STEREO MATCHING USING REDUCED-GRAPH CUTS

Ayman Zureiki   Michel Devy   Raja Chatila

LAAS-CNRS, 7 avenue du Colonel Roche, 31077 Toulouse France
{ayman.zureiki, michel.devy, raja.chatila}@laas.fr

ABSTRACT

Some recent stereo matching algorithms are based on graph cuts. They transform the matching problem to a minimisation of a global energy function. The minimisation can be done by finding out an optimal cut in a special graph. Different methods were proposed to construct the graph. But all of them, consider for each pixel, all possible disparities between minimum and maximum values. In this article, a new method is proposed: only some potential values in the disparity range are selected for each pixel. These values can be found using a local analysis of stereo matching. This method allows us to make wider the disparity range, and at the same time to limit the volume of the graph, and therefore to reduce the computation time.

***Index Terms*—** Stereo vision, matching, graph cut

1. STEREO VISION

Binocular stereo vision use two images taken by two cameras. A preliminary phase of calibration is needed to estimate the different parameters of a stereo rig: the parameters of projection of each camera (for a pin hole geometric model) and the spatial relationship between the two cameras. This knowledge allows us to calculate the 3D coordinates of a point from its projections in the two images by a simple triangulation.

Stereo matching is the problem of finding in the left and right images the homologous primitives, i.e. the primitives which are the projections of the same entity in the scene. A stereo matching is a minimisation problem, either for many energy functions representing the local costs of each matching supposed independent of the others, or for a unique global energy function, representing the global cost of matching between the primitives of the images. Chambon [1] develop in her thesis a well detailed state of the art of stereo correspondence. A recent taxonomy of algorithms of stereo correspondence is written by Scharstein and Szeliski [2]. The authors of this taxonomy differentiate four elements in the methods of stereo matching: (i) The local cost of matching, (ii) The aggregation area while calculation the local cost, (iii) The optimisation method, (iv) Refinement of results.

The method developed at LAAS since 1995 [3] is a modified algorithm described by Faugeras [4]. This method is adopted for robotic applications, especially for the real time constraint. It can be classified under local methods, because there is not an optimisation phase. We try to adapt a global method of optimisation as second layer around our algorithm, to gain the advantages of the global minimisation while remaining within the real time constraint.

In the next section, we will resume related works on using graph cuts for stereo correspondence, then in section 3 we will detail the construction of the graph in order to put in evidence our contribution concerning its reduction, and in section 4, we focus on our implementation and experimental results before concluding.

2. STEREO CORRESPONDENCE AND GRAPH CUTS

The first global method based on graph cuts for stereo correspondence was introduced by Roy [5]. Stating from the 1D formulation of the order constraint used by the dynamic programming applied separately to each image line, Roy tried to find a more general 2D formulation for this constraint to be applied to all lines together. He proposed a *local coherence* constraint which suggests that the disparity function is locally smooth, which means that the neighbour pixels in all directions have similar disparities. He claims that the advantages of this constraint is that it can link not only neighbour pixels of one epipolar line but between lines. Roy applied this constraint of local coherence by defining a disparity matching cost which depends on the variation of intensities of matched pixels. In the case of two cameras, the matching cost is the squared difference of the intensities.

The next step in the method proposed by Roy is to resolve the optimal disparity map over all the image. This can be visualised as a 3D mesh composed of planes which are composed of image of nodes. There is a plane for each level of disparity, and each node represents a matching between two pixels in the original images.

The 3D mesh is then transformed into a graph of maximal flow by connecting each node to its four neighbours in the same plane by edges called occlusion edge, and with the two nodes in the neighbour planes with edges called disparity edges. Edges are not oriented. We add two special nodes: a source connected to all nodes in the plane of minimum disparity, and a sink connected to all nodes in the plane of maximum disparity. The weight of a disparity edge is equal to the mean value of matching costs of the two nodes. For occlusion edges, the weight is multiplied by a constant to control the smoothness of the optimal disparity map. A graph cut will separate the nodes in two sub-sets: the optimal disparity map is constructed by the assignment of each pixel with the most bigger value of disparity for which the corresponding node is still connected to the source.

Ishikawa and Geiger [6] pointed out that the method of Roy can deal only with convex maps. Thereby, it can only take into account linear penalties on disparity, which may lead to mediocre results due to over smoothing of disparities. They proposed a novel graph with oriented edges, then it is possible to reinforce the constraint of uniqueness and order. But, their method is also weak at discontinuities because of linear penalties.

Sub-optimal optimisation algorithms: Boykov, Veksler and Zabih [7] proposed another method to resolve the stereo corresponding using graph cuts. The authors showed that the problem of stereo corresponding can be formulated by a *Markov Random Field* (MRF). They showed that the estimate MAP (Maximum A Priori) of such MRF, can be obtained by a minimal multiway cut, using a maximum flow. The advantages of such a method is that it accepts non linear penalties of discontinuity, and then it gives more precise disparity maps especially near objects' edges. As the general problem of minimal multiway cut is NP-complete (see Dahlhaus et al. [8]), Boylov et al. decided to introduce an approximated algorithm, which can resolve iteratively some sub-problems until convergence.

This approach has a wider application spectrum than the one proposed by Roy or by Ishikawa and Geiger. But, it is iterative and sub-optimal, then its convergence speed and the quality of the obtained minimum must be supervised.

3. GRAPH CONSTRUCTION

A graph is a set of sites (called also node or vertex) connected by edges. In a weighted graph, each edge has a weight (capacity). Roy [5] used the graph cuts to determine the minimum of global energy function. Veksler [9] reformulated the problem as a *Labelling Problem*. In such a problem, we have a set of sites and a set of labels. The sites represent the features of the image (pixels, segments...), for which we want to estimate some quantity. The labels represent the quantities associated to these sites: intensity, disparity... Let $\mathcal{P} = 1, 2, \ldots, n$ be a set of n sites, and $\mathcal{L} = \{l_1, ..., l_k\}$ be a set of k labels. Labelling is defined by a map from $\mathcal{P}$ into $\mathcal{L}$:

$$f : \mathcal{P} \to \mathcal{L} : s_p \mapsto f_p = f(s_p) = l_i \tag{1}$$

We assign an energy function to the labelling map, here is a **General form of energy function**:

$$E(f) = E_{data}(f) + \lambda . E_{prior}(f) \tag{2}$$

Our description of the graph will be base on this formulation. First we will explain the construction of a full graph, and then we explain our contribution in reducing its size to accelerate the algorithm.

3.1. Cut in a Full Graph

Let us consider a linear potential map:

$$V_{\{p,q\}}(f_p, f_q) = u_{\{p,q\}} |f_p - f_q| \tag{3}$$

We will construct a graph to minimise the global energy:

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \lambda \sum_{\{p,q\} \in \mathcal{N}} u_{\{p,q\}} |f_p - f_q| \tag{4}$$

where $\mathcal{N}_p$ is the set of neighbour pixels of the pixel p, $\mathcal{N}$ is the set of neighbour pairs $\{p, q\}$, and D_p is the cost of matching of the pixel p with the corresponding value of disparity.

Let us define a graph $G = (V, E)$, in which V has two particular sites: a source s and a sink t. Let k be the number of possible matches (given by the disparity range). For each pixel p we assign a chain of nodes $p_1, p_2, ..., p_{k-1}$. These nodes are connected by edges called *t-link* and noted $t_1^p, t_2^p, ...,$ t_k^p where $t_1^p = [s, p_1]$, $t_j^p = [p_{j-1}, p_j]$ and $t_k^p = [p_{k-1}, t]$. For each *t-link*, we assign a capacity $K_p + D_p(l_j)$, where K_p is a constant satisfies the constraint (eq:5). For each pair of neighbour pixels p and q, the corresponding chains are related by edges called *n-link*, at levels $j \in \{1, 2, \ldots, k-1\}$: the *n-link* $\{p_j, q_j\}$ has a capacity $u_{\{p,q\}}$.

$$K_p > (k-1) \sum_{q \in \mathcal{N}_p} u_{\{p,q\}} \tag{5}$$

The capacity of an $s - t$ cut of the graph is the sum of capacities of all cut edges. Depending on the method of constructing the graph, the cut capacity is composed of two parts: the first is the sum of capacities of the cut *t-link* edges, and the second is the sum of capacities of cut *n-link* edges. In fact, the constant K_p allows us to assure the uniqueness of cut of each *t-link* chain, see [5] for a proof.

A graph cut consists in dividing the graph into two parts. The cut *t-link* edges form the surface of searched depth. The problem of graph cut can be solved using the maximum flow. Ford and Fulkerson [10] showed that the maximum flow from the source s to the sink t saturates a set of edges dividing the set of nodes in two parts S and T. The major problem of such method for robotic applications, is its huge execution time.

3.2. Cut of a Reduced Graph

To overcome the problem of execution time, we propose to construct a reduced graph: for each pixel we keep only some potential disparity values, resulting from a local method of stereo matching.

By mean of a local matching method (based on local similarity measurement, as SAD for example), we calculate for each left pixel p the costs of matching for all possible values in the disparity range $[d_{min}, d_{max}]$. Then, we choose the N best values (for illustration purposes, we choose $N = 4$ without lack of generality). The choice may be done according to different criteria, for example, with a classic ZNCC score, we keep the disparity values around the peak, or the N best local maxima (if they exist)... We will note our selected disparities for the pixel p as $d_{1,p}, d_{2,p}, \ldots, d_{N,p}$, and the costs of matching as $D_{\{p,d_{1,p}\}}, D_{\{p,d_{2,p}\}}, \ldots, D_{\{p,d_{N,p}\}}$. To reduce the size of the graph, for each chain (*t-link*), we remove all the nodes and edges except $N-1$ nodes (and N edges). Thus for each pixel p we construct a novel chain of nodes $\{p_{L_1}, p_{L_2}, \ldots, p_{L_{N-1}}\}$. These nodes are connected by edges (*t-link*) noted $\{t_1^p, t_2^p, \ldots, t_N^p\}$ with $t_1^p = [s, p_{L_1}]$, $t_2^p = [p_{L_1}, p_{L_2}]$, ..., $t_N^p = [p_{L_{N-1}}, t]$. The capacity of the *t-link* edge i is $C + D_{\{p,d_{i,p}\}}$, where C is a constant satisfing the constraint (eq:6).

$$C > N * \max_{\{p,q\}\in\mathcal{N}} (u_{\{p,q\}}) * |d_{max} - d_{min}| \qquad (6)$$

For each adjacent pixels p and q, the corresponding chains are related by edges (*n-link*) at the $(N-1)$ levels, with a capacity as in equation (7). The figure 1 illustrates a front projection of the graph.

$$\textit{Capacity of n-link at level } i = u_{\{p,q\}} * (|d_{i,p} - d_{i,q}| + 1) \qquad (7)$$

Minimised Energy in Reduced Graph: The global energy (eq:4) has two terms. The first term represents the intrinsic data energy, which translates the constraints of associating labels to the data. The second term aggregates the extrinsic energies (*prior energy*) which translate the constraints defined by the prior information. The constant λ can control the relative importance of the two terms. Hence, The prior energy appears in the weights associated to the *n-link* edges in the graph. In the reduced graph, we do distinguish between two types of prior information, the first translates the information acquired by the local method and acts in the choice of nodes, while the second (smoothing) interferes in the penalties associated to the *n-link* edges. Thereby, we exploit the prior knowledge that the disparity of a pixel p has only N possible values (the most probable) in a new way. In fact, we consider that removing non potential nodes as a novel form of representing this prior knowledge.

4. IMPLEMENTATION, EXPERIMENTAL RESULTS

We find in the literature two approaches to solve the maximum flow [11]. The first is the algorithm of augmented path due to Ford and Fulkerson, and the second is *preflow-push*. We choose an implementation of the later called *push relabel*, proposed by Goldberg [12], included in *Boost Graph Library* [1]. Note that Boykov and Kolmogorov [13] have proposed a new algorithm, and they showed that on typical graphs for computer vision applications, their algorithm is 2 to 5 times more faster than the other algorithms.

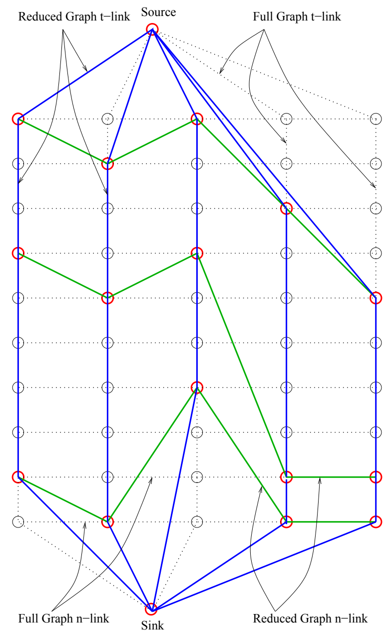


Fig. 1. Reduced Graph construction.

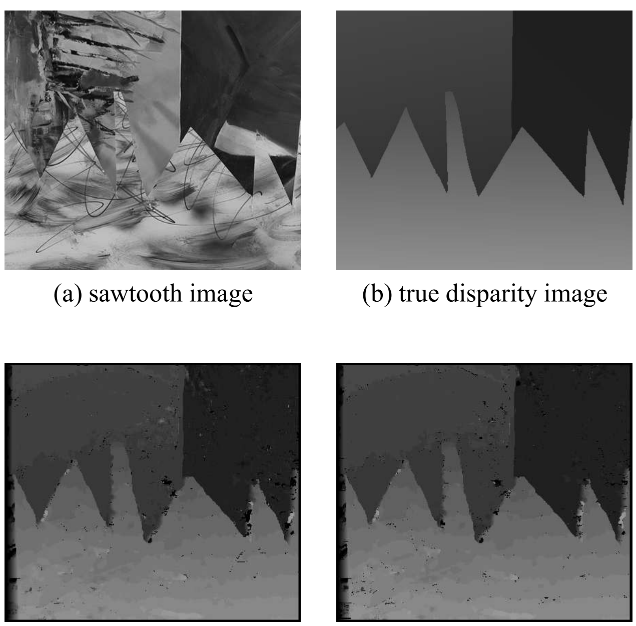
To illustrate the importance of our method of reduced graph, let W and H be the width and height of the image, $[0, d_{max}]$ be the disparity range, and N the number of best candidates given by the local matching method. Let v be the number of nodes and e be the number of edges. In the original graph, we have: $v = WH(d_{max} - 1) + 2$, $e \simeq 6v = 6WH(d_{max} - 1)$. The theoretical complexity of *push-relabel* is $\mathcal{O}(ve \log(v^2/e))$ [11]. In the case of full graph, the complexity is $\mathcal{O}(W^2H^2d_{max}^2 \log(WHd_{max}))$, while with the reduced graph, it becomes $\mathcal{O}(W^2H^2N^2 \log(WHN))$. We notice clearly that when $N \ll d_{max}$, our algorithm needs less memory and it is more faster. As an example, for $W = H = 512$, $d_{max} = 32$ and $N = 4$, in the full graph there are $v \simeq 8e6$ nodes and $e \simeq 50e6$ edges, whereas in the reduced graph has only $v \simeq 1e6$ nodes and $e \simeq 6e6$ edges. We notice that the full graph can not be manipulated on normal machines (with ordinary memory), whereas the reduced graph can be treated in acceptable time. So the reduced graph algorithm is more faster, in spite of a supplementary phase of calculation of local costs (which can be neglected compared with the execution time of graph cut algorithm).

We have evaluated the graph cuts, in the both cases: full graph and reduced one. Here are some results using the image sawtooth [2] (see figure 2-a). We used a local criterion based on SAD with a centred window of size 7. The test is done on

[1] http://www.boost.org/libs/graph/doc/index.html

a P4 with 3GHz and 512MB of RAM. For *sawtooth*, of size $434 * 380$ and 20 levels of disparities, the calculation time is about 15 seconds with $N = 4$, and 50 seconds for $N = 5$. With the full graph, we could not test the method with the full size image (memory explosion). Hence, we tested it with a half-sized image ($217 * 190$), the execution time is about 150 seconds, whereas with the reduced graph it is 4 and 5 seconds for N equal to 4 and 5 respectively.

In figure 2: (b) presents the true disparity image, (c) and (d) present the disparity images obtained respectively by the reduced graph and full graph. We can visually appreciate the quality: we have not yet done the qualitative test (as done in [2]). To be notice that for our robotics applications, we aim to use this algorithm in a mobile robot context: hence a reduced execution time will be very appreciated, even with a detriment of the quality of the disparity image.



(a) sawtooth image (b) true disparity image

(c) by reduced graph (d) by full graph

Fig. 2. experimental results.

5. CONCLUSION

We described in this article, our evaluation of graph cuts methods of stereo correspondence. The combination of a local method, able to select a reduced set of possible matches for each pixel, and a global method, based on the graph cuts algorithm, let us to achieve: (1) sensibly ameliorate the quality of disparity image obtained only by local method, and (2) to avoid the combinatorial explosion of the *Graph Cuts* method executed without preliminary reduction of the graph.

We will study some optimisations of our algorithm. Note that we work always on pre-rectified images, and we produce an integer disparity image: we will study how to ameliorate the precision (sub-pixel interpolation of disparity) and how to adapt this algorithm for not rectified images.

6. REFERENCES

[1] S. Chambon, *Mise en correspondance stéréoscopique d'images couleur en présence d'occultations*, Ph.D. thesis, Université Paul Sabatier, 2005.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.," *IJCV*, vol. 47, no. 1-3, pp. 7–42, 2002.

[3] P. Granjean and P. Lasserre, "Stereo vision improvments," in *IEEE ICAR*, Barcelona, 1995.

[4] O. Faugeras and T. Vieville et al., "Real-time correlation-based stereo : algorithm, implementations and applications," Tech. Rep. 2013, INRIA-Sophia Antipolis, 1993.

[5] S. Roy, "Stereo without epipolar lines: A maximum-flow formulation," *IJCV*, vol. 34, no. 2-3, pp. 147–161, 1999.

[6] H. Ishikawa and D. Geiger, "Occlusions, discontinuities, and epipolar lines in stereo," in *Proc. of the 5th ECCV*, 1998, pp. 232–248.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Markov random fields with efficient approximations," in *Proc. of the IEEE CVPR*, 1998, pp. 648–655.

[8] E. Dahlhaus and D. S. Johnson et al., "The complexity of multiterminal cuts," *SIAM Journal on Computing*, vol. 23, no. 4, pp. 864–894, 1994.

[9] O. Veksler, *Efficient graph-based energy minimization methods in computer vision*, Ph.D. thesis, Cornell University, 1999.

[10] L. Ford and D. Fulkerson, *Flows in Networks*, Princeton University Press, 1962.

[11] T. Cormen and R. Rivest et al., *Introduction to Algorithms, Second Edition*, The MIT Press, 2001.

[12] A. Goldberg and R. Tarjan, "A new approach to the maximum-flow problem," *Journal of the ACM*, vol. 35, no. 4, pp. 921–940, 1988.

[13] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," in *IEEE Trans. on PAMI*, 2004, vol. 26, pp. 1124–1137.