Confidence-Based Hierarchical Support Window for Fast Local Stereo Matching

Jae-Il Jung and Yo-Sung Ho

Abstract Various cost aggregation methods have been developed for finding correspondences between stereo pairs, but their high complexity is still a problem for practical use. In this paper, we propose a confidence-based hierarchical structure to reduce the complexity of the cost aggregation algorithms. Aggregating matching costs for each pixel with the smallest support window, we estimate confidence levels. The confidence values are used to decide which pixel needs additional cost aggregations. For the pixels of small confidence, we iteratively supplement their matching costs by using larger support windows. Our experiments show that our approach reduces computational time and improves the quality of output disparity images.

Keywords Cost aggregation • Hierarchical structure • Confidence • Low complexity • Stereo matching

1 Introduction

It is an old, but steady research topic to find dense correspondences from two or more images captured at different positions. It is called dense stereo matching, and its output is a disparity image, a set of the displacement vectors between correspondences. The dense disparity images have a number of applications such as robot vision, image-based rendering, and surveillance system, and their applicability is drastically increasing.

Although the principle of stereo matching seems to be straightforward, the ambiguity of images in practice makes stereo matching difficult. The ambiguity comes from homogeneous regions and periodic textures, and it is a main problem of

J.-I. Jung • Y.-S. Ho (✉)
Gwangju Institute of Science and Technology (GIST), 123 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, South Korea
e-mail: jijung@gist.ac.kr; hoyo@gist.ac.kr

J.S. Jin et al., *The Era of Interactive Media*,
DOI 10.1007/978-1-4614-3501-3_29, © Springer Science+Business Media, LLC 2013

Fig. 1 General flow of stereo matching

Step 1. Initial cost computation

Step 2. Cost aggregation

Step 3. Disparity optimization

Step 4. Disparity refinement (option)

stereo matching. Various stereo matching algorithms have been proposed to overcome this problem, and they generally have four steps [1] as shown in Fig. 1. Initial cost computation calculates matching costs for assigning different disparity hypotheses to different pixels, and cost aggregation spatially aggregate initial costs over support regions. Disparity optimization minimizes a predefined energy function locally or globally, and disparity refinement refines output disparity images.

Step 1 to Step 3 are mandatory, and Step 4 is optional. In general, stereo matching algorithms are categorized to two approaches: local and global algorithms. While the local methods use a local optimization method, the winner-takes-all (WTA), in Step 3, the global algorithms adopt global optimization methods such as graph-cut, belief propagation, and dynamic programming [2,3]. The global methods use additional prior knowledge such as smoothness and occlusion constraints. The global algorithms tend to show better results than the local algorithms, and they hold high ranks in the evaluating chart of Middle bury.

In both the local and global algorithms, the core part currently be watched is cost aggregation of Step 2. The performance of the local algorithms completely depends on cost aggregation, and it also plays an important role in the global algorithms. Therefore various cost aggregation algorithms have been proposed [9,10].

2 Cost Aggregation

Different cost aggregation algorithms use their own methods to select the support region and function for calculating new costs. Each method helps costs be robust to ambiguous regions in stereo images.

The easiest and oldest aggregation approach is to aggregate the cost of assigning disparity d to given pixel p with the average cost of all pixels in a square window, but it has the critical disadvantage that it is not able to handle disparity discontinuity boundaries. To overcome this property, the shiftable-window approach was proposed, which calculate matching cost with multiple square windows centered at different locations and selects the one having the smallest cost [5].

Other approaches use color segmentation based on the property that the depth discontinuity boundaries tend to co-locate the color discontinuity boundaries. The segmentation-based approaches select the sizes and shapes for support windows according to the segmentation results [6]. However it requires color segmentation

as a prior, which is an ill-posed problem and has high computational complexity. To remove these constraints, the edge detection-based approach was proposed [7].

Instead of searching for an optimal support window with arbitrary shape and size, the recently proposed adaptive-weight algorithm adjusts the support-weight of each pixel in a fix-sized square window [8]. However, the computation cost is high.

The recent cost aggregation algorithms show sufficiently good results without any global optimization, but its computational complexity is quite high. It is a crucial problem in practical use. Some algorithms are proposed for fast cost aggregation. Rhemann et al. smooth the cost volume with a weighted box filter, and the weights are based on the edge preservation [9,10]. Some researchers proposed its solution using graphics hardware [11], but it is not a root solution.

In this paper, we propose a confidence-based hierarchical structure to reduce a complexity of the cost aggregation algorithms. At first, we calculate data costs for each pixel with the smallest support window, and conduct confidence estimation. The confidence values are used to decide which pixels need additional cost aggregations.

3 Proposed Algorithm

The proposed algorithm uses the hierarchical structure and the confidence concept to reduce the computational complexity.

3.1 Hierarchical Structure

We can consider two types of hierarchical structure: pyramids of support window and image plane. Figure 2 shows these two methods. Type I and type II represent the pyramid of support window and pyramid of image plane, respectively.

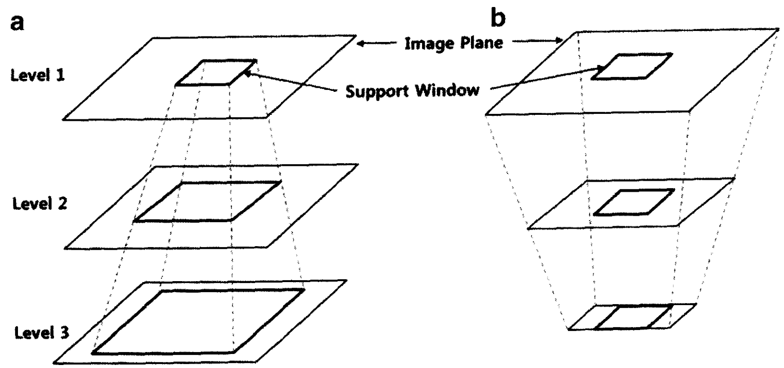


Fig. 2 Two types of hierarchical structure: (**a**) pyramid of support windows and (**b**) pyramid of image planes

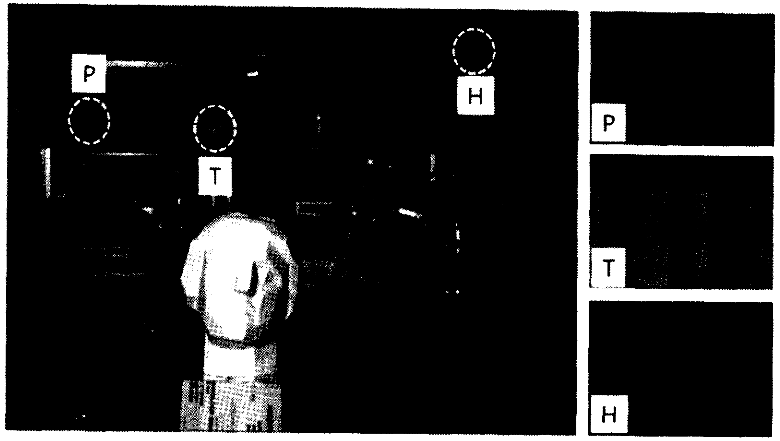


Fig. 3 Three texture types in *tsukuba* image: periodic (P), textured (T), and Homogeneous (H). The three images on the *right column* show the enlarged textures of the three parts

They have different advantages and disadvantages. While type II loses texture information and decrease the total disparity range in high levels, type I can keep them. However, type I induces higher computational complexity than type II.

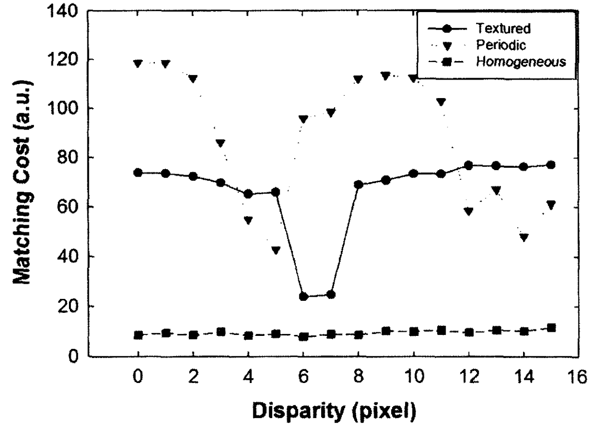
The problems of type II are very critical, because the texture information plays a core role in stereo matching and the reduced disparity range is very ambiguous to be directly used in lower levels. Therefore, we adopt type I in our algorithm and introduce the confidence-based structure to reduce computational complexity.

3.2 *Confidence Measurement*

The confidence in stereo matching means the probability that a certain pixel has a right disparity value. This concept is widely used to enhance the accuracy of stereo matching. However it is theoretically impossible to exactly distinguish between right and wrong disparities without ground-truth disparity images. We approximate the confidence value from given information such as matching costs and disparity pairs of stereo image. The algorithm estimating the confidence values with disparity pair is called cross checking process. Although it is very reasonable and shows reliable results, it takes doubled time since we should generate two disparity images for left and right images. Therefore we exclude cross checking in this paper, and use matching costs.

A captured image contains various textures, and their shape and density are unpredictable. In general, it is known that high frequency textures help accurate stereo matching, but this knowledge is invalid when the textures are periodic. Figure 3 shows *tsukuba* image from middle bury, which has various textures. We marked three parts on the image, which have special textures. The three images

Fig. 4 Matching costs of the three texture types: textured, periodic, and homogeneous



on the right column show the enlarged textures of each part. The part P and part T have high frequency textures, but the texture of the part P is periodic. The "P" and "T" are initial letters of "periodic" and "textured", respectively. The part H has low frequency textures, and the "H" is an initial letter of "Homogeneous". It is known that homogeneous regions and periodic-textured regions induce inaccurate stereo matching.

We extract each matching costs according to disparity candidates so as to analysis the relationship between texture types and matching costs. Figure 4 shows the matching cost, calculated by the adaptive support weight, according to disparities. The textured region has an exact minimum, but the minimums of others are ambiguous. The homogeneous region shows low matching costs on the whole, but notable minimum does not exist. Although there are notable local minimums in the periodic textures, it is not clear which minimum is correct.

The WTA algorithm, of course, can find the disparity having the smallest matching cost, but we cannot conclude this disparity is correct because these are various unexpected textures disturbing cost aggregation.

In our proposed algorithm, we use the confidence value to distinguish the pixels with reliable matching costs in whole pixels and stop aggregating costs in next level of the pyramid of support window. It helps to reduce the computational time. Therefore, it is very important to extract reliable pixel with an accurate measure. Some confidence measurements were proposed such as the first minimum cost value, and the ratio of the first minimum cost to the second minimum cost. Since our main goal is to extract very accurate pixels, we define the confidence measure as (1).

$$Confidence(i) = 255 \cdot \exp \lambda (1^{st}C(i) - 2^{nd}LC(i)) \tag{1}$$

where $1^{st}C(i)$ is the smallest matching cost of pixel i, and $2^{nd}LC(i)$ is the second local minimum matching cost of pixel i. λ and constant value 255 are scaling and normalizing factors. Figure 5 shows the original *tsukuba* image, its disparity image,

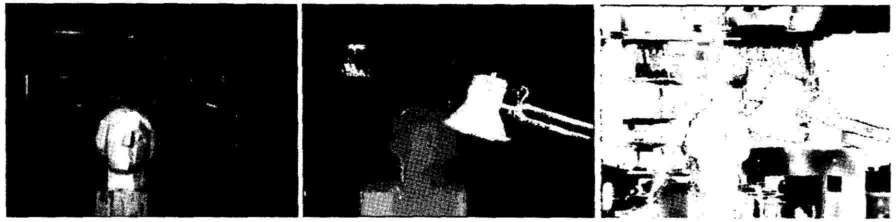


Fig. 5 Original *tsukuba* image, its corresponding disparity image, and confidence image

and confidence image. The brightness of the confidence image represents the reliability of the corresponding disparity value. Our confidence measure is able to distinguish not only homogeneous regions but also periodic regions.

3.3 *Confidence-Based Hierarchical Cost Aggregation*

With the confidence image and hierarchical support window, we can reduce the computational complexity of cost aggregation. The following code is the pseudo code for our algorithm.

```
For each Support_Window_Size, from small size,
  For each pixel i,
    if(Flg(i) == FALSE)
      Calculate Cost_new(i) using Cost_Aggregation
      Calculate Confidence(i) from Data(i)
  Refine Confidence(i)
  For each pixel i,
    if(Confidence(i) > Th)
      set Flg(i) = TRUE
    if else
      Cost_old(i) + = Cost_new(i)
Disparity with WTA
```

At first, we select the number of level and initiate the matching cost. We aggregate the matching cost with the smallest support window, and calculate the confidence values of whole pixels. The confidence image is refined by using Gaussian smooth filter to consider neighbor confidence values. If the confidence value is greater than a certain threshold value, we set the flag for the current pixel with "TRUE". This flag is used to decide whether the current pixel needs additional aggregation process or not. In the next level, we only aggregate costs for pixel whose flag is "FALSE". These processes are repeated until cost aggregation process with the largest support window is done. The final disparity image is obtained by using WTA.

Figure 6 shows the disparity and flag images according to each level. As the level increases, the "TRUE" flag, which means the current pixel does not need

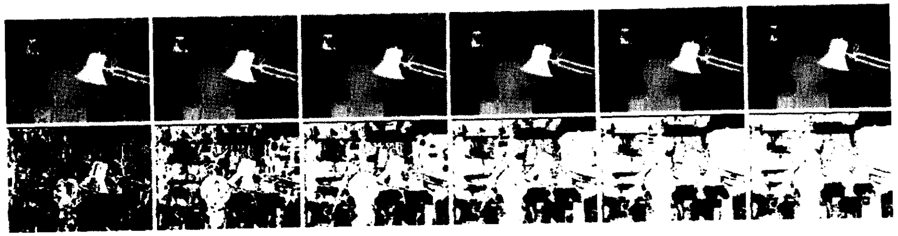


Fig. 6 Change of disparity and flag images according to increased hierarchical level

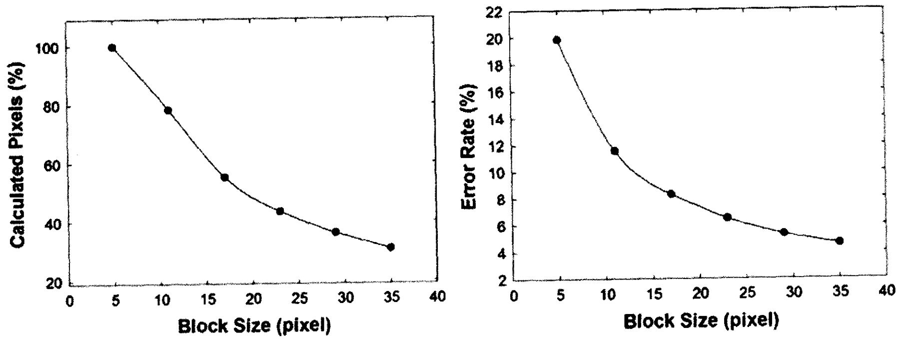


Fig. 7 Calculated pixels and error rates according to support window size

additional aggregation, also increase. The quality of the final disparity image is also improved.

We measure the number of calculated pixel and error rates at each level, and demonstrate them in Fig. 7. As the size of support window increases, the number of calculated pixels decreases. It means we can skip many pixels in the higher level which needs more time than lower level. The accuracy of stereo matching is also improved as the level increases.

Figure 8 shows the relationship between the number of levels and computational complexity. While a number of levels induces time-saving, the error rates are also incresed. In this paper, we set the number of levels with three to five according to the maximum support window size.

4 Experimental Results

In order to evaluate the performance of our proposed algorithm, we used the Middlebury stereo benchmark [12]. In our test run, the algorithm's parameters are set to constant values. The window size is chosen to be 19, 35, and 71. We did not handle occlusion regions.

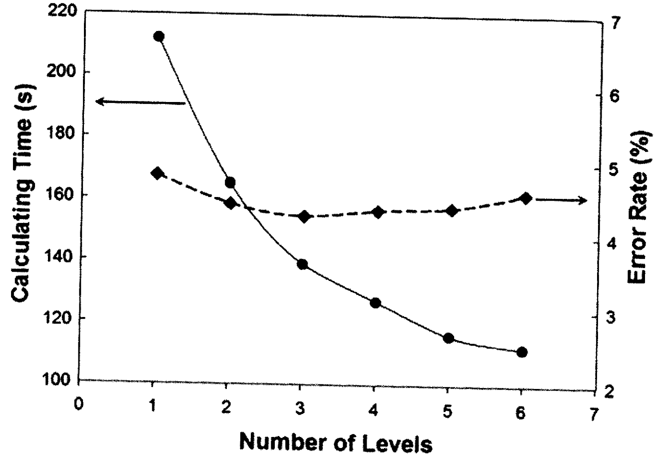


Fig. 8 Calculating time and error rates according to the number of levels

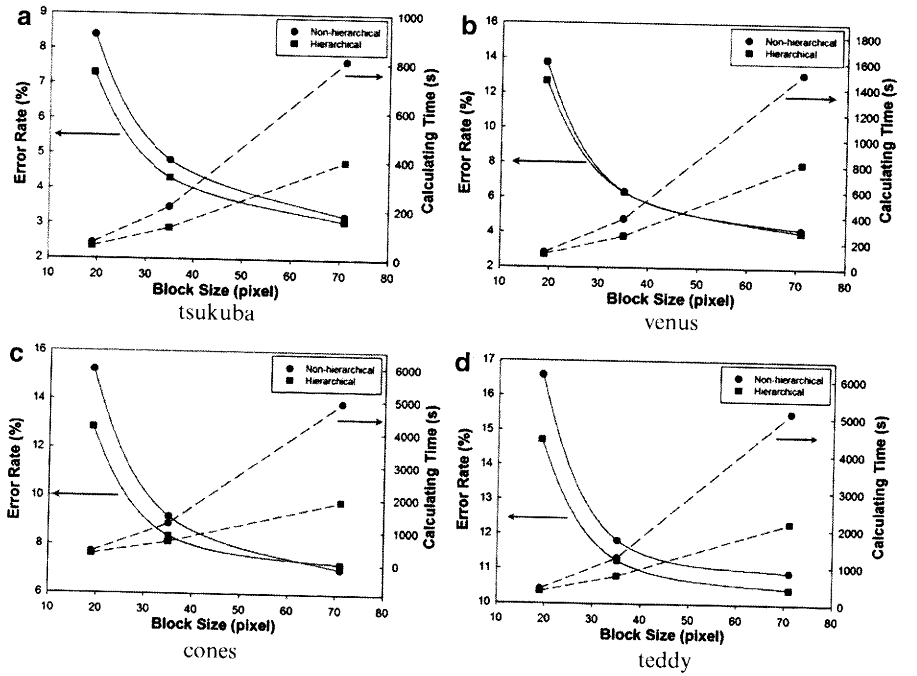


Fig. 9 Comparison between non-hierarchical and hierarchical cost aggregation

Figure 9 shows the comparison between non-hierarchical and hierarchical cost aggregation. The calculating times of the proposed algorithm are shorter than those of the non-hierarchical approaches. Especially the gaps increase when the maximum size of support window is large. In most test images, the error rates of non-occluded regions of the proposed method are smaller than those of the conventional approaches. It is a basic assumption that the pixels in the support window have similar disparity value. Sometimes large support window opposes this assumption. However our algorithm prevents this problem.

Table 1 Objective evaluation for the proposed algorithm with Middlebury test bed

	Non-hierarchical				Hierarchical			
Test image	Non-occ	All	Disc	Time (s)	Non-occ	All	Disc	Time (s)
Tsukuba	4.81	6.78	8.82	212	4.35	6.22	8.41	127
Venus	6.32	7.47	11.69	377	6.31	7.54	10.31	245
Cones	9.17	17.81	17.53	1,230	8.36	17.88	16.81	677
Teddy	11.81	18.31	22.27	1,226	11.24	19.22	21.52	739

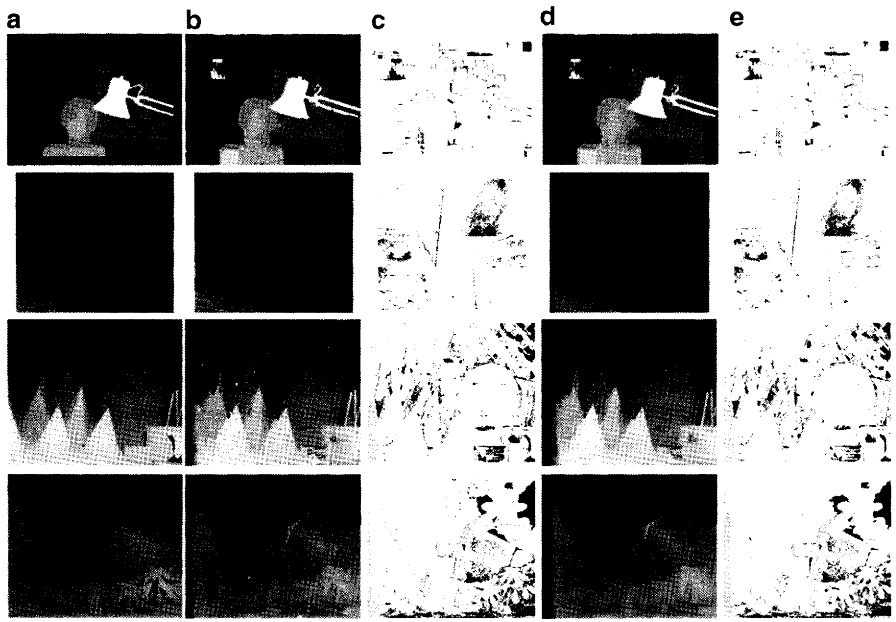


Fig. 10 Final disparity and error images obtained by non-hierarchical and hierarchical cost aggregation: (**a**) ground-truth, (**b**) and (**c**) disparity and error images of non-hierarchical cost aggregation, (**d**) and (**e**) disparity and error images of hierarchical cost aggregation

Table 1 shows the detail bad pixel ratios and operation times of both approaches when the support size is 34. Our algorithm reduces the computational complexity and improves accuracy in non-occluded regions and near discontinuities. Figure 10 demonstrates our results along with corresponding error images. One can see that our algorithm performs well in the reconstruction of disparity borders, while it also finds correct disparities for regions of low texture.

5 Conclusions

The recently proposed cost aggregation algorithms are very effective, but they have high computational complexity. In this paper, we proposed the confidence-based hierarchical structure to reduce the complexity of cost aggregation. The proposed method adopts the pyramid of support window, and confidence values based on matching costs. It aggregates the matching cost with the small support, the size of support increases according to confidence values. A variety of experiments show that our proposed algorithm can reduce the computational complexity of cost aggregation and slightly improve the accuracy of disparity images. In addition, our algorithm can be applied to any cost aggregation algorithm and its performance can be improved by further adjusting the parameters such as pyramid structures and confidence measure.

Acknowledgement This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011–0030822).

References

1. Scharstein, D. and Szeliski, R.: A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms. Int. J. Comput. Vis., 47(1–3):7–42 (2002)
2. Boykov, Y., Veksler, O., and Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts. IEEE TPAMI, 23(11):1222–1239 (2001)
3. Ohta, Y. and Kanade, T.: Stereo by Intra- and Interscanline Search Using Dynamic Programming. IEEE TPAMI, 7(2):139–154 (1985)
4. Forstmann, S., Ohya, J., Kanou, Y., Schmitt, A., and Thuering S.: Real-time stereo by using dynamic programming.. In Proc. CVPR Workshop on Real-time 3D Sensors and Their Use, pp. 29–36. Washington, DC (2004)
5. Veksler, O.: Fast variable window for stereo correspondence using integral images. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 556–561 (2003)
6. Gong, M. and Yang, R.: Image-gradient-guided Real-time Stereo on Graphics Hardware. In Proc. International Conference on 3-D Digital Imaging and Modeling, pp. 548-555. Ottawa (2005)
7. Wang, L., Kang, S. B., Shum H.-Y., and Xu, G.: Cooperative Segmentation and Stereo Using Perspective Space Search. In Proc. Asian Conference on Computer Vision, pp. 366-371. Jeju Island (2004)
8. Yoon, K.-J. and Kweon, I.-S.: Locally Adaptive Support-weight Approach for Visual Correspondence Search. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 924-931 (2005)
9. Rhemann1 C., Hosni1 A., Bleyer M., Rother C., Gelautz M.: Fast Cost-Volume Filtering for Visual Correspondence and Beyond. in Proc. The IEEE Conference on Computer Vision and Pattern Recognition, pp. 3017-3024 (2011)
10. Gong, M., Yang, R., Wang, L., and Gong, M.: A Performance Study on Different Cost Aggregation Approaches Used in Real-Time Stereo Matching. Int. J. Comput. Vis. 75(2), 283–296 (2007)

11. Yang, R., Pollefeys, M., and Li, S.: Improved Real-time Stereo on Commodity Graphics hardware. In Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshop on Realtime 3D Sensors and Their Use, Washington, DC (2004)
12. Middlebury, http://vision.middlebury.edu/stereo