Efficient Stereo Matching using Histogram Aggregation with Multiple Slant Hypotheses

Michel Antunes and João P. Barreto

Institute of Systems and Robotics,
Faculty of Sciences and Technology,
University of Coimbra,
3030 Coimbra, Portugal
(michel,jpbar)@isr.uc.pt

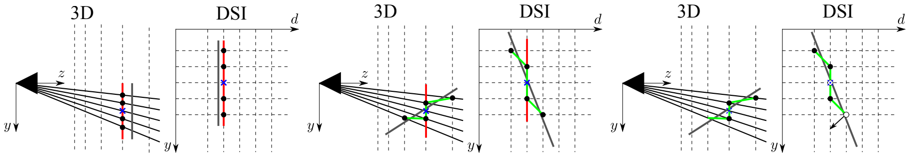
Abstract. This paper presents an enhancement to the recent framework of histogram aggregation [1], that enables to improve the matching accuracy while preserving a low computational complexity. The original algorithm uses a fronto-parallel support window for cost aggregation, which leads to inaccurate results in the presence of significant surface slant. We address the problem by considering a pre-defined set of discrete orientation hypotheses for the aggregation window. It is shown that a single orientation hypothesis in the Disparity Space Image is usually representative of a large interval of possible 3D slants, and that handling slant in the disparity space has the advantage of avoiding visibility issues. We also propose a fast recognition scheme in the Disparity Space Image volume for selecting the most likely orientation hypothesis for aggregation. The experiments clearly prove the effectiveness of the approach.

1 Introduction

Dense stereo matching consists in assigning to each pixel in one view the corresponding pixel in the other view [2]. This requires using a matching cost for comparing image pixel locations and quantifying their likelihood of being a correspondence. A particular stereo method can be said to be local or global, depending on the strategy that is used for obtaining the final disparity map [2]. This article focus exclusively in local methods, that aggregate the matching cost over a support region in the Disparity Space Image (DSI) [3], as a way to enforce spatially coherence and improve the final depth estimates.

It is well known that the aggregation window must be aligned with the surface of the pixel being analyzed in order to maximize the matching performance [4–8]. Their objective is invariably the estimation of the orientation of the 3D plane that approximates the surface region that is projected in the pixel, or group of pixels, under analysis. This usually involves the estimation of sub-pixel matches for each hypothesized planar region. Thus, these algorithms tend to be complex and time consuming.

This paper proposes a simple but effective approach for increasing the robustness to surface slant during stereo cost aggregation. Our strategy consists in



(a) FP Aggreg. (b) Aggreg.+Slant (c) HA+Slant

Fig. 1. Overview: (a) The fronto-parallel (FP) aggregation for a pixel (blue cross) is performed along the discrete disparity plane (red) closest to the true disparity value (grey). (b) Most stereo methods rely on FP aggregation windows (red), that completely disregard surface slant (grey). We propose an aggregation scheme that accommodates surface slant by considering a pre-defined set of possible orientations for the support window. Slant hypotheses involving sub-pixel disparities are efficiently approximated by discrete directions of aggregation in the DSI (green). (c) Histogram aggregation (HA)[1] is more efficient than standard cost aggregation. In this case, the selected aggregation directions in neighboring pixels are used to estimate the disparity of the reference pixel. Note that, if a neighboring pixel is assigned with a different aggregation directions (white), then its contribution will fall in a different bin of the histogram.

avoiding the errors in pixel matching caused by surface slant without having to explicitly infer the normal orientation of the original 3D surfaces in the scene. We explore the DSI and propose the discretization of slanted aggregation windows as it is done for disparity vs. 3D depth (Fig. 1). It is demonstrated that an initial small set of aggregation orientations improves the stereo aggregation even for surfaces contained in the scene that are only approximated by those orientations. In order to improve the efficiency of the proposed stereo aggregation, we use a simple and fast recognition scheme for selecting the most appropriate aggregation orientation $\boldsymbol{\alpha}$ for each pixel-disparity pair $(\mathbf{p}, d)$. The histogram aggregation technique [1] is used, which is conceptually different from the standard cost aggregation in those cases where only one aggregation orientation is considered for each pixel-disparity pair. In a certain sense, we enhance the histogram aggregation technique proposed in [1] with slant information, boosting the accuracy at the expense of a small computational overhead. The experimental results in terms of integer pixel disparity accuracy are very close to [8] (highly ranked in Middleburry), but with several orders of magnitude less computation time.

2 Related work

In recent years, three main research topics concerning cost aggregation were addressed: (i) handling depth discontinuities [9]; (ii) reducing the computational complexity [1]; and (iii) handling surface slant [4–8]. The first issue is solved by the adaptive weight strategy of Yoon and Kweon [9], while the second was recently address in [1] by eliminating redundant computations. We focus on the 3D slant issue and, to keep computation tractable, on the second by following similar sampling schemes as [1]. The stereo methods that take the surface slant

into account can be roughly divided into four distinct categories: (1) the methods that use fronto-parallel stereo for the initialization e.g. [6]; (2) the methods whose objective is to assign a 3D plane to each image pixel from a pre-defined set of plane hypotheses e.g. [4]; (3) approaches that fit 3D planes using image segmentation e.g. [5]; and (4) the algorithm recently proposed by Bleyer et al. in [8] that estimates a 3D plane at each pixel onto which the support region is projected.
Our new stereo aggregation strategy is more closely related to the second group, however with two conceptual differences: (i) we work in the DSI without the ambition of correctly estimate the 3D slant. In practical terms, we avoid interpolation issues, at the expense of no explicit sub-pixel matching accuracy; and (ii) we propose the quantization of the 3D plane space, and by doing it in the DSI, we are able to cover the slant space with less plane samples, as well as to implicitly handle visibility/impossible configuration issues.

3 Formulation of local stereo using histogram aggregation

Let's consider the general stereo problem, where the goal is to assign to each pixel $\mathbf{p} = (x_p, y_p)$ in the left image $\mathtt{I}$ a disparity d from a pre-defined set of discrete values $\mathbf{D} = [0, ..., D-1]$. This assignment implicitly associates $\mathbf{p}$ with the pixel $\mathbf{p'} = (x_p - d, y_p)$ in the right image $\mathtt{I}'$. As in [8], we choose as pixel matching cost the so-called truncated color and gradient differences (TD)

$$c(\mathbf{p}, d) = (1-\beta)\max(\tau_{col} - ||\mathtt{I}_\mathbf{p} - \mathtt{I}'_{\mathbf{p'}}||, 0) + \beta \max(\tau_{grad} - ||\Delta\mathtt{I}_\mathbf{p} - \Delta\mathtt{I}'_{\mathbf{p'}}||, 0),$$

where $||\mathtt{I}_\mathbf{p} - \mathtt{I}'_{\mathbf{p'}}||$ corresponds to the L_2-distance of the RGB colors of pixels $\mathbf{p}$ and $\mathbf{p'}$, $||\Delta\mathtt{I}_\mathbf{p} - \Delta\mathtt{I}'_{\mathbf{p'}}||$ is the L_2-distance of the gray-value gradients, the parameter β balances the influence of color and gradient, and τ_{col} and τ_{grad} serve to truncate the cost in order to improve robustness near discontinuities.

The cost aggregation is defined as a joint histogram voting [1]:

$$\mathsf{C}(\mathbf{p}, d) = \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \omega(\mathbf{p}, \mathbf{q}) c(\mathbf{q}, d),$$

with C being the aggregated DSI, $\mathcal{N}(\mathbf{p})$ denoting the pixel neighborhood of $\mathbf{p}$ defined by the size B of the aggregation window, and $\omega(\mathbf{p}, \mathbf{q})$ corresponding to the adaptive support weighting function proposed in [9]. This function is defined as:

$$\omega(\mathbf{p}, \mathbf{q}) = \exp\left(-\frac{\sqrt{(\mathtt{I}_\mathbf{p} - \mathtt{I}_\mathbf{q})^2}}{\delta_{col}} - \frac{\sqrt{(\mathbf{p} - \mathbf{q})^2}}{\delta_{sp}}\right),$$

with δ_{col} and δ_{sp} being constant parameters.

The complexity of histogram-based cost aggregation can be substantially reduced by applying the following sampling strategies [1]:

1. Selection of disparity hypotheses: The idea is to independently select for each pixel $\mathbf{p}$ a small subset of disparity hypotheses that have better support. This is accomplished by using a small square window for filtering the cost $c(\mathbf{p}, d)$ along the disparity dimension, and then choosing the $P\%$ local maxima of the obtained 1-D signal. The result is a subset $\mathbf{D}^{\mathbf{p}}_{P} = \{P\% \text{ best disparities of } \mathbf{p}\}$ comprising the disparities to be further considered in subsequent steps.

2. Spatial Sampling: The image grid is sampled by a factor of $S \times S$ that enables to further reduce the complexity of the stereo aggregation.

Taking into account the sampling strategies, the aggregated cost C can be re-written as:

$$\mathsf{C}_{P,S}(\mathbf{p}, d) = \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \omega(\mathbf{p}, \mathbf{q}) c(\mathbf{q}, d) o_P(\mathbf{q}, d) s_S(\mathbf{q}) \tag{1}$$

where

$$o_P(\mathbf{q}, d) = \begin{cases} 1 & \text{if } d \in \mathbf{D}^{\mathbf{q}}_{P} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad s_S(\mathbf{q}) = \begin{cases} 1 & \text{if } \mathbf{q}\%S = 0 \\ 0 & \text{otherwise} \end{cases}$$

4 Aggregation with different window orientations $\boldsymbol{\alpha}$

Let's assume a rectified stereo setup, with a relative camera translation of $\mathbf{t} = (b, 0, 0)^{\mathsf{T}}$, and a generic scene point $\mathbf{P} = (X_p, Y_p, Z_p)^{\mathsf{T}}$ that lies in a surface with normal $\mathbf{n} = (n_x, n_y, n_z)^{\mathsf{T}}$. This surface can be locally approximated by a plane that defines an homography H mapping points $\mathbf{p}$ in the left view into points $\mathbf{p}'$ on the right view [10]:

$$\mathsf{H} = \left(\mathsf{I}_{3\times 3} + \frac{\mathbf{t}\,\mathbf{n}^{\mathsf{T}}}{h}\right) = \begin{pmatrix} 1 + b\frac{n_x}{h} & b\frac{n_y}{h} \; b\frac{n_z}{h} \\ \mathsf{0}_{2\times 1} & \mathsf{I}_{2\times 2} \end{pmatrix} , \tag{2}$$

with $h = \mathbf{P} \cdot \mathbf{n}$. If $\mathbf{p}$ and $\mathbf{p}'$ are the images of $\mathbf{P}$, then $\mathbf{p}' = \mathsf{H}\mathbf{p}$ and the stereo disparity is

$$d_p = x_p - x_{p'} = -\frac{n_x b}{h} x_p - \frac{n_y b}{h} y_p + \frac{n_x b x_p + n_y b y_p + h d_p}{h} \tag{3}$$

with $y_p = y_{p'}$ indicating the image row coordinate. Consider now a generic image point (x, y) in the neighborhood $\mathcal{N}(\mathbf{p})$ that is still the projection of the plane surface. By applying the homography of Eq.2 comes that the disparity d of this neighboring point differs from d_p by

$$\Delta_d = d - d_p = \alpha_x (x - x_p) + \alpha_y (y - y_p) \ , \ \alpha_x = -\frac{n_x b}{h} \ , \ \alpha_y = -\frac{n_y b}{h} . \tag{4}$$

Eq.4 shows that the orientation of the aggregation window in the DSI must be in accordance with the 3D surface slant. A standard window along a constant disparity direction cannot account for the variation Δ_d in the neighborhood of the pixel under analysis. The ideal window must be slanted around a vertical axis by an angle with tangent α_x, and a horizontal axis by an angle with tangent α_y. Henceforth, we will parametrize the orientation of the aggregation window by

$\boldsymbol{\alpha}=(\alpha_x, \alpha_y)$, with $\boldsymbol{\alpha}=(0,\,0)$ being the standard situation of aggregation along a constant disparity. Remark that windows in the DSI with the same orientation correspond to different 3D surface slants, depending on the coordinates of $\mathbf{P}$, the focal length f and the baseline b.

Most existing stereo methods that handle surface slant explicitly estimate a 3D plane for each pixel onto which the neighborhood is projected (see Sec. 2). In order to accomplish this, they need to analyze for each 3D point $\mathbf{P}$ being considered, if the hypothesized surface is visible in both cameras. It can be shown that this visibility issue is implicitly solved for each pair $(\mathbf{p}, d)$ in the DSI using the parameterization $\boldsymbol{\alpha}$, and by limiting α_x and α_y to the ranges $\alpha_x \in [-1, 1[$ and $\alpha_y \in [-1, 1]$, respectively. Moreover, and since our objective is not to accurately estimate the surface slant for each pixel, we consider horizontal and vertical surface slants separately. Following [4], by horizontal slant we mean the surfaces on which the disparity changes as we move along the x-axis, which is related to the aggregation orientation α_x. Similarly, the disparity on a vertical slanted surface varies as we move along the y-axis, corresponding to the orientation α_y.

The DSI is inherently a discrete 3D space so that considering continuous window orientations in the DSI requires the interpolation of the cost volume. This provides depth estimations at a sub-pixel accuracy level, however with the drawback of increased computational cost. We avoid the interpolation issues and propose a simple scheme for voxelizing slanted windows in the DSI, where the incremental disparity between successive pixels is given by

$$\varDelta_d = (int)(\boldsymbol{\alpha} \cdot (\mathbf{p} - \mathbf{q})^{\mathsf{T}}). \tag{5}$$

We assume the working ranges $\alpha_x \in [-1, 1[$ and $\alpha_y \in [-1, 1]$ and consider vertical and horizontal surface slants separately. Following this, it can be verified that using the voxelization proposed in Eq.5, there are B–1 distinguishable horizontal and B distinguishable vertical aggregation patterns for a window of size B.

5 Histogram aggregation with multi-slant hypotheses

This section describes a new scheme for cost histogram aggregation that takes into account the surface slant. This is achieved by considering a pre-defined set of N window orientations in the DSI. In addition, we propose a simple recognition approach for selecting the best aggregation direction for each pixel, and discuss the differences between using standard and histogram aggregations in conjunction with orientation selection.

5.1 Cost aggregation in the $(\mathbf{p}, \boldsymbol{d}, \boldsymbol{\alpha})$ domain

In order to accommodate surface slant in the framework of histogram aggregation, we reformulate the function of Eq. 1 to consider an additional dimension $\boldsymbol{\alpha}=(\alpha_x, \alpha_y)$ that accounts for the orientation of the support window:

$$\mathsf{C}_{r,P,S}(\mathbf{p}, d, \boldsymbol{\alpha}) = \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} \omega(\mathbf{p}, \mathbf{q}) c(\mathbf{q}, d + \varDelta_d) h_{r,P}(\mathbf{q}, d + \varDelta_d, \boldsymbol{\alpha}) s_S(\mathbf{q}), \tag{6}$$

with Δ_d, that depends on $\boldsymbol{\alpha}$, being given in Eq. 5. The look-up table o_P for the disparity selection is now replaced by $h_{r,P}$ that enables selecting the aggregation direction in addition to disparity selection. Remark that the histogram voting is only performed if $(d+\Delta_d) \in \mathbf{D}$.

5.2 Sampling the space of the aggregation orientations $\boldsymbol{\alpha}$

We propose a simple and fast recognition approach for an efficient implementation of the histogram aggregation formulated in Eq. 6. The objective is to select for each pixel $\mathbf{p}$ and disparity d, the best aggregation orientation among the hypotheses in the configuration $\mathbf{A}_N$ under consideration. The recognition is accomplished by correlating the cost $c(\mathbf{p}, d)$ with the slanted window of size R defined by the $\boldsymbol{\alpha}$ hypothesis. It is important to distinguish between the size B of the aggregation window and the size R of the recognition window. This operation is carried whenever the parameter r is set (Eq.6) and is formalized by the following scoring function

$$\rho(\mathbf{p}, d, \boldsymbol{\alpha}) = \frac{\sum\limits_{\mathbf{q} \in \mathcal{N}_R(\mathbf{p})} c(\mathbf{q}, d + \boldsymbol{\alpha} \cdot (\mathbf{p} - \mathbf{q})^{\mathsf{T}})}{\sum\limits_{\mathbf{q} \in \mathcal{N}_R(\mathbf{p})} \left(d + \boldsymbol{\alpha} \cdot (\mathbf{p} - \mathbf{q})^{\mathsf{T}}\right) \in \mathbf{D}}, \tag{7}$$

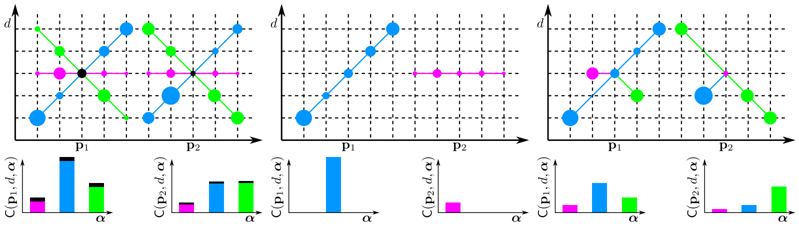
For each pixel $\mathbf{p}$ and disparity d, the orientation $\boldsymbol{\alpha}$ with highest score defines the set $\mathbf{A}_r^{\mathbf{p},d} = \{\text{best } \boldsymbol{\alpha} \text{ for } (\mathbf{p}, d)\}$. In the case the parameter r is zero, then $\mathbf{A}_r^{\mathbf{p},d} = \mathbf{A}_N$ and all orientations are considered for the aggregation. The new look-up table h is defined as:

$$h_{r,P}(\mathbf{p}, d, \boldsymbol{\alpha}) = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} \in A_r^{\mathbf{p},d} \wedge d \in D_P^{\mathbf{p}}. \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Remark that the selection of P percentage of the most likely disparities d for each pixel $\mathbf{p}$ ($D_P^{\mathbf{p}}$) only makes sense in conjunction with orientation selection ($r = 1$). In this case, the scoring function ρ is the new metric for choosing the best disparities.

5.3 Standard aggregation vs. Histogram aggregation for $r = 1$

There is a difference between standard [9] and histogram aggregation [1] when the aggregation orientation is pre-selected ($r = 1$). As shown in Fig. 2, for $r = 0$ both approaches obtain the same aggregated cost $\mathsf{C}(\mathbf{p}, d, \boldsymbol{\alpha})$, corresponding to the sum of all neighboring costs along the N orientations $\boldsymbol{\alpha}$. However, if the recognition parameter is set to $r = 1$, then for standard aggregation the cost $\mathsf{C}(\mathbf{p}, d)$ is obtained by aggregating the neighborhood of $\mathbf{p}$ along the assigned orientation $\boldsymbol{\alpha}$ for $(\mathbf{p}, d)$. In histogram aggregation, each neighbor votes along the orientation to which it was assigned. This means that the N bins $\mathsf{C}(\mathbf{p}, d, \boldsymbol{\alpha})$ of $(\mathbf{p}, d)$ are voted by the neighboring pixels for which the aggregation direction $\boldsymbol{\alpha}$ intersects $(\mathbf{p}, d)$.



(a) $r = 0$ (b) $r = 1$ Standard (c) $r = 1$ Histogram

Fig. 2. Differences between standard and histogram aggregation using 3 aggregation orientations (magenta, blue and green). We show two examples for two different reference points $\mathbf{p}_1$ and $\mathbf{p}_2$ (black). (b,c) Blue slant assigned to $\mathbf{p}_1$ and magenta to $\mathbf{p}_2$.

Table 1. We use 4 aggregation configurations.

	FP	3 Vert.			5 Vert.+2 Hor.							7 Vert.+4 Hor.										
	$\mathbf{A}_1$	$\mathbf{A}_3(R=5)$			$\mathbf{A}_7(R=5)$							$\mathbf{A}_{11}(R=11)$										
α_x	0	0	0	0	0	0	0	0	0	−.5	.5	0	0	0	0	0	0	0	−.5	−.2	.2	.5
α_y	0	−1	0	1	−1	−.5	0	.5	1	0	0	−1	−.5	−.2	0	.2	.5	1	0	0	0	0

6 Experimental Results

In this section, we will describe the methodology for the experimental evaluation, study the performance of the proposed stereo aggregation using different sets of aggregation orientations $\mathbf{A}_N$, and compare the proposed method against one state-of-art method.

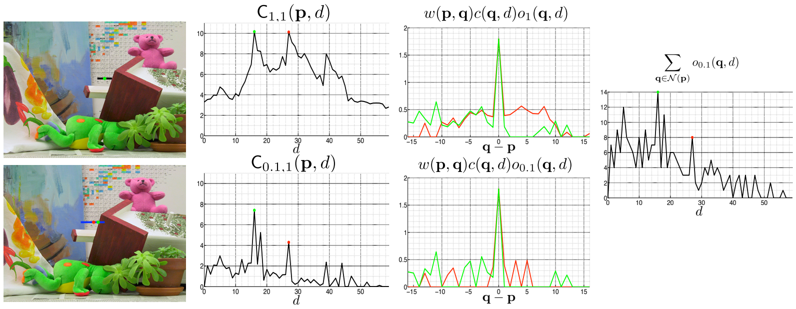
Following the standard evaluation, the disparity maps are scored by counting the number of *nonocc* (pixels in non-occluded regions), *all* (all pixels), and *disc* (visible pixels near occluded regions) pixels that differ in more than one pixel from the ground truth. The experiments are performed on the Middlebury datasets [2, 11, 12]. Concerning the possible orientations for aggregation, we only assume vertical or horizontal slants for the windows. The Tab. 1 specifies the configurations $\mathbf{A}_N$ to be considered, indicating the $\boldsymbol{\alpha} = (\alpha_x, \alpha_y)$ values that define the orientations of the N window hypotheses. We will show that in general our small discrete set of orientations $\boldsymbol{\alpha}$ will be able to approximate different *continuous* 3D slants in the scene. The experiments will compare the results for the 4 configurations $\mathbf{A}_N$ of Tab. 1 in an attempt to assess the influence of the number of considered direction hypotheses for aggregation. Please note that for $\mathbf{A}_3$ and $\mathbf{A}_7$ a window of $R = 5$ is sufficient for the recognition step; while for $\mathbf{A}_{11}$ we must use $R = 11$ to discriminate between the different aggregation orientations.

6.1 Comparison of different aggregation configurations $\mathbf{A}_N$

We show in Tab. 2 the results of the disparity labeling for nonocc pixels in 4 stereo pairs. As expected from [1], the selection of the best disparities improves the disparity estimation in most cases. The authors of [1] justify this as *unnecessary disparity candidates contaminate the aggregation process*. We reinforce this

Table 2. Comparison of 4 configurations $\mathbf{A}_N$(Tab.1). No spatial sampling is applied ($S = 1$). The under-script values correspond to conventional aggregation (Sec.5.3).

	Stereo Pair	Teddy			Cones			Wood1		
	P	0.1	1		0.1	1		0.1	1	
	r	1	1	0	1	1	0	1	1	0
FP	$\mathbf{A}_1$	5.29	6.04		2.95	3.4		8.18	10.6	
SLANT	$\mathbf{A}_3$	**2.84**	$\mathbf{3.05}_{4.21}$	**3.32**	**2.71**	$\mathbf{2.98}_{2.94}$	3.48	4.19	$12.6_{3.52}$	7.73
	$\mathbf{A}_7$	4.93	$8.41_{6.03}$	3.88	2.93	$3.74_{4.02}$	4.09	6	$18.7_{3.60}$	7.54
	$\mathbf{A}_{11}$	5.89	$13.3_{4.87}$	3.78	3.4	$9.85_{3.15}$	**3.09**	**1.78**	$\mathbf{4.43}_{1.89}$	**2.06**



(a) Input Images (b) DSI along **p** (c) Adapt. Weight (d) d count

Fig. 3. Disparity selection before histogram aggregation decreases the errors in ambiguous regions and near discontinuities. (a) Top: I with aggregation window (black) centered in the pixel **p** under analysis (green). Bottom: I′ with matching candidates (blue); green and red points are respectively correct and false matches. (b) Aggregated DSI results for pixel **p**. The disparity selection (bottom) avoids the existence of multiple maxima (top) that create ambiguity. (c) Adaptive aggregation for the neighboring pixels of **p** [9]. Green corresponds to the correct disparity, while red corresponds to a false match. If no disparity selection is used (top), the two cost aggregation results will be similar because of the low texture of the roof in the case of the correct disparity. The disparity selection (bottom) removes for the false match non-discriminative contributions caused by the textured wall in background. (d) The correct disparity for **p** is voted more times in $D_P^{\mathbf{q}}$.

observation using Fig. 3. The pixels in ambiguous regions vote in the aggregation histogram in a chaotic manner. However, the main point is that even in ambiguous image regions the correct disparity for **p** appears more times as local maxima in the neighborhood $\mathcal{N}_p$ than other disparities, so that the disparity selection step leads to an improved disparity voting.

Effect of considering various aggregation orientations ($r = 0$) Considering various aggregation orientations improves the accuracy in the majority of the cases when compared to fronto-parallel aggregation. This does not happen for one case ($\mathbf{A}_7$) in the cones dataset, however this scene does not contain any slanted planes, and more aggregation orientations tend to amplify the chaotic voting referred previously.

Table 3. Evaluation in Middlebury (Consulted in 11/2012.). We set $(P=0.1, r=1)$.

	Algorithm	Rank	Tsukuba			Venus			Teddy			Cones			Runtime
			nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	(Tsukuba)
	PatchMatch[8]	12	2.09_{66}	2.33_{52}	9.31_{63}	0.21_{22}	0.39_{18}	2.62_{31}	2.99_{1}	8.16_{8}	9.62_{2}	2.47_{5}	7.8_{9}	7.11_{7}	$\approx 60s$
$S=1$	HistAggr+TD	25	2.44_{72}	2.69_{56}	9.17_{62}	0.25_{30}	0.34_{16}	3.24_{38}	5.29_{16}	10.7_{22}	14_{16}	2.95_{24}	8.59_{24}	8.24_{25}	$16.9s$
	HistAggr+TD+Slant	17	2.38_{70}	2.62_{56}	9.33_{64}	0.26_{32}	0.36_{17}	3.32_{41}	2.84_{1}	8.19_{9}	8.51_{1}	2.71_{13}	8.16_{16}	7.52_{13}	$18s$
$S=3$	HistAggr+TD	30	2.27_{70}	2.52_{55}	9.14_{62}	0.24_{28}	0.31_{13}	2.92_{35}	5.90_{19}	11.6_{31}	15.4_{22}	3.16_{29}	8.81_{30}	8.73_{34}	$1.7s$
	HistAggr+TD+Slant	**22**	$\mathbf{2.25}_{68}$	$\mathbf{2.50}_{55}$	$\mathbf{9.77}_{68}$	$\mathbf{0.29}_{34}$	$\mathbf{0.37}_{17}$	$\mathbf{3.30}_{41}$	$\mathbf{3.44}_{3}$	$\mathbf{8.82}_{13}$	$\mathbf{9.77}_{4}$	$\mathbf{2.90}_{20}$	$\mathbf{8.40}_{20}$	$\mathbf{7.97}_{20}$	**2s**

Effect of selecting one slant hypothesis ($r = 1$) There are two different effects that must be accounted (Tab. 2): (i) the effectiveness of the recognition scheme in selecting the most suitable orientation hypothesis $\boldsymbol{\alpha}$, and (ii) the effect of the histogram aggregation. It can be observed that the results tend to be significantly worse than for $(r = 0, P = 1)$. This is not because of the slant selection process, but rather because of the fact that histogram aggregation is not effective without disparity sampling. We show in under-script the results when there is aggregation orientation selection, but the aggregation is performed in the standard manner (see Sec. 5.3). The accuracy degrades slightly but doubts concerning the effectivness of the recognition can be discarded. Finally, and as can be seen in column $(r=0, P=0.1)$, the histogram aggregation is effective if we use both disparity sampling and slant selection.

There are two take home messages considering histogram aggregation taking into account surface slant: (1) The first is that slant selection in histogram aggregation works always well if the surface slants contained in the scene are well approximated by the hypothesis considered in $\mathbf{A}_N$. Otherwise, the decision process can assign different values $\boldsymbol{\alpha}$ to points on the same 3D surface that are equally well approximated by the discrete aggregation directions. This creates contradictory contributions in the histogram voting for neighboring pixels, enhancing the ambiguity (Fig. 3); and (2) The second observation is that the previous effect can be compensated by pursuing both slant selection and disparity sampling. The disparity sampling discards the contributions of neighbors of (x, d) for which the decision of slant can be equally fitted by more than one hypothesis, so that their votes are diluted in the histogram voting.

6.2 Evaluation in Middlebury

We compare the proposed aggregation with PatchMatch[8] as being one of the most accurate local algorithms that take into account the surface slant. The results are presented in Tab. 3. HistAggr+TD corresponds to the fronto-parallel aggregation ($\mathbf{A}_1$), whereas HistAggr +TD+Slant takes into account 3 aggregation orientations ($\mathbf{A}_3$). Our algorithm combines the advantages of both, the accuracy of PatchMatch by considering surface slant hypotheses, and the speed of the histogram aggregation technique. We dramatically improve with respect to fronto-parallel HistAggreg+TD at the expense of a computational overhead of $15-20\%$. We take the first position in the ranking for the Teddy stereo pair, which is more relevant since it is the only one containing considerable slant. This is achieved with less than 1/3 of the runtime of PatchMatch. The spatial sampling $S = 3$ is just slightly more inaccurate but with a speedup of 30×. As finally remark, we propose to use **HistAggr+TD+Slant** with $\mathbf{S = 3}$, being the best compromise between accuracy and runtime.

7 Conclusions

This paper presented a new histogram aggregation framework that accounts for surface slant. The strategy consisted in choosing the most suitable aggregation direction within a pre-defined set of discrete hypotheses. The experimental results were highly ranked in the Middleburry benchmark. The approach is able to combine high matching accuracy with small computational overhead when compared to [1]. On the other hand, we converge to the accuracy of PatchMatch [8], but with much less computation time.

Acknowledgements

The authors acknowledge the Portuguese Science Foundation (FCT) for the generous funding through grants SFRH/BD/47488/2008 and PTDC/EIA-CCO /109120/2008.

References

1. Min, D., Lu, J., Do, M.N.: A revisit to cost aggregation in stereo matching: How far can we reduce its computational redudancy? ICCV (2011)
2. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV (2001)
3. Szeliski, R., Scharstein, D.: Sampling the disparity space image. PAMI (2004)
4. Ogale, A.S., Aloimonos, Y.: Stereo correspondence with slanted surfaces: Critical implications of horizontal slant. In: CVPR. (2004)
5. Klaus, A., Sormann, M., Karner, K.: Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: ICPR. (2006)
6. Zhang, Y., Gong, M., Yang, Y.H.: Local stereo matching with 3d adaptive cost aggregation for slanted surface modeling and sub-pixel accuracy. In: ICPR. (2008)
7. Bleyer, M., Rother, C., Kohli, P.: Surface stereo with soft segmentation. In: CVPR. (2010)
8. Bleyer, M., Rhemann, C., Rother, C.: Patchmatch stereo - stereo matching with slanted support windows. In: BMVC. (2011)
9. Yoon, K., Kweon, I.S.: Adaptive support-weight approach for correspondence search. PAMI (2006)
10. Ma, Y., Soatto, S., Kosecka, J., Sastry, S.S.: An Invitation to 3-D Vision: From Images to Geometric Models. SpringerVerlag (2003)
11. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: CVPR. (2007)
12. Hirschmuller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: CVPR. (2007)