A *locally global* approach to stereo correspondence

Stefano Mattoccia
Department of Electronics Computer Science and Systems (DEIS)
Advanced Research Center on Electronic Systems (ARCES)
University of Bologna, Viale Risorgimento 2, 40136 - Bologna, Italy
stefano.mattoccia@unibo.it

Abstract

A novel approach to deal with the stereo correspondence problem induced by the implicit assumptions made by cost aggregation (CA) strategies is proposed. CA relies on the implicit assumption that disparity varies smoothly within neighboring points except at depth discontinuities and state-of-the-art CA strategies adapt their support to image content by classifying each pixel based on geometric and photometric constraints. Our proposal explicitly models this behavior from a different perspective, by gathering for each point, multiple assumptions that locally would be made by a hypothetical variable CA strategy. This framework enables to derive a function that locally captures the plausibility of the underlying geometric and photometric constraints independently enforced by supports of neighboring points. Experimental results confirm the effectiveness of our proposal.

1. Introduction

Dense stereo vision algorithms aim at inferring the disparity field of a scene acquired by two or more cameras. Determining correspondences in two or more images (*i.e.* the *correspondence problem*) is at the core of each stereo algorithm. Consequently, this topic has been extensively researched by the computer vision community (see [11] for a review). According to [11] most dense stereo algorithms perform four steps: (1) *costs computation*, (2) *costs aggregation* (CA), (3) *disparity optimization* and (4) *disparity refinement* aimed at detecting/replacing wrong disparity assignments. Stereo algorithms are classified [11] in *local approaches* (typically performing steps 1,2 (and 4)) and *global* approaches (typically performing steps 1,3 (and 4)). The assumption made by both categories is that the scene is piecewise smooth. Local approaches implicitly model this assumption [11, 5, 14] aggregating costs of neighboring points within a *support* window (also referred to as *kernel* or *correlation* window) while global approaches explicitly model this assumption [11] by performing a disparity optimization on the entire stereo pair. Local algorithms based on CA strategy are typically faster than global approaches and more suited (mainly due to limited memory requirements) for hardware/real-time implementation. However, although state-of-the-art local approaches based on variable CA strategies achieved excellent results [11, 5, 14], in most cases they are outperformed in terms of the accurateness by global approaches. Nevertheless, it is noteworthy that several global approaches deploy CA (*e.g.* the three current top performing algorithms on the Middlebury web site [10]) to improve their effectiveness. Our proposal was inspired by the observation that state-of-the-art CA strategies, although capable of obtaining excellent results [5, 14] by adapting the support to image content, process points independently without taking into account mutual dependencies among the neighboring points.

In this paper we provide a methodology for tackling the correspondence problem from a different point of view compared to the known approaches. We propose a framework that explicitly models the mutual relationships among neighboring points deriving a point based function that locally captures the underlying geometric and photometric structure of the scene. As our methodology *locally* performs a sort of *global* reasoning restricted to the neighboring points but does not perform any disparity optimization amenable to conventional global approaches, our proposal is potentially suited for hardware/real-time implementation.

2. Related work

Stereo vision is a very broad topic which has been extensively surveyed by Scharstein and Szeliski in [11]. Although our approach does not fit perfectly within the class of local approaches defined in [11] it was inspired by recent advance in the area of variable CA strategies that were recently analyzed and evaluated in [14, 5]. Among these the *adaptive weights* approach proposed by Yoon and Kweon

2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops
978-1-4244-4441-0/09/$25.00 ©2009 IEEE

[17], inspired by the Gestalt principles and based on spatial and color proximity constraints, deserves particular attention. This method deploys squared fixed supports but adapts weights to image content according to a joint bilateral filter applied on the support of the reference and target images. Improvements to the adaptive weight approach were proposed in [12], deploying segmentation as additional cue, and in [9], deploying a regularized range filter computed on a block basis. Dima and Lacroix [3] proposed a methodology for improving the performance of indoor stereo; they use the disparity map provided by a *correlation* algorithm with fixed support recording for each pixel the five local maxima and a measure of reliability. They then perform grouping of pixels with similar disparity, deploying a blob-coloring algorithm, to reduce the effects of noise and filtering outliers. Another approach that considers multiple local maximum of the correlation curve, proposed for multi-view stereo, is [2]. In this paper, for each point, the authors extract a set of potential candidates by the NCC curve and then solve a multi-label discrete MRF model. Zitnick and Kanade [18] proposed an iterative algorithm that updates the match values by diffusing support among neighboring points deploying a 3D support area. They explicitly model the continuity and uniqueness assumptions originally proposed by Marr and Poggio [18]. The continuity assumption is also explicitly modeled by approaches deploying segmentation (*e.g*. [6] for the *disparity refinement* step). Yang *et al*. [16] proposed an iterative approach aimed at refining low-resolution range images deploying bilateral filtering and sub-pixel interpolation. Improvements to sub-pixel accuracy were also reported for stereo images. In the stratified approach proposed by Kostková and Sára [8] the support adapts to high-correlation structure in disparity space determined by a pre-matching step. Finally, we point out a different but interesting methodology pertinent with the idea of point plausibility that was proposed in [1].

3. Proposed approach

By assuming rectified binocular stereo pairs, the two key observations that motivated our approach can be outlined by observing Figure 1. For a portion of the reference image R of the Tsukuba stereo pair (near the lamp) are shown 8 supports that would be used in determining the disparity of 8 blue points by a CA strategy based on a 5×5 support.

The first key observation is that each of these supports includes the same red point, and for each blue point, once that a correspondence is set, an implicit assumption concerned with the disparity of the red point is made. Assuming fronto-parallel supports, as done by the state-of-the-art CA strategies [5, 14], the disparity of the red point would be assumed each time to be equal to the disparity of the blue point. This implies that in this scenario 25 (potentially different) assumptions are independently made for the disparity of the red point without taking into account the evidence that the 3D surfaces are piecewise smooth. Moreover, it is worth observing that in this scenario 25 disparity assumptions are also made for the correspondent 25 supports in the target image T.

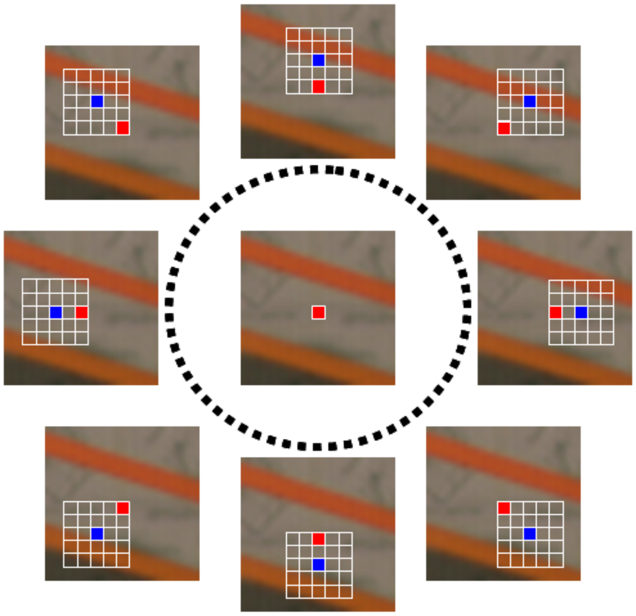

Figure 1. The same red point is included by the supports deployed to determine the correspondences of different blue points (*i.e*. the central points for which the disparity is being estimated in the neighborhood of the red point). For different blue points a disparity assumption is implicitly enforced for the same red point. **[Best viewed with colors]**

The second key observation is that the state-of-the-art cost aggregation strategies [14, 5] implicitly provide a mean for classifying each point belonging to their support by enforcing spatial and photometric constraints. Some of these approaches adapt the shape of the support to image content while others, more effectively, use a fixed support weighting each point according to image content [17, 9, 12, 4, 5].

By combining these two observations we define a framework that, for each point, allows for deriving a *plausibility* function that explicitly captures the geometric and photometric constraints implicitly and independently enforced within the neighboring points. It is worth observing that, compared to local approaches, we tackle the correspondence problem from a different point of view. Local approaches evaluate how well a certain disparity assumption fits with the points belonging to the support. Differently, we gather the multiple assumptions independently made by the supports that include a given point (*e.g*. the red point in Figure 1). Nevertheless, although the proposed framework was developed and tested using local algorithms, it could be deployed with any dense stereo algorithm.

3.1. Plausibility of point based correspondences

To formalize our framework we start by defining the *plausibility* of the disparity assumption made for each element of the support. For this, consider the situation depicted in Figure 2, where the supports for reference (S_f) and target image (S'_f) given a certain disparity hypothesis d are shown. We assume that the scene is piecewise smooth and, similarly to state-of-the-art CA strategies, we locally model the scene with fronto-parallel surfaces[1]. Under these assumptions the plausibility of points[2] g and g' being part of the supports S_f (centered in f) and S'_f (centered in f') at disparity d can be modeled by the following three events:

$\mathbf{E^R_{fg}}$: point $g \in S_f$ belongs to S_f. This event encodes the belief that g belongs to the fronto-parallel support centered in f. Assuming that the scenes are piecewise smooth, the plausibility of this event is related to the color proximity between f and g. For this event, we adopt the prior assumption that points closer to the central point are more relevant deploying a spatial proximity constraint.

$\mathbf{E^T_{f'g'}}$: point $g' \in S'_f$ belongs to S'_f. This event encodes the belief that g' belongs to the fronto-parallel support centered in f'. It is modeled similarly to E^R_{fg}.

$\mathbf{E^{RT}_{gg'}(d)}$: points $g \in S_f$ and $g' \in S'_f$ have disparity d and $-d$, respectively. This event encodes the belief that points g and g' are homologous and is related to the color proximity between g and g'. As we have no prior knowledge concerned with the disparity field we assume a uniform prior for $E^{RT}_{gg'}(d)$.

Given a certain color space (*e.g.* RGB in our case), let Δ^ψ a function that encodes color proximity between points f, g (and points f', g') and Δ^ω a function that encodes the color proximity between points g, g'. By Bayes' rule the posterior probability of joint event E^R_{fg}, $E^R_{f'g'}$,$E^{RT}_{gg'}(d)$ results:

$$\begin{aligned} P(E^R_{fg}, E^T_{f'g'}, E^{RT}_{gg'}(d) \mid \Delta^\psi_{fg}, \Delta^\psi_{f'g'}, \Delta^\omega_{gg'}) \\ \propto P_P(E^R_{fg}, E^T_{f'g'}, E^{RT}_{gg'}(d)) \cdot \\ P_L(\Delta^\psi_{fg}, \Delta^\psi_{f'g'}, \Delta^\omega_{gg'} \mid E^R_{fg}, E^T_{f'g'}, E^{RT}_{gg'}(d)) \end{aligned} \quad (1)$$

being P_P and P_L the prior and the likelihood, respectively. If we assume for simplicity that E^R_{fg}, $E^T_{f'g'}$ and $E^{RT}_{gg'}(d)$ are independent events[3] equation (1) can be written as:

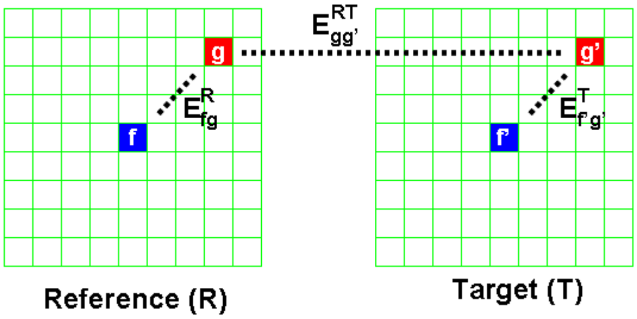

Figure 2. Supports S_f and $S_{f'}$ for points $f \in R$ and $f' \in T$ given a certain disparity hypothesis d. **[Best viewed with colors]**

$$\begin{aligned} P(E^R_{fg}, E^T_{f'g'}, E^{RT}_{gg'}(d) \mid \Delta^\psi_{fg}, \Delta^\psi_{f'g'}, \Delta^\omega_{gg'}) \\ \propto P_P(E^R_{fg}) \cdot P_L(\Delta^\psi_{fg} \mid E^R_{fg}) \cdot \\ P_P(E^T_{f'g'}) \cdot P_L(\Delta^\psi_{f'g'} \mid E^T_{f'g'}) \cdot \\ P_P(E^{RT}_{gg'}(d)) \cdot P_L(\Delta^\omega_{gg'} \mid E^{RT}_{gg'}(d)) \end{aligned} \quad (2)$$

For what concerns the priors, we set $P_P(E^R_{fg})$ (and $P_P(E^T_{f'g'})$) according to the following spatial proximity constraint:

$$P_P(E^R_{fg}) = e^{-\frac{\Delta_{f,g}}{\gamma_s}} \quad (3)$$

with Δ being the Euclidean distance between f,g and γ_s a parameter that controls the spatial proximity constraint. For what concerns $P_P(E^{RT}_{gg'}(d))$ we assume no prior knowledge about the appearance of g and g' in the two images setting $P_P(E^{RT}_{gg'}(d))$ as uniform.

We assume Lambertian materials and that the image formation process is affected by additive i.i.d. Gaussian noise. Let $I(p)$ a vector encoding the color intensity of point p (*e.g.* the three components $I(p)_R, I(p)_G, I(p)_B$ of the color space). For what concerns $P_L(\Delta^\psi_{fg} \mid E^R_{fg})$ (and $P_L(\Delta^\psi_{f'g'} \mid E^T_{f'g'})$) we model the color proximity constraints between f, g (and f', g') within the support according to this cue:

$$\Delta^\psi_{fg} = \sqrt{\sum_{c \in R,G,B} (I_c(f) - I_c(g))^2} \quad (4)$$

Similarly, for what concerns $P_L(\Delta^\omega_{gg'} \mid E^{RT}_{gg'}(d))$ we model the color proximity constraints between two potential corresponding points g, g' according to this cue:

[1] Our framework, although not straightforward, could be extended to more complex and realistic surface models.

[2] Given f, g and d, using fronto parallel supports f' and g' are unambiguously determined (*e.g.* $f' = f - d$ and $g' = g - d$)

[3] Doing this, we explicitly ignore conditional dependencies between the events that would be investigated in future research

$$\Delta^{\omega}_{gg'} = \sqrt{\sum_{c \in R,G,B} (I_c(g) - I_c(g'))^2} \tag{5}$$

To increase robustness (4) and (5) are truncated at ρ. Finally, from (2), (3), (4), (5) and assuming that residuals (4) and (5) are distributed according to Gaussian distributions, the plausibility of points g and g' of being part of supports S_f and $S'_{f'}$ at disparity d results:

$$P(E^R_{fg}, E^T_{f'g'}, E^{RT}_{gg'}(d) \mid \Delta^{\psi}_{fg}, \Delta^{\psi}_{f'g'}, \Delta^{\omega}_{gg'})$$
$$\propto e^{-\cdot\frac{\Delta_{f,g}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}{fg}}{\gamma_c}} \cdot e^{-\cdot\frac{\Delta_{f',g'}}{\gamma_s}} \cdot e^{-\frac{\Delta^{\psi}{f'g'}}{\gamma_c}} \cdot e^{-\frac{\Delta^{\omega}{gg'}}{\gamma_t}} \tag{6}$$

being γ_c and γ_t two parameters that control the behavior of $P_L(\Delta^{\psi}_{fg} \mid E^R_{fg})$, $P_L(\Delta^{\psi}_{f'g'} \mid E^T_{f'g'})$ and $P_L(\Delta^{\omega}_{gg'} \mid E^{RT}_{gg'}(d))$. Parameters ρ, γ_s, γ_c and γ_t are set empirically.

Therefore, once that a correspondence at disparity d between f and f' is set, (6) quantifies for each point of the two supports the plausibility of the disparity hypothesis d implicitly assumed for S_f and $S'_{f'}$. It is important to note that, differently by the weighting function deployed for CA in [17], the plausibility (6) includes the term $e^{-\frac{\Delta^{\omega}{gg'}}{\gamma_t}}$ that quantifies the likelihood of color proximity between gg'. Using this term rather than the matching cost $\Delta^{\omega}_{gg'}$ allows for rendering the contribution brought in by E^R_{fg}, $E^T_{f'g'}$ and $E^{RT}_{gg'}(d)$ homogeneous. Hereafter, in order to simplify the notation, we define the plausibility of point $g \in R$ with respect to the central point f given the disparity d (*i.e.* (6)) as $P^R_{f \rightharpoonup g}(d)$. Analogously for the target image T, we define (6) as the plausibility of point $g' \in T$ with respect to the central point f' given the disparity $-d$ as $P^T_{f' \rightharpoonup g'}(-d)$. It is important to note that the plausibility is symmetrical with respect to the two images (*i.e.* $P^R_{f \rightharpoonup g}(d) = P^T_{f' \rightharpoonup g'}(-d)$) but not so with respect to the same image (*i.e.* $P^R_{f \rightharpoonup g}(d) \neq P^R_{g \rightharpoonup f}(d)$ and $P^T_{f' \rightharpoonup g'}(-d) \neq P^T_{g' \rightharpoonup f'}(-d)$).

3.2. Gathering the plausibility of multiple disparity hypotheses

Let us assume that a disparity d has been set for the point $f \in R$ by a dense stereo algorithm based on a fronto-parallel CA strategy. In this case each point $g \in S_f$ belonging to the support is implicitly assumed at the same disparity d and receives, with respect to the central point f, a certain amount of plausibility given by (6). The plausibility of each point belonging to S_f for other values of disparity turns out to be zero. In these circumstances, when the correspondence for f at disparity d is set, the same assumption is made for each point $g' \in S'_{f'}$ belonging to the target image and centered at $f' = f - d$.

Since one point, say $g \in R$, is included by the supports of other points in its neighborhood (more precisely, it is included by all the points belonging to S_g, referred to as the *active support* of g) several assumptions concerned with its disparity (with consequent induced plausibility given by (6)) are made by the supports of the neighboring points within S_g. The number of points within the active support of each point g is given by the cardinality of set S_g (referred to as $\#S_g$); therefore, point g receives $\#S_g$ (potentially) different disparity assumptions by the neighboring points. Obviously, the only disparity assumption allowed is $d \in \{d_{min}, d_{max}\}$. This situation, for the point g of the reference image and with an active support of cardinality $\#S_g = 49$, is depicted in Figure 3.

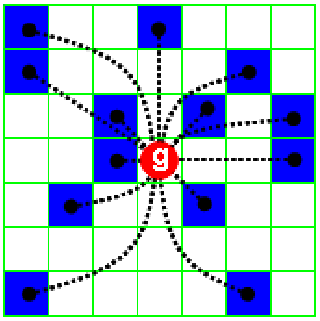

Figure 3. Red point g: the plausibilities of disparity assumption d made by other points within the *active support* S_g. In this case, (S_g of size 7×7) the red point has been included, and consequently assumed (blue points) at disparity d, by the support of 13 out of 49 points of S_g. For the other points within the active support (*e.g.* $49 - 13$), the plausibility of disparity d conveyed to g is zero as g was never assumed at disparity d by the supports of these neighboring points (empty cells).**[Best viewed with colors]**

Each time a correspondence between f and f' is set, the disparity assumption d is also made in the target image. However, in this case the situation is a little more complex as the points belonging to the active support of f' (*i.e.* centered in $f - d$ in T) are not distributed uniformly as in the reference image. In fact the neighboring points are mapped in the target image according to the algorithm that assigns the disparity, therefore points in the target image might be not included by the supports of the neighboring points in T (*e.g.* points belonging to occlusions with respect to the target image). However, the fact that one point in the target image is not included by the supports of its neighboring points simply emphasizes the evidence of its reduced plausibility. With this difference, the methodology outlined applies also to the target image.

Therefore, we gather for each point (say $g \in R$) and for each disparity d the plausibility derived by the implicit assumption locally made by the algorithm that assigns the disparity. This allows for deriving for each point $g \in R$ and for each disparity d $\in \{d_{min}, d_{max}\}$ an *accumulated plausibility* for the reference image referred to as Ω^R(g|d).

Analogously, gathering information from the target image allows for defining the accumulated plausibility Ω^T(g'|-d). Formally, we define Ω^R(g|d) and Ω^T(g'|-d) as:

$$\Omega^R(g \mid d) = \sum_{i \in S(g)} P^R_{i \rightharpoonup g}(d) \qquad (7)$$

$$\Omega^T(g' \mid -d) = \sum_{i' \in S'(g')} P^T_{i' \rightharpoonup g'}(-d) \qquad (8)$$

Finally, (7) and (8) are normalized by the overall plausibility of a point. With the realistic assumption that surfaces are piecewise smooth, plausibility for point g incoming from the neighboring points and induced by the assumptions made including g in their supports, turns out to be a powerful cue for determining locally consistent disparity fields. Nevertheless, we propose two additional improvements. The former consists in detecting unreliable assignments made by the algorithm that assign disparities: violation of uniqueness is a simple and well known approach deployed here for detecting disparity assignments in the occluded regions that might lead to perturbation of (7) and (8). The latter consists in cross validating the assumption made on R and T, encoded by (7) and (8), by means of the following (with respect to the reference image):

$$\Omega^{RT}(g \mid d) = \Omega^R(g \mid d) \cdot \Omega^T(g - d \mid -d) \qquad (9)$$

We conclude by pointing out that, although (7) and (8) rely on the disparity assignments made by a dense stereo algorithm (with computational complexity constrained by the disparity range), the computation of (7) and (8) is independent of the disparity range deployed because it depends only on the size of the active support. However, it noteworthy that the normalization, the cross validation and the detection of the maximum value of the accumulated plausibility are dependent on the disparity range.

4. Experimental results

In this section, the performance of the proposed methodology using the Middlebury dataset [10] is assessed[4]. For a better evaluation of the effectiveness of our proposal, referred to as *Locally Consistent* (LC) stereo, and for maintaining fairness with the local approaches, the disparity maps concerned with LC were obtained by means of the following point based WTA strategy (*i.e.* without aggregation of accumulated plausibility (9)):

$$\hat{d} = \underset{d \in \{d_{min}, d_{max}\}}{\text{argmax}} \; \Omega^{RT}(g \mid d) \qquad (10)$$

Parameters ALL, NOCC and DISC are defined according to the Middlebury web site [10]. ALL is the error computed on the whole image, NOCC is the error computed on the whole image excluding the occluded regions and DISC is the error computed within the discontinuity regions. As (9) encodes the plausibility of the multiple disparity assumptions locally made by a hypothetical CA strategy performing a sort of *local reasoning* but not a true global optimization/refinement step [11], we believe that it would not be fair to compare the disparity maps computed by our basic approach with those provided by approaches that are based on such optimization schemes. For these reasons we found that framework [14], aimed at assessing the performance of the state-of-the-art CA strategies, allows for optimally assessing the performance of our proposal. Nevertheless, although our approach (as well as the approaches considered in [14]) in its current form does not explicitly deploy any optimization and/or refinement of the disparity maps (*i.e.* it neither handles occlusions nor performs any refinement step aimed at filling incorrect disparity assignments), for the sake of completeness we also include the ALL error that renders the results of the framework [14] (otherwise restricted to NOCC and DISC errors) comparable to the results available on Middlebury [10].

Table 1 reports, according to [14], the performance of the top five performing CA strategies (namely: *SegmentSupport* (SS) [12], *Adaptive Weights* (AW) [17], *Segmentation Based* (SB) [4], *Reliability* (Rel) [7] and *Variable Windows* (VW) [15]). For each algorithm we computed NOCC, ALL and DISC errors according to the Middlebury website (the NOCC and DISC errors adopted by framework [14] are highlighted in boldface). Table 1 also includes two additional CA strategies referred to as Fast Bilateral Stereo (FBS) [9] and Fixed Window (FW) deployed for the evaluation of LC. FBS is a fast and accurate CA strategy with accuracy (see Table 1 and [9]) comparable to the top two performing algorithms [12, 17], while FW is an implementation of the fixed window approach. We deployed these two approaches for testing our proposal because FBS and FW are representative of the class of best performing and worst performing CA strategies, respectively. In the tables, for both approaches and for LC, the subscript refers to the radius of the support deployed. As the focus here is not on disparity refinement/interpolation, similar to [14], we found it to be fair and more effective to compare the results of our approach with the raw CA strategies available in [13] which are reported in Table 1. We do not consider here the results of SS [12] and AW [17] available on the Middlebury website as, in these cases, the raw disparity maps were processed by refinement/interpolation steps that significantly altered the results of the underlying raw CA strategy.

Table 2 reports the accuracy yielded by LC (with parameters $\gamma_s = 12, \gamma_c = 30, \gamma_t = 25, \rho = 69$ and the active support of radius 19[5]) using the disparity assumptions made

[4] Additional results: www.vision.deis.unibo.it/smatt/lc_stereo.htm

[5] This implies that, at least for reference image, each point is subject to 1521 assumptions concerned with its disparity

	Tsukuba			Venus			Teddy			Cones		
	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC
SS† [12]	**2.15**	4.04	**7.22**	**1.38**	3.0	**6.27**	**10.5**	19.7	**21.2**	**5.83**	16.4	**11.8**
AW† [17]	**4.66**	6.68	**8.25**	**4.61**	6.18	**13.3**	**12.7**	21.6	**22.4**	**5.5**	16.0	**11.9**
SB† [4]	**2.25**	2.86	**8.87**	**1.37**	2.31	**9.4**	**12.7**	20.1	**24.8**	**11.1**	19.2	**20.1**
Rel† [7]	**5.08**	6.94	**17.9**	**3.92**	5.5	**13.9**	**18.9**	27.0	**29.9**	**11.3**	20.7	**18.3**
VW† [15]	**3.12**	4.86	**12.4**	**2.42**	3.87	**17.7**	**25.9**	25.5	**21.2**	**29.6**	27.3	**18.3**
FBS$_{19}$ [9]	**2.95**	4.75	**8.69**	**1.29**	2.87	**7.62**	**10.71**	19.8	**20.82**	**5.23**	15.3	**11.34**
FW$_4$	**7.33**	9.32	**17.8**	**13.2**	14.5	**29.6**	**18.4**	26.4	**30.2**	**12.6**	21.7	**20.1**

Table 1. Accuracy according to the methodology defined by the Middlebury web site [10, 11] and (in boldface) according to [14]. The table reports the accuracy of the five top performing [14] state-of-the-art CA strategies [12, 17, 4, 7, 15], FBS [9] and FW. The disparity maps tagged with symbol †, available in [13] - section *"Original"*, are concerned with the original matching cost proposed by the authors of each paper and computed with the optimal parameters (see [14, 13] for details).

LC parameters	Tsukuba			Venus			Teddy			Cones		
$\gamma_s = 12, \gamma_c = 30, \gamma_t = 25, \rho = 69$	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC
1) LC$_{19}$ (U=ON, C=ON) + FBS$_{19}$	**1.77**	3.44	**5.92**	**0.27**	1.74	**1.77**	**9.30**	18.3	**17.9**	**4.75**	15.1	**10.5**
2) LC$_{19}$ (U=OFF, C=ON) + FBS$_{19}$	**1.90**	3.72	**5.85**	**0.38**	1.92	**2.50**	**9.25**	18.5	**18.0**	**4.61**	15.0	**10.1**
3) LC$_{19}$ (U=ON, C=OFF) + FBS$_{19}$	**1.76**	3.48	**5.68**	**0.34**	1.86	**2.50**	**9.63**	18.7	**18.6**	**5.06**	15.3	**11.3**
4) LC$_{19}$ (U=OFF, C=OFF) + FBS$_{19}$	**1.94**	3.89	**5.81**	**0.54**	2.11	**3.83**	**9.85**	19.0	**19.3**	**5.20**	15.6	**11.5**

Table 2. Accuracy obtained by LC (active support of radius 19) deploying the disparity assumptions made by FBS$_{19}$ for different configurations of U (uniqueness validation ON/OFF) and C (cross validation (9) ON/OFF).

by the FBS approach[6]. For completeness, the table reports the accuracy using four different configurations: uniqueness validation enabled/disabled (U={ON,OFF}) and cross validation (9) enabled/disabled (C={ON,OFF})). From Table 2 we can observe that LC, although deployed without any explicit CA strategy (*i.e.* (10)), concerning NOCC and DISC errors always dramatically improves the accuracy of the FBS approach as well as the accuracy of other top performing state-of-the-art approaches [12, 17, 4, 7, 15] reported in Table 1. Improvements are always notable on the whole dataset but are particularly evident for the Venus image where the error in the non-occluded areas (NOCC) drops to 0.27 (U=ON,C=ON) and at discontinuities (DISC) drops to 1.77 (*e.g.* for comparison, on the same image AW reports for NOCC and DISC 4.61 and 13.3 while SS reports 1.38 and 6.27). Although of limited interest – as all the considered approaches do not deal explicitly with the occluded regions – considering the ALL parameter, for each configuration reported in Table 2, LC outperforms (with the exception of SB, on the Tsukuba image) all the methods reported in Table 1. Table 2 also reports that enabling uniqueness validation and cross validation (U=ON,C=ON) provides the overall best performance. By comparing Tables 1 and 2 we also note that the improvements brought in by our proposal are more evident with the Tsukuba and Venus images, as these are mostly made of fronto-parallel and slanted surfaces, which better fit the model adopted for defining the plausibility. Figure 4 shows the disparity maps on the whole dataset for AW, VW, FBS$_{19}$ and LC$_{19}$. When we compare the disparity maps generated by LC with those of the state-of-the-art approaches, we observe that, although we do not explicitly deploy any smoothness term, our method provides a strong regularization of the disparity maps and improves the performance near the depth discontinuities. This means that the proposed accumulated plausibility is capable, on a point basis, to capture relevant cues concerned with the structure of the scene within the active support.

To assess the potential of our approach we evaluated the proposed methodology (with an active support of radius 19) deploying the disparity assumption made by FW - one of the low-performing CA approaches. Although (7) and (8) are modeled according to the behavior of an hypothetical CA strategy that is capable of adapting to the image content it was interesting to test the proposed methodology with an approach (FW) that ignored this cue. Table 3 reports the inferred accuracy of the disparity maps by deploying the accumulated plausibility proposed and computed using the disparity assumptions made by FW. The accuracy of the original approach (referred to as FW$_4$) on the four images of the dataset are reported in Table 1. Comparing Tables 1 and 3 we note that, on the whole dataset, LC dramatically improves the unreliable disparity maps provided by FW$_4$ (*e.g.* on Venus with U=ON and C=ON, NOCC and

[6] FBS parameters, see [9], $W = 39$, $w = 3$, $\gamma_s = 11$, $\gamma_c = 12$, TAD threshold 75.

LC parameters	Tsukuba			Venus			Teddy			Cones		
$\gamma_s = 74, \gamma_c = 20, \gamma_t = 32, \rho = 121$	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC
1) LC$_{19}$ (U=ON, C=ON) + FW$_4$	**3.19**	5.05	**9.85**	**0.57**	2.13	**5.30**	**10.6**	19.6	**22.1**	**5.52**	15.9	**12.3**
2) LC$_{19}$ (U=OFF, C=ON) + FW$_4$	**3.07**	4.92	**9.63**	**0.66**	2.22	**5.11**	**10.6**	19.7	**21.8**	**5.30**	15.7	**11.6**
3) LC$_{19}$ (U=ON, C=OFF) + FW$_4$	**3.17**	5.13	**9.06**	**0.75**	2.31	**6.94**	**11.1**	20.1	**23.1**	**5.82**	16.1	**12.7**
4) LC$_{19}$ (U=OFF, C=OFF) + FW$_4$	**3.09**	5.06	**8.87**	**0.90**	2.47	**7.18**	**11.3**	20.3	**23.0**	**5.87**	16.2	**12.6**

Table 3. Accuracy obtained by LC (active support of radius 19) deploying the disparity assumptions made by FW$_4$ for different configurations of U (uniqueness validation ON/OFF) and C (cross validation (9) ON/OFF).

DISC errors drop from 13.2 and 29.6 to 0.57 and 5.30, respectively). Enabling cross validation (C=ON) provides the overall best results and improvements are more evident on the Tsukuba and Venus images. Analyzing the results of Table 3 and Table 1 we note that the proposed methodology renders the unreliable disparity maps of FW$_4$ comparable, in most cases for NOCC, ALL and DISC, to those of the top performing approaches [12, 17, 4, 7, 15] and FBS$_{19}$ [9]. Figure 5 shows the disparity maps on the whole dataset for FW$_4$ for and LC$_{19}$ + FW$_4$. A qualitative analysis shows an evident improvement brought in by the proposed methodology (LC$_{19}$ + FW$_4$) compared to the original disparity maps. The regularization effect provided by the proposed methodology is even more evident in this case as the disparity measurements provided by FW$_4$ are noisier and more unreliable compared to FBS$_{19}$. Nevertheless, LC is capable of correctly capturing several cues present in the underlying structure of the scene. As for the execution time, with the exception of FBS$_{19}$ we deployed an unoptimized version of the code. With a 2.5 GHz Intel Core Duo processor, the whole process (FBS$_{19}$ + LC$_{19}$) required 13 seconds on Tsukuba and 37 seconds on Teddy. In this case the execution time is dominated by FBS$_{19}$ (5 and 24 seconds respectively). For FW$_4$ + LC$_{19}$, using an unoptimized version of the code on the same images we report 8 and 15 seconds for the overall process, respectively.

5. Conclusions

In this paper, we have proposed a novel methodology to deal with the correspondence problem that is motivated by the implicit assumptions made by the cost aggregation strategies. Our approach gathers the plausibility derived by the multiple assumptions made by a hypothetical adaptive CA strategy inducing a point based plausibility function that locally captures the underlying geometric and photometric structure of the scene. The experimental comparison of the proposed point based function, built upon the disparity assumptions of two diametrically opposite approaches, with state-of-the-art approaches highlights the effectiveness and the potential of our proposal.

References

[1] Y. Boykov, O. Veksler, and R. Zabih. A variable window approach to early vision. *IEEE Trans. PAMI*, 20(12):1283–1294, 1998.

[2] N. Campbell, G. Vogiatzis, C. Hernandez, and R. Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *ECCV 08*, pages 766–779, 2008.

[3] C. Dima and S. Lacroix. Using multiple disparity hypotheses for improved indoor stereo. In *ICRA*, pages 3347 – 3353. IEEE, May 2002.

[4] M. Gerrits and P. Bekaert. Local stereo matching with segmentation-based outlier rejection. In *Proc. CRV 2006*, pages 66–66, 2006.

[5] M. Gong, R. Yang, W. Liang, and M. Gong. A performance study on different cost aggregation approaches used in real-time stereo matching. *Int. Journal Computer Vision*, 75(2):283–296, 2007.

[6] H. Hirschmuller. Stereo processing by semi-global matching and mutual information. *IEEE Trans. PAMI*, 30(2):328–341, 2008.

[7] S. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *Proc. CVPR 2001*, pages 103–110, 2001.

[8] J. Kostková and R. Sára. Stratified dense matching for stereopsis in complex scenes. In *Proc. of BMVC2003*, pages 339–348, 2003.

[9] S. Mattoccia, S. Giardino, and A. Gambini. Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering. In *Proc. of ACCV2009 (to appear)*, 2009.

[10] D. Scharstein and R. Szeliski. Middlebury stereo vision. http://vision.middlebury.edu/stereo/.

[11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Jour. Computer Vision*, 47(1/2/3):7–42, 2002.

[12] F. Tombari, S. Mattoccia, and L. Di Stefano. Segmentation-based adaptive support for accurate stereo correspondence. In *Proc. IEEE Pacific-Rim Symposium on Image and Video Technology (PSIVT'07)*, 2007.

[13] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. www.vision.deis.unibo.it/spe/SPEHome.asp.

[14] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. In *CVPR08*, pages 1–8, 2008.

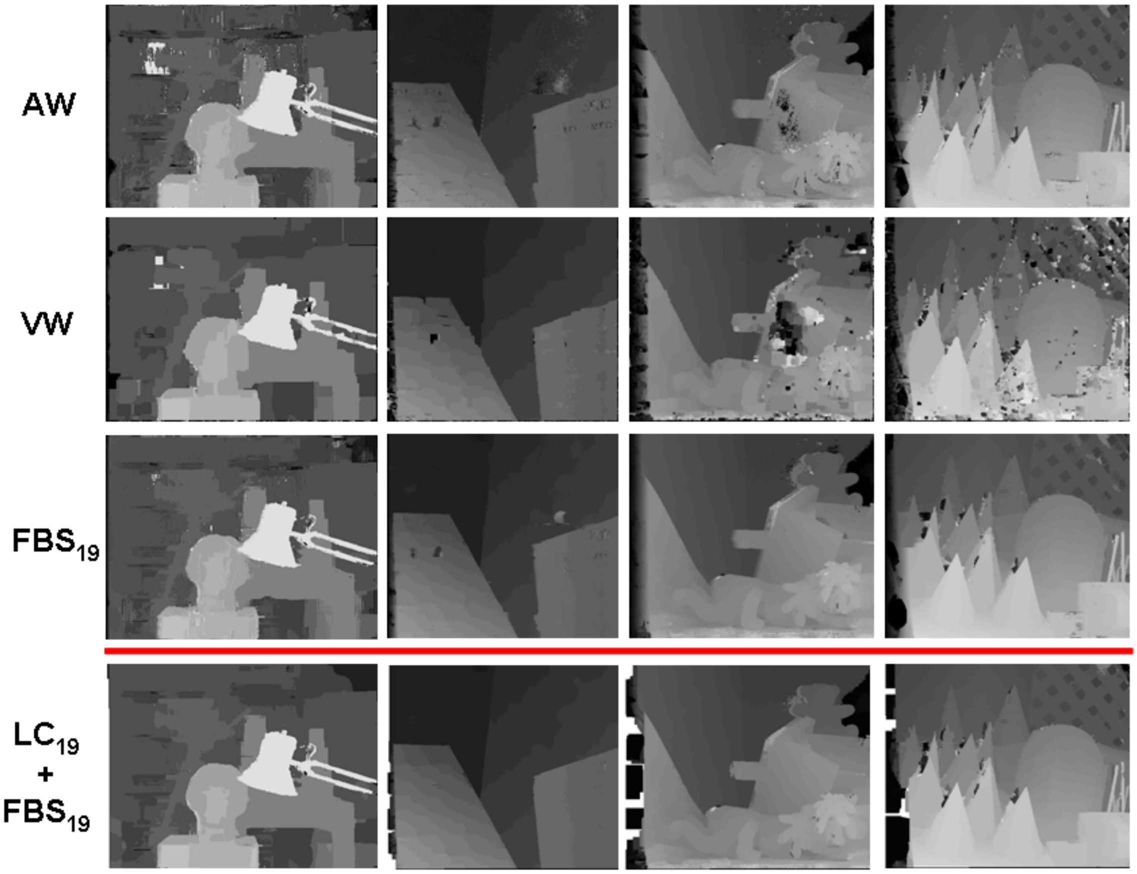

Figure 4. From top to bottom: disparity maps for AW [17], VW [15], FBS_{19} [9] and LC_{19} + FBS_{19} (U=ON,C=ON).

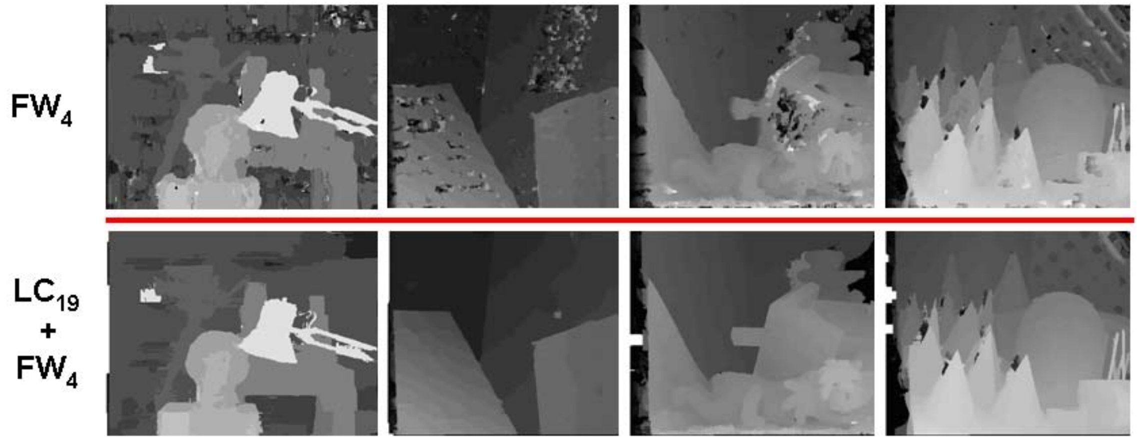

Figure 5. (Top) Disparity maps for FW_4 and (Bottom) for LC_{19} + FW_4 (U=ON,C=ON).

[15] O. Veksler. Fast variable window for stereo correspondence using integral images. In *Proc. CVPR 2003*, pages 556–561, 2003.

[16] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *Proc. of CVPR2007*, pages 1–8, 2007.

[17] K. Yoon and I. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. PAMI*, 28(4):650–656, 2006.

[18] C. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Trans. PAMI*, 22(7):675–684, 2000.