Depth map estimation: a region-based approach

Jean-Charles Bricola, Michel Bilodeau, Serge Beucher

DEPTH MAP ESTIMATION: A REGION-BASED APPROACH

Jean-Charles Bricola, Michel Bilodeau, Serge Beucher

Mines ParisTech
Centre de Morphologie Mathématique
35 rue Saint-Honoré, 77300 Fontainebleau, France

Abstract

This paper presents an approach to the estimation of depth from stereo images which exploits correspondences of image segments. A key motivation is to bypass pixel correspondences which are ambiguous in texture-less regions. One of the main difficulty though is to come up to an equivalent partitioning of both stereo images that facilitates region matching. The reference image which the depth is estimated for is first segmented using the watershed algorithm. Regions are then transferred to the other image of the stereo pair according to a correlation analysis of region contours. This transfer yields regional disparities which are significant as long as the contours for a given region are not due to an occlusion. Based on that observation, we propose an algorithm that detects occlusion contours along a region and that rectifies disparity accordingly.

1. Introduction

This article focuses on stereoscopic configurations that involve two images of a scene taken at different viewpoints. A depth map is computed for one image chosen from the stereo pair, namely the reference image. We refer to the other image as the second image. For a pixel in the reference image, there exists an epipolar line in the second image along which a correspondence may be found, provided the transformation between the coordinate systems of the stereo cameras is composed of a non null translation and no occlusion prevents the scene point from being projected in both stereo images [1, 2]. In most scenarios, rectified images will be considered, for which a pixel and its correspondence always share the same ordinate. The disparity is the measured offset between the pixels abscissa and is inversely proportional to the depth we aim at estimating.

The problem of depth map computation from stereo dates back to the 70s and has until now relied mostly on pixel correspondences between stereo images. Early approaches considered feature points and contour matching in that respect [3, 4]. Whilst the established correspondences were reliable, the measurements suffered from sparsity. Current methods on the contrary attempt to find a match for every pixel of the reference image. The correspondence is typically selected according to the smallest dissimilarity cost measured between image patches centred at the considered pixels. An exhaustive list of such measures is given in [5]. Because of matching uncertainties in homogeneous regions, the resulting disparity maps usually require a post-processing step that consists of minimizing an energy function. The latter usually leads to smooth transitions of disparities among neighbouring pixels except on edges where discontinuities are permitted [6] whilst ensuring that the disparities still agree with the reliable initial measurements.

The use of regions within depth estimation algorithms becomes increasingly popular and differentiates into super-pixels and layer-based categories. For instance, in [7], the reference image is over-segmented and disparities are estimated at the segment level whilst smoothness is enforced across neighbouring segments having close colour distribution. In [8] however, disparities are computed across image layers for which planar equations are estimated. Layers are obtained by means of colour segmentation and regions belonging to the same disparity plane according to an initial disparity map are merged. Our work falls into the layer-based category.
We propose an algorithm that relies on the disparity information along region contours and inside subregions to estimate depth maps. The gradient and markers generation required for the watershed segmentation of the reference image is first detailed in section 2. Section 3 introduces the concept of regional disparities. We show that the latter can be used to produce markers suitable for segmenting the second image of the stereo pair and producing instantaneous region matchings. We also discuss their relevance with respect to stereopsis. Section 4 explains how to overcome the shortcomings of regional disparities computed for initial segments and presents an algorithm that correct most disparity errors due to occluding contours. Results are finally exposed and evaluated in section 5.

2. Reference Image Segmentation

The watershed algorithm [9] partitions an image into a series of connected components satisfying each an homogeneity criterion and can be implemented effectively using hierarchical priority queues [10]. It has moreover the ability to be controlled by markers. This remarkable property enables the transfer of markers from the reference to the second image of the stereo pair, hence producing equivalent stereo partitions together with their matchings as explained in section 3. The watershed-based segmentation constitutes, therefore, a very suitable choice with respect to our application.
In short, watersheds are obtained by flooding a topological surface, typically the colour gradient of the image, from a set of markers. Each marker gives birth to a lake which a unique label is assigned to. Each lake in turn will lead to an image region. As the flooding goes on and uniformly increases in altitude, watersheds are constructed as soon as two lakes with different labels are on the verge of merging. More precisely the geodesic skeleton by influence zone of the lakes just before merging is computed inside the new lake obtained just after merging.
The segmentation result depends on the choice of appropriate markers. As far as the reference image is concerned, markers are chosen from the h-minima of the gradient in order to capture the regions of interest.

2.1. Colour and thick gradients

Colour is taken into account within a watershed segmentation by computing a colour gradient. Our current implementation just considers the supremum of gradients for the red, green and blue channels. However, there exists more meaningful representations which are worth mentioning [11]: in particular, the linear combination of the hue and luminance gradients according to the saturation could be used as a colour gradient instead as well as perceptual differences in CIE-La*b* space.
Transitions between two objects which span several pixels are captured by computing a thick gradient. The gradient values have indeed a significant impact when extracting markers from the h-minima of the gradient and when a hierarchical transform is applied on the watersheds. The thick gradient of size λ is computed

Symbol	Description
Morphological operators	
B and H are structuring elements	
H is an isotropic structuring element of elementary size	
$\delta_B(f)$	Dilation of f: $\delta_B(f)[x] = \sup_{x \in B} f[x]$
$\varepsilon_B(f)$	Erosion of f: $\varepsilon_B(f)[x] = \inf_{x \in B} f[x]$
$D_g^1(f)$	Geodesic dilation of f under mask g: $D_g^1(f) = \inf(\delta_H(f), g)$
$\mathtt{R}_g(f)$	Geodesic reconstruction of g from marker f: $\mathtt{R}_g(f) = D_g^{+\infty}(f) = D_g^1\left(\ldots\left(D_g^1(f)\right)\ldots\right)$
$\mathtt{R}_g^*(f)$	Dual geodesic reconstruction of g from marker f: $\mathtt{R}_g^*(f) = -\mathtt{R}_{-g}(-f)$

Table 1: Notation

according to equation 1:

$$g_\lambda(f) = (\delta_{\lambda H}(f)) - (\varepsilon_{\lambda H}(f)) \tag{1}$$

Hence, if a transition between two objects requires λ pixels to complete, the thick gradient g_λ at the middle of the transition equals the difference between the smallest and the highest values of the transition. In comparison with the morphological gradient of size $\lambda = 1$, pixels on the edges are assigned a gradient magnitude similar to the one expected for a perfect transition between two regions whilst the magnitude remains small for irrelevant variations of intensities. Figure 1(b) shows an example of thick gradient of size $\lambda = 4$ computed for an image composed of 1920×1080 pixels.

2.2. Markers from h-minima

Markers control the behaviour of the watershed segmentation: ideally they should lie at the bottom of the topological surface being flooded and discriminate every region to be segmented as well. To this end, we use the h-minima of the gradient [12]. Let g be the gradient of the reference image and h a positive constant. The h-minima are computed according to equation 2.

$$\mathtt{min}(g,h)[x] = \begin{cases} 1 & \text{if } \left(\mathtt{R}_g^*(g+h) - g\right)[x] > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

In other words, x is a h-minima of g if and only if there exists on g a strictly descending path from x to a minima of g, say m, such that $g(x) - g(m) > h$. By definition, h-minima contain the local minima of the function as well. h is chosen to be sufficiently large so as to ignore small catchment basins that are due to noise in the gradient. This choice implies that h-minima occasionally overlap contours at low gradient values.
In order to solve that problem, we compute a distance function p by construction of successive erosions applied over the binary mask corresponding to the h-minima. Thin structures appearing on the mask will appear as valleys in function p. New markers are given by the connected components resulting from the indicator function $p - \mathtt{R}_p(\alpha p) > 0$ for $0 \leqslant \alpha < 1$. Small values of α encourage markers splitting at very narrow valleys only and reduce slightly the overall size of the markers, getting them away from the borders as shown in figure 1(c). Note that contrary to an erosion or an opening, that approach to process h-minima preserves small markers.

3. Region transfer and disparities

In this section, we explain how regional disparities are estimated for every region discovered in the reference image. We also discuss their relevance and usefulness with respect to depth estimation and equivalent stereo segmentation.

3.1. Regional disparities

Let $\mathcal{R} = \{R_1, \cdots, R_n\}$ be the set of partitions obtained for the reference image. $R_i(x, y)$ is an indicator value that equals 1 for every pixel (x, y) belonging to region R_i, 0 otherwise. Also let g_R and g_S be the colour gradients for the reference and second images respectively.

For each region R_i, the translation $\mathbf{t}^{(i)}$ that provides the best overlap between the gradients of the reference image and the shifted second image within R_i is estimated. As stereo images are assumed to be rectified, the translation only occurs along the x-axis, so $\mathbf{t}^{(i)} = (d^{(i)}, 0)$. Optimal translations are determined by equation 3.

$$d^{(i)} = \arg\min_{d} \sum_{(x,y)|R_i(x,y)=1} |g_R(x, y) - g_S(x + d, y)| \tag{3}$$

From equation 3, we finally obtain for each region R_i its regional disparity $|d^{(i)}|$.

The sum of absolute differences (SAD) is considered here as a cross-similarity measure and the gradients as inputs for that measure. It is of course possible to take the luminance or a hue channel in place of the gradient. Doing so is preferable when source images have no illumination changes and do not suffer from noise as this will increase the support, hence the accuracy of the correlation.

Whilst regional disparities are relevant to fronto-parallel regions, they only provide an average estimate of the disparities among other regions. Furthermore regional disparities are influenced by contour disparities themselves. This is a shortcoming with respect to regions for which contours are due to occlusions as shown in figures 2(b) and 2(f). These observations justify the disparity refinement step exposed in section 4.

3.2. Equivalent stereo segmentation

Estimating regional disparities and generating markers for segmenting the second image of the stereo pair are interestingly two equivalent problems. A simple method for obtaining an equivalent segmentation of the second image alongside a matching between regions is described below.

Let M_i be an indicator function representing the pixels of the second image which are spanned by the marker corresponding to region R_i in the first image. Then:

$$R'_i(x, y) = R_i\left(x - d^{(i)}, y\right) \tag{4}$$

$$M_i(x, y) = R'_i(x, y)\left(1 - \sup_{j \neq i} R'_j(x, y)\right) \tag{5}$$

Each marker is resized using a distance function as described in section 2 with the exception of markers labelling which remains unchanged, even if splitting occurs. Finally, the watershed segmentation driven by these markers and the the second image colour gradient is performed. Regions inherit from the marker labels which yields region matchings between the two images. These matchings potentially provide a mean to detect re-

gions occluded in the second image as well as a correlation measure based on the intersection volume between matched regions.

4. Disparity refinement

The refinement algorithm proposed below aims at detecting occlusion contours and rectifying mistaken regional disparities accordingly. It first assigns to every region two disparities: one that is representative of the disparity near the left contour of the region, the other of the right contour. To do so, regions obtained for the reference image are vertically split into left and right subregions and regional disparities for these subregions are estimated. Second, neighbourhood relationships between regions are defined: $\mathcal{N}_l(R_i)$ includes all regions $R_j \in \mathcal{R}$ that are adjacent to the left subregion of R_i and the same goes for $\mathcal{N}_r(R_i)$ with respect to the right subregion of R_i. Third, algorithm 1 is executed for every region R_i:

Algorithm 1 Regional disparity refinement

1: **function** REFINEDISPARITIES($\mathcal{R}$, i)
2: $d_l \leftarrow$ `getLeftContourDisparity`(R_i)
3: $d_r \leftarrow$ `getRightContourDisparity`(R_i)
4: **if** $d_l > d_r + \tau$ **then**
5: **for all** $R_j \in \mathcal{N}_l(R_i)$ **do**
6: $d_r^{(j)} \leftarrow$ `getRightContourDisparity`(R_j)
7: **if** $|d_l - d_r^{(j)}| < \tau$ **then**
8: $d_l \leftarrow d_r$
9: **end if**
10: **end for**
11: **else if** $d_r > d_l + \tau$ **then**
12: **for all** $R_j \in \mathcal{N}_r(R_i)$ **do**
13: $d_l^{(j)} \leftarrow$ `getLeftContourDisparity`(R_j)
14: **if** $|d_r - d_l^{(j)}| < \tau$ **then**
15: $d_r \leftarrow d_l$
16: **end if**
17: **end for**
18: **end if**
19: **return** (d_l, d_r)
20: **end function**

The intuition behind this algorithm is the following: when a region is assigned to very different left and right disparity values (i.e. when algorithm condition 4 or 11 is fulfilled for a given threshold τ), either the highest disparity value is due to an occluding contour or the region is simply not fronto-parallel. To verify whether we fall into the first category, we check whether condition 7 or 14 is true. In the affirmative, we assign to the entire region the disparity that is the smallest between the left and right disparities. Sample results are shown in figures 2(c) and 2(g).

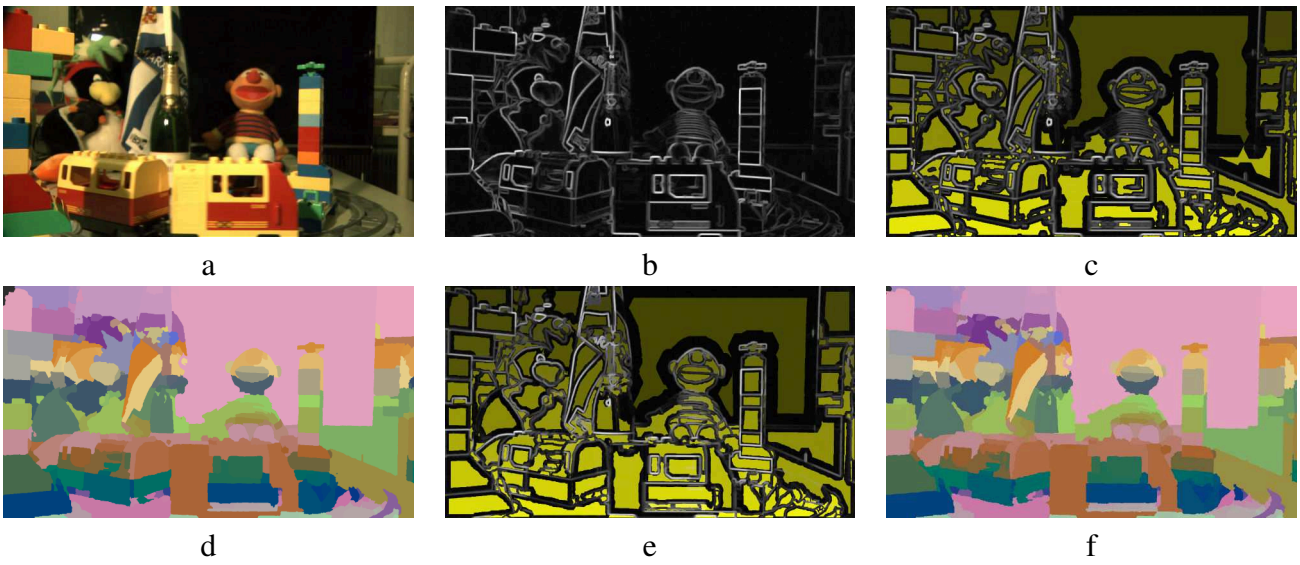

Figure 1: Watershed-based segmentation of stereo images. From the reference image (a) a colour thick gradient is computed (b). The h-minima of the gradient are rescaled using a distance function and serve as markers (c) used to drive the watershed segmentation of the reference image (d). The second image markers (e) are obtained by translating each region from the reference image according to its regional disparity and hence produce an equivalent segmentation (f) for the second image.

5. Results and Discussion

Figures 1(a–c) show the main stages of the watershed segmentation procedure for a reference image (namely colour gradient computation and markers generation). Figure 1(d) represents the resulting partitioning of the reference image. Thanks to the regional disparities computed for that image, markers are generated in order to capture the regions of the second image, i.e. they fall into the catchment basins of the second image gradient (figure 1(e)). Since each marker is assigned to the label of the reference image region from which it originates, we obtain an equivalent labelled partitioning (figures 1(d) and 1(f)) from which region matchings are obtained. Figure 2 presents brute regional and refined disparity measurements mapped onto the regions of the reference image. It clearly appears that refined disparities give more insight on the stereopsis than regional disparities from initial image segments alone. The mistaken assignments are mainly due to occluding contours: on figure 2(f), one can see that the disparity value assigned to the pig has been transferred to the surrounding region. The same goes for the cone as shown in figure 2(b). The disparity refinement algorithm splits those ambiguous regions in two and discovers brutal changes of disparity values for the left and right resulting subregions. Since one of the subregion seems to inherit from the disparity value of a neighbour region, the disparity of that subregion is hence corrected using the disparity found for the other subregion. Figures 2(c) and 2(g) result from the refinement algorithm. Considering direct neighbourhood only is however a strong limitation of the refinement algorithm as noticed for the region between the jar and the brush shown in figure 2(c): since that region is homogeneous and completely surrounded by occlusion contours, both left and right subregions have indeed invalid disparities. Future work on our framework will aim at considering more adequate models to overcome this kind of issue.

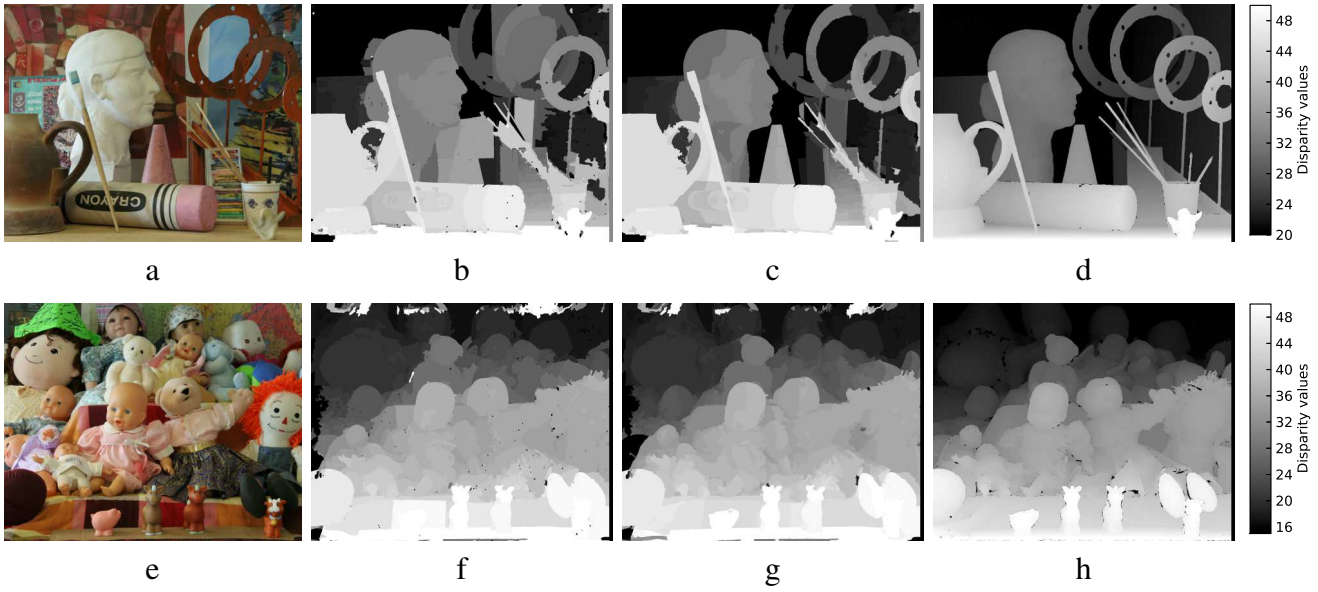

Figure 2: Disparity maps estimated over Middlebury stereo datasets `art` and `dolls` [13]. (a) and (e): Reference images. (b) and (f): Regional disparity maps. (c) and (g): Refined disparities mapped onto image subregions. (d) and (h): Ground truth disparities.

6. Conclusion

We have presented an approach that exploits regions and the disparities along their contours to estimate depth maps from stereo. The proposed algorithm relies on the concept of regional disparities and yields visually acceptable results with respect to ground truth disparities. Still it only partially solves two problems arising out of this approach: the need to accurately detect contours which are due to occlusions and a way to compute the evolution of disparities across non fronto-parallel surfaces. Our future work is bound to concentrate on the resolution of these shortcomings.

Acknowledgments

This work has been performed in the project PANORAMA, co-funded by grants from Belgium, Italy, France, the Netherlands, and the United Kingdom, and the ENIAC Joint Undertaking.

References

[1] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2004.

[2] Radu Horaud and Olivier Monga, *Vision par ordinateur: outils fondamentaux*, chapter Vision stéréoscopique, Hermes, 2nd edition, 1995.

[3] Marsha Jo Hannah, *Computer Matching of Areas in Stereo Images*, Ph.D. thesis, Stanford, CA, USA, 1974.

[4] Yuichi Ohta and Takeo Kanade, "Stereo by intra-and inter-scanline search using dynamic programming," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 2, pp. 139–154, 1985.

[5] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[6] Pascal Fua, "A parallel stereo algorithm that produces dense depth maps and preserves image features," *Machine vision and applications*, vol. 6, no. 1, pp. 35–49, 1993.

[7] C Lawrence Zitnick and Sing Bing Kang, "Stereo for image-based rendering using image oversegmentation," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 49–65, 2007.

[8] Michael Bleyer and Margrit Gelautz, "A layered stereo matching algorithm using image segmentation and global visibility constraints," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 3, pp. 128–150, 2005.

[9] Serge Beucher, *Segmentation d'Images et Morphologie Mathématique*, Ph.D. thesis, Ecole Nationale Supérieure des Mines de Paris, 1990.

[10] Serge Beucher and Nicolas Beucher, "Hierarchical queues: general description and implementation in mamba image library," Tech. Rep., Center of Mathematical Morphology - Mines ParisTech, 2011.

[11] Jesús Angulo López, *Morphologie mathématique et indexation d'images couleur: application à la microscopie en biomédecine*, Ph.D. thesis, Ecole Nationale Supérieure des Mines de Paris, 2003.

[12] Serge Beucher, "Maxima and minima: a review," Tech. Rep., Center of Mathematical Morphology - Mines ParisTech, 2013.

[13] Heiko Hirschmuller and Daniel Scharstein, "Evaluation of cost functions for stereo matching," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.