A dynamic programming algorithm for perceptually consistent stereo

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

Accessed
November 21, 2017 10:51:50 PM EST

Citable Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:33957073

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)
A dynamic programming algorithm for perceptually-consistent stereo

Jialiang Wang
Daniel Glasner
and
Todd Zickler

TR-02-17

Computer Science Group
Harvard University
Cambridge, Massachusetts
A dynamic programming algorithm for perceptually-consistent stereo

Jialiang Wang  
Harvard University  
jialiangwang@h.harvard.edu

Daniel Glasner  
AiCure *  
dglasner@gmail.com

Todd Zickler  
Harvard University  
zickler@seas.harvard.edu

1. Introduction

This document provides details of the dynamic programming algorithm discussed in “Towards perceptually-consistent stereo: a scanline study” [3]. For the motivation of the algorithm, see that paper.

2. Objective

On a horizontal scanline of a rectified pair of stereo images, we assume a discrete domain for the visual field, indexed by \( n \in \{1 \ldots N\} \). For notational convenience, we choose a flipped coordinate system, with \( n = 1 \) the right-most pixel and \( n = N \) the left-most one. We define a small set of spatial basis functions \( \{B_b(n)\}_{b=1 \ldots M} \) and the disparity at pixel \( n \) as \( d(n) = \sum_{b=1}^{M} \ell(b)B_b(n) \) with shape coefficients \( \ell = \{\ell(b)\}_{b=1 \ldots M} \). The \( M \)-dimensional shape space is quantified, so there are \( L \) possible choices of vector \( \ell \). Assume we have matching cost \( C(n,\ell) \in [0,1] \) defined at all \((n,\ell)\). Suppose \( \{s_i\} \) are breakpoint locations between 1 and \( N \), and with \( s_{i-1} < s_i \). We define a half-occulsion function \( k(s_i,\ell_{s_i},\ell_{s_i+1}) \) that dictates the spatial location of a half-occulsion boundary according to the following rule (see the blue dashed line in Figure 1 for an example):

\[
k(s_i,\ell_{s_i},\ell_{s_i+1}) =
\begin{cases}
s_i + d_{\ell_{s_i}}(s_i) - d_{\ell_{s_i+1}}(s_i), & \text{if } d_{\ell_{s_i+1}}(s_i) < d_{\ell_{s_i}}(s_i) \\
\text{s_i}, & \text{otherwise}.
\end{cases}
\]

We want to simultaneously find the number of breakpoints, the breakpoint locations \( \{s_i\} \), and the per-segment shape coefficients \( \{\ell_{s_i}\} \) between each sequential pair of breakpoints. Mathematically, we want to find the global minimum of the objective function:

\[
\sum_i \rho_c(s_{i-1},s_i,\ell_{s_{i-1}},\ell_{s_i}) = \\
\sum_i \rho_c(s_{i-1},s_i,\ell_{s_{i-1}},\ell_{s_i}) + \lambda_1 \rho_b(s_{i-1},\ell_{s_{i-1}},\ell_{s_i}) + \lambda_2
\]

where \( \rho_c(s_{i-1},s_i,\ell_{s_{i-1}},\ell_{s_i}) \) is the matching cost of segment \( [s_{i-1},s_i] \), \( \rho_b(s_{i-1},\ell_{s_{i-1}},\ell_{s_i}) \) is a boundary cost defined at breakpoint \( s_{i-1} \), and \( \lambda_1,\lambda_2 \) are constants.

The matching cost of a segment is the sum of the matching cost over the unoccluded portion of the segment:

\[
\rho_c(s_{i-1},s_i,\ell_{s_{i-1}},\ell_{s_i}) = \sum_{n=k(s_{i-1},\ell_{s_{i-1}},\ell_{s_i})}^{s_i-1} C(n,\ell_{s_i}).
\]

3. Optimization by dynamic programming

Let \( opt(u,\ell_u) \) denote the scalar cost associated with the optimal disparity sub-profile over interval \([1,u]\) with the constraint that the final segment, the one containing \( u \), has shape \( \ell_u \). For the first pixel we have \( opt(1,\ell) = C(1,\ell) + \lambda_2 \) for all \( \ell \), and from there, we can visit the remaining pixels \( u \in \{2 \ldots N\} \) in sequence, recursively computing the \( L \) values of \( opt(u,\cdot) \) at each pixel \( u \). For each pair \((u,\ell_u)\), we search for the optimal location of the previous breakpoint \( v \). We set \( v = 1 \) if having only one segment is optimal for interval \([1,u]\). If \( v \neq 1 \), \( v \) will necessarily be a breakpoint between the shape of the final segment \( \ell_u \) and a different shape, say \( \ell_v \), of the segment before it (see Figure 1). Thus, we can write the recursion as

\[
\text{opt}(u,\ell_u) = \min_{v} (\rho(v,u,\ell_v,\ell_u) + \text{opt}(v,\ell_v)),
\]

where \( \Gamma \) is a subset of pairs \((v,\ell_v)\) within \([1,u-1] \times [1 \ldots L]\) that satisfy several constraints (to be discussed in Section 4). To be able to build each valid subset \( \Gamma \) during recursion, we also maintain a record of the optimal beginnings (the minimizers of Equation 2):

\[
\text{arg}(u,\ell_u) = \arg \min_{v} (\rho(v,u,\ell_v,\ell_u) + \text{opt}(v,\ell_v)).
\]

This data structure has size \( N \times L \times 2 \). In our notation, \( \text{arg}(u,\ell_u) = \text{arg}(u,\ell_u) \) means that among all possible sub-profiles defined on interval \([1,u]\) with final segment shape \( \ell_u \), the one with the lowest cost (and cost equal to \( \text{opt}(u,\ell_u) \)) is smooth over interval \([v,u]\) and has a breakpoint at \( v \) marking a transition to shape \( \ell_v \).

*Most work was done when Daniel Glasner was at Harvard University.
combinations are excluded from consideration by imposing several constraints. First, we enforce a half-occlusion constraint. Second, to ensure geometric ordering, we disallow any profile for which the previously-computed optimal sub-profile to \((v, \ell_v)\) cannot be in the green region per half-occlusion constraint. Second, to ensure geometric ordering, we disallow \((v, \ell_v)\) for which the previously-computed optimal sub-profile to \((v, \ell_v)\) (red line) has a leftmost breakpoint \(x\) such that \((x, d_{\ell_v}(x))\) is in the orange region (Eqn. 4).

Optimal sub-profile to \((u, \ell_u)\)

Algorithm 1: Find optimal disparity profile (from [3])

1: for all \(\ell \in [1, L]\) do
2: \(\text{opt}(1, \ell) \leftarrow C(1, \ell) + \lambda_2\)
3: \(\text{arg}(1, \ell) \leftarrow (1, \ell)\)
4: end for
5: for \(u \leftarrow 2\) to \(N\) do
6: for \(\ell_u \leftarrow 1\) to \(L\) do
7: \(\Gamma \leftarrow \text{BUILDVALIDSUBSET}(u, \ell_u, \text{arg}(u, \ell_u))\)
8: \(\text{opt}(u, \ell_u) \leftarrow \min \{ \rho(v, u, \ell_v, \ell_u) + \text{opt}(v, \ell_v) \}\)
9: \(\text{arg}(u, \ell_u) \leftarrow \arg \min \{ \rho(v, u, \ell_v, \ell_u) + \text{opt}(v, \ell_v) \}\)
10: end for
11: end for
12: \(Z, u \leftarrow N\) \(\triangleright\) initialize trace back
13: \(\Theta, \ell_u \leftarrow \arg \min \{ \text{opt}(N, \ell) \}\)
14: while \(u > 1\) do
15: \((u, \ell_u) \leftarrow \text{arg}(u, \ell_u)\)
16: APPEND\((Z, u)\)
17: APPEND\((\Theta, \ell_u)\)
18: end while

Once the recursion terminates at pixel \(N\), we trace the optimal profile by using the \(\text{arg}\) data structure to accumulate the profile’s breakpoints and shape-transitions, from the last one at pixel \(N\) to the first one at pixel 1. Algorithm 1 provides pseudo code (from [3]).

4. Constraints

To discuss the constraints, we first need to define two types of breakpoints: We call a breakpoint an \textit{occluding breakpoint} if \(d_{\ell_u}(v) < d_{\ell_v}(v)\); otherwise we call it a \textit{non-occluding} one. In Figure 1, the breakpoint at \(v\) is an occluding one while the one at \(x\) is not.

There are three constraints we enforce in our algorithm. We first introduce all three constraints in Section 4.1, and then discuss the limitations of our “ordering constraint” in Section 4.2.

4.1. Description of constraints

\textbf{Disparity bounds.} We specify a valid disparity range \([d_{\text{min}}, d_{\text{max}}]\), such that \(d_{\text{min}} < d_l(n) < d_{\text{max}}\). Following this constraint, some models \(l\) are only valid at a subset of the pixel locations.

\textbf{Half-occlusion constraint.} We enforce a half-occlusion constraint which imposes a hard lower bound \(K\) on the size of the visible portion of segments that include half-occlusion. This type of segment occurs after all occluding breakpoints. Mathematically, this constraint is \(u - k(v, \ell_v, \ell_u) \geq K\) if \(d_{\ell_u}(v) < d_{\ell_v}(v)\). For example, the green triangular region in Figure 1 is excluded from \(\Gamma\).

\textbf{Ordering constraint.} As in most dynamic programming approaches to scanline stereo, our algorithm is only made feasible when the output preserves geometric ordering [1], which, as shown in Figure 2, means that the ordering of matched points is preserved between the left and the right images. In our algorithm, we enforce this with a constraint that is sufficient but not necessary. At an occluding breakpoint, we require:
\[ v - x > d_{\ell_v}(x) - d_{\ell_u}(v) \text{ with } (x, \ell_x) = \text{arg}(v, \ell_v) \quad (4) \]

This constraint requires that the occluding segment, e.g., segment \([x, v]\) in Figure 1, is long enough to guarantee that geometric ordering is preserved. It can be visualized as a requirement on the second breakpoint \(x\) from \(u\), and the disparity \(d_{\ell_v}(x)\) at that point. For example, any \((v, \ell_v)\) pair whose previous optimal breakpoint (stored in \text{arg}) falls into the orange region in Figure 1 is excluded from \(\Gamma\).

Imposing this constraint is sufficient to guarantee geometric ordering. Figure 3 shows the depth planes corresponding to the disparity segments with shape coefficients \(\ell_u\) and \(\ell_v\) from Figure 1. \(V\) and \(X\) are the back projection of breakpoints \(v\) and \(x\) in the left camera. Let \(d_{\ell_u}(v)\), \(d_{\ell_v}(x)\) be the disparities of background pixel \(v\) and foreground pixel \(x\) respectively. We know that [2],

\[ d_{\ell_u}(v) = f \frac{w}{Z_v} \quad \text{and} \quad d_{\ell_v}(x) = f \frac{w}{Z_x}. \]

for focal length \(f\) and baseline \(w\).

If point \(V\) in the scene is not visible from the right camera, then geometric ordering is guaranteed. We define \(V'\) as the intersection of the ray through \(V\) and a line through \(X\) that is parallel to baseline \(w\). From similar triangles,

\[
\frac{|V'X|}{w} > \frac{Z_v - Z_x}{Z_v} \\
\iff \quad |V'X| > w \left(1 - \frac{Z_x}{Z_v}\right) = w \left(1 - \frac{d_{\ell_v}(v)}{d_{\ell_v}(x)}\right) \\
\]

Using similar triangles again, we have,

\[
\frac{v - x}{f} = \frac{|V'X|}{Z_x} \\
\iff \quad v - x = \frac{|V'X|}{Z_x} \frac{f}{Z_x} > w \left(1 - \frac{d_{\ell_v}(v)}{d_{\ell_v}(x)}\right) \frac{d_{\ell_v}(x)}{w} \\
> d_{\ell_v}(x) - d_{\ell_u}(v),
\]

of which is exactly the constraint of Equation 4.

### 4.2. Limitations of the ordering constraint

Our ordering constraint is sufficient but not necessary to guarantee geometric ordering. This means that there are some profiles that do not violate geometric ordering but could not be reconstructed by our algorithm. Figure 4 is an example. Figure 4(a) shows the scene geometry and 4(b) shows the disparity profile. Our constraint excludes this due to the short segments \([x, v]\), even though the true scene geometry does not violate geometric ordering.

Another example is Figure 5, which results from the fact that the optimization we do is over the previous breakpoint \(v\) instead of a joint optimization over the two previous breakpoints. In this case, \text{arg}(v, \ell_v)\) stores the black profile to \((v, \ell_v)\) since it has the lowest cost, and the blue one is not stored. Thus, at \((u, \ell_u)\), the black profile is excluded from \(\Gamma\) because it violates the ordering constraint. As a result, the algorithm selects the red profile which has a different previous breakpoint \((v', \ell_v')\) instead of the blue one, although the blue one has lower cost and preserves geometric ordering.

**Acknowledgements.** This work was funded by National Science Foundation awards IIS-1212928 and IIS-1618227.
Figure 4. An example that cannot be recovered by our algorithm even though it does not violate the geometric ordering. (a) the depth scene and (b) the corresponding disparity profile.

Figure 5. An example of the limitations of the ordering constraint. The red profile is selected instead of the blue one, even though neither one violates geometric ordering, and the blue one has lower cost.

References

