HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

Toward perceptually-consistent Stereo: A scanline Study

Jialiang Wang, Daniel Glasner and Todd Zickler
Project site: vision.seas.harvard.edu/stereo

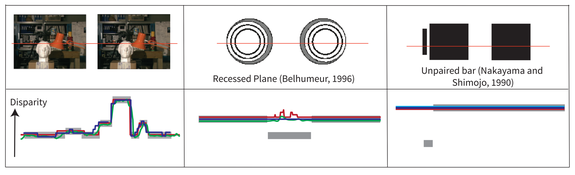
Motivation

Two types of information exist in a stereo pair:

1) matched regions: depth
2) unmatched regions: location and magnitude of depth discontinuities

People can perceive depth even when matching information is absent or very weak:

However, state-of-the-art computer vision systems cannot:

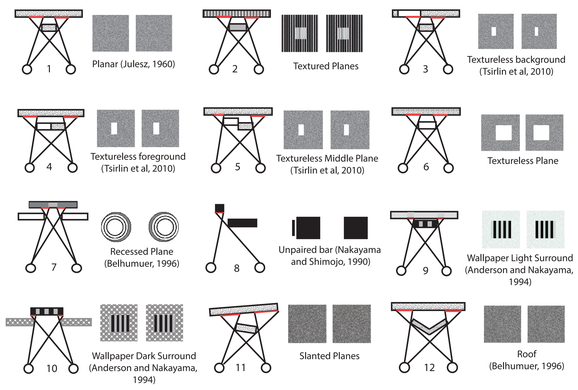


Goal: design a scanline stereo system that

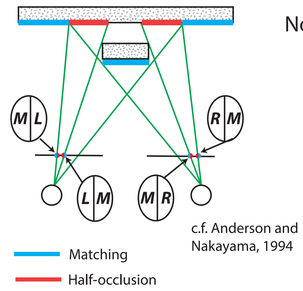
1) works on images with limited matching information
2) improves results in half-occluded regions of natural stereo pairs
3) inspire improvements to future 2D algorithms

Perceptual dataset

vision.seas.harvard.edu/stereo

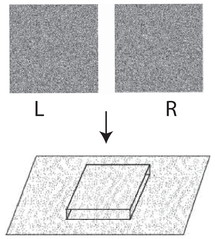


Correlation and Decorrelation Cues



c.f. Anderson and Nakayama, 1994

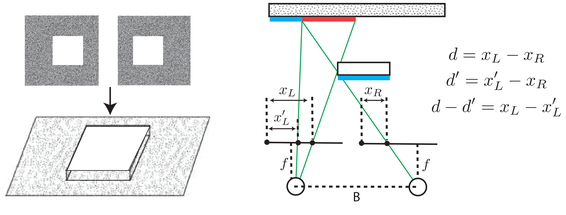
No intensity edge information



In disparity space (cyclopean), we can identify the occluding contours and half-occlusion size using correlation gradients:

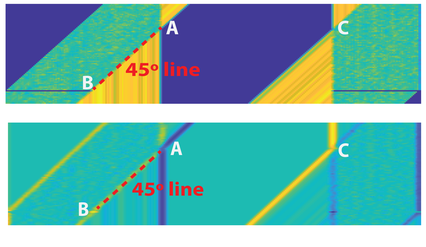
- A, C: occlusion-inducing boundaries
- B, D: half-occlusion boundaries

For images that are not pure random-dots, such as the textureless plane stimuli:



$$d = x_L - x_R$$
$$d' = x'_L - x_R$$
$$d - d' = x_L - x'_L$$

DSI on the left image

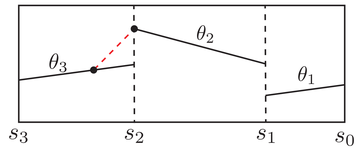


Scanline Formulation

Represent disparity profile as a piecewise smooth function determined by:

(1) breakpoint locations $\mathcal{I} = \{s_0, s_1, \cdots, s_{|\mathcal{I}|}\}$

(2) shape parameters in each smooth piece $\{\theta_i\}$
e. g. piecewise constant: $\theta_i \in \mathbb{R}$
piecewise linear: $\theta_i \in \mathbb{R}^2$



Minimize:

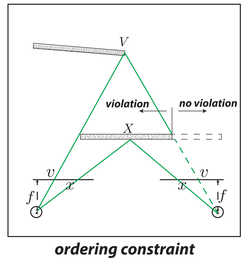
$$\mathcal{L}(\mathcal{I}, \Theta) = \sum_{i=1}^{|\mathcal{I}|} \rho_c(s_{i-1}, s_i, \theta_{i-1}, \theta_i) + \lambda_1 \rho_g(s_{i-1}, \theta_{i-1}, \theta_i) + \lambda_2$$

- ρ_c: Matching cost of segment $\{s_{i-1}, s_i\}$
- ρ_g: Boundary cost of breakpoint s_i
- λ_2: Geometric prior

$$\rho_c(s_{i-1}, s_i, \theta_{i-1}, \theta_i) = \sum_{n=k(s_{i-1}, \theta_{i-1}, \theta_i)}^{s_i - 1} C(n, \theta_i) \qquad k(\cdot) \text{ models } 45^o \text{ geometry}$$

$$\rho_g(s_{i-1}, \theta_{i-1}, \theta_i) = \begin{cases} 1 - G(s_{i-1}, \theta_i), & \text{if } d_{\theta_i}(s_{i-1}) \geq d_{\theta_{i-1}}(s_{i-1}) \text{ (e.g. } s_1\text{)} \\ G(s_{i-1}, \theta_{i-1}) - G(k(s_{i-1}, \theta_{i-1}, \theta_i), \theta_i), & \text{otherwise} \qquad \text{(e.g. } s_2\text{)} \end{cases}$$

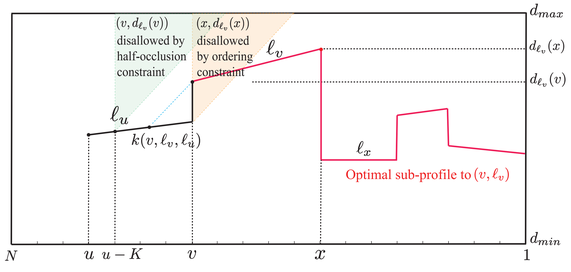
Dynamic Programming Optimization



ordering constraint

Algorithm 1 Find optimal disparity profile

```
for all ℓ ∈ L do
    opt(1, ℓ) ← C̃(1, ℓ) + λ2
    arg(1, ℓ) ← (1, ℓ)
end for
for u ← 2 to N do
    for ℓ_u ← 1 to L do
        Γ ← BuildValidSubset(u, ℓ_u, arg(u, ℓ_u))
        opt(u, ℓ_u) ← min_Γ (ρ(v, u, ℓ_v, ℓ_u) + opt(v, ℓ_v))
        arg(u, ℓ_u) ← arg min_Γ (ρ(v, u, ℓ_v, ℓ_u) + opt(v, ℓ_v))
    end for
end for
I, u ← N                                   ▷ initialize trace back
Θ, ℓ_u ← arg min_ℓ opt(N, ℓ)
while u > 1 do
    (u, ℓ_u) ← arg(u, ℓ_u)
    Append(I, u)
    Append(Θ, ℓ_u)
end while
```

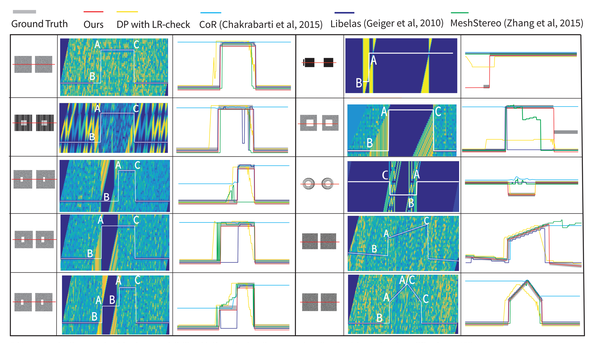


Results

Perceptual dataset:

Our algorithm:
- recovers the correct disparities for 10 of 12 perceptual stereo pairs

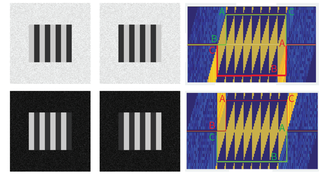
State-of-the-art algorithms:
- can perform descently in some stereo pairs that have matching information
- fail completely when matching information is absent



* Note: vertically stretched for better visibility. Half-occlusion lines are visualized with angles that differ from 45°

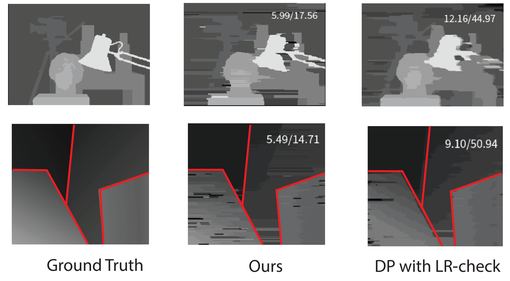
Failure cases:

Green curves: human observations
Red curves: our algorithm



- both explanations have gradient signals that are consistent with our gradient requirements
- perheps could be fixed using monocular cues

Tsukuba and Venus:



Acknowledgement: This work was supported by US NSF IIS-1212928 and IIS-1618227