

Adaptive Scale Selection for Hierarchical Stereo

Yi-Hung Jen
dedanann@cs.unc.edu

Enrique Dunn
dunn@cs.unc.edu

Pierre Fite-Georgel
pierre.georgel@gmail.com

Jan-Michael Frahm
jmf@cs.unc.edu

3D Computer Vision Group
Computer Science Department
University of North Carolina
Chapel Hill, NC, USA

Hierarchical stereo provides an efficient coarse-to-fine mechanism for disparity map estimation. However, common drawbacks of such an approach include the loss of high frequency structures not observable at coarse scale levels, as well as the unrecoverable propagation of erroneous disparity estimates through the scale space. This paper presents an adaptive scale selection mechanism to determine a suitable resolution level from which to begin the hierarchical depth estimation process for each pixel. The proposed scale selection mechanism allows us to robustly implement variable cost aggregation to reduce the variability of the photo-consistency measure across scale space. We also incorporate a weighted shiftable window mechanism to enable error correction during coarse to fine depth refinement. Our approach is aimed at

- recovering detailed structure lost during successive convolution and sub-sampling in the generation of a discrete scale space.
- resolving ambiguity in the initialization of succeeding levels from the up-sampled disparity image.

Variable Cost Aggregation : We use a weighted cost function which models surface membership priors. The weights are Gaussian determined based on color dissimilarity and the 2D distance to the center pixel (x_c, y_c) within an image template. They are define as follows:

$$W_f(x, y, I) = \exp\left(-\frac{\sqrt{\sum_{c=1}^h (I(x_c, y_c, c) - I(x, y, c))^2}}{k_c}\right) \cdot \exp\left(-\frac{\sqrt{(x-x_c)^2 + (y-y_c)^2}}{k_d}\right),$$

where k_c and k_d are weight parameters. The similarity measure used this weight combined with an L_1 norm:

$$C(x_c, y_c, d) = \sum_{(x, y) \in \mathcal{N}(x_c, y_c)} W_f(x, y, I_s) \cdot W_f(x + d, y, I_t) \cdot \|I_s - I_t\|,$$

where d corresponds to the disparity. Under this formulation, weighting the contribution by color distance conforms to the assumption that objects lying on the same surface are similar in color; whereas weighting by proximity models the correlation of disparity from neighboring pixels to the center pixel on the matching template.

Adaptive Scale Selection We select the most favorable scale to perform disparity search, for this we quantify the saliency of the surface prior at each level and search for an extremum in the scale space using a Laplacian of Gaussian (*LoG*) [1] on the weight image. We empirically demonstrate that the scale selection based on our weight function was more robust than directly using the color information.

Variable Cost Shiftable Window In order to obtain smooth disparity map, we compute the final disparity based on a local consensus in the spirit of shiftable windows [3]. An initial disparity estimate $\tilde{d}(c_x, c_y)$ is computed using the weighted cost function C and is subsequently replaced by the depth estimate of a pixel in their neighborhood if the weighted cost of that neighboring pixel is better. We express this updated depth as

$$d(c_x, c_y) = \tilde{d}\left(\arg \min_{(n_x, n_y) \in \mathcal{N}(c_x, c_y)} \frac{C(n_x, n_y, \tilde{d}(n_x, n_y))}{W_f(n_x, n_y, I_s) \cdot W_f(n_x + d, n_y, I_t)}\right)$$

Results Overview Scale selection enables fine structure recovery while providing a performance/cost trade-off between hierarchical stereo and full resolution processing. This tradeoff is dependent on the distribution

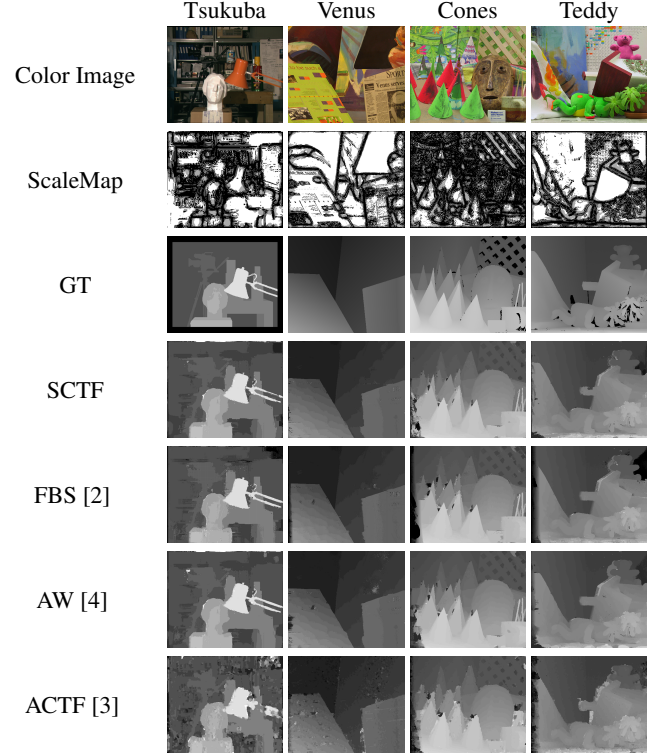


Figure 1: Disparity maps and scale maps for the approaches considered.

of scales within the scale map being utilized. Our weighted shiftable window enables robust error recovery during coarse-to-fine depth estimation. Both of these mechanisms are implemented within the context of variable cost aggregation.

- [1] Tony Lindeberg. Feature detection with automatic scale selection. In *International Journal of Computer Vision*, 1998.
- [2] Stefano Mattoccia, Simone Giardino, and Andrea Gambini. Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering. In *ACCV*, 2009.
- [3] Mikhail Sizintsev and Richard P. Wildes. Efficient stereo with accurate 3d boundaries. In *BMVC*, 2006.
- [4] Kuk-Jin Yoon and In So Kweon. Adaptive support-weight approach for correspondence search. *PAMI*, 28:650–656, 2006.

	Tsukuba			Venus			Cones		
	NOCC	ALL	DISC	NOCC	ALL	DISC	NOCC	ALL	DISC
SCTF	2.37	4.05	9.91	1.50	2.49	10.02	4.44	12.29	10.32
FBS	2.95	4.75	8.69	1.29	2.87	7.62	5.23	15.3	11.34
AW	1.86	3.67	8.26	2.09	3.20	11.89	4.49	12.59	9.98
ACTF	8.19	10.22	25.59	7.82	9.02	33.18	7.76	17.36	20.91

Table 1: Benchmark data for the considered cost aggregation methods. We represent the percentage of incorrect pixels in disparity map under different scenarios: NOCC concerns all pixel that are not occluded, ALL takes all pixels into consideration, and DISC considers only pixels around discontinuities. The proposed method is named SCTF.