MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
5-4

Smooth Energy Auto-Estimation for Graph Cuts Algorithm

Haibing Ren, Jiali Zhao
CASIA-SAIT Joint Lab
95 Zhongguangcun East Road, Beijing, P.R. China
haibing.ren@samsung.com
emily.zhao@samsung.com

Seokcheol Kee
Samsung Advance Institute of Technology Korea
sckee@samsung.com

Abstract

Graph cuts is one of the best algorithms for disparity estimation. In its energy function, the smooth energy that is necessary for achieving the disparity map with good smoothness and good discontinuity varies from images to images. Based on image texture information, this paper presents a novel approach to estimate smooth energy automatically and precisely. To train the parameters of representation, an algorithm is proposed to auto-estimate the precise smooth energy coefficient, given the stereo images and the ground-truth disparity map. The experimental results show that, with the estimated smooth energy, the disparity map with less gross error could be achieved. And the disparity map obtained by hierarchical algorithm with estimated smooth energy for each level is much better than the one with same smooth energy for each level.

1 Introduction

In stereovision, more and more algorithms formulated disparity estimation as a problem of energy minimization. Among these algorithms, graph cuts algorithm [1,2,4,6] is one of the best. For disparity estimation, graph cuts algorithm considers the disparity of each pixel to be label and disparity estimation is to obtain the disparity map f that could minimize the following equation:

$$E(f) = E_{data}(f) + E_{smooth}(f) \quad (1)$$

where $E_{data}(f)$ is the energy that measures the disagreement between the disparity map and the stereo images, $E_{smooth}(f)$ is the energy that measures the smoothness of the disparity map.

Assuming each pixel to be a mass, the graph cuts algorithm can be considered to be a spring-mass system with 2 kinds of forces, as Fig1.The data energy is like the force F that pulls the pixel to the position with maximal data coherence between the left and right image. And the smooth energy is similar to the spring force that pulls the pixel to be near its neighboring pixels.

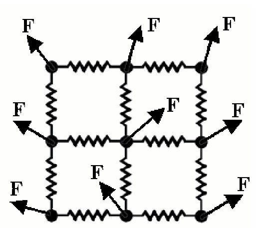


Figure 1. A spring-mass system.

The minimization of the energy function is transformed to the acquisition of the optimal α-expansion (or α-β-swap) that could minimize the energy function among all α-expansion (or α-β-swap). According to the current disparity map and α-expansion (or α-β-swap), a directed acyclic graph with source vertices and sink vertices is built. The min cut/max flow of the graph is equal to the minimum of the energy among all α-expansion (or α-β-swap).

Given stereo images and the disparity map, $E_{data}(f)$ can be easily calculated via the intensity difference between the corresponding pixels in left and right images. The calculation of smooth energy is the more nontrivial. In the traditional graph cuts algorithm, the smooth energy is formulated as follows.

$$E_{smooth}(f) = \sum_{\{p,q\}\in N} V_{p,q}(f_p, f_q).T(f(p) = f(q)) \quad (2)$$

$$T(a) = \begin{cases} 0 & a \text{ is true} \\ 1 & a \text{ is false} \end{cases} \quad (3)$$

Where p, q are neighboring pixels in the left image, $f(p)$ and $f(q)$ are the disparities of p, q, N is the set consisting the pairs of neighboring pixels, $V_{p,q}(f_p, f_q)$ is the penalty as equation(4).

$$V_{p,q}(f_p, f_q) = \begin{cases} \lambda & I_L(p) - I_L(q) > 5 \\ 2\lambda & I_L(p) - I_L(q) < 5 \end{cases} \quad (4)$$

Where λ is called the smooth energy coefficient in this paper and it is the key to the smooth energy. The following will be focused on auto-estimating the precise smooth energy coefficient. In graph cuts algorithm, λ is very important:

- As the $E_{data}(f)$ and $E_{smooth}(f)$ are two different metrics, λ should be able to adjust the scale relationship between the data energy function and smooth energy.
- Different stereo images need different λ to achieve the disparity map with good smoothness and good discontinuity. For different stereo images, the same λ may cause disparity maps over-smooth or under-smooth.

How to determine the coefficient for summing up two different metrics in energy function is a traditional hard problem. In most algorithms such as dynamic programming and relaxation algorithms for epipolar line matching [5], snake algorithm[7] for contour detection, the coefficients are determined by experience and experiments. In most graph cuts algorithm[1,2,6], λ is also determined by experience and experiments.

This paper presents a novel algorithm to estimate smooth energy coefficient automatically and precisely. First, this paper gives a parametric representation between image average texture and the optimal smooth energy coefficient. To train the parameters of representation, an algorithm is proposed to auto-estimate the precise smooth energy coefficient, given the stereo images and the ground-truth disparity map. Then, with the estimated smooth energy coefficient for each level, hierarchical graph cuts algorithm is built to decrease the computation time. The experimental results show that with the estimated smooth energy, the disparity map with good smoothness and good discontinuity could be achieved. And the disparity map obtained by hierarchical algorithm with estimated coefficient for each level is much better than the one with const coefficient for each level.

2. Smooth energy coefficient estimation

With different smooth energy coefficients, different disparity maps could be achieved. Some will be over-smooth and some will be under-smooth. Only via suitable smooth energy coefficient, could the disparity map with good smoothness and discontinuity be estimated. This paper gives an algorithm to estimate the coefficient automatically and precisely from the texture information.

2.1 The importance of the smooth energy coefficient

The smooth energy coefficient is very important for graph cuts algorithms for disparity estimation. With different coefficients, different results could be obtained. Even the same stereo images of different resolutions need different coefficients to achieve the disparity map with good smoothness and good discontinuity. There are two examples and the following is the first one:

(a)Left image 384*288

(b)Ground-truth disparity map

(c) Over-smooth disparity map

(d)Under-smooth disparity map

(e)Good Disparity map

Figure2. Disparity maps via different smooth energy coefficients.

Where Fig2(a) is the left one of the 'head and lamp' stereo images from the website of Middlebury College and (b) is the ground-truth disparity map. (c) is the over-smooth disparity map with the gross error 6.58%, which is caused by large smooth energy coefficient. (d) is under-smooth one with the gross error 7.13%, which is caused by small coefficient. The best disparity map (e) with the gross error 4.64% could only be obtained with the specific smooth energy coefficient. Here the gross error is the proportion between the error area and the all image area (not limited in the central area as [1,2]).

Even the same stereo images of different resolutions need the different smooth energy coefficient to achieve the disparity map with good smoothness and good discontinuity. The 'head and lamp' stereo images are also tested and the following figure is obtained.

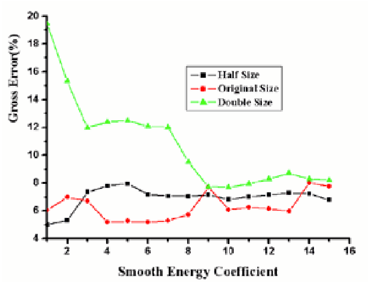


Figure3. Coefficients needed to achieve the minimal gross errors for different resolutions are different.

Where the horizontal axis is the smooth energy coefficients and the vertical axis is the gross error of the disparity map. The original size refers to the original 'head and desk' stereo images with the size 384*288. The half size refers to the half-size 'head and desk' stereo images with the size 192*144, and double size refers to the double-size 'head and desk' stereo image with the size 768*576. The half size and double size images are obtained by the commercial software Photoshop. In Fig2, the optimal coefficients for half size, original size and double size are completely different.

2.2 Smooth energy coefficient and image texture

For the image area with unique texture, disparity could be obtained via pixel matching. Unfortunately, most areas in the image have no unique texture. Therefore, smooth constraint is added to the global-optimization algorithm in order that the disparities of these areas are similar to its neighbor areas. So disparity map is affected by 2 factors: pixel correspondence and smoothness.

For the image with much texture, pixel correspondence plays a more important role than smoothness. For the image with less texture, the smoothness should be more important. Smooth energy coefficient is used to determine how important the smoothness is in the graph cuts algorithm. Obviously, smooth energy coefficient is related with the image texture.

In my algorithm, a texture term is defined as equation(5) to describe the quantity of the texture in the image.

$$T(x,y)=\sqrt{\frac{1}{x_k}\sum_{i=1}^{x_k}\frac{\left(I(x+i,y)-I(x-i,y)\right)^2}{2i}+\frac{1}{y_k}\sum_{i=1}^{y_k}\frac{\left(I(x,y+i)-I(x,y-i)\right)^2}{2i}} \quad (5)$$

Where x_k and y_k is the width and height of the window. With statistical algorithm, the relationship between the smooth energy coefficient and image average texture is formulated as equation(6).

$$E=\sum_{i=m}^{n}a_i t^i \quad (6)$$

Where $n>0>m$, E is the smooth energy coefficient for estimation, t is the image average texture.

2.3 Parameters estimation

There are two problems in estimating the parameters a_i ($i=m,..,n$): one is lack of training samples; the other is how to obtain the optimal smooth energy coefficient, given the stereo images and ground-truth disparity map.

In my experiments, there are only 7 base samples. Each base sample has a pair of stereo images with ground-truth disparity map. Via scaling and segmenting the base samples, more training samples with stereo images and ground-truth disparity map could be obtained.

With these training samples, the optimal smooth energy coefficient should be estimated. For simplicity, the coefficients can be considered to be integers and each possible value could be searched. But it costs a lot of computation time if the range is very large and the result is only a rough value rather than a precise one. In this paper, an iteration algorithm is utilized to obtain the precise coefficient.

$$E_{k+1}=E_k-\alpha\cdot M(D_k,G) \quad (7)$$

where E_k is the smooth energy coefficient in kth iteration, G is the ground-truth disparity map, D_k is the disparity map estimated by graph cuts with E_k in kth iteration. Of course, D_k is dependent on E_k. $M(D_k,g)$ is a function that measure the smoothness between D_k and G. If D_k is more smooth than G, $M(D_k,G)>0$,else $M(D_k,G)<0$. And α is a step coefficient. In my paper, $M(D_k,G)$ is implemented via simply measuring the edge information of the disparity map.

With the iteration algorithm, some training samples with stereo images and suitable smooth energy coefficient could be obtained. Some of the training samples are as following figure:

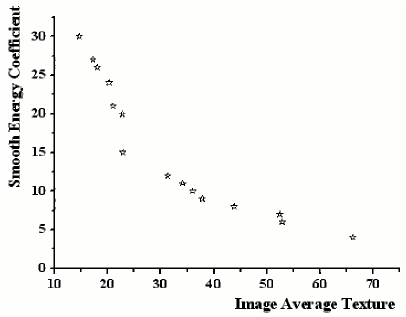


Figure 4. Some training samples.

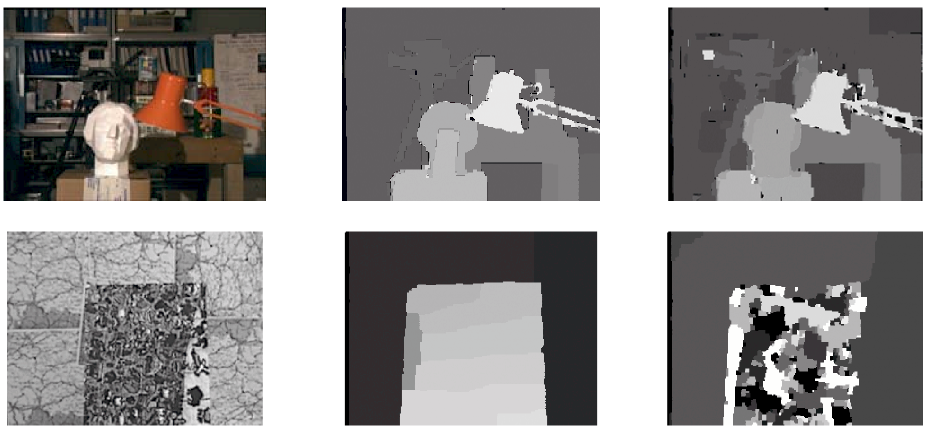
With the estimated smooth energy coefficient, the parameters a_i ($i=-m,..,n$) could be estimated via maximum likelihood estimate.

3 Hierarchical graph cuts algorithm

It is easy to understand that the hierarchical structure could decrease the computation time for disparity estimation. But different levels of hierarchical structure need different smooth energy coefficients, which is impossible for the traditional graph cuts algorithm. That's to say, smooth energy estimation makes it possible for hierarchical graph cuts algorithm to achieve less computation time and high precision.

4 Experimental results

In this section, some experiments are carried out to compare my algorithm and [4]. The followings are the images and disparity maps:



(a)Image (b) Disparity map via my algorithm (f)Disparity map via traditional algorithm [4]

Figure 5. The examples of disparity estimation results.

Where the size of first image is 768*576 and disparity range is [0,31]; the size of second image is 284*216 and disparity range is [0, 29]; the size of third image is 360*240 and disparity range is [0,7]. The average gross error of the disparity map via my algorithm is 5.12% while the gross error via traditional algorithm is 13.12%. The computation time of smooth energy coefficient estimation costs less 1% of traditional graph cut algorithm.

The hierarchical graph cuts algorithms with estimated and same smooth energy for each level are also compared. With the double size 'head and lamp' stereo images as an example, the gross error of the disparity map obtained by the hierarchical graph cuts algorithm with same energy for each level is as follows:

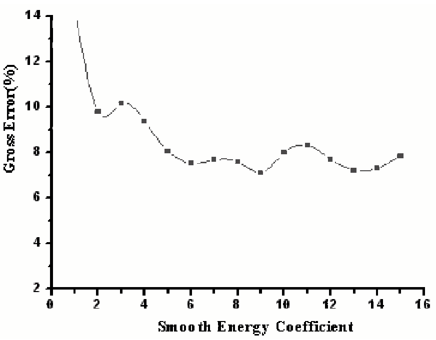


Figure6. The relation between the gross error and smooth energy coefficient

where the horizontal axis is the smooth energy coefficients and the vertical axis is the gross error of the disparity map. The minimal gross error among all possible smooth energy coefficients is 7.12%. With estimated smooth energy for each level, the gross error is 5.46%. The hierarchical graph cuts algorithm cost 12.03 seconds, merely 8.98% of the traditional graph cuts computation time.

5 Conclusion

In graph cuts algorithm for disparity estimation, the optimal smooth energy varies from images to images. In order to obtain the disparity map with good smoothness and good discontinuity, this paper presents a novel algorithm to estimate the smooth energy coefficients automatically and precisely based on the image average texture. The hierarchal graph cuts algorithm with the estimated smooth energy for each level is also presented, whose performance is much better than the one with const smooth energy. Besides, this algorithm can also be applied to other energy functions with different metrics.

Reference

[1] Yuri, Boykov, Olga Veksler and Ramin Zabih. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume: 23, Issue: 11, Nov. 2001, Pages:1222-1239.

[2] Kolmogorov Vladimir, Zabih, Zabih. Computing visual correspondence with occlusions using graph cuts. Eighth IEEE International Conference on Computer Vision. Volume: 2, 7-14 July 2001, Pages:508-515.

[3] Stan Birchfield, Carlo Tomasi. Multiway cut for stereo and motion with slanted surfaces. The Proceedings of the Seventh International Conference on Computer Vision. Kerkyra, Greece, 1999.Volume: 1, 20-27 Sept. 1999, Pages:489 - 495.

[4] Stan Birchfield, Carlo Tomasi. Depth Discontinuities by Pixel-to-Pixel Stereo. Proceedings of the 1998 IEEE International Conference on Computer Vision. Bombay, India, 4-7 Jan. 1998, Pages:1073 – 1080.

[5] D. H. Ballard and C. M. Brown. Computer Vision, Prentice-Hall Inc. Englewood Cliffs, New Jersey, 1982.

[6] Yuri, Boykov, Kolmogorov, Vladimir. Computing Geodesics and Minimal Surfaces via Graph Cuts. Ninth IEEE International Conference on Computer Vision, Oct. 13-16, 2003, Pages:26 – 33.

[7] M. Isard and A. Blake. Active contours. Springer-Verlag, 1998.