Automatic Disparity Search Range Estimation for Stereo Pairs of Unknown Scenes

Jana Kostková and Radim Šára

Center for Machine Perception, FEE, Czech Technical University
Karlovo nám. 13, 121 35 Prague, Czech Republic
{kostkova,sara}@cmp.felk.cvut.cz

Abstract

Known disparity search range is crucial for stereo matching tasks, since many algorithms require the disparity search range to be known. Searching over the whole disparity range (i.e. $[-\text{image.width}, \text{image.width}]$) is not only very time consuming (mainly for large images), but even many stereo algorithms do not perform well with unspecified disparity search range. Therefore, automatic estimation of disparity search range for unknown stereo image pairs is highly desired. In low-level image processing (i.e. without knowing any information about the captured scene) this task is very difficult. We propose an approach based on Confidently Stable Matching, which is fast, precise and robust. We demonstrate the algorithm properties on benchmark image sets with known disparity search range as well as on unknown complex scenes.

1 Introduction

Stereo matching gained a big interest of researches over the last four decades due to its importance in computer vision. Its core problem is to establish the correspondences between the binocularly visible regions in the input images. The results are usually presented as a disparity map, where the disparity represents the shift between the corresponding points from the input images. Typically, it is assumed that the input images are rectified, i.e. that the epipolar lines coincide with the corresponding image rows, which reduces the matching problem from 2D to 1D search problem, where the correspondences are found only in the respective image rows.

In general, one of the crucial parameters for the overall success of an arbitrary matching algorithm is the disparity search range. This parameter defines the minimal and maximal disparity in the given stereo image pair, which reduce the search of corresponding point over all the tentative pairs to only pairs admissible by this range. Unknown disparity search range can be replaced by searching over the whole disparity range (i.e. $[-\text{image.width}, \text{image.width}]$),

*This work has been supported by the European Union under project IST-2001-32184, by the Czech Ministry of Education under project MSM 212300013, by The Grant Agency of the Czech Republic under project GACR 102/03/0440 and by The Grant Agency of the Czech Technical University under project CTU 0306413.
Figure 1: Demonstration of the importance of disparity search range specification. In the first row, the disparity maps are computed using the given disparity search range, in the bottom row, the solution is searched out over the whole image. The disparity is colour-coded: small disparities (background) are dark blue, large disparities (foreground) are red, unassigned correspondences are gray. The input left image is shown in Fig. 6.

but it is typically much more time consuming (especially for the large images) and moreover, many algorithms decrease their performance if the disparity search range is not specified. In Fig. 1, results of the same algorithms with specified disparity search range (top row) are compared with results without it (bottom row), the input left image can be found in Fig. 6. We can see that in the case of the Confidently Stable Matching approach [8] the non-specified range has decreased significantly the matching density and increased the running time about 14× (from 5.96 s to 83.93 s), while the accuracy has been preserved. In the case of the Graph Cuts approach [3], with non-defined search range the algorithm does not converge (it is necessary to limit the number of iterations manually) and its performance decreases (c.f. the wrong correspondences around the car). Finally, in the case of the Winner-take-all algorithm [10], the running time has been increased 7× (from 1.1 s to 7.4 s), and the algorithms performance significantly decreased. The other way is to design algorithms which find correspondences without specifying the search range. Such approaches work typically on multiresolutions [1, 7, 6], but they tends to create false positives in the results, which we want to avoid.

Well-known and benchmark stereo image pairs are typically provided with already given disparity search range. However, we usually want our algorithms to be run on our data sets and images where we have no information available. One possibility is to estimate the disparity search range manually, which might not be very rigorous. Or we can use special equipment [9], which cannot be applied in general. Nevertheless, for the automatic image processing, which is nowadays an important topic (e.g. automatic 3D scene reconstruction from a set of uncalibrated images), the automatic disparity search range estimation is inevitable. In low-level
vision, i.e. the image processing based only on the pixels and their intensity values, this task is very difficult, however. Therefore, not much work on this topic has been done. To our knowledge, only one approach dedicated directly to automatic disparity search range estimation has been published [2]. Their approach is based on spatial correlation statistical analysis between stereo pairs, so called image variograms, which show the function of similarity (normalised SSD) between the stereo images depending on the increasing disparity. The first significant minimum in this function determines the maximum disparity. The main drawbacks of this approach are following: (1) they do not estimate the disparity search range itself, but only the maximum expected disparity between the stereo pair, (2) they assume only positive disparities between the images, however negative disparity is a common phenomenon (introduced during the rectification process), and (3) there is no guarantee that the algorithm would not neglect the thin foreground objects (of high disparity).

We propose an approach which is completely automatic and able to estimate disparity search range (i.e. the minimal and maximal disparity between the input images). We assume the images are already rectified as stereo approaches in general do. Our approach is based on the Confidently Stable Matching [8] algorithm due to its qualities [5]: on the strict confidence level it gives disparity maps with (almost) no false positives and mismatches. They are very sparse, which however does not make any problems for this application. The described properties allow to estimate the search range from the computed disparity map. The given scene geometry and properties are obviously unknown, however to get good results, we assume the objects of the scene which are at margins of the disparity range fulfill the following conditions: (1) they are of a strong/good texture, (2) this texture is unique in the search range, and (3) they are big enough to be recognised as objects. In such scenes we are able to guarantee to get the search range correctly. Nevertheless, if these conditions are not fulfilled, the search range will be estimated to objects which meet these criteria (and thus it will be a bit narrowed), since the objects violating these conditions are typically of low interest (e.g. in the case of scene reconstruction we even want to avoid them). For speeding up the process, we propose a hierarchical algorithm, where the disparity search range is refined from the initial estimate on the coarse level up to the fine resolution.

The paper is organised as follows: In the next Section, we describe our approach to automatic disparity search range estimation, and introduce the Confidently Stable Matching and its properties useful for our application. In Sec. 3, its experimental performance is demonstrated and finally, conclusions are given in Sec. 4.

2 Estimation Algorithm

In the case of automatic image processing two aspects have to be observed while designing new approaches: time efficiency and results accuracy. Both of them are obvious but crucial for applicability of the method. To achieve this in our algorithm we profit from the remarkable properties of the Confidently Stable Matching approach [8]: very low error rates [5], and good statistical properties of its results, which are discussed in Sec. 2.1.

In our approach, we have decided to use a hierarchical method due to two reasons: time
demands and matching sparsity. Estimating the search range on the whole images is slow and the matching is very sparse (as long as the image size increases the matching density rapidly decreases). On the other hand, estimation on the small sub-sampled images (where then the search range is only enlarged according to the sub-sampling factor) is fast, however sub-sampling might cause huge distortions in the images (e.g. skipping thin or small objects due to high frequencies filtering). The results demonstrating this observation are shown in Fig. 2. The images are of a random texture, the level 1 represents the optimal texture on the camera resolution. We can see the matching density decreases on the both sizes from this level. The running time constantly increases with increasing image size. Thus, we propose a coarse-to-fine estimation: we sub-sample the images to a small size which allows to compute the disparity map very fast and get the initial disparity search range estimate. Then in each step, the images are enlarged and the search range is refined from the improved disparity map.

In proposing the estimation algorithm two problems, which need to be solved, arise: the shape of the sub-sampling pyramid and the threshold for the acceptable matching density.

In the subsequent paragraphs we introduce the Confidently Stable Matching algorithm, describe the two main problems in the estimation algorithm and discuss their possible solutions and finally we summarise the whole estimation approach.

2.1 Confidently Stable Matching

This approach, proposed by Šára [8], is able to find the largest unambiguous subset of the matching on a given (by the user) confidence level. The matching process is based on the stability principle: it solves an optimisation task which is defined on mutual occlusion graph $G = (T, E)$ in which the vertex set $T$ is the set of all tentative matches (pairs) and $(t, s)$ is an edge in $E$ if pairs $t$ and $s$ are mutually exclusive, i.e. cannot be both elements of the same matching due to occlusion. We use uniqueness and ordering constraints as the occlusion model. Every pair $t$ in $T$ thus has a set of competitors (neighbours in $G$) $N(t)$ which we call the inhibition zone of $t$.

\[ N(t) = \{(k, l) \mid k = i \text{ or } l = j, (k, l) \neq (i, j)\}. \]

1Inhibition zone for matchings: if $t = (i, j)$ then $N(t) = \{(k, l) \mid k = i \text{ or } l = j, (k, l) \neq (i, j)\}$. 

Figure 2: Dependence of the matching sparsity and running time on the image resolution. The resolution of the optimal texture corresponds to the level 1.
Figure 3: Dependence of the matching properties on the parameter values: (a) matching accuracy, (b) matching sparsity and (c) the running time.

statistics. We say a pair \( t \in T \) is confidently excluded by another pair \( e \in T \) if \((t, e) \in E\) and \(c(t) \leq c(e) - \Delta(t, e)\). The value of \(\Delta(t, e)\) is related to the confidence interval widths of \(c(t)\) and \(c(e)\). Confidently stable subset \(M\) is the largest subset of \(T\) such that every pair not in \(M\) has either no unambiguous competitors in \(M\) (due to a circular exclusion conflict) or is confidently excluded by a competitor in \(M\). Simply and somewhat imprecisely, all pairs not in \(M\) are either ambiguous in \(c(\cdot)\) or confidently occluded by some strongly better match in \(M\). If exclusion takes into account uniqueness the stable subset is a (univalued) matching. For precise definitions, existence and uniqueness theorems, and the algorithm see [8].

The properties of this algorithm are: low false positive and mismatch errors, which is paid by higher matching sparsity. The numeric values of these errors depend only on the confidence level parameters: the wider the confidence interval is, the more accurate and sparser the matching is. This property makes the Confidently Stable Matching (CSM) approach well suitable for the automatic disparity search range estimation, since it requires accurate results, while low density does not make any problems.

The confidence level of the matching is determined by two parameters: \(\alpha\) and \(\beta\), where \(\alpha = \alpha_0 \cdot \sigma^2\) and \(\sigma^2\) is the image intensity variation. The values \(\alpha_0 = 0\), \(\beta = 0\) put no constraints to the matching, which is thus denser but with more errors. Increasing these parameters makes the matching more accurate but sparser with increasing running time. Thus, we have to make a trade-off between accuracy, sparsity and speed, nevertheless, this parameter is not critical for the methods success. We have derived the parameters value based on our ground-truth evaluation method [5]. The method studies different errors separately, which allows to measure rigorously the properties of the algorithm’s results. We have run the evaluation experiment on images with the optimal texture (of the highest texture contrast) while changing the parameters values. Based on the results shown in Fig. 3 we decided to set the parameters to \(\alpha_0 = 0.015\) and \(\beta = 0.1\) (which correspond to the parameters value of 30 from Fig. 3 plots), since it reaches the optimum: the accuracy (Fig. 3(a)) is about 99.7\%, while about 30\% of the pixels is matched (Fig 3(b)) and the running time (Fig. 3(c)) is good enough. It has been verified also on various images that it is able to capture the objects with significant texture, so it might lose indistinct features at disparity search range margins, which however are typically hard to match anyway. The excellent property of the CSM approach is that on the given confidence
Figure 4: CSM matching properties on the confidence level of $\alpha_0 = 0.015$ and $\beta = 0.1$.

Figure 5: CSM matching features extraction: two left images show the matching results on the untextured scene, the two right images show the results on the well textured scene.

level the error rates are independent on the texture contrast, as it has been shown in [5, 4]. In our case of $\alpha_0 = 0.015$ and $\beta = 0.1$ we can guarantee: no false positives (Fig. 4(a)), only about $0.3\%$ of mismatches (Fig. 4(c)), and reliably sufficient density, see Fig. 4(b).

The another advantage of CSM is that it does not need a prior detection of matchable image features: it automatically recovers them as a side-effect of the matching process. To demonstrate this property, let us assume that we have a complex scene with completely untextured objects, the algorithm then finds the correspondences only on the object boundaries, i.e. the only features in the image, resulting in sparse matching. However, while having well textured scene, the results are dense. Both these examples are shown in Fig. 5. Therefore, we need not to be worried about getting incorrect disparity search range estimate even in difficult scenes.

CSM disparity map has two remarkable properties very useful for our approach: constant mismatch error and specific disparity distribution. The mismatch rate has been already discussed and the results can be found in Fig 4(c). Our experiment showed that only about $0.3\%$ of matched correspondences are incorrect independently on the texture contrast, defining the error percentile. The second excellent property of the CSM disparity map is the disparity distribution. The disparity histogram typically consists of two components: (1) the component corresponding to image data (objects in the scene), and (2) the component corresponding to disparity errors (mismatches). The first one typically arises at disparities where the objects are (correct disparities), represented by a high number of pixels. The latter one arises at all disparities, it is uniformly distributed and represented by very low number of pixels. The histogram normalised to the number of pixels in the input images is shown in Fig 4(d). By combination of both these properties we can specify the valid disparities (corresponding to objects) in the
resulting matching. We can assume that disparities represented by less pixels than this error percentile correspond to mismatches, and consequently such marginal disparities (if exist) can be excluded from disparity search range. Fig. 4(d) shows such situation: the red line stands for the given percentile level and the disparities below it are excluded.

2.2 Sub-sampling Hierarchy

The sub-sampling process plays the role of import in the algorithm since it affects significantly the results (there is a trade-off between the input images size, loosing the scene details and the running time, cf. Fig 2). The number of levels in coarse-to-fine estimation process should not be high, but still reasonable. We have experimentally derived that four levels are optimal: the first one represents the smallest image, where the disparity map is computed over the whole search range to get the initial estimate, the fourth one correspond to the original input images. For time efficiency we have used the ratio $1 : 2$ as the sub-sampling factor between the constituent levels.

Very important parameter for the method’s success is the selection of the smallest image size (the coarse level). Having it very small, there would not be anything left in the sub-sampled images from the original scene. On the other hand, the large size of the images may cause that the search range will not be estimated at all since there would be too many competitors for each pair (because of not yet specified search range), and thus the disparity map will be sparser than required, see Fig. 2(a). The subsequent enlarging of the image will not improve the estimate and the algorithm will return the whole image range. Because of these reasons, the setting the adequate smallest image size is crucial for the good disparity search range estimation. We have experimentally verified that the optimal minimal image size should be around 100 pixels, since on this size in general images the main scene structures are still preserved, while there are not many competitors and thus it is possible to get acceptable initial estimate. Thus, the images are sub-sampled to the closest size to that depending on the original image size and the sub-sampling factor. However, due to various kinds of images with distinct properties, an interesting solution would be to define the smallest image size based on the frequencies in the input images, which will be a topic of our future research.

2.3 Minimal Matching Density

The matching density is directly determined by the given confidence level and the discriminability of matching features. Since the confidence level is fixed in our approach, the density depends on the matching features discriminability. As the image resolution increases, the matching feature discriminability decreases (texture is above the optimal level in general images) and the number of matching competitors for each pair increases. Consequently, the matching density decreases with the increasing image resolution, see Fig. 2(a): on the lowest hierarchy level it is typically between $30\%$ and $50\%$, while to the highest level it rapidly decreases up to less than $1\%$. Based on this experiment, we have determined the minimal acceptable matching density (to guarantee reliable data) as follows: the percentage of assigned correspondences $t_{density}$ is required to be at least $2\%$ of all the input image pixels. If the re-
sulting disparity map does not fulfill this condition, the algorithm terminates and the search range is only enlarged to the original image size based on the sub-sampling ratio. The process is terminated since enlarging the images will only increase the number of competitors of each pair, i.e. decrease the disparity map density, and therefore it will not upgrade the estimate.

2.4 Algorithm Overview

The input for our algorithm is the stereo image pair and it returns the disparity search range \([d_{\text{min}}, d_{\text{max}}]\). The estimation process can be summarized as follows:

1. Subsample the input images to the smallest size
2. Solve the matching problem over the whole disparity search range
3. Estimate the initial disparity search range
4. Enlarge the input images and the search range
5. Solve the matching problem using the estimated search range
6. if the disparity map is sparser than \(t_{\text{density}}\) goto 4
7. Eliminate disparities corresponding to the error percentile
8. Estimate new precise disparity search range
9. if the image size is not equal to the original one goto 4

3 Results

We would like to demonstrate the estimation procedure on benchmark image Parking Meter, used for the motivation example in Fig. 1. It has been estimated in three steps (since the last one was too sparse). The results are shown in Fig. 6: first row corresponds to the sub-sampled input images, the middle one to the estimated disparity map while the bottom one to the disparity histograms. They are organized from left to right in the increasing resolution order. On the top of disparity maps there is depicted the matching density, we can see that the highest has been achieved in the second sub-sampling level.

We have tested our algorithm on various benchmark images of different scenes: real as well as artificial. The results are shown in Tab. 1. As it can be seen, the disparity search range is in all experiments estimated accurately (cf. with Ground-truth column). Also the running times are low, even for very large images, such as St. Martin (the disparity search range estimation lasted less than 11 minutes). We can see that for the St. Martin image pair the lower bound of the search range has not been estimated well: the furthest object in the image is the cross on the rotunda roof and since it is represented by very low number of pixels, their disparities have been rejected from the search range. However, in the automatic processing we can never guarantee that such small objects will not be missed.
Figure 6: Demonstration of the estimation procedure: first row shows the input images, the middle one computed disparity maps, and the bottom one the disparity histograms. The disparity map is colour-coded as it has been described in Fig. 1.

4 Conclusions

In this paper we have proposed an approach to automatic disparity search range estimation. It profits from the good properties of Confidently Stable Matching algorithm, which is fast and precise. Our experiments showed that the estimated search range covers the significant image features (i.e. of a good and strong texture and big enough), while it might miss some indistinct image objects at the disparity margins. Such objects however typically cannot be matched at all, and consequently our approach allows to extract from the image its parts of interest, which is desired by various applications of automatic image processing.
<table>
<thead>
<tr>
<th>Image Pair</th>
<th>Image Size</th>
<th>Disparity Search Range</th>
<th>Running Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parking Meter</td>
<td>480 × 512</td>
<td>-512 ... 512</td>
<td>0 ... 30</td>
</tr>
<tr>
<td>Map</td>
<td>216 × 284</td>
<td>-284 ... 284</td>
<td>0 ... 29</td>
</tr>
<tr>
<td>Birch</td>
<td>484 × 640</td>
<td>-640 ... 640</td>
<td>0 ... 55</td>
</tr>
<tr>
<td>Shrub</td>
<td>480 × 512</td>
<td>-512 ... 512</td>
<td>10 ... 30</td>
</tr>
<tr>
<td>Pentagon</td>
<td>512 × 512</td>
<td>-512 ... 512</td>
<td>-10 ... 10</td>
</tr>
<tr>
<td>Sawtooth</td>
<td>380 × 434</td>
<td>-434 ... 434</td>
<td>0 ... 19</td>
</tr>
<tr>
<td>Venus</td>
<td>383 × 434</td>
<td>-434 ... 434</td>
<td>0 ... 19</td>
</tr>
<tr>
<td>Lab Scene</td>
<td>288 × 384</td>
<td>-384 ... 384</td>
<td>0 ... 15</td>
</tr>
<tr>
<td>Test Scene</td>
<td>351 × 571</td>
<td>-571 ... 571</td>
<td>-90 ... -43</td>
</tr>
<tr>
<td>St. Martin</td>
<td>2184 × 1846</td>
<td>-1846 ... 1846</td>
<td>-35 ... 50</td>
</tr>
</tbody>
</table>

Table 1: Estimated search ranges on standard images

References


