

Multi-label depth estimation for graph cuts stereo problems

Nicolas Papadakis* and Vicent Caselles†

June 10, 2010

Abstract

We describe here a method to compute the depth of a scene from a set of at least two images taken at known view-points. Our approach is based on an energy formulation of the 3D reconstruction problem which we minimize using a graph-cut approach that computes a local minimum whose energy is comparable (modulo a multiple constant) with the energy of the absolute minimum. As usually done, we treat the input images symmetrically, match pixels using photoconsistency, treat occlusion and visibility problems and we consider a spatial regularization term which preserves discontinuities. The details of the graph construction as well as the proof of the correctness of the method are given. Moreover we introduce a multi-label refinement algorithm in order to increase the number of depth labels without significantly increasing the computational complexity. Finally we compared our algorithm with the results available in the Middlebury database.

1 Introduction

Depth estimation is a highly studied problem in computer vision. Applications such as stereomovies, 3D reconstruction or novel view synthesis are indeed all based on the same basic ingredient: finding the depth of a scene seen from several cameras. All the methods giving good results on this problem and developed within the past 10 years can not be described here, as it would require half a book [11, 16, 26, 27, 28, 30, 29, 31]. Let us only underline that the approaches giving the more accurate results are often based on graph cuts [17, 18, 20] and belief propagation [16, 27, 30] minimizations, according to Middlebury database. In this work, we are interested in the seminal works on graph cut algorithm by Boykov, Kolmogorov *et. al.* in a series of papers [4, 5, 17, 18, 20]. More precisely, in the application of graph cuts to the depth estimation problem as presented in the last three papers.

Our approach can be considered as a combination of the ideas of [17, 18, 20] with a slightly different energy presentation. We believe that the description of the energy models relative to stereo problems that can be found in the graph cuts literature [17, 18, 20] presents some lacks of details that makes its implementation harder. Introducing an energy function that models the problem of depth estimation from a set of images in the more general way, we explain how the graph representability of the involved energies determines the occlusion management, and finally we recover the energy of [18] as a specific case. In this paper we give a presentation of the problem which is well suited to fill-in all details of the graph construction. They will be given in the Appendix.

*nicolas.papadakis@barcelonamedia.org. Barcelona Media, Avenida Diagonal 177, 08017 Barcelona, Spain.

†vicent.caselles@upf.edu. Universitat Pompeu Fabra, Carrer de Roc Boronat 138, 08018 Barcelona, Spain.

The proposed energy is based on a multi-label formulation of plane sweeping. It contains four terms: a photo-consistency matching cost for couples of corresponding pixels in pairs of images, a penalty on the occluded regions, a visibility constraint term and a regularization term. As usually in graph cut approaches, we use a regularization term based on the Potts model combined with a contours mask. When considering the actual best techniques dedicated to disparity estimation, it appears that most of them are strongly dependent on a previous segmentation of the images [27, 28, 31]. We believe that segmentations may be difficult to control in real images and they do not always provide the results that one would like. For this reason we choose a contours mask, which is easier to control, specially if contours are computed as significant edges in the sense of [8] (the ones we used, although a good implementation of the Canny edge detector could also be used). This information can be easily added to the regularization term of the energy.

We finally consider the problem of increasing the depth resolution by means of a multi-label refinement approach. The use of hierarchical schemes or warping techniques to obtain accurate estimations is a common technique in the variational (and partial differential equation) approaches to optical flow computation or stereo [1, 2, 6, 10, 15, 21] (see also [25] and references therein), where perturbations around the current solution permit to obtain faster and more accurate solutions, but it has not been adapted to the graph cuts context. We propose an adaptation here for our energy model.

Let us describe the plan of the paper. In Section 2 we recall some basic facts about image homographies induced by planes in the 3D scene. In Section 3 we present the energy used to model the depth estimation problem. We then describe in Section 4 the minimization of the proposed energy with a graph cut algorithm and we discuss in Section 5 the extension of the method to a multi-label refinement of the depth estimation. In Section 6, we finally show results obtained on stereo benchmark data sets, that present significant improvements with respect to the original paper of [18]. The final Appendix contains the details that permit an easy construction of the associated graph for each α expansion step of the minimization algorithm.

2 Preliminaries

To fix our notation, we assume that we have $N \geq 2$ cameras, one of them taken as the reference one. Thus, we consider N images $I_i : \Omega_i \rightarrow \mathbb{R}^m$, where $\Omega_i \subseteq \mathbb{R}^2$ denotes the domain of I_i , $i = 1 \dots N$, and $m = 1$ in case of gray level images or $m = 3$ for color images. We denote by $\mathcal{I} = \{(i, j) \in \{1, \dots, N\}^2, i \neq j\}$, the set of oriented pairs of images.

Our energy functional is based on the formulation of plane sweeping [7]. A plane of the 3D scene induces an homography between each pair of images, and the projections of image points which lie on the plane are matched by this homography. Thus, we may sweep the scene with a family of parallel ordered planes Π_λ , each one indexed by a parameter $\lambda > 0$ indicating its depth, and only the points of the scene lying on Π_λ will be matched in two images I_i and I_j , $i \neq j$, by the corresponding induced homography \mathcal{H}_λ^{ij} (see [12], page 326 for more details). The homography \mathcal{H}_λ^{ij} can be written as $\mathcal{H}_\lambda^{ij} = (\mathcal{H}_\lambda^j)^{-1} \mathcal{H}_\lambda^i$ where \mathcal{H}_λ^k is the homography between the k image plane and Π_λ , $k = 1, \dots, N$. Obviously, this is the case only if the point is visible in both cameras. It is important in practice that the sweeping planes are not too slanted with respect to the cameras. Notice also that the model only considers a point approximation to the object in the scene without taking into account its orientation (of its tangent plane) with respect to the camera. A better approximation could be obtained by comparing patches around each pixel. Nevertheless, if the object's orientation with respect to each camera is very different we could

have difficulties in comparing the information of both patches, if we do not take into account the object's relative deformation.

As usual, the matching is measured by a photo-consistency term and a constant penalty is included for occlusions. These two terms are complemented by the regularization term which penalizes neighboring pixels with similar intensity level or color and having different depth labels. The coherence between the estimations obtained from different cameras is finally enforced with a visibility constraint. The N images are treated in a symmetric way.

Before giving the precise description of our energy, let us recall the parameterization of the homographies induced by a family of sweeping planes.

2.1 The image homographies induced by a plane in the scene

We assume that the camera model is the finite projective camera model. Let $(X, Y, Z)^\top \in \mathbb{R}^3$ be the coordinates in the world coordinate system \mathcal{R}_w , which are mapped by the camera to the coordinates $(\tilde{x}, \tilde{y}, \tilde{z})^\top \in \mathbb{R}^3$ in the camera coordinate system \mathcal{R}_c . If \vec{C} denotes the vector of coordinates of the camera center in \mathcal{R}_w , and $R = (r_{ij})$ denotes the 3×3 rotation matrix representing the orientation of the camera coordinate frame, then, using homogeneous coordinates, we have

$$(\tilde{x}, \tilde{y}, \tilde{z}, 1)^\top = \begin{bmatrix} R & \vec{t} \\ \vec{0}^\top & 1 \end{bmatrix} (X, Y, Z, 1)^\top, \quad (1)$$

where the classical notation $\vec{t} = -R\vec{C} = (t_1, t_2, t_3)^\top$ represents the relative translation of both reference systems written in \mathcal{R}_c , and $\vec{0} = (0, 0, 0)^\top$. If K denotes the 3×3 calibration matrix containing the intrinsic parameters of the camera and $(x, y, 1)^\top$ the homogeneous image coordinates, then

$$(x, y, 1)^\top = K[R, \vec{t}](X, Y, Z, 1)^\top.$$

The 3×4 matrix $P = K[R, \vec{t}]$ is the projection matrix.

Let us recall the equations of the homography between the image plane and a plane sweeping the scene. Let Π_λ be a 3D plane given in \mathcal{R}_c by the equation

$$[0, 0, 1, -\lambda][\tilde{x}, \tilde{y}, \tilde{z}, 1]^\top = 0. \quad (2)$$

Using (1) the equation of Π_λ in \mathcal{R}_w is

$$[0, 0, 1, -\lambda] \begin{bmatrix} R & \vec{t} \\ \vec{0}^\top & 1 \end{bmatrix} [X, Y, Z, 1]^\top = [r_{13}, r_{23}, r_{33}, t_3 - \lambda][X, Y, Z, 1]^\top. \quad (3)$$

Given the N images $I_i : \Omega_i \rightarrow \mathbb{R}$, $i = 1, \dots, N$, obtained by N cameras whose calibration matrices are K^i and whose projection matrices are $P^i = K^i[R^i, \vec{t}^i]$, we first choose a camera $k \in \{1, \dots, N\}$ as the reference camera of the planes in \mathbb{R}^3 . In our applications we take the middle image as reference $k = \lfloor N/2 \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. We have the plane Π_λ defined by its coefficients $[r_{13}^k, r_{23}^k, r_{33}^k, t_3^k - \lambda]$. One can then define the projections \mathcal{H}_λ^i induced by the plane Π_λ that maps a pixel $p_i \in \Omega_i$ to the scene point $\mathcal{H}_\lambda^i p_i$ lying to the 3D plane Π_λ . Next, we can compute the homography $\mathcal{H}_\lambda^{ij} = \mathcal{H}_\lambda^j (\mathcal{H}_\lambda^i)^{-1}$ induced by the plane Π_λ . This homography maps a pixel $p_i \in \Omega_i$ to a pixel $p_j = \mathcal{H}_\lambda^{ij} p_i \in \Omega_j$ with respect to the 3D plane Π_λ . Such homographies are in practice used to find the correspondences between images i and j with respect to the 3D plane Π_λ . Using [12], page 326, we can also write the homography \mathcal{H}_λ^{ij} as

$$\mathcal{H}_\lambda^{ij} = K^j [\tilde{R} - \tilde{t} n^\top / (t_3^k - \lambda)] (K^i)^{-1},$$

where $\tilde{R} = R^j (R^i)^{-1}$, $\tilde{t} = -R^i \vec{t}^i + \vec{t}^j$, and $n = [r_{13}^k, r_{23}^k, r_{33}^k]^\top$.

3 The energy formulation

As explained in Section 2 our energy functional is based on a variational formulation of plane sweeping. It contains four terms: a photoconsistency matching cost for couples of corresponding pixels in pairs of images, a penalty on the occlusion regions, a visibility constraint term and a regularization term which penalizes the neighboring pixels with similar intensity level or color to have different depth labels. Moreover, the images are treated in a symmetric way. For each pair of images $(i, j) \in \mathcal{I}$, let $M_{ij} \subseteq \Omega_i$ be the set of pixels in Ω_i which are occluded in image j . We want to find the depth $\lambda(p)$ associated to each $p \in \Omega_i$, $i = 1, \dots, N$, and the sets of occluded pixels M_{ij} . Note that these sets of occluded pixels can be estimated from the knowledge of the depth. As the problem will be solved in a discrete framework via a graph cut approach, we assume that the depth $\lambda(p)$ takes its values in a predefined discrete set of possible depths contained in the range $[\lambda_{min}, \lambda_{max}]$.

Let us define the four energy contributions. The matching cost is

$$\mathcal{E}_{\text{match}}(\lambda) := \sum_{(i,j) \in \mathcal{I}} \sum_{p \in \Omega_i} D(p, \mathcal{H}_{\lambda(p)}^{ij}) T(\lambda(p) = \lambda(\mathcal{H}_{\lambda(p)}^{ij})) \quad (4)$$

where $T(\cdot)$ is 1 if its argument is true and 0 otherwise. This energy considers the scene points $\mathcal{H}_{\lambda(p)}^i p$, with $p \in \Omega_i$, that are visible in both images I_i and I_j . The matching cost $D(p, q) : \Omega_i \times \Omega_j \mapsto \mathbb{R}$ measures the difference between the colors $I_i(p)$ and $I_j(q)$ of the pixels $p \in \Omega_i$ and $q \in \Omega_j$. We will detail this function latter on as its minimization by graph cuts involves some restricted class of matching functions.

The cost of occluded regions is

$$\mathcal{E}_{\text{occ}}(\lambda) := \gamma \sum_{(i,j) \in \mathcal{I}} \sum_{p \in \Omega_i} T(\lambda(p) > \lambda(\mathcal{H}_{\lambda(p)}^{ij})), \quad (5)$$

where $\gamma \geq 0$ is the occlusion parameter. Indeed, if $\lambda(p) > \lambda(q)$, where $q = \mathcal{H}_{\lambda(p)}^{ij} p$, then the scene point $\mathcal{H}_{\lambda(p)}^i p$ is occluded by the scene point $\mathcal{H}_{\lambda(q)}^j q$ in the image I_j . The cost γ penalizes the occlusions.

We say that a labeling depth function $\lambda : \cup_{i=1}^N \Omega_i \rightarrow [\lambda_{min}, \lambda_{max}]$ satisfies the *visibility constraints* if

$$\forall q \in \Omega_j, j=1 \dots N, \text{ for which } \exists p \in \Omega_i \text{ with } i \neq j \text{ such that } q = \mathcal{H}_{\lambda(p)}^{ij} p, \text{ then } \lambda(q) \leq \lambda(p). \quad (6)$$

We will refer to the (p, q) visibility constraint when the relation (6) holds. This constraint means that a pixel p of an image i can be occluded by a pixel q in image j , if and only if the depth of q is smaller than the depth of p . We denote by \mathcal{A}_{vis} the set of admissible depth labelings that verify relation (6).

As previously explained, in case that $\lambda(q) < \lambda(p)$, then the 3D point $\mathcal{H}_{\lambda(p)}^i p$ of the scene is visible in image I_i but not in image I_j . Thus $p \in M_{ij}$. If $\lambda(q) = \lambda(p)$, then the 3D point $\mathcal{H}_{\lambda(p)}^i p$ is visible in both images. Hence, the energy that permits to impose the visibility constraint is

$$\mathcal{E}_{\text{vis}}(\lambda) := U \sum_{(i,j) \in \mathcal{I}} \sum_{p \in \Omega_i} T(\lambda(p) < \lambda(\mathcal{H}_{\lambda(p)}^{ij})), \quad (7)$$

where $U \rightarrow \infty$ is a large scalar that prevents the solution to violate the visibility constraint.

To define the regularization energy, let us consider a decreasing function $F : [0, \infty) \times [1, \infty) \rightarrow [0, \infty)$, described in the next paragraph, that penalizes the discontinuities of the depth labeling λ in the homogeneous regions of the image. The neighborhood $\mathcal{N}_{p_i}^\epsilon$, $p_i \in \Omega_i$, is defined as

$$\mathcal{N}_{p_i}^\epsilon = \{q \in \Omega_i, q \neq p_i, |q - p_i| \leq \epsilon\}, \quad \epsilon > 0. \quad (8)$$

In practice we have taken $\mathcal{N}_{p_i} = \mathcal{N}_{p_i}^{\sqrt{2}}$, that corresponds to a 8-neighborhood system. The regularization energy is

$$\mathcal{E}_{\text{reg}}(\lambda) := \alpha \sum_{i=1}^N \sum_{p_i \in \Omega_i} \sum_{q_i \in \mathcal{N}_{p_i}} F(|I_i(p_i) - I_i(q_i)|, |p_i - q_i|) |\lambda(p_i) - \lambda(q_i)|, \quad (9)$$

where $\alpha \geq 0$ is the regularization parameter. Finally, we want to minimize the energy \mathcal{E} defined as:

$$\min_{\lambda \in \mathcal{A}_{\text{vis}}} \mathcal{E}(\lambda) := \mathcal{E}_{\text{match}}(\lambda) + \mathcal{E}_{\text{occ}}(\lambda) + \mathcal{E}_{\text{vis}}(\lambda) + \mathcal{E}_{\text{reg}}(\lambda). \quad (10)$$

Remark 1 In practice, when induced by an arbitrary plane in \mathbb{R}^3 , the homography transformation \mathcal{H}_λ^{ji} does not necessarily map integer to integer coordinate values. As the graph cut approaches are discrete, given a pixel $p \in \Omega_i$ we have to take the integer coordinates of the corresponding pixel $\lfloor \mathcal{H}_\lambda^{ij} p_i \rfloor$. As a consequence, one has to choose the value of \mathcal{H}_λ^{ji} such that $p_i = \lfloor \mathcal{H}_\lambda^{ji} \lfloor \mathcal{H}_\lambda^{ij} p_i \rfloor \rfloor$. Nevertheless, notice that this problem disappears when computing integer disparities between pairs of rectified images, as pixel to pixel correspondences are involved in this case.

Regularization function

The regularization function $F(|I_i(p_i) - I_i(q_i)|, |p_i - q_i|) |\lambda(p_i) - \lambda(q_i)|$ is composed of two parts. The first one, $F(|I_i(p_i) - I_i(q_i)|, |p_i - q_i|)$, weights the discontinuities of the labeling taking into account the image gradient. This function is big when its first argument $|I_i(p_i) - I_i(q_i)|$ is small. The second argument $|p_i - q_i|$ enables weighting the influence of pixel q_i , the neighbor of pixel p_i , with respect to their Euclidean distance. The larger $|p_i - q_i|$, the smaller F . In our experiments, following [17, 18, 20], we have taken

$$F(r_1, r_2) = \begin{cases} 1/r_2 & \text{if } r_1 > \sigma \\ \kappa/r_2 & \text{otherwise} \end{cases} \quad (11)$$

with $\kappa > 0$ and $\sigma > 0$. Notice that the function F is always positive since it is better in practice to include a minimum penalty for the label discontinuities. Hence, the parameter κ permits to specify the range of values of the function F . With respect to [20], we have empirically set $\kappa = 3$ and $\sigma = 5$. Let us point out that we have also tested a continuous version of the regularization function such as $(1 + \kappa \exp(-r_1^2/\sigma^2))/r_2$, but curiously it leads to worst results in practice.

Let us mention that this regularization term is not new and is commonly used [3, 5, 10] (see also [25] and references therein).

The second term, $|\lambda(p_i) - \lambda(q_i)|$, is a total variation type term. It can be replaced by $\min(|\lambda(p_i) - \lambda(q_i)|, b)$ (in practice we set $b = 2$). Note that in the case $b = 1$ we recover the Potts model, $T(\lambda(p_i) \neq \lambda(q_i))$, which does not weight the amplitude of the depth discontinuities.

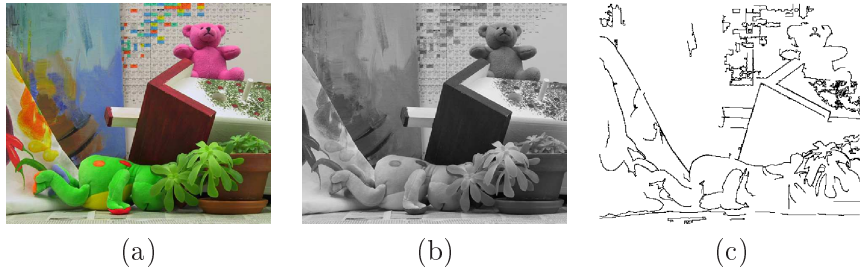


Figure 1: **Extraction of the significant level lines of the image:** An input image (a) is shifted to grey value (b). Applying [8], the significant level lines of the grey level image are extracted to give a map of contours (c).

Moreover, we use the information given by the significant edges of the image in the sense of [8]. This is done by extracting the so-called significant (in terms of length and contrast) level lines of the images I_i , as illustrated in figure 1.

We then dispose of a set of masks of contours $C_i : \Omega_i \mapsto \{0; 1\}$, with $C_i(p_i) = 1$ if p_i belongs to the detected contours in image I_i . We then use this additional information by considering $F(|I_i(p_i) - I_i(q_i)|, |p_i - q_i|\tau(p_i))$. The scalar $\tau(p_i)$ is set to 1, excepting for pixels belonging to a contour (i.e. $C_i(p_i) \neq C_i(q_i)$). In this case, we have $\tau(p_i) = \tilde{\tau} > 1$, with $\tilde{\tau} = 1.3$ in practice. This process allows to de-penalize the depth discontinuities on the detected contours C_i . It thus improves the estimation at the boundaries of image discontinuities.

4 Minimization of the energy using a graph cut algorithm

We now propose to solve the minimization problem (10) using a graph cut method with an α -expansion algorithm [5]. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}_d)$ be a directed graph. The set of vertices $\mathcal{V} = \{v_i\}$ corresponds to the set of pixels $p_i \in \mathcal{P}$, where $\mathcal{P} = \cup \Omega_i$, with two additional distinguished terminal vertices $\{S, T\}$ called the source and the sink. The set of edges linking the vertices is denoted by \mathcal{E}_d .

A cut $\mathcal{C} = \{\mathcal{V}^S, \mathcal{V}^T\}$ is a partition of the set of vertices such that $S \in \mathcal{V}^S, T \in \mathcal{V}^T$. The cost of the cut is the sum of the weights of the edges between a vertex in \mathcal{V}^S and a vertex in \mathcal{V}^T . A minimum cut is a cut with a minimum cost. The minimum cut can be found by computing the maximal flow using the Ford and Fulkerson algorithm [9]. Any cut can be described by a set of binary variables $\{x_i\}_{i=1, \dots, n}$, one for each vertex in $\mathcal{V} = \{v_i\}_{i=1, \dots, n}$, so that $x_i = 0$ when $v_i \in \mathcal{V}^S$ and $x_i = 1$ when $v_i \in \mathcal{V}^T$. Thus, if the graph represents an energy, this energy can be viewed as a function of the n binary variables $\{x_i\}$.

To minimize the energy (10) we need to create a graph adapted to the energy $\mathcal{E}(\lambda)$ such that the minimum cut corresponds to its minimum. Since the depth $\lambda(p)$ takes a discrete set of values, the global minimum of (10) could be computed with a minimum cost multiway cut, but this problem is known to be NP-hard [5]. If we fix the occlusion regions, the energy has the form

$$\mathcal{E}(f) = \sum_{p \in \mathcal{P}} C(p, f(p)) + \sum_{(p, q) \in \mathcal{N}} V_{p, q}(f(p), f(q)) \quad (12)$$

where \mathcal{P} is the set of pixels, $\mathcal{N} \subseteq \mathcal{P} \times \mathcal{P}$ is any neighborhood system on pixels, $f(p)$ is the label associated to p , $C(p, f(p))$ is the cost of assigning the label $f(p)$ to pixel p (a function derived from the data), and $V_{p, q}(f(p), f(q))$ is the cost of assigning the labels $f(p), f(q)$ to the adjacent

pixels p, q and is used as a regularization term. While the global minimum of the energy can be computed in the case of binary variables, the situation is more complex for a general discrete set of labels. It can be efficiently computed if the labels are consecutive integers and $V_{p,q}$ is the L^1 metric [13, 14, 22]. It can also be computed if $V_{p,q}$ is a convex functional, by considering the binary set under the graph of f [13]. The computation of the global minimum in the case of a discontinuity preserving regularization is more complex. In the case of the Potts model $V_{p,q}(f(p), f(q)) = T(f(p) \neq f(q))$ where $T(\cdot)$ is 1 only if its argument is true, the computation of the global minimum is NP-hard [5]. For that reason, several particular strategies have been designed in order to compute local minima which are within a known multiplicative factor of the global minimum energy. Between them, we mention the α -expansion algorithm [5]. Let us describe it.

Assume that the set of labels is a finite set \mathcal{L} . Given a labeling $\mathcal{P} \rightarrow \mathcal{L}$ and a particular label $\alpha \in \mathcal{L}$, we say that the labeling \tilde{f} is within an α -expansion of f if $\tilde{f}(p) = f(p)$ each time $\tilde{f}(p) \neq \alpha$. If $V_{p,q}$ is a metric for each pair (p, q) , then the problem of finding the minimum of the energy in the class of all α -expansions of f can be formulated as the minimization of an energy functional depending on a finite set of binary variables. Then the algorithm realizes cycles through the set of labels $\alpha \in \mathcal{L}$ finding each time the lowest energy α -expansion move from the current labeling. It adopts \tilde{f} as the new current labeling if \tilde{f} has a lower energy than the current one. The algorithm ends when there is no more α -expansion move with lower energy, for any label α .

To minimize the energy (10) we construct a graph whose vertices correspond to the pixels of the N images. The labeling is identified with the depth function λ that takes a finite number of values $\mathcal{L} \subseteq [\lambda_{min}, \lambda_{max}]$. If λ is the current labeling and $\alpha \in \mathcal{L}$, we associate a binary value x to each vertex p so that $x(p) = 1$ if p adopts label α and $x(p) = 0$ if it keeps the old label $\lambda(p)$. We then write the energy of a labeling $\tilde{\lambda}(p)$ within an α -expansion of λ as an energy function depending on a finite set of binary variables.

Following [19], a function E of n binary variables x_i is in the class \mathcal{F}^2 if it can be represented as

$$E(x_1, \dots, x_n) = \sum_k E^k(x_k) + \sum_{k < l} E^{k,l}(x_k, x_l).$$

We say that $E^{k,l}$ is regular if

$$E^{k,l}(0, 0) + E^{k,l}(1, 1) \leq E^{k,l}(0, 1) + E^{k,l}(1, 0). \quad (13)$$

If $E^{k,l}$ are non-regular, the minimization of E is NP-hard [19]. In practice, the edges of the graph corresponding to the energies are built from the values of $E^k(0)$, $E^k(1)$, $E^{k,l}(0, 0)$, $E^{k,l}(0, 1)$, $E^{k,l}(1, 0)$ and $E^{k,l}(1, 1)$, as explained in [19].

4.1 Reformulation of the energy

Let us fix a current labeling λ which satisfies the visibility constraints and $\alpha \in \mathcal{L}$. We are going to describe in detail the α -expansion moves in order to minimize (10) and write the energy as a function in the class \mathcal{F}^2 with regular energies $E^{i,j}$. The data matching cost, the occlusion cost and the visibility constraint are jointly modeled by two binary energies that are regular. In order to introduce them, let us define the set A^α as

$$A^\alpha = \{(p_i, p_j), \text{ with } p_i \in \Omega_i, p_j \in \Omega_j, (i, j) \in \mathcal{I}, \text{ such that } p_j = H_\alpha^{ij} p_i\}$$

and the set A^0 as

$$A^0 = \{(p_i, p_j), \text{ with } p_i \in \Omega_i, p_j \in \Omega_j, (i, j) \in \mathcal{I}, \text{ such that } p_j = H_{\lambda(p_i)}^{ij} p_i \text{ and } \lambda(p_i) \neq \alpha\}.$$

Recalling that the functions in \mathcal{A}_{vis} verify the relation (6), we define the admissible class

$$\mathcal{A}^{\lambda, \alpha} := \{\tilde{\lambda} \in \mathcal{A}_{vis}, \text{ such that } \tilde{\lambda}(p) \in \{\lambda(p), \alpha\} \forall p \in \cup_{i=1}^N \Omega_i\}$$

We are going to introduce two regular energies E_{ij}^α and E_{ij}^0 associated to the sets A^α and A^0 so that for any $\tilde{\lambda} \in \mathcal{A}^{\lambda, \alpha}$ the energy $\mathcal{E}(\tilde{\lambda})$ defined in (10) can be written as

$$\mathcal{E}(\tilde{\lambda}) = \tilde{\mathcal{E}}(\tilde{\lambda}) + \mathcal{E}_{reg}(\tilde{\lambda}), \quad (14)$$

where

$$\tilde{\mathcal{E}}(\tilde{\lambda}) = \sum_{(i,j) \in \mathcal{I}} \sum_{(p_i, p_j) \in A^\alpha} E_{ij}^\alpha(\tilde{\lambda}(p_i), \tilde{\lambda}(p_j)) + \sum_{(i,j) \in \mathcal{I}} \sum_{(p_i, p_j) \in A^0} E_{ij}^0(\tilde{\lambda}(p_i), \tilde{\lambda}(p_j)). \quad (15)$$

Now, observe that we may write $E_{ij}^\alpha(\tilde{\lambda}(p_i), \tilde{\lambda}(p_j))$, $E_{ij}^0(\tilde{\lambda}(p_i), \tilde{\lambda}(p_j))$ as energies depending on binary variables if we associate the label $x(p) = 1$ when $\tilde{\lambda}(p) = \alpha$ and $x(p) = 0$ when $\tilde{\lambda}(p) = \lambda(p)$. From now on we shall use this convention. Observe also that if $p_i \in \Omega_i$, then there may be only two pixels associated to p_i in Ω_j , namely $p_j := H_{\lambda(p_i)}^{ij} p_i$ and $p'_j := \mathcal{H}_\alpha^{ij} p_i$. Therefore we may write

$$\tilde{\mathcal{E}}(\tilde{\lambda}) = \sum_{(i,j) \in \mathcal{I}} \sum_{p_i \in \Omega_i} [E_{ij}^\alpha(x(p_i), x(p'_j)) T((p_i, p'_j) \in A^\alpha) + E_{ij}^0(x(p_i), x(p_j)) T((p_i, p_j) \in A^0)], \quad (16)$$

The energies E_{ij}^α and E_{ij}^0 jointly model the matching and occlusion costs and the visibility constraints. When restricted to the class $\mathcal{A}^{\lambda, \alpha}$ the problem (10) is equivalent to

$$\min_{\tilde{\lambda} \in \mathcal{A}^{\lambda, \alpha}} \tilde{\mathcal{E}}(\tilde{\lambda}) + \mathcal{E}_{reg}(\tilde{\lambda}). \quad (17)$$

The construction of the edges corresponding to the regularization term of the energy (10) is classical [19] and we will not detail it here.

4.2 Description of E_{ij}^α

Let us first describe E_{ij}^α , $(i, j) \in \mathcal{I}$, the energy associated to A^α . For all $(p_i, p_j) \in A^\alpha$, i.e. $p_j = \mathcal{H}_\alpha^{ij} p_i$, this energy models:

- (i) the matching cost $D(p_i, p_j)$ between p_i and p_j ,
- (ii) the occlusion cost γ of a pixel p_i by p_j ,
- (iii) the visibility constraint between p_i and p_j . That is, $\tilde{\lambda}(p_j) \leq \tilde{\lambda}(p_i)$.

For the sake of clarity, we describe in Appendix A the different possible geometric configurations of the pair of points $p_i \in \Omega_i$, $p_j \in \Omega_j$ such that $(p_i, p_j) \in A^\alpha$.

4.3 Description of E_{ij}^0

Let us now detail E_{ij}^0 , $(i, j) \in \mathcal{I}$, the energy associated to A^0 . For all $(p_i, p_j) \in A^0$, i.e. $p_j = \mathcal{H}_{\lambda(p_i)}^{ij} p_i$, this energy models:

- (i) the matching costs, relative to the labels different from α , between p_i and p_j ,
- (ii) the occlusion cost γ of a pixel p_i by p_j ,
- (iii) the visibility constraint between p_i and p_j . That is, $\tilde{\lambda}(p_j) \leq \tilde{\lambda}(p_i)$.

For the sake of clarity, we describe in Appendix B the different possible geometric configurations of the pair of points $p_i \in \Omega_i$, $p_j \in \Omega_j$ such that $(p_i, p_j) \in A^0$.

4.4 Definition of the matching cost function

From the previous cases described in appendices A and B, it appears that the functions $E_{i,j}^\alpha$ and $E_{i,j}^0$ are regular depending on the choice of the matching function D . In particular, these energies are regular in the sense of relation (13) as soon as $D(p_i, p_j) \leq \gamma$, for all $p_i \in \Omega_i$, $p_j \in \Omega_j$, $(i, j) \in \mathcal{I}$. This condition means that the matching cost can not exceed the occlusion cost. This strong constraint is however necessary to minimize correctly the energy with graph cut.

We then define the matching cost function as:

$$D(p_i, p_j) := \min\left(\sum_{l=1}^m |I_i^l(p_i) - I_j^l(p_j)|, \gamma\right),$$

where I_i^l is the l -color channel of image $I_i : \Omega_i \rightarrow \mathbb{R}^m$. In [18], the occlusion parameter is automatically chosen as

$$\gamma = \sup_{i,j,p_i,p_j} D(p_i, p_j).$$

Such a model considers a too high occlusion cost, and we prefer to set γ as a parameter to be chosen.

Remark 2 As the homography transformation does not necessarily map integer to integer coordinate values, for each pixel $p \in \Omega_i$, a bilinear approximation is used to compute $I_j^l(\mathcal{H}_\lambda^{ij} p)$.

Remark 3 The construction of the graph can be equivalently simplified by first modifying the matching cost between $p_i \in \Omega_i$ and $p_j \in \Omega_j$ as

$$\tilde{D}(p_i, p_j) := D(p_i, p_j) - \gamma \tag{18}$$

with $\gamma > 0$, and then resetting the occlusion cost γ to zero. In this case, we exactly recover the matching cost function of [18]. In practice, this trick enables to pay a null cost for all occlusions cases. This modification reduces the complexity of the graph construction, as a lot of edges have now a null weight. In this paper, contrary to [18], we choose to present the energies with positive matching and occlusion costs, in order to give an intuitive meaning to the different variables and parameters involved.

4.5 Properties of the constructed energies

Proposition 1 *If λ satisfies the visibility constraints, so does any $\tilde{\lambda} \in \mathcal{A}^{\lambda, \alpha}$, $\alpha \in \mathcal{L}$, such that $\tilde{\mathcal{E}}(\tilde{\lambda}) < \infty$. In particular, the minimum of (17) satisfies them.*

From the definition of the energies E_{ij}^α , E_{ij}^0 , it is clear that if $\mathcal{H}_{\tilde{\lambda}(p_i)}^{ij} p_i = p_j$, then $\tilde{\lambda}(p_j) \leq \tilde{\lambda}(p_i)$. Thus the above statement follows.

Proposition 2 *The formula (15) holds. Indeed, for each (p_i, p_j, p'_j) where $p_j = \mathcal{H}_{\lambda(p_i)}^{ij} p_i$, $p'_j = \mathcal{H}_{\alpha p_i}^{ij} p_i$, we have that $E_{ij}^\alpha(x(p_i), x(p'_j)) + E_{ij}^0(x(p_i), x(p_j))$ is*

- = $D(p_i, p_j)$ if there is a matching between p_i and p_j ,*
- = $D(p_i, p'_j)$ if there is a matching between p_i and p'_j ,*
- = γ if the configuration represents an occlusion of p_i by p_j or by p'_j , , i.e. $p_i \in M_{ij}$,*
- = $+\infty$, if there is a violation of visibility either of (p_i, p_j) , (p_j, p_i) , (p_i, p'_j) , or (p'_j, p_i) .*

The proof of this Proposition consists in checking for each configuration of (p_i, p_j) all possible configurations of (p_i, p'_j) consistent with it. Since this checking is immediate but tedious we omit the details (they can be easily checked using the full description of the energies E_{ij}^α and E_{ij}^0 given in the appendix. Just notice that the only possibilities are the ones described above: given α , p_i can either match with p_j or with p'_j , it can be occluded by p_j or p'_j , or there may be a violation of the visibility constraint in the configuration we examine. The occlusion of p_j or p'_j by p_i will pay γ when we compute the configuration associated to the cameras (j, i) .

The α -expansion algorithm

The α -expansion is summarized in Algorithm 1.

Algorithm 1 *Depth estimation*

- Initialize with a configuration respecting the visibility constraint: $\lambda(p) = \lambda_{max}$, $\forall p \in I_i$, $i = 1, \dots, N$.*
- Repeat until convergence, for $\alpha = \lambda_{max}, \dots, \lambda_{min}$,*
 - 1 Construct the edges relative to the terms E_{ij}^α and E_{ij}^0 , $i, j = 1, \dots, N$.*
 - 2 Construct the regularization edges.*
 - 3 Cut the graph.*
 - 4 Labeling: if, after the cut, a vertex v is associated to the sink, then $\lambda(p) = \alpha$.*

5 A multi-label refinement algorithm for depth estimation

Using a very large set of possible labels improves the accuracy of the estimation. Nevertheless, the computational cost of the process increases with the considered number of labels. Another important point is the memory storage, as pre-computing all the possible matching costs $D(p, q)$ may lead to huge memory allocations in case of large sets of labels. One solution to get around

this problem could be to recompute the necessary matching costs for each α -expansion, but it would drastically increase the computational time. Another approach consists in considering hierarchical algorithms.

Hierarchical schemes based on a multi-resolution image representation or warping methods are commonly used in the context of variational methods or partial differential equation approaches to optical flow computation or stereo [1, 2, 6, 10, 15, 21] (see also [25] and references therein) where perturbations around the current solution permit to obtain faster and more accurate solutions. For graph cut optimization, such kind of approach would nevertheless induce the propagation of errors at the boundaries of objects. In order to treat this problem, a complex specific process dealing with the visibility constraint at the depth discontinuities should then be applied at each level of the image pyramid. For this reason, we choose to define a multi-label refinement scheme of the depth in order to improve the quality of the estimation while keeping an attractive computational time.

5.1 Labeling refinement

After applying the algorithm of the previous section, we obtain a first depth estimation $\lambda(p) \in \mathcal{L} = [\lambda_{min}; \lambda_{max}]$ for all $p \in \Omega_i, i = 1, \dots, N$. As the set of labels \mathcal{L} is in practice discrete, finite and ordered, we can denote by L the number of its elements and rewrite it as:

$$\mathcal{L} = \{\lambda_l\}_{l \in \{0, \dots, L-1\}},$$

with $\lambda_0 = \lambda_{min}$ and $\lambda_{L-1} = \lambda_{max}$. In order to make this discrete estimation more accurate, we propose a multi-label refinement process in order to consider a larger number of possible depths. We then refine the set \mathcal{L} by a factor M and we consider the following set of labels:

$$\mathcal{L}_M = \{\lambda_l\}_{l \in \{0, \frac{1}{M}, \dots, \frac{M-1}{M}, 1, \dots, L-2 + \frac{M-1}{M}, L-1\}}.$$

Next, the multi-label depth refinement is only done on the neighborhood of the values of the first depth estimation λ . More precisely, in order to reduce the computational cost of this process, we want to take advantage of the first estimation to restrict the number of possible labels for each pixel. Considering a pixel p whose current depth estimation is $\lambda(p) = \lambda_{l^p} \in \mathcal{L}$, the new admissible set of labels for this pixel is:

$$\mathcal{L}(p) := [\lambda_{l^p-1}, \lambda_{l^p+1}] \cap \mathcal{L}_M.$$

5.2 Multi-label energies

A slightly modified α -expansion algorithm is applied for all $\alpha \in \mathcal{L}_M$.

5.2.1 Domain of the multi-label α -expansion

We first define $\Omega_i^\alpha \subseteq \Omega_i, i = 1, \dots, N$, as the sets of pixels that can take the label α :

$$\Omega_i^\alpha = \{p \in \Omega_i \text{ such that } \alpha \in \mathcal{L}(p)\}. \quad (19)$$

In the new graph, we include one vertex for each pixel $p \in \Omega_i^\alpha$. That is, we only consider the pixels that can take the label α . This reduces significantly the size of the current graph. For all the vertices corresponding to $p \in \Omega_i^\alpha, i = 1, \dots, N$, the graph is then built as explained in the previous section.

5.2.2 Boundary of the multi-label α -expansion

As the regularization term (11) measures the discrepancy between the label of two neighbor pixels, we have to include this information on the boundary of the regions Ω_i^α . In order to keep the continuity of the current estimation $\boldsymbol{\lambda}$, we define the boundary set B_i^α of Ω_i^α as

$$B_i^\alpha = \{p \in \Omega_i \setminus \Omega_i^\alpha \text{ such that } \exists q \in \mathcal{N}_p \cap \Omega_i^\alpha, \mathcal{N}_p \text{ being the neighborhood of } p\}. \quad (20)$$

Hence, for each image i , one vertex is added for each pixel $p \in B_i^\alpha$. For the vertices corresponding to pixels $p \in \Omega_i^\alpha$, $i = 1, \dots, N$, the graph is built as explained in the previous section. For the pixels p belonging to the boundary set B_i^α , we need a new energy term $\mathcal{E}_{\text{bound}}$. For any labeling $\tilde{\boldsymbol{\lambda}}$ within an α -expansion of the current labeling $\boldsymbol{\lambda}$, this energy is defined as

$$\mathcal{E}_{\text{bound}}(\tilde{\boldsymbol{\lambda}}) := U \sum_{p \in B_i^\alpha} T(\tilde{\boldsymbol{\lambda}}(p) = \alpha), \quad (21)$$

with $U > 0$ being a large constant. Hence, for each image i , one vertex is added for each pixel $p \in B_i^\alpha$. As before, this energy can be written with respect to the binary variable $x(\tilde{\boldsymbol{\lambda}}(p))$ with a graph representable energy E_{bound} :

$$\mathcal{E}_{\text{bound}}(\tilde{\boldsymbol{\lambda}}) = \sum_{p \in B_i^\alpha} E_{\text{bound}}(x(\tilde{\boldsymbol{\lambda}}(p))).$$

This energy sets $E_{\text{bound}}(1) = \infty$ and $E_{\text{bound}}(0) = 0$ for the vertices corresponding to pixels $p \in B_i^\alpha$. Indeed, a pixel $p \in B_i^\alpha$ cannot take the label α , since from definitions (19) and (20), $\alpha \notin \mathcal{L}(p)$.

5.2.3 Visibility in the multi-label α -expansion

Moreover, some additional constraints must be added to respect the visibility constraint. Assume that the current label of $q \in \Omega_j \setminus \Omega_j^\alpha$ is $\boldsymbol{\lambda}(q)$ and there exists $p = \mathcal{H}_{\boldsymbol{\lambda}(q)}^{ji} q \in \Omega_i^\alpha$. Due to the visibility constraint, the pixel p can not always change its label to α . Indeed, if p is currently occluding q or if p and q are currently matching, some violation of the visibility constraint could appear as q is not involved in the current α -expansion, since $\alpha \notin \mathcal{L}(q)$. As a consequence, during the α -expansion, the pixel p can only change its label to the value α if $\boldsymbol{\lambda}(p) > \boldsymbol{\lambda}(q)$ and $\alpha > \boldsymbol{\lambda}(q)$. In the other cases, the visibility constraint can be violated for the pixel q . The set of pixels that can violate the visibility constraint is denoted by C^α and defined as

$$C^\alpha := \{p = \mathcal{H}_{\boldsymbol{\lambda}(q)}^{ji} q \in \Omega_i^\alpha, \text{ such that } q \in \Omega_j \setminus \Omega_j^\alpha \text{ and } \boldsymbol{\lambda}(p) \leq \boldsymbol{\lambda}(q) \text{ or } \alpha \leq \boldsymbol{\lambda}(q)\}.$$

We then design a last visibility energy $\mathcal{E}_{\text{vis}}^C$ for the pixels $p \in C^\alpha$. For any labelling $\tilde{\boldsymbol{\lambda}}$ within an α -expansion of the current labelling $\boldsymbol{\lambda}$, this energy is defined as

$$\mathcal{E}_{\text{vis}}^C(\tilde{\boldsymbol{\lambda}}) := U \sum_{p \in C^\alpha} T(\tilde{\boldsymbol{\lambda}}(p) = \alpha), \quad (22)$$

with $U > 0$ being a large constant. As before, this energy can be written with respect to the binary variable $x(\tilde{\boldsymbol{\lambda}}(p))$ with a graph representable energy E_{vis} :

$$\mathcal{E}_{\text{vis}}(\tilde{\boldsymbol{\lambda}}) = \sum_{p \in B_i^\alpha} E_{\text{vis}}(x(\tilde{\boldsymbol{\lambda}}(p))).$$

This energy sets $E_{\text{vis}}^C(1) = \infty$ and $E_{\text{vis}}^C(0) = 0$ for the vertices corresponding to the pixels $p \in C^\alpha$.

In practice, the multi-label refinement process presents one main advantage: the memory necessary to store the possible matching costs is significantly reduced, as each pixel can only take a reduced range of depth labels at each level of resolution. This property makes the process faster than considering directly the whole set of possible labels, since it allows pre-computing the different matching costs even for a final huge number of labels.

If the memory is big enough to pre-compute all the matching costs, we have observed that, for the same final amount of labels, the multi-label refinement is always faster than the direct estimation. The speed gain is nevertheless small (up to 25%).

6 Application to disparity estimation

Considering the planes in \mathbb{R}^3 corresponding to disparity values, one can apply the depth algorithm to the estimation of disparity from pairs of rectified images. Indeed, from the focal distance f of the camera and the distance D between optical centers, the disparity d is linked to the depth λ through the relation $d = Df/\lambda$. Then the corresponding homography matrix between the images 1 and 2 is given by

$$\mathcal{H}_\lambda^{12} = \begin{bmatrix} 1 & 0 & Df/\lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We have applied such a process to the Middlebury benchmark disparity data [23, 24] composed of four datasets and obtained some significant improvements comparing to the original method of [18]. Namely, as presented in Figure 2, the introduction of the distance between depth in the regularization term allows recovering some smoother disparity maps. Our approach has been first applied with one label per integer disparity value, that is, with L possible labels (we refer to it in Tables 1 and 3 as 'Exp. L'). Then, we have doubled the number of labels, by considering real disparities (0, 0.5, 1, 1.5, ...) in order to reach subpixel accuracy. We have computed the disparity with $2L$ labels in two different ways:

- A direct estimation¹ with $2L$ labels ('Exp. 2L')
- A refined estimation using 'Exp. L' as the first resolution, followed with a multi-label refinement corresponding to $M = 2$ ('Exp. 2L + ML').

Notice that the results presented here have been obtained with the same parameters for the four sequences and all the experiments: occlusion $\gamma = 17$ and regularization $\alpha = 3$. Tuning independently the parameters for each pair of images obviously leads to smaller errors (in that case the mean error has been reduced up to 5.5). The computational time corresponding to these results is from 10 seconds for the Tsukuba images up to one minute for the Teddy and Cones images on a standard PC.

We show in Table 1 the percentage of pixels of the image with a disparity error greater than 1, for the non occluded parts (Nocc), the whole image (all), and the region around the discontinuities (Disc). As illustrated by the last column of this table, the simple method 'Exp. L' performs very well with respect to the state-of-the-art techniques. We have only quoted the

¹The direct estimation has been done with a pre-computing of the matching cost when it was possible to store in memory all the information.

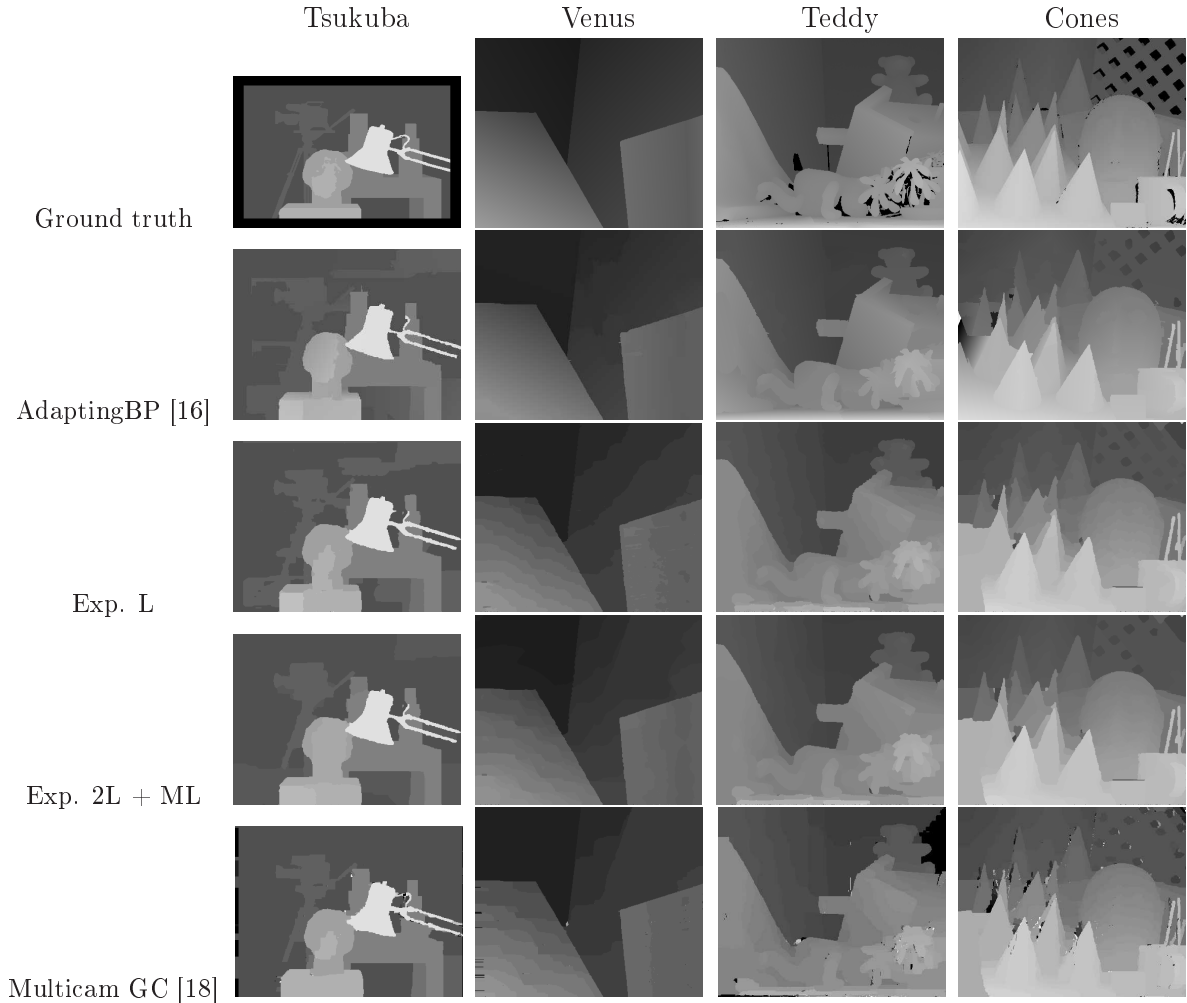


Figure 2: **Visualization of the results obtained on the Middlebury benchmark datasets.** We compare the ground truth with the results obtained by four methods. First, the method of [16] having the best quantitative evaluation, then the results obtained by our approach, without and with a multi-label refinement of depth and finally the results of the original paper [18].

techniques with very good performance, as well as all the approaches based on graph cuts. Let us underline that the current total number of evaluated methods on the website is nearby 50 and they can not be all reported here. It is important to underline that the only methods that present some better quantitative results all require a segmentation of the regions of the images. We nevertheless believe that such a pre-processing may fail and badly constrain the disparity estimation in real cases.

The obtained results are quite interesting when considering the two subpixel results. First of all, with the same parameters as before, we observe an increase of the error in Table 1 for these two approaches. This observation is quite natural for the label refinement approach 'Exp. 2L + ML'. Indeed, the second level of resolution is realized in the neighborhood of the previous estimation 'Exp. L' that was considering integers disparities. As a consequence, there is no gain on the error rate corresponding to the threshold 1. In the case of the direct estimation 'Exp. 2L', the increase of the error may be intuitively explained by the fact that too many labels are possible for each pixel. From the high combinatory of this labeling problem, the process seems

Method	Sequences												Mean
	Tsukuba			Venus			Teddy			Cones			
	Nocc	All	Disc	Nocc	All	Disc	Nocc	All	Disc	Nocc	All	Disc	
AdaptingBP [16]	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32	4.23
CoopRegion [28]	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13.0	2.79	7.18	8.01	4.41
DoubleBP [30]	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	9.63	2.90	8.78	7.79	4.19
OutlierConf [29]	0.88	1.43	4.74	0.18	0.26	2.40	5.01	9.12	12.8	2.78	8.57	6.99	4.60
AdaptOvrSegBP [27]	1.69	2.04	5.64	0.14	0.20	1.47	7.04	11.1	16.4	3.60	8.96	8.84	5.59
SymBP+occ [26]	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9	5.92
Exp. L	0.9	1.32	4.82	0.45	0.84	3.32	6.46	11.8	17.05	4.34	10.55	10.37	6.04
AdaptDispCalib [11]	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72	6.10
OverSegmBP [31]	1.69	1.97	8.47	0.50	0.68	4.69	6.74	11.9	15.8	3.19	8.81	8.89	6.11
Exp. 2L	0.92	1.33	4.97	2.23	2.64	4.48	7.10	12.67	17.99	5.34	11.43	12.31	6.95
Exp. 2L + ML	0.99	1.4	5.36	0.63	1.08	4.23	8.16	13.87	19.5	5.25	11.27	11.91	6.97
GC+occ [17]	1.19	2.01	6.24	1.64	2.19	6.75	11.2	17.4	19.8	5.36	12.4	13.0	8.26
MultiCamGC [18]	1.27	1.99	6.48	2.79	3.13	3.60	12.0	17.6	22.0	4.89	11.8	12.1	8.31
GC [4]	1.94	4.12	9.39	1.79	3.44	8.75	16.5	25.0	24.9	7.70	18.2	15.3	11.4

Table 1: Comparison of the quantitative results on the Middlebury benchmark datasets with a disparity error threshold of 1.

to get stuck in a bad local minimum of the energy. Indeed, as previously mentioned, graph cut algorithms find local minima which are within a known multiplicative factor of the global minimum energy [5]. When observing the energy of the different solutions found by our process (see Table 2), this explanation seems reasonable.

Experience	Tsukuba	Venus	Teddy	Cones
Exp. L	-746894	-533160	-947660	-705721
Exp. 2L	-743147	-455284	-936695	-648662
Exp. 2L + ML	-746922	-534561	-959592	-712202

Table 2: Comparison of the energy value of the local minimum found by the graph cut optimization for the different tested approaches. The values are negative as the matching cost is negative from relation (18). The multi-label refinement approach always finds the local minimum with the smallest energy.

If we now look at a smaller error threshold, the results are quite different. We show in Table 3 the percentage of pixels of the image with a disparity error greater than 0.5, for the non occluded parts (Nocc), the whole image (all), and the boundaries of the discontinuities (Disc). It appears that the multi-label refinement process 'Exp. 2L + ML' really decreases this rate with respect to the first level of estimation 'Exp. L', with some significant improvements on the Tsukuba image data. On the other hand, the direct estimation with the full label resolution 'Exp. 2L' does not improve the quality of the estimation. It is also important to see that the order of the methods giving the best results changes when modifying the threshold.

Method	Sequences												Mean
	Tsukuba			Venus			Teddy			Cones			
	Nocc	All	Disc	Nocc	All	Disc	Nocc	All	Disc	Nocc	All	Disc	
AdaptOvrSegBP [27]	5.98	6.56	9.09	3.66	3.96	13.2	13.0	18.9	26.4	9.48	14.9	17.2	11.9
OverSegmBP [31]	7.75	8.17	13.8	4.33	4.73	16.8	13.2	19.3	27.5	6.53	12.6	14.0	12.4
CoopRegion [28]	18.9	19.5	21.2	1.38	1.77	9.39	10.6	15.3	22.6	5.36	10.6	13.0	12.5
Exp. 2L + ML	2.96	3.61	8.51	7.71	8.33	14.46	14.97	21.41	28.06	10.8	16.59	18.8	13.02
AdaptingBP [16]	9.1	19.3	17.4	4.84	5.08	7.84	12.8	16.7	26.3	7.02	13.2	14.0	13.6
DoubleBP [30]	18.7	19.1	15.8	7.82	8.22	11.3	14.4	19.9	24.4	11.8	17.6	19.7	15.7
SymBP+occ [26]	20.7	21.6	19.5	5.96	6.27	10.2	15.7	20.9	31.7	11.4	17.5	18.9	16.7
Exp. L	15.45	15.88	16.19	8.89	9.44	13.57	15.72	22.17	30.41	14.48	20.22	22.15	17.05
OutlierConf [29]	24.7	25.0	17.4	8.01	8.27	13.7	15.6	20.5	28.7	10.9	17.3	17.4	17.3
GC+occ [17]	6.10	7.11	14.6	10.7	11.3	16.9	23.7	30.1	34.6	12.2	19.2	21.9	17.4
AdaptDispCalib [11]	24.6	24.7	21.3	7.14	7.56	15.0	18.8	25.2	29.7	9.21	15.1	16.7	17.9
Exp. 2L	11.77	12.24	14.84	14.19	14.69	15.15	18.36	24.63	32.5	14.88	20.42	21.96	17.97
MultiCamGC [18]	6.56	7.55	15.7	16.4	16.8	17.6	24.3	30.4	36.9	12.0	18.8	21.2	18.7
GC [4]	7.71	9.82	17.4	13.5	15.0	20.9	31.5	38.5	38.9	15.9	25.5	24.7	21.6

Table 3: Comparison of the quantitative results on the Middlebury stereo benchmark datasets with a disparity error threshold of 0.5.

As illustrated on the table 2, the multi-label refinement approach always finds the local minimum with the smaller energy. This table supports our model. Indeed, the smaller is the

energy of the solution found, the better is the error with respect to the Middlebury most accurate threshold of 0.5 (see Table 3).

From these experiments, it seems better to restrict the number of possible labels for the first level of estimation and increase it in a second turn. With respect to a direct full estimation, the multi-label refinement approach allows us to obtain very accurate results while speeding-up the process.

7 Conclusion and future works

In this paper, we have addressed the problem of depth estimation of a scene seen from multiple cameras. After defining a functional relying on photoconsistency, occlusion, visibility and regularization, we have fully detailed its minimization using a graph cuts optimization. We have finally introduced a multi-label refinement scheme that enables to speed up the process while still giving very accurate results on the Middlebury benchmark dataset.

This work is our first step to address the problem of view synthesis from a virtual point of view in the case of static scenes. The other ingredients that will be needed are the transfer of information to the new virtual view point, the handling of occlusions by using the information from other cameras, the matting approach and the eventual inpainting of the missing details.

Acknowledgment This work was partially funded by Mediapro through the Spanish project CENIT-2007-1012 i3media and by the Centro para el Desarrollo Tecnológico Industrial (CDTI), within the Ingenio 2010 initiative. Nicolas Papadakis acknowledges support from the Torres Quevedo program of the Ministerio de Educación y Ciencia in Spain. Vicent Caselles also acknowledges partial support by 'Fundació Barcelona Media', by PNPGC project, reference MTM2006-14836. and by "ICREA Acadèmia" prize for excellence in research funded by the Generalitat de Catalunya.

A Energy E^α

We now describe the different possible geometric configurations of the pair of points $p_i \in \Omega_i$, $p_j \in \Omega_j$ which have current labels $\lambda_i = \lambda(p_i)$ and $\lambda_j = \lambda(p_j)$ and such that $(p_i, p_j) \in A^\alpha$, namely $\mathcal{H}_\alpha^{ij} p_i = p_j$. Let v_i denotes the vertex of pixel p_i and v_j the one of pixel p_j . Let us recall that the source is S and the sink T . The weight of a directed edge that goes from vertex a to vertex b is denoted by e_{ab} . In practice, we have $a \in \{\mathcal{V}, S\}$ and $b \in \{\mathcal{V}, T\}$.

We also recall that the energy $E_{ij}^\alpha(x(p), x(q))$ involves $x(p)$ (resp. $x(q)$), the binary label of $p \in \Omega_i$ (resp. $q \in \Omega_j$). In an α -expansion algorithm, if $x(p) = 1$, then p will take the depth label α . If $x(p) = 0$, p will keep its current depth label λ .

Finally, remark that the energy E_{ij}^α does not consider its arguments p_i and p_j in a symmetric way. Namely, it counts the occlusion of p_i by p_j but does consider the occlusion of p_j by p_i . This second case will thus be treated with the energy E_{ji}^α . Otherwise, we would penalize twice the occlusion. For the visibility constraint this claim does not hold anymore, as we can sum infinites without changing the result.

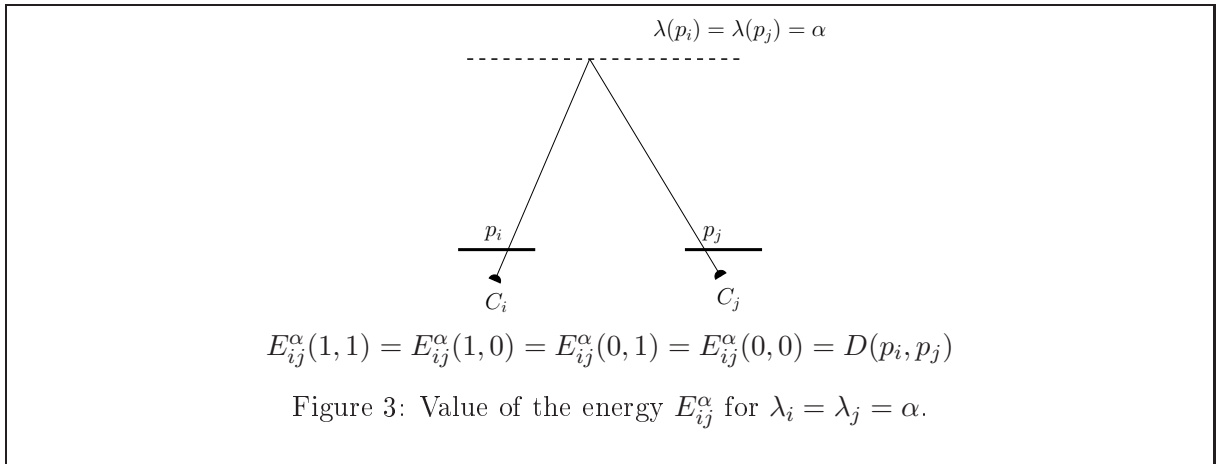
We will illustrate the different cases occurring during the α -expansion algorithm for a pixel $p_i \in \Omega_i$ of the camera C_i and its corresponding pixel $p_j = \mathcal{H}_\alpha^{ij} p_i \in \Omega_j$ of the camera C_j .

The following cases, that give a full description of E_{ij}^α , have to be considered:

- $\lambda_i = \lambda_j = \alpha$. As illustrated in figure 3, this configuration means that p_i and p_j already correspond with the label α and nothing will change, so

- * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$,
- * $E_{ij}^\alpha(1, 0) = D(p_i, p_j)$,
- * $E_{ij}^\alpha(0, 1) = D(p_i, p_j)$,
- * $E_{ij}^\alpha(0, 0) = D(p_i, p_j)$.

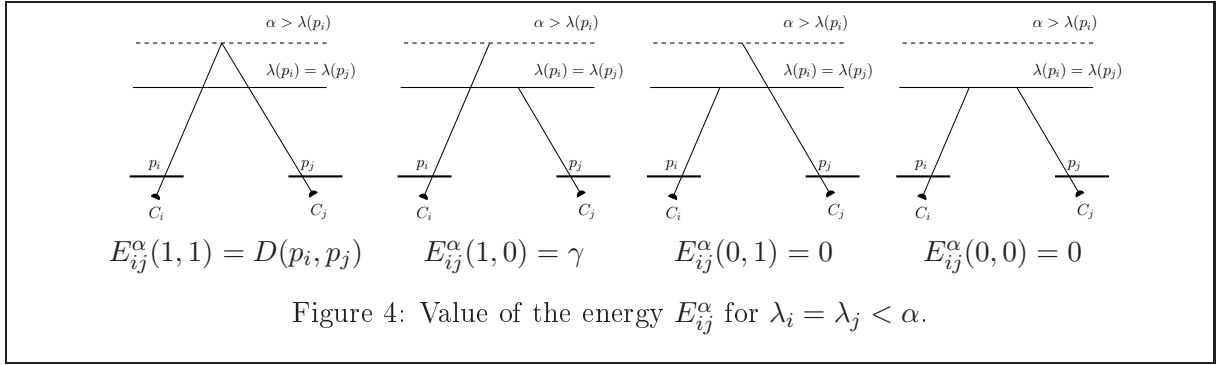
In terms of edges: $e_{Sv_i} = e_{v_i T} = D(p_i, p_j)$.



- $\lambda_i = \lambda_j \neq \alpha$. It means that p_i and p_j did not correspond with the current label. We have to distinguish two cases:

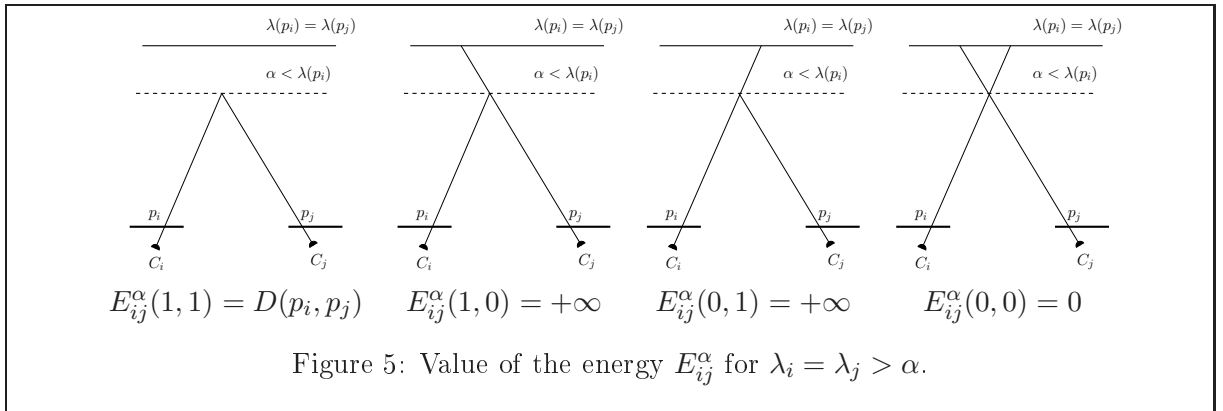
- if $\lambda_i < \alpha$, then
 - * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, the cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(1, 0) = \gamma$, p_j will occlude p_i in the image I_j (i.e. $p_i \in M_{ij}$),
 - * $E_{ij}^\alpha(0, 1) = 0$, p_i will occlude p_j (i.e. $p_j \in M_{ji}$),
 - * $E_{ij}^\alpha(0, 0) = 0$.

These cases are illustrated in figure 4. Notice that the third situation does not violate the visibility constraint since we do not have that $\mathcal{H}_{\lambda_i}^{ij} p_i = p_j$. In terms of edges: $e_{Sv_i} = D(p_i, p_j)$, $e_{v_j v_i} = \gamma - D(p_i, p_j)$.



- if $\lambda_i > \alpha$, then
 - * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, the cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(1, 0) = +\infty$, impossible. Otherwise, as $\mathcal{H}_\alpha^{ij} p_i = p_j$ and $\tilde{\lambda}_j = \lambda_j > \tilde{\lambda}_i = \alpha$, this would contradict the visibility constraint for (p_i, p_j) ,
 - * $E_{ij}^\alpha(0, 1) = +\infty$, impossible. Otherwise this would violate the visibility constraint for (p_j, p_i) ,
 - * $E_{ij}^\alpha(0, 0) = 0$.

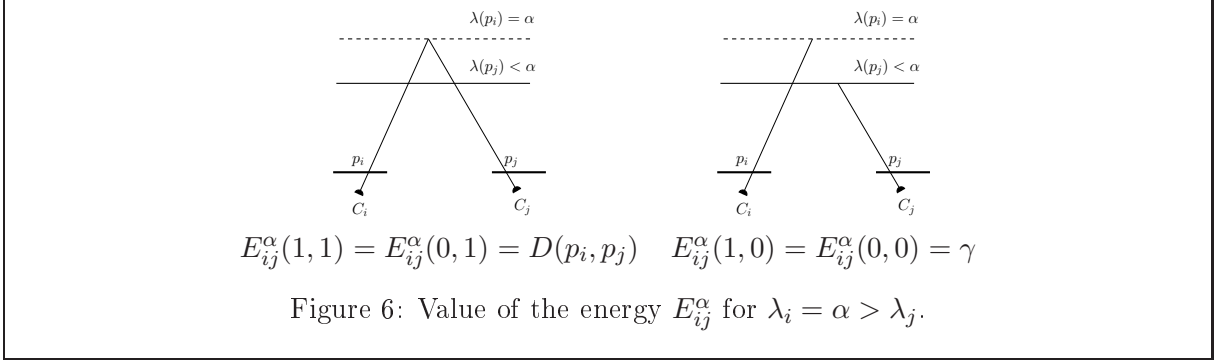
In terms of edges: $e_{Sv_j} = D(p_i, p_j)$, $e_{v_i v_j} = e_{v_j v_i} = \infty$. These cases are illustrated in figure 5.



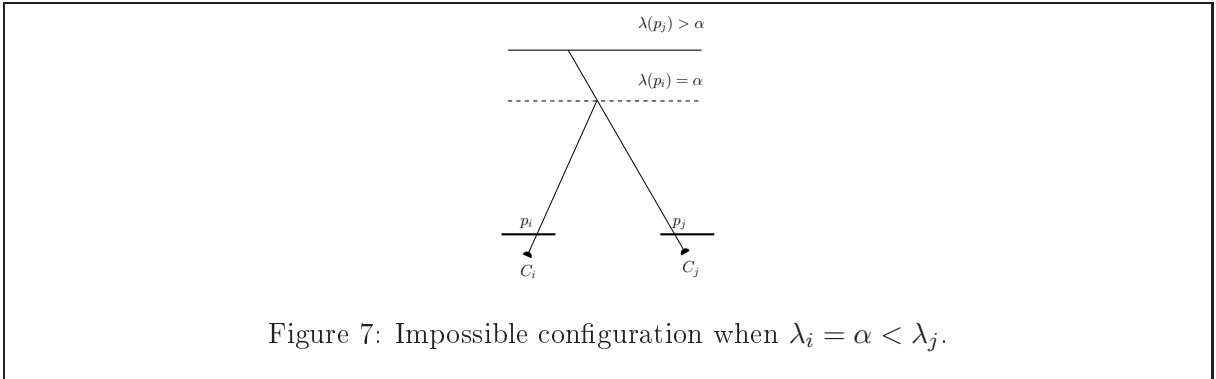
- $\lambda_i = \alpha \neq \lambda_j$. We have to distinguish two cases:

- $\lambda_j < \alpha$. It means that p_j was previously occluding p_i . Then
 - * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(1, 0) = \gamma$, p_j still occludes p_i ,
 - * $E_{ij}^\alpha(0, 1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(0, 0) = \gamma$, p_j still occludes p_i .

In terms of edges: $e_{Sv_j} = D(p_i, p_j)$, $e_{v_jT} = \gamma$. These cases are illustrated in figure 6.



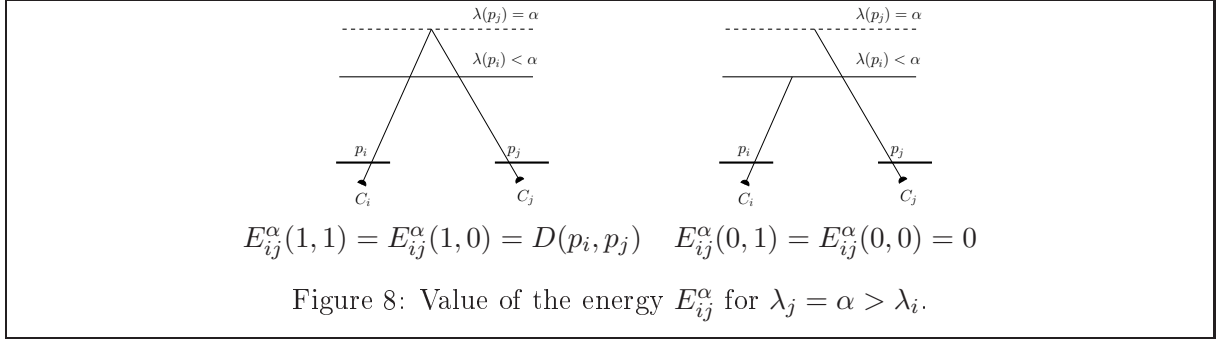
- $\lambda_j > \alpha$. As illustrated in figure 7, this configuration is impossible since we are assuming that λ satisfies the visibility constraint.



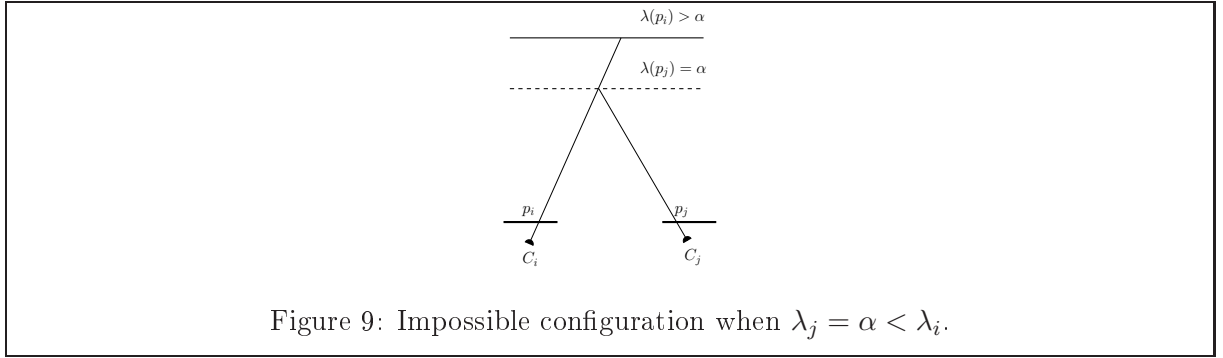
- $\lambda_j = \alpha \neq \lambda_i$. We have to distinguish two cases:

- $\lambda_i < \alpha$. It means that p_i was previously occluding p_j . Then
 - * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(1, 0) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(0, 1) = 0$, p_i still occludes p_j ,
 - * $E_{ij}^\alpha(0, 0) = 0$, p_i still occludes p_j ,

Notice that the last two cases do not violate the visibility constraint, since even if $\tilde{\lambda}_j > \tilde{\lambda}_i$, it does not hold that $\mathcal{H}_{\tilde{\lambda}_i}^{ij} p_i = p_j$. In terms of edges: $e_{Sv_i} = D(p_i, p_j)$. These cases are illustrated in figure 8.



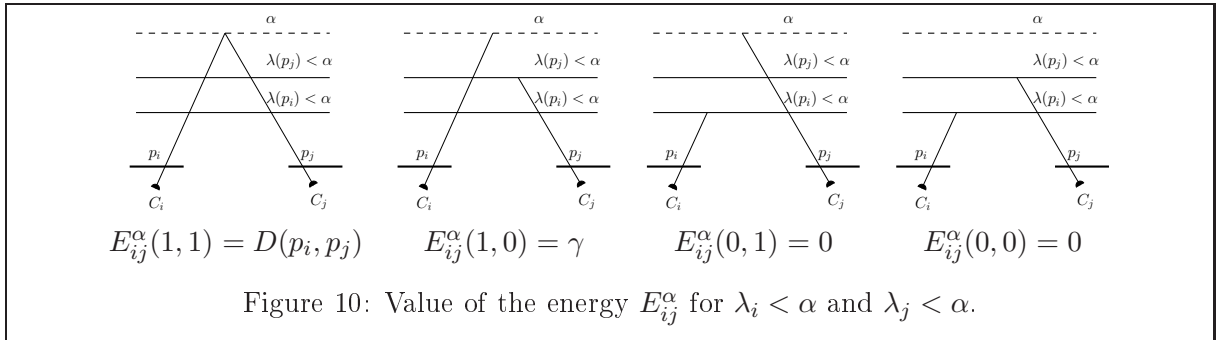
- $\lambda_i > \alpha$. As illustrated in figure 9, this configuration is impossible since we are assuming that λ satisfies the visibility constraint and this situation means that $\mathcal{H}_\alpha^{j_i} p_j = p_i$ while $\lambda_i > \lambda_j = \alpha$.



- $\lambda_i \neq \alpha$, $\lambda_j \neq \alpha$ and $\lambda_i \neq \lambda_j$. We have to distinguish four cases:

- $\lambda_i < \alpha$ and $\lambda_j < \alpha$. Then
 - * $E_{ij}^\alpha(1,1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
 - * $E_{ij}^\alpha(1,0) = \gamma$, p_j will occlude p_i ,
 - * $E_{ij}^\alpha(0,1) = 0$, p_i will occlude p_j ,
 - * $E_{ij}^\alpha(0,0) = 0$.

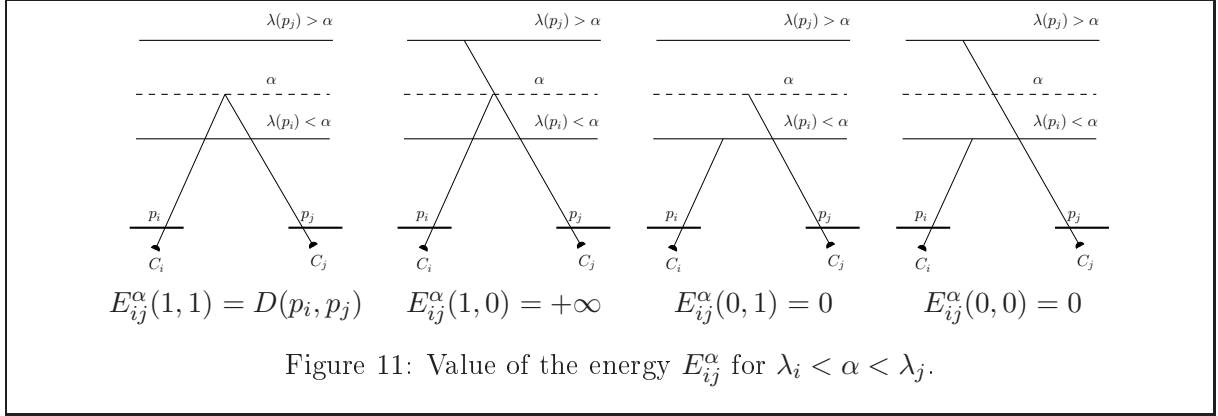
As above, the third situation does not violate the visibility constraint. In terms of edges: $e_{Sv_i} = D(p_i, p_j)$, $e_{v_jv_i} = \gamma - D(p_i, p_j)$. These cases are illustrated in figure 10.



- $\lambda_i < \alpha < \lambda_j$. Then

- * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
- * $E_{ij}^\alpha(1, 0) = +\infty$, otherwise this would violate the visibility constraint for (p_i, p_j) ,
- * $E_{ij}^\alpha(0, 1) = 0$, p_i will occlude p_j ,
- * $E_{ij}^\alpha(0, 0) = 0$.

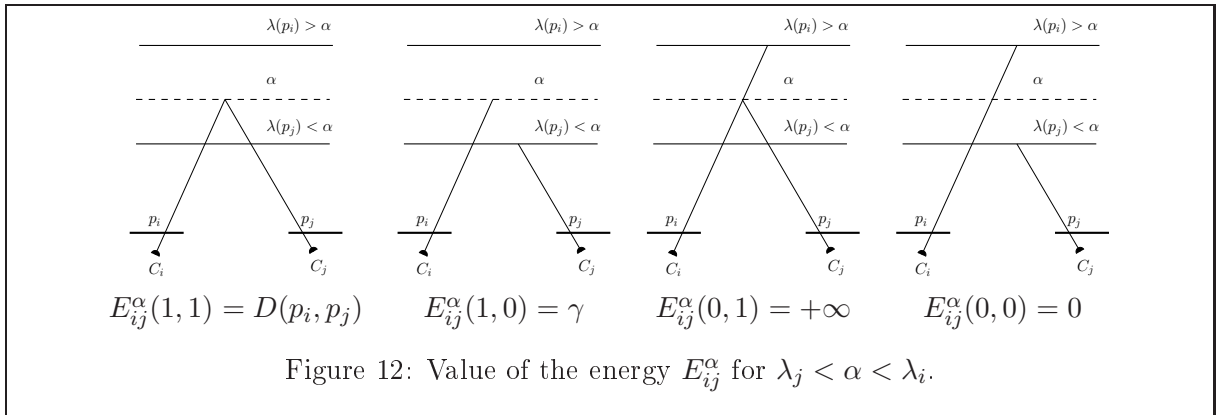
Notice again that the last two cases do not violate the visibility constraint. In terms of edges: $e_{Sv_i} = D(p_i, p_j)$, $e_{v_jv_i} = \infty$. These cases are illustrated in figure 11.



- $\lambda_j < \alpha < \lambda_i$

- * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
- * $E_{ij}^\alpha(1, 0) = \gamma$, p_j will occlude p_i in the image I_j ,
- * $E_{ij}^\alpha(0, 1) = +\infty$, otherwise we would violate the visibility constraint for (p_j, p_i) since we have $\mathcal{H}_\alpha^{j_i} p_j = p_i$ and $\tilde{\lambda}_j = \alpha < \tilde{\lambda}_i = \lambda_i$,
- * $E_{ij}^\alpha(0, 0) = 0$.

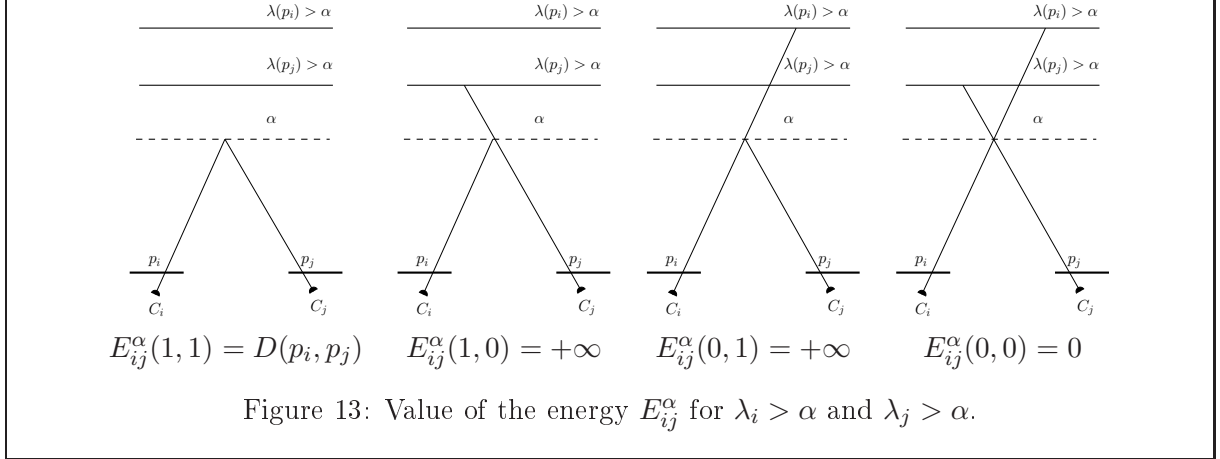
Observe that the last case does not violate the visibility constraint. In terms of edges: $e_{Sv_i} = D(p_i, p_j)$, $e_{v_jv_i} = \gamma - D(p_i, p_j)$, $e_{v_i v_j} = +\infty$. These cases are illustrated in figure 12.



- $\lambda_i > \alpha$ and $\lambda_j > \alpha$. Then

- * $E_{ij}^\alpha(1, 1) = D(p_i, p_j)$, cost of matching between p_i and p_j ,
- * $E_{ij}^\alpha(1, 0) = +\infty$, otherwise we would violate the visibility constraint for (p_i, p_j) ,
- * $E_{ij}^\alpha(0, 1) = +\infty$, otherwise we would violate the visibility constraint for (p_j, p_i) ,
- * $E_{ij}^\alpha(0, 0) = 0$.

Observe that the last case does not violate the visibility constraint. In terms of edges: $e_{Sv_i} = D(p_i, p_j)$, $e_{v_i v_j} = e_{v_j v_o} = \infty$. These cases are illustrated in figure 13.



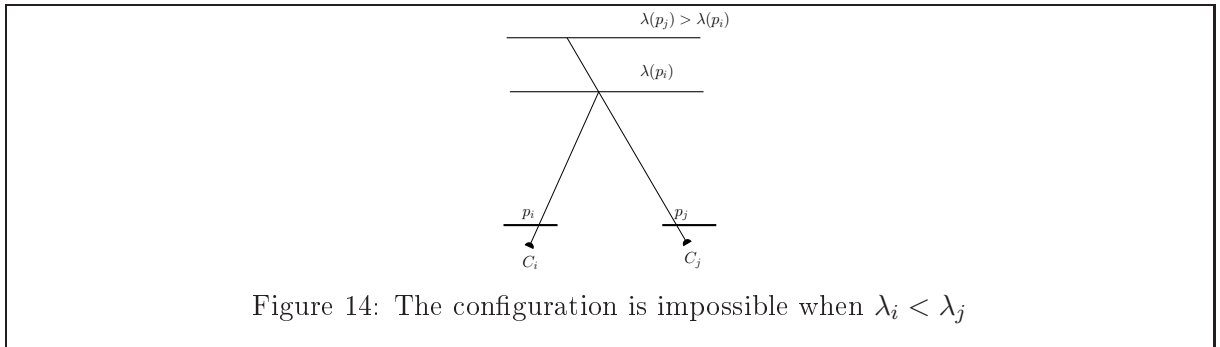
From the two cases $\lambda_i = \lambda_j < \alpha$ and $\lambda_i \neq \alpha$, $\lambda_i < \alpha$, $\lambda_j < \alpha$, it appears that the energy E_{ij}^α verifies the regularity condition (13) as soon as $D(p_i, p_j) \leq \gamma$.

B Energy E^0

Once again, v_i denotes the vertex of pixel p_i and v_j the one of pixel p_j . The weight of a directed edge that goes from vertex a to vertex b is denoted by e_{ab} . In practice, we have $a \in \{\mathcal{V}, S\}$ and $b \in \{\mathcal{V}, T\}$. We also recall that the energy $E_{ij}^0(x(p), x(q))$ involves $x(p)$ (resp. $x(q)$), the binary label of $p \in \Omega_i$ (resp. $q \in \Omega_j$). In an α -expansion algorithm, if $x(p) = 1$, then p will take the depth label α . If $x(p) = 0$, p will keep its current depth label.

We now describe the different possible geometric configurations for the pair of points $p_i \in \Omega_i$, $p_j \in \Omega_j$ which have the current labels $\lambda_i = \lambda(p_i) \neq \alpha$ and $\lambda_j = \lambda(p_j)$ and such that $(p_i, p_j) \in A^0$. We should underline that all the configurations $\lambda_i < \lambda_j$ are impossible as λ satisfies the visibility constraints and $\mathcal{H}_{\lambda_i}^{ij} p_i = p_j$.

From figure 14, we see that the cases $\lambda_i < \lambda_j$ are impossible.



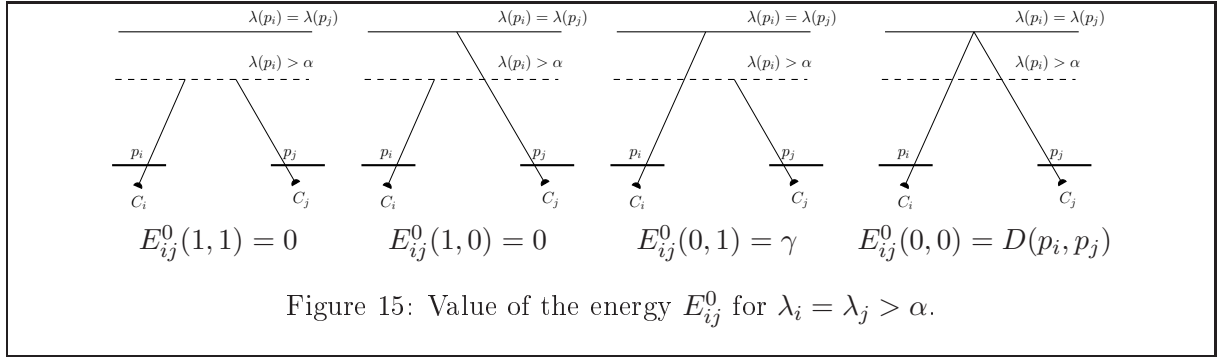
We then have to consider:

- $\lambda_i = \lambda_j \neq \alpha$. This configuration means that p_i and p_j correspond with the current label. We have to distinguish two cases:

- $\lambda_i > \alpha$. Then

- * $E_{ij}^0(1, 1) = 0$,
- * $E_{ij}^0(1, 0) = 0$, visibility is not violated, $p_j \in M_{ji}$ and the cost γ will be paid when considering the pair (j, i) ,
- * $E_{ij}^0(0, 1) = \gamma$, p_j will occlude p_i in the image I_i , i.e. $p_i \in M_{ij}$,
- * $E_{ij}^0(0, 0) = D(p_i, p_j)$, cost of keeping the matching between p_i and p_j .

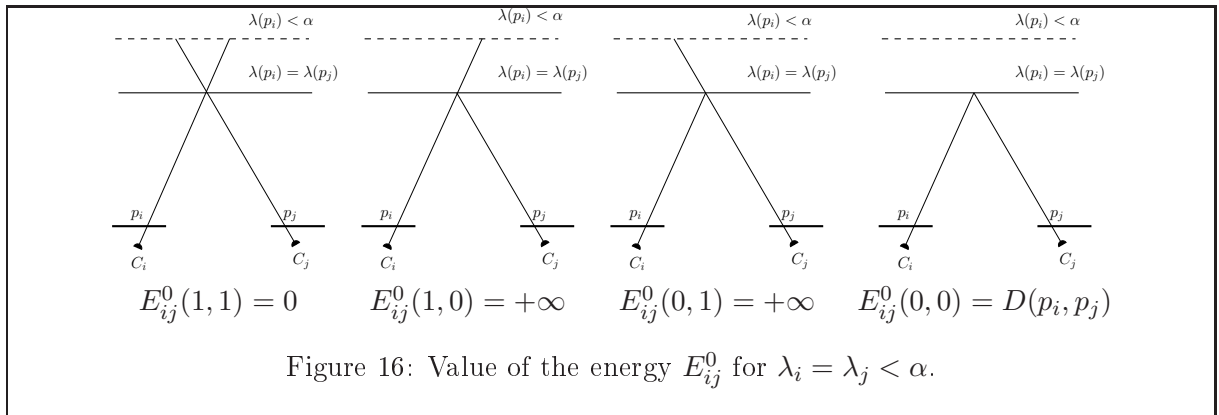
In terms of edges: $e_{v_i T} = D(p_i, p_j)$, $e_{v_i v_j} = \gamma - D(p_j, p_j)$. These cases are illustrated in figure 15.



- $\lambda_i < \alpha$. Then

- * $E_{ij}^0(1, 1) = 0$,
- * $E_{ij}^0(1, 0) = +\infty$, otherwise we would violate the visibility constraint for (p_j, p_i) ,
- * $E_{ij}^0(0, 1) = +\infty$, otherwise we would violate the visibility constraint for (p_i, p_j) ,
- * $E_{ij}^0(0, 0) = D(p_i, p_j)$, cost of keeping the matching between p_i and p_j .

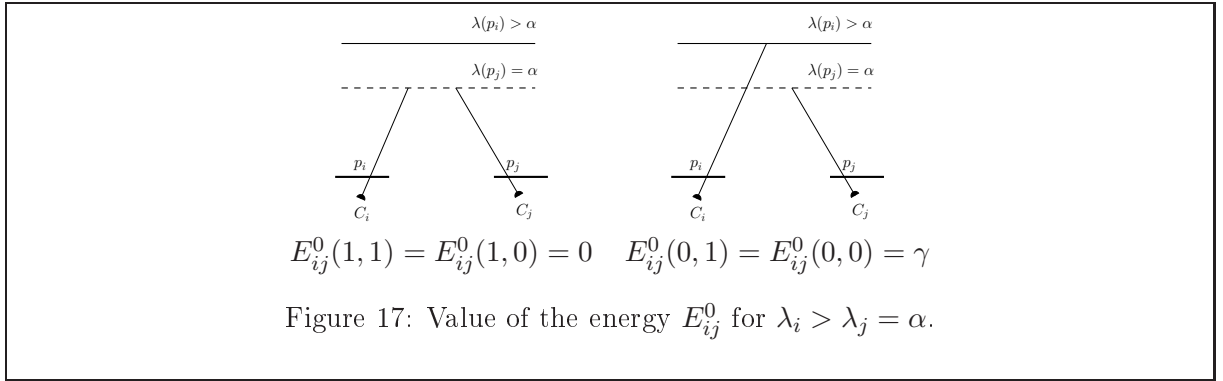
In terms of edges: $e_{v_i T} = D(p_i, p_j)$, $e_{v_i v_j} = e_{v_j v_i} = \infty$. These cases are illustrated in figure 16.



- $\lambda_i \neq \lambda_j = \alpha$. We have to distinguish two cases:

- $\lambda_i < \alpha$ This case is impossible, as it would mean that $\lambda_i < \lambda_j$.
- $\lambda_i > \alpha$ It means that p_j was occluding p_i in the image I_i . Then
 - * $E_{ij}^0(1, 1) = 0$,
 - * $E_{ij}^0(1, 0) = 0$,
 - * $E_{ij}^0(0, 1) = \gamma$, since still p_j occludes p_i in the image I_i , i.e. $p_i \in M_{ij}$,
 - * $E_{ij}^0(0, 0) = \gamma$, since still p_j occludes p_i in the image I_i , i.e. $p_i \in M_{ij}$.

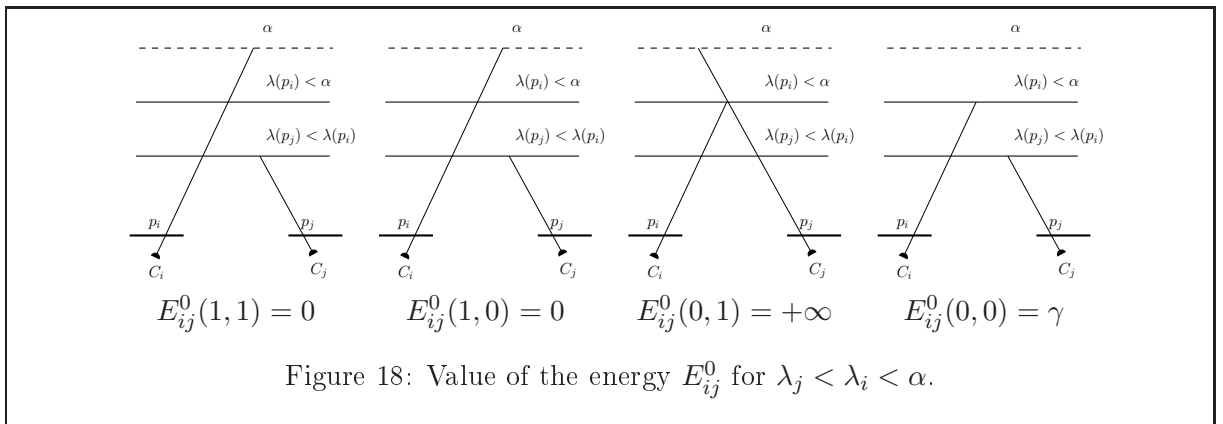
In terms of edges: $e_{v_i T} = \gamma$. These cases are illustrated in figure 17.



- $\lambda_i \neq \lambda_j$, $\lambda_i \neq \alpha$. We have to distinguish three cases (as the cases $\lambda_i < \lambda_j$ are impossible):

- $\lambda_j < \lambda_i < \alpha$. It means that p_j was occluding p_i in the image I_i . Then
 - * $E_{ij}^0(1, 1) = 0$,
 - * $E_{ij}^0(1, 0) = 0$,
 - * $E_{ij}^0(0, 1) = \infty$, otherwise we would violate the visibility constraint for (p_i, p_j) ,
 - * $E_{ij}^0(0, 0) = \gamma$, since p_j still occludes p_i in the image I_i , i.e., $p_i \in M_{ij}$.

In terms of edges: $e_{v_i T} = \gamma$, $e_{v_i v_j} = \infty$. These cases are illustrated in figure 18.



- $\lambda_j < \alpha < \lambda_i$. Then

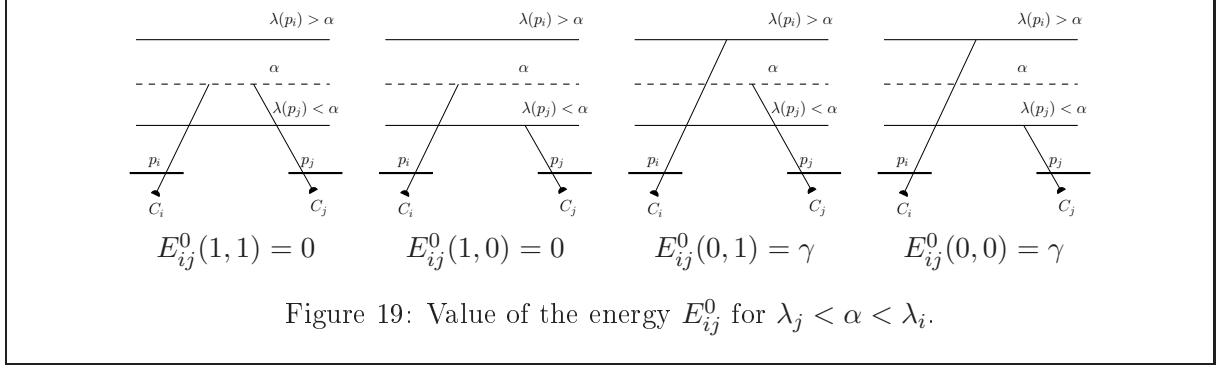
- * $E_{ij}^0(1, 1) = 0$,

- * $E_{ij}^0(1, 0) = 0$,

- * $E_{ij}^0(0, 1) = \gamma$, since p_j still occludes p_i in the image I_i , i.e., $p_i \in M_{ij}$,

- * $E_{ij}^0(0, 0) = \gamma$, since p_j still occludes p_i in the image I_i , i.e., $p_i \in M_{ij}$,

In terms of edges: $e_{v_i T} = \gamma$. These cases are illustrated in figure 19.



- $\alpha < \lambda_j < \lambda_i$. Then

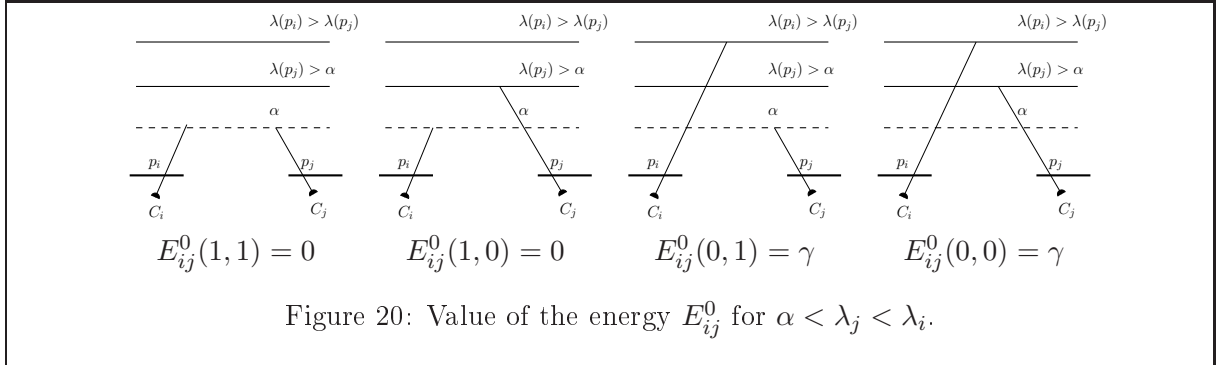
- * $E_{ij}^0(1, 1) = 0$,

- * $E_{ij}^0(1, 0) = 0$,

- * $E_{ij}^0(0, 1) = \gamma$, since p_j still occludes p_i in the image I_i , i.e., $p_i \in M_{ij}$,

- * $E_{ij}^0(0, 0) = \gamma$, since p_j still occludes p_i in the image I_i , i.e., $p_i \in M_{ij}$.

In terms of edges: $e_{v_i T} = \gamma$. These cases are illustrated in figure 20.



From case $\lambda_i = \lambda_j > \alpha$, it appears that the energy E_{ij}^0 verifies the regularity condition (13) as soon as $D(p_i, p_j) \leq \gamma$.

References

- [1] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the Second European Conference on Computer Vision*, Springer-Verlag, pages 237–252, 1992.

- [2] M.J. Black, and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.
- [3] A.F. Bobick and S.S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, 33(3):181–200, 1999.
- [4] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *IEEE Conf. Comp. Vis. and Pat. Recog. (CVPR'98)*, pages 648–655, 1998.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pat. Anal. Mach. Intell.*, 23(11):1222–1239, 2001.
- [6] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Comput. Vis. (ECCV'04)*. Springer-Verlag, 2004.
- [7] R. Collins. A space-sweep approach to true multi-image matching. In *In Proc. Comp. Vis. and Pat. Recog. Conf. (CVPR'96)*, pages 358–363, 1996.
- [8] A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by Helmholtz principle. *J. of Math. Imag. and Vis.*, 14:271–284, 2001.
- [9] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian J. of Mathematics*, 8:399–404, 1956.
- [10] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1):35–49, 1993.
- [11] Z. Gu, X. Su, Y. Liu, and Q. Zhang. Local stereo matching with adaptive support-weight, rank transform and disparity calibration. *Pat. Recog. Letters archive*, 29(9):1230–1235, July 2008.
- [12] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [13] H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Trans. on Pat. Anal. and Mach. Intell.*, 25(10):1333–1336, 2003.
- [14] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *Europ. Conf. Comp. Vis. (ECCV'98)*, volume 1, pages 232–248, 1998.
- [15] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920–932, 1994.
- [16] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Int. Conf. on Pat. Recog. (ICPR '06)*, pages 15–18, 2006.
- [17] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions via graph cuts. In *IEEE Int. Conf. Comp. Vis. (ICCV'01)*, volume 2, pages 508–515, 2001.

- [18] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Europ. Conf. Comp. Vis. (ECCV'02)*, pages 82–96, 2002.
- [19] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. on Pat. Anal. Mach. Intell.*, 26(2):147–159, 2004.
- [20] V. Kolmogorov, R. Zabih, and S. Gortler. Generalized multi-camera scene reconstruction using graph cuts. In *Energy Minimization Methods in Comp. Vis. and Pat. Recog.*, pages 501–516, 2003.
- [21] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155, 2002.
- [22] S. Roy and I. J. Cox. A maximum flow formulation of the n-camera stereo correspondence problem. In *IEEE Int. Conf. Comp. Vis. (ICCV'98)*, pages 492–502, 1998.
- [23] D. Scharstein. High-accuracy stereo depth maps using structured light. In *Conf. on comp. Vis. and Pat. Recog. (CVPR'03)*, pages 195–202, 2003.
- [24] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. of Comp. Vis.*, 47(1):7–42, 2002.
- [25] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 131-140, 2001.
- [26] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum. Symmetric stereo matching for occlusion handling. In *Conf. Comp. Vis. and Pat. Recog.n (CVPR'05)*, volume 2, pages 399–406, 2005.
- [27] Y. Taguchi, B. Wilburn, and L. Zitnick. Stereo reconstruction with mixed pixels using adaptive over-segmentation. In *Conf. Comp. Vis. and Pat. Recog. (CVPR'08)*, 2008.
- [28] Z. F. Wang and Z. G. Zheng. A region based stereo matching algorithm using cooperative optimization. In *Conf. Comp. Vis. and Pat. Recog. (CVPR'08)*, 2008.
- [29] L. Xu and J. Jia. Stereo matching: An outlier confidence approach. In *Europ. Conf. Comp. Vis. (ECCV'06)*, 2008.
- [30] Q. Yáng, L. Wang, R. Yang, H. Stewénus, and D. Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. *IEEE Trans. on Pat. recog. and Mach. Intell.*, 31(3):492–504, 2009.
- [31] C. L. Zitnick and S. B. Kang. Stereo for image-based rendering using image over-segmentation. *Int. J. of Comp. Vis.*, 75(1):49–65, October 2007.