Cost Aggregation with Anisotropic Diffusion in Feature Space for Hybrid Stereo Matching

YONSEI UNIVERSITY
DIGITAL IMAGE MEDIA LAB

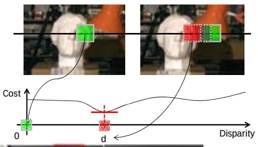
Bumsub Ham*, Dongbo Min**, Kwanghoon Sohn*
Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea*
Advanced Digital Science Center, Singapore**

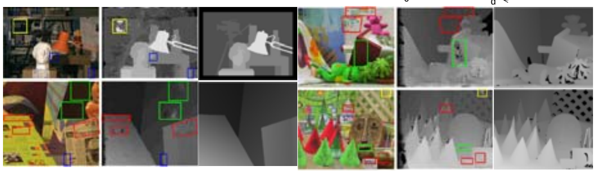
ICIP 2011

1. INTRODUCTION

Stereo Matching

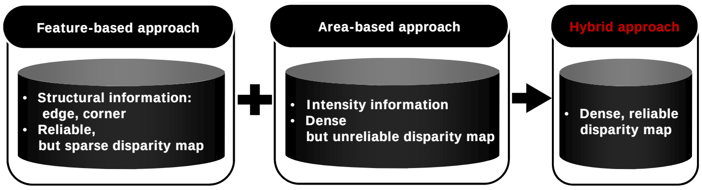
- Objective;
  - Find the correspondence points between stereo images
- Challenge:
  - Matching ambiguities at homogeneous and occlusion regions





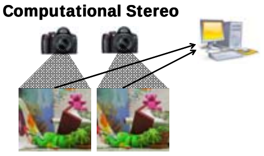
: Depth Discontinuities : High Textured Regions : Low Textured Regions : Repetitive Patterns : Occlusions

Motivation



2. DESIGN OF CORRESPONDENCE MATCHING

Computational Stereo



Modeling; Monocular and binocular cues

Impossible to perfectly model several depth cues!

Design of Correspondence Matching

It was shown that depth can be perceived in the absence of monocular depth and familiarity cues and of all binocular cues except for disparity.

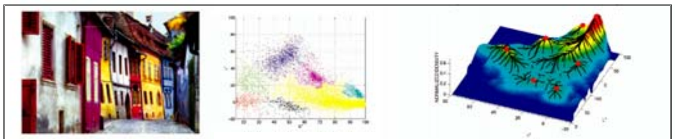
The correspondence of objects and patterns in the two retinal projections can be established with actual recognition of the objects and patterns. This pattern matching is based on some relatively simple processes of **finding *connected clusters* formed by adjacent points of similar brightness.**

[B. Julesz, "Binocular Depth Perception without Familiarity Cues," *Science*, 1964]

3. FEATURE SPACE

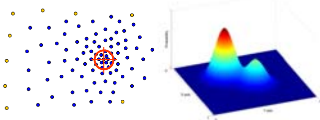
Feature Space

- **Feature space**; Connected clusters formed by neighboring points with similar brightness
- **Assumption**; **Two points are more likely to have similar depth if two feature spaces are similar.**
- Feature space have arbitrary shapes and sizes → HARD TO MEASURE A SIMILARITY



[D. Comaniciu and P. Meer, "Mean Shift: A Robust Aproach Toward Feature Space Analysis, *IEEE Trans. Pattern Anal. and Mach. Intell.*, 2002.]

- **Alternative way;**
  Compare a representative value
  → **local mode, i.e., local maxima in local histogram**



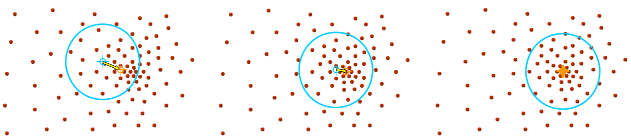
Feature Space Similarity (Coherence)

- **Two feature spaces are coherence if they are similar.**
- Finding local mode; mean shift:
  Iteratively computes mean vector, followed by the translation of the kernel by using the mean shift vector.

$$\mathbf{I}_\mathbf{p} = (\mathbf{p}^\mathrm{T}, \mathbf{c}_\mathbf{p}^\mathrm{T})^\mathrm{T} \quad \mathbf{p} = (x, y)^\mathrm{T} \quad \mathbf{c}_\mathbf{p} = (L, a, b)^\mathrm{T}$$

$$\mathbf{m}(\mathbf{I}_\mathbf{p}) = \frac{\sum_{\mathbf{s}\in N_\mathbf{p}} \mathbf{I}_\mathbf{s}\, g_p(\|\mathbf{p}-\mathbf{s}\|)\, g_c(\|\mathbf{c}_\mathbf{p}-\mathbf{c}_\mathbf{s}\|)}{\sum_{\mathbf{s}\in N_\mathbf{p}} g_p(\|\mathbf{p}-\mathbf{s}\|)\, g_c(\|\mathbf{c}_\mathbf{p}-\mathbf{c}_\mathbf{s}\|)} - \mathbf{I}_\mathbf{p}$$

$$\tilde{\mathbf{I}}_\mathbf{p} = (\tilde{\mathbf{p}}^\mathrm{T}, \tilde{\mathbf{c}}_\mathbf{p}^\mathrm{T})^\mathrm{T}$$



- **Feature space:** $\mathbf{F}_\mathbf{p} = (\mathbf{p}^\mathrm{T}, \tilde{\mathbf{c}}_\mathbf{p}^\mathrm{T})^\mathrm{T}$

4. COST AGGREGATION VIA ANISOTROPIC DIFFUSION

Motivation; Anisotropic Diffusion

Disparity gradient is a function of feature similarity, i.e., more dissimilar features allows larger disparity gradients.

$$\|\nabla \mathbf{D}\| \propto 1/h(\|\nabla \mathbf{F}\|)$$

[K. Prazdny, "Detection of Binocular Disparities," *Biol. Cybern.*, 1985.]

- Observation;
  - It coincides with our assumption: **Two points are more likely to have similar depth if two feature spaces are similar.**
  - The role of the function $h(x)$ is the same as the "edge-stopping" function (monotonically decreasing function) in anisotropic diffusion.

4. COST AGGREGATION VIA ANISOTROPIC DIFFUSION

Cost Aggregation via Anisotropic Diffusion on Feature Space

- $\mathbf{E}$; 2D cost plane (a section of initial 3D matching cost volume)
- R, T; Reference and target image

$$\frac{\partial \mathbf{E}}{\partial t} = \nabla \cdot \left(g_f(\|\nabla \mathbf{F}^\mathrm{R}\|)\, g_f(\|\nabla \mathbf{F}^\mathrm{T}\|) \nabla \mathbf{E} \right)$$

Inter-coherence = Intra-coherence of the reference image ✖ Intra-coherence of the target image

- **Two points are likely to have similar depth spaces as both intra-coherences are high.**
- It also means that they **belong to the similar feature space** in both images.

- Discretization
  $\mathbf{q} = \mathbf{p} + \mathbf{d}, \mathbf{r} = \mathbf{s} + \mathbf{d}$; the corresponding points of $\mathbf{p}$ and $\mathbf{s}$ in the reference image with $\mathbf{d} = (d, 0)^\mathrm{T}$

$$\mathbf{E}^{t+1}(\mathbf{p},\mathbf{d}) = \mathbf{E}^t(\mathbf{p},\mathbf{d}) + \lambda \sum_{\mathbf{s}\in N_\mathbf{p}} g_f(\|\nabla \mathbf{F}^\mathrm{R}(\mathbf{s},\mathbf{p})\|)\, g_f(\|\nabla \mathbf{F}^\mathrm{T}(\mathbf{r},\mathbf{q})\|) \nabla \mathbf{E}^t(\mathbf{s},\mathbf{p})$$

$$\nabla \mathbf{F} \approx \mathbf{F}_\mathbf{s} - \mathbf{F}_\mathbf{p} \equiv \nabla \mathbf{F}(\mathbf{s},\mathbf{p})$$

Feature Confidence

- **$\mathbf{F}$ can be referred to as sets of diffused local mode in a feature space.**
- A point $\mathbf{I}_\mathbf{p} = \mathbf{F}_\mathbf{p} = (\mathbf{p}^\mathrm{T}, \tilde{\mathbf{c}}_\mathbf{p}^\mathrm{T})^\mathrm{T}$ which is inherently located at local mode is more reliable. → **distinct feature**

$$\mathbf{E}^{t+1}(\mathbf{p},\mathbf{d}) = \mathbf{E}^t(\mathbf{p},\mathbf{d}) + \lambda \sum_{\mathbf{s}\in N_\mathbf{p}} c_\mathbf{R} c_\mathbf{T}\, g_f(\|\nabla \mathbf{F}^\mathrm{R}(\mathbf{s},\mathbf{p})\|)\, g_f(\|\nabla \mathbf{F}^\mathrm{T}(\mathbf{r},\mathbf{q})\|) \nabla \mathbf{E}^t(\mathbf{s},\mathbf{p})$$

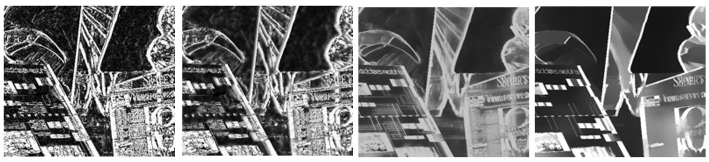
$c_\mathbf{R} = g_p(\|\tilde{\mathbf{s}} - \mathbf{s}\|)\, g_c(\|\tilde{\mathbf{c}}_\mathbf{s}^\mathrm{R} - \mathbf{c}_\mathbf{s}^\mathrm{R}\|)$; feature confidence of the reference image

$c_\mathbf{T} = g_p(\|\tilde{\mathbf{r}} - \mathbf{r}\|)\, g_c(\|\tilde{\mathbf{c}}_\mathbf{r}^\mathrm{T} - \mathbf{c}_\mathbf{r}^\mathrm{T}\|)$; feature confidence of the target image

Advantages

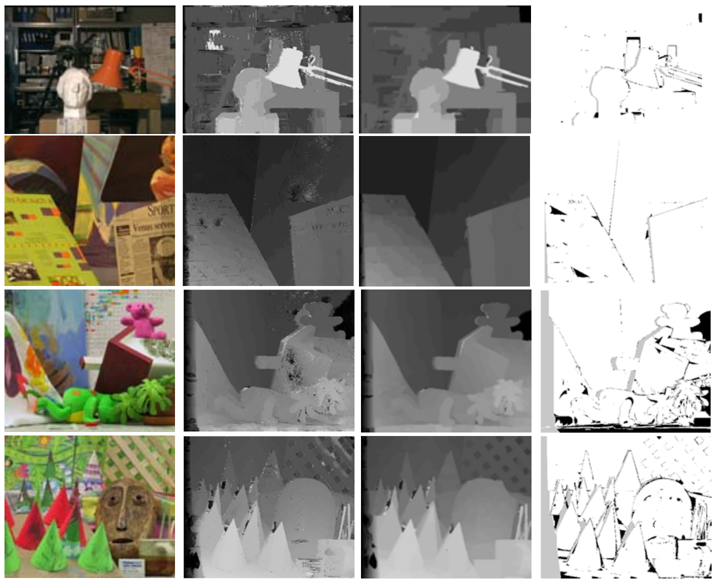
- Similar features are grouped together → **costs vary smoothly within a same depth level.**
- Different weights → according to the relative importance of the features → **better discriminative power** for different depth levels.
- A window can move *dynamically* in constructing the feature space → **propagate the information into neighborhood very well.** → homogeneous regions can be successfully handled.
- The anisotropic diffusion on feature space → a dense feature matching in the viewpoint of cost aggregation → **Hybrid method**

5. EXPERIMENTAL RESULTS



Results of aggregated cost in 'Venus' for (from left to right) initial matching cost, anisotropic diffusion [1], Adaptive weight [2], proposed method when disparity is 0.

- It diffuses pixels inside same depth levels while preventing pixels from being diffused across different depth levels.
- Distinct discrimination is observed across the different depth levels.



Results for (from top to bottom) 'Tsukuba', 'Venus', 'Teddy' and 'Cone'. (from left to right) reference images, adaptive weight [2], proposed method, error maps.

OBJECTIVE EVALUATION [3]

Algorithm	Tsukuba		Venus		Teddy		Cone	
	NonOcc	*Disc*	*NonOcc*	*Disc*	*NonOcc*	*Disc*	*NonOcc*	*Disc*
Segment support	2.28	7.5	1.21	5.88	10.99	22.01	5.42	11.83
Proposed method	1.8	7.27	1.13	4.92	11.2	23.2	5.6	12.4
Adaptive weight	4.66	8.25	4.61	13.3	12.7	22.4	5.5	11.9
Variable Windows	4.1	10.79	10.66	9.94	13.93	25.53	7.24	13.86
Reliability	5.14	18.31	3.86	11.51	16.96	30.62	13.52	21.55

[1] D, Min and K. Sohn, "Cost Aggregation and Occlusion Handling with WLS in Stereo Matching," *IEEE Trans. Image Processing*, 2008.
[2] K. Yoon and I. Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006.
[3] F. Tombari, S. Mattoccia and L. Stefano, "Classification and Evaluation of Cost Aggregation Methods for Stereo Correspondence," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.