Fast Multiple-baseline Stereo with Occlusion

Marc-Antoine Drouin
Martin Trudeau
Sébastien Roy
{drouim,trudeaum,roys}@iro.umontreal.ca

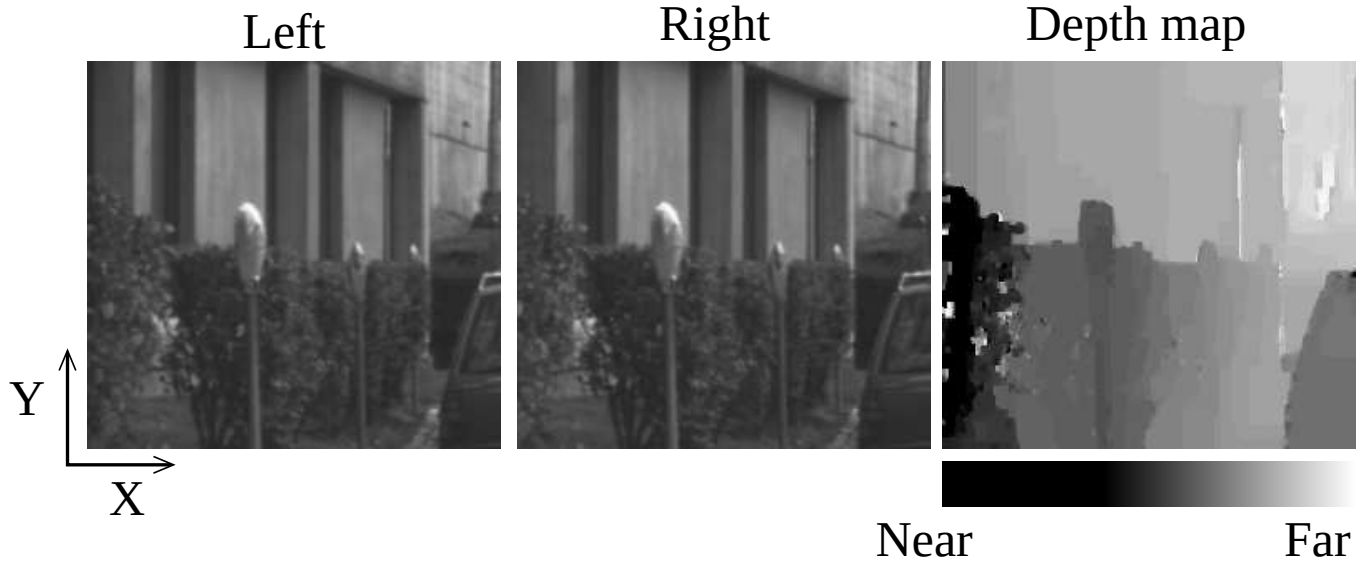
June 2005

Université de Montréal

Overview

- Introduction.
- Previous Works.
- Observation.
- Our Algorithm.
- Experimental Results.
- Conclusion.

Dense Stereo



2 cameras

- For each pixel in the left image we try to find the corresponding pixel in the right image.
- The resulting displacement for that pixel (disparity) relates to the distance between the object and the reference camera.

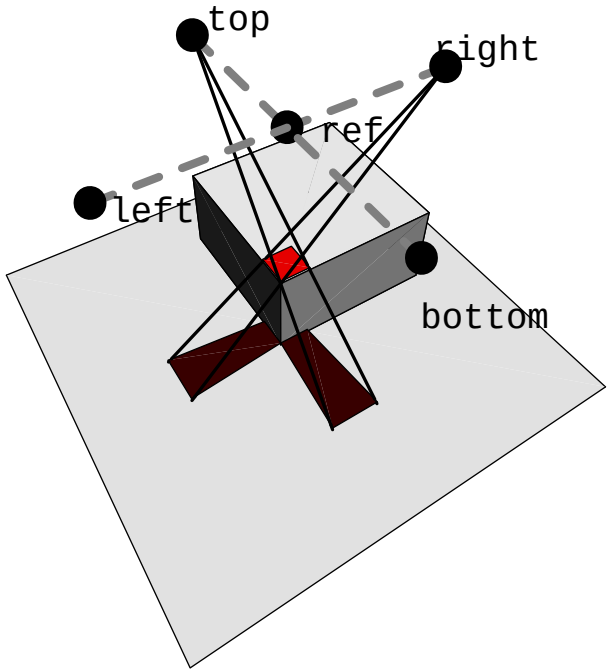
Dense Stereo

$$E(f) = \sum_{\mathbf{p} \in \mathcal{P}} \underbrace{e(\mathbf{p}, f(\mathbf{p}))}_{likelihood} + smoothing.$$

2 cameras

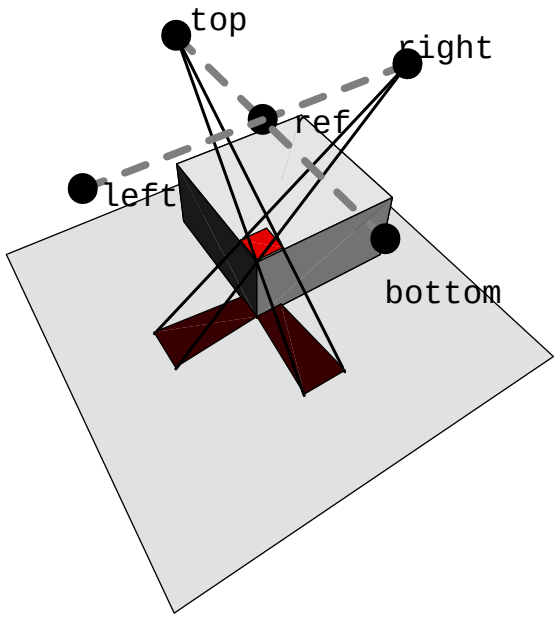
- $\mathcal{P}$: set of reference pixels.
- f : disparity map.
- Hypothesis : for each reference pixel corresponds a supporting pixel.

Camera Configuration



- 5 cameras in cross configuration.
- Disparity map is computed for the central camera.
- In red, examples of occlusion.

Disparity and Visibility Maps



Reference camera

- One disparity for each pixel.
- One visibility mask for each pixel.
- i.e. mask $(0, 0, 0, 1)$.

Multi-camera and Occlusion

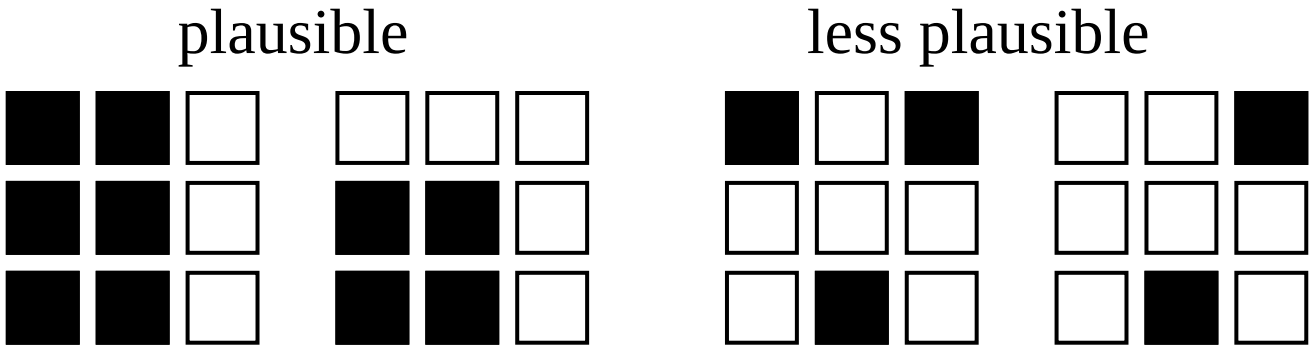
$$E(f,g) = \sum_{\mathbf{p}\in\mathcal{P}} e(\mathbf{p}, f(\mathbf{p}), g(\mathbf{p})) + smoothing.$$

with

$$g(\mathbf{p}) = V(\mathbf{p}|f(\mathbf{p}), f) \quad \forall \mathbf{p} \in \mathcal{P}$$

- f disparity map.
- g visibility mask map.

Nakamura96



Occlusion

- Some masks are very probable.
- Some masks are improbable.
- We can pre-compute a sub-set $\mathcal{M}_h$ of plausible masks.

Nakamura96, Park97, Kang01 and Besnerais04

$$g_f^*(\mathbf{p}) = \arg \min_{\mathbf{m} \in \mathcal{M}_h} e(\mathbf{p}, f(\mathbf{p}), \mathbf{m})\, w(\mathbf{m})$$

then,

$$E(f, g_f^*) = \sum_{\mathbf{p} \in \mathcal{P}} e(\mathbf{p}, f(\mathbf{p}), g_f^*(\mathbf{p})) + smoothing$$

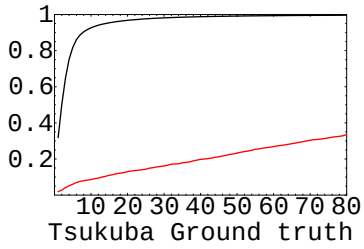
Occlusion

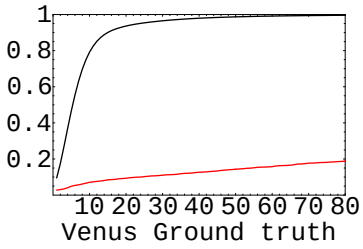
- Hypothesis : photo-consistency $\Rightarrow$ correct visibility.
- Visibility is heuristic.

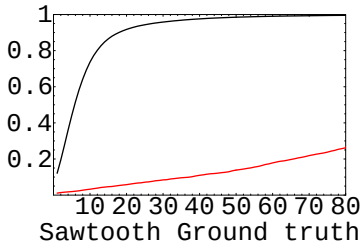
Occlusion Zones in Stereo

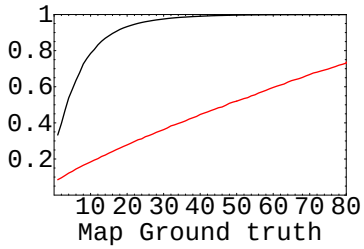
Cumulative histogram of likelihood term

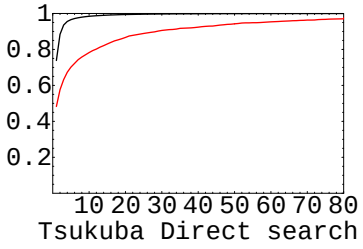
- Black : non-occluded pixels.
- Red :occluded pixels.

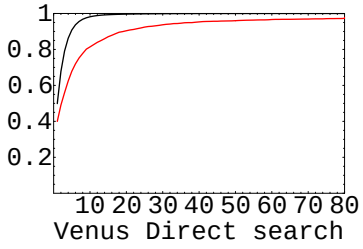


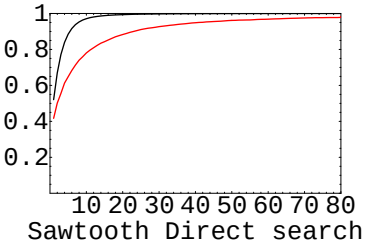


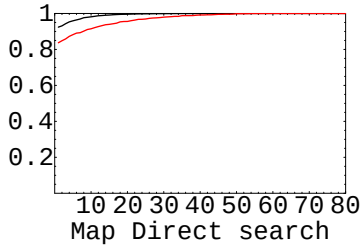












- Photo-consistency $\not\Rightarrow$ geo-consistency.
- Show the limitation of heuristic approaches.

Geo-consistency

All masks are consistent with the scene geometry.

$$g(\mathbf{p}) \;\leq\; V(\mathbf{p}|f(\mathbf{p}), f) \quad \forall \mathbf{p} \in \mathcal{P}$$

Nakamura96

- Using an occluded camera $\Rightarrow$ important artifact.
- Not using a visible camera $\Rightarrow$ no impact.

Kolmogorov02, Faugeras98 and Drouin05

$$E(f,g) = \sum_{\mathbf{p}\in\mathcal{P}} e(\mathbf{p}, f(\mathbf{p}), g(\mathbf{p})) + smoothing$$

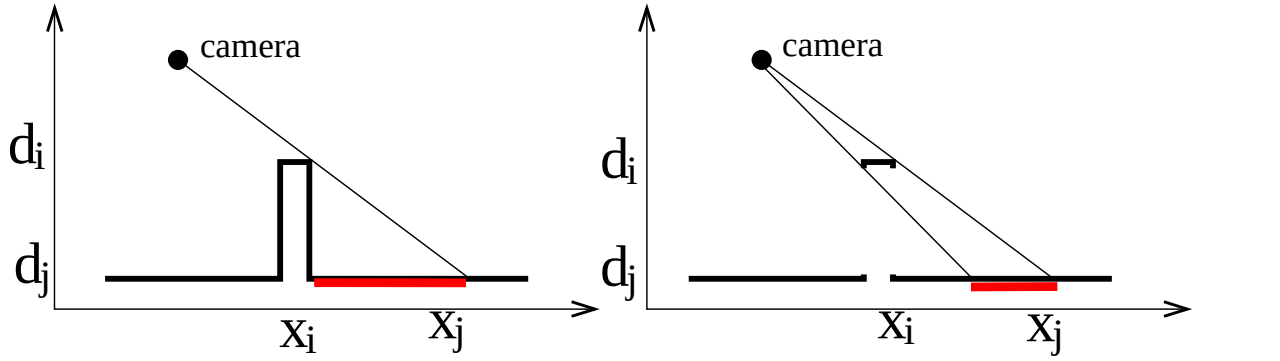
with

$$g(\mathbf{p}) \leq V(\mathbf{p}|f(\mathbf{p}), f) \quad \forall \mathbf{p} \in \mathcal{P}$$

Occlusion

- Kolmogorov : jumps from one geo-consistent configuration to another.
- Faugeras : level set (continuous framework).
- Drouin : starts from a non geo-consistent solution and converges to one which is.
- One common feature : hard to solve.

Disparities and Occlusions



continuous not continuous

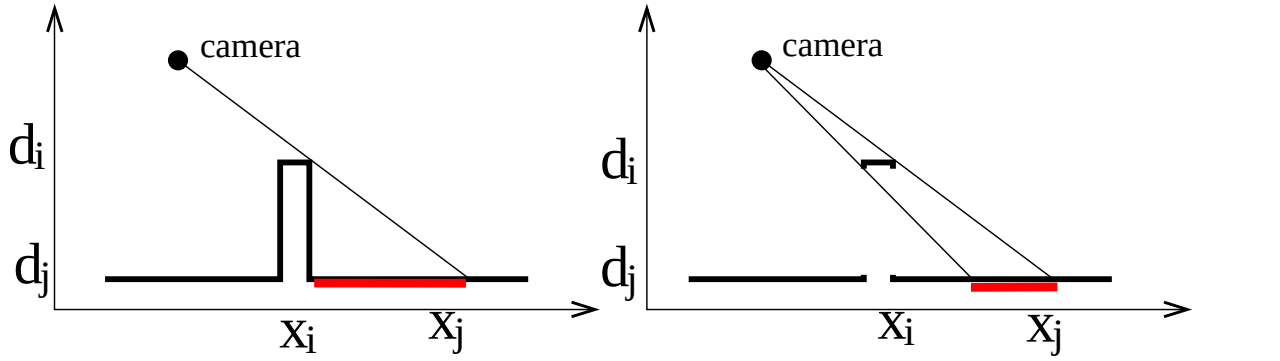
Continuous representation

- Occlusion : $x_i + d_i \geq x_j + d_j$

Discontinuous representation

- Occlusion : $x_i + d_i = x_j + d_j$

Disparities and Occlusions



continuous not continuous

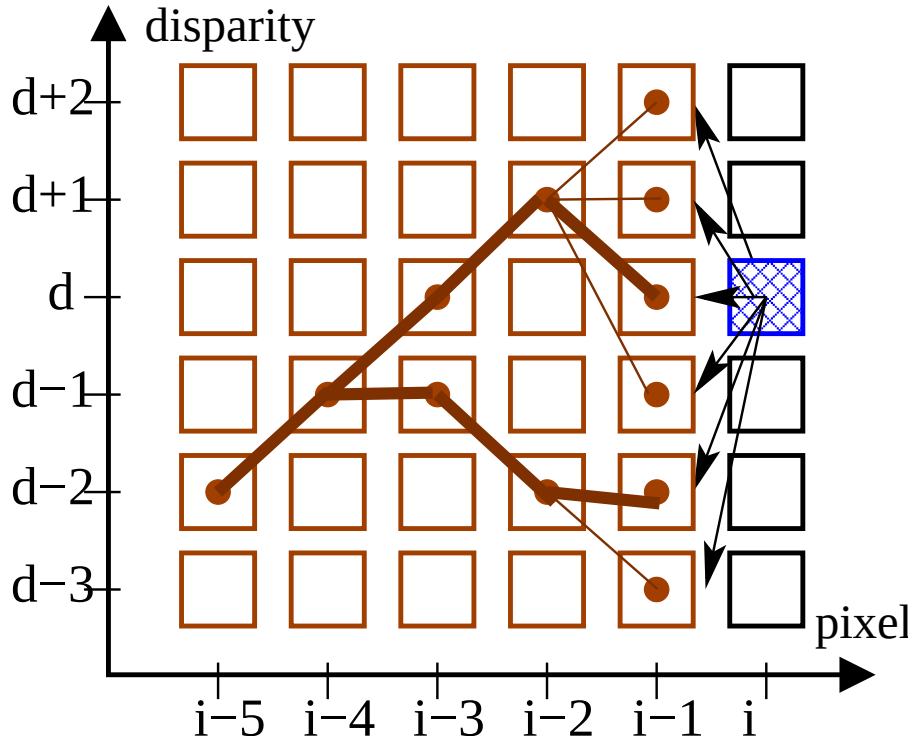
Continuous representation

- Occlusion occurs when

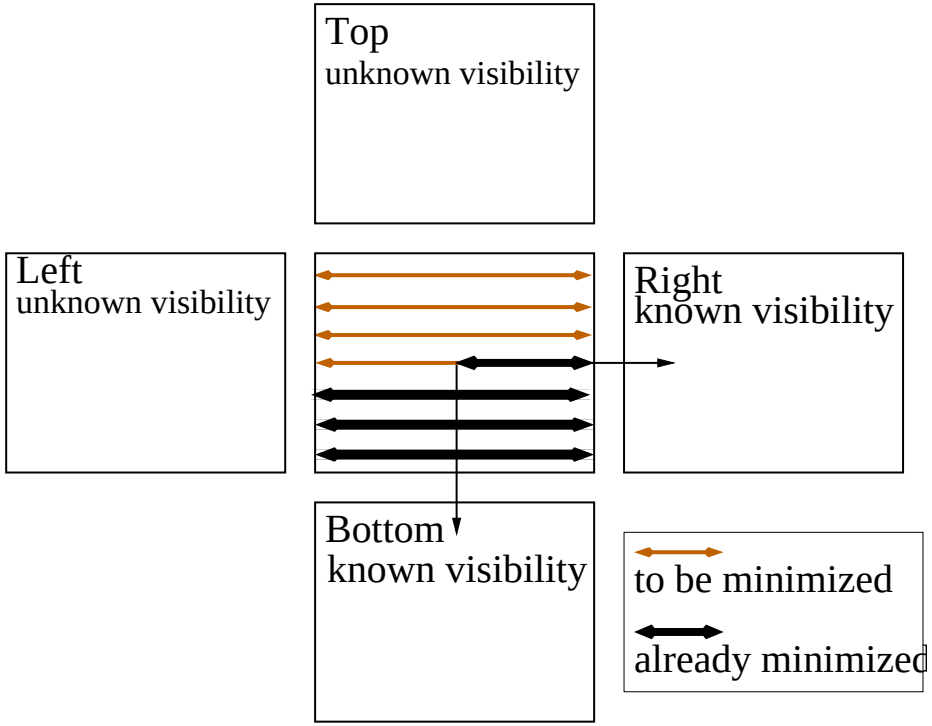
$$\max_{0 \le k < j} (k + d_k) \ge j + d_j$$

- Occlusion at j depends on visibility at $< j$.
- Efficiently computed.

Dynamic Programming



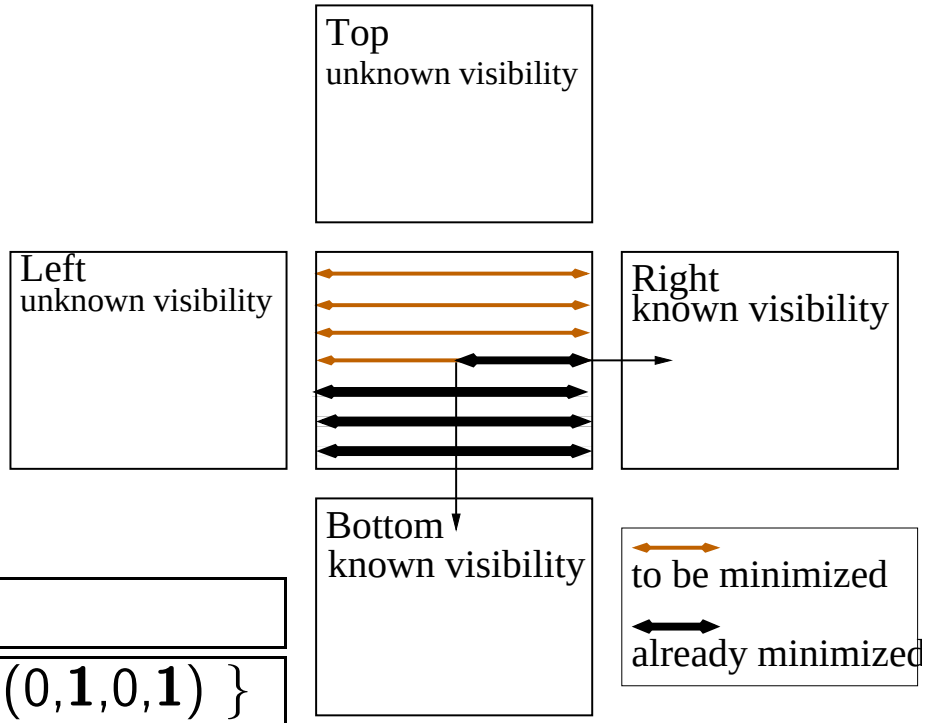
Visibility Masks



Occlusion

- Cameras can be split in 2 sets $\mathcal{C}_G$ and $\mathcal{C}_H$.
- 2 sets of masks are build $\mathcal{M}_G$ and $\mathcal{M}_H$.
- Sets depend on the order in which lines are processed.

Visibility Masks



Masks
$\mathcal{M}_g = \{$ (0,**1**,0,**0**),(0,**0**,0,**1**),(0,**1**,0,**1**) $\}$
$\mathcal{M}_h = \{$ (1,**0**,0,**0**),(0,**0**,1,**0**) $\}$

Occlusion

- Camera order (left, right, top, bottom).
- In bold : cameras belonging to $\mathcal{C}_g$.

Energy Function

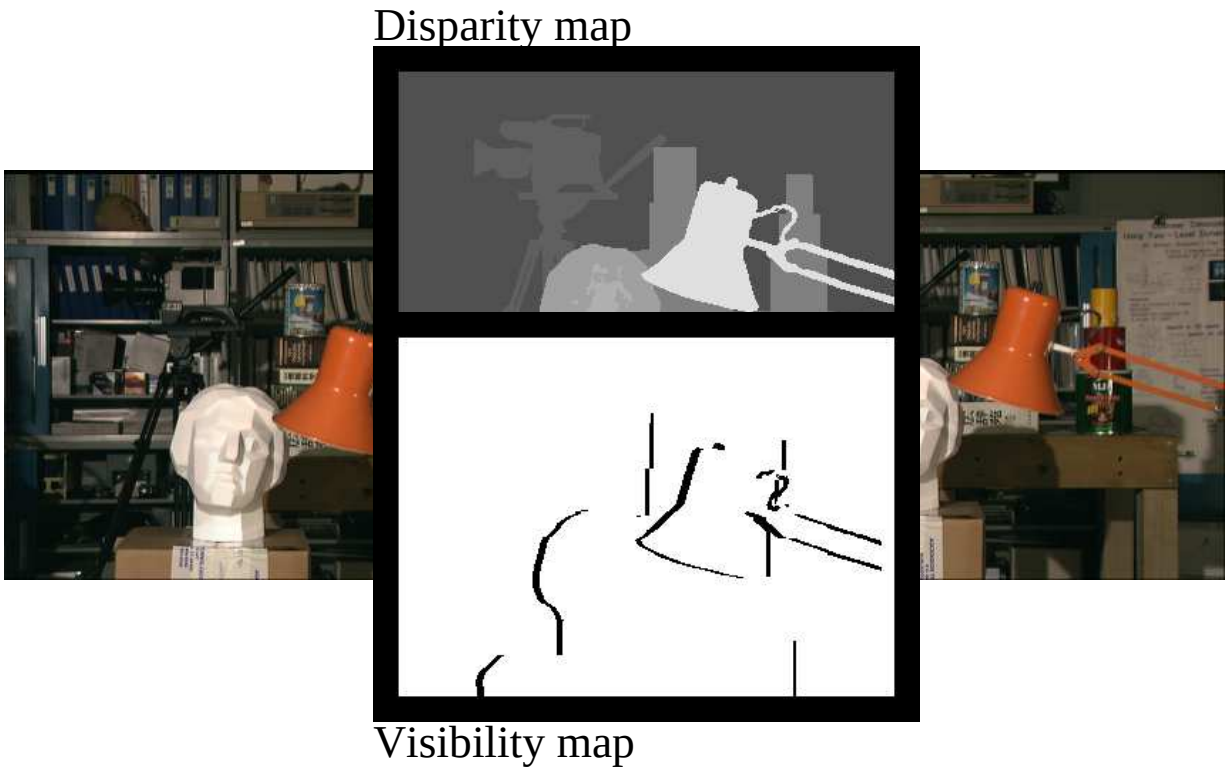
$$E(f,g) = \sum_{\mathbf{p}\in\mathcal{P}} e(\mathbf{p}, f(\mathbf{p}), g(\mathbf{p})) + smoothing$$

with

$$g(\mathbf{p}) = \begin{cases} \text{a mask in } \mathcal{M}_g & \text{if a camera in } \mathcal{C}_g \text{ is visible} \\ \arg\min_{m\in\mathcal{M}_h} e(\mathbf{p}, f(\mathbf{p}), m) & \text{otherwise} \end{cases}$$

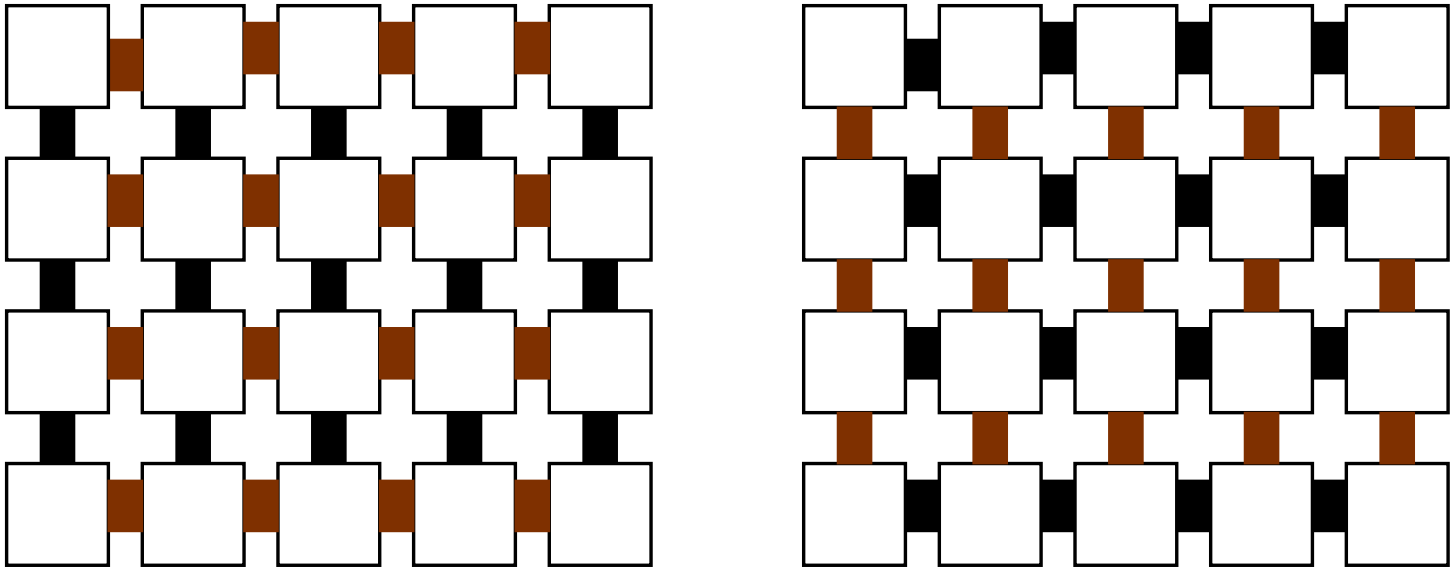
Configuration of low energy.

Disparity and Visibility smoothing



- Difference of depth between two neighbor pixels.
- Change in the set of masks ($\mathcal{M}_h$ and $\mathcal{M}_g$).
- Smoothing function may have any shape.

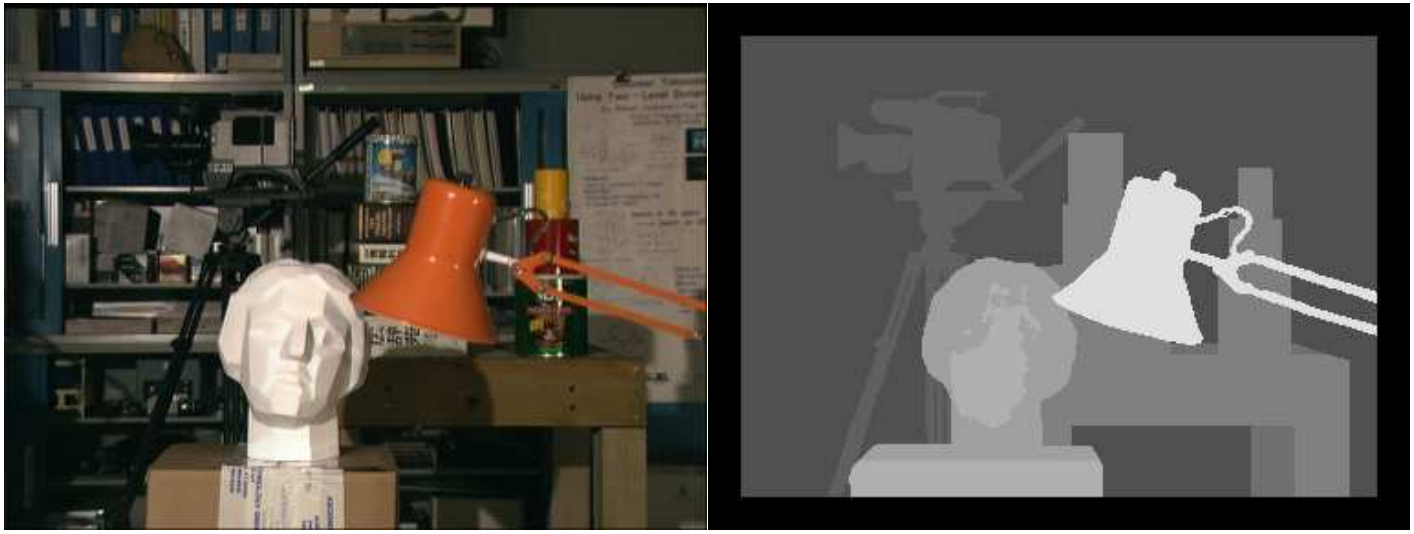
2 Steps Smoothing



Active Smoothing
Passive Smoothing

Iterative Dynamic Programming (Leung04)

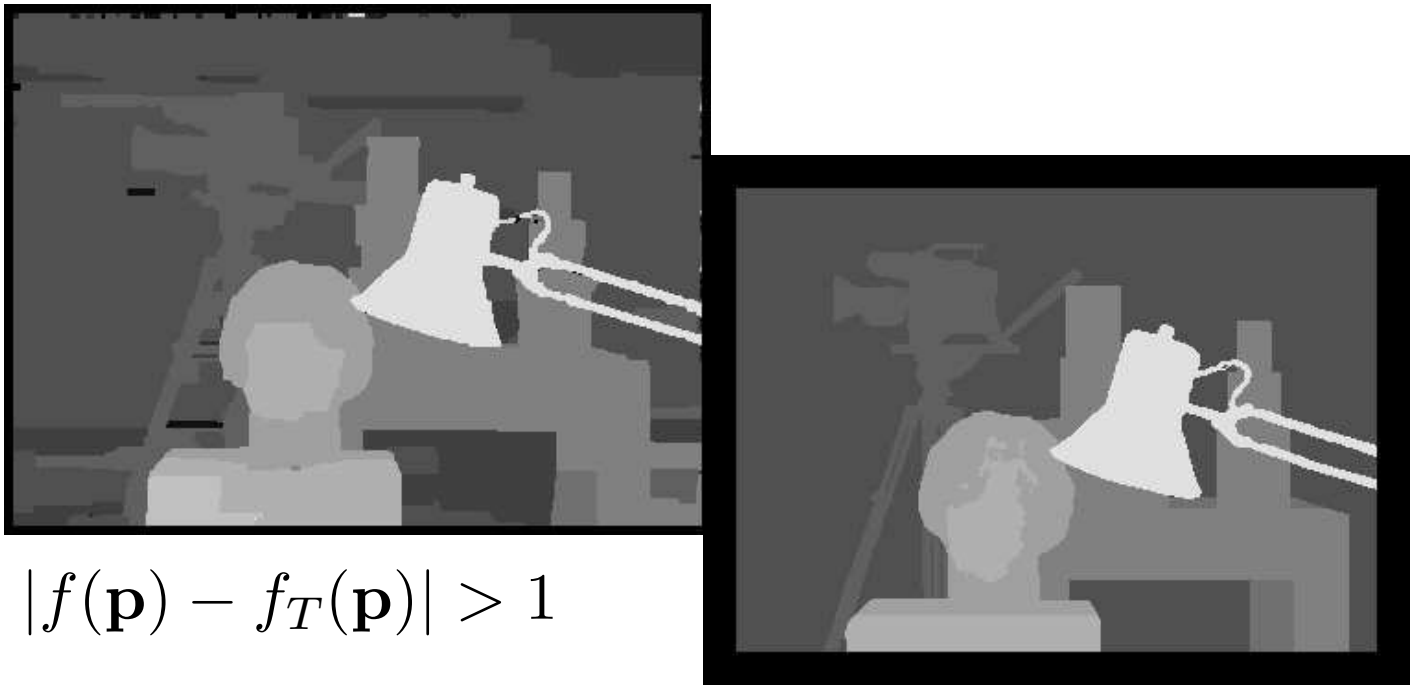
Experimental Results



Tsukuba Head and Lamp

- 384×288 with 16 disparity steps.
- 5 images in cross shape configuration were used.

Experimental Results



$|f(\mathbf{p}) - f_T(\mathbf{p})| > 1$

- An error of 1 could be the result of discretization.
- Standard metric.

Experimental Results

Algorithms	**Error**
Ours + IDP (16 iterations)	1.57%
Ours + IDP (4 iterations)	1.67%
Nakamura96+ Graph Cut	1.77%
Ours + IDP (1 iteration)	1.82%
Kolmogorov02	2.30%
Nakamura96 + IDP (12 iterations)	2.35%
Drouin05 +BNV	2.46%

Experimental Results

Middlebury sequence

- 334×383 with 20 disparity steps.
- 6 scenes with 7 images each in single baseline configuration were used.

Experimental Results

algorithms	Middlebury sequence						average
	barn1	barn2	bull	poster	venus	sawtooth	
Graph Cut (no occlusion)	3.5 %	3.1 %	0.7 %	3.7 %	3.4 %	3.3%	3.0%
IDP (no occlusion)	3.0 %	4.9%	1.2%	6.0 %	5.8%	3.7%	4.1%
Drouin05 +Graph Cut	0.8 %	**0.6 %**	**0.4 %**	**1.1 %**	**2.4 %**	1.1 %	1.3%
Nakamura96 + Graph Cut	1.4 %	1.5 %	0.9 %	**1.1 %**	4.0 %	1.5%	1.7%
Ours +IDP	**0.7 %**	3.9 %	0.8 %	4.0 %	5.3%	**1.0 %**	2.6%
Nakamura96 + IDP	1.6 %	6.0 %	1.9 %	4.5 %	7.4%	2.2 %	3.9%

- The camera configuration is not favorable to our approach.

Experimental Results

Tsukuba sequence

- 320×240 with about 24 disparity steps.
- 5 images in cross shape configuration were used.

Conclusion

Summary

- Hybrid between geo-consistent and heuristic approaches.
- Fast and can easily be parallelized.
- Code can be download from :
  `www.iro.umontreal.ca/~drouim/`

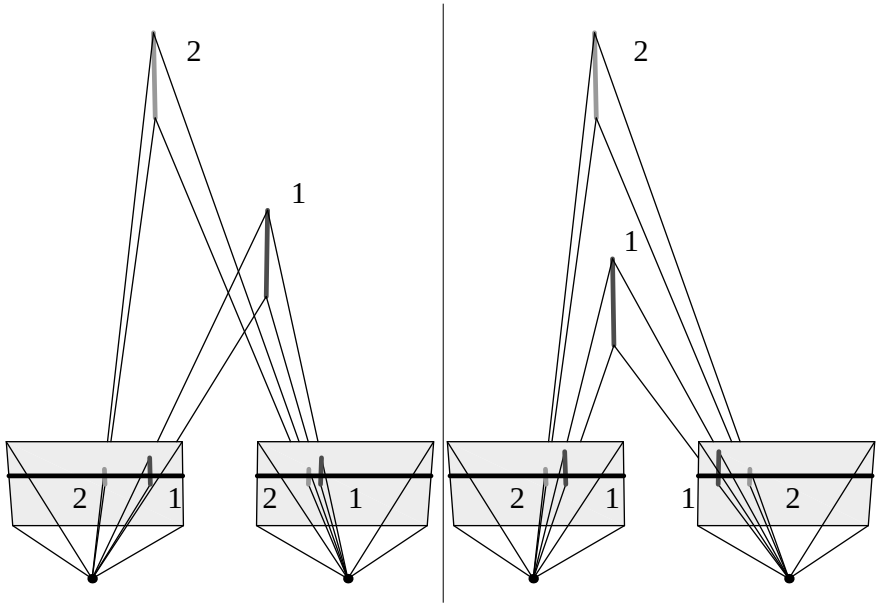
Future work

- Generalizing to arbitrary camera configurations.

Wish list

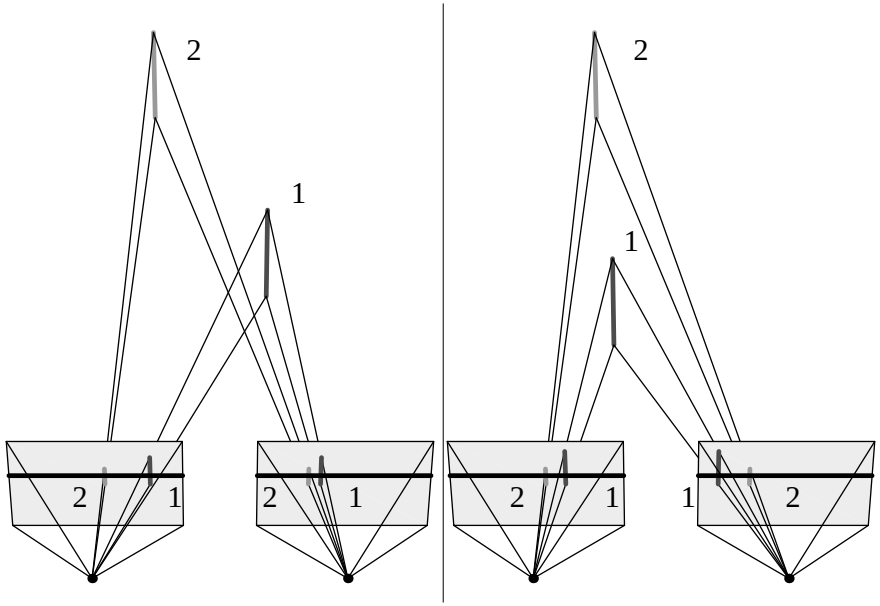
- Designing an hardware implementation in FPGA.

Ordering Constraint



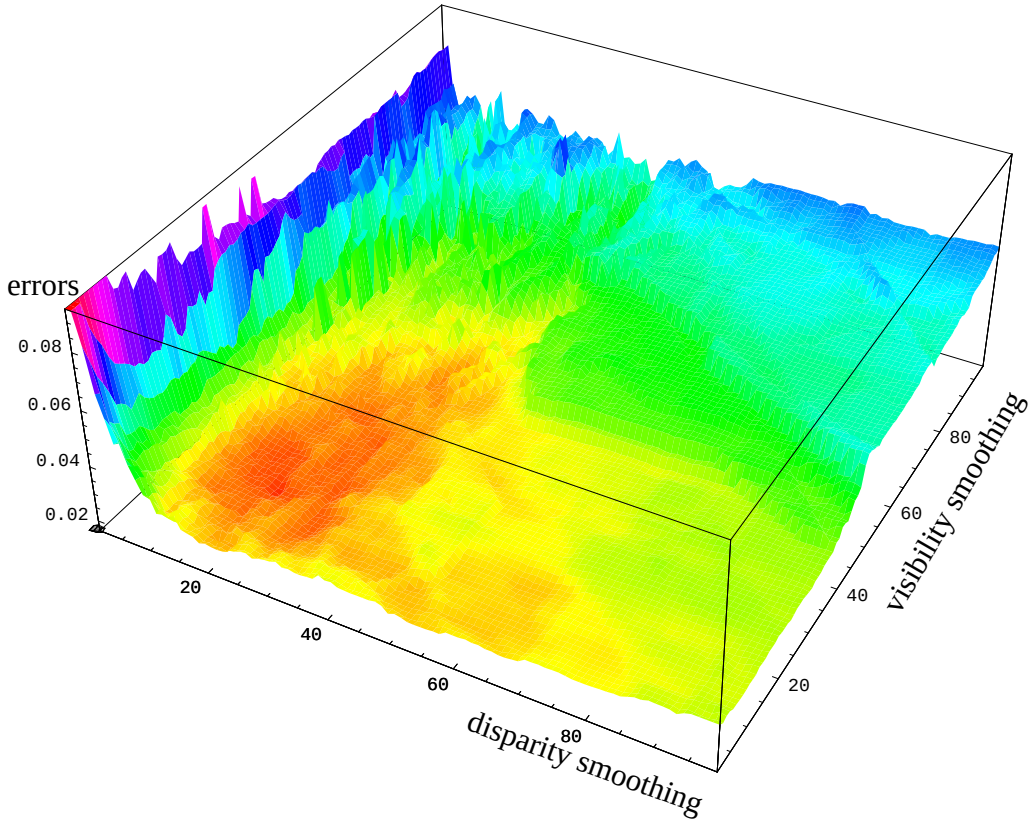
- The order in which two objects are encounter along an epipolar line does not change.
- Not always true.

Ordering Constraint



- Continuous mesh $\Rightarrow$ ordering constraint.
- On the masks but **not** on the geometry.

Experimental Results



Resistance to change of the smoothing parameter.