Disparity from stereo-segment silhouettes of weakly-textured images

Babette Dellen[1]
bkdellen@bccn-goettingen.de
Florentin Wörgötter
worgott@bccn-goettingen.de

Bernstein Center for Computational Neuroscience (BCCN) Göttingen, Max-Planck-Institute for Dynamics and Self-Organization
BCCN Göttingen, University Göttingen

In stereo vision, 3D information is reconstructed from two images of the same scene taken from different viewpoints. Different approaches to stereo disparity estimation have been extensively compared in several studies [1, 3]. Most stereo algorithms perform well in textured image areas, but often fail when there is only weak texture, due to the correspondence problem. Here local matching fails, and, as a consequence, global methods do not deliver correct disparities either, simply because the energy functions used in global methods remain under-constrained. However, stereo from weakly textured images is important for many applications, which take place in urban or industrial settings, where little texture exists, requiring novel solutions to the stereo problem.

While being ill-suited for stereo analysis, weakly-textured image parts can easily be used for color-based segmentation and, in addition, it is often also possible to find unique *segment correspondences* between images. The goal of the present study is to recover disparity in weakly-textured image parts by using correspondent image segmentation together with an interpolation algorithm based on a spring-mass model. To this end, we constrain interpolation (the springs and the masses) by the disparities of the segment-edges as well as by the vague, remaining stereo information inside a segment, which we can still recover using conventional stereo algorithms. This way we can regenerate rather accurate disparity information in regions that are usually quite resistant to stereo analysis, such as some images from the 2006 Middlebury stereo dataset.

In the proposed method, the stereo image is first decomposed into corresponding regions, i.e. stereo segments. Such a decomposition can be seen in Fig. 1(a-d) for a stereo image showing a cluttered scene. Stereo segmentation is achieved via a conjoint spin relaxation process based on the superparamagnetic clustering of data [4], where links between pixels in different frames are established via a precomputed disparity map obtained with a standard algorithm (see Fig. 1(e)).

For each stereo segment, segment silhouettes are computed in both frames. Unique correspondences of silhouette-edge points are searched, and the respective disparities are calculated. While many segment edges represent real object boundaries, some edges are in fact caused by an occlusion and thus do not represent a “true” edge. These occluded edges need to be identified to avoid erroneous interpolation results. Occluded areas are estimated by finding the approximate depth ordering of the scene. The segment-center disparity for each stereo segment is computed from the distance between the left and right segment center. The stereo segments are ordered according to their segment-center disparity. An edge of a segment that is adjacent to an edge of another segment with larger disparity (thus, it is closer) is then assumed to be caused by an occlusion, and an occlusion map is obtained (see Fig. 1(f)) and an occlusion-free segment-edge disparity map is derived (see Fig. 1(g)).

Additionally, texture inside segments can be exploited by applying a window-based stereo algorithm which operates strictly inside stereo segments, thus occlusion problems do not arise. Confidence values are computed and only those disparities are used that have a high confidence, resulting usually in sparse disparity maps (see Fig. 1(h)).

The sparse disparity information from both the segment-silhouette edges and inner-segment texture are used to interpolate disparity values inside segment area (see Fig. 1(i-k)). For this purpose, we used an interpolation algorithm based on a spring-mass model. Each pixel of the image is described by a mass point, connected via elastic springs to its nearest neighbor. However, if the neigbor belongs to a different segment, the connection is cut. The amplitude of the mass point represents its disparity value. The sparse disparity maps are used to define data forces which push the respective masses towards the precomputed value. A damping force is included to drive the system to a local minimum. The corresponding system of equations is numerically solved.

[1]BD is also with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, Spain. This work has received support from the BMBF funded BCCN Göttingen and the EU Project PACO-PLUS.

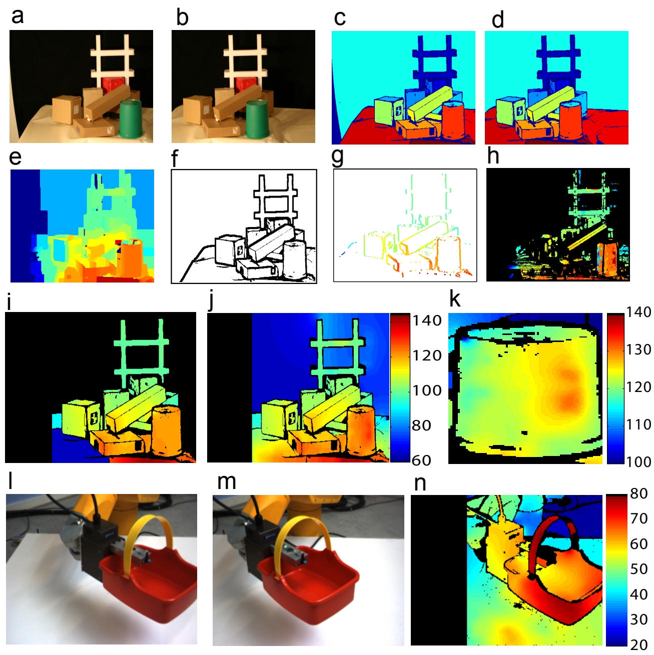

Figure 1: Results for weakly-textured stereo pairs. (a) Left image of *Cluttered scene*. (b) Right image. (c) Segmentation of left image. (d) Segmentation of right image. (e) Initial disparity map used for segment linking. (f) Estimated occluded areas. (g) Segment-edge disparities. (h) Disparities in inner-segment areas. (i) Interpolation result using edge disparities only. (j) Interpolation result using both edge and inner-segment disparities. (k) Enlarged result of trash can. (l) Left image of *Basket*. (m) Right image. (n) Estimated disparities.

Despite the elaborated structure of the algorithm, the results for various stereo-image pairs could be obtained with the same parameter set. For the Middlebury stereo images, disparity maps with an average error between 0.81 and 3.97 pixels were obtained. Even for very weakly textured scenes such as *basket*, the algorithm succeeded in capturing the basic 3D structure of the scene. We also demonstrated that the algorithm is applicable to stereo images characterized by large disparities, which often pose a problem to other stereo algorithms (see Fig. 1(l-n)).

Color segments have been used before to improve disparity estimation [2, 5]. By contrast, our method is based primarily on the computation of disparities from stereo-segment silhouettes, which, to our knowledge, have not been utilized before. Texture cues are used then as an additional cue to improve results, but are not mandatory.

[1] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.

[2] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Intl. Conf. Pattern Recognition*, pages 15–18, 2006.

[3] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.

[4] C. von Ferber and F. Wörgötter. Cluster update algorithm and recognition. *Physical Review E*, 62:1461–1664, 2000.

[5] Q. Yang, C. Engels, and A. Akbarzadeh. Near real-time stereo for weakly-textured scenes. In *British Machine Vision Conference*, 2008.