Dense Disparity Maps
Respecting Occlusions and Object Separation
Using Partial Differential Equations

Dennis Maier, Andreas Rößle, Jürgen Hesser, and Reinhard Männer

Lehrstuhl für Informatik V, Universität Mannheim
B6, 23-29, C, 68131 Mannheim, Germany,
Dennis.Maier@ti.uni-mannheim.de

Abstract. In this work, we present substantial enhancements to solve the stereo correspondence problem using a minimization and regularization formulation with a partial differential equations approach. For the first time it allows to respect occlusions and separation of objects.
We introduce a boundary condition that makes it possible to estimate disparities of arbitrarily shaped regions and thus to explicitly handle occlusions. After calculating a dense disparity map, we detect occlusions and object boundaries, cut the disparity map at these boundaries, and minimize the energy functional once more resulting in more accurate estimates. We show that we can achieve a speed up of a factor of four by rectified image pairs and a correlation based algorithm to calculate an initial estimate. In addition, a priori knowledge provided as region of interest or location of object boundaries further improves both the speed and the quality of the estimation. The results demonstrate that the quality is greatly improved with the proposed approach.

1 Introduction

Numerous algorithms to solve the stereo correspondence problem have been proposed in recent years. Feature based methods first extract certain corners or line segments from the images then establish correspondences between them, resulting in a sparse disparity estimation [11]. Area based methods on the other hand can produce dense disparity maps by correlating image patches, which work well in relatively textured areas, implicitly assuming a locally front-to-parallel scene [6]. More recently, research focused on energy based methods that minimize an energy functional including a regularization formulation. These global methods produce dense disparity maps and do not suffer from regions with low texture content. They assume a single smooth surface. A recent evaluation showed that these method yield much better results than feature or area based methods [10].

The problem of current partial differential equation (PDE) formulations, however, is that they implicitly have to estimate a disparity for every pixel in the image. As some pixels might not be visible in the other image due to occlusion the regularization forces a disparity estimate based on the local neighborhood [9]. The regularization leads to smooth artificial connections between

foreground and background objects. Anisotropic smoothing of the disparities alleviates this problem but does not remove it. Furthermore, the regularization will force wrong disparities in the vicinity of object boundaries even for pixels that are visible in both images because it has to generate a single smooth surface.

The PDE approach presented in [1] produces an anisotropically smoothed dense disparity map even in regions with little texture. The authors note that results are not perfect when occlusions are present as is the case in almost all complex images. The goal of this paper is to build on the anisotropic energy based method, but to respect occlusions and object separation. This will result in more accurate disparities at object boundaries. It is therefore necessary to be able to estimate disparities for arbitrarily shaped image regions possibly containing holes.

The remainder of this paper is organized as follows. In section 2 we first describe the energy based method presented in [1] on which our new approach is based. We then introduce a boundary condition in section 3 which allows various enhancements and improvements over the original algorithm as presented in section 4. Experimental results are given in section 6. Finally, conclusions and an outlook are given in section 7.

2   Disparity Estimation

The algorithm of [1] produces dense disparity maps using global anisotropic diffusion. The authors make use of the epipolar geometry to better estimate disparities. The point x' corresponding to a pixel x in the one image has to lie on its epipolar line, which can be described by the following equation

$$x'^\top F x = 0, \tag{1}$$

where F is a homogeneous 3×3 matrix known as the fundamental matrix [4].

Assuming that corresponding pixels in the two images have equal color values, we seek a vector function $h(x,y) := (u(x,y), v(x,y))^\top$ such that

$$I_i(x,y) = I_i'(x + u(x,y), y + v(x,y)) \quad i = 1, 2, 3, \tag{2}$$

where I_i respective I_i' represent the 3 color channels of image I respective I'.

The following auxiliary functions are defined:

$$\begin{aligned} a(x,y) &= f_{11}x + f_{12}y + f_{13}, \\ b(x,y) &= f_{21}x + f_{22}y + f_{23}, \\ c(x,y) &= f_{31}x + f_{32}y + f_{33}, \end{aligned} \tag{3}$$

where f_{ij} are the entries of the fundamental matrix. The epipolar line is parameterized using a function $\lambda(x,y)$, then the disparity term depends just on one parameter.

To estimate the disparities $\lambda(x,y)$, the following energy functional is minimized:

$$E(\lambda) = E_D(\lambda) + E_S(\lambda). \tag{4}$$

The data term E_D is defined as

$$E_D(\lambda) = \sum_{i=1}^{3} \int_\Omega (I_i(x,y) - I'_i(x + u(\lambda), y + v(\lambda)))^2 \mathrm{d}x\mathrm{d}y \qquad (5)$$

while the smoothness term E_S is

$$E_S(\lambda) = C \int_\Omega (\nabla\lambda)^\top D(\nabla I_{max})\nabla\lambda \mathrm{d}x\mathrm{d}y \qquad (6)$$

where Ω is the image domain and C is a positive constant and the matrix $D(\nabla I_{max})$ is given by:

$$D(\nabla I_{max}) = \frac{1}{\mid \nabla I_{max}\mid^2 + 2\nu^2} \left\{ \begin{bmatrix} \frac{\partial I_{max}}{\partial y} \\ -\frac{\partial I_{max}}{\partial x} \end{bmatrix} \begin{bmatrix} \frac{\partial I_{max}}{\partial y} \\ -\frac{\partial I_{max}}{\partial x} \end{bmatrix}^\top + \nu^2 \mathrm{Id} \right\} \qquad (7)$$

where Id is the identity matrix and

$$\nabla I_{max}(x,y) = \{\nabla I_{i_0}(x,y), \qquad \text{such that}$$
$$\|\nabla I_{i_0}(x,y)\| \geq \|\nabla I_i(x,y)\| \quad \forall i = 1,2,3)\}.$$

The energy functional establishes a balance between the equality of the pixel intensities of corresponding points and the regularity of the disparity. The smoothness term is anisotropic. In regions with homogeneous intensities disparities are smoothed in all directions, in areas with color edges smoothing is mainly along the edge not across it. To compute the minima of the above functional the associated Euler-Lagrange partial differential equation is used.

Applying a gradient descent method to the Euler-Lagrange equation requires to solve the following parabolic differential equation:

$$\frac{\partial\lambda}{\partial t} = C \, \mathrm{div}(D(\nabla I_{max})\nabla\lambda)$$

$$+ \sum_{l=1}^{3} \left(I_l(\overrightarrow{x}) - (I'_l)^\lambda(\overrightarrow{x})\right) \frac{a \left(\frac{\partial I'_i}{\partial y}\right)^\lambda (\overrightarrow{x}) - b \left(\frac{\partial I'_i}{\partial x}\right)^\lambda (\overrightarrow{x})}{\sqrt{a^2 + b^2}} \qquad (8)$$

Its asymptotic state for $t \to \infty$ gives the estimated disparity. The authors of [1] have proven that if I, I' are sufficiently regular functions, there exists a unique solution of that equation for any initial solution. To calculate the disparities, the parabolic differential equation is approximated by finite differences. The resulting linear system of equations has a tridiagonal structure and is solved iteratively with a symmetric Gauß-Seidel algorithm. The linear implicit scheme allows using large time steps in order to increase the convergence rate. For more details refer to [1].

3   Boundary Condition

Every pixel with less than eight neighbors is a boundary pixel. The algorithm of [1] assumes a complete 8-neighborhood. These neighbors are used to calculate the finite difference approximations of the second derivatives in the smoothness term. If one of these eight neighbors is not present the algorithm cannot compute the smoothness term.

Therefore we introduce a boundary condition. We use the Neumann boundary condition so pixels on the boundary can have locally varying values and no energy will be transferred across the boundary: $\partial f / \partial n = 0$.

It becomes necessary to uniquely determine the direction of the boundary normal to be able to compute the derivative of the boundary normal. Arbitrarily shaped areas (that result from user defined regions of interest or because occluded pixels have been eliminated) can lead to non-unique normals of boundary pixels. Prerequisite for a unique direction is a connected boundary in the 8-neighborhood, where connected means that all valid pixels are vertically or horizontally linked. Furthermore, an 8-neighborhood may contain a maximum of 5 invalid pixels. If one of these criteria is not met, the pixel will be marked as invalid. This procedure is iterated until there are only valid pixels. This process may invalidate pixels, but guarantees that we can apply the Neumann boundary condition.

Using central differences to compute the derivatives of the boundary normal requires support points which lie in invalid regions. We mirror existing values at the boundary normal to fill these missing values, then use these central differences to approximate the Neumann boundary condition.

4   Proposed Method

4.1   Initial Estimate

In general, the solution of the Euler-Lagrange partial differential equation will not be unique. Thus, the necessary initial disparity estimate should be close to the true disparity in order to avoid convergence to a local minimum. The authors of [1] propose a pyramidal strategy. The algorithm as described in section 2 is applied to each level of an image pyramid, where the initial estimate for each pyramid level is extrapolated from the result of the previous level. This approach yields good results but is quite time consuming.

Instead, we propose an efficient correlation based stereo algorithm [6] as basis to compute the initial estimate. The resulting disparity map will be sufficiently accurate to serve as a starting point for the energy based estimation. However, the correlation based stereo algorithm typically does not produce estimates for all pixels in the image due to occlusions or texture-less regions. Assuming it is better to produce a sparse disparity map than a wrong one, the correlation based algorithm tries to eliminate false matches with a left to right consistency check and uniqueness validation. This disparity map must be filled since we need

616

estimates for every pixel in the image in order to be able to apply the iterative algorithm described in section 2.

We assume that the disparity map is based on the left image. In this case, scene points that are only visible in the left image and occluded in the right image can only appear at the left side of objects. Therefore, missing estimates are assigned to the background in the following way: each missing estimate is set to the closest available estimate to the left of it, if there is no such estimate its value is set to the closest available estimate to the right of it. Afterwards, still missing estimates are filled from the top respectively bottom of the image.

4.2   Rectified Images

Our correlation based algorithm uses rectified images [7], which is essential for an efficient implementation. The fundamental matrix for rectified images is substituted into (3), which leads to a simpler formula, because some terms in the LSE can be dropped.

Note that rectified images usually contain a priori known regions that cannot be matched into the other image, since the epipolar lines for these pixels do not intersect the other image. Time spent on calculating disparities for these a priori known regions is wasted. This is one reason for the introduction of a region of interest, see section 4.5.

4.3   Occlusion Detection

Occlusions occur in almost all complex images. Yet there is no perfect way to detect them [3]. Occlusions influence our image understanding regarding distance estimation and object segmentation [2]. It is necessary to detect occluded image areas because it is not possible to compute disparity estimates in these regions. In global methods the smoothness term plays a dominant role in occluded areas. Discontinuity preserving methods allocate the occluded region partly to the foreground object and partly to the background object with a steep disparity gradient in between which introduces an artificial edge in the 3D-reconstruction. A comparison of methods to detect occlusions can be found in [3].

The algorithm of [1] does not model occlusions which leads to large errors in the disparities of occluded pixels and also in their neighborhood. Although occluded pixels do not have corresponding pixels in the other image, the algorithm must yield an estimate for their disparity, which is largely derived from the smoothness term, trying to find correspondences so that the energy functional is minimal. If the disparity gradient is very large at pixels close to boundaries, the smoothness term will lead to wrong, locally smoothed disparities at those pixels.

For the results in this paper, we use a combination of the bimodality method and the "matching goodness jumps" method to detect occlusions. The bimodality method [5] uses the fact, that occlusions occur at disparity edges. Since we assume that our disparity maps are based on the left image, occlusions can only occur in areas where the disparity gradient is positive. This identifies possible

locations of occlusions but does not help us in identifying the size of the occluded region, i.e. we do not know how many pixels left of the detected location are occluded. To overcome this problem we use the "matching goodness jumps" method, which states that bad matching pixels have a high probability of being occluded. We examine for every scanline pixels left of large positive disparity gradients. If the color intensity difference lies above a threshold, the pixel will be added to the occluded area and we continue in the same way with the next pixel left of it. The occluded area ends as soon as there is a pixel which is being judged as not occluded.

4.4   Object Separation

The smoothness term forces a smooth surface over the entire image. Although the method uses anisotropic diffusion respecting discontinuities, foreground and background objects still become artificially connected. However, the disparities in the vicinity of object boundaries will have much steeper gradients as if one would use isotropic smoothing. This makes it well suited to detect object boundaries. If we are able to detect these boundaries, we can "cut" the disparity map at these locations, introducing new boundary pixels in the problem formulation.

One expects the smoothness term to relax the surface at the object boundaries, if one takes the resulting disparity map as new initial estimate and then minimizes the energy functional. The algorithm then does not force a smooth surface over the disconnected pixels, so disparities should converge to their true value. This will result in more accurate estimates.

The disparity gradient serves us as a criterion where to cut the disparity map. If the gradient is above a threshold the disparity map is cut at that location. The threshold is determined from the global gradient histogram. This way we try to cut the artificial connection of objects at their greatest disparity gradient.

4.5   Region of Interest

Sometimes one might be interested in only a certain region of the image. Excluding pixels from the calculation will accelerate the algorithm. Image preprocessing may already exclude uninteresting image areas, but this way only rectangular areas may be selected. Since we are now able to compute disparities for arbitrarily shaped regions, a more detailed selection can be done. For example, a circular region may be selected for stereo endoscope images. The lens undistortion and rectification stage may give another criterion for a region of interest.

We implemented an automatic background detection for images of objects in front of a uniform background. This is useful, for example, in turntable sequences to extract the object of interest.

Furthermore, a priori knowledge of object boundaries can be included in a straight forward way. We simply remove pixels on the boundaries from the region of interest. This will automatically disconnect the disparities along the boundary and will help to improve the quality of the results.

5 Algorithm Summary

Having described all enhancements to the original approach, we are now ready to summarize our complete algorithm. The steps of the algorithm are:

1. Rectify the stereo image pair if the camera system was not in standard geometry.
2. Compute initial disparities using an efficient correlation based algorithm.
3. Optionally perform background detection and respect user provided region of interest.
4. Minimize the energy functional with boundary condition.
5. Detect occlusions as described in section 4.3.
6. Cut the disparity map along disparity gradients to separate objects.
7. Minimize the energy functional with boundary condition to relax disparities at object boundaries.

6 Results

All calculations for this paper were performed on a Pentium III at 800MHz with 512MB RAM. We used the software of [8] and our own implementation for the proposed algorithm to compare the two methods. First, we examine the time of the pyramidal strategy versus the correlation based method for the computation of the initial disparity estimate. Columns 3 and 4 of table 1 list the times needed to obtain the initial disparity estimates for a few image pairs. The window size was 7×7 and the disparity range was set to -50 to 50 for the correlation based algorithm. Both methods have runtime proportional to the image size. The correlation method is significantly faster, and disparities at object edges are not smoothed as in the pyramidal approach. However, missing disparities must be filled with guessed values as described in section 4.1.

Table 1. Execution times.

Image	Size	Pyramid	Correlation	Non-Rectified	Rectified
Synthetic	285 x 206	15 sec	< 1 sec	35 sec	11 sec
Tsukuba	384 x 288	28 sec	< 1 sec	90 sec	30 sec
Dinosaur	761 x 595	100 sec	< 3 sec	450 sec	136 sec

A further significant speed up was possible due to the simplification that arose because we use rectified images. The speed up could be achieved because the simple geometry of rectified images allowed to eliminate some terms and now only a linear instead of a bilinear interpolation of the original algorithm is necessary. Columns 5 and 6 of table 1 show the timings for step 4 of our algorithm. Using rectified images increases the speed of the algorithm by a factor

of three. Including the time needed for the calculation of the initial disparity map, the new algorithm is four times faster than the original. Computation time can be even further reduced if it is not necessary to compute disparities for every pixel of the image.

Table 2. Runtimes with and without region of interest.

Image	ROI	Full Image	Restricted
Synthetic	63.5%	11 sec	10 sec
Endoscope	46.1%	34 sec	24 sec
Dinosaur	15%	136 sec	61 sec

Table 2 illustrates the speed up of restricting the area for the calculation. As expected, calculation is faster when the area can be restricted, however the speed up is not proportional to the percentage of excluded pixels. Although 85% of the pixels could be excluded for the dinosaur image, only a speed up of about 55% could be achieved, because the number of forward and backward eliminations in the Gauß-Seidel method may vary due to different initial estimates.

The result of the occlusion detection and the object separation can be seen in figure 1. Our algorithm could detect most occluded pixels, however it also marks more pixels as occluded than necessary. The proposed method emphasizes on speed and results in a reasonable occlusion detection. However, any other method to detect occlusions may be used in step 5.

Fig. 1. Occlusion detection and object separation. One image of the stereo pair (left), disparity map with occlusions marked red (middle), and disparity map with occlusions and object separation marked red (right).

The cuts in the disparity map created by step 6 of our algorithm greatly improve the results because they disconnect foreground and background objects and enable a more accurate estimation of the disparities in the vicinity of the cuts. Occlusion detection and cutting the disparity map both work towards this

goal. For better visualization, we use a random dot stereogram to demonstrate the qualitative improvements of our approach. The stereogram contains two distinct disparity levels, a square elevated over a plane. Figure 2 shows from left to right 3D-models generated from the disparity maps of the original algorithm, of our approach with automatic occlusion detection and object separation, and of our approach with ideal occlusions and object separation. The image on the right of figure 2 illustrates that a priori knowledge clearly improves the quality of the results even more. In this case the location of occluded pixels and the object separation was provided as region of interest.

Fig. 2. 3D-models generated from disparity maps. Result of original algorithm (left), our proposed method (middle), and our method with ideal occlusions and cuts (right).

Figure 3 shows a plot of the disparities along a scanline through the center of the stereogram for the original algorithm and for our proposed method. The dotted line shows the smoothed disparities of the original algorithm. The solid line represents the disparity estimates of the proposed method. After the occlusion detection (notice the gap between the disparities for pixel 59 and 70) and object separation the disparities relaxed to their true values and the two objects became disconnected.
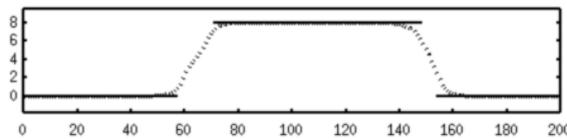
Fig. 3. Disparities along a scanline through the center of the stereogram. Original algorithm (dotted) and proposed method (solid).

The occlusion detection and the object separation generate new boundary pixels. The disparities in the vicinity have not yet adjusted to the new situation. Although the iterative disparity estimation is executed a second time, we found empirically that about half the number of iterations of the first step is sufficient because the disparity estimates already converged except for pixels in the vicinity of the newly introduced boundaries. As a rule of thumb, step 7 of our algorithm increases the total computation time by about 30%.

7   Conclusions

We have presented an algorithm to compute stereo correspondences for color images based on the algorithm presented in [1]. The use of a boundary condition as proposed in this paper enables us to dissect the disparity map and therefore to explicitly handle occlusions and to disconnect objects at their boundaries. Thus there is no smoothing over object boundaries provided they are correctly detected. This results in more accurate disparity estimates.

Using an efficient correlation based stereo algorithm to calculate the initial estimate speeds up the calculation significantly. The simplifications in the discretized parabolic equation due to the standard geometry of rectified image pairs further increased the speed by a factor of three.

The occlusion detection used in this paper is fast and simple but may discard too many pixels. To further improve the algorithm, one could use another strategy to detect occlusions. The simple implementation of the separating stage may introduce cuts more than one pixel wide. Cutting only at the steepest location would result in a slightly denser disparity map.

The proposed algorithm is faster and more accurate than the one presented in [1]. It may be used if accurate, dense disparity maps are required and computation time of about a minute for a PAL image can be tolerated.

References

1. Luis Alvarez and Javier Sánchez: 3-D Geometry reconstruction using a color image stereo pair and partial differential equations. Cuadernos del Instituto Universitario de Ciencias y Tecnologas Ciberneticas, Vol. 6, pp. 1–26, 2000.
2. B. Anderson: The role of partial occlusions in stereopsis. *Nature*, 367, pp. 365–368, 1994.
3. Geoffrey Egnal and Richard P. Wildes: Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches. *IEEE TPAMI*, 24(8), August 2002.
4. Richard Hartley and Andrew Zisserman: Multiple View Geometry in computer vision. Cambridge University Press, 2000.
5. J. Little, W. Gillett: Direct evidence for occlusion in stereo and motion. *ECCV*, pp. 336–340, 1990.
6. Karsten Mühlmann, Dennis Maier, Jürgen Hesser, Reinhard Männer: Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. *International Journal of Computer Vision*, 47(1), pp. 79–88, May 2002.
7. M. Pollefeys, R. Koch, L. Van Gool: A simple and efficient rectification method for general motion. *Proc. of the 7th ICCV*, pp. 496–501, Sept. 1999.
8. Javier Sánchez: StereoFlow Software. Demo Page, `http://serdis.dis.ulpgc.es/~jsanchez/research/software/StereoFlow/StereoFlow.htm`
9. D. Scharstein, R. Szeliski; Stereo matching with nonlinear diffusion. *IJCV*, 28(2), pp. 155–174, 1998.
10. D. Scharstein, R. Szeliski: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1), pp. 7–42, May 2002.
11. O. Veksler: Semi-Dense Stereo Correspondence with Dense Features. *IEEE Workshop Stereo and Multi-Baseline Vision*, pp 149–157, 2001.