Stereo under Sequential Optimal Sampling: A Statistical Analysis Framework for Search Space Reduction

Yilin Wang, Ke Wang, Enrique Dunn, Jan-Michael Frahm
University of North Carolina at Chapel Hill
{ylwang,kewang,dunn,jmf}@cs.unc.edu

Abstract

We develop a sequential optimal sampling framework for stereo disparity estimation by adapting the Sequential Probability Ratio Test (SPRT) model. We operate over local image neighborhoods by iteratively estimating single pixel disparity values until sufficient evidence has been gathered to either validate or contradict the current hypothesis regarding local scene structure. The output of our sampling is a set of sampled pixel positions along with a robust and compact estimate of the set of disparities contained within a given region. We further propose an efficient plane propagation mechanism that leverages the pre-computed sampling positions and the local structure model described by the reduced local disparity set. Our sampling framework is a general pre-processing mechanism aimed at reducing computational complexity of disparity search algorithms by ascertaining a reduced set of disparity hypotheses for each pixel. Experiments demonstrate the effectiveness of the proposed approach when compared to state of the art methods.

1. Introduction

Dense stereo disparity/depth estimation methods commonly rely on the exhaustive enumeration of the photo-consistency cost volume attained from an *a priori* determined set of disparity/depth hypotheses. This is inherently inefficient given that, in the absence of scene structure priors, all pixels share a common hypotheses search space designed to cover the entire scene volume. To avoid exhaustive sampling we propose a novel framework to reduce the candidate hypotheses per pixel. The reduction especially benefits high resolution stereo for modern high resolution digital cameras or satellite terrain heightmap estimation shown as in Figure 1 (with an additional computational burden due to the rational polygonal camera model (RPC) [3] during the photo-consistency cost computation).

Recent randomized [1] and structured [8] sampling schemes are efficient and robust mechanisms for disparity estimation. The concepts of sparsity and propagation are recurring themes across these efficiency driven optimizations of the basic disparity search framework. The underlying property being exploited is that of scene structure regularity due to the assumption of local depth correlations among adjacent pixels. These assumptions are typically encoded as predetermined sampling distributions or data propagation schemes. Since these broad *a priori* assumptions are in general error prone we, instead, favor the explicit adaptive sampling of the disparity (or depth) search space to build incremental models of the local scene structure.

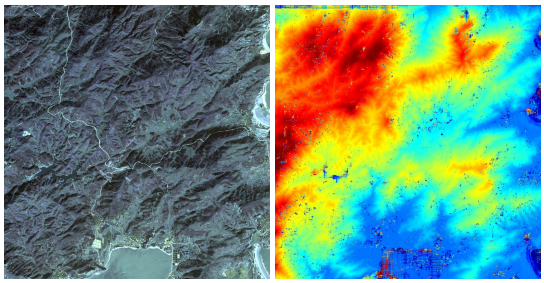


Figure 1. Left: Stereo disparity for high resolution satellite images (150M) is a computationally intensive task evaluating upwards of 1000 hypotheses per pixel. Right: Output from our prosed SOS$^+$ framework, space reduction enables an order of magnitude reduction in computational complexity.

We propose an efficient sampling scheme for building an accurate model of the local disparity structure. Our sampling scheme strives to minimize the number of sampling computations while providing statistical guarantees of coverage sufficiency based on the sequential probability ratio test (SPRT). This data driven sampling scheme enables an adaptive framework for optimal disparity sampling leading to a turnkey solution for reduced complexity disparity sampling along with a high-efficiency seed propagation. We utilize the output of our optimal pixel based sampling to estimate local planar surface approximations and deploy a seed propagation framework leveraging our reduced disparity search space. Figure 2 depicts an overview of our developed system, which attains state of the art performance.

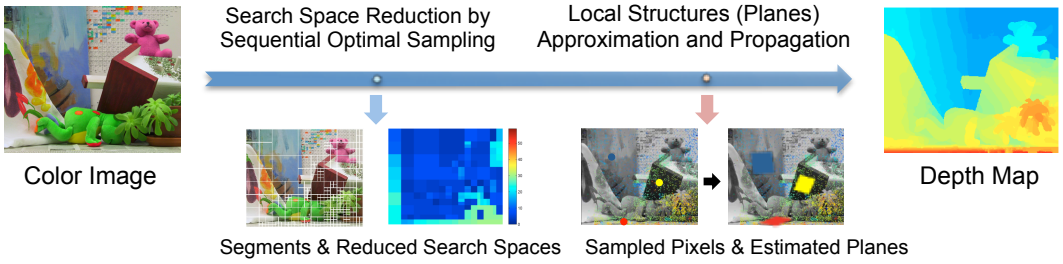


Figure 2. Overview of our proposed approach

2. Related Work

High computational cost is commonly required to obtain disparity maps with satisfying quality, such as clear boundaries, by using global methods (e.g. Graph Cuts and Belief Propagation [13]) or adaptive support aggregation window [20]. The recent explosion in image resolution has brought efficiency to the forefront of requirements for high quality stereo. Algorithms for reducing the burden of matching cost aggregation while retaining matching accuracy include non-local aggregation [19], cross-based aggregation[7], and fast cost-volume filters [10]. Search space reduction for stereo offers a complexity reducing framework to avoid the exhaustive evaluation of the cost volume (i.e. reducing the number of matching cost computations). Our proposed statistical analysis sampling framework falls into this category.

Hierarchical stereo (HS) [14, 12] is a multi-resolution approach for deterministic search space reduction, but lacks adaptability to fine structures. Veksler [15] compared the effect of using the disparities obtained by HS, dynamic programming and local stereo for limiting the disparity range, and concluded that reduction by local stereo resulted in almost no loss in accuracy with a significant efficiency improvement over HS and dynamic programming. We operate at the original pixel resolution, using probabilistic modeling to adapt to local scene structures and provide a bound for omitting such structures from the disparity search space.

Wang *et al*. [17] proposed a search space reduction method for MRF stereo based on estimating a putative disparity map from pixel-wise photo-consistency. Estimation reliability is verified through left-right consistency and reliable pixel estimates are propagated to the entire image. The disparity variability in the local neighborhood is then used to determine a candidate depth range for each pixel. In contrast, we model the uncertainty in the local photoconsistency to make a decision regarding the inclusion of a sampled depth into a set of locally representative estimates.

Histogram Aggregation (HA) [8] reduces the computational complexity of the the disparity estimation by combining a pixel-wise likelihood histogram aggregation scheme with sparse image sampling. To be robust against noisy matching cost, the pixel-wise depth candidates for each seed in the sampling grid are attained by selecting a fixed number of local extrema in the seed cost function. The resulting set of extrema is used in a spatial voting framework to propagate the depths to the entire image. Our approach does not make hard *a priori* assumptions about the pixel cost behavior, instead we build a statistical model of the confidence of the global extremum for each sampled pixel cost function to incrementally discern among reliable and unreliable depths.

PatchMatch (PM) [1] is a fixed propagation scheme with random depth initialization. With assumed sufficient sampling of the local structure (i.e. depths), the spatial propagation and pairwise hypothesis comparison message passing framework effectively converges in few iterations. PM treats each pixel as a seed, which propagates its best disparity to other pixels.Our approach focuses on achieving sufficient local depth sampling sufficiency and develops an adaptive propagation framework to communicate not only the current best estimate of a given pixel but also to incorporates information regarding its candidate set.

SDDS [18] is a sparse sampling framework that has a similar goal to our proposed scheme. For each local patch, it randomly computes the matching cost for one pixel per disparity, and then finds a candidate set by repeated sampling with a constant number (3 or 4) of iterations. In contrast, our sampling scheme balances sampling completeness and efficiency in a statistical framework, enabling the adaptive termination of local structure sampling.

3. Our Optimal Disparity Sampling Scheme

We strive to improve the efficiency of stereo approaches by reducing the search space. To this end, our approach focuses on eliminating most incorrect candidate disparities (or depths) by stereo depth sampling. Such elimination enables, in principle, stereo approaches leverage the reduced search space without significant quality degradation. Stereo depth sampling aims to attain a representative scene structure by exploring the photo-consistency cost volume. A sampling operation refers to determining a pixel's depth estimate from the enumeration of its cost function across the entire depth range. We develop a sampling scheme aimed at *efficiently* ascertaining a *compact* and *sufficient* representa-

tion of local scene structure in terms of a reduced set of candidate disparities D. The sufficiency of our reduced search space entails that D contains all distinct disparities in a local neighborhood, while compactness pertains to the cardinality of D. Conversely, the efficiency of the sampling process hinges on the number of sampling operations required to determine the all the elements in D.

It is well known that local photo-consistency is a noisy measurement. Since the best matching cost does not always corresponds to the correct disparity, noisy depth observations will be mistakenly added to the candidate disparity set. A naive sampling scheme is to specify a fixed sampling ratio and perform random sampling (RS) for a predetermined portion of pixels. Such open loop sampling disregards the observed local structure and may be arbitrarily inefficient or insufficient, i.e. oversampling in simple and flat regions while undersampling in regions with complicated and overlapping structures. Our adaptive sampling scheme overcomes these limitations by building an incremental model of the local disparity candidate set D and relying on SPRT to determine an optimal stopping criteria for the random sampling within each local neighborhood.

3.1. Sequential Probability Ratio Test

SPRT is a pairwise likelihood-based hypothesis testing technique commonly used in decision theory. The work of Chum and Matas [2] is an example of the use of SPRT within robust model estimation frameworks. Given two hypotheses H_0 and H_1, along with sequential observations $x_k(k = 1, ..., n)$, suppose the corresponding likelihoods for these two hypotheses $P(x_k|H_{i=\{0,1\}})$ are already known. In the SPRT model, testing is controlled by the accumulated likelihood ratio L:

$$L_n = \prod_{k=1}^{n} \frac{P(x_k|H_0)}{P(x_k|H_1)} = L_{n-1} \cdot \frac{P(x_k|H_0)}{P(x_k|H_1)}. \qquad (1)$$

Given thresholds T_1 and T_0 ($T_1 \leq T_0$), for each observation the SPRT model will be one of following three states: 1) $L \geq T_0$: stop testing and accept H_0; 2) $L \leq T_1$: stop testing and accept H_1; 3) $T_1 < L < T_0$: wait for a new observation. The SPRT model is completely data driven and optimal in the sense that it minimizes the number of samples needed arrive at a given decision [16].

We now describe how the likelihood thresholds T_0 and T_1 are determined *a priori* through user-defined decision error bounds e_0 and e_1. The errors e_0 and e_1 correspond to the admissible probability of erroneously accepting H_0 and H_1, respectively. For illustration, without loss of generality, we assume H_0 is the correct hypothesis and a given threshold T_1. Given the set of all possible sequences of consecutive observations x_k, the likelihood error e_1 is the proportion observation sequences leading to the acceptance of H_1 (i.e. $L_n \leq T_1$). A corresponding argument also applies to H_1 and T_0 to describe e_0. Accordingly, e_0 and e_1 quantify the representative ability of the likelihood function L_n and can be pre-specified *a priori*. The thresholds T_0 and T_1 are related to the likelihood errors by $T_0 \leq \frac{1-e_0}{e_1}$ and $T_1 \geq \frac{e_0}{1-e_1}$ [16]. The above modeling describes a general framework for robust and efficient sampling, we now describe the framework in the context of disparity estimation.

3.2. Sequential Optimal Sampling for Stereo

Recall that a depth sampling model finds a set of candidate depth D for each local patch by sampling K pixels in the search space S and evaluating their matching costs for a given aggregation window. We design a scheme that dynamically adjusts the value of K according to the previous observations. The observation x_k is represented by the cost profile of the kth randomly selected pixel p, comprised by the matching cost for all candidate depths $d \in S$. Our matching cost is computed based on color and gradient values as in Bleyer et. al. [1].

In our SPRT scheme, hypothesis H_0 is that the current disparity set D is sufficient with a probability α, and H_1 is that D should be expanded to $D'(\supseteq D)$. Different from the standard SPRT scheme, our model performs an incremental test, where hypotheses H_0 and H_1 keep changing until D is accepted. $P(x_k|H_0)$ and $P(x_k|H_1)$ are the likelihoods to accept D and D' respectively. Since D' is a superset of D, we have $P(x_k|H_0) \leq P(x_k|H_1)$. Accordingly, the accumulated likelihood ratio L is monotonically decreasing for any given pair of hypotheses H_0 and H_1. Hence, we can see that our SPRT model based on evolving hypotheses is similar to a "one-sided" model, which only considers whether to expand the current D by comparing the accumulated likelihood ratio L with the threshold T_1.

A candidate depth set D is α-sufficient if it has not been updated in N consecutive observations, and our optimal sampling aims to find an α-sufficient depth set with the minimum samples. Values for hypotheses errors e_0 and e_1 are attained from two user parameters: α_{suff} (sufficiency) and α_{conf} (confidence), where α_{suff} is an acceptable accuracy of the reduced disparity set (e.g. for $\alpha_{\text{suff}} = 90\%$, the reduced set contains the true depths of at least 90% of the pixels within the sampling block), and α_{conf} is an estimated probability of finding a new depth through N independent samples, e.g. our confidence on the 90% sufficiency assertion. Given the two parameters, since $\alpha_{\text{conf}} = 1 - (\alpha_{\text{suff}})^N$, we have $N = ln(1-\alpha_{\text{conf}})/ln(\alpha_{\text{suff}})$, which is an overestimate as it assumes sampling with replacement.

We call this SPRT-based sampling scheme Sequential Optimal Sampling (SOS), which is able to sample less and provide a significantly tighter search space for each local image region. Figure 3 shows the flowchart of the SOS model, where given the original depth set D, the model s-

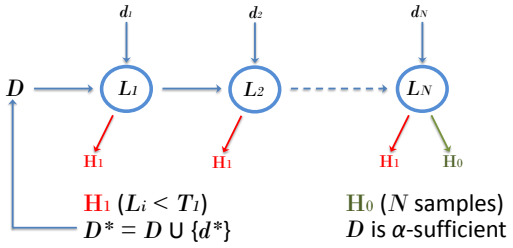


Figure 3. Sequential Optimal Sampling (SOS) model. Given a depth set D and a new observations d_i, if the accumulated likelihood drops below threshold T_1, D is augmented. Sampling halts after N consecutive samples without updates to D.

tate L (the accumulated likelihood ratio) is updated by each new observation d. To improve the sensitivity to photo consistency noise we design our likelihood function to be more tolerant to costs slightly greater than the minimum matching cost. Let $c(x,d)$ be the matching cost for pixel x at depth d, $\bar{c}(x)$ the mean cost for x across all depth hypotheses, and d^* the depth with the minimum matching cost. We define the score for each depth d as:

$$s(d,x) = \exp(-1 + \frac{\bar{c}(x) - c(x,d)}{\bar{c}(x) - c(x,d^*)}) \tag{2}$$

The range of the score $s(d,x)$ is $(0,1]$, with higher scores being sought. Thus, even if the current best disparity estimate d^* is not included in the current depth set D, if there exists some $d \in D$ whose score is relatively high, we delay the inclusion of d^* into D until we have gathered more sampling evidence to mitigate spurious depth outlier estimates. Accordingly, in this new support test model, newly encountered depths will be stored in a candidate pool instead of being directly added into D, and the sample information (the coordinates and the cost curve) will be recorded and will then be used in updating the candidate depth set.

The likelihood $P(x|H)$ is defined as follows:

$$P(x|H) = \frac{\max_{d\in D} s(d,x)}{\sum_{\tilde{D}} \max_{d\in\tilde{D}} s(d,x)} \tag{3}$$

where $\tilde{D}$ is an arbitrary subset of the entire search space. Let $D' = D \bigcup \{d'\}$ be the superset of current reduced depth candidate set D, which will be used to replace D. Recall that H_0: D is α-sufficient, and H_1: D' is α-sufficient. Then the likelihood ratio is

$$\frac{P(x|H_0)}{P(x|H_1)} = \frac{\frac{\max_{d\in D} s(d,x)}{\sum_{\tilde{D}} \max_{d\in\tilde{D}} s(d,x)}}{\frac{\max_{d\in D'} s(d,x)}{\sum_{\tilde{D}} \max_{d\in\tilde{D}} s(d,x)}} = \frac{\max_{d\in D} s(d,x)}{\max_{d\in D'} s(d,x)}$$
$$\geq \frac{\max_{d\in D} s(d,x)}{\max_{d\in D\bigcup\{d^*\}} s(d,x)} = \max_{d\in D} s(d,x)$$

where equality is achieved when $d' = d^*$. Then the accumulated likelihood ratio L is

$$L_n = \prod_{i=1}^{n} \frac{P(x_i|H_0)}{P(x_i|H_1)} \geq \prod_{i=1}^{n} \max_{d\in D} s(d,x_i) = L_n^* \tag{4}$$

The lower bound of the accumulated likelihood ratio L_n^* is used to replace L_n in the test, i.e. D will be updated if L_n^* is less than a predefined threshold T_1. Since there will be several different depths in the candidate set, we exploit the recorded sample information to accumulate the likelihood ratio for each candidate, and choose the one with the lowest value as d' to expand the current depth set D. According to the SPRT theory [16], the lower bound for threshold T_1 is given by $T_1 \geq \frac{e_0}{1-e_1}$ where e_0 and e_1 are errors for hypotheses H_0 and H_1 respectively. Since D' is the superset of D, which yields $e_0 \geq e_1$, we choose $T_1 = \frac{e_0}{1-e_0} \geq \frac{e_0}{1-e_1}$ to be a more strict threshold, and $e_0 = 1 - \alpha_{\text{suff}}$ according to the definition of α-sufficiency.

3.3. Bounding the Output Search Space

The number of attained disparities through sequential optimal sampling for a given image region is dependent on the observed scene structure. Accordingly, image regions covering a large number of disparities will be assigned a larger set of candidate depths. However, the efficiency of many stereo algorithms (such as Belief Propagation) is heavily impacted by the maximum size of all candidate sets. To make the candidate depth sets more balanced over the entire image, we introduced a quad-tree based recursive sampling strategy to leverage sampling patch consistent with the true depth distribution, which attains disparity sets of bounded cardinality K across the entire image. This adaptive sampling scheme guarantees that the search space of each pixel won't exceed K, so the upper bound of computation cost for stereo is limited. In order to achieve such behavior, when we detect that the current set of K depths is still incomplete, instead of adding a new depth, the patch is recursively partitioned into four sub-blocks and the optimal sampling process restarted for each new partition. Figure 4 shows an image automatically segmented by constrained optimal sampling. Note how complex regions are sampled by smaller windows while large homogeneous regions like the background remain unchanged.

3.4. Recovering Isolated Structures

Sequential optimal sampling is designed to achieve disparity set α-sufficient for each given region. Accordingly, our coverage may be incomplete whenever there are isolated structures comprising a region coverage near or below the $1-\alpha$ confidence threshold. The shortcoming of sampling in regular square blocks is that some isolated regions (such as the tip of the lamp in Figure 4) might be missed. Although the portion of such areas is much smaller with

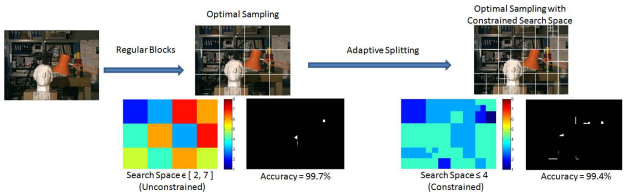


Figure 4. Optimal sampling with constrained search space. Limiting the size of the disparity setby spatial partitioning incurs in a marginal accuracy penalty at partition boundaries.

respect to the entire patch, it is still worth to recover the missed search space for some boundary sensitive applications. In this paper, the search space for the local patch will be propagated to small neighboring regions, so that isolated areas (as the errors shown in Figure 4) will have a joint search space that contains the correct depth. The length of the propagation is set as 10% (called propagation ratio) of the patch size. In practice such propagation results in a small dilation of each sampling block aimed at mitigating the boundary effects of our block partitioning scheme.

4. Stereo under Optimal Sampling

The SOS scheme proposed in Section 3 reduces the search space for each pixel. Accordingly, the proposed method can be deemed as a generic complexity reducing pre-processing step for stereo estimation algorithms. The computational benefits of performing such pre-processing depend on the choice of the stereo disparity algorithm being deployed. Clearly, exhaustive search stereo achieves a linear speedup with respect to the search space compression ratio, while global methods may benefit at rate proportional to their algorithmic complexity. Besides the reduced search space, our optimal sampling scheme adaptively separates the image into proper sub-regions, whose local structures are also inherently represented by the candidate disparity set and the spatial distribution of the corresponding sampled pixels. To fully utilize these cues, here we propose an efficient stereo approach by propagating exploited structures.

4.1. Local Structure Approximation

A local image patch may contain one or more distinct structures (e.g. fronto-parallel or slanted planes), most of which should be sufficiently covered by sampled pixels under SOS. We first group sampled pixels into connected components based on their color similarity and spatial distance. Then, each component is treated as a distinct structure and locally approximated by oriented planes. Namely, each sampled pixel s will be assigned with the disparity having the minimum matching cost in the reduced disparity set, and fitted with an oriented plane by RANSAC-like estimation [9]: for each iteration we randomly select three sampled pixels from the same component, and generate a plane $pl(x, y)$: $a \cdot x + b \cdot y + c$, where x and y are coordinates of the seed, d is the corresponding best disparity, and a, b, and c are estimated coefficients. Aggregating the weighted matching cost for pixel s from all the seeds belonging to the same component with respect to the generated plane, and keep the plane if its cost is better than the current cost (the default cost is computed from the fronto-parallel plane). Once a new plane is found, we also check the neighboring planes pl^+ and pl^-, which are generated by increasing and decreasing c with 1 disparity, and keep the one with the best matching cost. We repeat sampling, until no better matching cost is found for $R = 10$ consecutive samples. In this way, we can find the best planes for all sampled pixels, which will be used as propagation seeds.

4.2. Local Structure Propagation

We use the sampled pixel positions s and their associated local plane estimates pl_s as the input for a spatial propagation scheme. Each seed p_s propagates its plane $pl_s(x, y)$ to its four neighbors, and each neighbor p_n will discard the invalid planes whose disparities are not in p_n's search space. For a valid plane, p_n will simply set pl_s as its local plane if p_n has not received any plane before. Otherwise, p_n compares the matching cost of pl_s against the cost of its current best plane, and if the new cost is better p_n will update its local plane by the best of $\{pl_s^-, pl_s, pl_s^+\}$. When the local best plane is updated, the new plane will also be propagated to p_n's neighbors. Pixels updating their local plane will be the seeds in the next iteration, and the propagation will stop when a steady state is reached. Since SOS splits a local patch into compact regions having reduced candidate depths, our seed propagation corresponds to the implicit smoothness of the local patches. Namely, many pixels directly "borrow" the plane estimate from neighboring seeds. Matching cost is only estimated for pixels already having an assigned local plane (either by seed initialization or subsequent propagation) and are receiving a contradicting disparity estimate from one of their neighbors. The influence of incorrect seed estimates is restricted to those regions containing such erroneous disparities.

4.3. Disparity Post-Processing

Next we perform a disparity refinement where the reliable pixel disparity estimates are identified through left-right cross-validation and unreliable pixels near the left image boundary are assigned the median of the neighboring reliable pixels. Remaining unreliable pixels are interpolated according to the method proposed in Hirschmueller et al. [5]. Then, errors in the textureless regions are mitigated by a) segmenting the image into connected components based on color similarity, b) identifying segments having small number of disparities (< 10) where there is dominant disparity ($\geq 50\%$ of pixels), and c) propagate dominant

disparity to entire connected component. Finally, weighted median filtering is used to smooth the disparity map.

5. Experiments

We evaluate our search space reduction through SOS sampling and present results of its use by our proposed propagation scheme as well as in combination with other stereo algorithms. For ground truth evaluation and benchmarking we used the the Middlebury Stereo datasets [11]. All algorithms were implemented in C++ and executed on an Intel Xeon CPU W3540 2.93GHz. The default aggregation window size is 3×3 for our depth sampling preprocessing step. Matching cost computation parameters are set to the default parameters proposed in [1]. The SOS stopping parameters were set to $\alpha_{\text{suff}} = 0.90$ and $\alpha_{\text{conf}} = 0.95$.

5.1. Search Space Reduction from SOS

We compare our (SOS) scheme and SOS with constrained search space $\|D\| \leq 5$ (SOS-C) against a Random Sampling scheme RS(X), which randomly selects pixels with the fixed sampling ratio $X = \{0.005, 0.01, 0.05, 0.1\}$ in each patch and uses their disparities to form the reduced search space. The reduced search space is evaluated in three aspects: cardinality, accuracy, and redundancy. The results of SOS are evaluated on non-overlapping blocks of default size 50×50. Our evaluation is based on the average data of the five test images: tsukuba, venus, teddy, cones, and art.

Compactness. Figure 5 (column 1) compares the reduced search space for SOS and SOS-C against RS(X) with multiple fixed sampling ratios. Both SOS and SOS-C consistently provide smaller search spaces, irrespective of patch size. Moreover, our proposal found more compact disparity sets than the random sample variants geared at performing less sampling (e.g. RS0.005).

Accuracy. We analyze the fraction of pixels whose ground truth disparity is present in the reduced set. In Figure 5 (column 2) we can see the accuracy of SOS and SOS-C are always above 95% with arbitrary matching windows sizes and patch sizes, showing that our optimal schemes are able to obtain a stable accuracy by adjusting the sampling ratio according to local structures, providing more flexibility than random sampling with a predefined ratio.

Redundancy Figure 5 (column 3) measures by the average number of wrong disparities in the reduced disparity set. Our optimal sampling models consistently mitigate redundancy (less than 1 spurious disparities in the candidate set), improving over any random sampling scheme. Note the two high accuracy (nearly 100%) schemes RS 0.05 and RS 0.1 offer large redundancy (2% accuracy improvement with more than 20 spurious depths).

Sampling efficiency. We focus on the total number of samples required to estimate the local structure. The sampling ratios for SOS and SOS-C are shown in Figure 5 (column 4), and we observe a stable ratio around 1%. Figure 6 shows the final patches generated by SOS-C and the corresponding sampling density in each of the patches. In general, the block size reveals the complexity of the local structure and all pixels in the image have a bounded (i.e. turnkey) reduced disparity set. Moreover, SOS-C successfully detects image regions with complex structure and recursively partitions said region. Accordingly, flat regions with few disparities are represented by relatively large blocks.

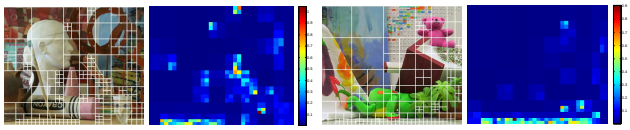


Figure 6. Local structures exploited by optimal sampling with constrained search space. Left of pair images show spatial partitioning while right of the pair images show the sampling ratio.

Experiments show that the SOS schemes outperform the fixed ratio random sampling RS(X) schemes. Processing times of SOS-C for tsukuba, venus, teddy, cones and art are 21ms, 42ms, 106ms, 114ms, and 193ms respectively. Thus, SOS and SOS-C are reliable light-weight sampling schemes suitable as a stereo complexity reduction pre-process.

5.2. SOS$^+$: Stereo under SOS

We now evaluate the performance of our SOS-based propagation framework (SOS$^+$) as well the coupling of SOS as a pre-processing step for a variety of stereo algorithms. We compare the performance of our propagation-based stereo against two efficiency driven state of the art disparity sampling techniques PatchMatch (PM) [1]and HistogramAggregation (HA) [8]. As an additional baseline we include typical local and global stereo methods: Exhaustive search (EX) and Belief Propagation (BP) under the complete and the reduced disparity search space estimated through SOS (PM+S, HA+S, EX+S, and BP+S). The reduced search space is generated by SOS-C on 100×100 blocks with a maximum size of the disparity set of $|D| = 5$ and using propagation ratio $\gamma = 0.1$.

Stereo on Fronto-Parallel Planes To enable leveled comparison against algorithms working under the fronto-parallel assumption we modify SOS$^+$ and PM for compliance to this assumption. The default window size for cost aggregation is 11, except for BP (no explicit cost aggregation). For HA (position-dependent), the spatial ratio is 3, and aggregation window is 31 (the default value used in [8], which is similar to aggregate cost from 11×11 pixels). For fronto-parallel PatchMatch (PM(FP)), the maximum number of iterations is four, and in each iteration the disparities are propagated starting from the top-left to the bottom-right, and then they are propagated back to the top-left. For BP, the maximum number of iterations is fifteen.

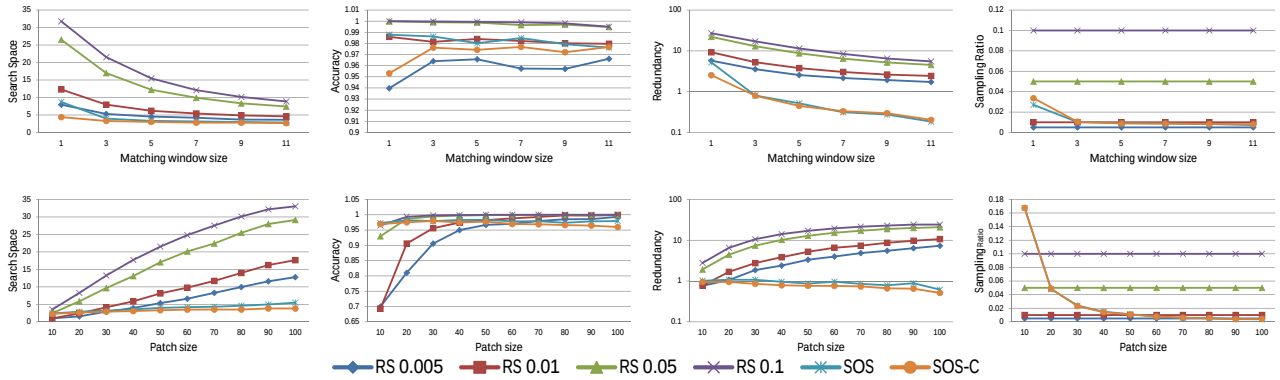
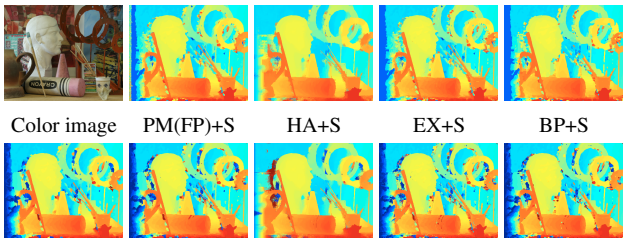


Figure 5. Sequential optimal sampling (SOS) vs. random sampling (RS) for various matching window size (top) and sampling neighborhood size (bottom). Columns 1 to 3: cardinality, accuracy, redundancy of the reduced search spaces, and Column 4: sampling ratio.



Color image PM(FP)+S HA+S EX+S BP+S

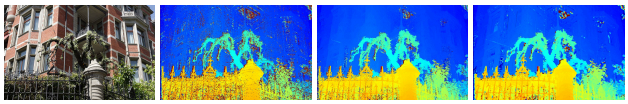
SOS$^+$(FP) PM(FP) HA EX BP

Figure 7. Raw disparity maps for various stereo methods.

Time (s)	SOS$^+$(FP)	PM(FP)+S	PM(FP)	HA+S	HA	EX+S	EX	BP+S	BP
Tsukuba	**0.33**	1.76	1.91	1.40	1.80	2.54	5.13	7.88	34.00
Venus	**0.41**	2.49	2.80	2.10	3.09	3.81	9.48	14.57	78.30
Teddy	**0.89**	2.93	3.28	2.65	6.41	5.20	24.61	35.03	652.95
Cones	**0.88**	2.89	3.29	2.72	6.38	5.18	23.05	41.32	649.59
Art	**1.13**	3.26	3.72	3.70	8.21	7.76	31.22	102.38	1221.36
Books	**0.94**	3.11	3.5	3.06	8.13	6.27	32.34	58.33	1192.62

Table 1. Processing time for various stereo methods.

Figure 7 shows samples of raw disparity maps generated by the various stereo algorithms, and the corresponding processing times are listed in Table 1. We observe no significant quality loss between stereo algorithms under reduced and entire search spaces, while the processing time on reduced spaces is smaller than using the entire space, for PM(85%), HA(50%), EX(20%), and BP(6%). In principle, the computational overhead of SOS may be comparable to a real-time stereo method. Hence, slower global methods will gain the most speedup benefits. Since the accuracy of SOS search space reduction is above 95% (Fig 5) and can be tuned through parameter manipulation.Note that SOS$^+$(FP) evaluates on reduced search spaces corresponding to local structures (exploited by optimal sampling), which will converge quickly and many pixels just receive the propagated disparity values without any matching cost computation. These results indicate SOS is more efficient than sampling methods not exploiting local scene structure.

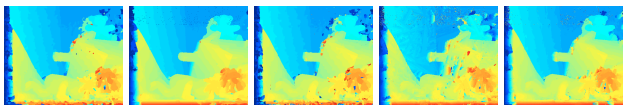


Color image PM PM(FP)+S SOS$^+$(FP)

Figure 8. Raw disparity maps for high resolution ($21M$) images.

We also compare SOS$^+$(FP) and PM(FP) on the high resolution ($21M$) images of Kim *et al.* [6] with a large candidate disparities set of 250 disparities. After our optimal sampling, the average search space is reduced to 8.4 disparities. Figure 8 shows the raw disparity map for PM(FP), PM(FP)+S, and SOS$^+$(FP), with the corresponding processing time 534s, 419s, and 167s. We also can see PM has many outliers around the bush regions which have been successfully removed by our sampling so that PM+S has much fewer outliers. While the goal of our SOS sampling scheme is to enable attainment of the same results as exhaustive disparity search from a reduced search space, in this case increased accuracy is a byproduct of our optimal local structure estimates due to the removal of ambiguous and wrong disparities.

Stereo with Oriented Plane In the next experiment, we investigate the effect of aggregating matching cost across oriented planes by comparing our SOS$^+$ algorithm against PatchMatch. The aggregation window for both method is 31×31 (similar as the default value in [1]). Two types of PM are tested: the first one, PM(NPR), propagates the randomly initialized planes, and the second one (PM) incorporates the iterative plane refinement [1]. The raw disparity maps are shown in Figure 9, and the corresponding processing times are: SOS$^+$(FP) 0.89s, SOS$^+$ 17.41s, PM(FP) 3.28s, PM(NPR) 220.54s, and PM 747.97s. SOS$^+$ is able to account for the slanted surfaces (the ground in teddy), but is slower than the fronto-parallel version SOS$^+$(FP). Without the iterative plane refinement step, there are many ambiguous regions that can not be recovered by PM's propagation scheme, but recovered by our SOS$^+$.



SOS$^+$(FP) SOS$^+$ PM(FP) PM(NPR) PM

Figure 9. Raw disparity maps comparison for SOS and PM with their fronto-parallel version SOS$^+$(FP) and PM(FP), and PM(NPR) is the PM without plane refinement.

	Tsukuba (nocc,all)	Venus (nocc,all)	Teddy (nocc,all)	Cones (nocc,all)	APBP (%)
SOS$^+$	(1.45,**1.63**)	(**0.21**,0.32)	(3.13,8.45)	(**2.43**,**7.10**)	**4.30**
PM[1]	(2.09,2.33)	(0.21,0.39)	(**2.99**,**8.16**)	(2.47,7.80)	4.59
SOS$^+$(FP)	(1.58,1.81)	(0.21,**0.31**)	(5.67,11.0)	(2.57,7.70)	5.37
NLF[19]	(1.47,1.85)	(0.25,0.42)	(6.01,11.6)	(2.87,8.45)	5.48
AW[20]	(**1.38**,1.85)	(0.71,1.19)	(7.88,13.3)	(3.97,9.79)	6.67
SG [4]	(3.26,3.96)	(1.00,1.57)	(6.02,12.2)	(3.06,9.75)	7.50
SDDS[18]	(3.31,3.62)	(0.39,0.76)	(7.65,13.0)	(3.99,10.00)	7.19
HA[8]	(2.47,2.71)	(0.74,0.97)	(8.31,13.8)	(3.86,9.47)	7.33

Table 2. Disparity map evaluation for non occlusion(nocc), all regions, and average percent bad pixels (APBP).

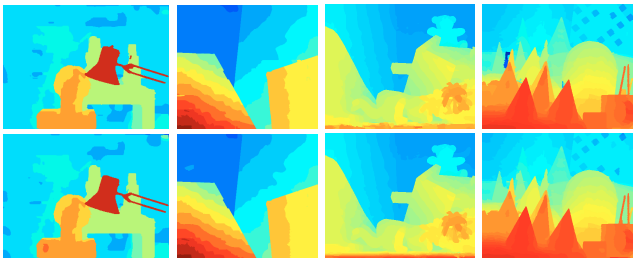


Figure 10. Refined output for SOS$^+$(FP) (top) and SOS$^+$ (bottom).

Evaluation for Refined Disparity Maps Figure 10 shows the refined disparity maps for our constrained SOS$^+$(FP) and SOS$^+$ algorithms, and the quality evaluation are listed in Table 2, with the rank 23 and 10 in the Middlebury benchmark for SOS$^+$ and SOS$^+$(FP) respectively. The quality of the SOS$^+$(FP) algorithm is similar to PatchMatch (rank 22) with main differences coming from the ground region of the teddy image, which can only be recovered by using oriented planes. However, the processing time of SOS(FP) is much faster than other investigated stereo algorithms. In practice, SOS$^+$ is more suitable for the scene with large slanted surfaces, and SOS$^+$(FP) is more efficient for time-sensitive applications.

6. Conclusion

We introduced a novel approach to reduce the disparity search space for stereo based on the Sequential Ratio Probability Test from the sequential decision theory. Our method avoids unnecessary evaluation of irrelevant disparities for pixels of an image. Moreover, our method can be combined with a large variety of existing stereo estimation methods. The propagation-based stereo scheme integrated with the SOS is more efficient than state-of-art stereo methods. As shown in our experimental evaluation, our method maintains the quality of the exhaustive disparity estimation at significantly lower computational costs.

Acknowledgement Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government.

References

[1] M. Bleyer, C. Rhemann, and C. Rother. Patchmatch stereo - stereo matching with slanted support windows. In *BMVC*, 2011. 1, 2, 3, 6, 7, 8

[2] O. Chum and J. Matas. Optimal randomized ransac. *PAMI*, 2008. 3

[3] G. Dial and J. Grodecki. Rpc replacement camera models. In *ASPRS*, 2005. 1

[4] H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *CVPR*, 2005. 8

[5] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2008. 5

[6] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.* 7

[7] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. In *GPUCV*, 2011. 2

[8] D. Min, J. Lu, and N. D. Minh. A revisit to cost aggregation in stereo matching. In *ICCV*, 2011. 1, 2, 6, 8

[9] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm. Usac: A universal framework for random sample consensus. *IEEE Trans. Pattern Anal. Mach. Intell.* 5

[10] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *CVPR*, 2011. 2

[11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 2002. 6

[12] M. Sizintsev. Hierarchical stereo with thin structures and transparency. In *CRV*, 2008. 2

[13] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *ICCV*, 2003. 2

[14] G. Van Meerbergen, M. Vergauwen, M. Pollefeys, and L. Van Gool. A hierarchical symmetric stereo algorithm using dynamic programming. *Int. J. Comput. Vision*, 2002. 2

[15] O. Veksler. Reducing search space for stereo correspondence with graph cuts. In *BMVC*, 2006. 2

[16] A. Wald. *Sequential Analysis*. Dover, 1947. 3, 4

[17] L. Wang, H. Jin, and R. Yang. Search space reduction for mrf stereo. In *ECCV*, 2008. 2

[18] Y. Wang, E. Dunn, and J.-M. Frahm. Increasing the efficiency of local stereo by leveraging smoothness constraints. In *3DIMPVT*, 2012. 2, 8

[19] Q. Yang. A non-local cost aggregation method for stereo matching. In *CVPR*, 2012. 2, 8

[20] K. J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. PAMI*, 28:650–656, 2006. 2, 8