolution operations between images templates. The algorithm performance is, thus, very dependent on the correlation and on the search window sizes. Small correlation windows usually generate maps that are more sensitive to noise, but less sensitive to occlusions, better defining the objects (Hirschmüller, 2001).

Local stereo correspondence methods are in general fast algorithms, so can be used for real time applications. However, they are exposed to many failure sources, in particular occlusions or variations of intensity between the two images. In fact, these situations can produce many false matches. In addition, because of the absence of any constraint between matches, adjacent pixels can have very different disparities, which can be particularly remarkable in scenes having vertical lines (edges of an open door for example). Global methods try to overcome these problems.

### *2.2 Global stereo matching methods*

Global matching methods give very accurate results but are very time and computational demanding because of their iterative nature. In global stereo matching approaches, match of each pixel is computed based on the match of other pixels in the image. Global approaches try to define a global model of observed scene and to minimise a global cost function. Matches for pixels of one line or pixels of the whole image, are searched at the same time. In a global method, the matching between a pixel in the left image and a pixel in the right image does not depend only on their neighbours, but also on the matches of their neighbours. Hence, the match of a pixel influences the matches of its neighbour pixels. This influence is modelled by regularisation constraints on the matches set. Some methods are based only on the epipolar constraint to transform the bidimensional matching problem into one-dimensional problem, as in dynamic programming (Belhumeur, 1996; Cox, 1992). Other methods address the bidimensional problem by taking into account, inter-linesnal problem by taking into account, inter-lines constraints, i.e., compatibility constraints between matchings provided on every epipolar lines, as in graph cuts algorithms (Boykov et al., 1998; Ishikawa and Geiger, 1998). The global regularisation aims to reduce the sensibility of stereo correspondence to ambiguities caused by occlusions, poor local texture or fluctuation of illumination. This improvement has a cost, which is the increasing of algorithms complexity, and in consequence, a longer execution time, in addition to some secondary effects due to this regularisation (smoothing). One of the most important global approaches is stereo matching using dynamic programming (Ohta and Kanade, 1985; Cox et al., 1996; Bobick and Intille, 1999). The principle of dynamic programming, introduced by

Bellman (1957), allows resolving optimisation problems having an objective function as a sum of monotone non-decreasing functions of resources. In practice, this means that we can infer the optimal solution of a problem using optimal solutions of sub-problems. The dynamic programming applied to stereo matching searches for a path of minimal cost through a matrix composed of possible matches. To reduce the complexity, this technique is applied on two sets of points of the same epipolar line. Thus, the stereo correspondence is applied successively to find matchings for all pixels of a line of one image with pixels located on its epipolar line in the other image (Ohta and Kanade, 1985). The majority of stereo matching dynamic programming methods try to match pixels between epipolar lines in both images without taking into account inter-lines consistency. Hence, they do not use the bidimensional nature of the problem. To overcome this drawback, and to take into consideration bidimensional continuity constraint, a solution has been proposed using the graph theory. Graph cuts applied to stereo matching were proposed by Roy and Cox (1998), and then reformulated by Veksler (1999) in which the matching problem is considered as a minimisation of an energy function. The first global method based on graph cuts for stereo correspondence were given (Roy and Cox, 1998; Ishikawa and Geiger, 1998; Roy, 1999). Afterwards, the iterative graph-cuts algorithms were introduced in Kolmogorov and Zabih (2001, 2002a, 2002b).

To resume, the matching problem can be seen as a minimisation problem. Local approaches try to minimise separately many energy functions, representing local matching costs supposed independent between different entities to be matched: a local cost depends on similarity constraints between these entities. Global approaches try to minimise a unique energy function taking into account all matching costs: this global cost integrates local matching costs, and also compatibility costs expressing how consistent are matchings computed on a line or on the whole image. Local methods are very fast and fit to real time applications but are very sensible to disturbances. Global methods gives more accurate results in a longer execution time that explodes exponentially if high image resolution is used. In our approach, we try to maximise separately many energy functions, representing global matching costs calculated using possibility distributions that will be detailed in next section.

## 3 Possibility distributions

In stereo matching problems, we have a pair of pictures of the same scene taken from different positions, and possibly orientations, and the goal is to discover corresponding points, that is, pixels in both images that are projections of the same scene point. The most intuitive way of doing that is by comparing groups of pixels of the two images to obtain a similarity value. After similarities are computed, one may or may not include restrictions and calculate the matching that maximises the global similarity. Our proposal assumes

1 ordering constraint

2 uniqueness of the correct matching.

In general, given a point in one image, the comparison is not made with all points of the other image. Using the epipolar restriction, only pixels on a certain line in one image are the corresponding candidates of a pixel in the other one. The orientation of

this line depends only on the relative orientation of the two cameras. The test images used in the current work have a horizontal epipolar line, thus pixels are searched only in such direction. We suppose that the rectification is exact, which means that the disparity depends only on the column index of the pixel: the pixel $(r, c)$ in the left image is matched to the pixel $(r, c + d)$ in the right image. We define a 3D disparity space which dimension are $r$, $c$ and $d$ respectively to designate row, column and disparity. Each element $(r, c, d)$ of the disparity space is projected to the pixel $(r, c)$ in the reference image and $(r, c + d)$ in the matching image. The element $(r, c, d)$ refers to the pairing of the pixel $(r, c)$ of the reference image and the pixel $(r, c + d)$ of the matching image. $\gamma(r, c, d)$ refers to the matching cost of the pairing $(r, c, d)$.

Local stereo matching methods search separately for the best match of each pixel starting from one image (e.g., the left one) without taking into account the matches of other ones. The matching cost between two pixels is based only on similarity measurements of the local intensity function. However, in a global method, the matching between a pixel in the left image and a pixel in the right image does not depend only on their neighbours, but also on the matches of their neighbours. Hence, the match of a pixel influences the matches of its neighbour pixels. This influence is modelled by regularisation constraints on the matches set.

In many applications, where high precision is needed, local stereo matching methods need a step of refinement to include the stereoscopic constraints which is time costing. In our stereo matching method, stereoscopic ordering and uniqueness constraints are included directly in the matching score. The matching score between two pixels depends on the similarity measurement and on the matches of their neighbours. Whence the appellation of semi-local matching method.

Using fuzzy sets and based on the similarity assumption, we define a possibility of matching distribution. Based on this distribution and using the uniqueness and ordering constraints we define a possibility of unmatching distribution. The matching and unmatching possibility distributions are used to define the matching score.

### *3.1 Possibility of matching*

Similarity assumption assumes that the projections of the same physical point have comparable light intensities. The major part of stereo vision approaches uses this assumption in a statistic similarity criterion that calculates the difference of illumination between two areas around the two pixels tested for the correspondence. This statistic measure based on numerical distances of intensities is particularly disturbed by changes caused by non-ergodic phenomena like the change of the point of view, partial occlusion, sampling, scanning..., which can hardly be modelled by simple normal laws. In our approach, we propose to model the similarity assumption by a fuzzy measure more robust to noise and changes. This measure expresses the degree of membership of two pixels to a same grey class. We define a grey scale classification of pixels. Three classes are defined; *black pixels*, *white pixels* and *average pixels*. Membership functions of these grey classes, given (1), are Gaussian centred in 0, 127.5 and 255.

$$\mu_{class}(m) = exp\left(-\frac{(I(m) - c_{class})^2}{2\sigma^2_{class}}\right) \tag{1}$$

$I(m)$ is the intensity at the pixel m, $c_{class}$ and $\sigma_{class}$ are respectively the centre and the standard deviation of the class under consideration. Based on this classification we give the following proposition.

Proposition 3.1 The pairing of two pixels $m_1$ and $m_2$ projections of the same physical point M on the stereo images, is 'possible' if the two pixels belong to the same grey class. That means ($m_1$ is *black* AND $m_2$ is *black*) OR ($m_1$ is *white* AND $m_2$ is *white*) OR ( $m_1$ is *average* AND $m_2$ is *average*).

*Definition 3.1:* Considering two pixels $m_1$ and $m_2$ from the two stereo images, we define $\Pi(m_1, m_2)$ as the *possibility of matching* between the two pixels by expressing Proposition 3.1 using classical fuzzy logic operators. $\Pi(m_1, m_2)$ is a measure of co-membership to a same grey class. It reflects how much it is 'possible' to have $m_1$ and $m_2$ as corresponding pixels. $\Pi(m_1, m_2)$ is given by (2).

$$\Pi(m_1, m_2) = \max \begin{pmatrix} \min(\mu_{black}(m_1), \mu_{black}(m_2)), \\ \min(\mu_{average}(m_1), \mu_{average}(m_2)), \\ \min(\mu_{white}(m_1), \mu_{white}(m_2)) \end{pmatrix} \tag{2}$$

$\mu_{class}(m)$ is the degree of membership of the pixel $m$ to the class under consideration. The possibility of matching ranges between 0 and 1.

*Notation:* Thereafter, we will use the notation: $\Pi(r, c, d)$=$\Pi(m_1, m_2)$ with $m_1 = (r, c)$ and $m_2 = (r, c, d)$.

### *3.2 Possibility of unmatching*

Supposing that the observed objects are opaque and the disparity is not significant, the uniqueness constraint assumes that an object whose projection is a pixel on the first image has a projection that is a pixel in the second image. Using such a constraint reduces the number of potential matches of a pixel in the reference image. This constraint can be used only to modify an initial pairing distribution to have a new distribution with less violation to the uniqueness constraint. Referring to the possibilities of matching, a match $(r, c, d)$ violates the uniqueness constraint if there is a match $(r, c, d')$ with $d' \neq d$ and $\Pi(r, c, d) < \Pi(r, c, d')$.

*Definition 3.2:* Considering two pixels $m_1 = (r, c)$ and $m_2 = (r, c + d)$ from the two stereo images, we define $\Pi_U\overline{(r, c, d)}$ as the *possibility of unmatching relatively to the uniqueness constraint*. $\Pi_U\overline{(r, c, d)}$ reflects how much the pairing of the pixel $(r, c)$ in the reference image and the pixel $(r, c + d)$ in the matching image violates the uniqueness constraint. $\Pi_U\overline{(r, c, d)}$ is given by (3).

$$\Pi_U\overline{(r, c, d)} = \sup_{d' \neq d} \{\Pi(r, c, d') > \Pi(r, c, d)\} \tag{3}$$

Under some conditions defined in Faugeras (1993), the ordering of pixels is preserved across the images. This constraint can be formulated by the following proposition.

Proposition 3.2 Considering two pixels $m_1$ and $m_2$, respectively from the reference and the matching image. If $m_1$ and $m_2$ are projections of the same physical point $M$ then all the pixels on the right (respectively left) of the pixel $m_1$ are on the right (respectively left) of the pixel $m_2$.

By extension we can express the dual negative proposition as follow.

Proposition 3.3 Considering two pixels $m_1$ and $m_2$, respectively from the reference and the matching image. If $m_1$ and $m_2$ are projections of the same physical point $M$ then all the pixels on the right (respectively left) of the pixel $m_1$ can not be on the left (respectively right) of the pixel $m_2$.

In other words, a match $(r, c, d)$ violates the ordering constraint if there is a match $(r, c', d')$ that verifies: $(c < c' \ \ AND \ \ c + d > c' + d') \ \ OR \ \ (c > c' \ AND \ c + d < c' + d') \ AND \ (\Pi(r, c, d) < \Pi(r, c', d'))$. Based on this analysis, we give the following definition.

*Definition 3.3:* Considering two pixels $m_1 = (r, c)$ and $m_2 = (r, c + d)$ from the two stereo images, we define $\Pi_O\overline{(r, c, d)}$ as the *possibility of unmatching relatively to the ordering constraint*. $\Pi_O\overline{(r, c, d)}$ reflects how much the pairing of the pixel $(r, c)$ in the reference image and the pixel $(r, c + d)$ in the matching image violates the ordering constraint. $\Pi_O\overline{(r, c, d)}$ is given by (4).

$$\Pi_O\overline{(r, c, d)} = \max \begin{pmatrix} \sup\limits_{\substack{c' > c \\ d' < d - (c' - c)}} \{\Pi(r, c', d') > \Pi(r, c, d)\} \\ \sup\limits_{\substack{c' < c \\ d' > d + (c - c')}} \{\Pi(r, c', d') > \Pi(r, c, d)\} \end{pmatrix} \tag{4}$$

*Definition 3.4:* We define the *global unmatching possibility* by merging unmatching possibilities relatively to uniqueness constraint and ordering constraint. Global unmatching possibility, given by (5), expresses how much the pairing of the pixel $(r, c)$ in the reference image and the pixel $(r, c + d)$ in the matching image violates the stereoscopic constraints.

$$\Pi\overline{(r, c, d)} = \max(\Pi_U\overline{(r, c, d)}, \Pi_O\overline{(r, c, d)}) \tag{5}$$

## 4 Semi-local fuzzy stereo matching

We assume here a binocular stereo pair and images in standard form, i.e., with corresponding epipolar lines lying on corresponding image scanlines. Should the latter assumption not be verified, a suitable transformation, known as rectification (Fusiello et al., 2000; Trucco and Verri, 1998) can be applied to obtain a pair of images in standard form from the original ones. Hence, in local algorithms, given a point in the reference image, the homologous point is selected by searching along the corresponding scanline in the other image, and within a certain disparity range, for the point that minimises (maximises) an error (similarity) function, representing the degree of dissimilarity (similarity) between two small regions centred at the points under examination.

### *4.1 Matching cost and disparity computation*

The goal of improving the reliability of the disparity measurements provided by stereo matching algorithms can be accomplished by introducing constraints to the basic matching core in a refinement step. Since we are interested in a fast algorithm, suited to realtime stereo applications, a major guideline of this work has been to include constraints in the matching cost. To this end, rather than carry out additional calculations, we try to exploit the information related to match reliability which is already embodied into unmatching possibilities used to compute the matching score.

Suppose that $\gamma(r, c, d)$ is the matching cost between a window of size $(2n+1)(2n+1)$ centred at coordinates $(r, c)$ in the left image and the corresponding window centred at $(r, c+d)$ in the right image:

$$\gamma(r,c,d) = \frac{1}{(2n+1)(2n+1)} \sum_{i,j=-n}^{n} \frac{\Pi(r+i, c+j, d)}{1+\Pi(r+i, c+j, d)} \tag{6}$$

We use the matching costs to determine a disparity value $d(r, c)$ for each pixel $p = (r, c)$ of the reference view. We use the winner- takes-all strategy. Hence, the disparity is computed by

$$d(r,c) = \arg\max_{d \in D} \gamma(r,c,d) \tag{7}$$

where $D$ represents the set of all allowed disparities.

### *4.2 Occlusion handling*

Up to this point we have ignored the occlusion problem. Occlusion is a critical and difficult phenomena to be dealt with by stereo algorithms. With any reasonably complex scene there exist occluded pixels that have no correct match. In order to detect occlusions, we try to identify occlusions by examining the magnitude of the converged match costs. Since no correct match exists in areas of occlusion, all match costs corresponding to occluded pixels should be small since uniqueness and ordering constraints are embodied into matching score. In addition, provided mutually occluded areas within the disparity range generally do not have similar intensities, all match costs corresponding to occluded pixels will be small. If a matching cost is below a threshold, the pixel is labelled as occluded.

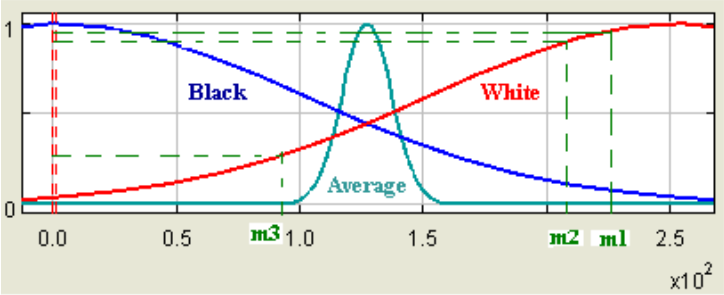
## 5 Implementation and results

In our work, images are acquired by a pre-calibrated stereo rig. In addition, our stereo matching algorithms considers that the left and right images are rectified, which means that epipolar lines are horizontal and have the same line index in the left and right images. Thus, the rectified images come from the pre-calibrated cameras are the input of the stereo algorithms. We suppose that the rectification is exact, which means that the disparity depends only on the column index of the pixel: the pixel $(r, c)$ in the left image is matched to the pixel $(r, c+d)$ in the right image. The proposed technique was implemented as a C++ library and a collection of test programmes. This library generates disparity maps using the default correlation methods and our approach.

### 5.1 Fuzzy sets definition

For all the experiments, we set the standard deviations of the pixel classes defined for equation (1) as follow: $\sigma_{black} = \sigma_{white} = 7.071$ and $\sigma_{average} = 2.236$. These values were empirically determined after many experiments on MATLAB fuzzy toolbox and give significant values of the matching possibilities. Figure 1 shows the membership functions for three grey classes. The two pixels $m_1$ and $m_2$ have comparable light intensities and the possibility of matching is $\Pi(m_1, m_2) = 0.9$, however for $m_1$ and $m_3$, $\Pi(m_1, m_3) = 0.26$.

**Figure 1** Membership functions of the *black*, *white* and *average* grey classes (see online version for colours)




### 5.2 Disparity estimation

The algorithm presented is exploited to its maximum capacity in terms of the stereo correspondence estimation performance. Four popular synthetic images are chosen from the database of the University of Middlebury. The relevant disparity maps are shown in Figures 2 and 3. In addition, error images are also calculated for each of the estimated disparity maps that simply are the absolute difference, defined in (8), between the ground truth and estimated disparity maps in terms of grey scale intensity values, as shown in Figures 2(d) and 3(d). The absolute error can be expressed as

$$E = |d_G(r,c) - d_E(r,c)|_{\forall r,c} \tag{8}$$

where $d_G(r,c)$ is the discrete ground truth disparity map, whereas $d_E(r,c)$ is the estimated one. In order to find the statistical deviation of the estimated disparity maps from the provided ground truth disparity, two statistics are calculated as

$$E = \sqrt{\frac{1}{N}\sum_{r,c} |d_G(r,c) - d_E(r,c)|^2} \tag{9}$$

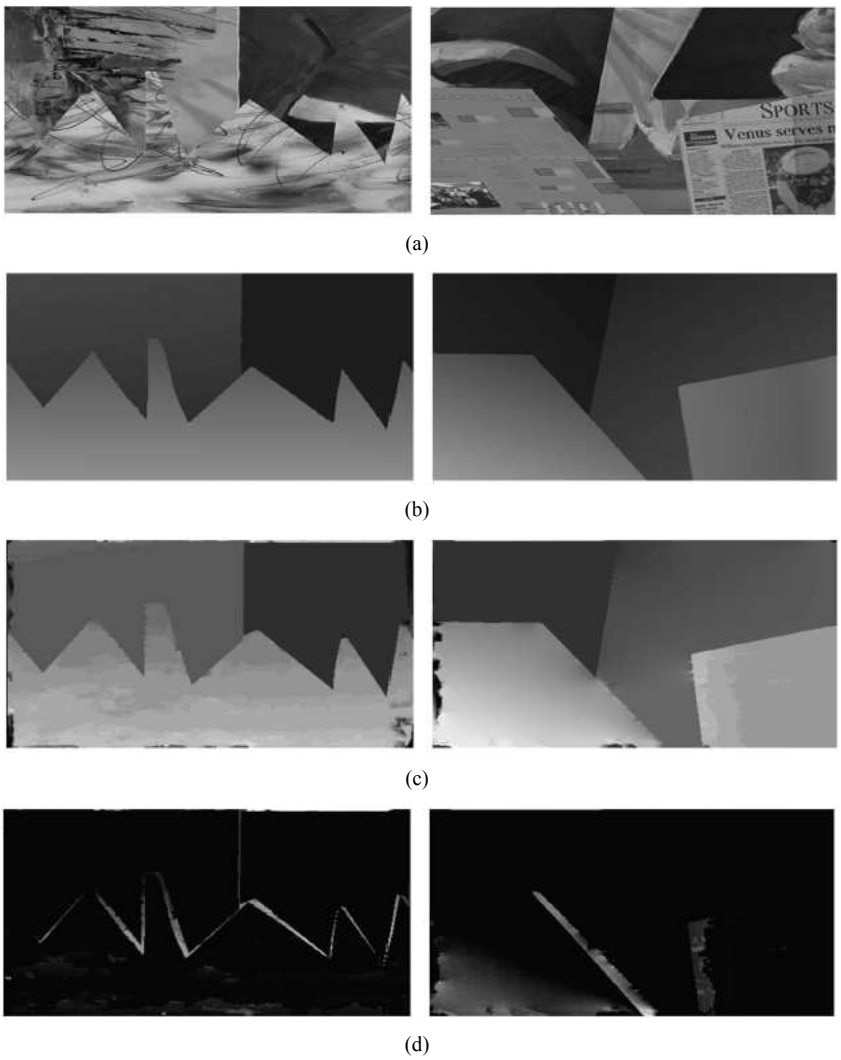
and

$$B = \frac{1}{N}\sum_{r,c} |d_G(r,c) - d_E(r,c)|^2 > \xi \tag{10}$$

where $R$ and $B$ represent the *root mean squared error* (RMSE) and *percentage of bad disparities* (PBD), respectively. $N$ represents the total number of pixels in the

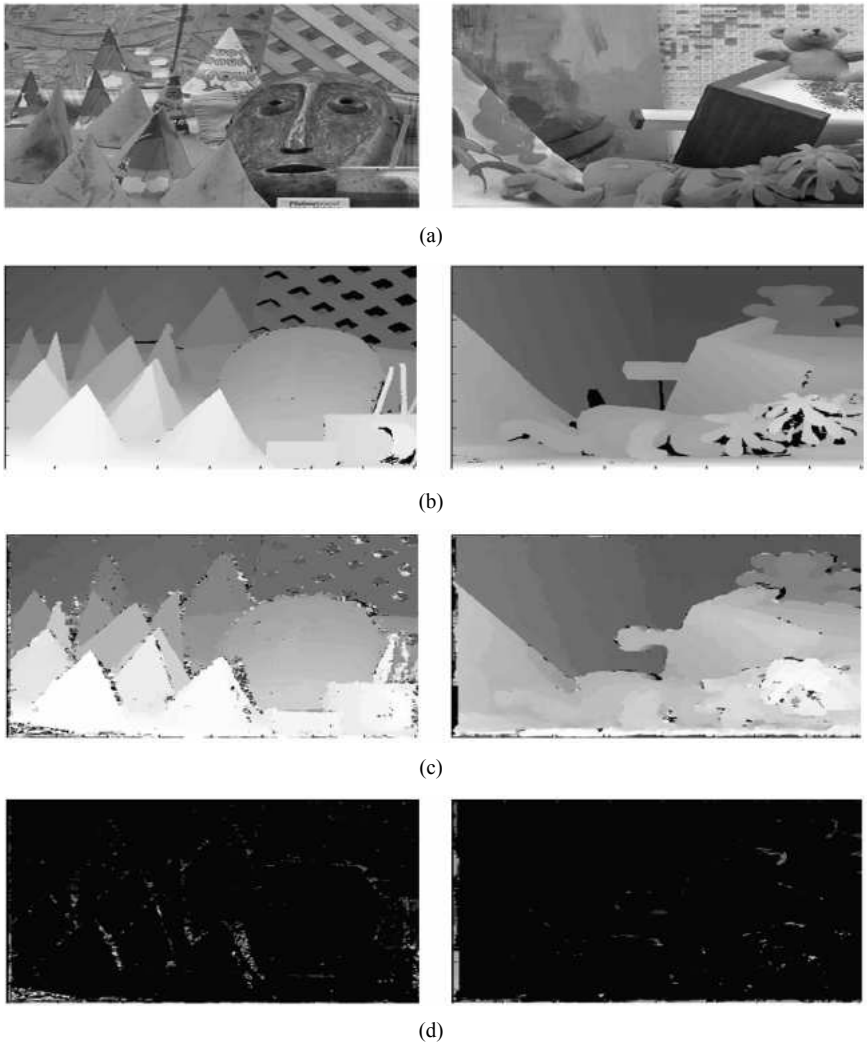
input image whereas $\xi$ represents the acceptable deviation of the estimated disparity value from the ground truth and is fixed to 1 in this particular work. The images are taken into consideration with different complexities, in terms of pixel intensity variation and surface boundaries. First pair of stereo images is shown in Figure 2 with related ground truth disparity maps, estimated disparity maps and the error between the ground truth and estimated disparity maps. As it can be seen in Figure 2 the edges of the discontinuities are extracted to high accuracy and estimated disparity is very much similar to the ground truth disparity, visually. The RMSE and PBD score for Sawtooth and Venus are $R = 1.7486$, $B = 0.0218$ and $R = 0.1027$, $B = 0.0374$, respectively.

**Figure 2** (a) Right images of the Sawtooth (left) and Venus (right) stereo pair (b) ground truth disparity maps (c) estimated disparity maps (d) disparity error



Similarly, another pair of disparity maps are shown for images Cones and Teddy and related RMSE and PBD scores are $R = 2.8427$, $B = 0.0930$ and $R = 1.8574$, $B = 0.0727$, respectively, as shown in Figure 3.

**Figure 3** (a) Right images of the Cones (left) and Teddy (right) stereo pair (b) ground truth disparity maps (c) estimated disparity maps (d) disparity error



### 5.3 Disparity errors comparison

To further validate the claims about the performance of the proposed algorithm a comparison is performed between the proposed algorithm and a number of selected algorithms from the literature. Eight algorithms are chosen, known for their performance, within the computer vision research community. These estimated disparity maps are related to the images *Cones*, *Venus* and *Teddy*. The chosen algorithms for comparison purpose are "double-bp (Yang et al., 2006), graph cuts (Scharstein, 2002),

infection (Olague et al., 2006), layered (Zitnick et al., 2004), scanline optimisation (Scharstein, 1998), SSD min. filter (Scharstein et al., 2002) and symmetric-occlusion (Sun et al., 2005)". The calculated statistics, i.e., R and B, for the analysis of comparative performance with respect to the estimated results are shown in Table 1 and Figures 4 to 6. It is obvious from Table 1 and Figures 4 to 6 that the proposed algorithm has performed best in the case of *Venus* image. However, in case of *Cones* and *Teddy* images the proposed algorithm has ranked 3rd, though very competitive to the algorithm ranked 1. Specifically in case of B, the proposed algorithm has outperformed all other algorithms. This reflects the true consistency and robustness of the proposed algorithm as number of bad disparity values estimated are lowest in all cases.

**Table 1** A comparison of the estimated disparity with a number of existing well known algorithms

| *Algorithms* | *Cones (Serie 1)* | | *Venus (Serie 2)* | | *Teddy (Serie 3)* | |
|---|---|---|---|---|---|---|
| | *R* | *B* | *R* | *B* | *R* | *B* |
| Estimated | $3.39764_3$ | $0.2138_1$ | $2.0018_1$ | $0.2248_1$ | $2.7588_3$ | $0.2115_1$ |
| Double-Bp | 3.4898 | 0.2329 | $2.2114_3$ | $0.2860_3$ | 2.9360 | 0.2852 |
| Graphcut | 4.9694 | 0.2732 | 3.3977 | 0.3065 | 5.6912 | 0.3314 |
| Infection | 4.2949 | $0.2147_3$ | 4.4952 | 0.3119 | 4.5092 | $0.2439_3$ |
| Layered | 4.6167 | 0.2638 | 3.1955 | 0.3186 | 4.3622 | 0.3096 |
| Realtime-Gpu | $3.2784_2$ | 0.2456 | $2.0780_2$ | $0.2609_2$ | $2.7535_2$ | 0.2815 |
| Scanline opt. | 5.4622 | 0.2989 | 4.2491 | 0.3090 | 5.7917 | 0.3538 |
| SSD. min. filter | 4.5599 | 0.2248 | 3.7330 | 0.2960 | 5.9532 | 0.3111 |
| Sym. occlusion | $3.1457_1$ | $0.2145_2$ | 15.7478 | 1.000 | $2.6445_1$ | $0.2421_2$ |

**Figure 4** Comparison of estimated disparity map with existing algorithms for image *Venus*

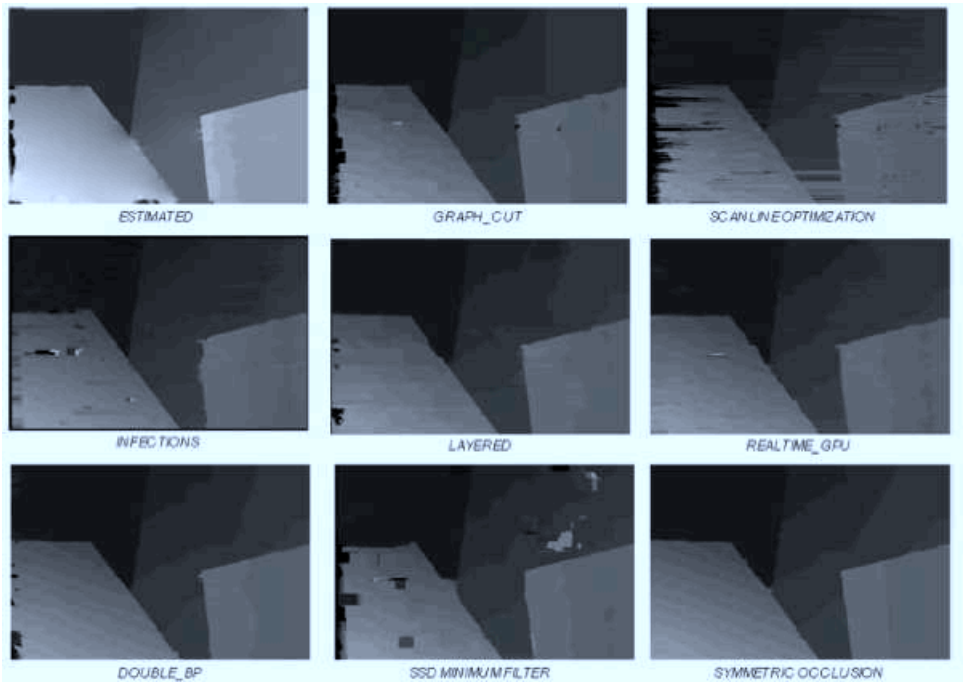


**Figure 5** Comparison of estimated disparity map with existing algorithms for image *Cones*

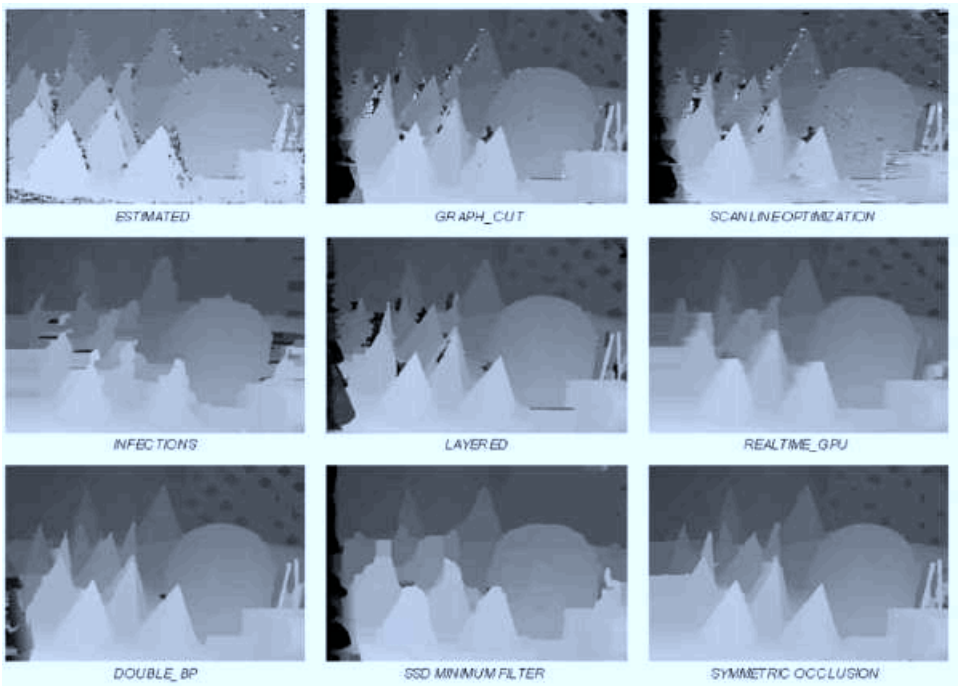
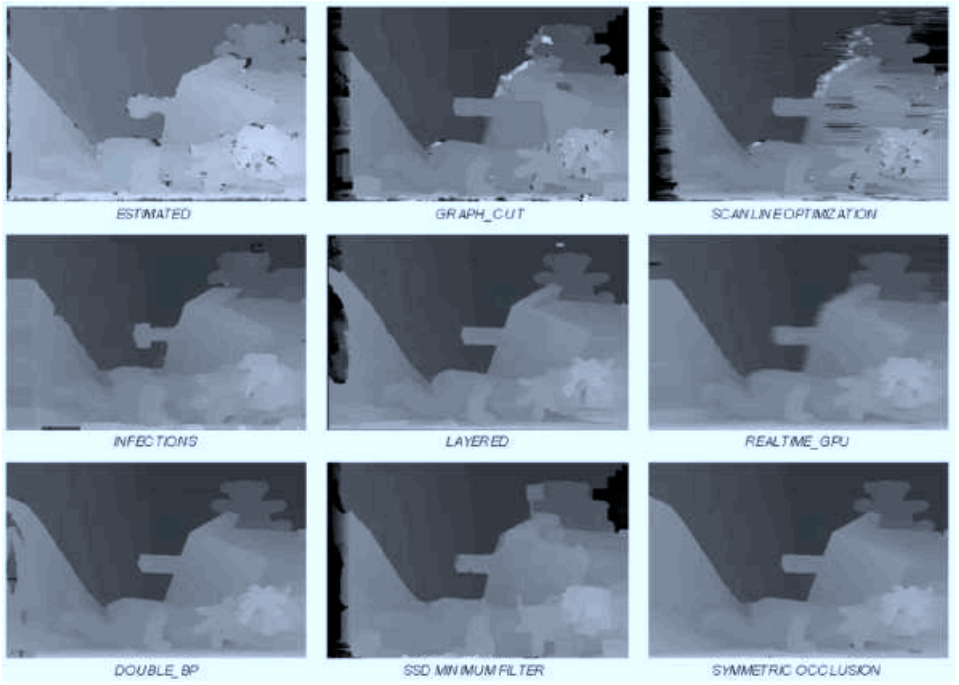



**Figure 6** Comparison of estimated disparity map with existing algorithms for image *Teddy*



### *5.4 Performance evaluation*

To evaluate the performance of these methods, we have used the image *Sawtooth*. Four algorithms (*graphcut, realtime-Gpu, double-Bp, SSD. min. filter*) are considered for the comparison. The test is done on a P4 with 3 GHz and 512 MB of RAM. Table 2 gives the execution time for the image *Sawtooth* with an image size of 434 $\times$ 380. We have used a window size $n = 4$ in this test. This table puts in evidence the enormous gain in execution time provided by our algorithm compared to global matching approaches and its suitability for realtime applications.

**Table 2** Comparison of execution time in seconds with existing stereo matching algorithms for *Sawtooth* image with an image size of 434 $\times$ 380

| *Algorithms* | *Execution time* |
|---|---|
| Our algorithm | 0.97 s |
| Graphcut | 15 s |
| Realtime-Gpu | 1.03 s |
| Double-Bp | 3.17 s |
| SSD. min. filter | 1.37 s |

## 6 Conclusions and perspectives

Local matching methods aim to find a matching for a given pixel without taking into account neighbour pixels correspondences. Global methods try to define a global model of the observed scene and to minimise a global cost function. Local stereo correspondence methods are in general fast algorithms, so can be used for realtime applications. However, they are exposed to many failure sources, in particular occlusions or variations of intensity between the two images. The global methods aim to reduce the sensibility of stereo correspondence to ambiguities caused by occlusions, poor local texture or fluctuation of illumination. This improvement has a cost, which is the increasing of algorithms complexity, and in consequence, a longer execution time. Our ultimate goal in this work was to find a compromise between realtime requirements and quality results. For this purpose, we have proposed a new approach to stereo matching, combining local method with global constraints. A new similarity measure is calculated based on fuzzy classification of grey scale levels and used to define a matching possibility distribution. Uniqueness and ordering constraints are used to define an unmatching possibility distribution based on the matching possibilities. The definition of a global matching cost combining the defined possibilities, let us achieve two goals:

- sensibly ameliorate the quality of disparity images as described in results section
- reduce the execution time in comparison with area-based methods that use a refinement step or bidirectional matching.

Hence, SLSM algorithm can be considered a viable alternative to area-based algorithms with optimisation step, in particular with big images and large disparity ranges, as it is the case of many current stereo applications.

However, our analysis and results confirm the need for further research aimed at dealing with the typical problems of area-based stereo algorithms. In particular, we

plan to embody specific techniques to cope with the borderlocalisation problem into an SLSM-based framework. Other optimisations are currently studied. Note that we work always on pre-rectified images, and we produce an integer disparity image: we will study how to adapt this algorithm in order to find stereo matching on non-rectified images and obtain sub-pixel disparity images.

## Acknowledgements

The authors would like to acknowledge the many helpful suggestions of three anonymous reviewers of this paper. The authors also thank the editor of this journal.

## References

Belhumeur, P.N. (1996) 'A Bayesian approach to binocular stereopsis', *Computer Vision*, Vol. 19, No. 3, pp.237–260.

Bellman, R. (1957) *Dynamic Programming*, Princeton University Press, Princeton.

Bobick, A.F. and Intille, S.S. (1999) 'Bobick', *International Journal of Computer Vision*, Vol. 33, No. 3, pp.181–200.

Boykov, Y., Veksler, O. and Zabih, R. (1998) 'Markov random fields with efficient approximations', in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.648–655.

Boykov, Y., Veksler, O. and Zabih, R. (1999) 'Fast approximate energy minimization via graph cuts', in *Proceedings of International Conference on Computer Vision (ICCV),*, Vol. 1, pp.377–384.

Chen, Q. and Medioni, G. (1999) 'A volumetric stereo matching method: application to image based modeling', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

Cox, I.J. (1992) 'Stereo without disparity gradient smoothing: a Bayesian sensor fusion solution', in *Proceedings of British Machine Vision Conference (BMVC)*, pp.337–346.

Cox, I.J., Hingorani, S.L., Rao, S.B. and Maggs, B.M. (1996) 'A maximum likelihood stereo algorithm', *Computer Vision and Image Understanding*, Vol. 63, No. 3, pp.542–567.

Faugeras, O. (1993) *Three-Dimensional Computer Vision: a Geometric Viewpoint*, MIT Press, Cambridge, MA, USA.

Fusiello, A., Trucco, E. and Verri, A. (2000) 'A compact algorithm for rectification of stereo pairs', *Machine Vision Appl.*, Vol. 12, No. 1, pp.16–32.

Hirschmüller, H. (2001) 'Improvements in real-time correlation-based stereo vision', *Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV)*, Sao Luis, Brazil, 21–23 September, pp.16–32.

Horn, B.K.P. (1986) *Robot Vision*, The MIT Press, Cambridge, MA, USA.

Ishikawa, H. and Geiger, D. (1998) 'Occlusions, discontinuities, and epipolar lines in stereo', in *Proceedings of the 5th European Conference on Computer Vision (ECCV)*, Springer-Verlag, pp.232–248.

Julesz, B. (1962) 'Towards the automation of binocular depth perception', in *IFIP Congress*, pp.439–444.

Kolmogorov, V. and Zabih, R. (2001) 'Computing visual correspondence with occlusions using graph cuts', in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp.508–515.

Kolmogorov, V. and Zabih, R. (2002a) 'Multi-camera scene reconstruction via graph cuts', *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, Springer-Verlag pp.82–96.

Kolmogorov, V. and Zabih, R. (2002b) 'What energy functions can be minimized via graph cuts?', *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, Springer-Verlag pp.65–81.

Marr, D. (1982) *Vision-A Computational Investigation into the Human Representation and Processing of Visual Information*, The MIT Press, Cambridge, MA, USA.

Ohta, Y. and Kanade, T. (1985) 'Stereo by intra- and inter-scanline search using dynamic programming', in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 7, No. 2, pp.139–154.

Olague, G., Fernández, F., Pérez, C. and Lutton, E. (2006) 'The infection algorithm: an artificial epidemic approach for dense stereo correspondence', *Artificial Life*, Vol. 12, No. 4, pp.593–615.

Roy, S. (1999) 'A maximum-flow formulation of the n-camera stereo correspondence problem', *International Journal of Computer Vision*, Vol. 34, Nos. 2–3, pp.147–161.

Roy, S. and Cox, I. (1998) 'Stereo without epipolar lines: a maximum-flow formulation', in *Proceedings of the International Conference on Computer Vision (ICCV)*, pp.492–499.

Sára, R. (2002) 'Finding the largest unambiguous component of stereo matching', in *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, Springer-Verlag, pp.900–914.

Scharstein, D. (1998) 'Stereo matching with nonlinear diffusion', *International Journal of Computer Vision*, Vol. 28, No. 2, pp.155–174.

Scharstein, D. and Szeliski, R. (2002) 'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms', *International Journal of Computer Vision*, Vol. 47, No. 1, pp.7–42.

Sun, J., Li, Y., Kang, S.B. and Shum, H.Y. (2005) 'Symmetric stereo matching for occlusion handling', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

Trucco, E. and Verri, A. (1998) *Introductory Techniques for 3D Computer Vision*, Prentice Hall, Englewood Cliffs, NJ.

Veksler, O. (1999) 'Efficient graph-based energy minimization methods in computer vision', PhD thesis, Cornell University, Adviser: Ramin Zabih.

Yang, Q., Wang, L., Yang, R., Stewenius, H. and Nister D. (2006) 'Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling', *CVPR*, Vol. 2, pp.2347–2354.

Zitnick, L., Kang, S.B., Uyttendaele, M., Winder, S. and Szeliski, R. (2004) 'High-quality video view interpolation using a layered representation', *SIGGRAPH*, Vol. 2, pp.2347–2354.