ENERGY-MINIMIZATION BASED MOTION ESTIMATION USING ADAPTIVE SMOOTHNESS PRIORS

Tarik Arici[1] and Vural Aksakalli[2]
[1]*Department of Electrical Engineering, Istanbul Sehir University, Istanbul, Turkey*
[2]*Department of Industrial Engineering, Istanbul Sehir University, Istanbul, Turkey*

Keywords: Energy Minimization, Approximation Algorithms, Primal-dual Method, Motion Estimation.

Abstract: Energy minimization algorithms are used in low-level computer vision applications for labeling tasks such as stereo-disparity estimation, image restoration, motion estimation, and optical flow. The energy function involves terms that evaluate the goodness of a solution in terms of a prior knowledge in addition to data terms. The most widely used priors are smoothness-based priors, which enhance the quality significantly. However, the smoothness assumption is not valid across discontinuities (*e.g.* motion boundaries). We present a method to update the weights of smoothness terms using the dual problem when the approximation algorithm is iterative. The dual of the primal energy minimization problem is used to infer about the validity of the smoothness prior and impose it more correctly at each iteration. We demonstrate the effectiveness of this method against the state-of-the-art in the optical flow literature.

1 INTRODUCTION

Energy minimization refers to designing an energy function that describes the desired properties of a solution and its minimization to obtain a good solution. Most problems in computer vision such as motion/optical-flow estimation, stereo-disparity estimation, video synopsis, image formation modeling, texture segmentation are ill-posed (Boykov et al., 2001; Rav-Acha et al., 2008; Rav-Acha et al., 2006; Hofmann et al., 1996) in the Hadamard sense: many solutions exist and choosing one of the many solutions based solely on data results in over-fitting of the parameters. A model with over-fitted parameters is not good in prediction and also not effective in generalizing beyond data for truly explaining the observed phenomenon.

An energy function consists of a data term and a prior knowledge term to discriminate between the large number of solutions. Energy functions are often difficult to minimize because the solution space is generally exponentially large depending on the size of the problem. The data energy typically has a large number of non-convexities, which exacerbates the problem. One of the most commonly used prior knowledge is the smoothness of the solution, which is shown to be NP-hard (Boykov et al., 2001).

Energy minimization problems can be formulated as labeling problems. A typical energy is of the form

$$E(v) = \sum_{b \in B} D_b(l_b) + \lambda \sum_{b,a \in \mathcal{N}} V_{b,a}(l_b, l_a), \qquad (1)$$

where D_b is the data fidelity term, $V_{b,a}$ represents the prior knowledge as a penalty cost between two interacting sites b and a. B is the set of sites, $\mathcal{N}$ is the set of sites that interact with each other, and l_b is the label for site b. With n_b sites and n_l labels, the number of possible labellings is $n_b^{n_l}$.

NP-hard labeling problems will require exponential time to find the global minimum. To achieve practical algorithms, local search methods can be used to refine an initial solution iteratively (Papadimitriou and Steiglitz, 1998). When there are many local minimums, approximation algorithms that do a local search are likely to get stuck at a local minimum. The local minimums are created mostly due to the non-convexities in D_b. Fortunately, $V_{b,a}$ is generally convex because the prior knowledge is commonly a norm of a distance measure on the solution space or its derivatives. Hence, increasing the weight of $V_{b,a}$ in the energy will help with the convergence of the algorithm. But labels have discontinuities, which make $V_{b,a}$ invalid for specific interactions between sites and increasing its weight will falsely impose an invalid prior knowledge on the labels. To solve this problem, we propose to use the dual problem of the primal en-

Arici T. and Aksakalli V..
ENERGY-MINIMIZATION BASED MOTION ESTIMATION USING ADAPTIVE SMOOTHNESS PRIORS.
DOI: 10.5220/0003805202010207
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 201-207
ISBN: 978-989-8565-04-4

ergy problem. The primal problem for each iteration of a label is expressed as a linear programming problem and the equivalence of its Lagrangian to the energy under specific choices of the Lagrangian multipliers is shown. The dual solution is used to correctly impose the prior knowledge by adapting the weights in the smoothness prior via complementary slackness conditions.

In the next section we give a review of prior art in energy minimization. In Section 3, we present our primal-dual method for energy function design using a spatial smoothness prior. Finally, in Section 4, we demonstrate the power of our proposed method against the state-of-the-art in the optical flow literature.

2 RELATED WORK

Prior works on energy-minimization determines the energy function at the beginning by fixing the parameters manually, and focus on efficient minimization techniques. Energy-minimization is commonly used in computer vision applications (Rav-Acha et al., 2008; Rav-Acha et al., 2006), and it is applied to pixel level stereo-disparity estimation, which is similar to motion estimation in essence (Kolmogorov and Boykov, 2002; Boykov and Veksler, 2006; Veksler, 1999).

The majority of prior art focuses on the minimization problem rather than the design of the energy function. Graph-cut methods are powerful in quickly converging to good local minimums due to their ability to do large moves (Kolmogorov and Boykov, 2002; Boykov and Veksler, 2006; Veksler, 1999). However, they are computationally complex and are not hardware-friendly since they operate on node lists to find max flows in a graph. Modifying energy functions to avoid falsely imposing smoothness priors across discontinuities is proposed for object segmentation but the discontinuity is inferred using the contrast in the pixel intensities (Kohli and Torr, 2007). For optical flow changing the weights using the image gradients is also proposed (Werlberger et al., 2009). In these approaches, although the energy is adapted to the image (or data in general), it is still fixed throughout the iterations. Also, the adaptation may bias the solutions towards a bad local minimum since in some cases contrast in pixel intensity may not produce discontinuities/boundaries in the labels. Instead, we propose to modify the energy using the labels from the previous iteration, which becomes more reliable as the minimization algorithm improves the solution.

Belief propagation techniques are also used for minimizing the energy (Meltzer et al., 2005). The results achieved with belief propagation techniques are similar to graph-cut methods in quality.

Application of energy minimization to various computer vision problems is done by introducing new terms in the energy function to utilize prior information for the specific problem. For example, video synopsis tries to compress an image sequence in time, hence, requires the use of energy terms that imposes this via temporal energy terms (Rav-Acha et al., 2006). Another application such as unwrapping the surface of an object for video editing requires some other energy terms related to tracking and mosaic stitching (Rav-Acha et al., 2008).

In this work, we study the energy design aspect of energy minimization that uses first-order smoothness priors. We propose a primal-dual linear programming method that utilizes labels from the previous iteration to modify the smoothness priors in order to take the discontinuities into account.

3 PRIMAL-DUAL METHOD FOR ENERGY-MINIMIZATION

3.1 Formulation of a Move as a Linear Program

We define a move as a change of labels to a particular candidate label, ρ. A large move can change a group of site's labels to ρ, keeping all other labels the same as the pre-move labels l^*, while a standard-move can only change a single site's label to ρ.

We formulate the move problem for candidate label ρ as a linear programming problem (move-LP) in the canonical form

$$\min \sum_{b \in B} x_b D_b(\rho) + (1 - x_b) D) b(l_b^*) \qquad (2)$$

subject to

$$\begin{array}{lll} x_b - x_a \geq -p_{ba}, & x_a - x_b \geq -p_{ba}, & \forall \{b,a\} \in \mathcal{N}, \\ x_b \geq 0, & -x_b \geq -1, & \forall b \in B, \end{array}$$

where l^* is the pre-move labels, ρ is the candidate label of the move, and p_{ba} is the probability of b and a to be on different objects[1]. x is the variable of the move-LP, which is in $[0,1]$. $x_b = 1$ indicates that $l_b = \rho$, and $x_b = 0$ indicates that $l_b = l_b^*$. For $x_b \in (0,1)$, a fractional move is indicated, which can

[1]Generally, p_{ba}'s are neither known nor directly observed from data. However as discussed in Section 3.3 our technique does not require the knowledge of p.

be interpreted as b takes on ρ with probability x_b and retains l_b^* with probability $1-x_b$. The advantage of expressing the move problem as a linear problem is that it enables fractional moves during the iterations. To avoid confusion, it is important to note that the above equation is not a relaxed version of the energy minimization problem given in (1), but its Lagrangian with specific set of multipliers is identical to (1), which will be proved next.

The cost term in (2) is a weighted sum of data terms with labels ρ and l_b^*. Two inequality constraints are imposed on the move to achieve smoothness: $x_b - x_a \geq -p_{ba}$ and $x_a - x_b \geq -p_{ba}$. They are canonical form linear constraints for $|x_b - x_a| \leq p_{ba}$. With these constraints neighbor sites that are likely to move together are forced to move together depending on the probability, p_{ba}, of b and a to have different labels. Hence, smoothness is enforced on actions but not on the labels. However, a sequence of smooth moves will create a smooth label set in the end.

We next write the Lagrangian for the move-LP in (2),

$$\mathcal{L}(x,\beta,\gamma,\kappa) = \sum_{b\in B} x_b D_b(\rho) + (1-x_b)D_b(l_b^*) + \sum_{\{b,a\}\in\mathcal{N}} \beta_{ba}(-x_b + x_a - p_{ba}) + \beta_{ab}(-x_a + x_b - p_{ba}) + \sum_{\{b,a\}\in(N)} -\gamma_b x_b + \kappa_b(x_b - 1), \quad (3)$$

where β, γ, and κ are positive Lagrange multipliers associated with their corresponding inequality constraints.

Lemma 1. *A binary solution[2] x of the move-LP corresponds to a move from l^* to l^{LP}.*

Proof. By the definition of x, $l_b = \rho$ if $x_b = 1$, and $l_b = l_b^*$ if $x_b = 0$, which means a change of labels to ρ while keeping all other labels same as in l^*. Therefore, a binary x corresponds to a move from the previous label set l^* to a new label set l^{LP}, which is the solution to our linear program. □

Using Lemma 1, and the Lagrangian in (3) yields:

Theorem 2. *The Lagrangian of the move-LP with a binary solution x is equal to the energy, $E(l^{LP})$, if the Lagrangian multipliers satisfy*

$$\begin{array}{ll} \beta_{ba}p_{ba} = \beta_{ab}p_{ab} = -\frac{1}{2}\lambda V_{b,a}(l_b^*, l_a^*) & s.t.\ x_b = 0, x_a = 0 \\ \beta_{ba} = 0,\ \beta_{ab} = 0 & s.t.\ x_b = 1, x_a = 1 \\ \beta_{ba} = 0.\beta_{ab}(1-p_{ba}) = \lambda V_{b,a}(\rho, l_a^*) & s.t.\ x_b = 1, x_a = 0 \\ \beta_{ba}(1-p_{ba}) = \lambda V_{b,a}(l_b^*, \rho), \beta_{ab} = 0 & s.t.\ x_b = 0, x_a = 1 \\ \gamma_b = 0, \kappa_b = 0, & \end{array} \quad (4)$$

where l^{LP} is a move away from l^*.

Proof. We skip the proof due to space limitations.[3] □

3.2 Sub-optimality of the Energy

Theorem 2 shows that for a specific choice of the Lagrangian multipliers, the Lagrangian of the move-LP problem is equal to the energy under a binary move. The conditions for the Lagrangian multipliers for this equality may not be the best in terms of convergence and energy function design, because the Lagrangian multipliers are chosen without utilizing any information learnt from the iterations. By intelligently choosing the Lagrangian multipliers, contributions of the constraints in the Lagrangian can be better adjusted. This will especially improve the application of the smoothness constraints for sites that are in the vicinity of a discontinuity.

To achieve our goal, we utilize the *primal-dual schema* (Papadimitriou and Steiglitz, 1998). This technique constructs a feasible solution of the dual problem and an integer solution of the primal problem iteratively. Since a solution of the dual problem also provides lower bound on the primal problem by the weak duality, one can compute a feasible solution for the dual problem first which then can be used to find the corresponding primal solution via the complementary slackness conditions (Papadimitriou and Steiglitz, 1998).

Our method differs from the primal-dual schema by the way complementary slackness conditions (CSCs) are utilized. In the conventional primal-dual schema, CSCs are used to obtain the integer primal solution from the dual solution. However, we want to design a uniform algorithm, which has one type of computation kernels, which iteratively update the labels by substituting in the energy. Iterations of the same forms of energy function are more suitable for parallel implementations rather than an approach that involves minimizing an energy followed by solving the CSCs. In our method, CSCs are used to infer if a constraint of the move-LP is *tight* or *slack*. A slack constraint means the constraint is satisfied as a strict inequality, and a tight constraint means the constraint is satisfied with equality. By learning if a constraint is slack or tight via the dual feasible solution and the CSCs, the Lagrangian multipliers can be adjusted because a slack constraint means our prior information agrees with our current solution, and a tight constraint mean our prior information disagrees with our current

[2] A binary solution is defined such that each component of x is either zero or one.

[3] The proof can be found in the first author's Ph.D thesis.

solution. Compared to choosing the Lagrangian multipliers as specified by the condition of Theorem 2 in an ad hoc manner, our primal-dual method uses the feasible dual solution to determine the Lagrangian multipliers, which is then used in the Lagrangian that is to be minimized. We present the dual problem of energy minimization in the following section.

3.3 The Dual Problem

We start with rewriting the move-LP problem in (2) in matrix form for a compact representation. The primal problem is

$$\min_x c^T x + \bar{c}^T(\mathbf{1} - x) \tag{5}$$

subject to

$$\begin{aligned} Ax &\geq -p \\ x &\geq 0, -x \geq -1, \end{aligned}$$

where $c, \bar{c}, x \in \mathbb{R}^{|\mathcal{N}|}$, and $A \in \mathbb{R}^{|\mathcal{N}| \times |B|}$ such that $c_b = D_b(\rho)$ and $\bar{c}_b = D_b(v_b^*)$, and $A_{ki} = 1$, $A_{kj} = -1$ for any k^{th} pair $\{i, j\} \in \mathcal{N}$. It is straightforward to construct the remaining labels so that the above problem is equal to the original move-LP problem. The Lagrangian can be formed similar to (3):

$$\begin{aligned} \mathcal{L}(x, \beta, \gamma, \kappa) = & \quad c^T x + \bar{c}^T(\mathbf{1} - x) + \beta^T(-p - Ax) - \\ & \gamma^T x + \kappa^T(x - \mathbf{1}) \\ = & \quad (c - \bar{c} - A^T\beta - \gamma - \kappa)^T x - \beta^T p - \kappa^T \mathbf{1} \\ = & \quad m^T x - \beta^T p - \kappa^T \mathbf{1}, \end{aligned} \tag{6}$$

where m is substituted for $c - \bar{c} - A^T\beta - \gamma - \kappa$ for compactness. One can see that because of the way the constraints of the primal problem are utilized in the Lagrangian, the Lagrangian is always smaller than the primal problem's objective

$$\mathcal{L}(x, \beta, \gamma, \kappa) \leq c^T x + \bar{c}^T(\mathbf{1} - x), \tag{7}$$

given that x is a feasible point in (5).

The dual function is a minimization of (6) on x:

$$\begin{aligned} g(\beta, \gamma, \kappa) &= \inf_x \mathcal{L}(x, \beta, \gamma, \kappa) \\ &= \begin{cases} -\beta^T p \quad -\kappa^T \mathbf{1} & \text{if } m = 0 \\ -\infty & \text{otherwise} \end{cases} \end{aligned} \tag{8}$$

For the dual function to exist m must be equal to 0. It is straightforward to show that the dual function provides lower bounds on the more complex primal problem's optimal value by observing (7) and (4).

The dual problem is the maximization of the dual function, which becomes a minimization by changing the sign of the objective function:

$$\min_{\beta, \gamma, \kappa} \beta^T p + \kappa^T \mathbf{1} \tag{9}$$

subject to

$$c - \bar{c} - A^T\beta - \gamma - \kappa = 0, \tag{10}$$

The dual problem turns out to be the minimum cost flow problem with costs given as the probabilities of two sites belonging to different objects, which can be solved using the maximum flow algorithms. The required flow is specified by $c - \bar{c} - A^T\beta - \gamma - \kappa$, while flows across discontinuities (high p values) are discouraged. We are looking for a feasible solution of the dual problem, not necessarily the optimal solution. An optimal solution requires more computation and knowledge of p, which specifies the probability of two sites having different blocks, which is not known. p values can be inferred from data similar to (Kohli and Torr, 2007) by using intensity gradient, but in this work we will use a feasible solution of the dual problem that does not require the knowledge of p.

Dual feasibility condition is specified by the required flow condition given by

$$c - \bar{c} - A^T\beta - \gamma - \kappa = 0, \tag{11}$$

And β can be estimated by

$$\hat{\beta} = A^{\dagger}(c - \bar{c} - \gamma), \tag{12}$$

where $A^{\dagger}$ is the pseudo-inverse of A, κ is set to zero in the second equation, because the dual function that we want to minimize is an increasing function of κ[4]. Since $A^{\dagger}(c - \bar{c})$ can be negative, the non-negativity assumption of $\hat{\beta}$ required for the derivation of the dual function can be violated. Hence, we need to threshold its components to zero from below. Since γ is unknown, one can choose γ to minimize the need for this clamping operation. At any rate, we can rewrite (12) to get the dual feasibility condition (DFC)

$$\hat{\beta} = \max\{0, A^{\dagger}(c - \bar{c})\}. \tag{13}$$

Using a feasible solution for $\hat{\beta}$ of the dual problem, we can understand if the constraints of the primal problem are loose or tight, which can be used to re-design the energy function for the next iteration. Next section discusses this relation between $\hat{\beta}$ and the smoothness constraints, and also how it is used for energy re-design.

[4]We note that $\kappa = 0$ does not necessarily produce the optimal solution of the dual problem.

3.4 The Primal-dual Relation

First, we describe a method to solve for β. To have a unique solution for the DFC equation, $A^{\dagger}$ must be well-defined. However, by the way we defined the constraints in the move-LP each block pair $\{a,b\}$ has two constraints

$$x_b - x_a \geq -p_{ba} \tag{14}$$

$$x_a - x_b \geq -p_{ba}. \tag{15}$$

To find a relation between β and move-LP constraints, we utilize the dual complementary slackness conditions (Papadimitriou and Steiglitz, 1998). If the complementary slackness conditions are satisfied, any feasible solution x of the primal and β of the dual problem are optimal. The dual complementary slackness condition (DCSC) is given below.

$$\text{Either } \beta_{ba} = 0 \text{ or } A_{ba}x = -p_{ba}, \tag{16}$$

where $\{b,a\}$ is any site pair and A_{ba} is the row for the constraint associated with β_{ba}. The complementary slackness conditions are important in the design of efficient approximation algorithms for complex problems. These algorithms are called primal-dual based methods and they try to improve x and β by modifying them in a way that more of their components satisfy the complementary slackness conditions (Papadimitriou and Steiglitz, 1998).

Our goal for applying the primal-dual method is to re-design the energy function, which corresponds to the Lagrangian of the primal problem. The Lagrangian is showed to be equivalent to the energy function under the conditions on the Lagrangian multipliers given in Theorem 2. However, our choice of the constraints were imposed for all sites pairing with a site, independent of the discontinuity since the unknown l is what we want to estimate. Fortunately, $\hat{\beta}$ reveals some information on the connectedness of sites with their neighbors by the use of DCSC. DCSC dictates that if $\hat{\beta}$ is non-zero, then the constraint is binding and must be satisfied with equality. If the constraint had been relaxed, the interacting sites would have chosen different labels. This implies that the two sites are on different label segments. Hence, the constraint that is binding is conflicting with our smoothness of l assumption in that locality, and should be imposed less by decreasing its contribution in the energy. To this end, the Lagrangian multipliers must be inversely related to $\hat{\beta}$, for example by multiplying with a function $f(\hat{\beta})$ that is decreasing with $\hat{\beta}$ and has range $[0,1]$. Hence, by DCSC we propose to update the β values as

$$\beta_{ba}^{DCSC} = f(\hat{\beta}_{ba})\beta_{ba}, \tag{17}$$

where $\hat{\beta}$ is a feasible solution of the dual problem given in (13), and f is a decreasing function of $\hat{\beta}$. Replacing Lagrangian multiplier β with β^{DCSC} and substituting $\kappa = 0$ from the solution of the dual problem, Lagrangian in (3) after some simplification becomes

$$E(v) = \sum_{b \in B} D_b(v_b) + \lambda \sum_{\{b,a\} \in \mathcal{N}} f(\hat{\beta}) V_{b,a}(v_b, v_a). \tag{18}$$

The above energy formulation does not isotropically enforce a smoothness constraint, but adapts the weights of constraints in the energy with information derived from the data via the dual problem. This will enable us to obtain an energy function that is more powerful to explain labeling across discontinuities.

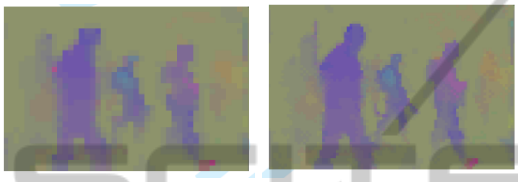
4 EXPERIMENT RESULTS

We present experiment results using the proposed method and compare our results with the state of the art in the optical flow literature. Our implementation of the proposed method is a hierarchical motion estimation algorithm that uses a full resolution and a half resolution image produced by down-sampling the full resolution image by two. For each half resolution block, a motion search is performed to pick the two best motion vectors to minimize a cost. The cost is a sum of absolute deviation (SAD) based cost and all motion vectors in a 2-D search window are evaluated to find the minimum cost vectors. This way an initial motion-vector field v^0 is created and refined using N standard-move iterations. Candidate vectors in reduced search window S is used to pick the best-two standard move for each b. S consists of 18 vectors obtained from a block's and its eight-connectivity neighborhoods' best-two vectors. Smoothness cost is derived from blocks in four-connectivity neighborhood and weights of the neighboring blocks is adapted using information from the dual problem. Each block is partitioned to quarter blocks so that a half-resolution quarter-block matches with a full resolution block in size. This will increase the reliability of the centered motion search in full resolution. Also, partitioning to quarter blocks increases the quality of half resolution vectors in general, since with a smaller block size block-based translational-motion model is less problematic for rotation, zooming, and motion boundaries. After the partitioning, standard-move iterations are again applied to refine further, before passing to full resolution. Execution of our algorithm on a sample instance of the underlying problem is illustrated in Figures 1 and 2. FR-FULL, FR-QUARTER, and FR-QUARTER^2 images in Figures 1 and 2 have resolutions of 16x16, 8x8, 4x4 pixel blocks respectively.

Table 1: Comparison of our DCSC-based method to Weickert et al.'s method on the Middlebury dataset. Mean endpoint error of our method is comparable, if not superior, to that of Weickert et al.'s method even though our method uses only half the pixel accuracy.

Mean Endpoint Error	Schefflera			Wooden			Grove			Urban			Teddy		
	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
Method of Weickert et al. (2005)	1.12	1.80	0.99	1.07	2.06	1.12	1.23	1.52	1.62	1.54	2.15	0.96	1.38	2.26	1.83
DCSC-based method	1.21	1.77	1.18	0.94	2.03	0.97	1.20	1.57	1.08	1.73	1.90	1.12	1.37	2.16	1.81

The super-imposed image in Figure 4 shows the video frame with color plane modulated using the estimated motion vectors.



(a) (b)

Figure 1: (a) FR-FULL and (b) FR-QUARTER.

(a) (b)

Figure 2: (a) FR-QUARTER^2 and (b) Super-imposed image.

We now compare our method's performance against that of Weickert et al.'s method (Weickert et al., 2005), which is a state-of-the-art technique in optical flow literature. We use the Middlebury dataset available on the Web at *http://vision.middlebury.edu/flow* for benchmarking purposes. This dataset has four types of data with different characteristics and error measures distinguished for motion discontinuities and textureless regions. Comparison of our DCSC-based method to Weickert et al.'s method is given in Table 1. Although performance of our method is comparable to that of Weickert et al. (2005) in mean endpoint error, our method performs better in 4 out of 5 discontinuity cases; as expected. Our motion vectors operate with only half pixel accuracy because we are searching in a reference image upscaled by two. It is therefore impressive that even though our current implementation uses only of half available pixels, our method performs better than that of Weickert et al. (2005) in 10 out of the total 15 cases. This shows that adapting the smoothness prior using the dual problem is a promising technique. It is, however, important to observe that we are not finding the optical flow, but rather performing motion estimation via motion vector searching. As future work, we plan to perform further benchmarking on different datasets and compare our method to other optical flow techniques in the literature.

5 CONCLUSIONS

Energy minimization enables incorporation of prior knowledge for improving parameter estimation. Unfortunately, the weights of the prior knowledge in the energy function is not known. We proposed to use a feasible solution of the dual of the primal minimization problem to adjust the weights, which improves the energy when the prior knowledge is not valid. In the case of a smoothness prior knowledge, which is commonly used in computer vision problems, the dual feasible solution is used to infer about the discontinuities via the dual complementary slackness conditions.

REFERENCES

Boykov, Y. and Veksler, O. (2006). *Mathematical Models of Computer Vision: The Handbook, Graph Cuts in Vision and Graphics: Theories and Applications*. Dover Publications.

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

Hofmann, T., Puzicha, J., and Buhmann, J. (1996). Unsupervised segmentation of textured images by pairwise data clustering. In *International Conference on Image Processing*.

Kohli, P. and Torr, P. (2007). Dynamic graph cuts for efficient inference in markov random fields. *Pattern Analysis and Machine Inteligence*, 29(12):2079–2088.

Kolmogorov, V. and Boykov, Y. (2002). An experimental comparison of mincut/max-flow algorithms for energy minimization in vision. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*.

Meltzer, T., Yanover, C., and Weisss, Y. (2005). Globally optimal solutions for energy minimization in stereo

vision using reweighted belief propagation. In *International Conference on Computer Vision*, pages 428–435.

Papadimitriou, C. H. and Steiglitz, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications.

Rav-Acha, A., Kohli, P., Rother, C., and Fitzgibbon, A. (2008). Unwrap mosaics: A new representation for video editing. *ACM Transactions on Graphics (SIGGRAPH 2008)*.

Rav-Acha, A., Pritch, Y., and Peleg, S. (2006). Making a long video short: Dynamic video synopsis. *IEEE Conference on Computer Vision and Pattern Recognition*.

Veksler, O. (1999). *Efficient Graph-Based Energy Minimizatation Methods in Computer Vision*. Cornell University.

Weickert, J., Bruhn, A., and Schnorr, C. (2005). Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231.

Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., and Bischof, H. (2009). Anisotropic huber-l1 optical flow. In *British Machine Vision Conference*.