Stereo Vision: Algorithms and Applications

Stefano Mattoccia

Department of Computer Science (DISI)
University of Bologna

stefano.mattoccia@unibo.it
www.vision.deis.unibo.it/smatt

January 12, 2013

Updates

- Jan 2015: added new 3D applications
- Added details about our stereo camera with FPGA processing
- November 21, 2011: added experimental results for "Linear stereo matching" (ICCV2011), Min et al's algorithm (ICCV2011), description of "Fast Segmentation driven (FSD)" (IC3D) algorithm and description of SGM
- May 19, 2011: added experimental results of FBS on the GPU [71] and the VisionSt stereo camera
- July 25, 2010: Linux and Windows implementations of the Fast Bilateral Stereo algorithm available at: www.vision.deis.unibo.it/smatt/fast_bilateral_stereo.htm
- April 20th, 2010: included descriptions and experimental results for papers [67], [68], [69]

The latest version of this document is available here:

http://www.vision.deis.unibo.it/smatt/Seminars/StereoVision.pdf

Outline

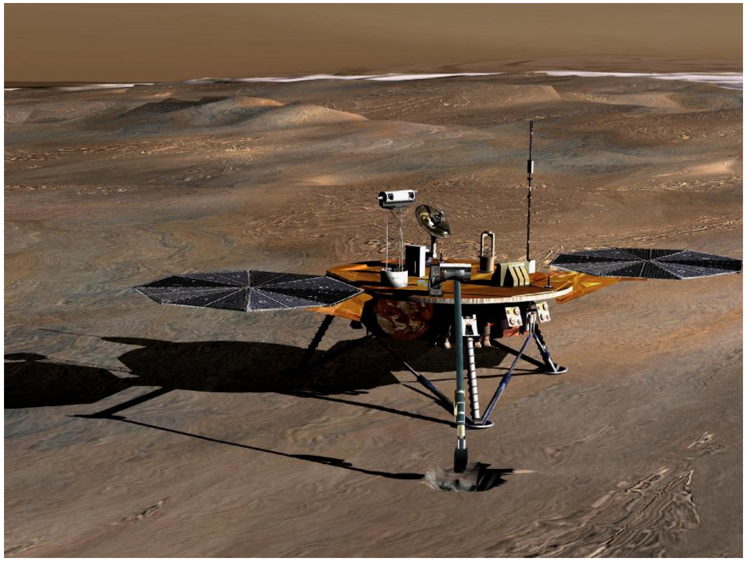
- Introduction to stereo vision
- Overview of a stereo vision system
- Algorithms for visual correspondence
- Computational optimizations
- Hardware implementation
- Applications

What is stereo vision ?

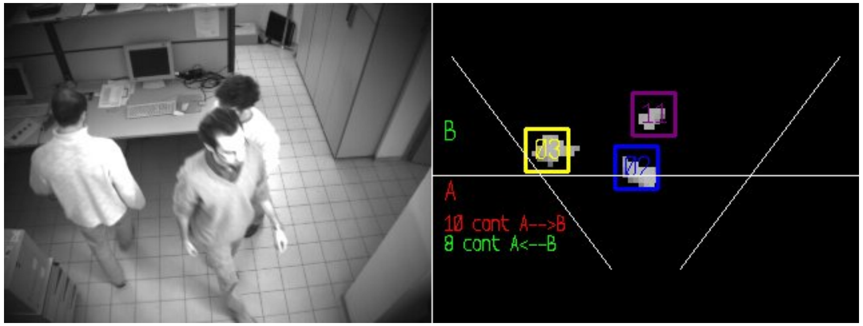
- Is a technique aimed at inferring depth from two or more cameras
- Wide research topic in computer vision
- This seminar is concerned with
  - binocular stereo vision systems
  - dense stereo algorithms
  - stereo vision applications
- Emphasis is on approaches that are (or might be hopefully soon) feasible for real-time/hardware implementation

Applications

www.nasa.gov

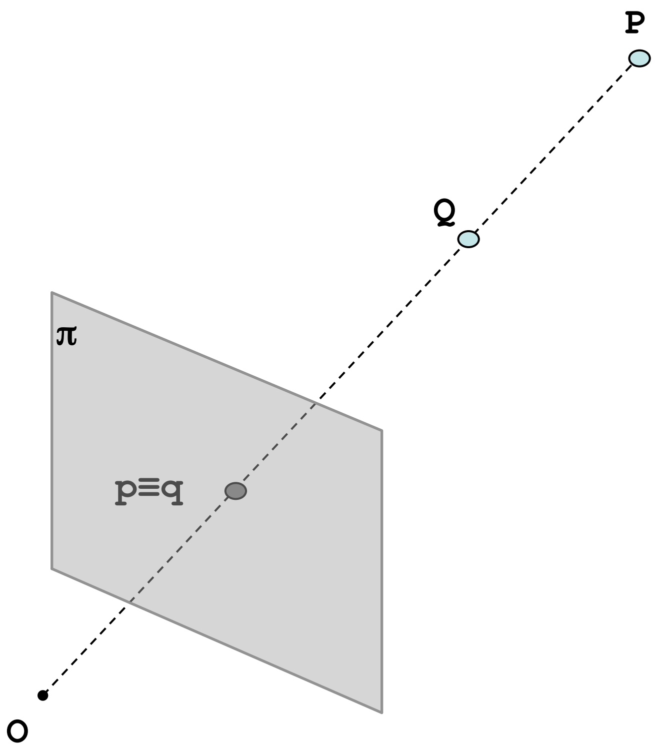


www.nasa.gov



www.vision.deis.unibo.it/smatt/stereo

Single camera



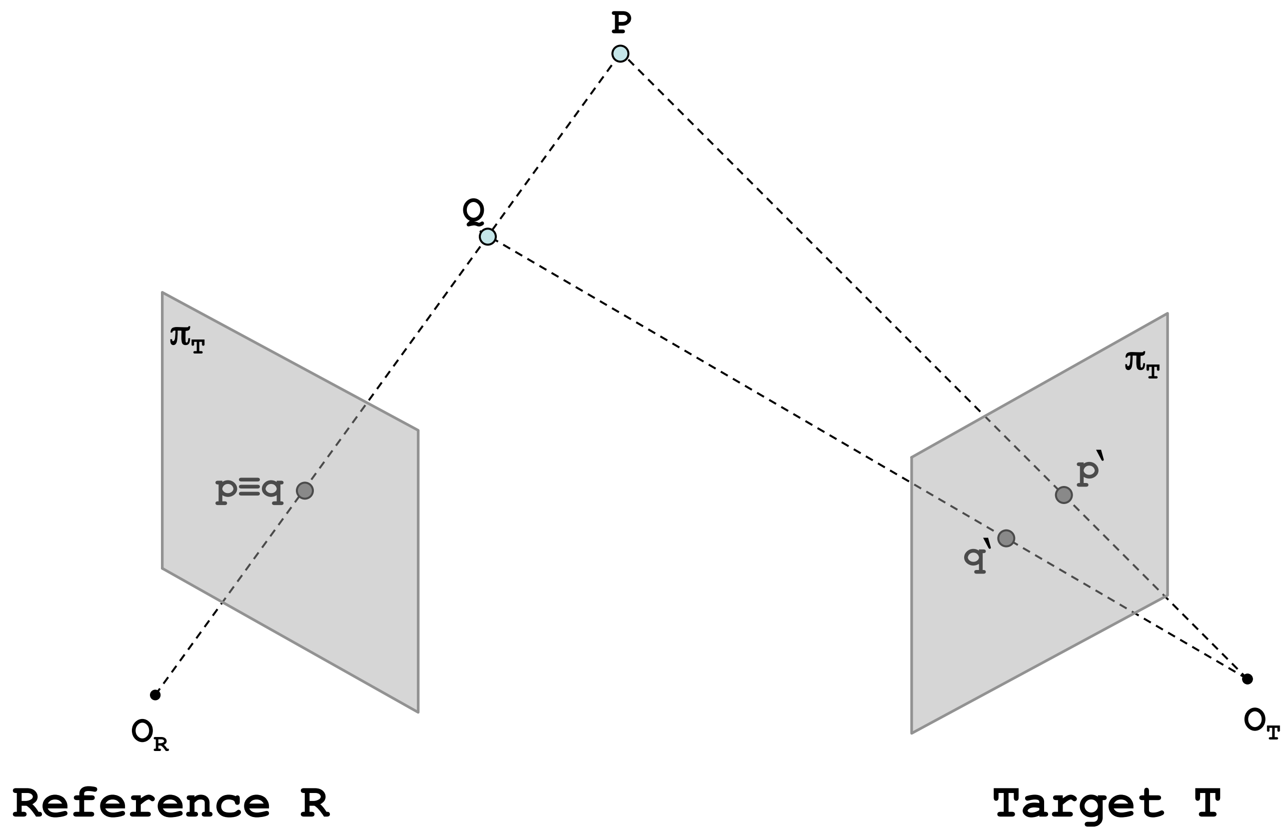
π: image plane

O: optical center

- Both (real) points (P and Q) project into the same image point (p ≡ q)
- This occurs for each point along the same line of sight
- Useful for optical illusions…

Courtesy of http://www.coolopticalillusions.com/

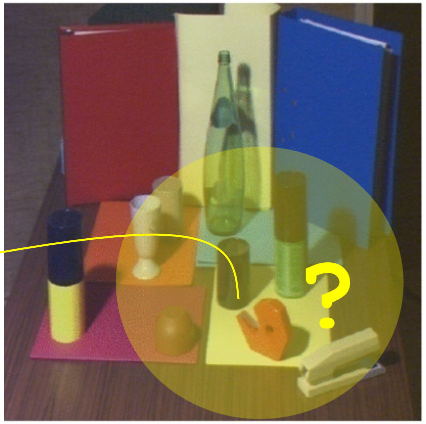
Stereo camera



With two (or more) cameras we can infer depth, by means of triangulation, if we are able to find corresponding (homologous) points in the two images

How to solve the correspondence problem ?

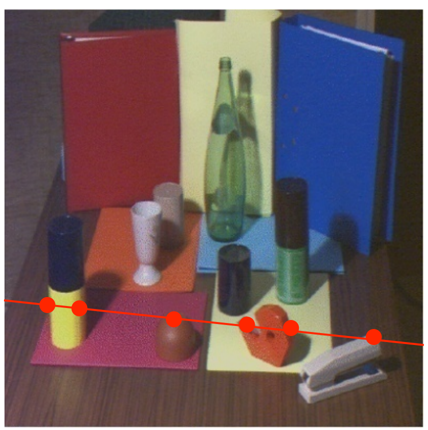
Reference (R)



Target (T)

2D search domain ?

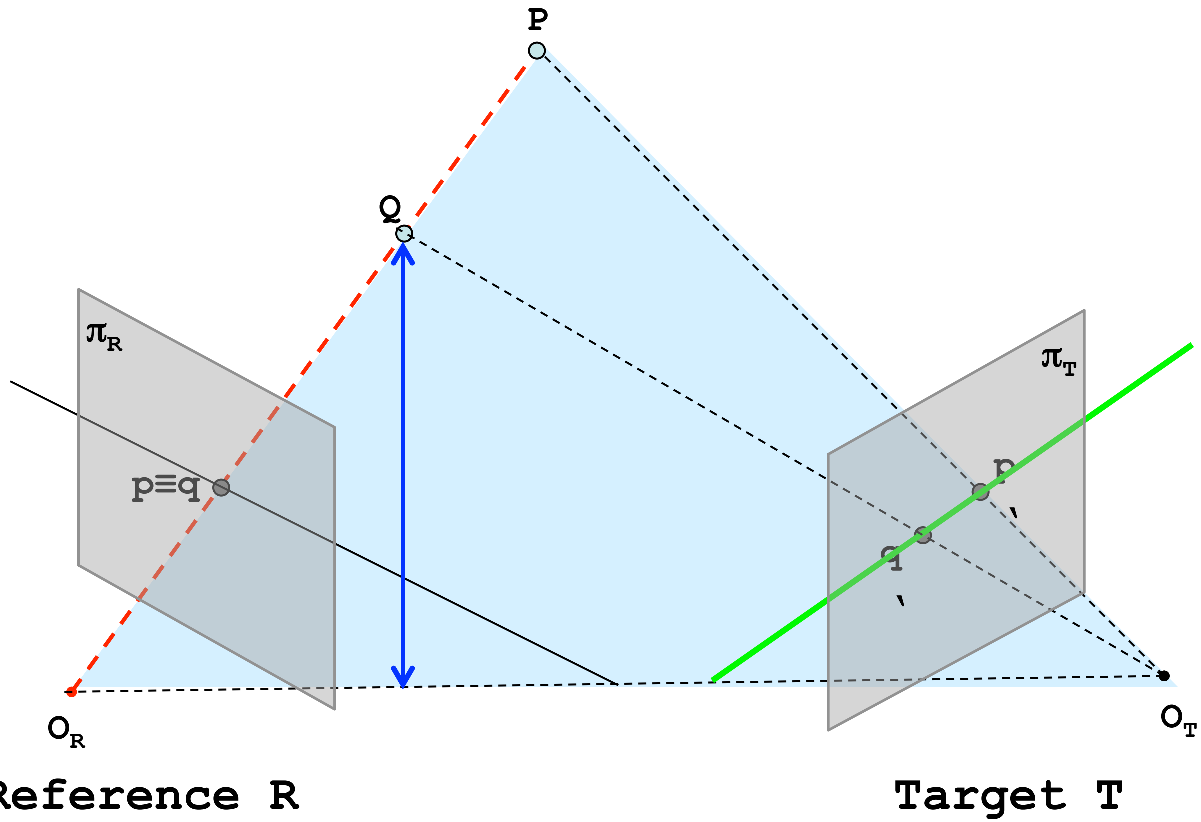
Reference (R)



Target (T)

No!! Thanks to the epipolar constraint

Epipolar constraint

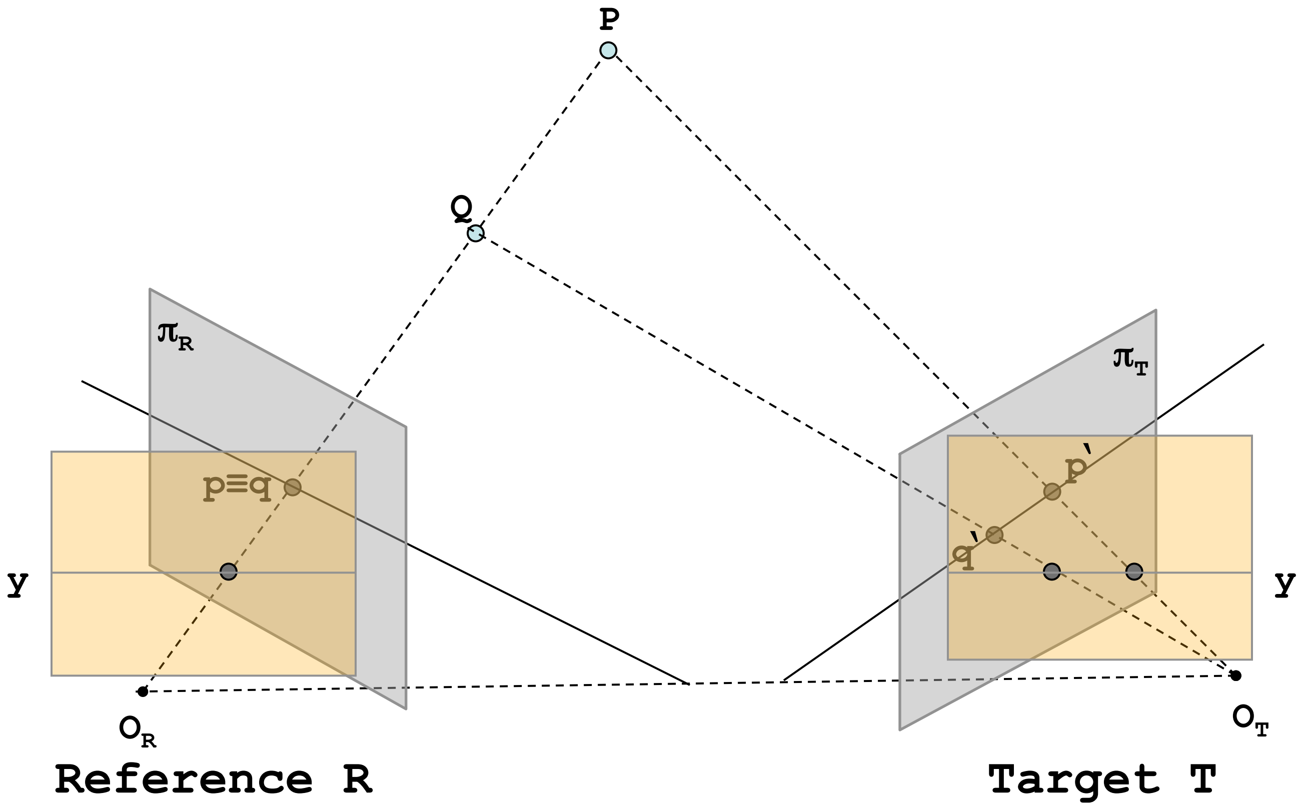


- Consider two points P and Q on the same line of sight of the reference image R (both points project into the same image point p≡q on image plane π_R of the reference image)

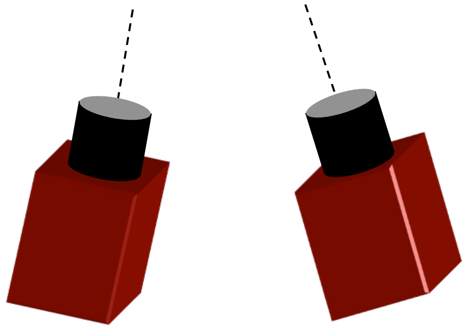
- The epipolar constraint states that the correspondence for a point belonging to the (red) line of sight lies on the green line on image plane π_T of target image

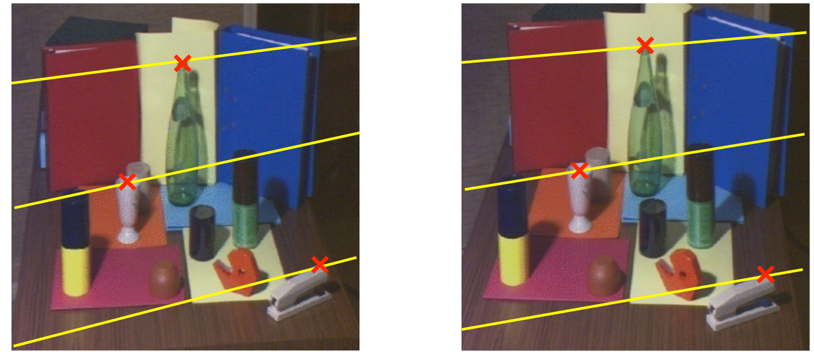
Stefano Mattoccia

Stereo camera in standard form

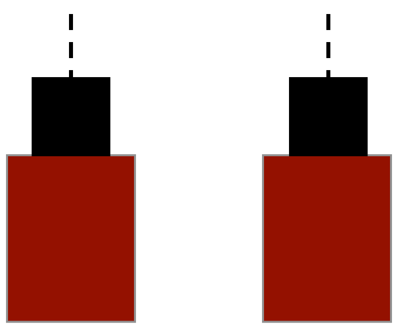


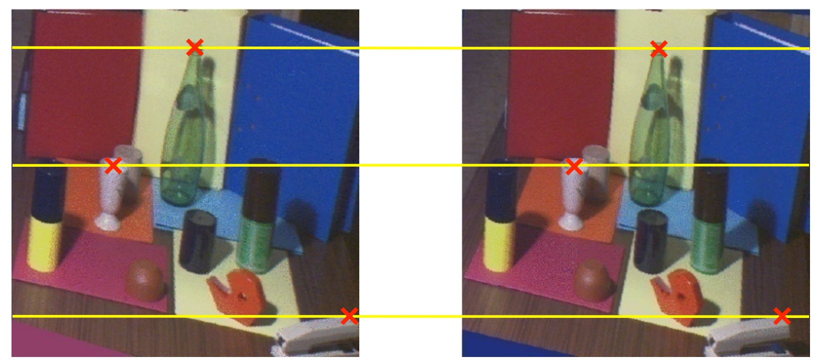
Once we know that the search space for corresponding points can be narrowed from 2D to 1D, we can put (virtually) the stereo rig in a more convenient configuration (standard form) - corresponding points are constrained on the same image scanline





Original stereo pair



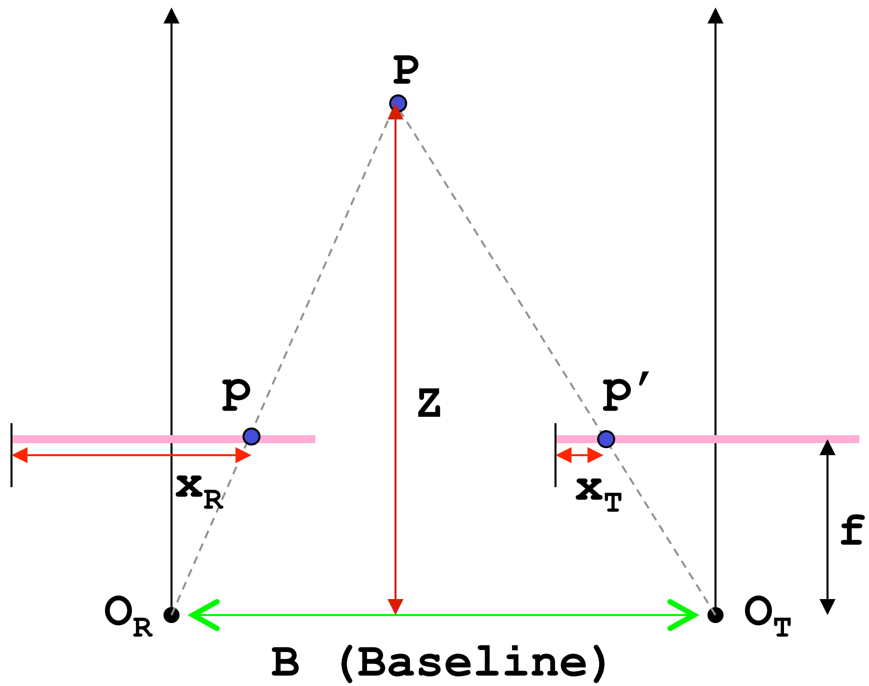


Stereo pair in standard form

Cameras are "*perfectly*" aligned
and with the same *focal length*

Stefano Mattoccia

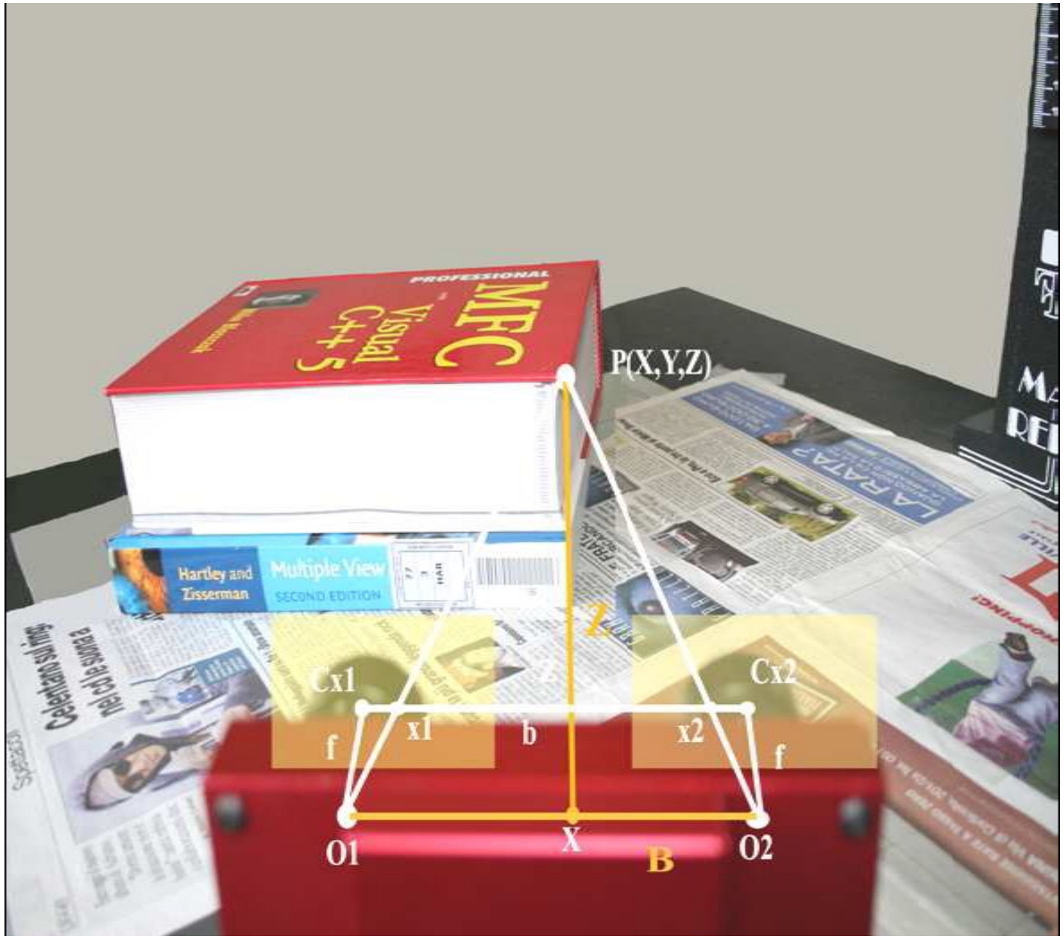
Disparity and depth



With the stereo rig in standard form and by considering similar triangles (PO_RO_T and Ppp'):

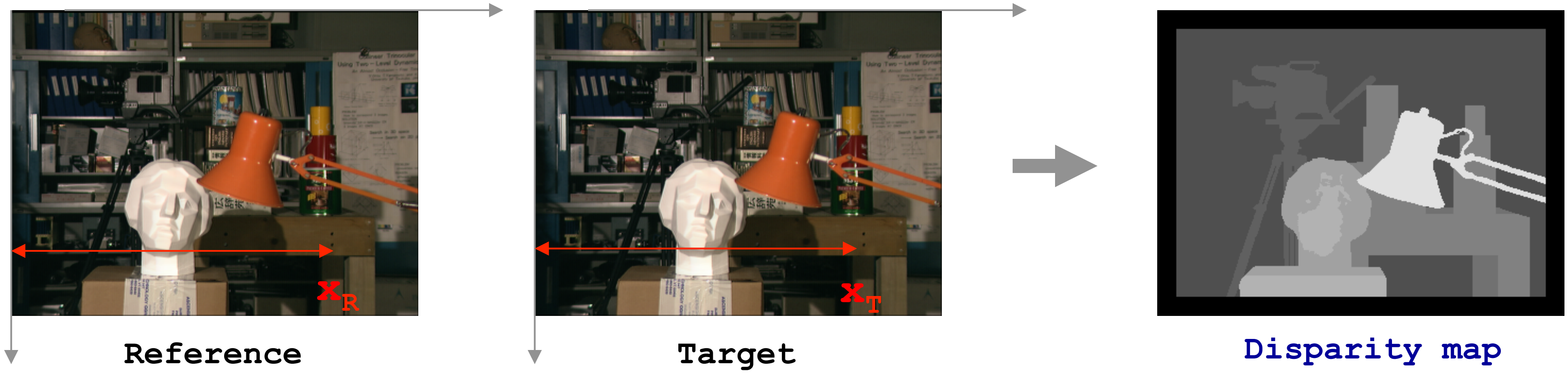
$$\frac{b}{Z} = \frac{(b + x_T) - x_R}{Z - f} \quad \Rightarrow \quad Z = \frac{b \cdot f}{x_R - x_T} = \frac{b \cdot f}{d}$$

$x_R - x_T$ is the disparity

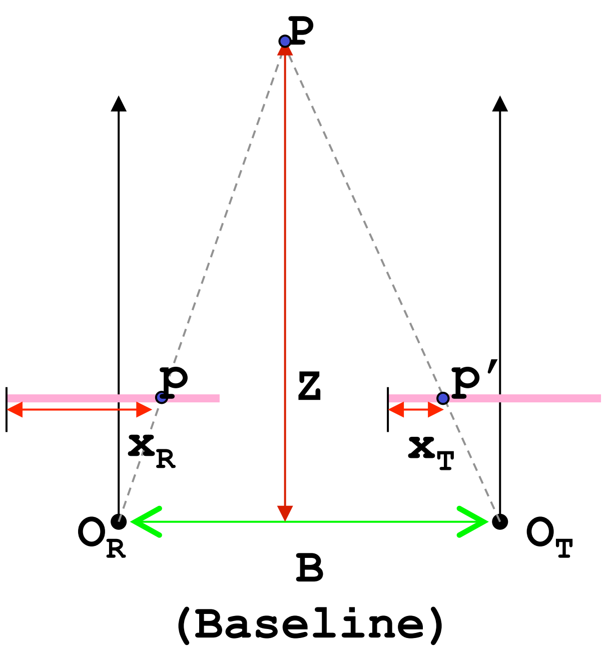


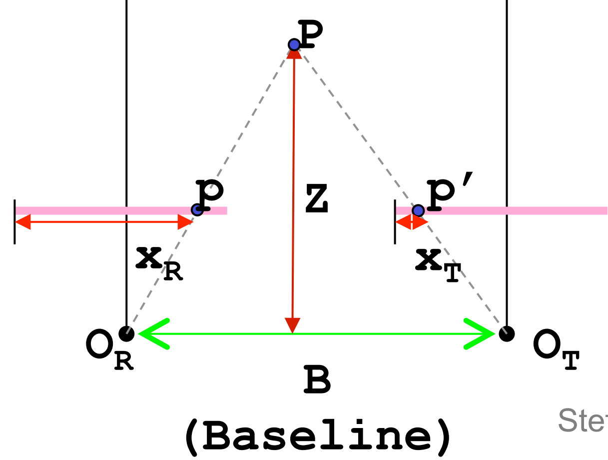
Disparity and depth

The disparity is the difference between the x coordinate of two corresponding points; it is typically encoded with greyscale image (closer points are brighter).



Disparity is higher for points closer to the camera

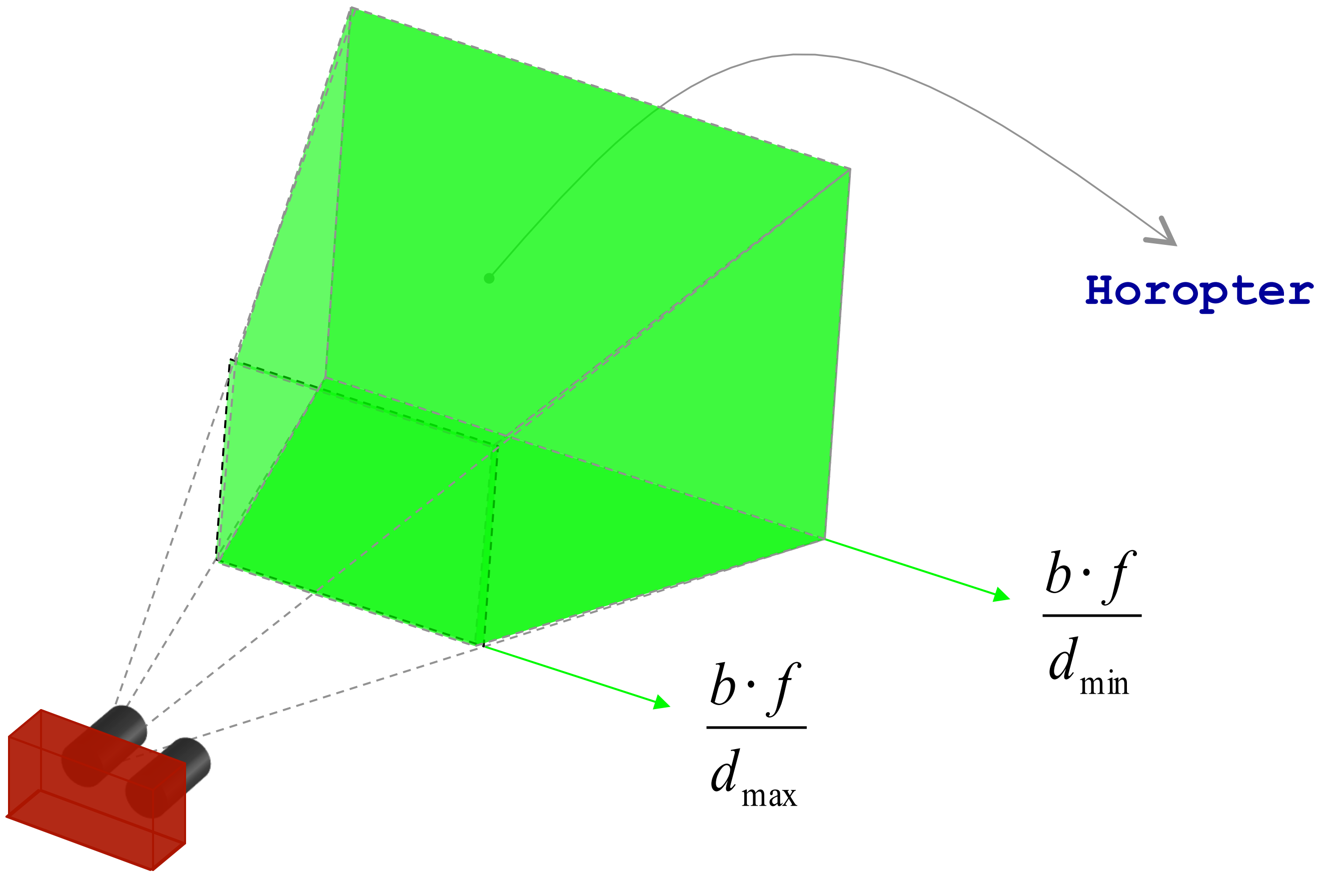


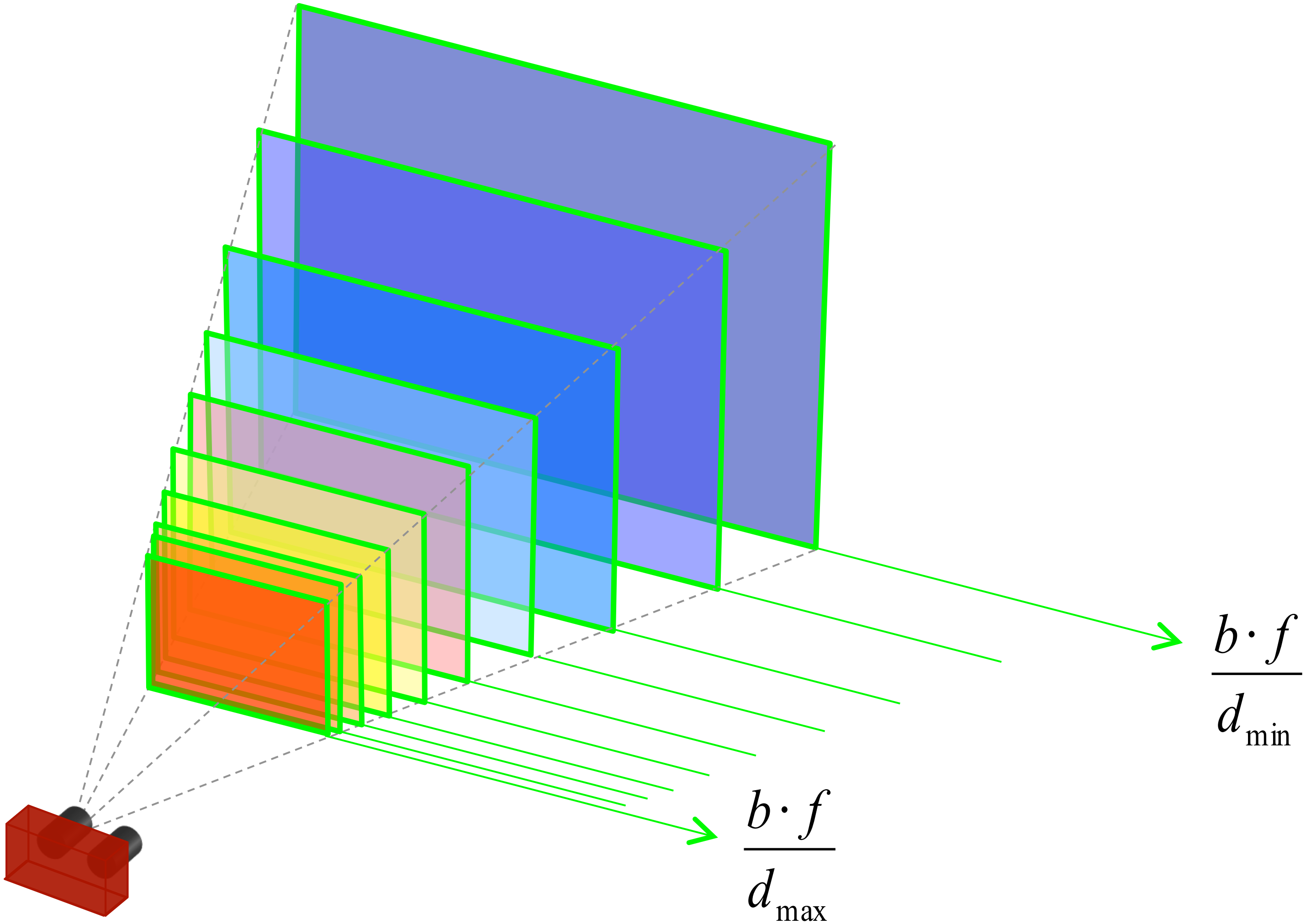


Stefano Mattoccia

Range field (Horopter)

Given a stereo rig with baseline b and focal length f, the range field of the system is constrained by the disparity range [d_{min}, d_{max}].

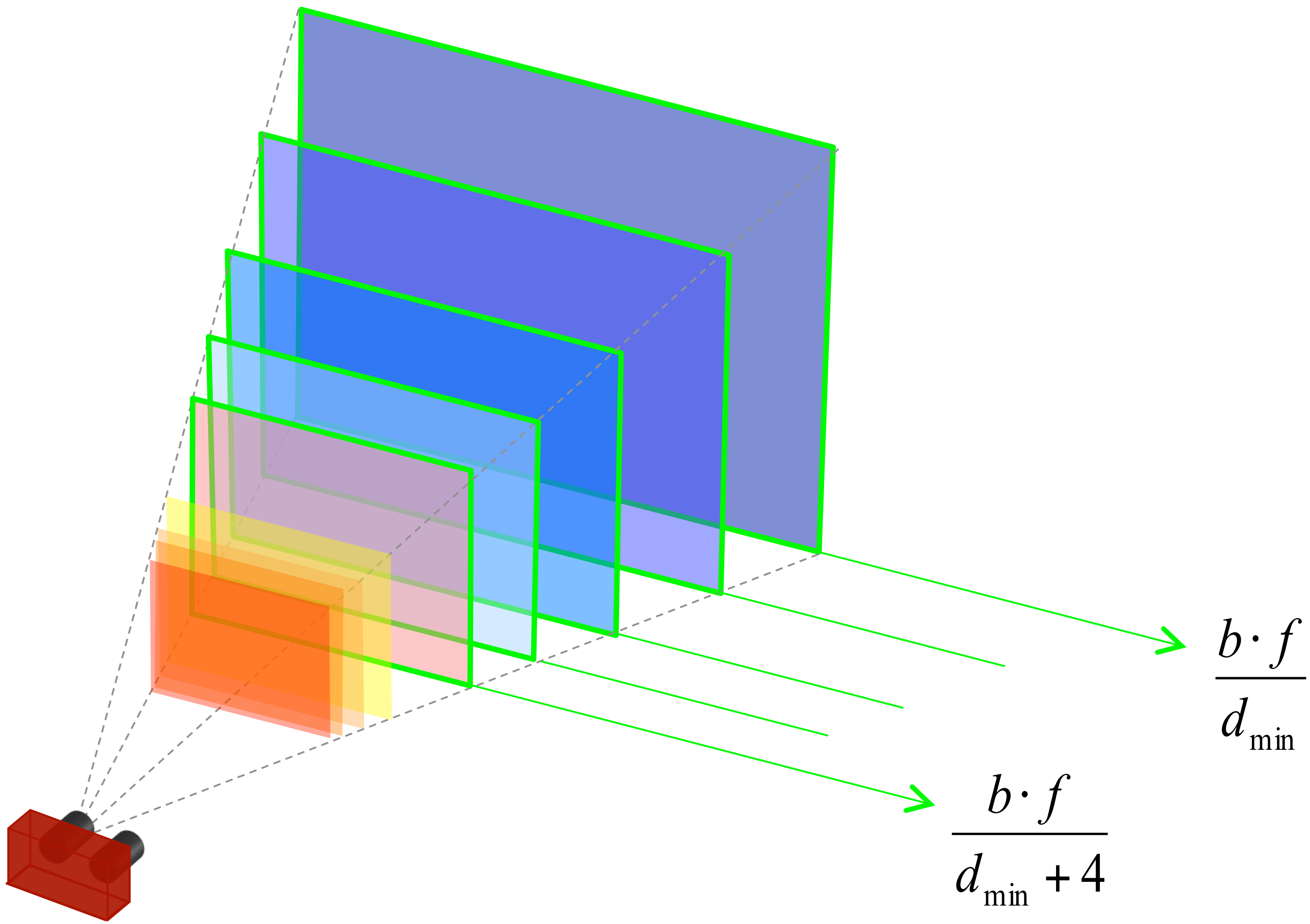




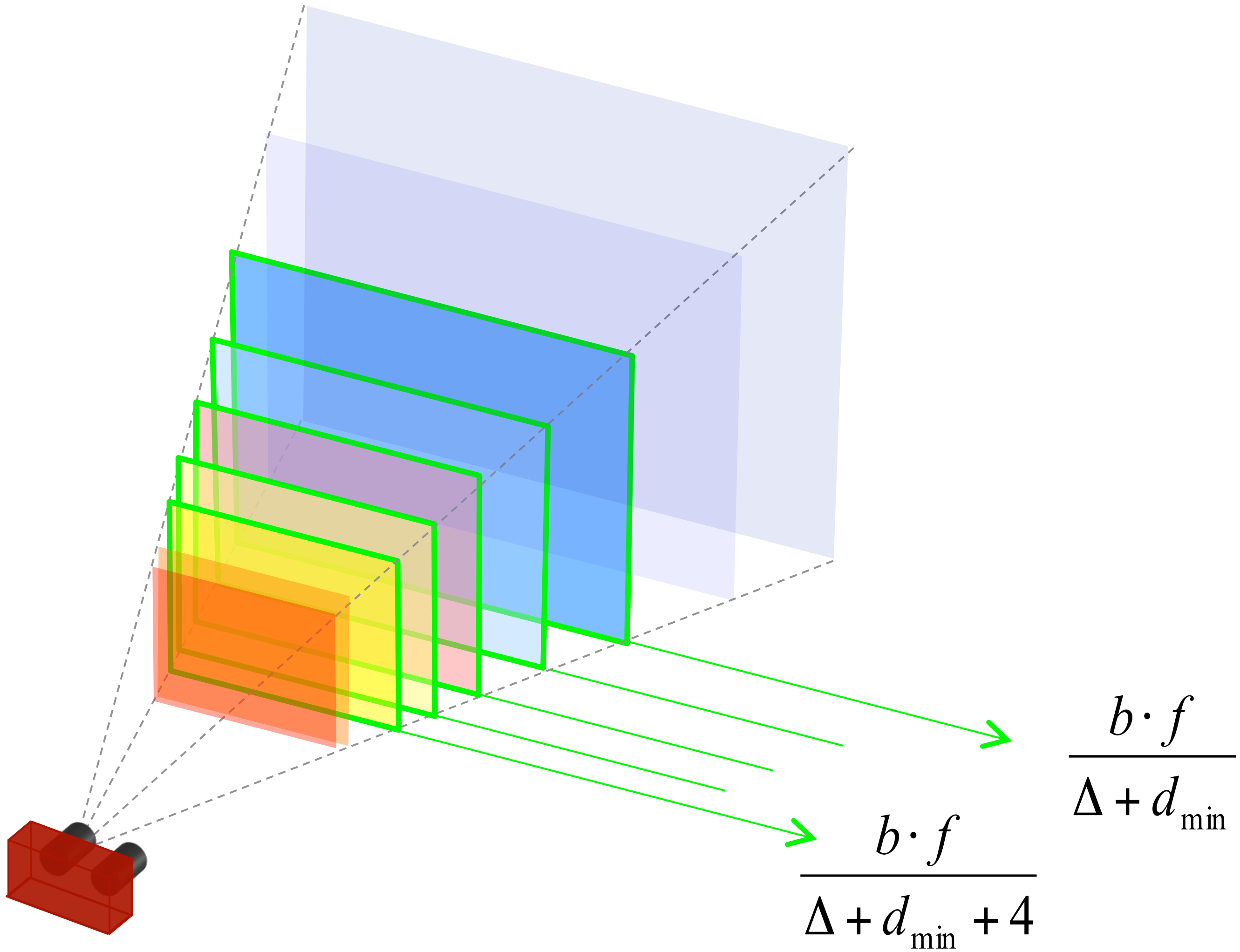
- Depth measured by a stereo vision system is discretized into parallel planes (one for each disparity value)

- A better (virtual) discretization can be achieved with subpixel techniques (see Disparity Refinements)

Stefano Mattoccia



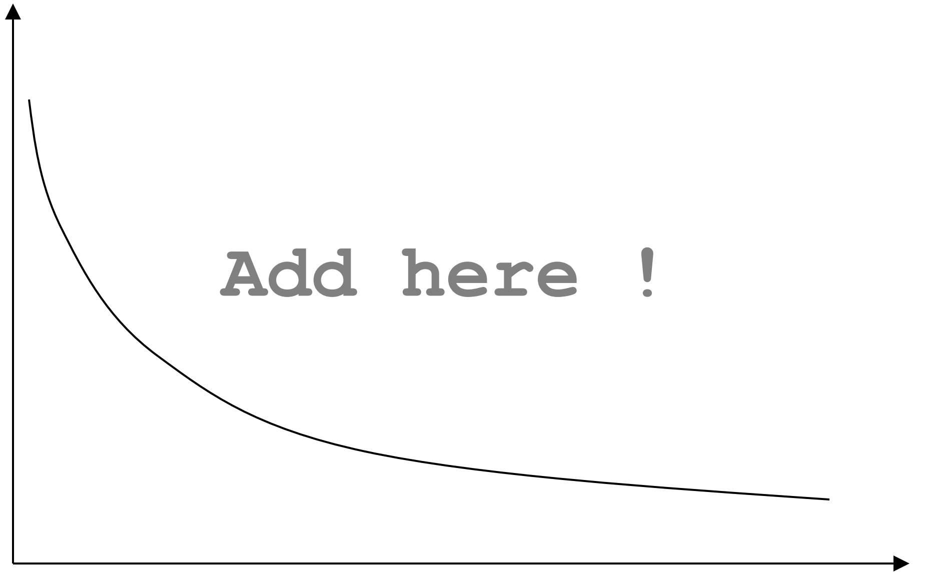
- **The range field (horopter) using 5 disparity values [d_{min}, d_{min}+4]**



- **Using 5 disparity values [Δ+d$_{min}$,Δ+d$_{min}$+4]**
- **With Δ>0, *horopter* gets closer and shrinks (depth and obviously area/volume)**

Stefano Mattoccia

Accuracy vs Resolution: quantitative analysis

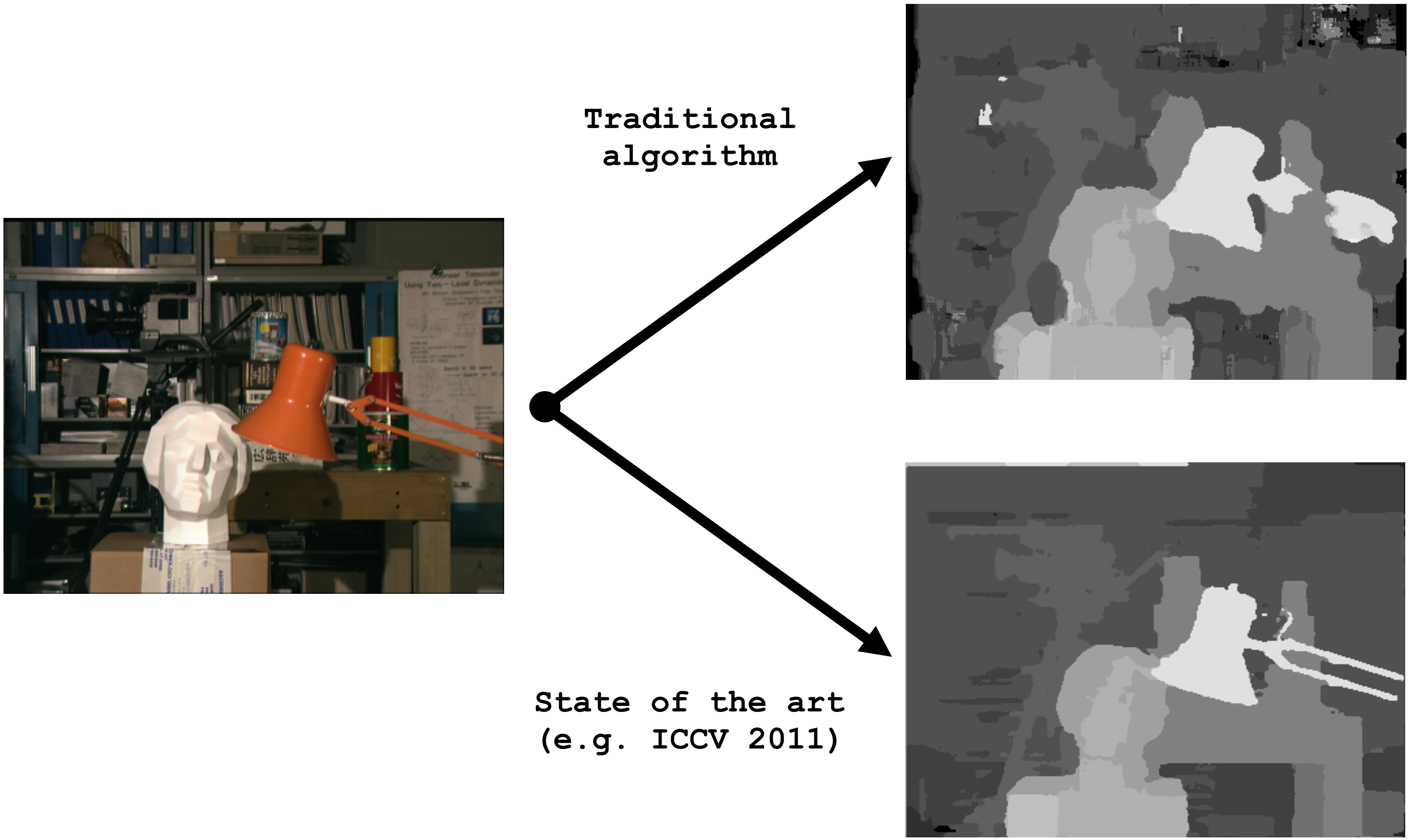


Color or greyscale sensors ?

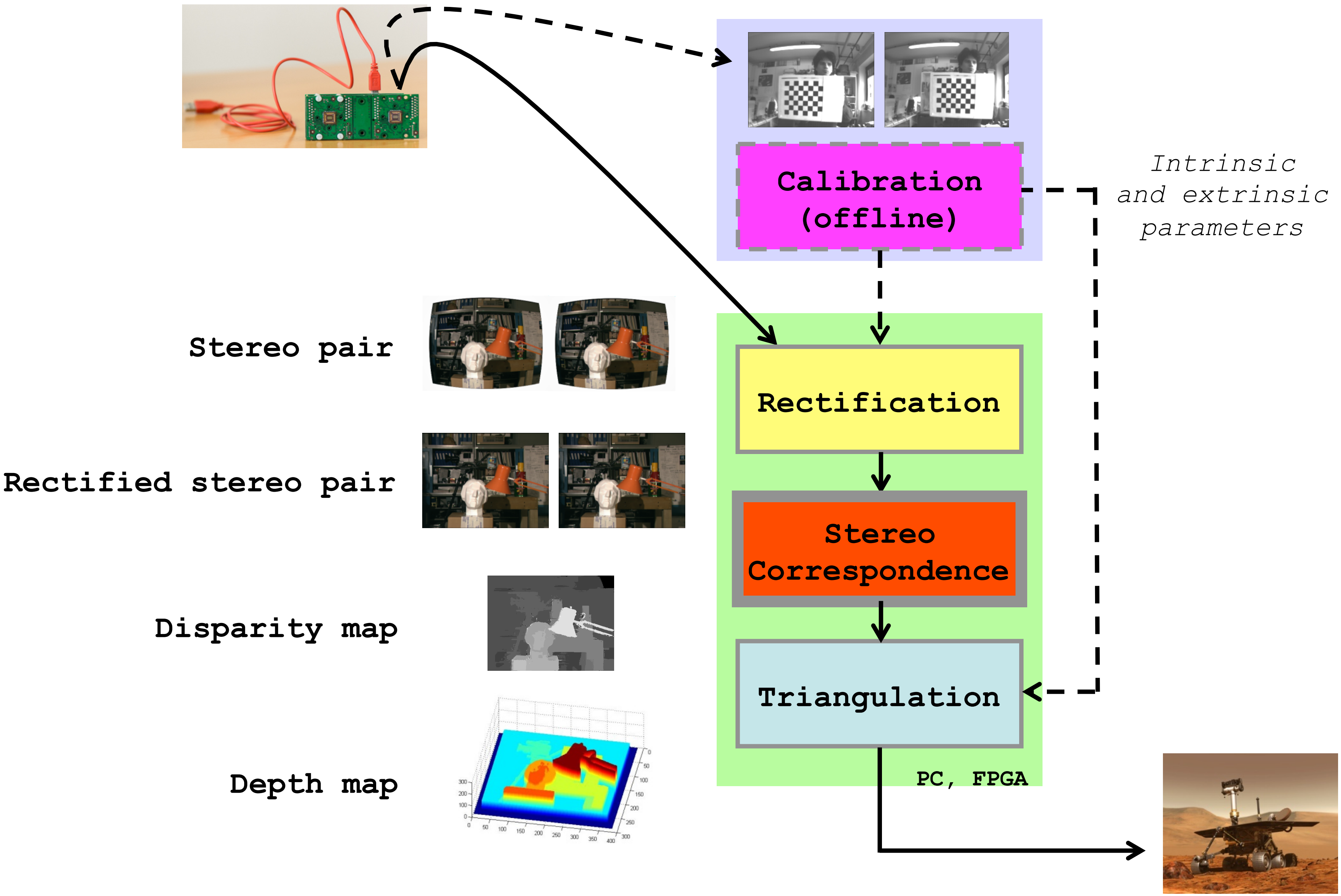
Insert here

Key module in stereo vision?

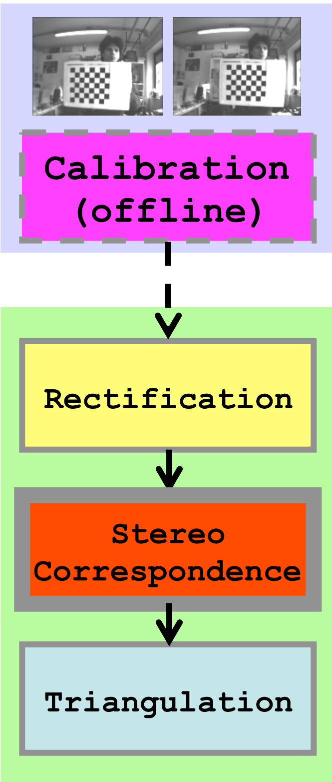
The algorithm is crucial in this technology



Overview of a stereo vision system

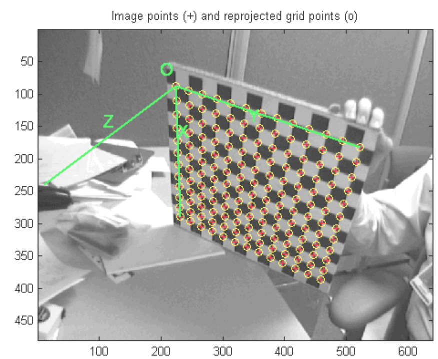


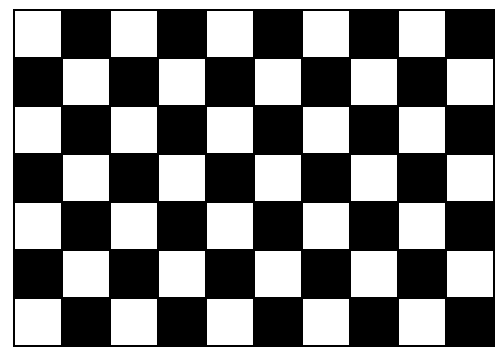
Calibration (offline)



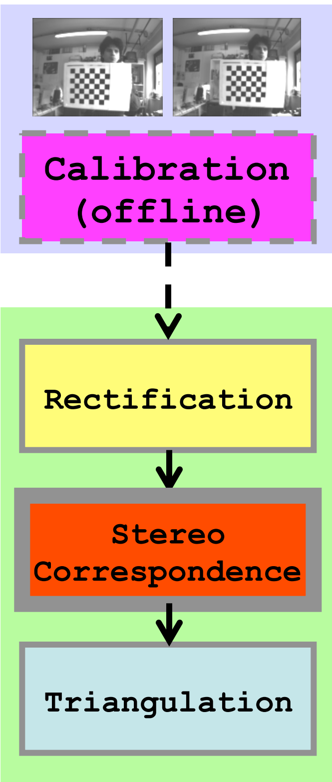
Offline procedure aimed at finding:

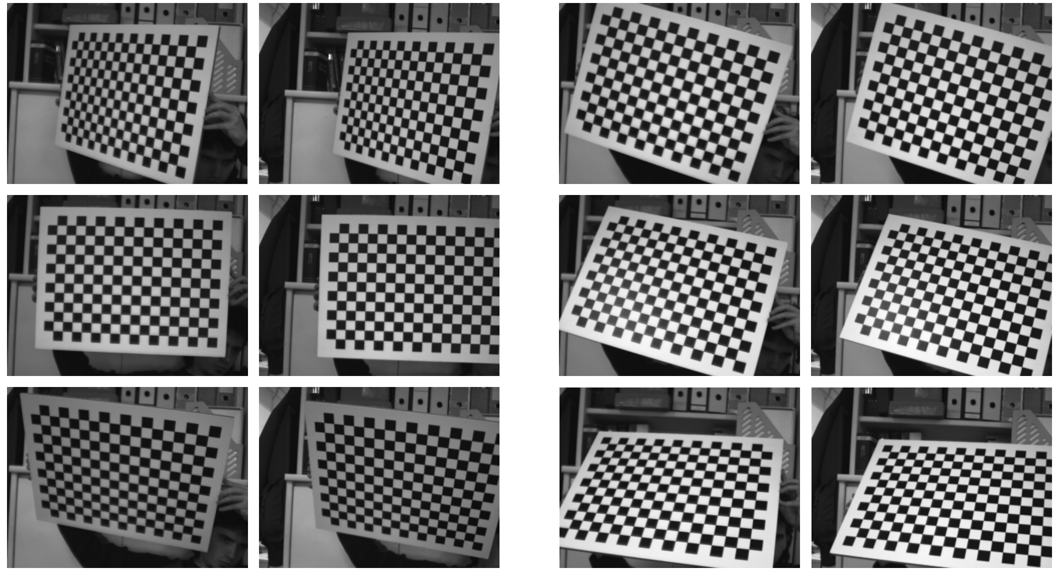
- Intrinsic parameters of the two cameras (focal length, image center, parameters of lenses distortion, etc)
- Extrinsic parameters (R and T that aligns the two cameras)





Calibration is carried out acquiring and processing 10+ stereo pairs of a known pattern (typically a checkerboard)

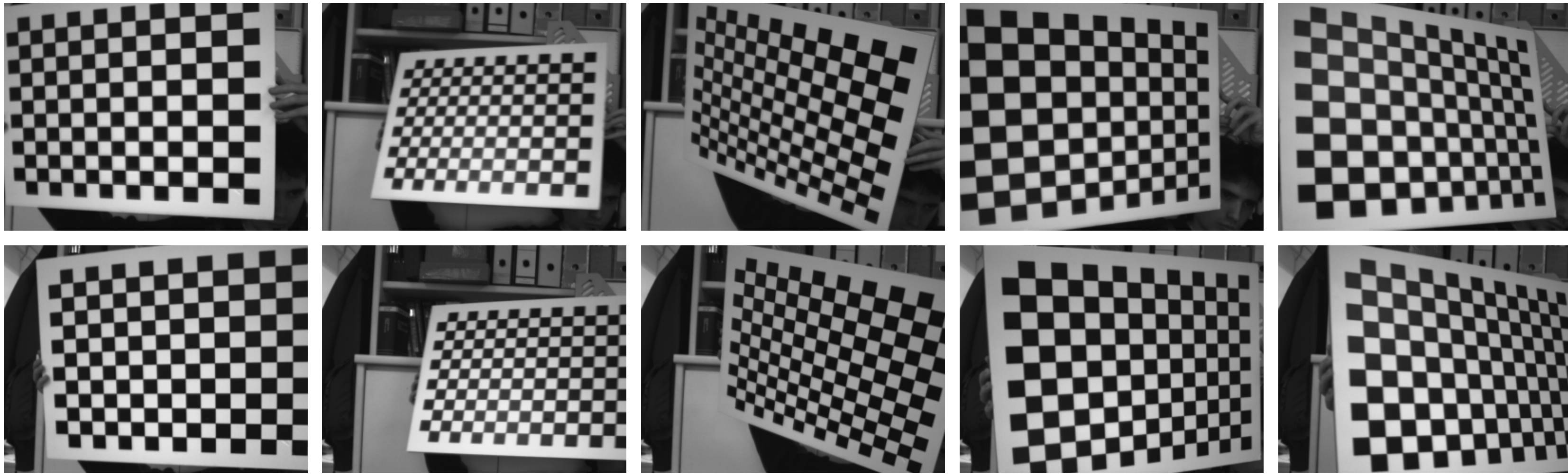




- Calibration is available in OpenCV [39] and Matlab [40]
- A detailed description of calibration can be found in [20,21,22]
- Next slides show 20 stereo pairs used for calibrating a stereo camera

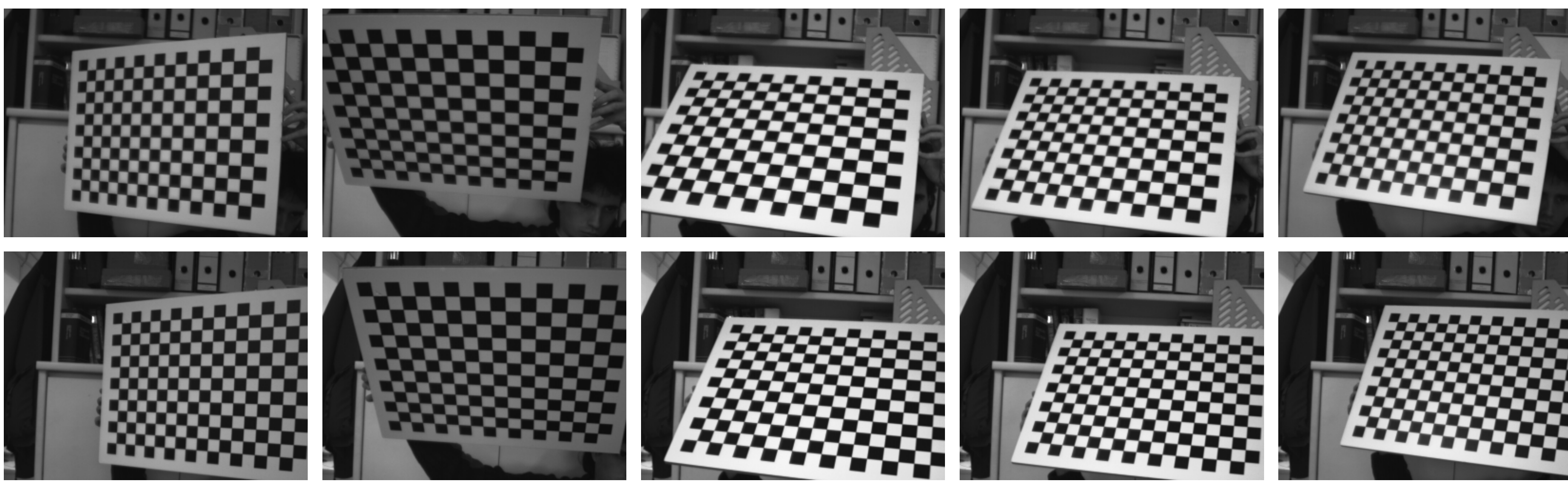
R

T



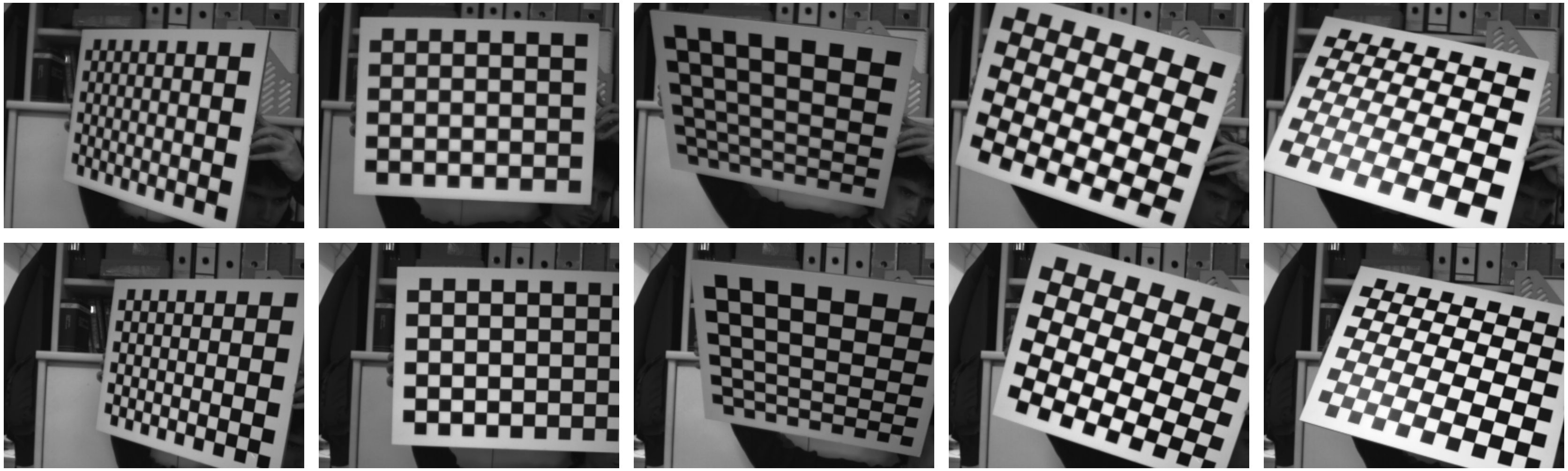
R

T



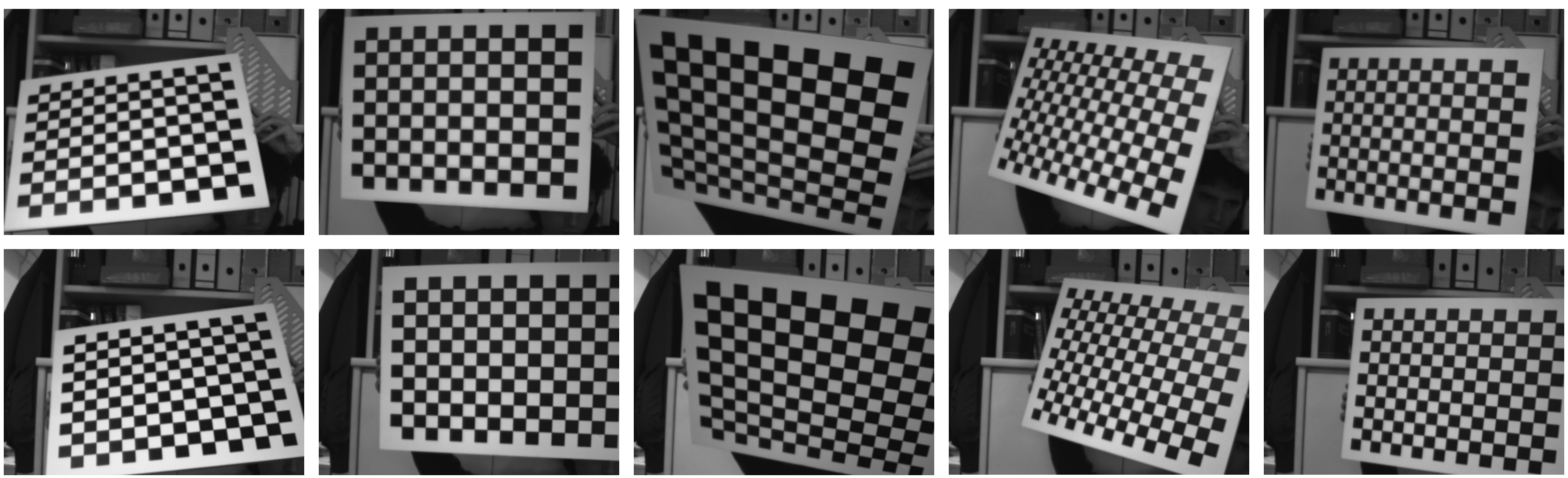
R

T



R

T

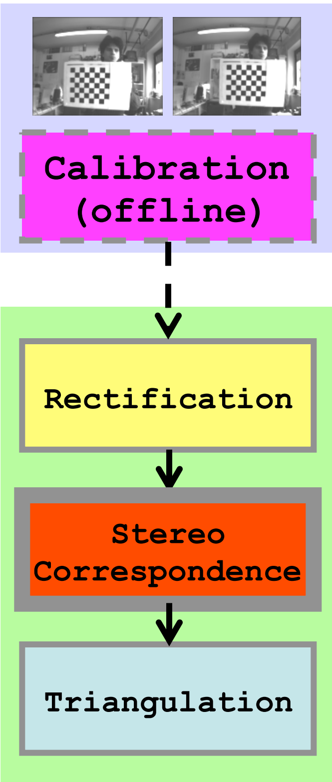


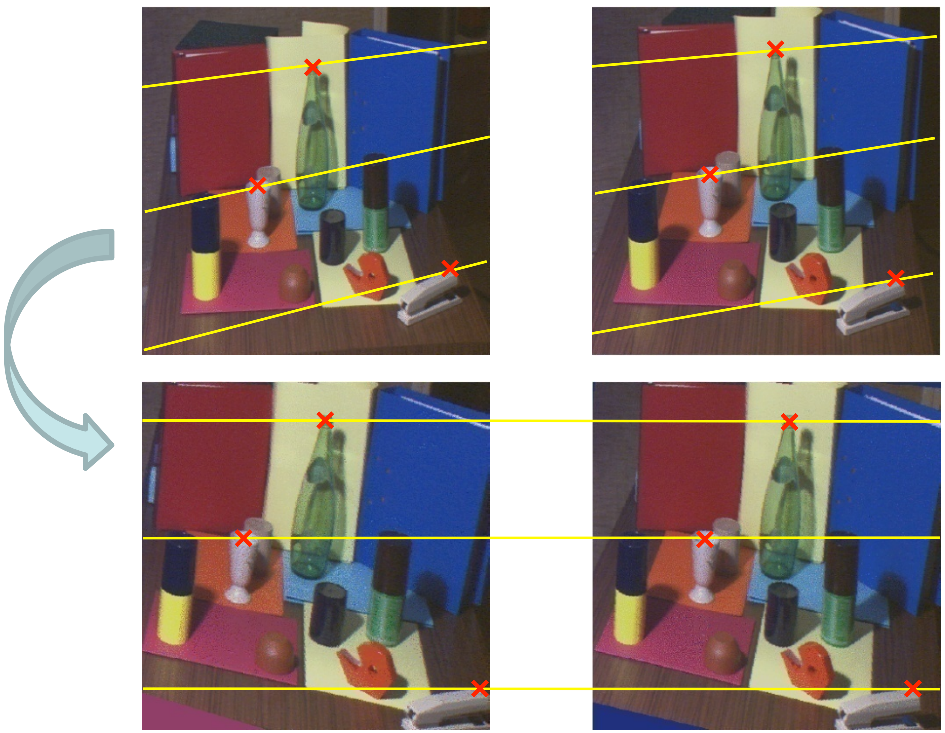
Rectification

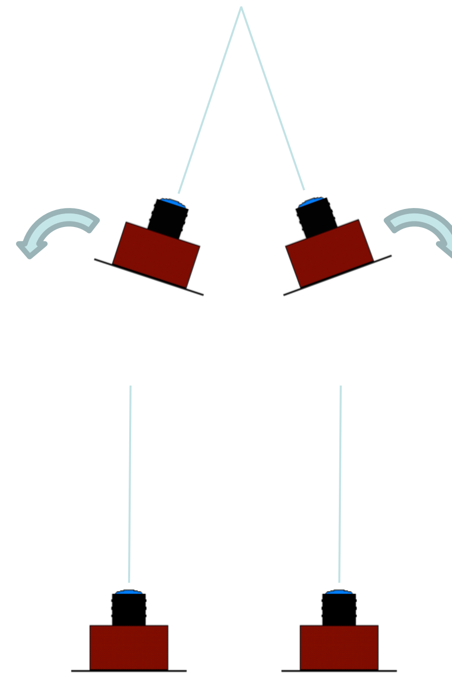
Using the information from the calibration step:

a) removes lens distortions

b) turns the stereo pair in standard form



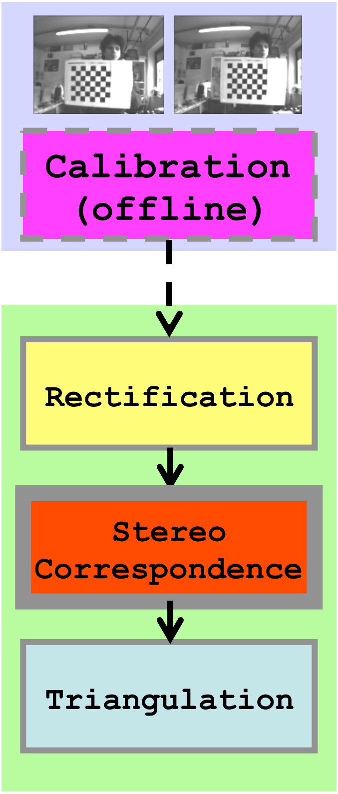


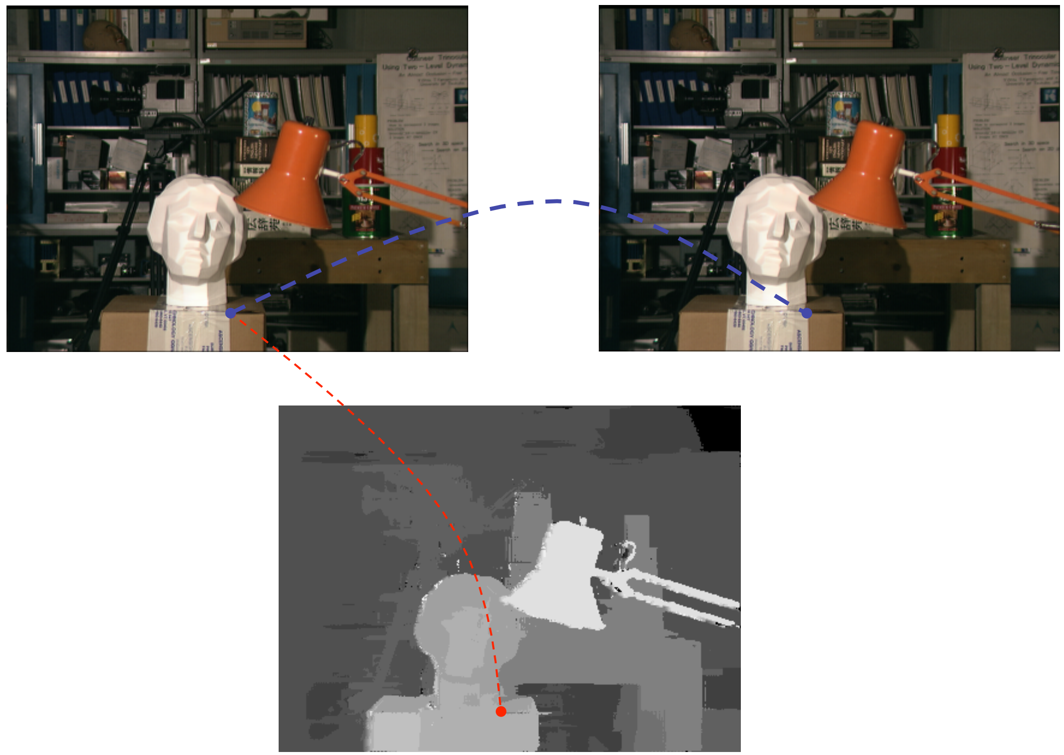


Stereo camera in standard form

Stereo correspondence

Aims at finding homologous points in the stereo pair.

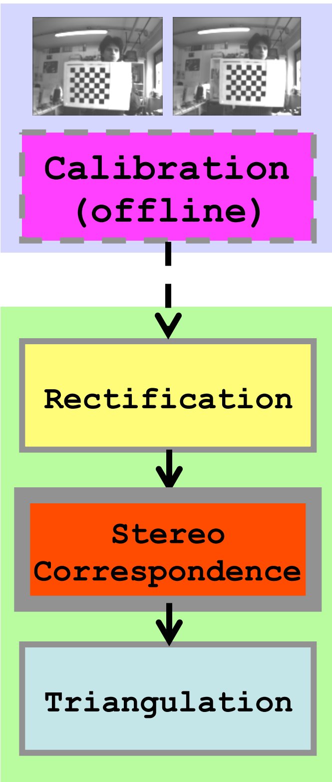




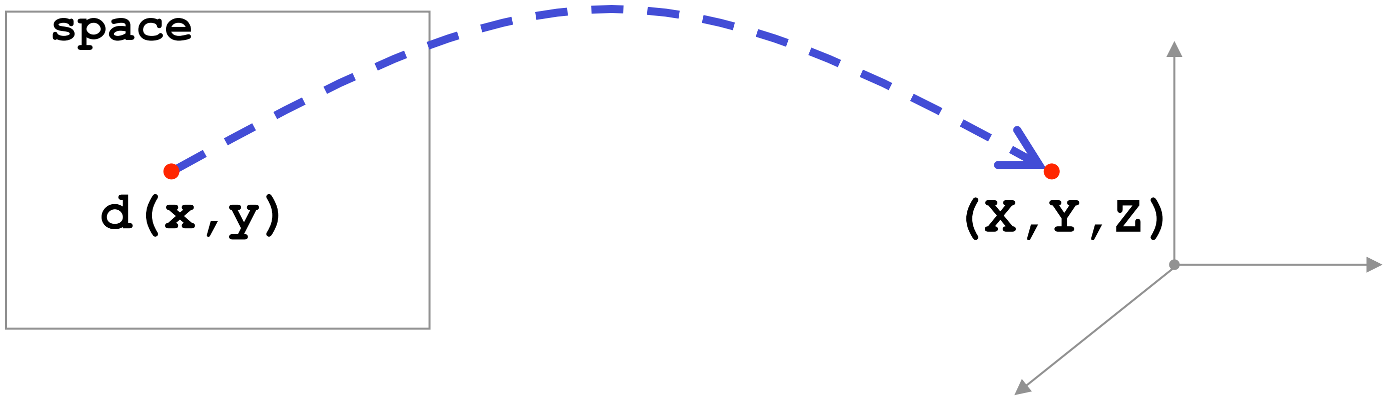
disparity map

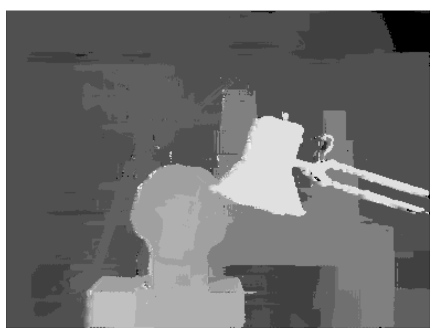
This topic will be extensively analyzed in the next slides...

Triangulation



Given the disparity map, the baseline and the Focal length (calibration): triangulation computes the position of the correspondence in the 3D space



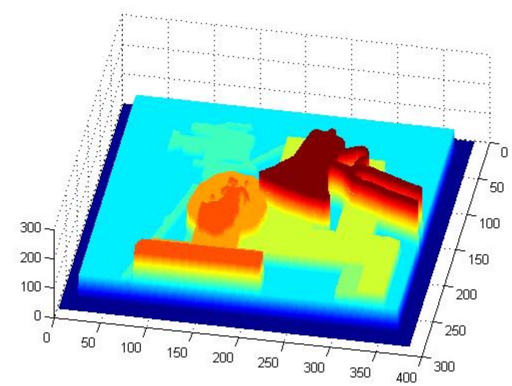


disparity map

$$Z = \frac{b \cdot f}{d}$$

$$X = Z \frac{x_R}{f}$$

$$Y = Z \frac{y_R}{f}$$



depth map

Datasets: stereo sequences

Sequences acquired with stereo cameras are available at:

http://www.vision.deis.unibo.it/smatt/stereo.htm

The datasets include:

- **calibration parameters**
- **original sequences**
- **rectified sequences**
- **disparity maps**

Architectures

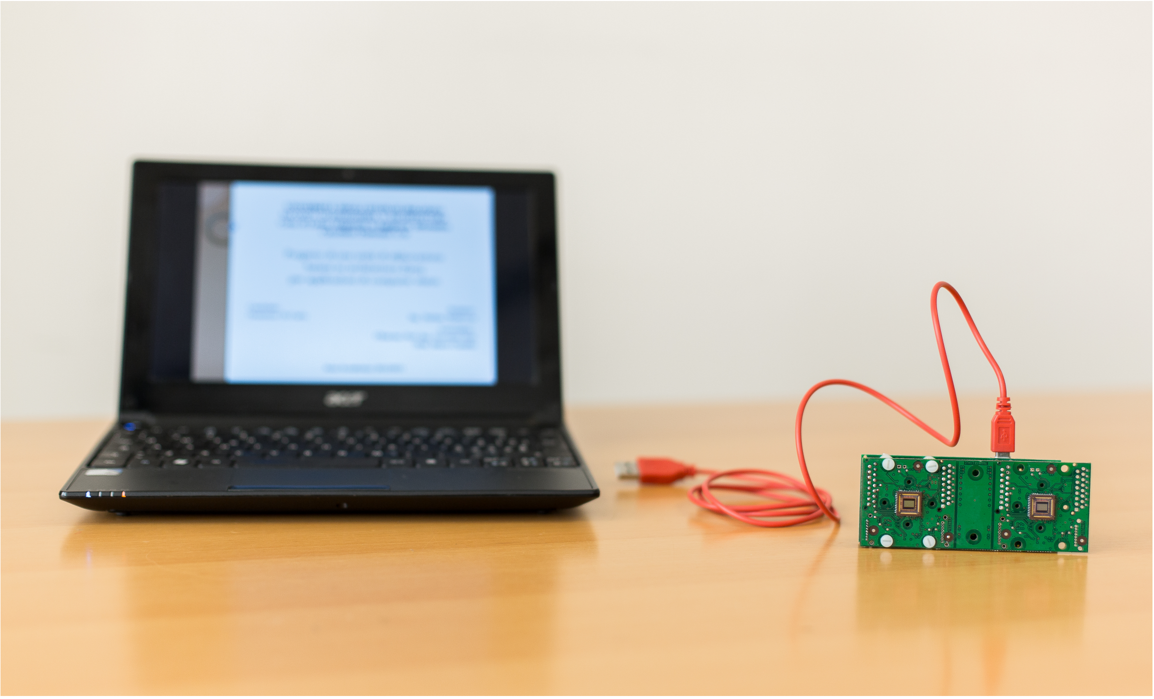
- Microprocessors
  - Floating Point (FP) units + SIMD
  - C/C++ (+ assembly)
  - power,cost and size are the main drawbacks

- Low power & low cost processor
  - C/c++
  - no FP
  - no SIMD (often)

- GPUs (Graphic Processing Units)
  - raw power
  - high power dissipation and cost
  - programming is difficult (CUDA and OpenCL help)

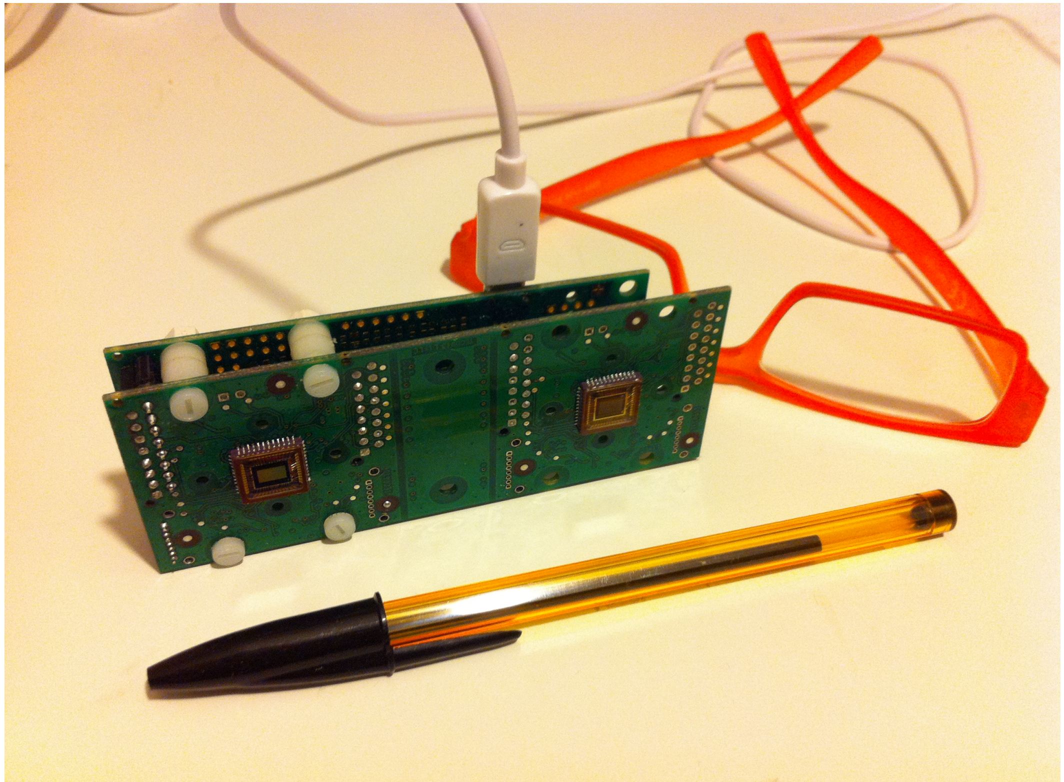
- FPGA (Field Programmable Gate Array)
  - efficient, low power (<1 W), low cost
  - programming language: VHDL
  - coding is difficult and tailored for specific devices

Our custom FPGA-based stereo camera 1/3



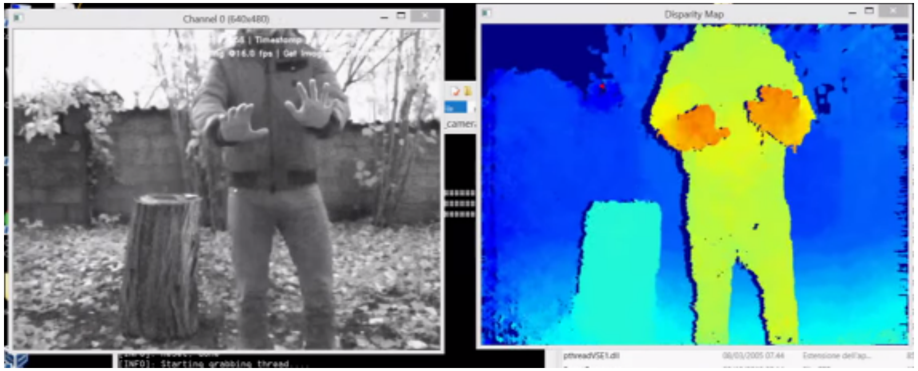
- We have designed a real-time stereo camera with depth maps computed according to state of the art algorithms
- Details: www.vision.deis.unibo.it/smatt
- Youtube channel: www.youtube.com/channel/UChkayQwiHJuf3nqMikhxAlw

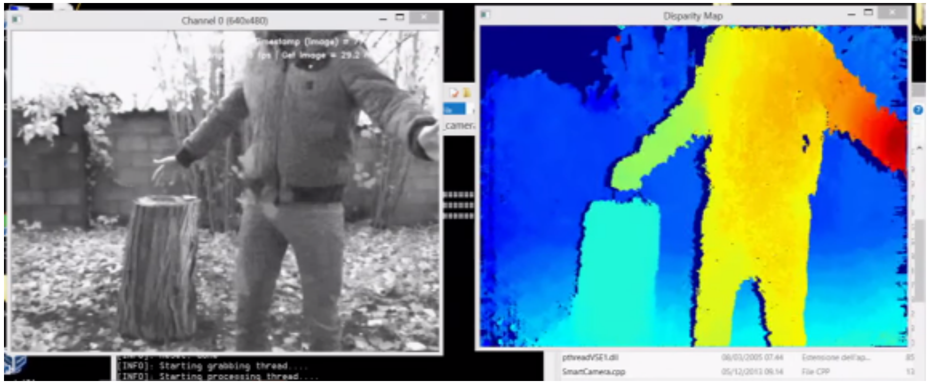
Our custom FPGA-based stereo camera 2/3

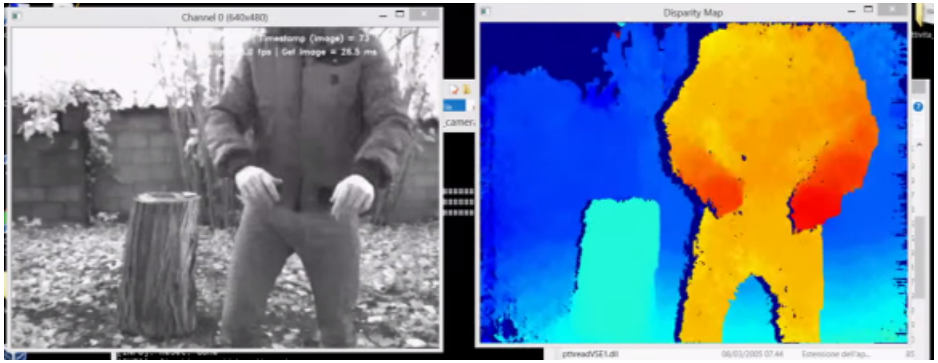


- Processing at 30+ fps (640x480)
- Power consumption: < 2.5 Watt
- Self powered via USB cable
- Weight: < 80 g with lenses and holders

Our custom FPGA-based stereo camera 3/3







www.youtube.com/watch?v=KXFWIvrcAYo

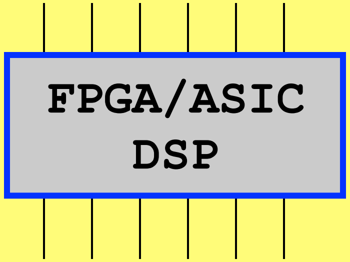
Some available stereo cameras

www.videredesign.com

www.visionst.com

www.valdesystems.com

www.focusrobotics.com



www.tyzx.com

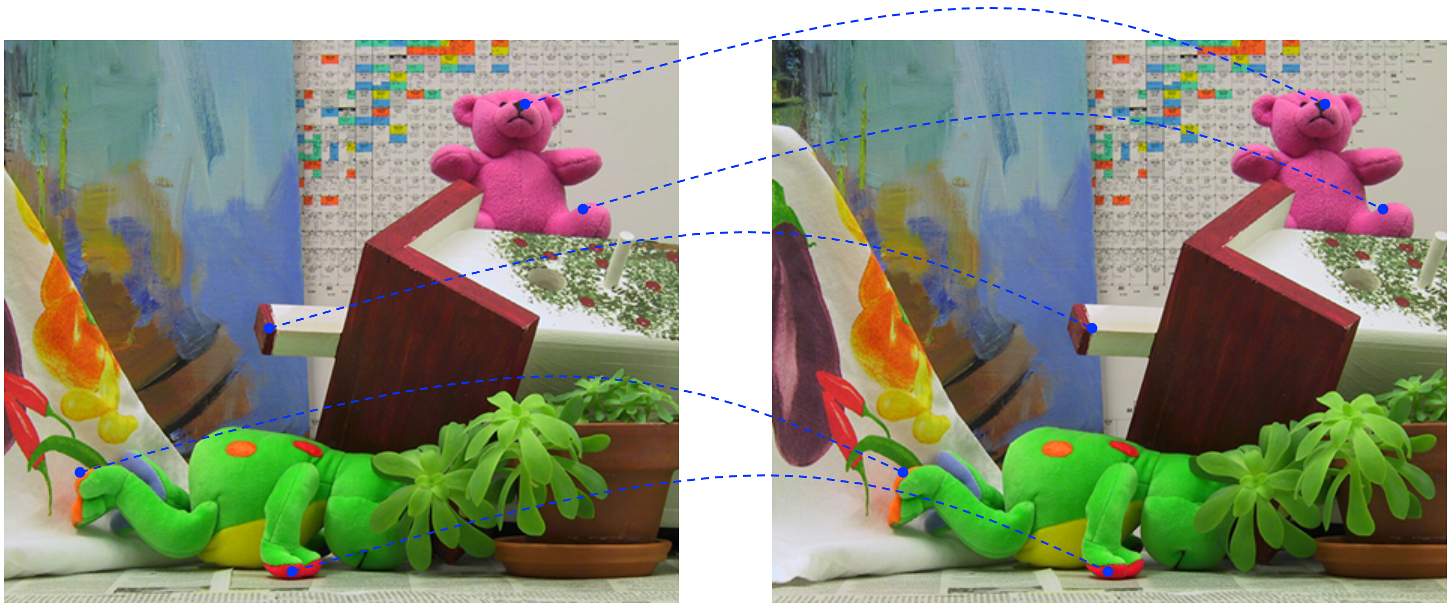
www.ptgrey.com

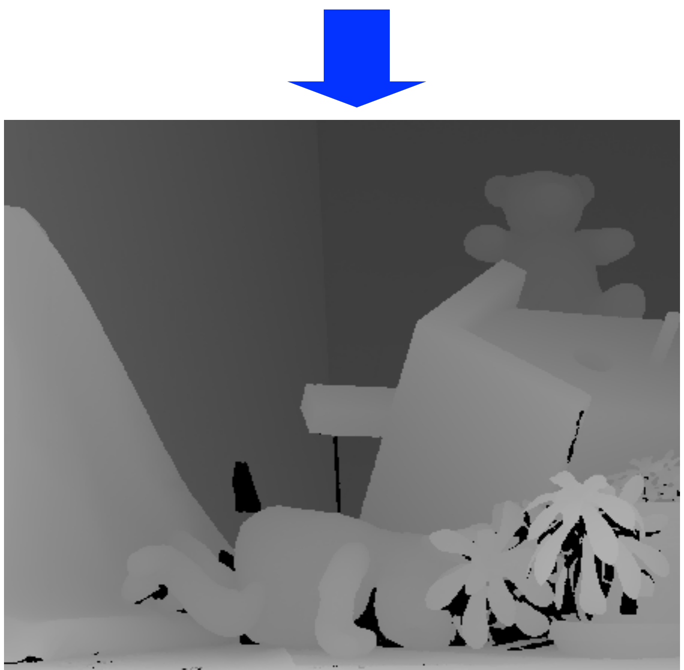
www.nvela.com

*

www.minoru3dwebcam.com

Why is stereo correspondence so challenging ?



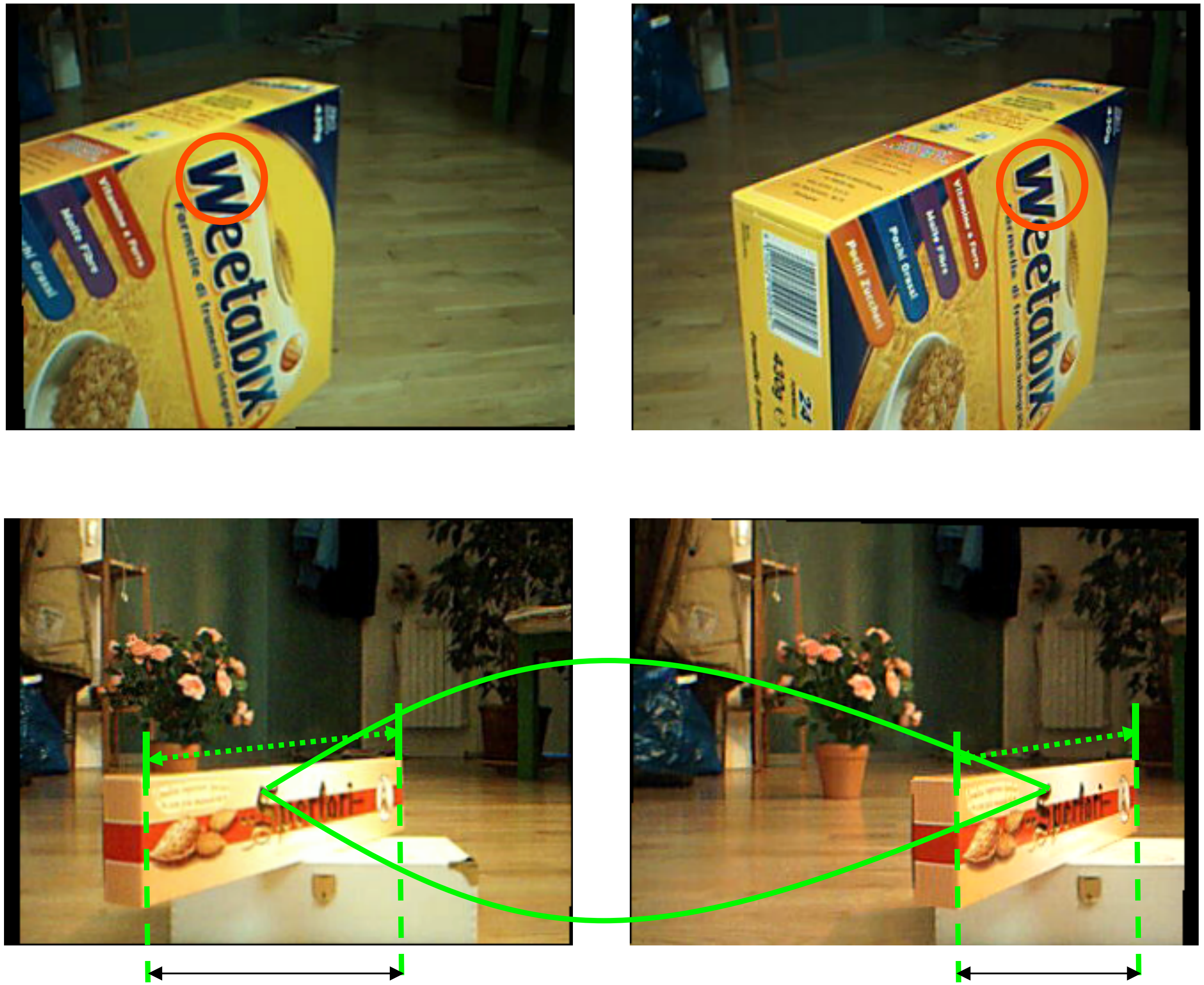


Next slides show common pitfalls…

Photometric distortions and noise

Specular surfaces

Foreshortening



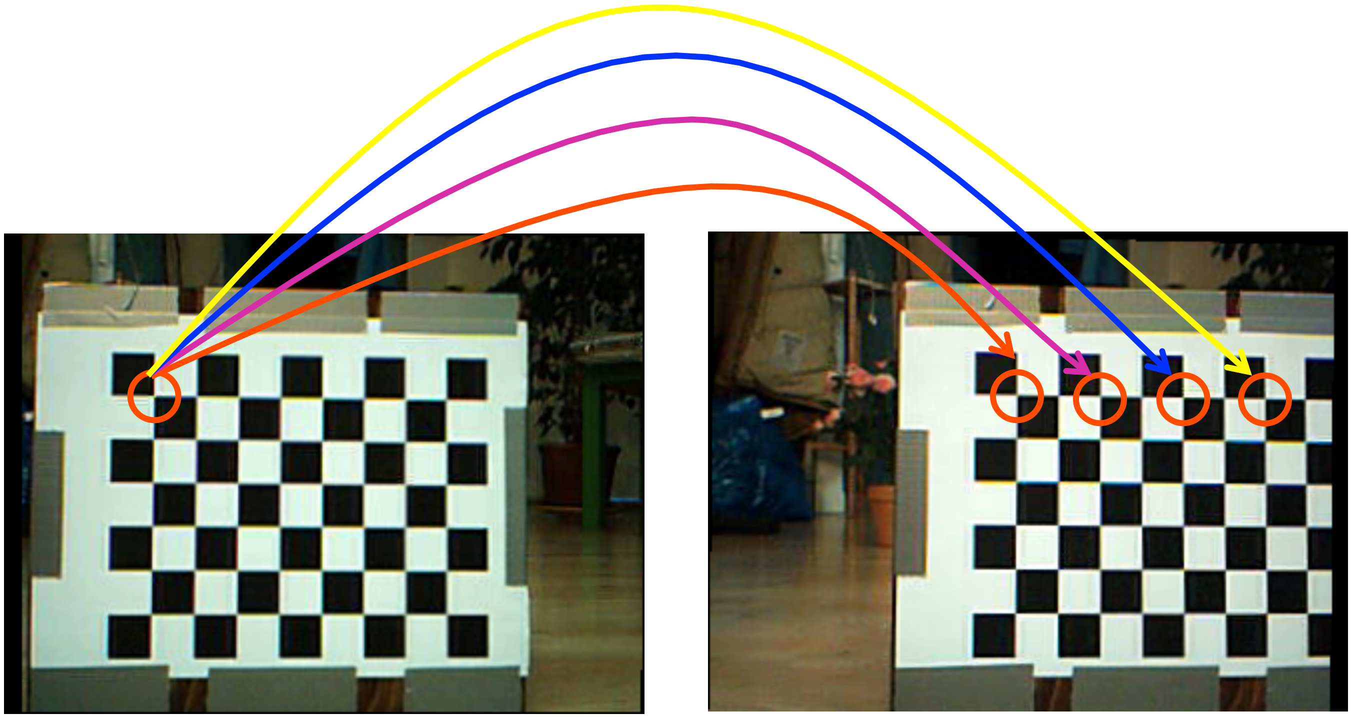
Uniqueness constraint ? :-(

Stefano Mattoccia

Perspective distortions

Uniform/ambiguous regions

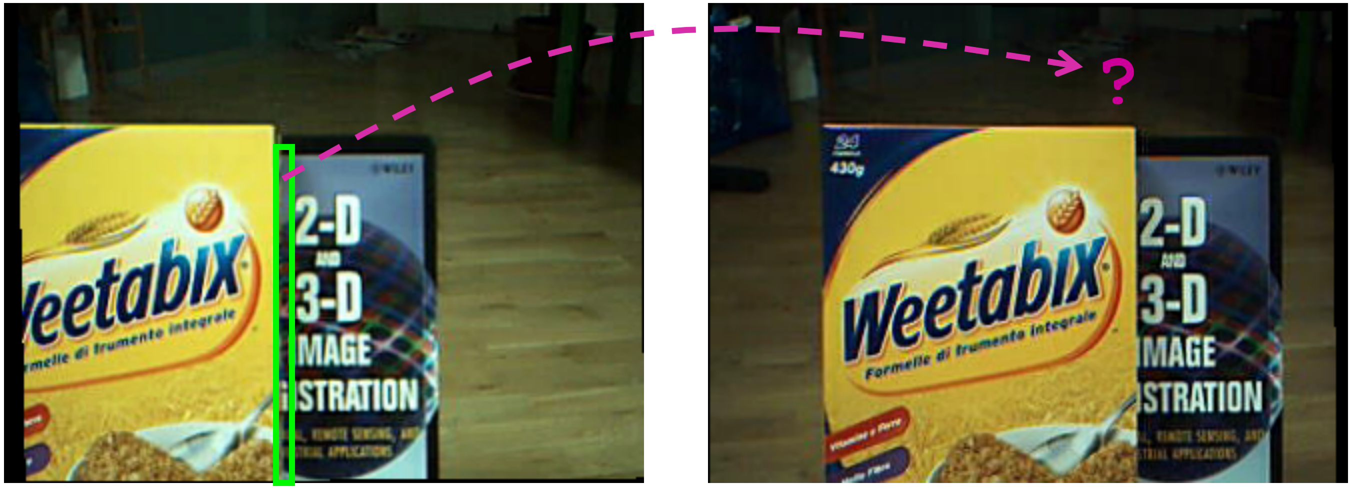
Repetitive/ambiguous patterns



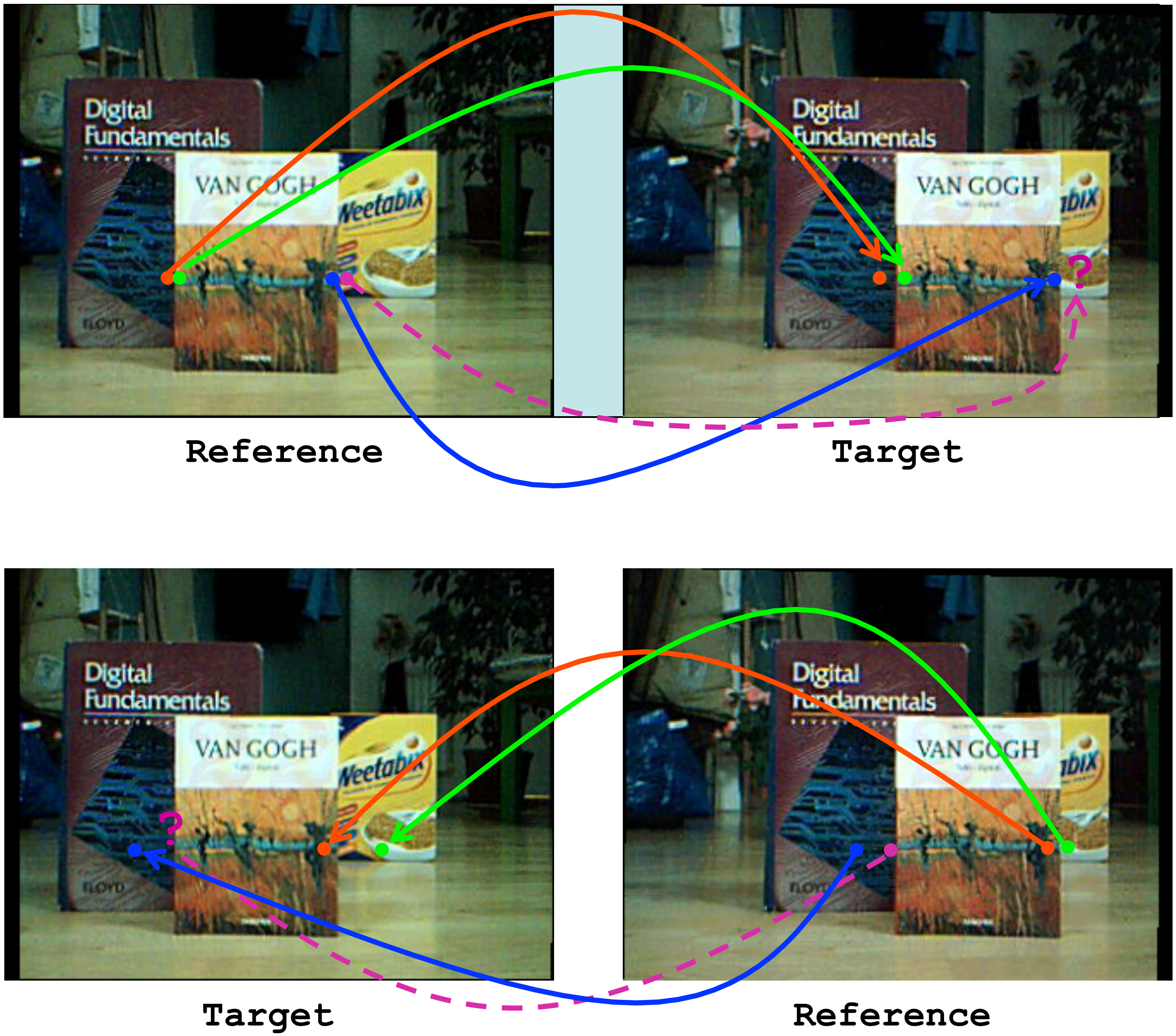
How to reduce ambiguity... ?

Transparent objects

Occlusions and discontinuities 1/2



Occlusions and discontinuities 2/2



Middlebury stereo evaluation

The Middlebury stereo evaluation site [15] provides a framework and a dataset (showed in the next slide) for benchmarking novel algorithms.

Scharstein and Szeliski provide:

- **a methodology for the evaluation of (binocular) stereo vision algorithms [11]**
- **datasets with groundtruth [11,15,17,18,19]**
- **online evaluation procedure and ranking [15]**

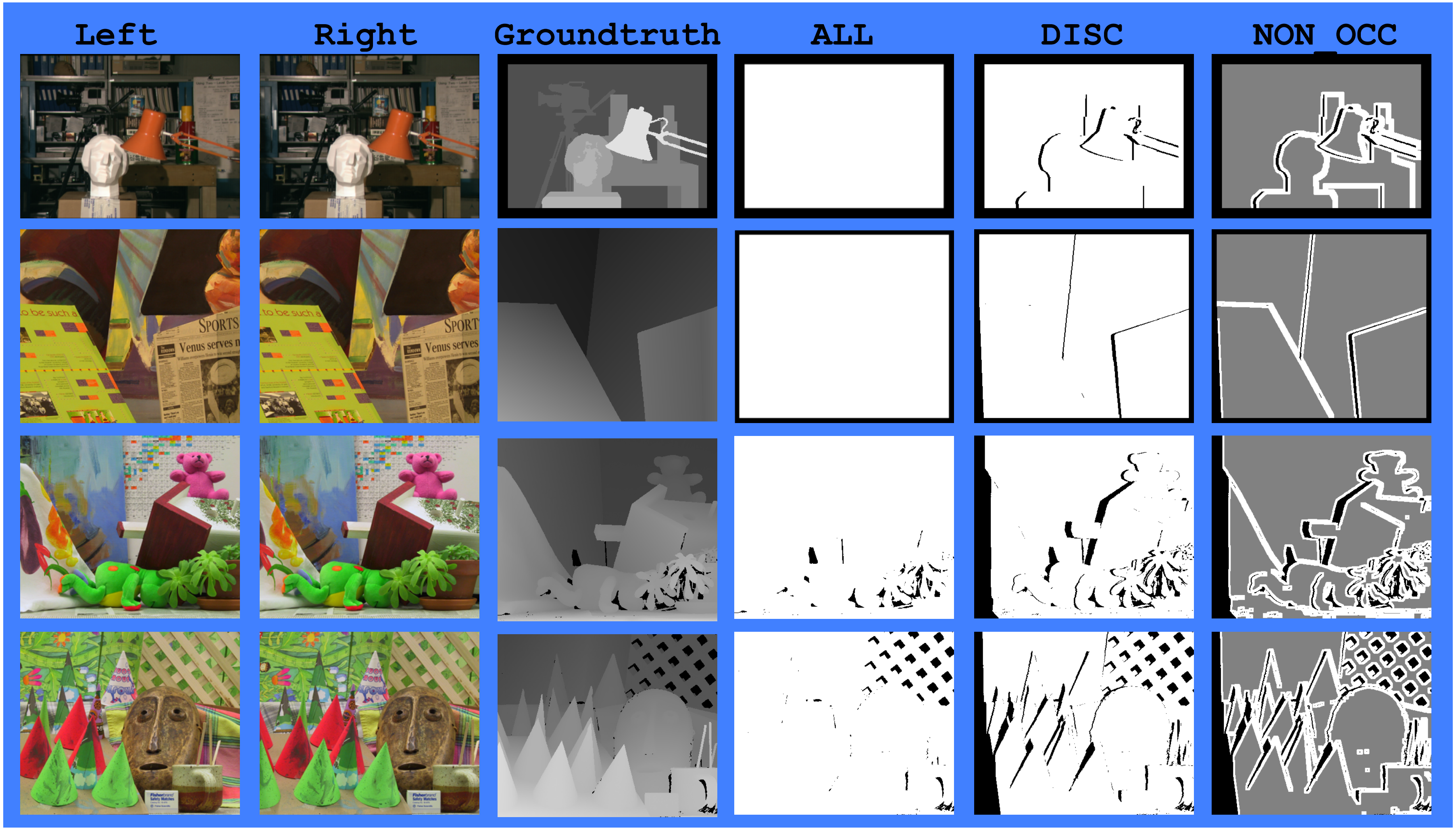
Datasets (with groundtruth) of stereo pairs affected by photometric distortions are also available in [15].

[15] D. Scharstein and R. Szeliski, http://vision.middlebury.edu/stereo/eval/

[11] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms" Int. Jour. Computer Vision, 47(1/2/3):7–42, 2002

Middlebury dataset (2003) [15]

Tsukuba, Venus, Teddy and Cones stereo pairs



D. Scharstein and R. Szeliski, http://vision.middlebury.edu/stereo/eval/

The correspondence problem

According to the taxonomy proposed in [11] most stereo algorithms perform (subset of) these steps:

1) Matching cost computation
2) Cost aggregation
3) Disparity computation/optimization
4) Disparity refinement

Local algorithms perform:
1 ⇒ 2 ⇒ 3 (with a simple Winner Takes All (WTA) strategy)

Global Algorithms perform:
1 (⇒ 2) ⇒ 3 (with global or semi-global reasoning)

Pre-processing (0)

Sometime is deployed a pre-processing stage mainly to compensate for photometric distortions.

Typical operations include:

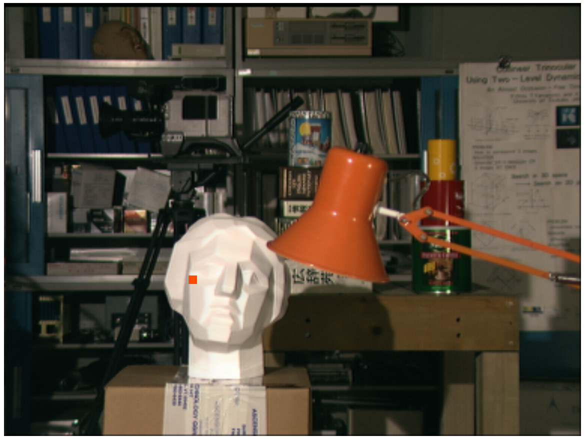
- **Laplacian of Gaussian (LoG) filtering [41]**
- **Subtraction of mean values computed in nearby pixels [42]**
- **Bilateral filtering [16]**
- **Census transform**

[41] T. Kanade, H. Kato, S. Kimura, A. Yoshida, and K. Oda, Development of a Video-Rate Stereo Machine International Robotics and Systems Conference (IROS '95), Human Robot Interaction and Cooperative Robots, 1995

[42] O. Faugeras, B. Hotz, H. Mathieu, T. Viville, Z. Zhang, P. Fua, E. Thron, L. Moll, G. Berry, Real-time correlation-based stereo: Algorithm. Implementation and Applications, INRIA TR n. 2013, 1993

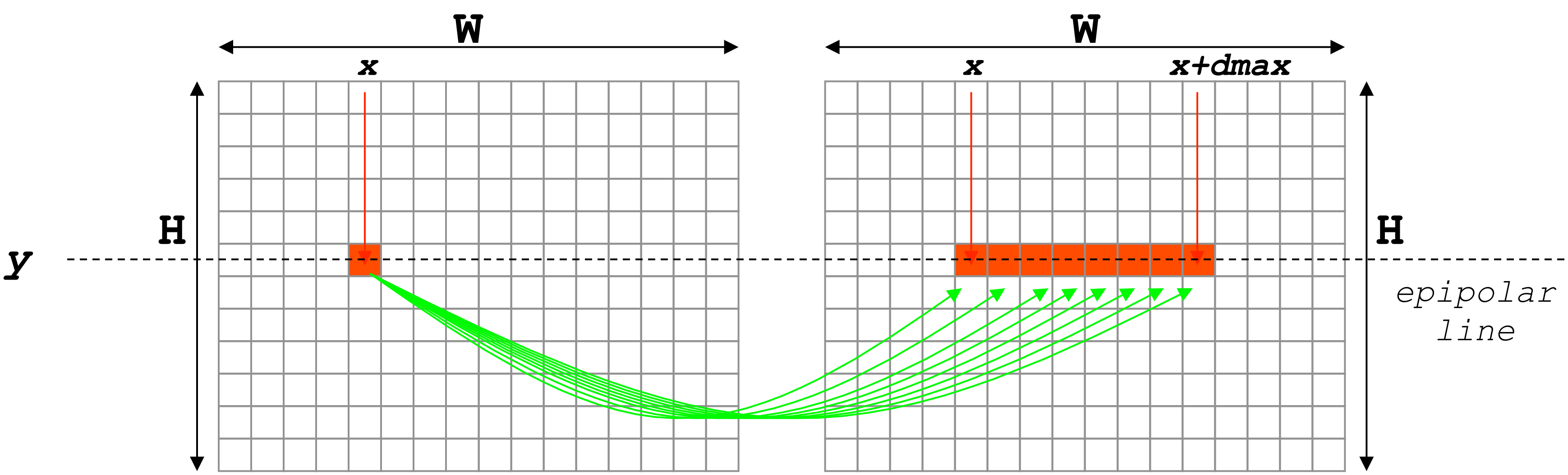
[16] A. Ansar, A. Castano, L. Matthies, Enhanced real-time stereo using bilateral filtering IEEE Conference on Computer Vision and Pattern Recognition 2004

The simplest (naive and unused) local approach:



Reference (R)

Target (T)

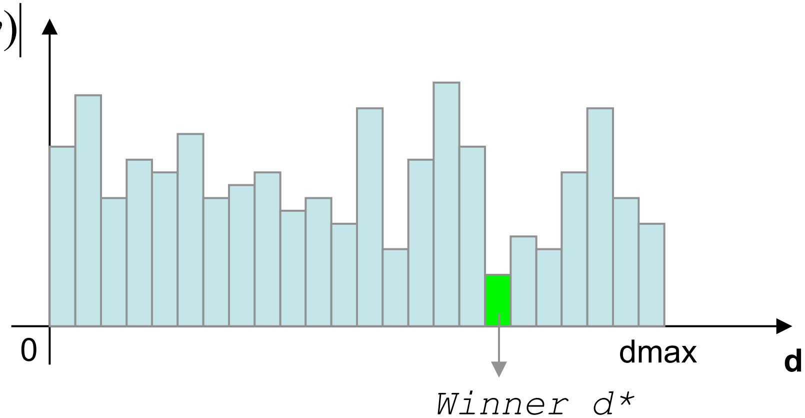


Reference (R)

Target (T)

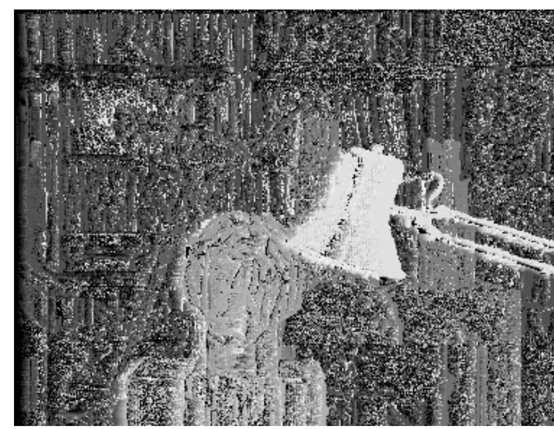
- **matching cost (1): pixel-based absolute difference between pixel intensities**
- **disparity computation (3): Winner Takes All (WTA)**

$$|I_R(x,y) - I_T(x+d,y)|$$



Reference

Groundtruth



Result (disappointing)

Stefano Mattoccia

How to improve the results of the naive approach ?

Basically exist two different (not mutually exclusive) strategies:

- **Local algorithms** use the simple WTA disparity selection strategy but reduce ambiguity (increasing the signal to noise ratio (SNR)) by aggregating matching costs over a support window (aka kernel or correlation window). Sometime a smoothness term is adopted. **Steps 1+2 (+ WTA)**

- **Global (and semi-global*)** algorithms search for disparity assignments that minimize an energy function over the whole stereo pair using a pixel-based matching cost (sometime the matching cost is aggregated over a support). **Steps 1+3**

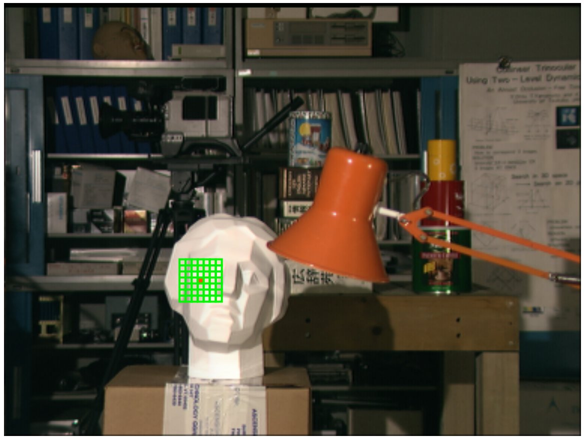
Both approaches assume that the scene is piecewise smooth. Sometime this assumption is violated...

This hypothesis is implicitly assumed by local approaches while it is explicitly modelled by global approaches

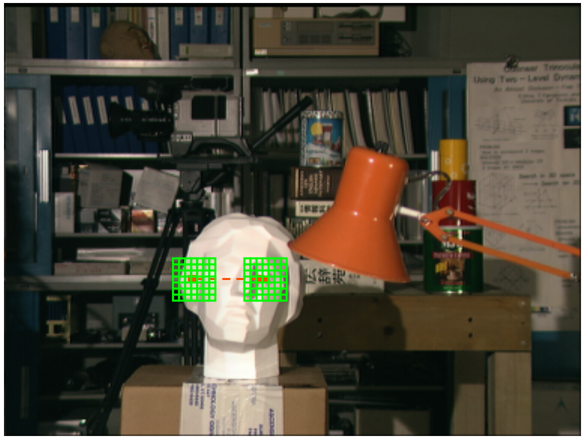
* subset of the stereo pair

Local approaches:

In order to increase the SNR (reduce ambiguity) the matching costs are aggregated over a support window



Reference (R)



Target (T)

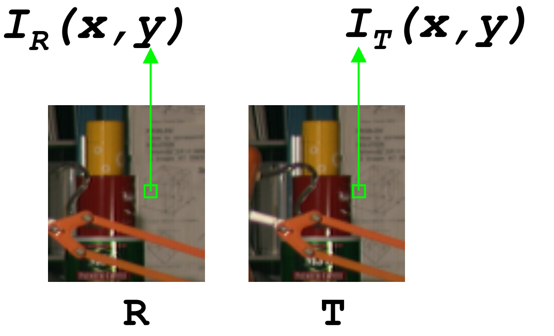
Global (and semi-global*) approaches:

Many algorithms search for the disparity assignment that minimize a certain cost function over the whole* stereo pair

$$E(d) = E_{data}(d) + E_{smooth}(d)$$

* subset of the stereo pair

Matching cost computation (1)



Pixel-based matching costs

- Absolute differences

$$e(x,y,d) = \left| I_R(x,y) - I_T(x+d,y) \right|$$

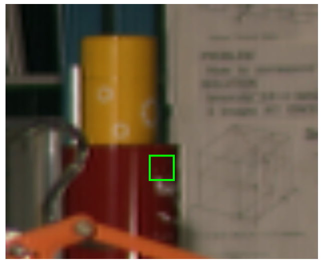
- Squared differences

$$e(x,y,d) = \left(I_R(x,y) - I_T(x+d,y) \right)^2$$

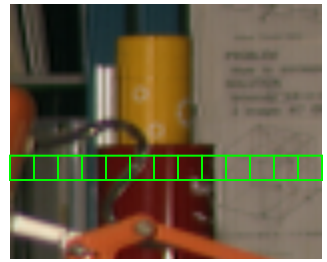
- Robust matching measures (M-estimators)
  - Limit influence of outliers
  - Example: truncated absolute differences (TAD)

$$e(x,y,d) = \min\left\{ \left| I_R(x,y) - I_T(x+d,y) \right|, T \right\}$$

- **Dissimilarity measure insensitive to image sampling (Birchfield and Tomasi [27])**

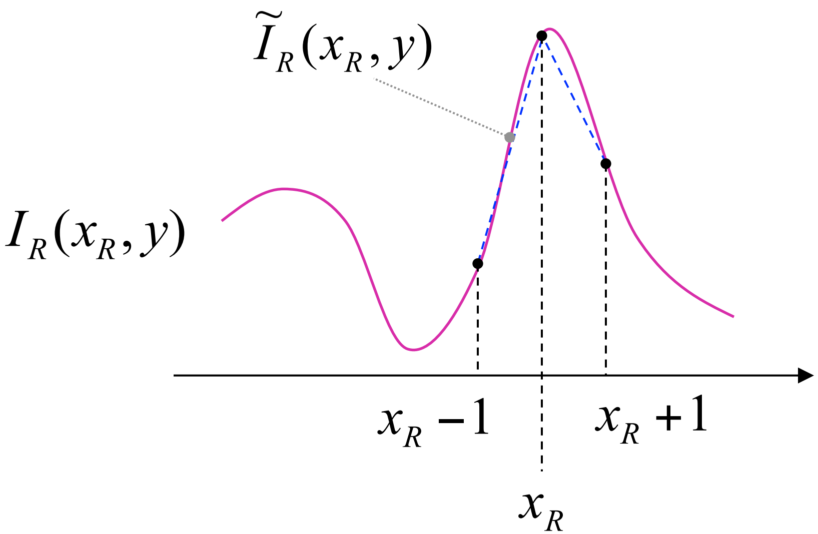


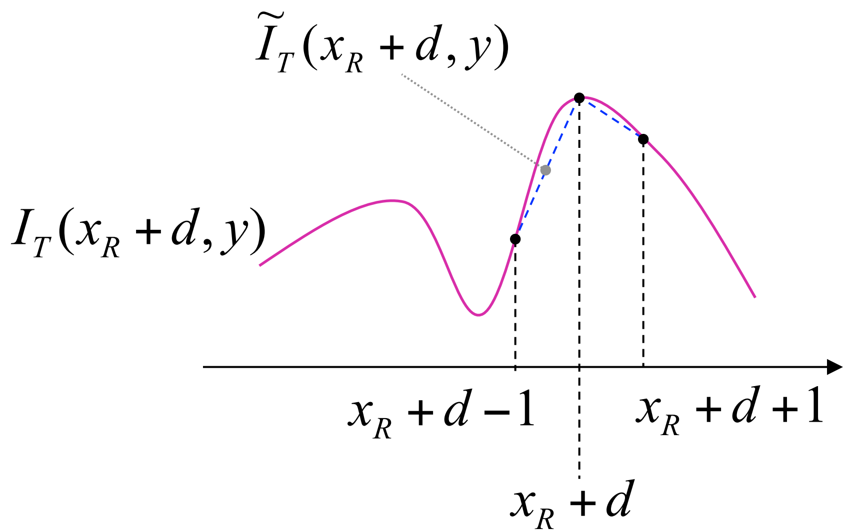
Reference (R)



Target (T)

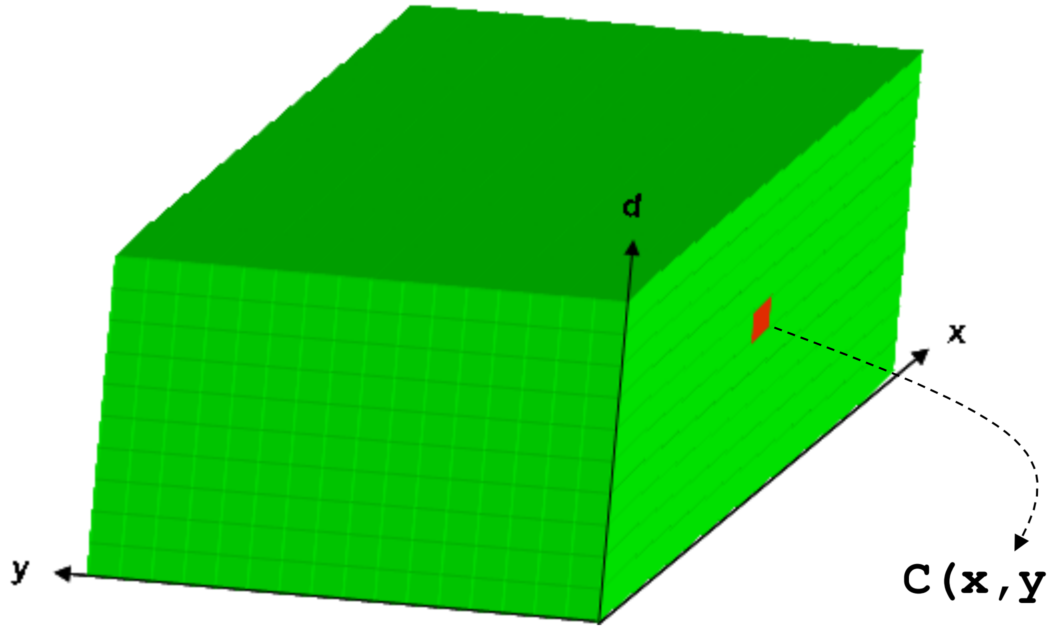
BT helps at depth and color discontinuities





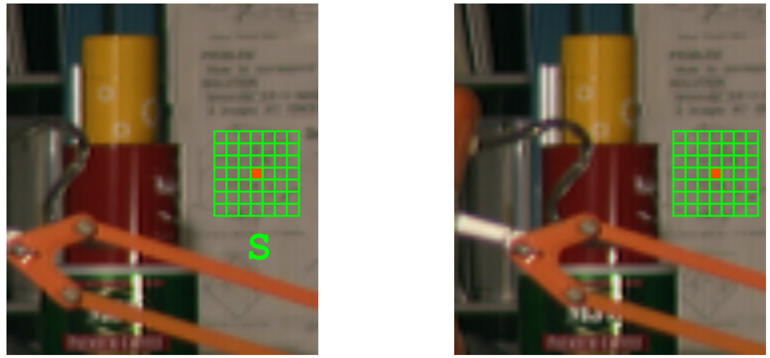
$$e(x_R, y, d) = \min\left\{ \min_{x_R - \frac{1}{2} \le x \le x_R + \frac{1}{2}} \left| I_R(x_R, y) - \tilde{I}_T(x + d, y) \right|, \min_{x_R - \frac{1}{2} \le x \le x_R + \frac{1}{2}} \left| I_T(x_R + d, y) - \tilde{I}_R(x, y) \right| \right\}$$

The Disparity Space Image (DSI) is a 3D matrix (WxHx(d_{max}-d_{min})



Each element C(x,y,d) of the DSI represents the cost of the correspondence between $I_R(x_R,y)$ and $I_T(x_R+d,y)$

Area-based matching costs:



- **Sum of Absolute differences (SAD)**

$$C(x,y,d)=\sum_{x\in S}\left|I_R(x,y)-I_T(x+d,y)\right|$$

- **Sum of Squared differences (SSD)**

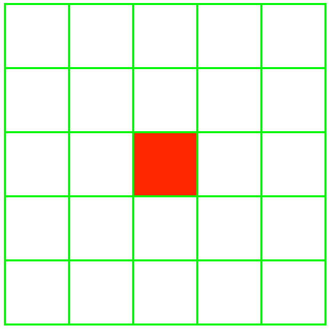
$$C(x,y,d)=\sum_{x\in S}\left(I_R(x,y)-I_T(x+d,y)\right)^2$$

- **Sum of truncated absolute differences (STAD)**

$$C(x,y,d)=\sum_{x\in S}\min\left\{\left|I_R(x,y)-I_T(x+d,y)\right|,T\right\}$$

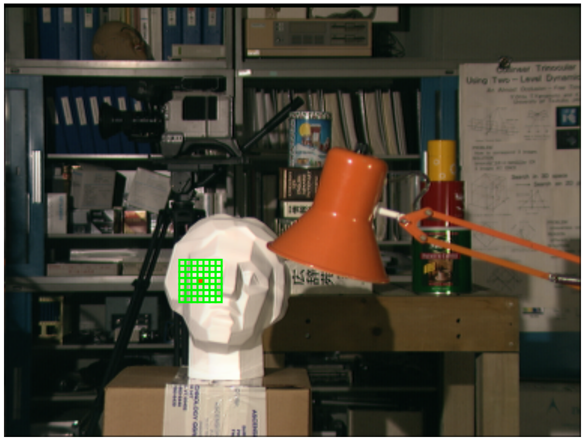
- **Normalized Cross Correlation [57]**
- **Zero mean Normalized Cross Correlation [58]**
- **Gradient based MF [59]**
- **Non parametric [60,61]**
- **Mutual Information [30]**
- **. . .**
- **Combination of matching costs**

Add content here
Area-based matching costs

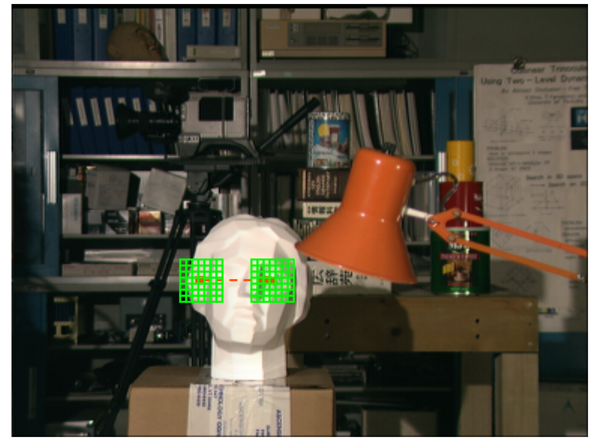


Cost aggregation (2)

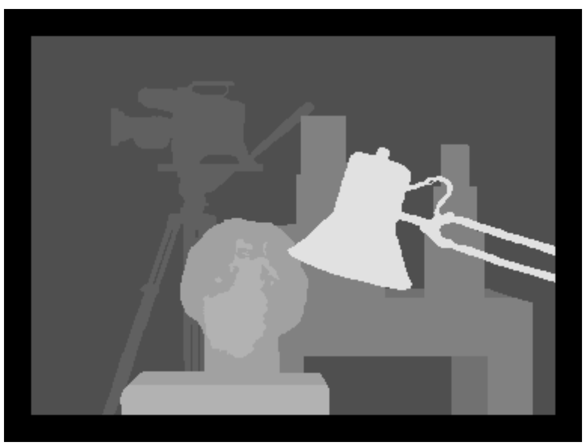
Let's start by examining the simplest Fixed Window (FW) cost aggregation strategy (TAD, disparity selection WTA)



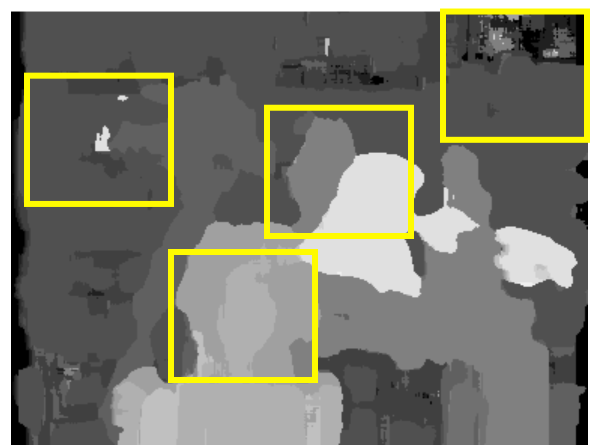
Reference (R)



Target (T)



Groundtruth

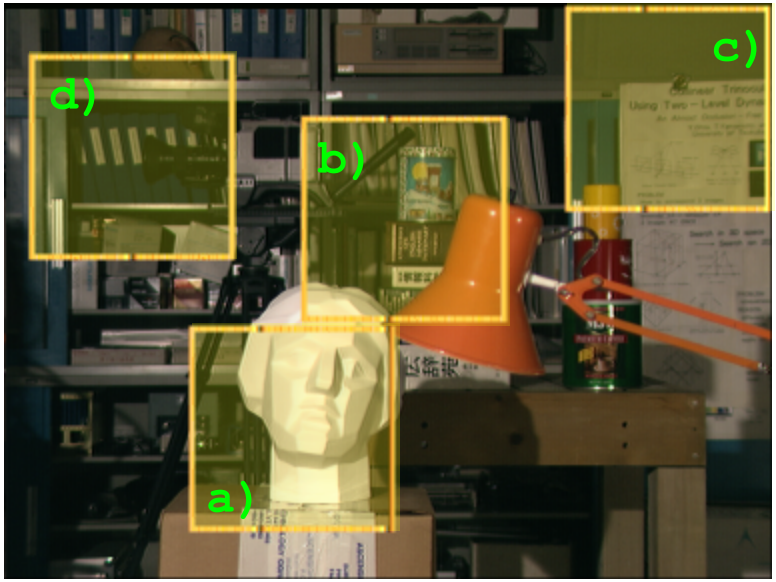


Fixed Window (FW)

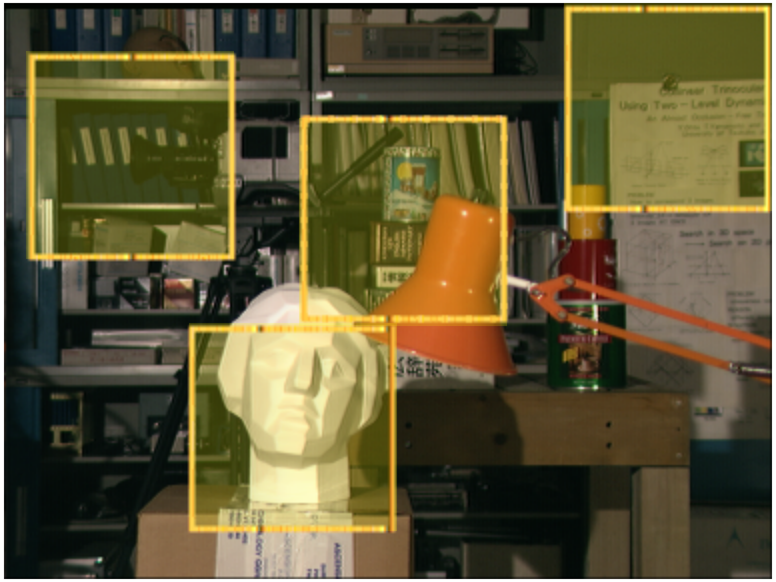
What's wrong with FW ?

Stefano Mattoccia

FW (with WTA reasoning) fails in most points for the following reasons:



Reference (R)



Target (T)

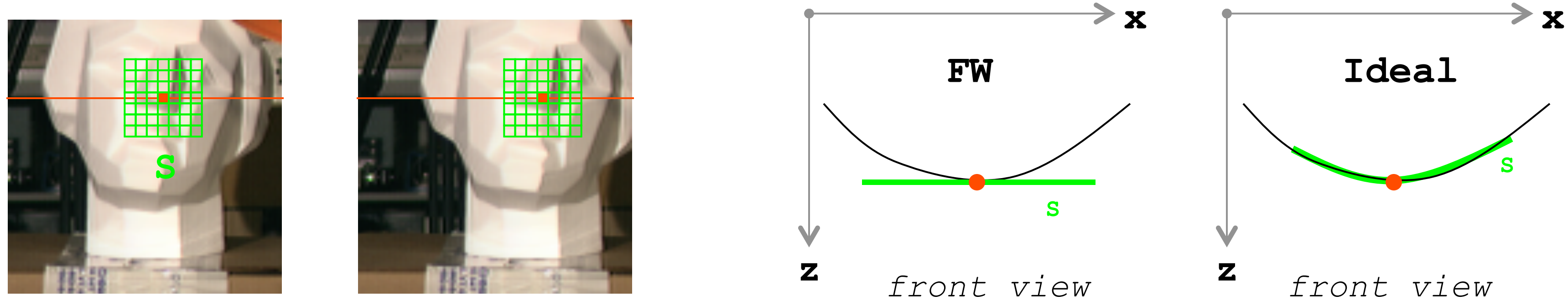
a) implicitly assumes frontal-parallel surfaces

b) ignores depth discontinuities

c) does not deal explicitly with uniform areas

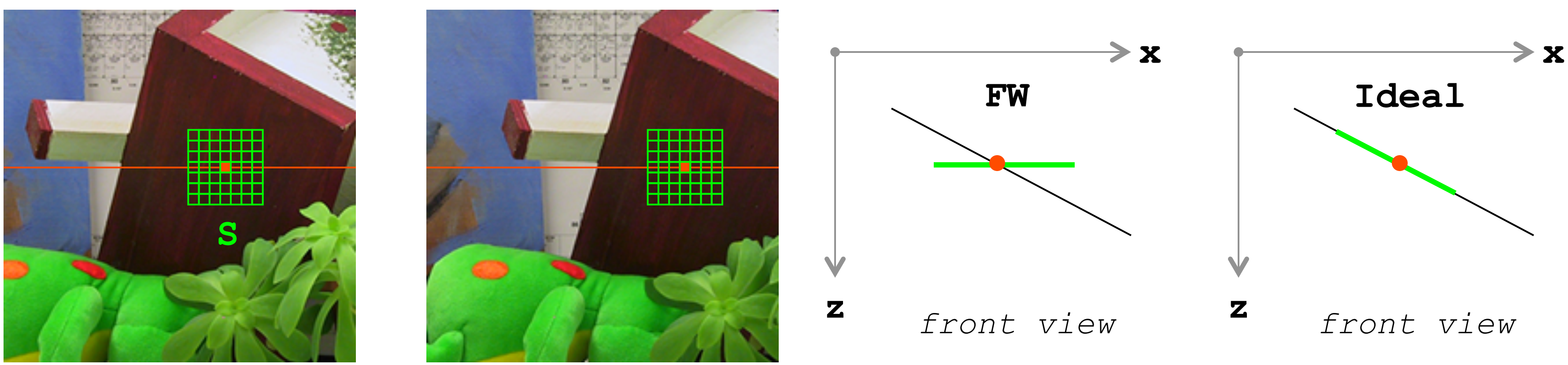
d) does not deal explicitly with repetitive patterns

a) FW implicitly assumes frontal-parallel surfaces



FW

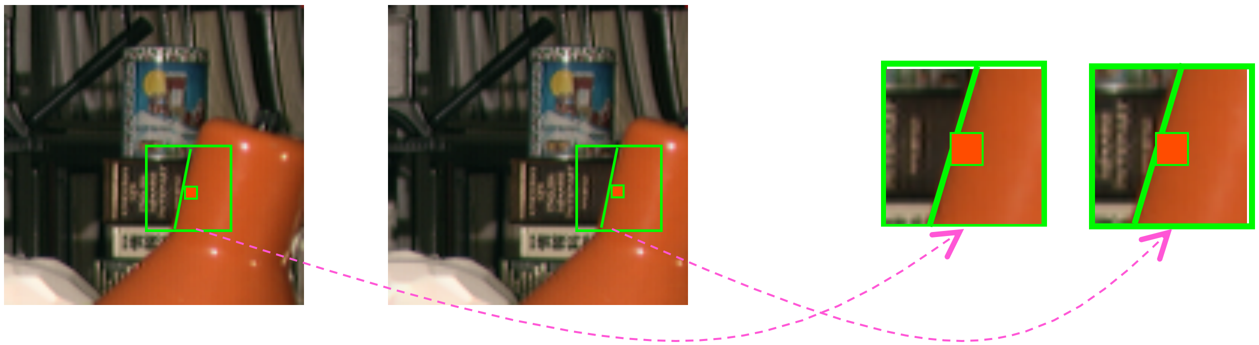
Often violated in practice: top figure, slanted surfaces (down), etc.



Nevertheless, almost all state-of-the-art cost aggregation strategies rely on the assumption that all the points belonging to the support share the same disparity (only few exceptions).

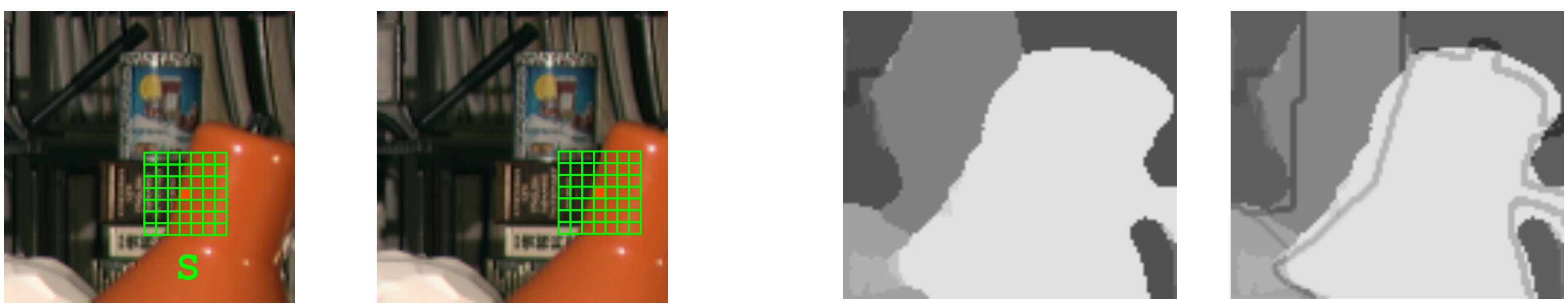
b) FW ignores depth discontinuities

Implicitly assuming frontal-parallel surface in the real scene is violated near depth discontinuities.



Background is misaligned !

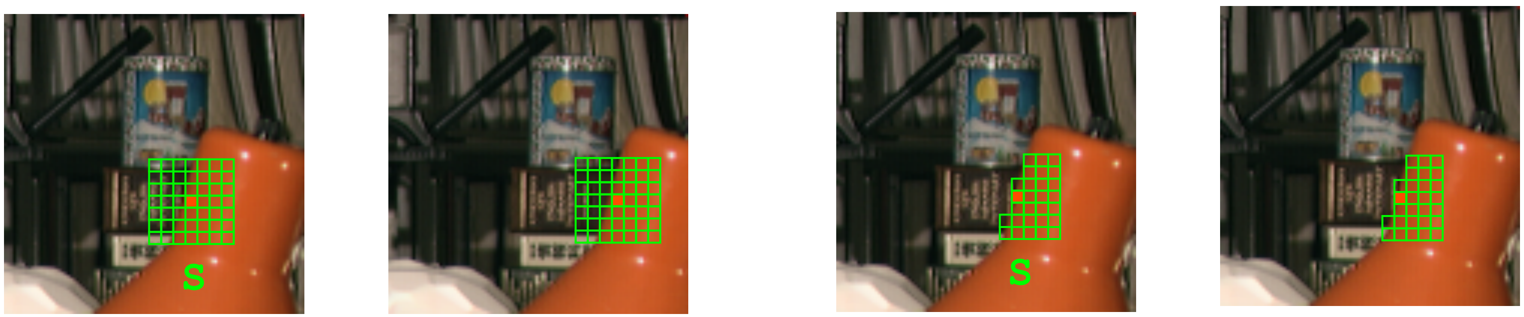
Aggregating the matching costs of two populations at different depth (aligned foreground and misaligned background (outliers)) results in the typical inaccurate localization of depth borders.



FW

Robust matching measures (TAD) can partially reduce the influence of outliers

State-of-the-art cost aggregation strategies aim at shaping the support in order to include only points with the same (unknown) disparity.



FW

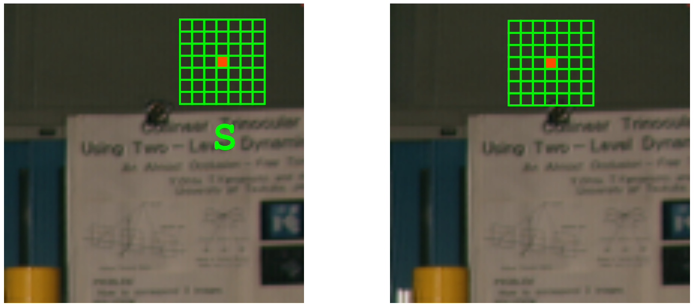
Ideal

For what concerns FW: decreasing the size of the support helps in reducing the border localization problem.

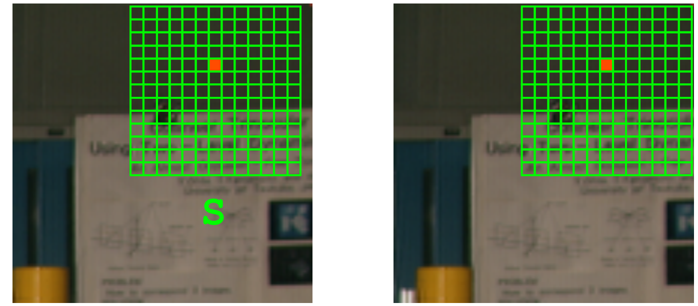
However, this choice renders the correspondence problem more ambiguous (especially when dealing with uniform regions and repetitive patterns, see the next slide).

In practice, for the FW approach the choice of the optimal size of the support is done empirically.

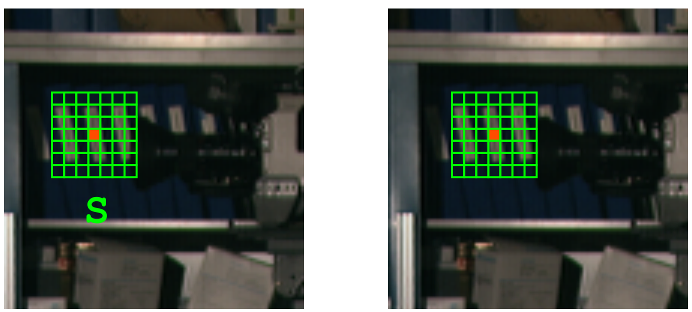
FW does not deal explicitly with ambiguous regions - uniform areas c) and repetitive patterns d)



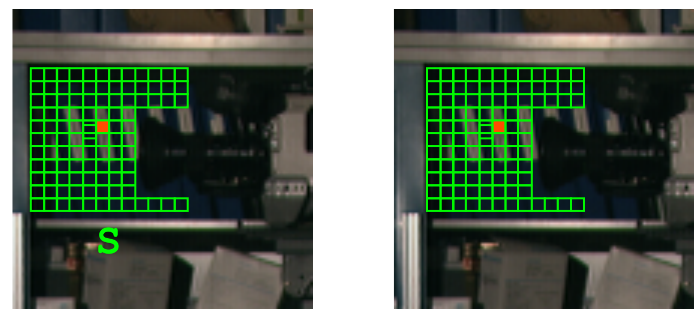
FW



Ideal



FW



Ideal

In both cases an ideal cost aggregation strategy should extend its support in order to include as much points at the same (unknown) depth as possible.

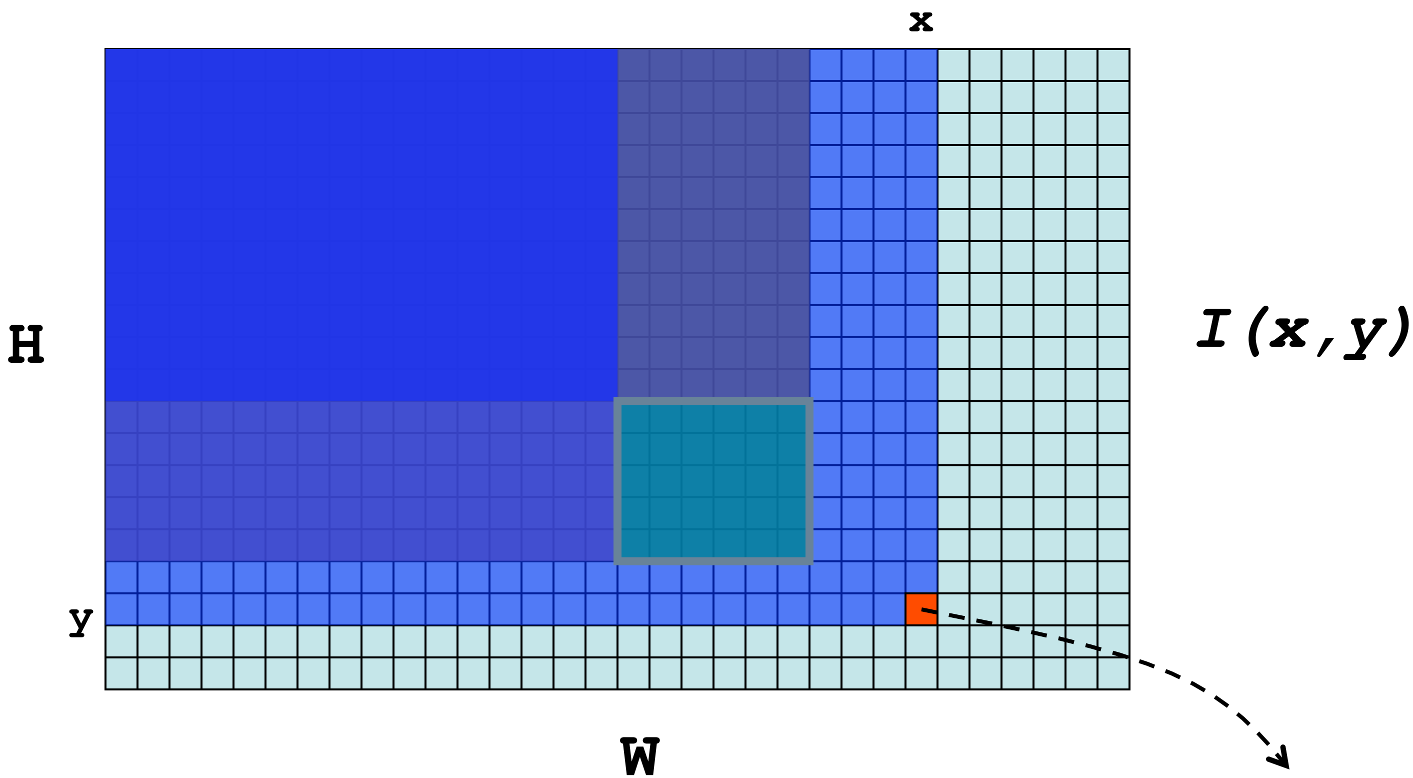
Quite surprisingly, in spite of its limitations, FW is widely adopted in practice (probably it is the most frequently used algorithm for real applications).

- **Easy/fast implementation**
- **Fast, thanks to incremental calculation schemes**
- **Runs in real-time on standard processors (SIMD)**
- **Has limited memory requirements**
- **Hardware implementations (FPGA) run in real-time with limited power consumption (<1W)**

Before analyzing more sophisticated approaches let's consider two optimization techniques used by FW and other algorithms:

- **Integral Images (II)**
- **Box-Filtering (BF)**

Optimization: Integral Images (aka Summmed Area Table)



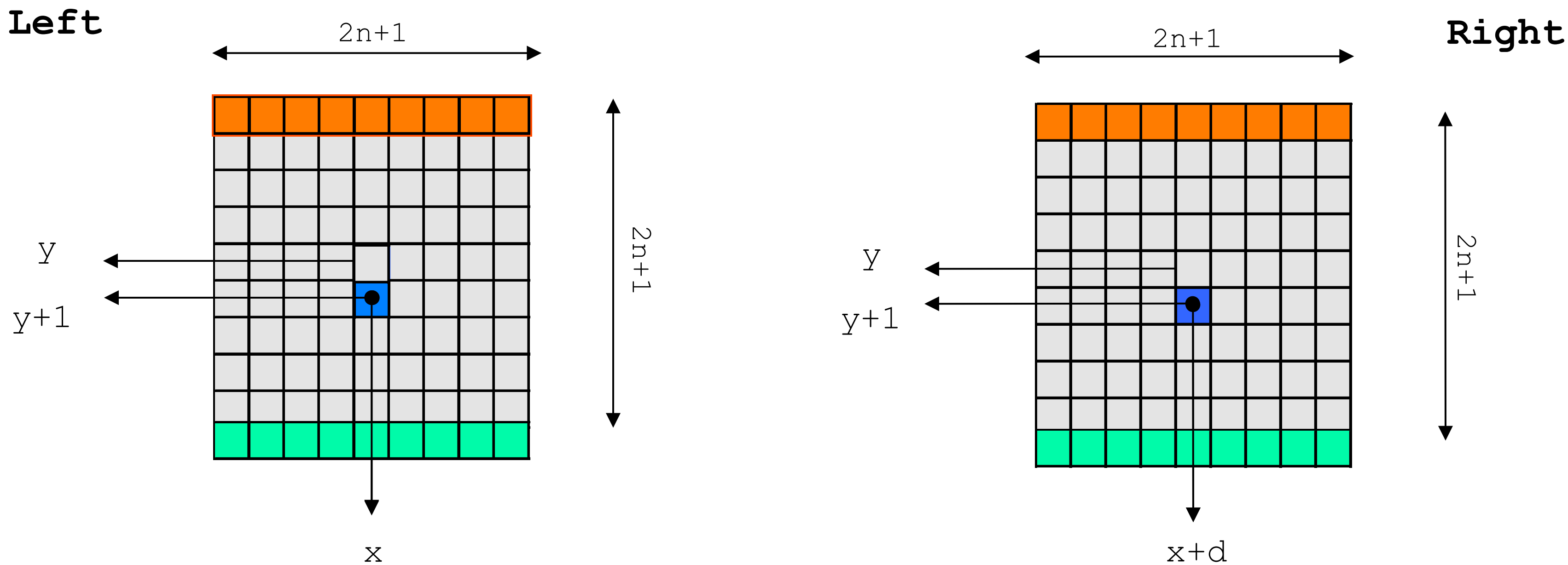
$$S(x,y) = \sum_{i<x,\, j<y} I(i,j)$$

$$S^2(x,y) = \sum_{i<x,\, j<y} I^2(i,j)$$

- **Straightforward extension to stereo (2 images)**

F. Crow, Summed-area tables for texture mapping, Computer Graphics, 18(3):207–212, 1984

Optimization: Box-Filtering 1/2



$$SAD(x,y,d) = \sum_{i,j=-n}^{n} \left| L(x+j, y+i) - R(x+d+j, y+i) \right|$$

$$SAD(x,y+1,d) = SAD(x,y,d) + U(x,y+1,d)$$

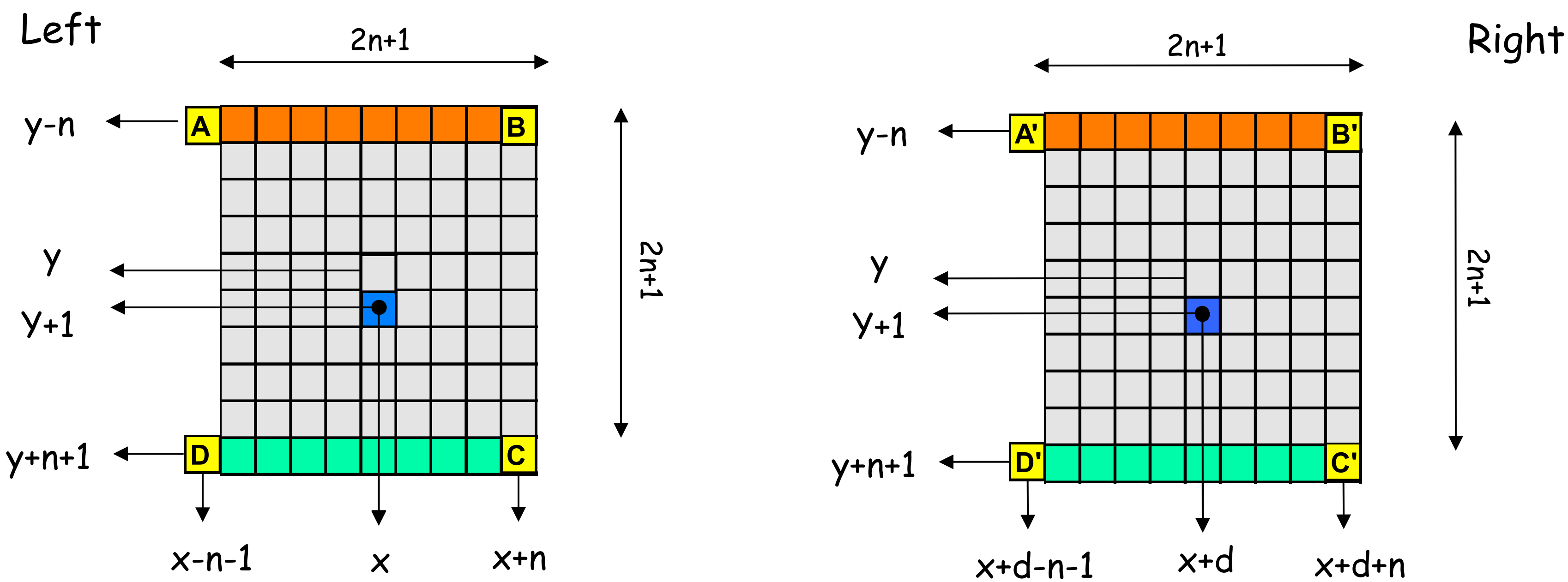
$$U(x,y+1,d) = \boxed{\sum_{j=-n}^{n} \left| L(x+j, y+n+1) - R(x+d+j, y+n+1) \right|} - \boxed{\sum_{j=-n}^{n} \left| L(x+j, y-n) - R(x+d+j, y-n) \right|}$$

M. Mc Donnel. Box-filtering techniques. Computer Graphics and Image Processing, 17:65–70, 1981

Optimization: Box-Filtering 2/2

$$SAD(x, y+1, d) = SAD(x, y, d) + U(x, y+1, d) \qquad d \in [0..d_{max}]$$

$$U(x, y+1, d) = \sum_{j=-n}^{n} |L(x+j, y+n+1) - R(x+d+j, y+n+1)| - \sum_{j=-n}^{n} |L(x+j, y-n) - R(x+d+j, y-n)|$$



$$U(x, y+1, d) = U(x-1, y+1, d) + |A - A'| - |B - B'| + |C - C'| - |D - D'|$$

$$SAD(x, y+1, d) = SAD(x, y, d) + U(x-1, y+1, d) + |A - A'| - |B - B'| + |C - C'| - |D - D'|$$

Box-Filtering Vs Integral Images

- Both require 4 operations per point

- *Integral images* can handle supports of different size

- *Integral Images* has overflow issues
  (for example, with int32 and $S^2 \Rightarrow$ WxH<256x256)

- *Integral images* is more demanding in terms of memory requirements. For single images:

  WxHxsizeof(data_type) Vs ≈ Wxsizeof(int32) for S^2

In practice, integral images may be convenient when supports of different size are required.

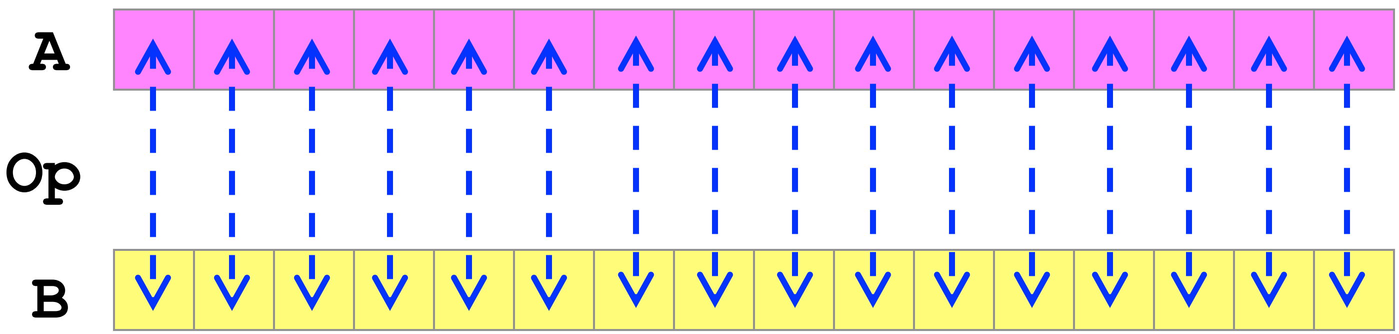
Extension of box-filtering to more complex shapes was proposed in [47].

Optimizations: Single Instruction Multiple Data (SIMD)

a Op **b**

Scalar computation 1 Op

It's a computation paradigm that that allow for processing with the same operation multiple data in parallel.



SIMD 16 Ops in parallel !

- **Several computer vision algorithms are suited for SIMD**
- **SIMD features are available in most current processors**
- **Intel processors SIMD available since Pentium (MMX)**
- **SIMD mapping is difficult (assembly)**

Single Matching Phase Algorithm [48,49]

- *Image type: grayscale*
- *Preprocessing: subtraction of mean values*
- *Matching cost (Step 1): Absolute Differences*
- *Aggregation strategy (Step 2): FW*
- *Disparity selection (Step 3): WTA*
- *Outlier detection: efficient strategy (later, Step 4)*
- *Discards uniform areas: yes, analyzing image variance*
- *Optimizations: box-filtering + SIMD instructions (SSE)*
- *Sub-pixel interpolation up to 1/16 of pixel (later)*
- *Runs in real-time on a standard PC*

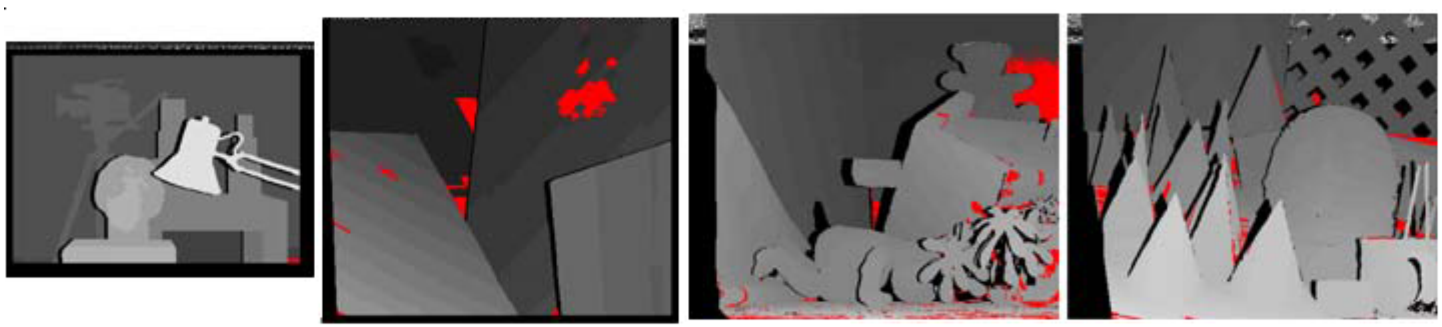
L. Di Stefano, M. Marchionni, S. Mattoccia, A fast area-based stereo matching algorithm
Image and Vision Computing, 22(12), pp 983-1005, October 2004

L. Di Stefano, M. Marchionni, S. Mattoccia, A PC-based real-time stereo vision system
Machine Graphics & Vision, 13(3), pp. 197-220, January 2004

How far can we go with more effective (frontal parallel) cost aggregation strategies ?

We made an experiment computing ideal frontal parallel supports using the ground truth.

With 43x43 max support, TAD and a WTA strategy:



Results (errors in red)

There is room for improvements...

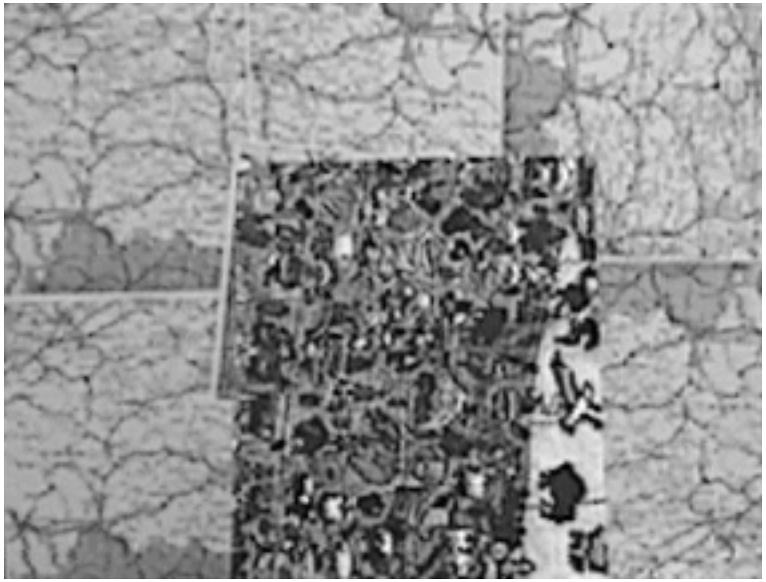
- Compared to pixel-based approaches the support aggregation (potentially) allows for improving robustness

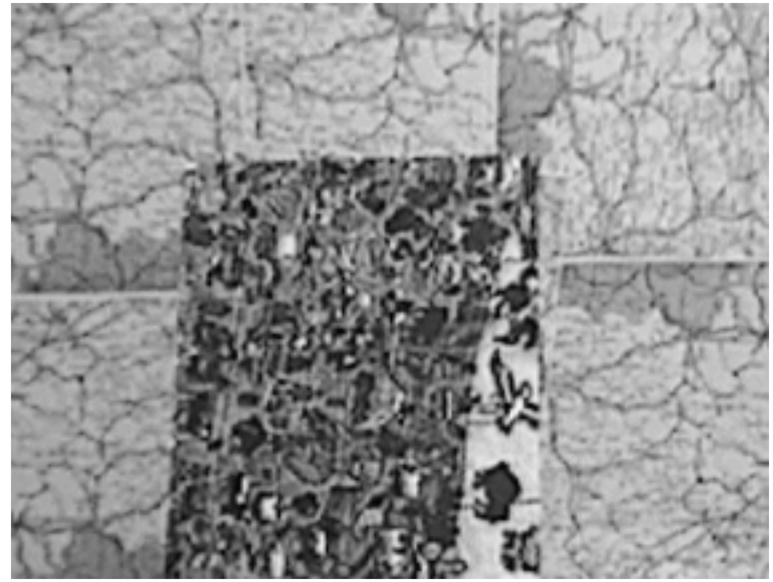
- An ideal (frontal parallel) cost aggregation strategies should include in the support only points with similar disparity:

    - expanding in regions at similar depth (left)

    - shrinking near depth discontinuities (right)

What about symmetric/asymmetric support, discontinuities and occlusions ?

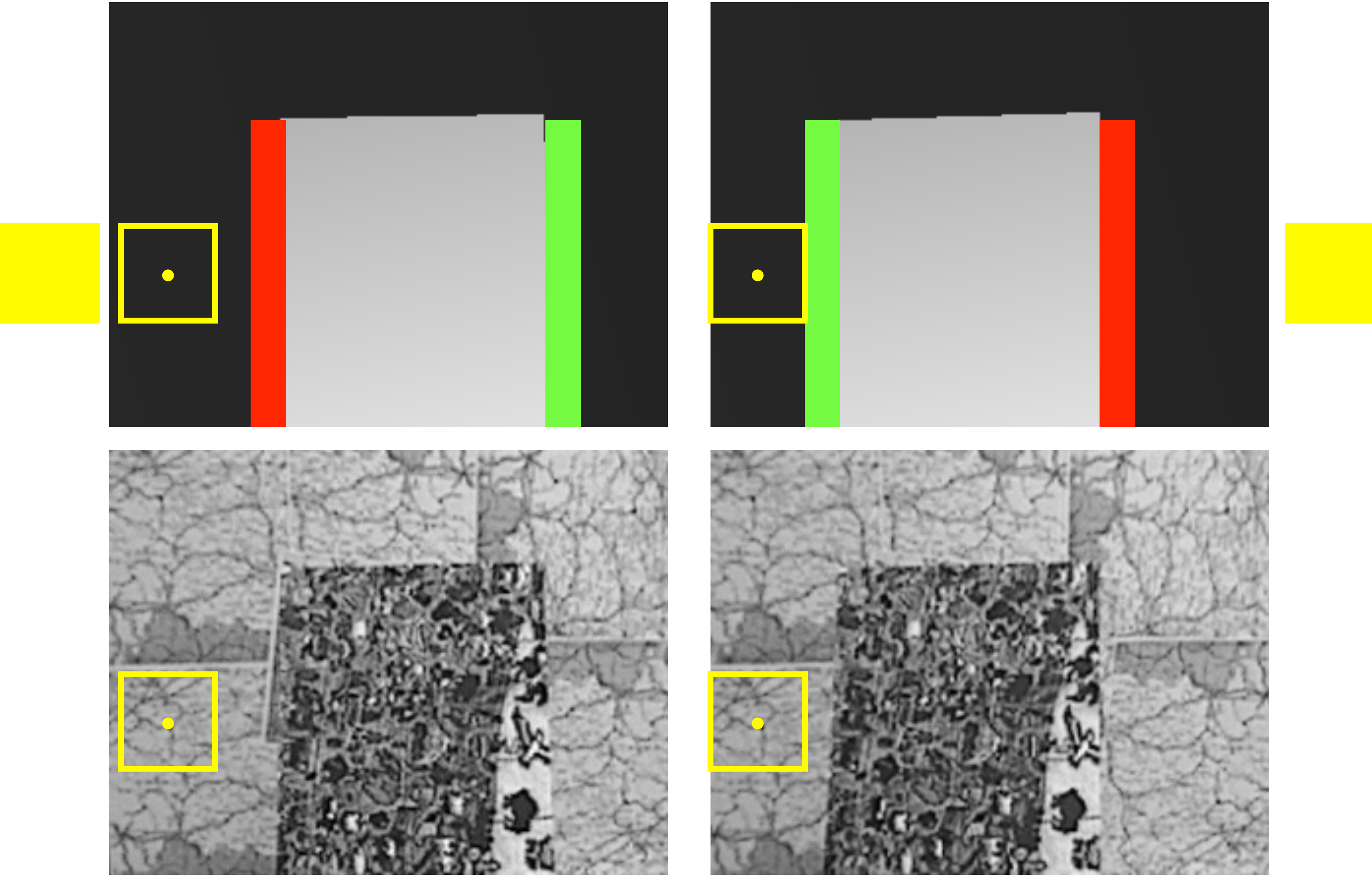




- (Unknown) Occlusions and discontinuities play a central role for support aggregation strategies. The next slides depict relevant cases using a simple object laying on a planar background

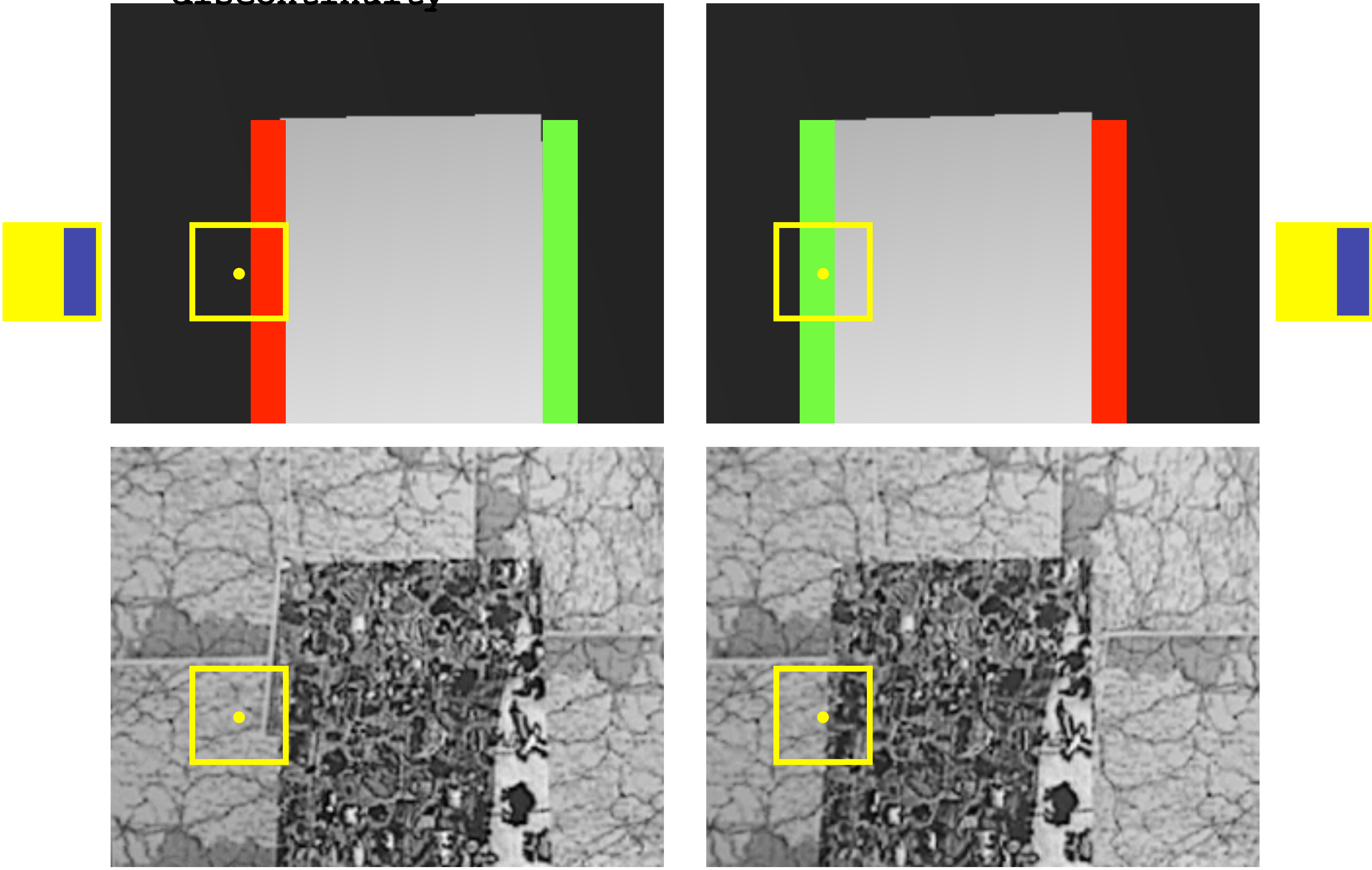
- Occlusions and discontinuities are strictly related

Case 1: no half occlusion, no discontinuity



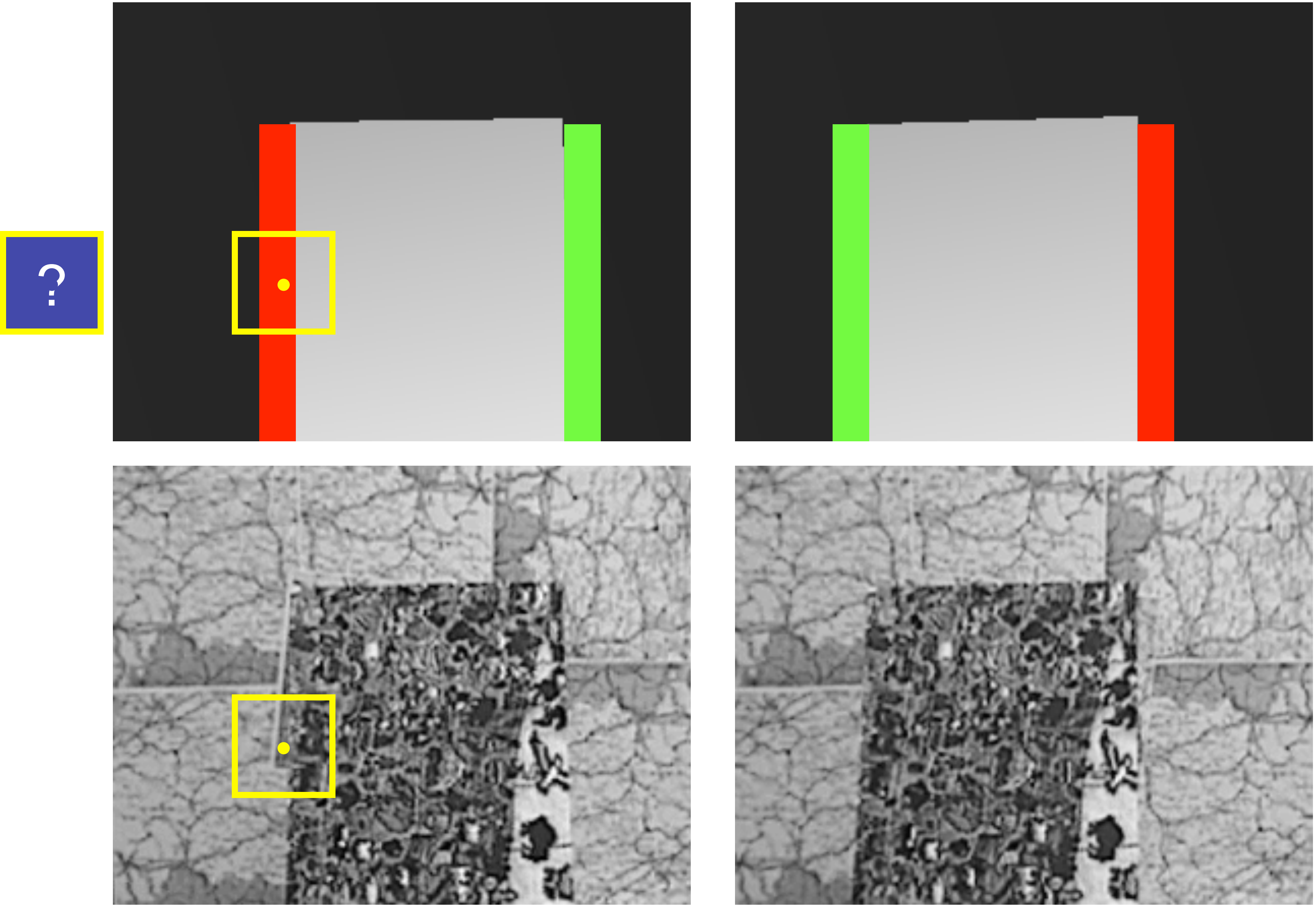
Stefano Mattoccia

Case 2: near half occlusion vs inside discontinuity

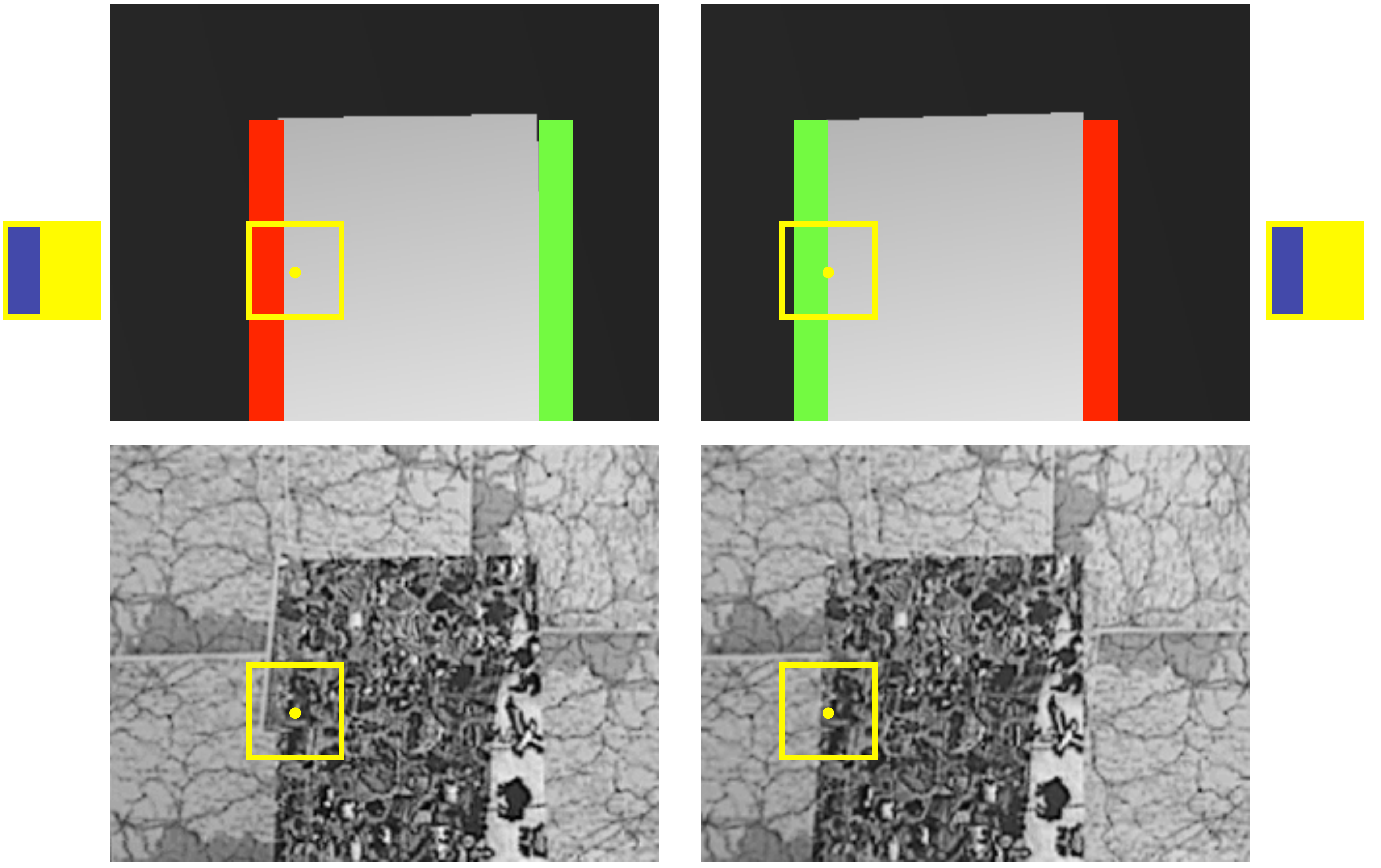


Stefano Mattoccia

Case 3: inside half occlusion vs any -> depth = occlusion !!

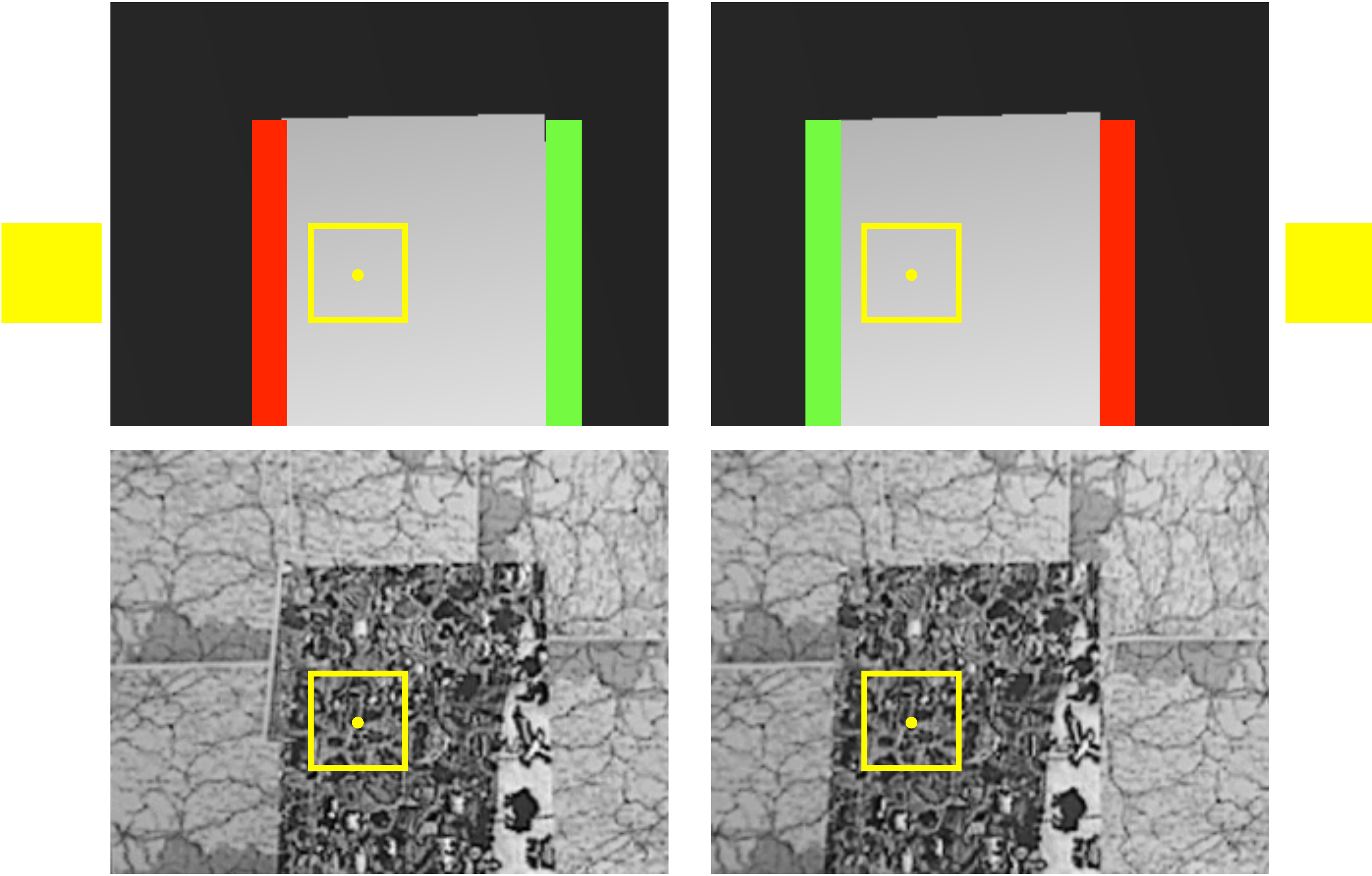


Case 4: near half occlusion vs near discontinuity



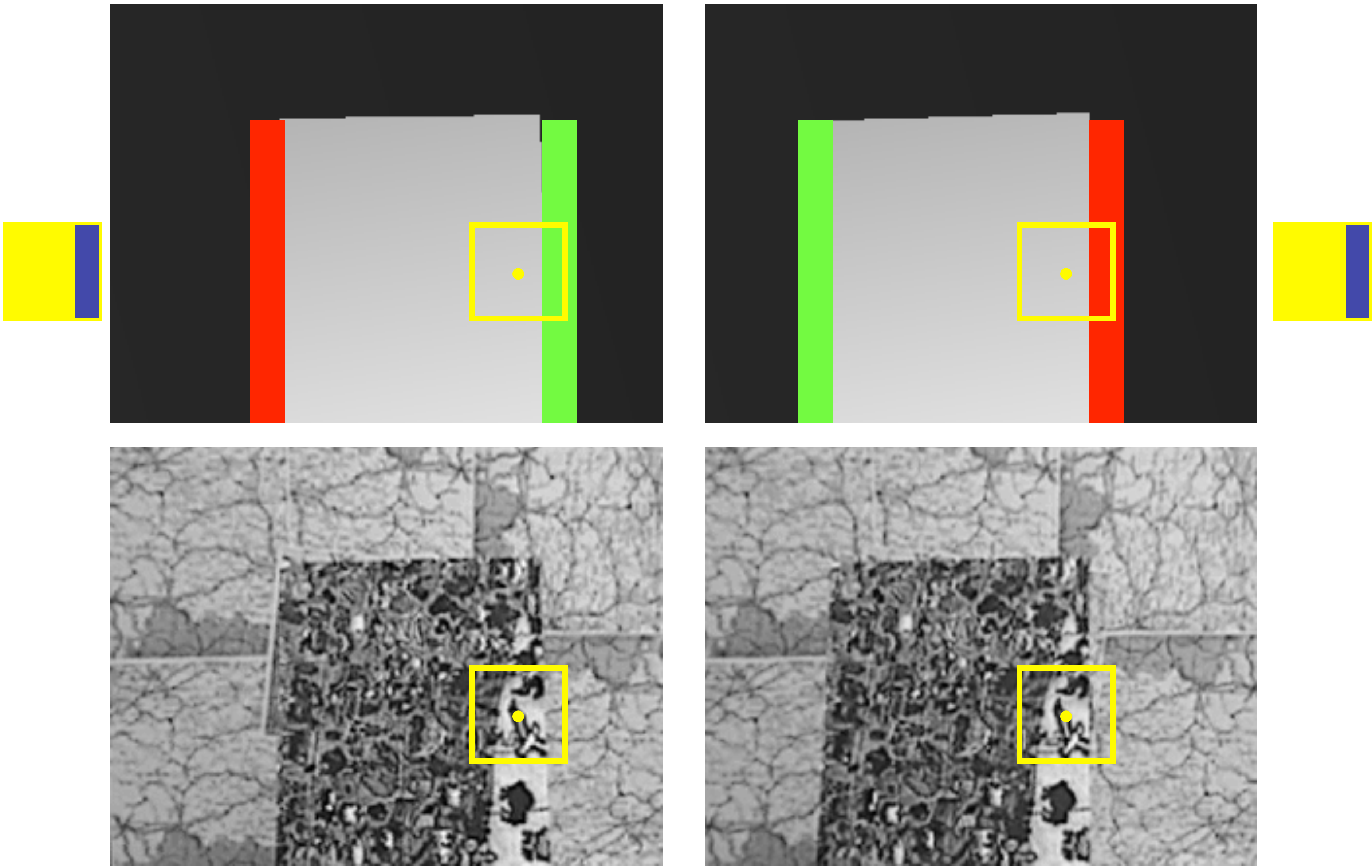
Stefano Mattoccia

Case 5: no half occlusion, no discontinuity



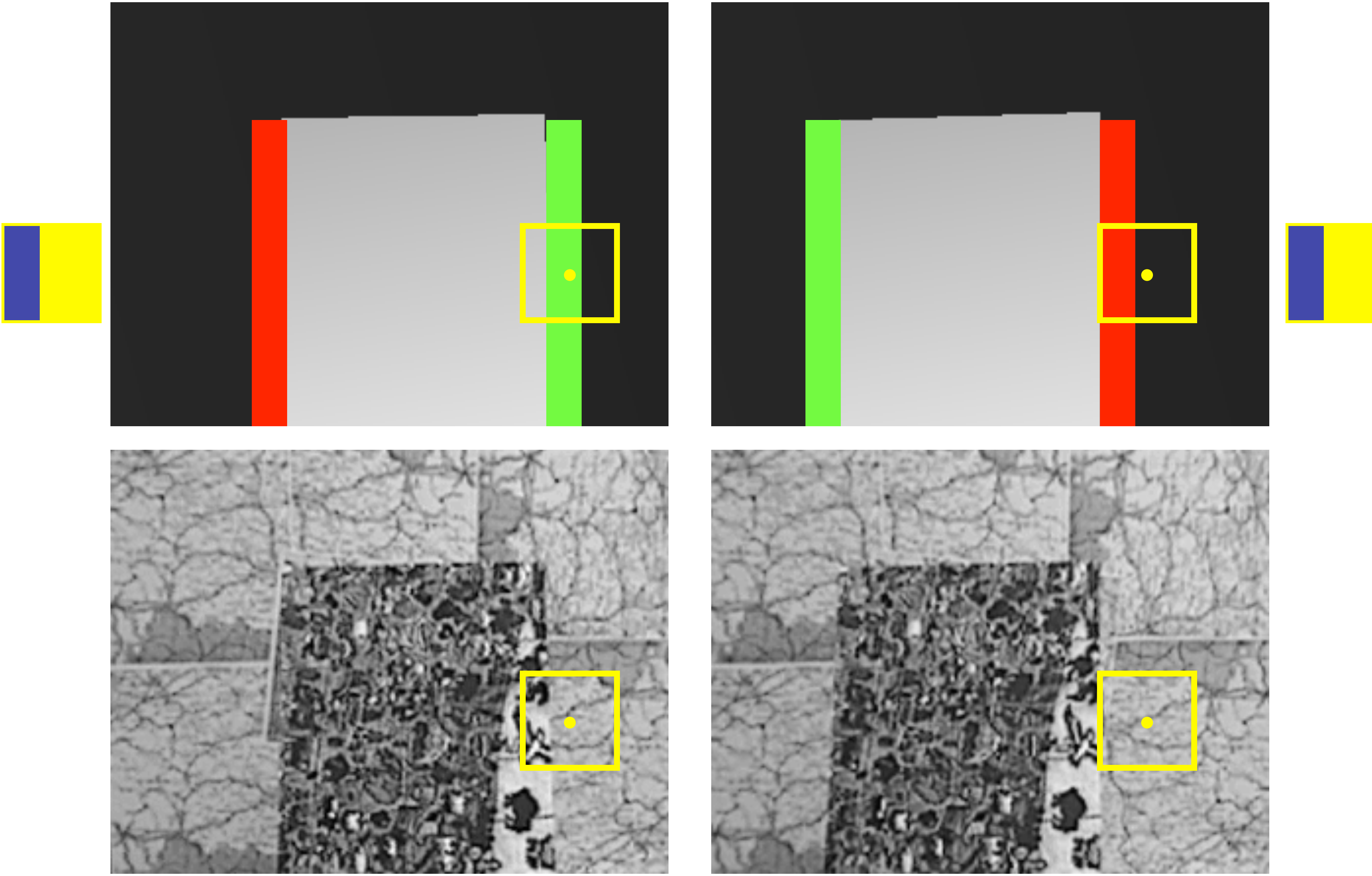
Stefano Mattoccia

Case 6: near discontinuity, near occlusion



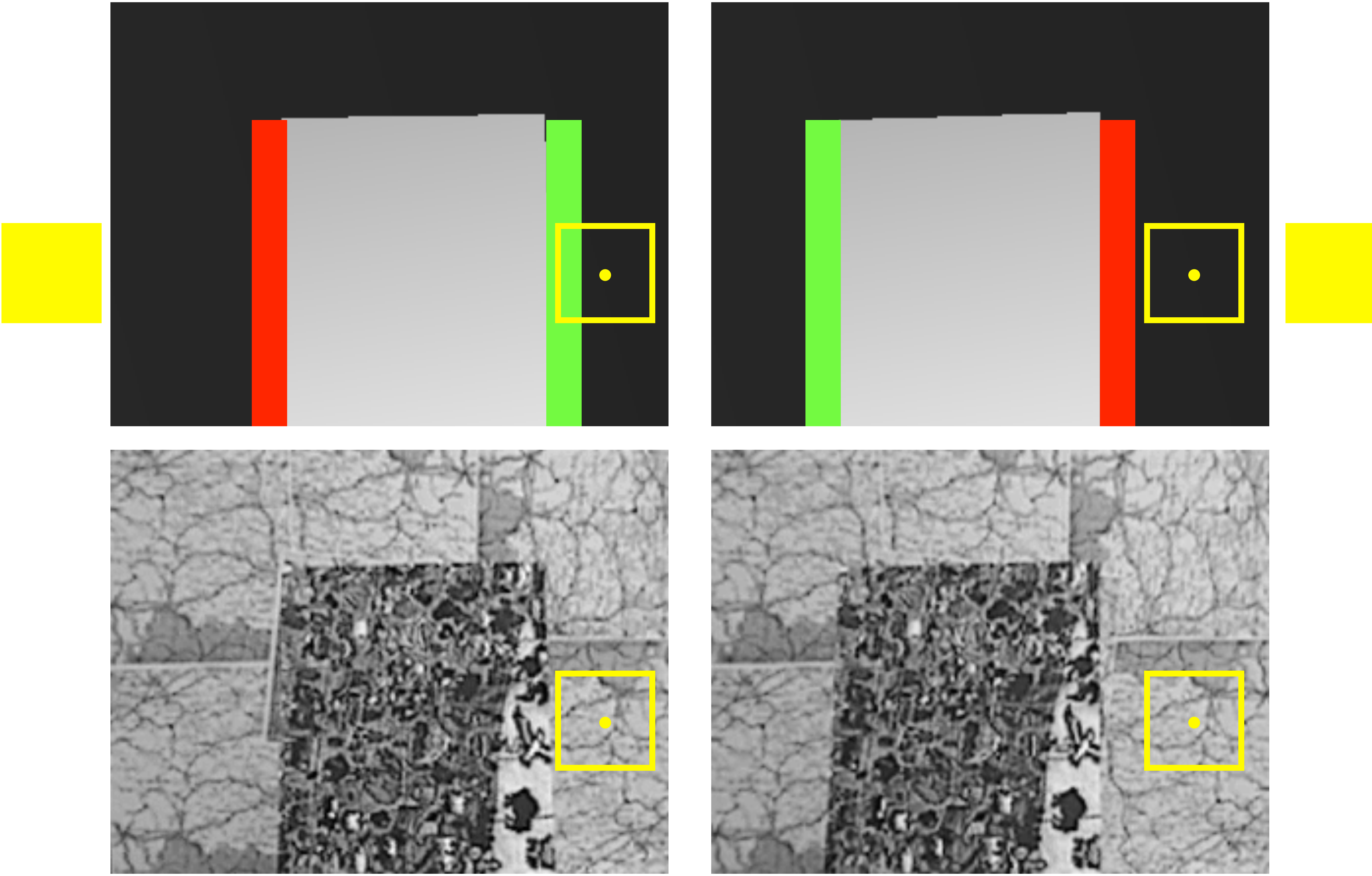
Stefano Mattoccia

Case 7: inside discontinuity, near occlusion

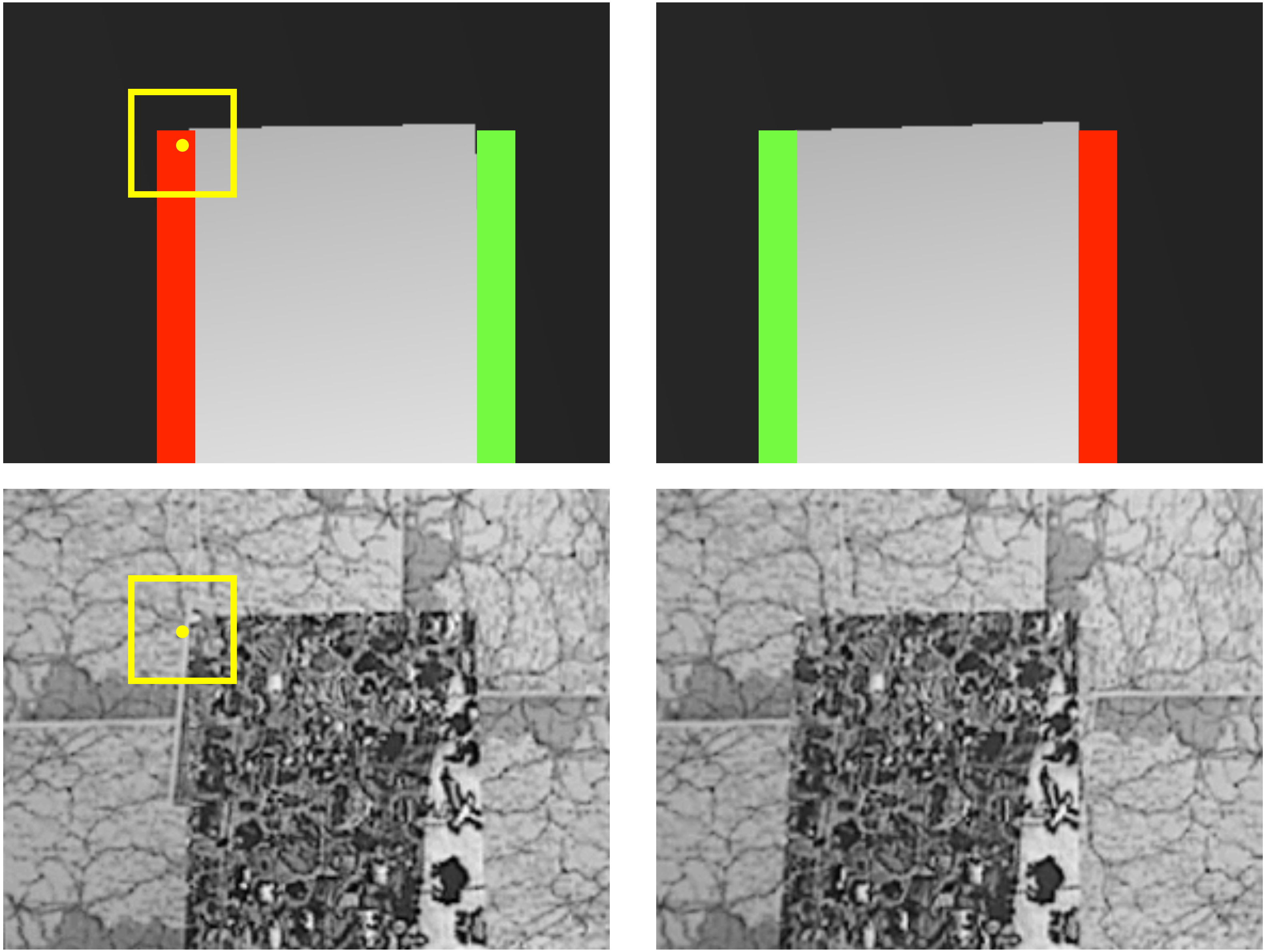


Stefano Mattoccia

Case 8: near discontinuity, no occlusion no discontinuity

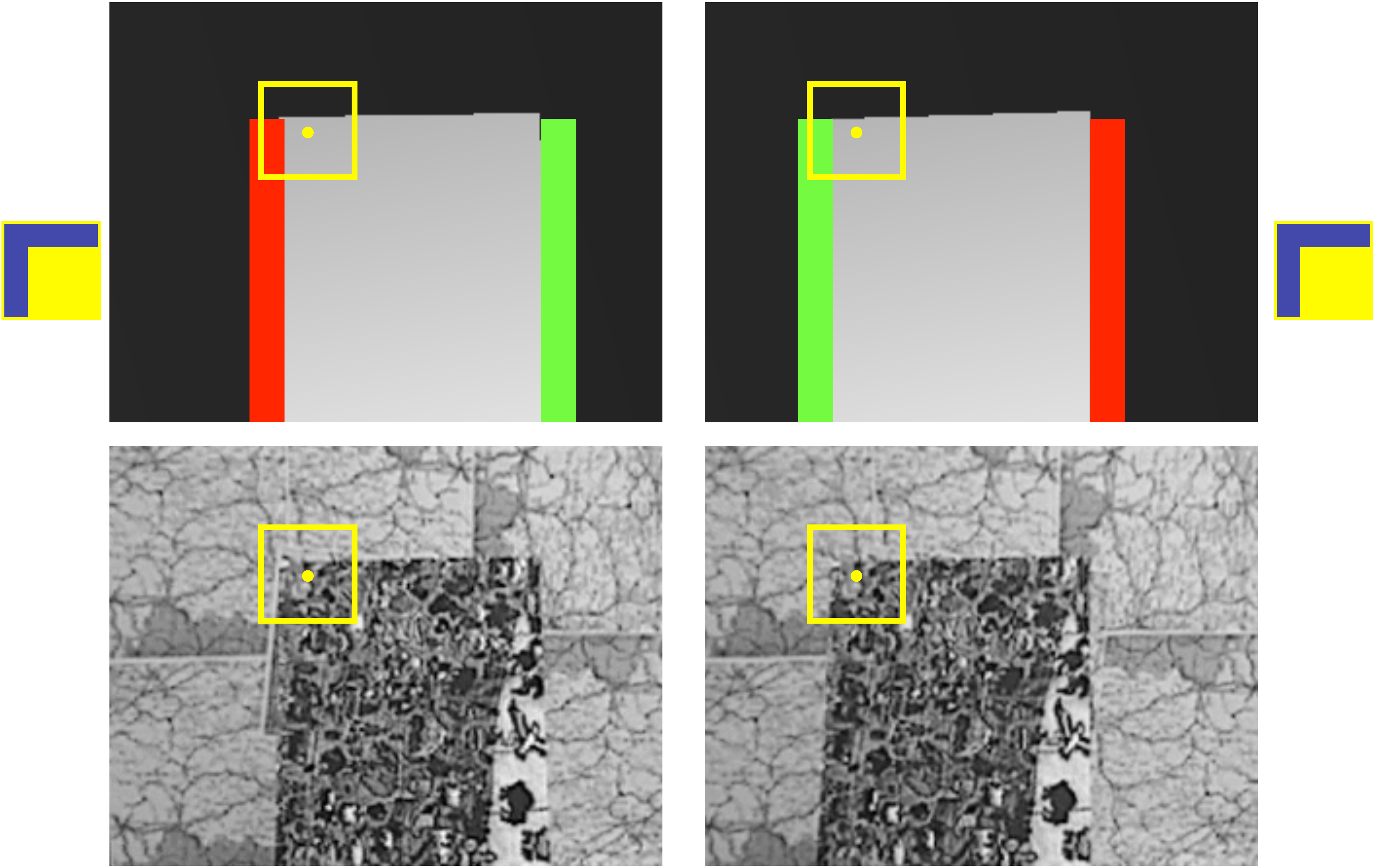


Case 9: inside occlusion vs any -> depth = occlusion !!



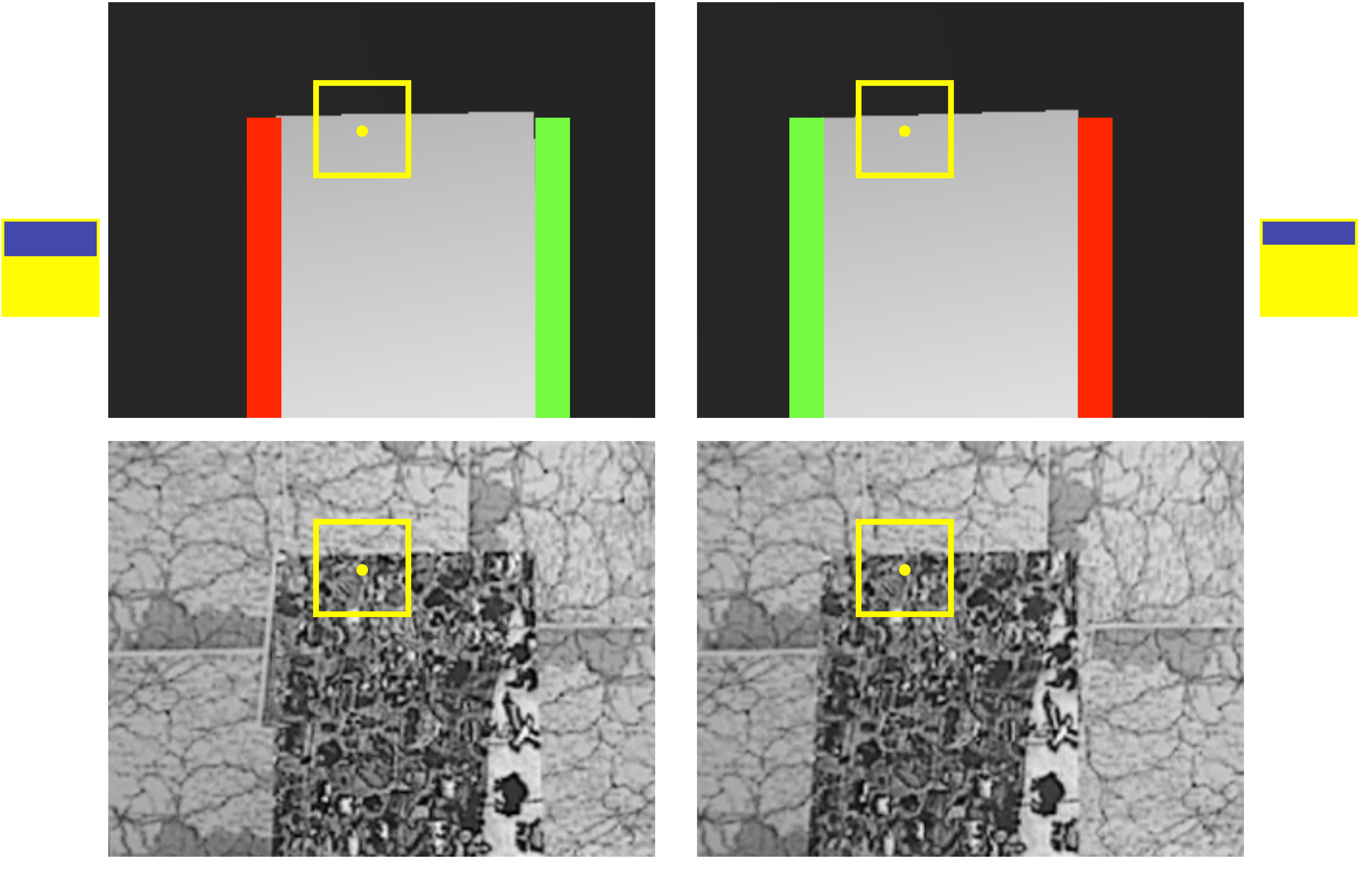
Stefano Mattoccia

Case 10: near occlusion and discontinuity vs near discontinuity



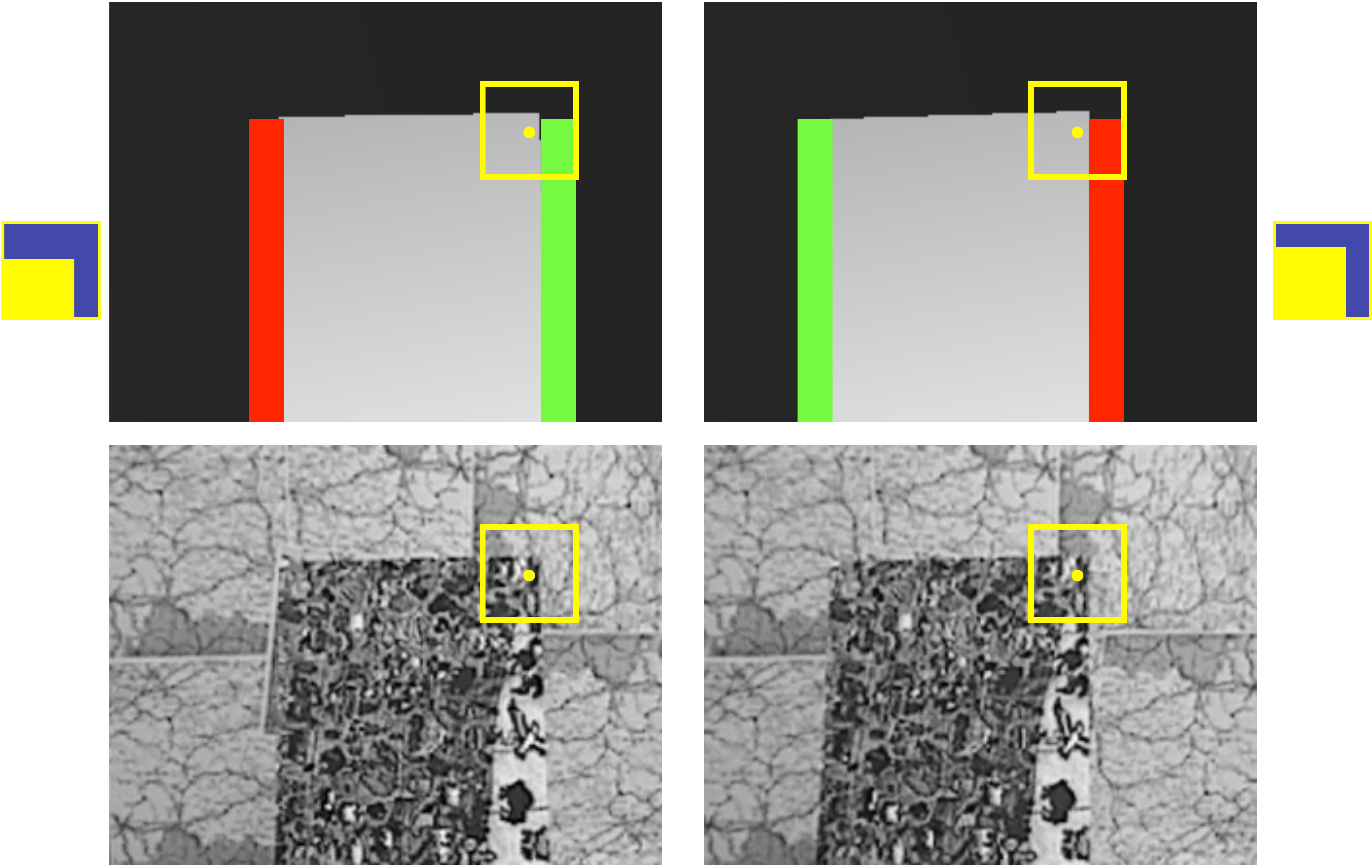
Stefano Mattoccia

Case 11: near discontinuity vs near discontinuity



Stefano Mattoccia

Case 12: near discontinuity vs near discontinuity and occlusion

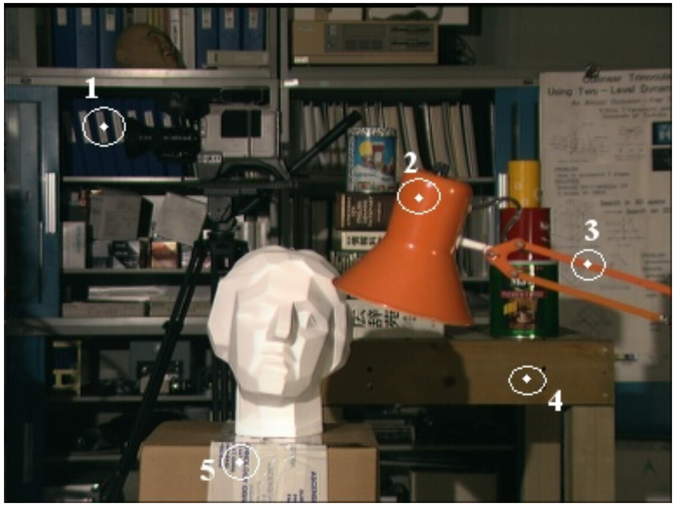


Stefano Mattoccia

Classification and evaluation of cost aggregation strategies for stereo correspondence

- In [1] we classified, implemented and evaluated (accuracy and execution time) 10+ state-of-the-art cost aggregation strategies

- Since the focus is on the cost aggregation strategy the evaluation methodology includes only DISC and NON_OCC

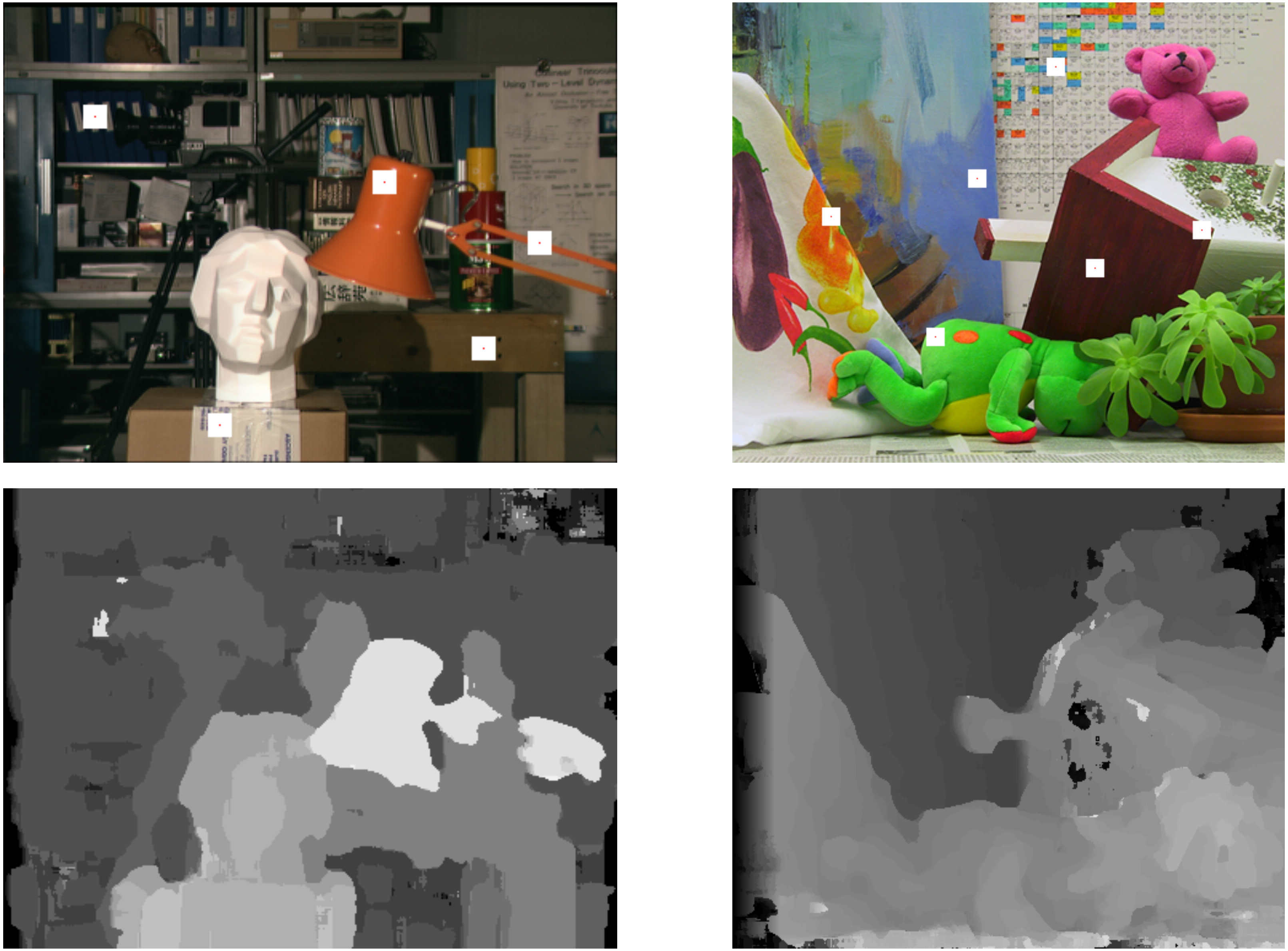


F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Classification and evaluation of cost aggregation methods for stereo correspondence, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2008)

Accompanying web site and software: www.vision.deis.unibo.it/spe/SPEHome.asp

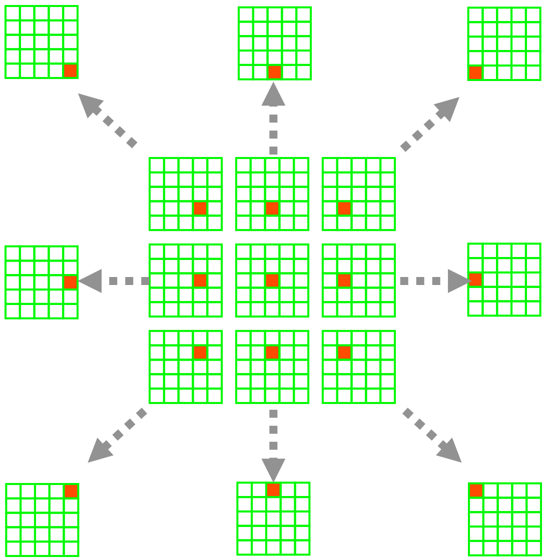
- **Analyzed a subset of relevant state-of-the-art cost aggregation strategies**
  - **position**
  - **shape**
  - **position and shape**
  - **weights**
- **Most of these techniques compute the support using a symmetric strategy**
- **Benchmarking platform: Intel Core Duo 2.14 GHz CPU**
- **Execution time: Teddy stereo pair (size 450x373) with and a disparity search range of 60.**
- **Optimizations: the same proposed by authors*, no SIMD, no multicores, etc**
- **The next slides describe most of these methods and some novel approaches not included in the paper (i.e. Fast Aggregation [64], Fast Bilateral Stereo (FBS) [65] and the Locally Consistent (LC) methodology [66])**

Fixed Window: results

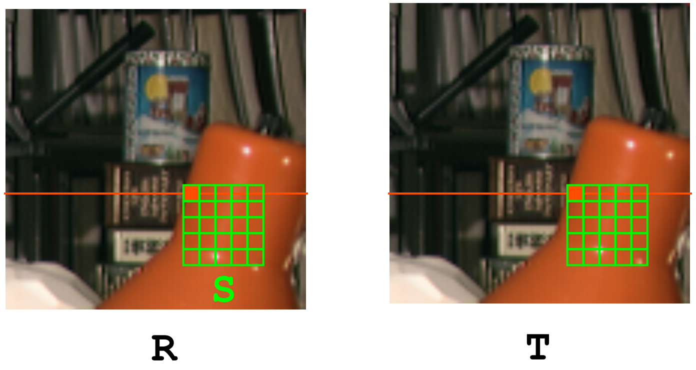


Shiftable Windows [11]

- This approach aims at reducing the border localization problem of FW not constraining the support to be centered on the central position
- Support is symmetric
- Execution time: 12 sec

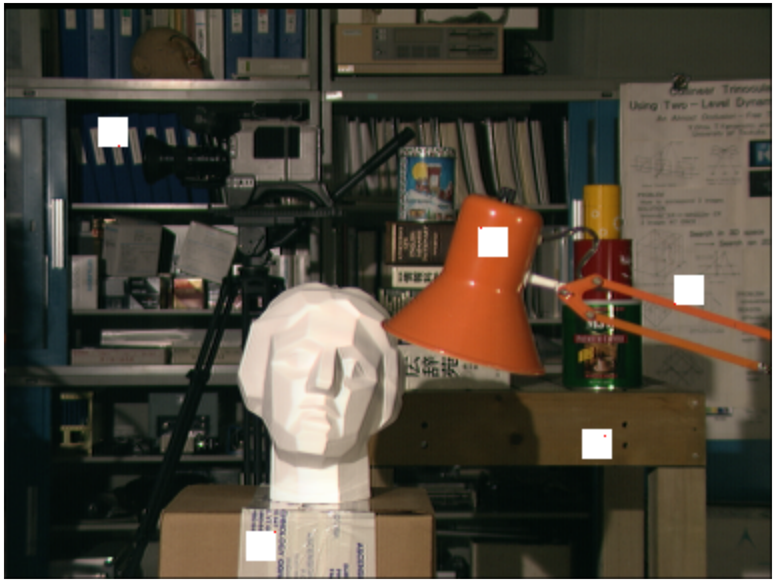


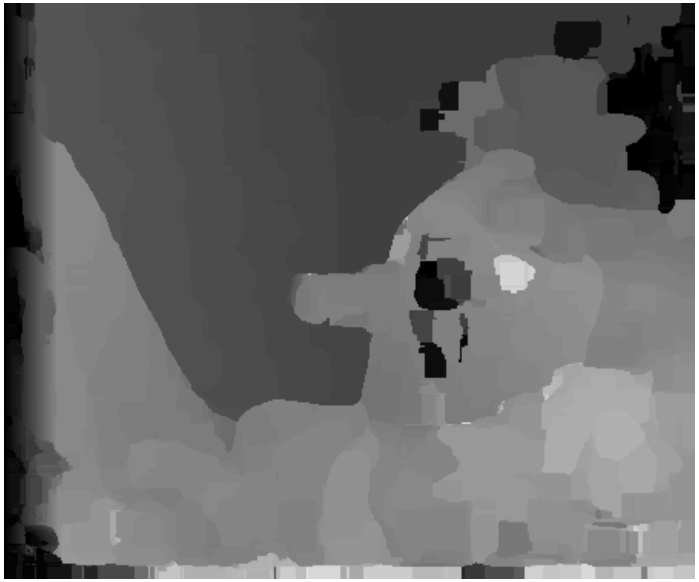
The position with the best score is selected



D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms Int. Jour. Computer Vision, 47(1/2/3):7–42, 2002

Shiftable Windows: results



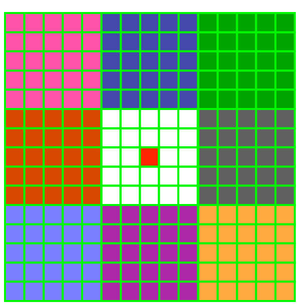


Stefano Mattoccia

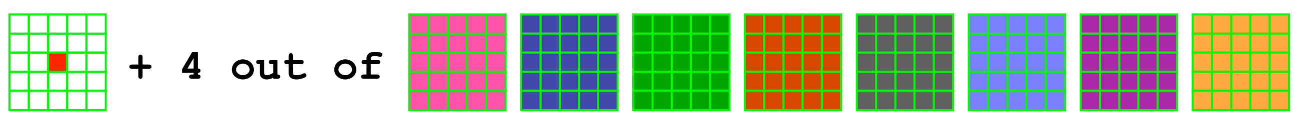
Multiple Windows [7]

- **The number of elements in the support is constant**
- **The shape of the support is not constrained to be rectangular**
- **Support is symmetric**
- **Proposed for 5, 9 and 25 sub-windows (5W, 9W and 25W)**
- **Execution time (9W): 11 sec (*)**

With 9 sub-windows (9W):

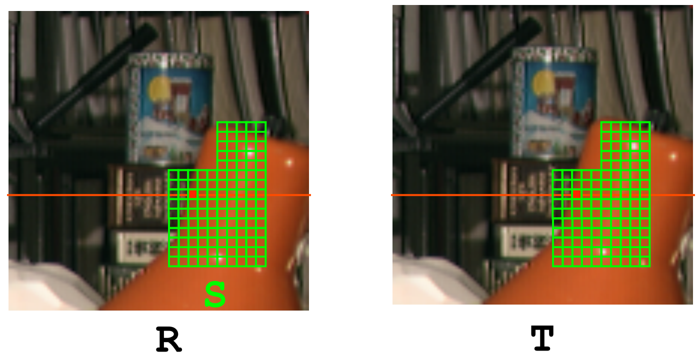


S

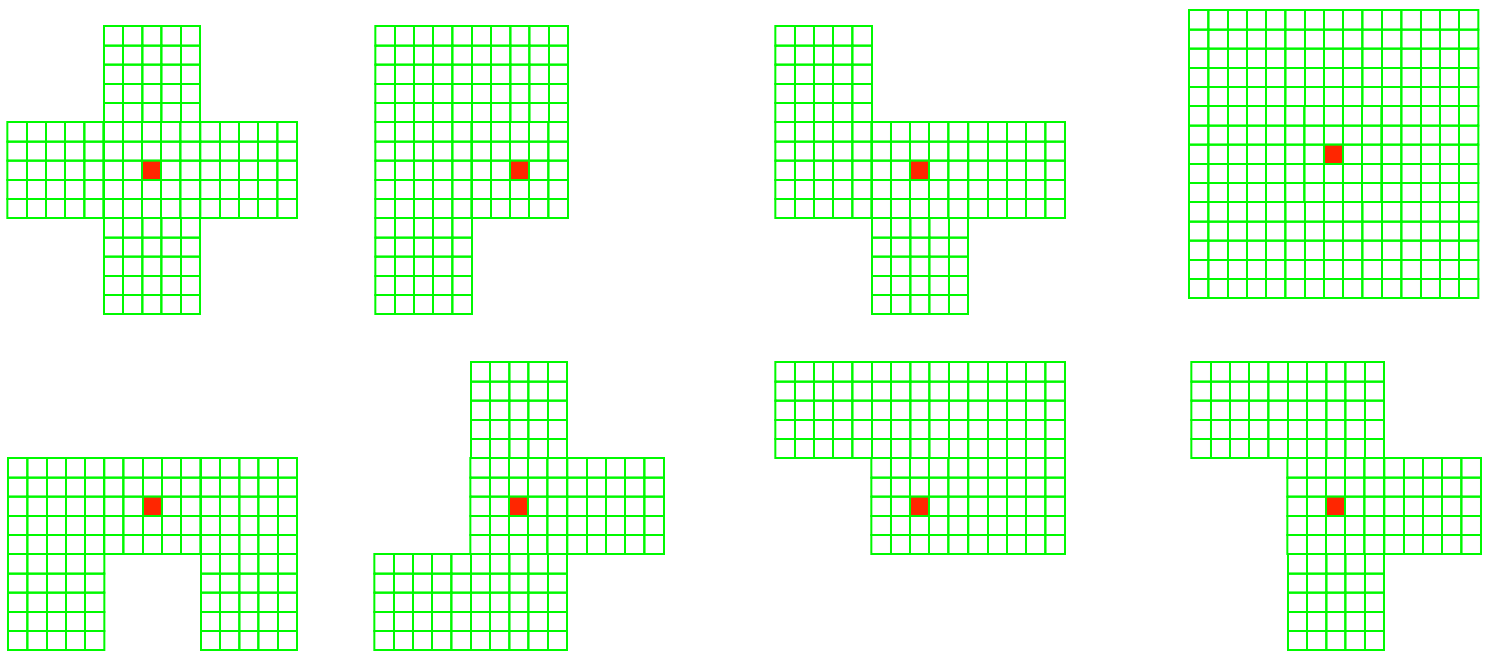


according to the matching cost computed over the single sub-windows

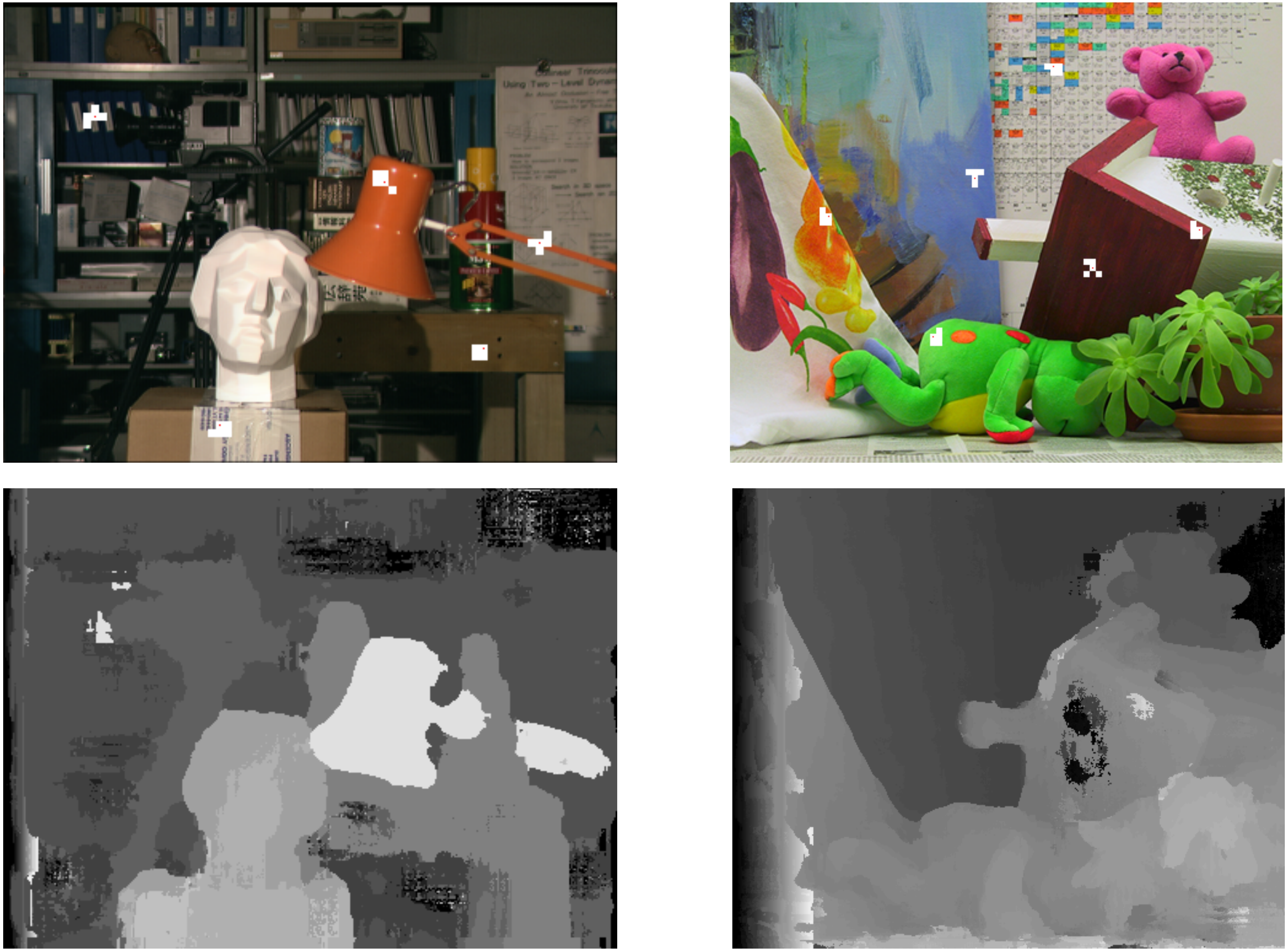
H. Hirschmuller, P. Innocent, and J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors
Int. Journ. of Computer Vision, 47:1–3, 2002



Support: some shapes (with 9 sub-windows)



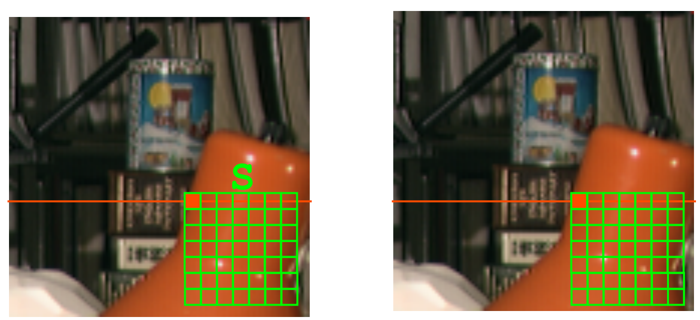
Multiple (9) Windows: results



Stefano Mattoccia

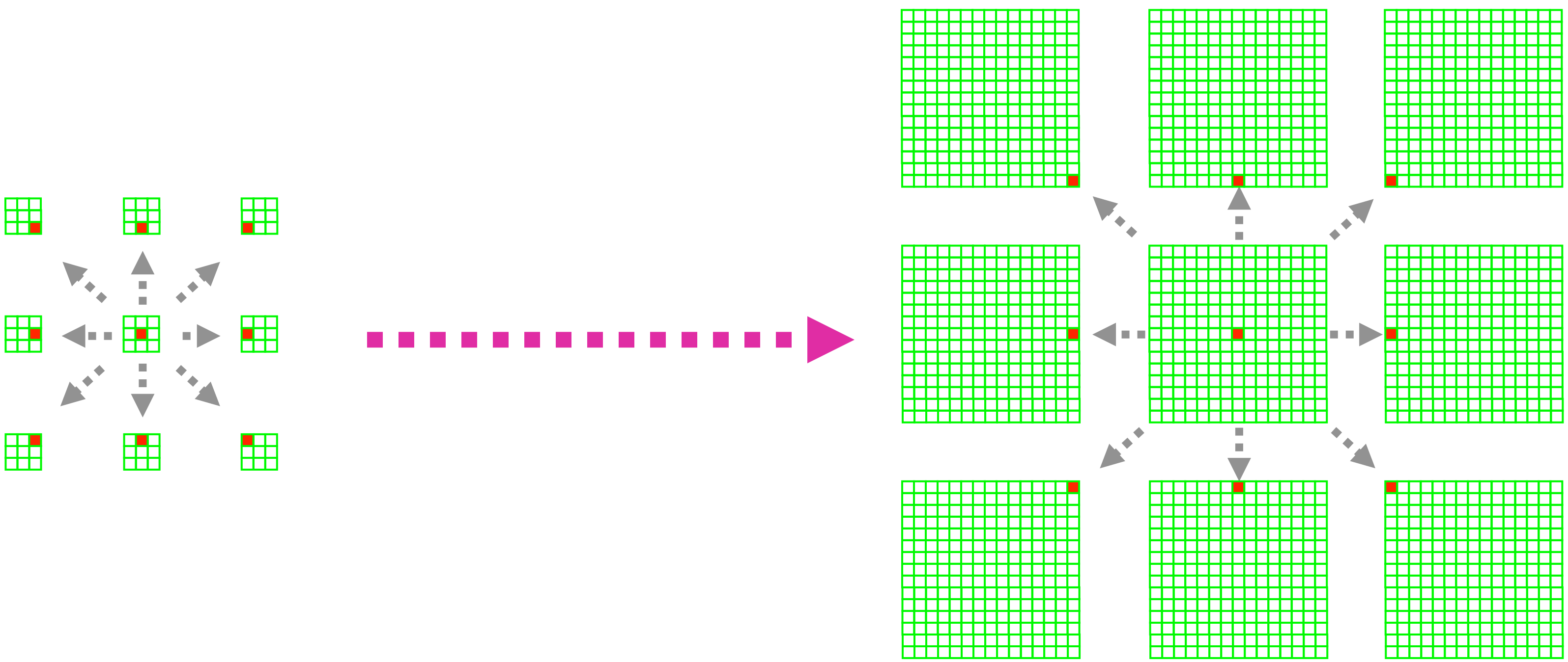
Variable Windows [12]

- **Pixel-based cost function: Birchfield and Tomasi**
- **Size of the support varies while shape is constrained (square)**
- **Position of the support changes (shiftable windows)**
- **Support is symmetric**
- **Efficient search based on a DP technique**
- **Execution time: 16 sec (good trade-off speed/accuracy)**



R **T**

O. Veksler, Fast variable window for stereo correspondence using integral images
In Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2003), pages 556–561, 2003



$$\bar{e} = \frac{\sum_{i,j \in S} e_d(i,j)}{|S|}$$

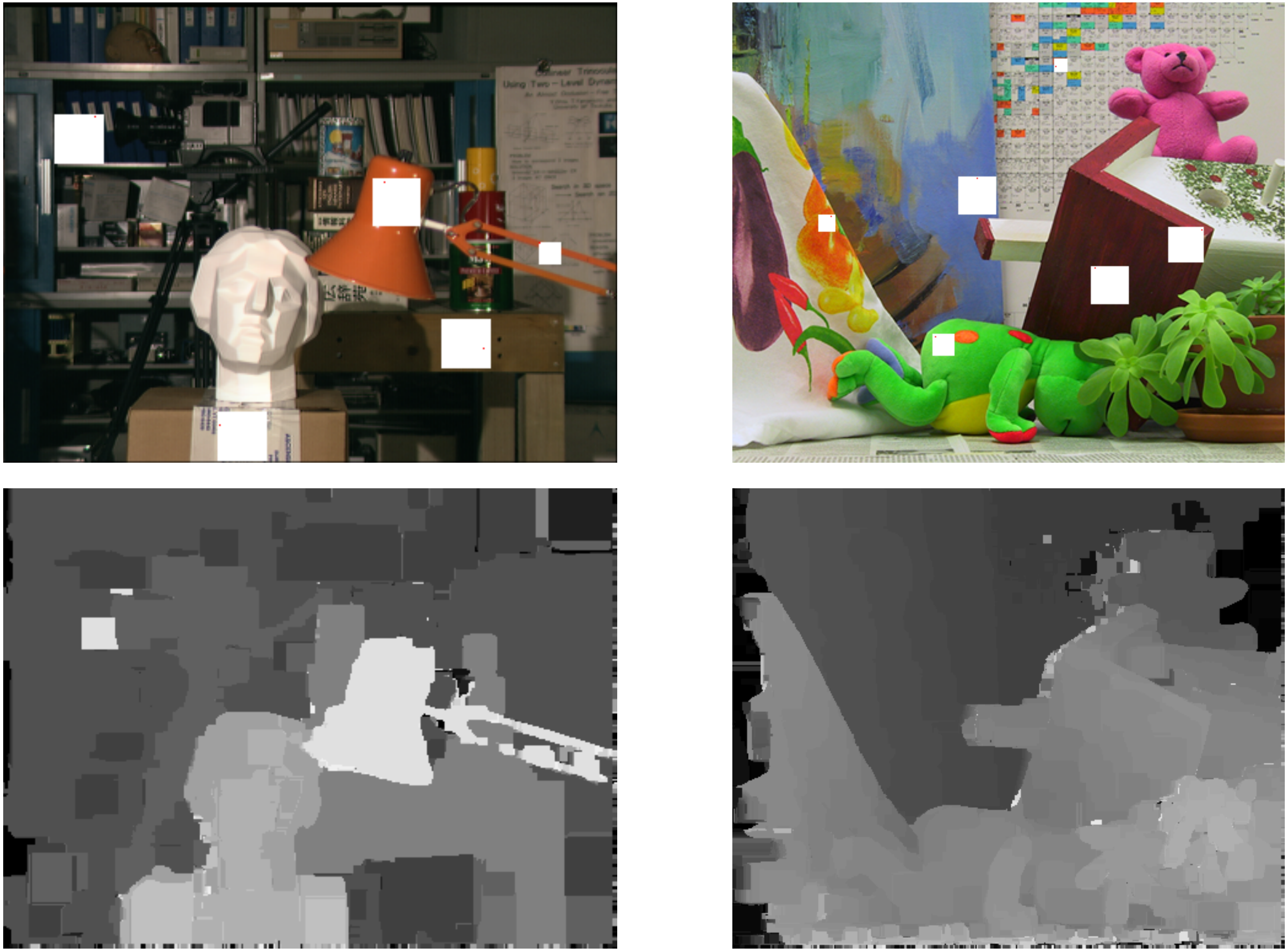
cardinality of the support

$$C_d(S) = \bar{e} + \alpha \cdot \mathrm{var}(\bar{e}) + \frac{\beta}{\sqrt{|S| + \gamma}}$$

This term favors large windows in uniform areas (where $\bar{e} + \alpha \cdot \mathrm{var}(\bar{e})$ is small)

α, β, γ : parameters of the algorithm

Variable Windows: results



Segmentation

Original

Segmented [50]

- Partitioning of the image in regions made of connected pixels with *similar* colors intensity

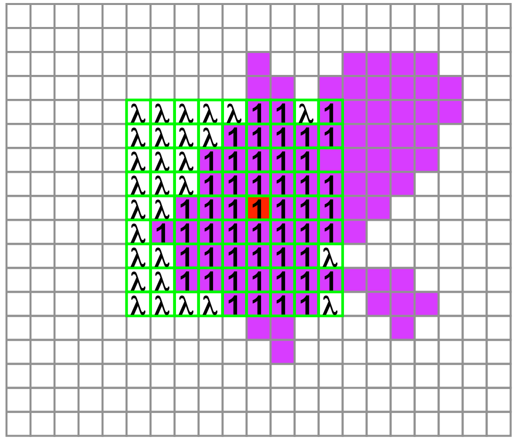
- Useful in stereo for cost aggregation, disparity refinement, outliers detection, etc

D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis
IEEE Transactions on Pattern Analysis and Machine Intelligence, 24:603–619, 2002

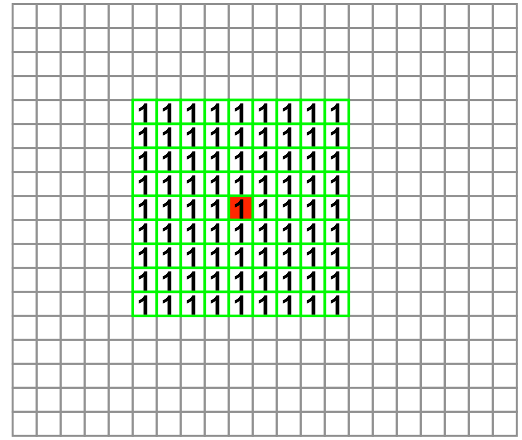
Segmentation based [5]

- Assumption: depth within each segment varies smoothly
- Segmentation of reference image (Not Symmetrical)
- Shape and size unconstrained (within max support)
- Pixel-based cost function: M-estimator
- Requires explicit segmentation
- Each cost is weighted 1 (same segment) or $\lambda \ll 1$ (different segment)
- Execution time: 2 sec (fast)

M. Gerrits and P. Bekaert, Local Stereo Matching with Segmentation-based Outlier Rejection
Proc. Canadian Conf. on Computer and Robot Vision, 2006



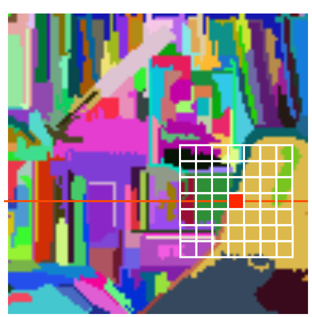
R



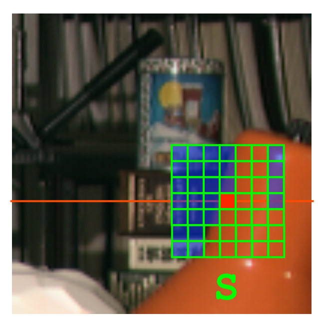
T

For each point within the maximum allowed support:

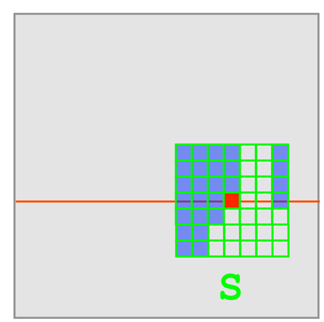
- points within the same segment of the central point (reference image) assume weight 1
- points outside are weighted λ<<1



R(Seg)

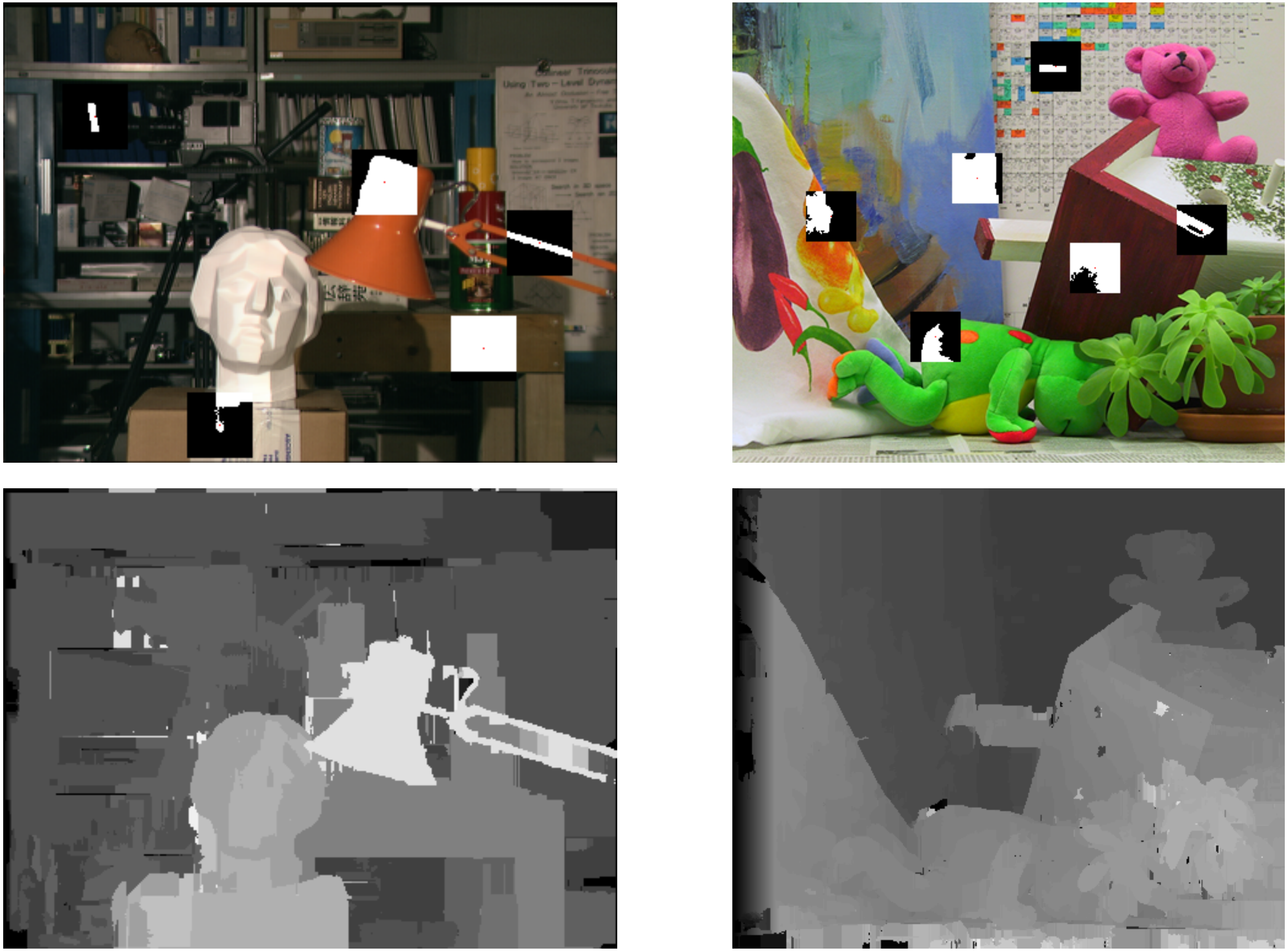


R



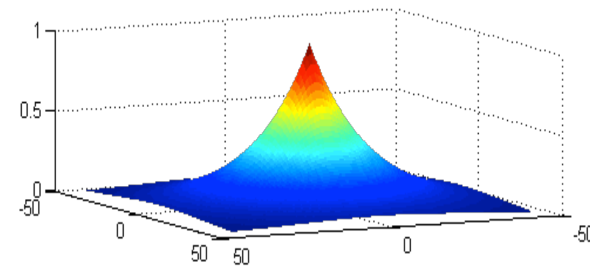
R

Segmentation based: results



Bilateral Filtering [51]

- **Edge preserving smoothing technique**
- **In the sum each element is weighted according to its spatial and color proximity (wrt the central point)**
- **Implicitly deploys segmentation**



Weight function

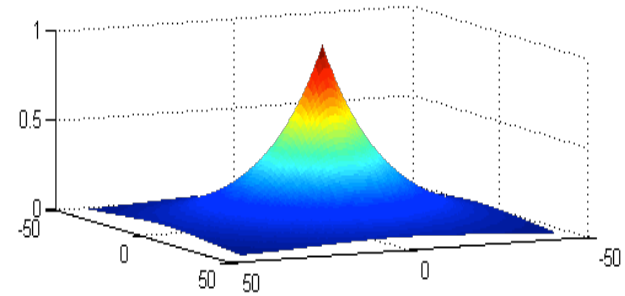
Original image

Conventional smoothing

Bilateral Filtering

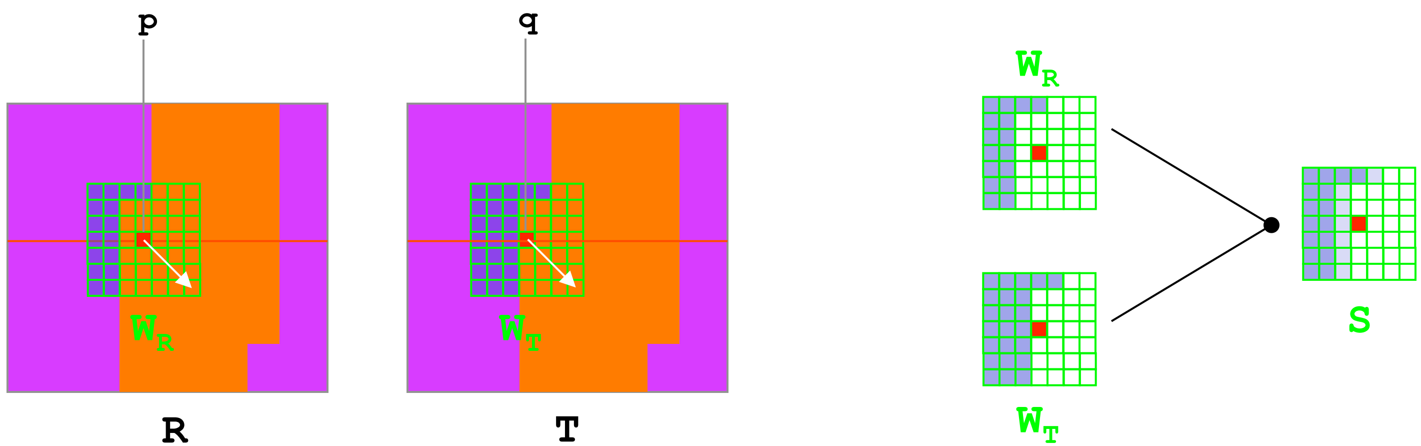
C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV98*, pages 839–846, 1998

Adaptive Weights [14]



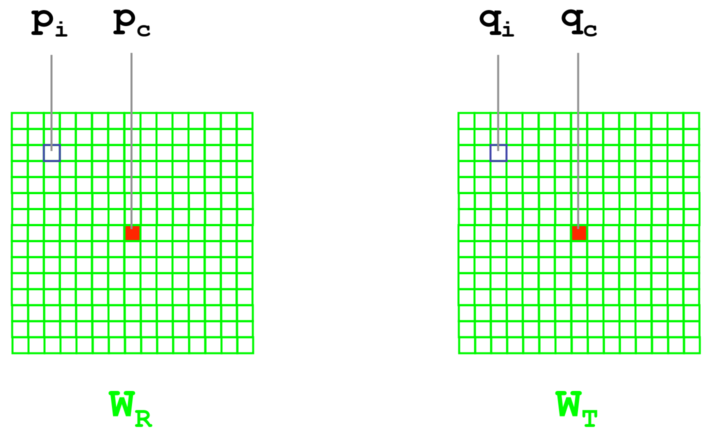
- Costs are symmetrically weighted by spatial and color proximity
- Implicitly deploys segmentation

- Pixel-based cost function: TAD
- Symmetric support
- Execution time: 17 minutes (very slow)



Simplified example (using only color proximity)

K. Yoon and I. Kweon. Adaptive support-weight approach for correspondence search IEEE PAMI, 28(4):650–656, 2006

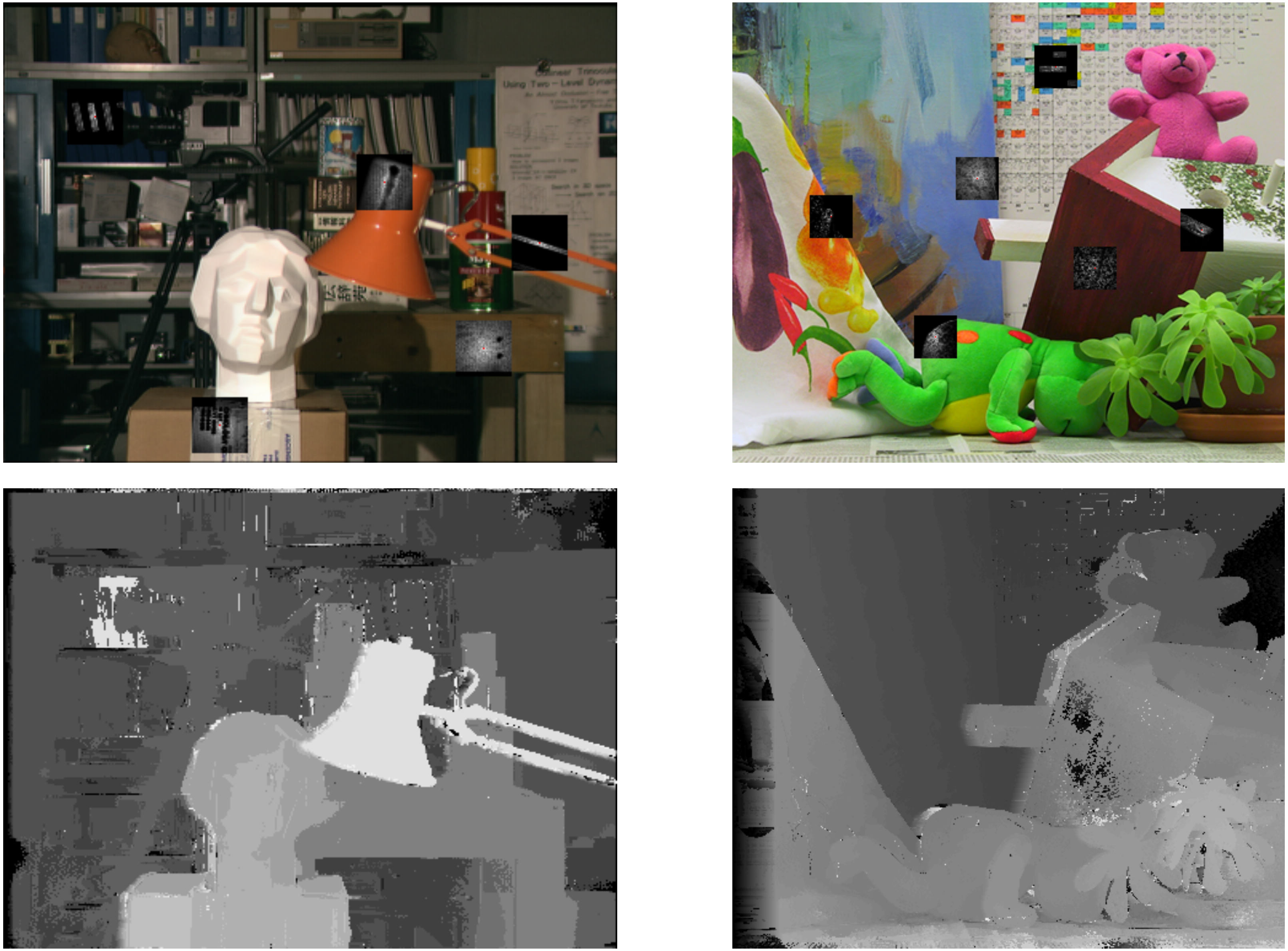


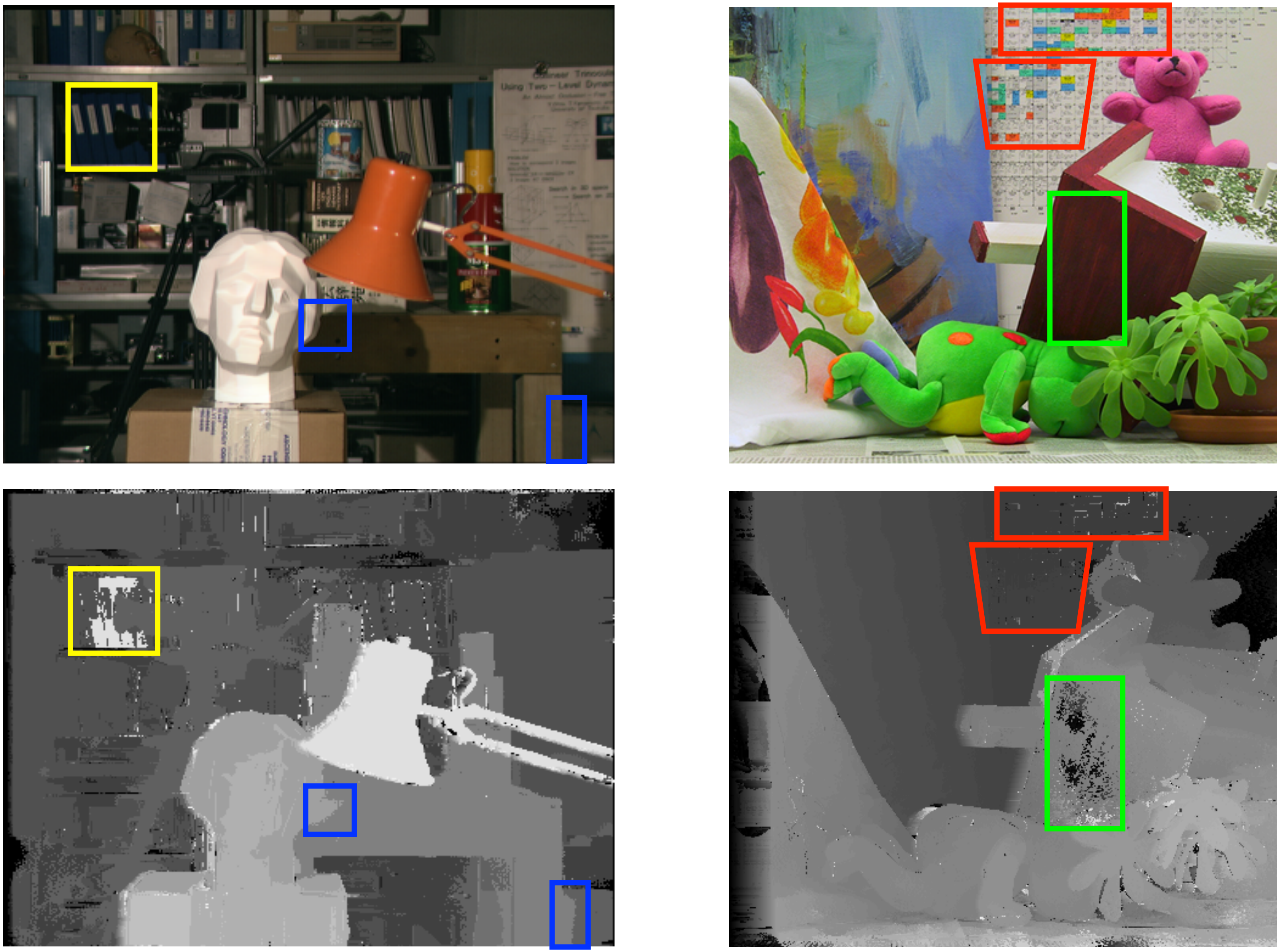
$$C\left(p_c, q_c\right) = \frac{\displaystyle\sum_{p_i \in WT, q_i \in W_T} w_r\left(p_i, p_c\right) \cdot w_t\left(q_i, q_c\right) \cdot TAD\left(p_i, q_i\right)}{\displaystyle\sum_{p_i \in W_R, q_i \in W_T} w_r\left(p_i, p_c\right) \cdot w_t\left(q_i, q_c\right)}$$

$$w_R\left(p_i, p_c\right) = e^{-\frac{d_p\left(p_i, p_c\right)}{\gamma_p}} e^{-\frac{d_c\left(I_R(p_i), I_R(p_c)\right)}{\gamma_c}}$$

$$w_T\left(q_i, q_c\right) = e^{-\frac{d_p\left(q_i, q_c\right)}{\gamma_p}} e^{-\frac{d_c\left(I_R(q_i), I_R(q_c)\right)}{\gamma_c}}$$

Adaptive Weights: results





Stefano Mattoccia

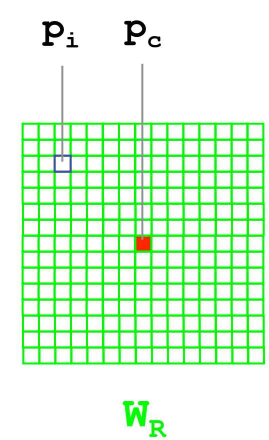
Segment Support [10]

- Segments both images
- Discard the spatial proximity assumption: weights rely only on segmentation and color proximity
- Cost function: TAD
- Symmetric support
- Execution time: 30 minutes (very slow)

Weights for reference (and target) image are assigned according to:

$$w_R^{'}(p_i, p_c) = \begin{cases} 1.0 & for \quad p_i \in S_c \\ e^{-\frac{d_c(I_R(p_i), I_R(p_c))}{\gamma_c}} & , otherwise \end{cases}$$

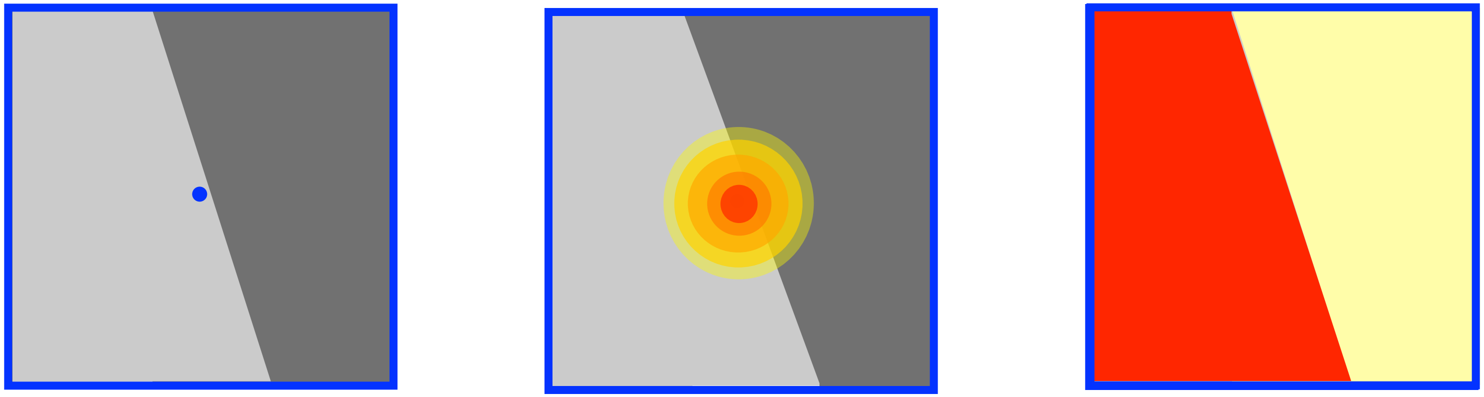
Sc segment that includes the central point



and then combined (symmetric support)

F. Tombari, S. Mattoccia, L. Di Stefano, Segmentation-based adaptive support for accurate stereo correspondence IEEE Pacific-Rim Symposium on Image and Video Technology (PSIVT 2007)

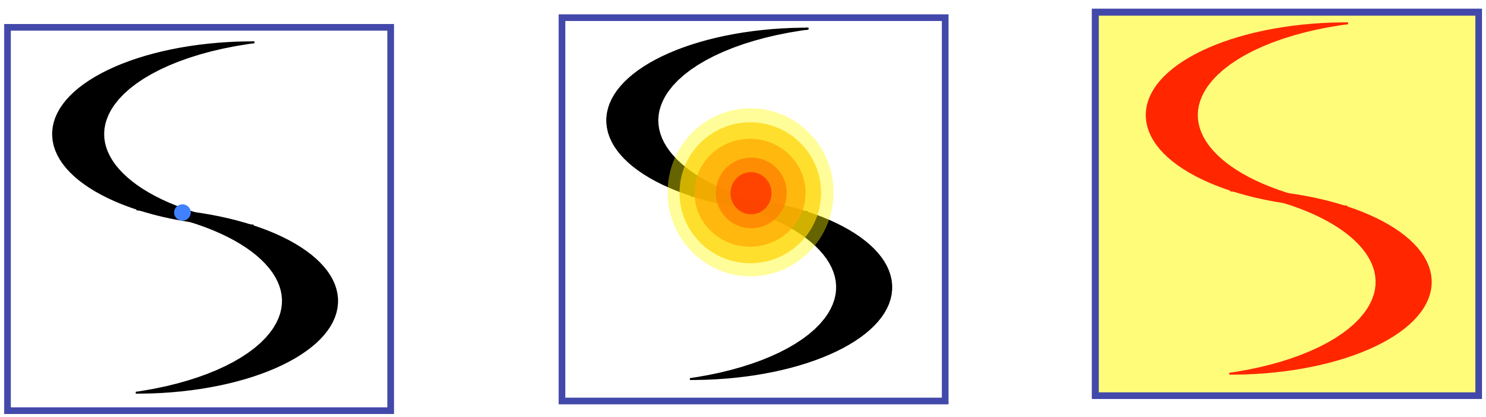
Depth borders



Adaptive weights

Ideal vs Segment Support

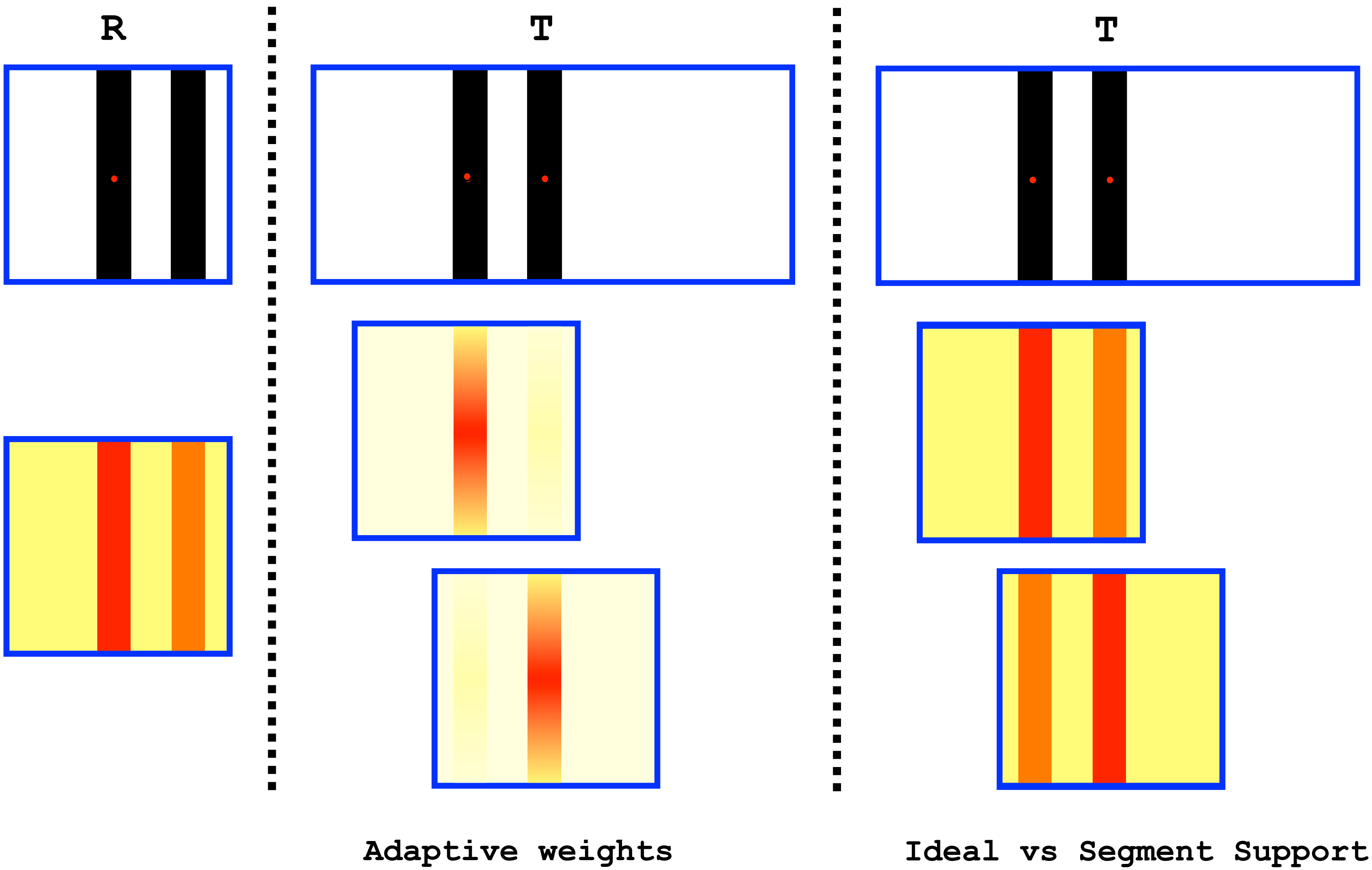
Planar regions



Adaptive weights

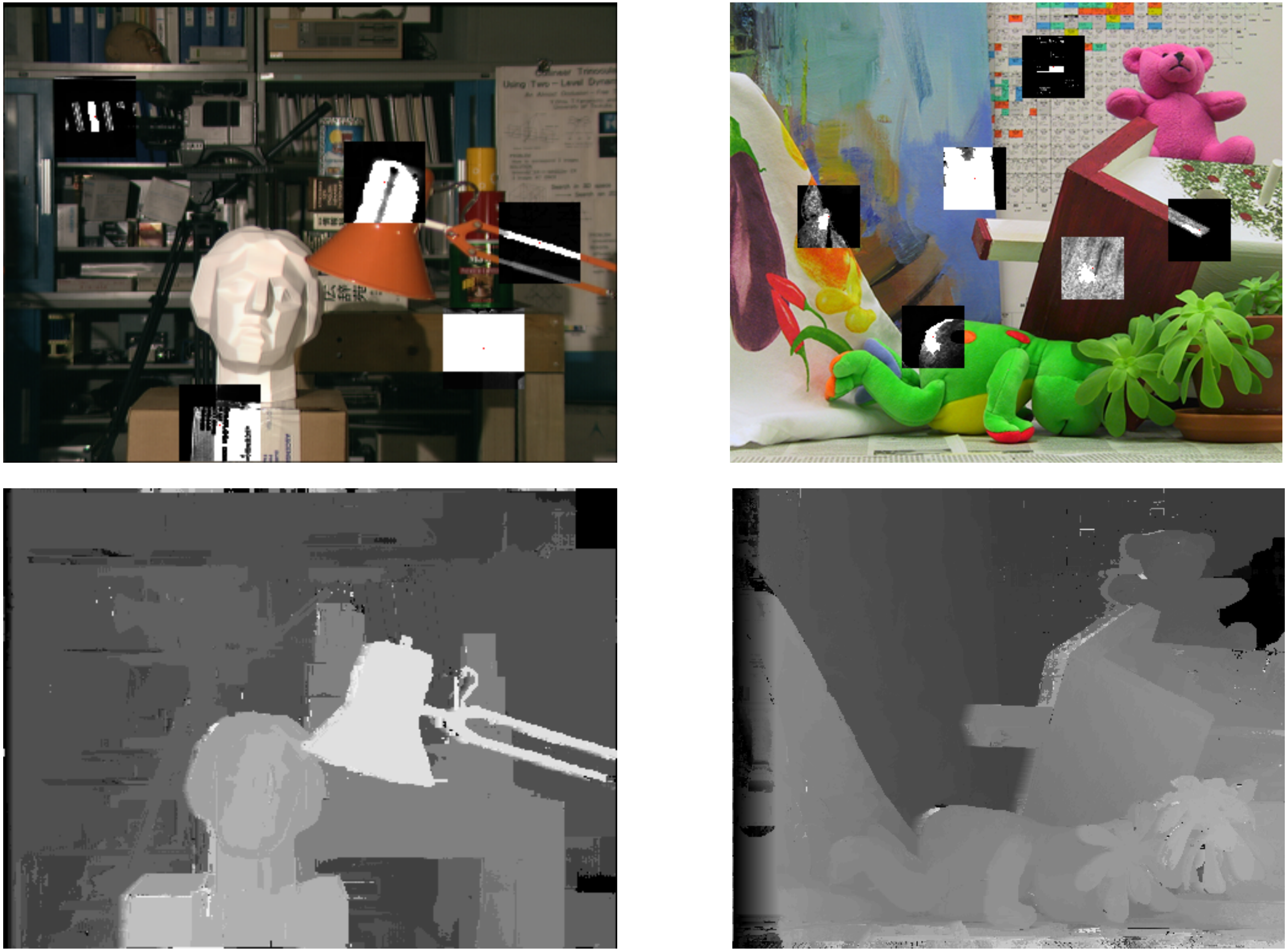
Ideal vs Segment Support

Repetitive patterns



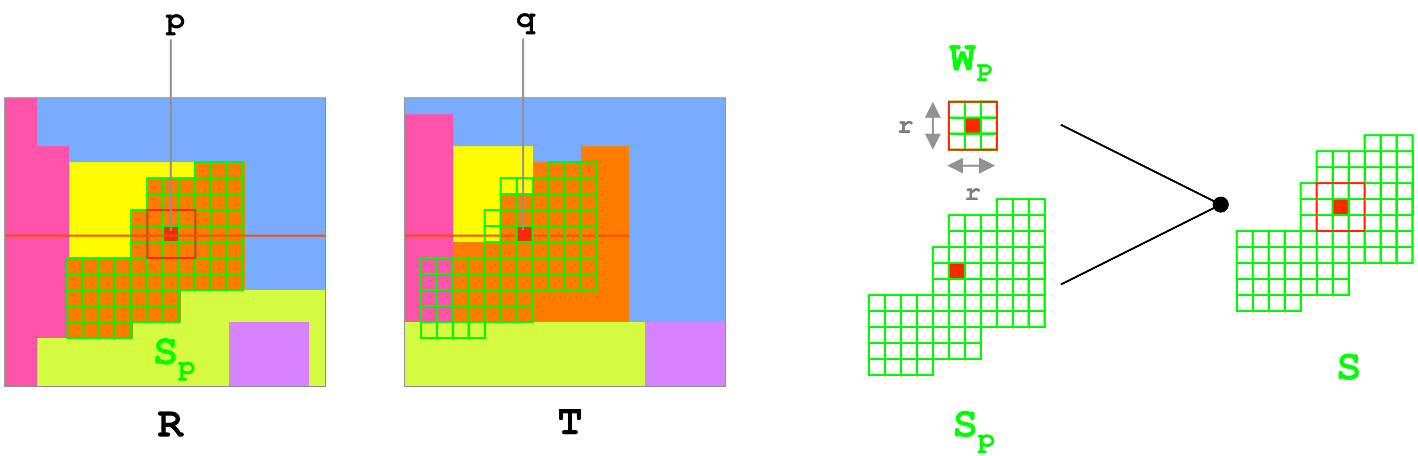
Stefano Mattoccia

Segment Supports: results



Fast Aggregation [64]

- Assumption: depth within each segment varies smoothly
- Cost function: TAD
- Segments only the reference image R
- Asymmetric support (reference image)
- Support extends to the entire segment (R)
- Fast: 0.6 sec (segmentation accounts for 40%-80%)



F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Near real-time stereo based on effective cost aggregation
International Conference on Pattern Recognition (ICPR 2008)

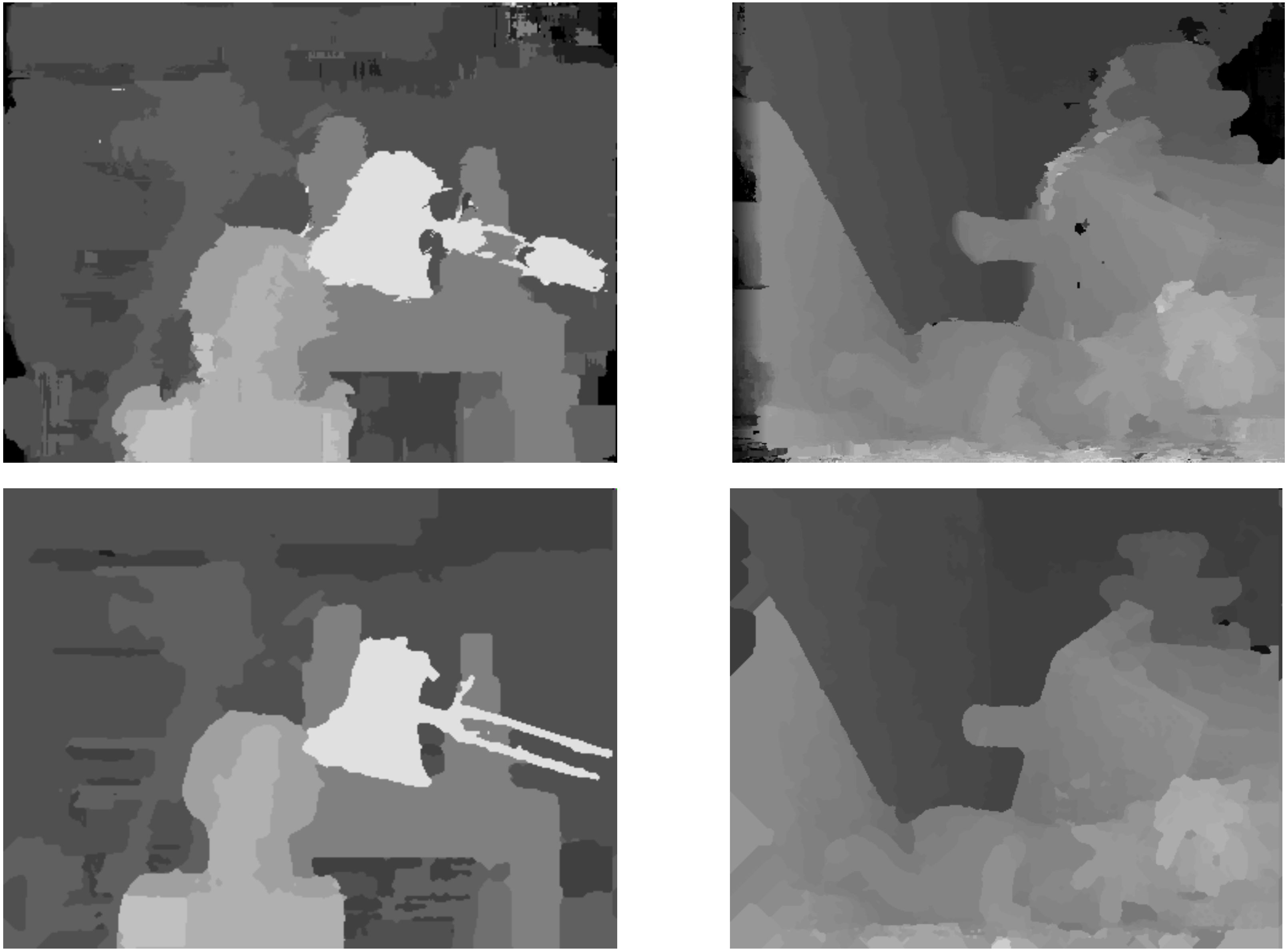
$$C_{agg}(p,q,d) = \frac{C_S(p,q,d)}{|Sp|} + \frac{C_W(p,q,d)}{r^2}$$

$$C_{S_p}(p,q,d) = \sum_{p_i \in S_p} TAD(p_i, q_{i+d})$$

$$C_{W_p}(p,q,d) = \sum_{p_i \in W_p} TAD(p_i, q_{i+d})$$

- **Cw tries to avoid 'segment locking'**
- **Cw may help in highly textured regions (small segments)**
- **However, Cw may introduce artifacts (discontinuities) since aggregation is performed on a fixed window**

Fast Aggregation: results



Fast Bilateral Stereo framework (FBS) [65]

- Symmetric support
- Combines accuracy of adaptive weights approaches with efficiency of traditional (correlative) approach
- Deploys a regularized range filter computed on a block basis of size wxw
- Increase noise robustness
- Efficient pixel-wise cost computation by means of integral-image/box-filtering schemes
- Results comparable to top performing approaches Segment Support and Adaptive Weights
- Fast: 32 sec on Teddy (w=3)
- Moreover, several trade-off speed vs accuracy are feasible: 14 sec (w=5) , 9 sec (w=7), 5 sec (w=9)

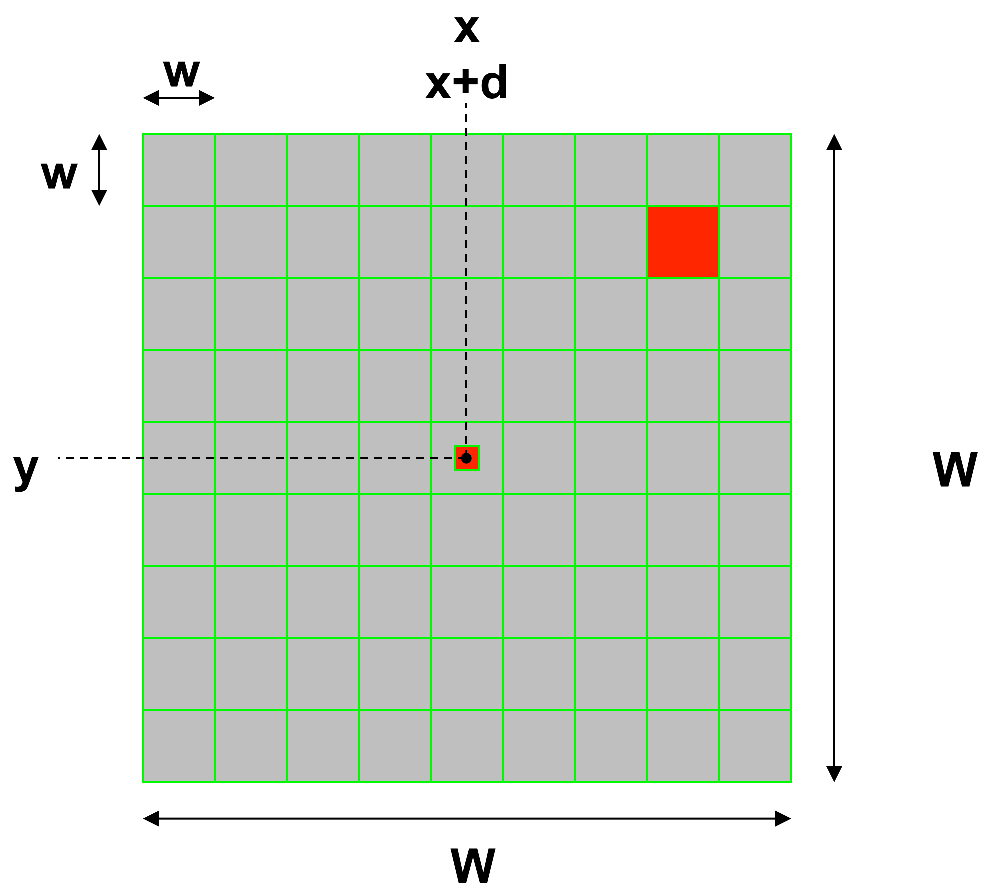
S. Mattoccia, S. Giardino,A. Gambini, Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering, Asian Conference on Computer Vision (ACCV2009)

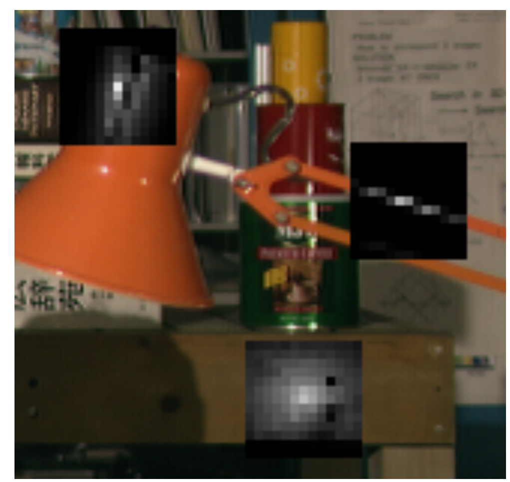
www.vision.deis.unibo.it/smatt/fast_bilateral_stereo.htm

- The range filter is computed on a block-basis deploying the average value within the block

- To avoid inaccurate localization of the discontinuities the central point is kept as reference

- Spatial filter computed on block basis

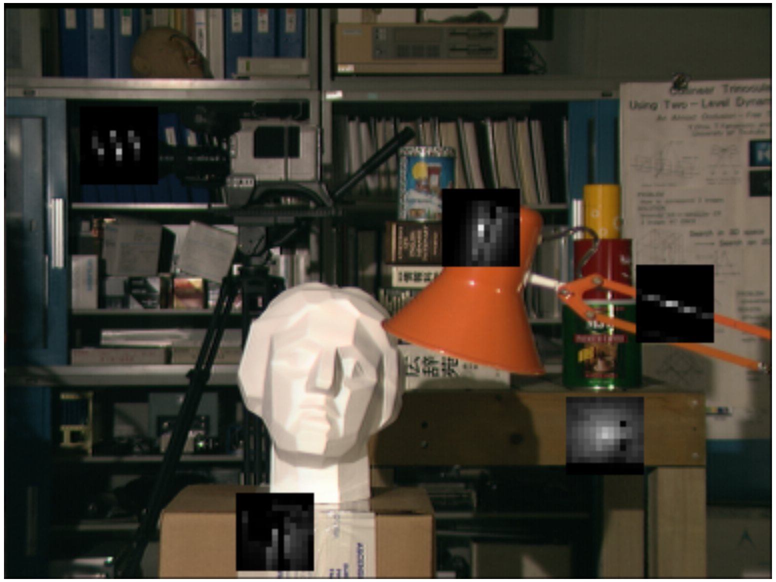




Three supports computed by Fast Bilateral Stereo

FBS (w=3) vs Adaptive Weights (AW)

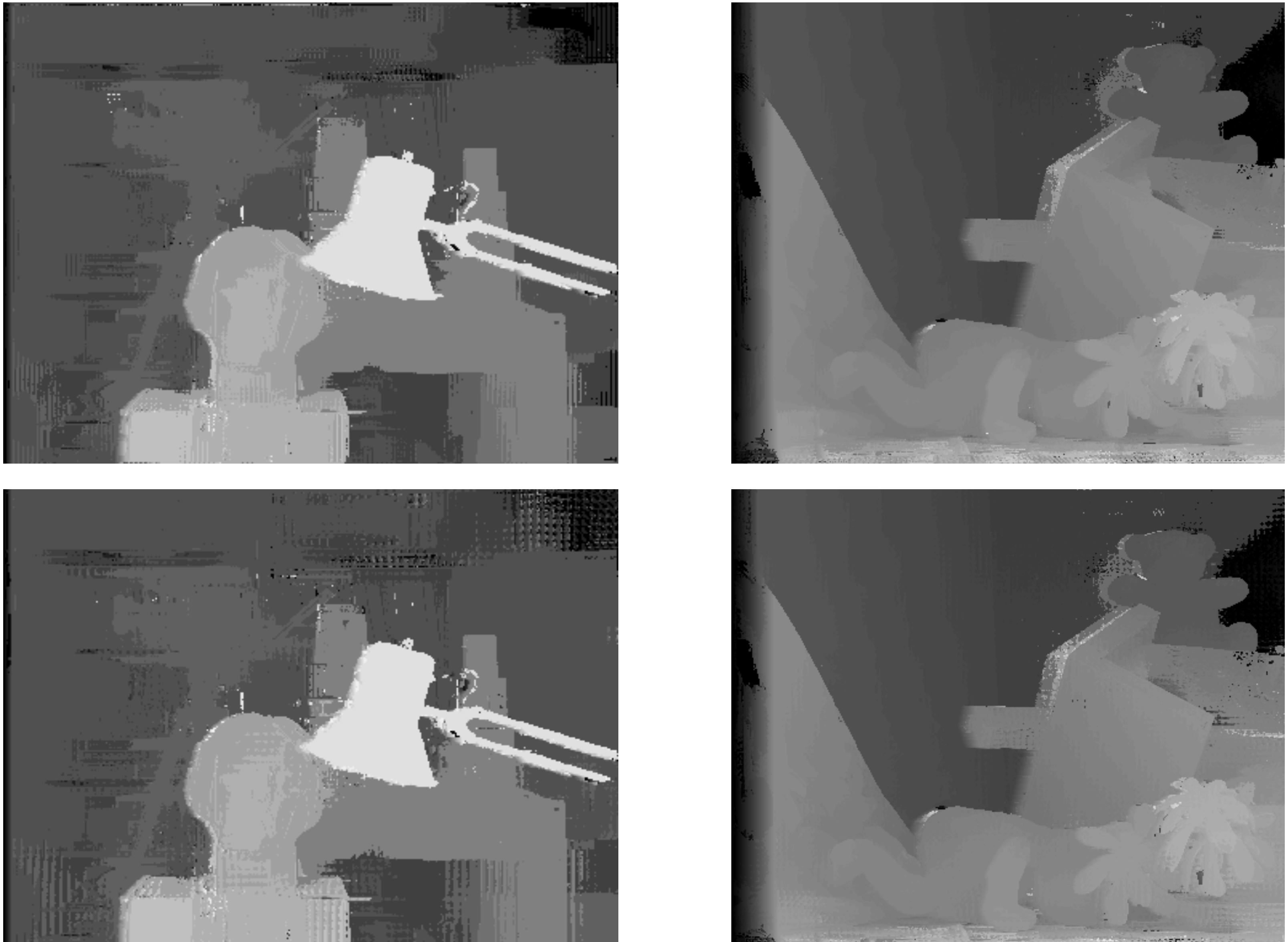
FBS



AW

Stefano Mattoccia

Fast Bilateral Stereo: results (w=3, w=5)



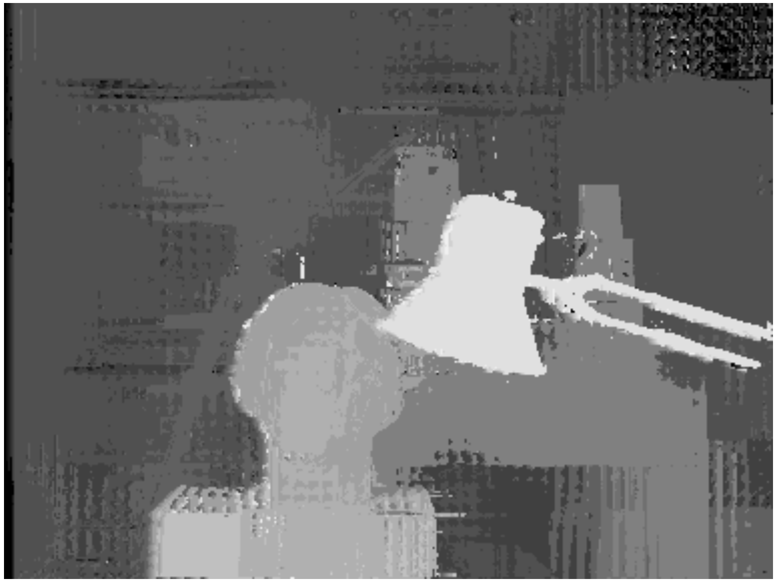
w=3

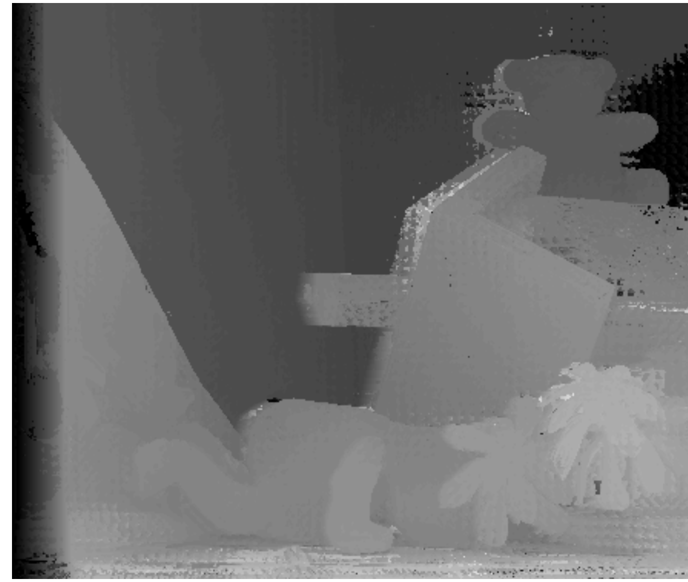
w=5

Stefano Mattoccia

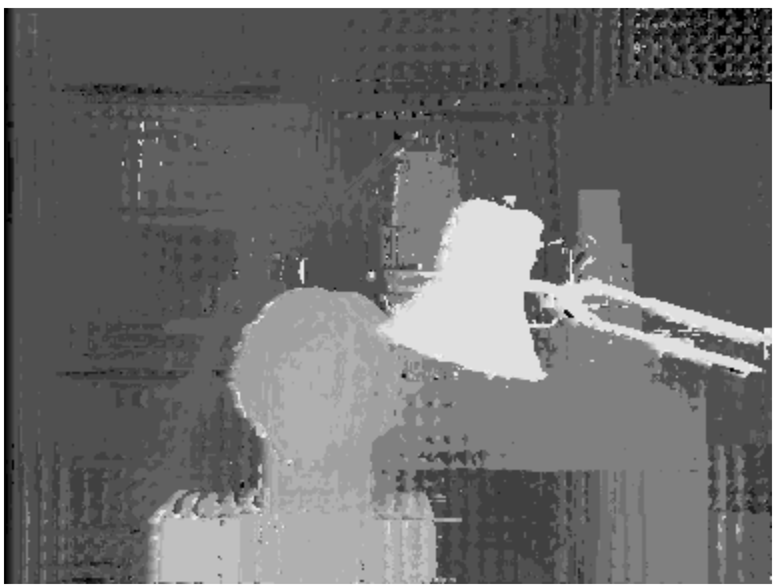
Fast Bilateral Stereo: results (w=7, w=9)

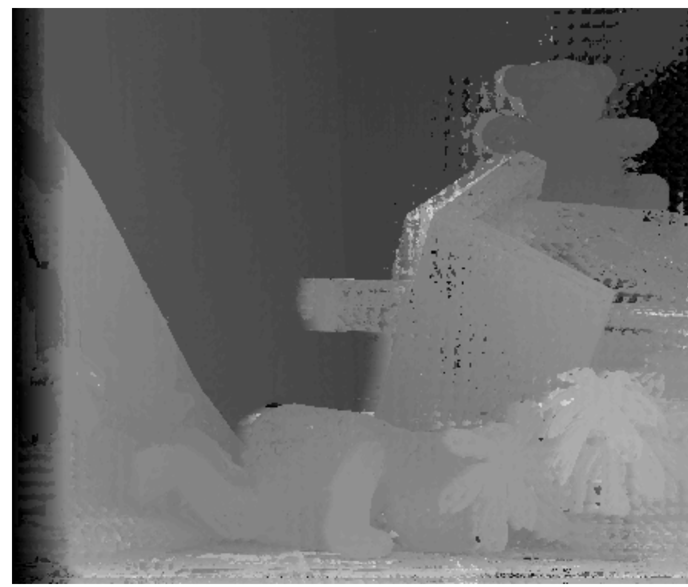
w=7





w=9





Fast Bilateral Stereo on the GPU [71]

- The *local nature* of the FBS algorithm allows to exploit parallel capabilities available in GPUs
- Compared to a single core CPU, on the Middlebury dataset, the implementation of FBS with CUDA enables to obtain:
  - 70X speed-up on an NVIDIA GEForce 460 GTX GPU
  - 100X speed-up on an NVIDIA Tesla C2070 GPU(*)

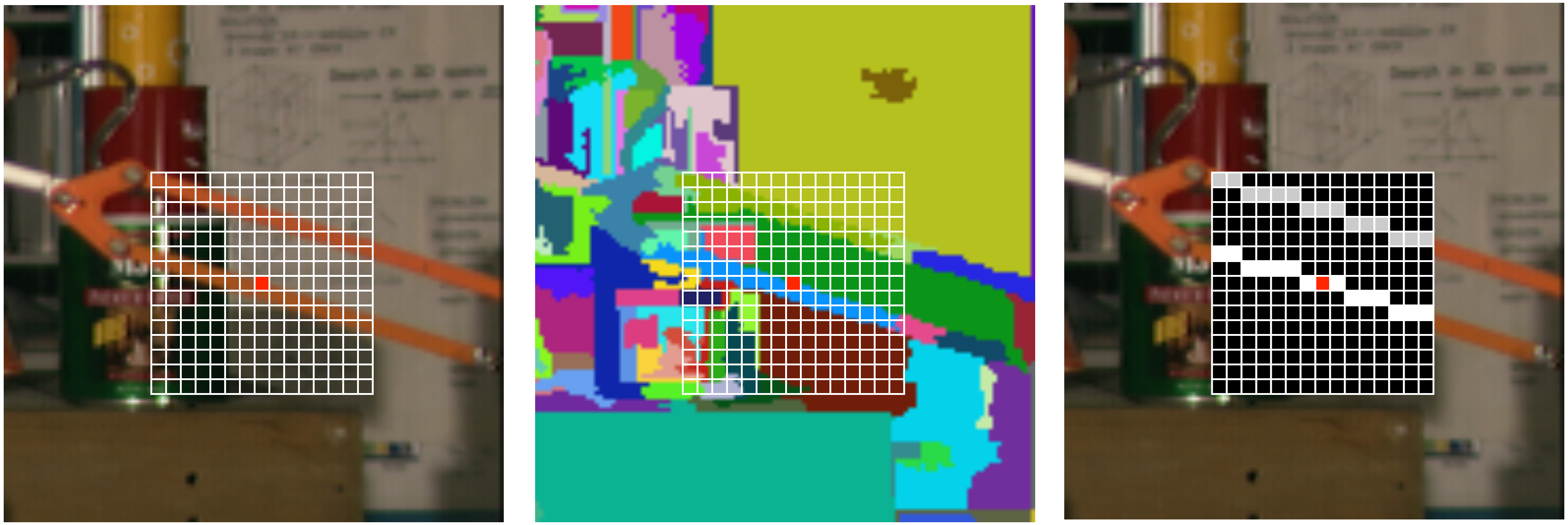
The measured execution time, with parameters w=3 and W=19, is (Teddy stereo pair): 300 ms for the GEForce 460 GTX and 200 ms on the Tesla C2070

Detailed results available in: www.vision.deis.unibo.it/smatt/FBS_GPU.html

(*) We acknowledge with thanks NVIDIA for the donation of the Tesla C2070

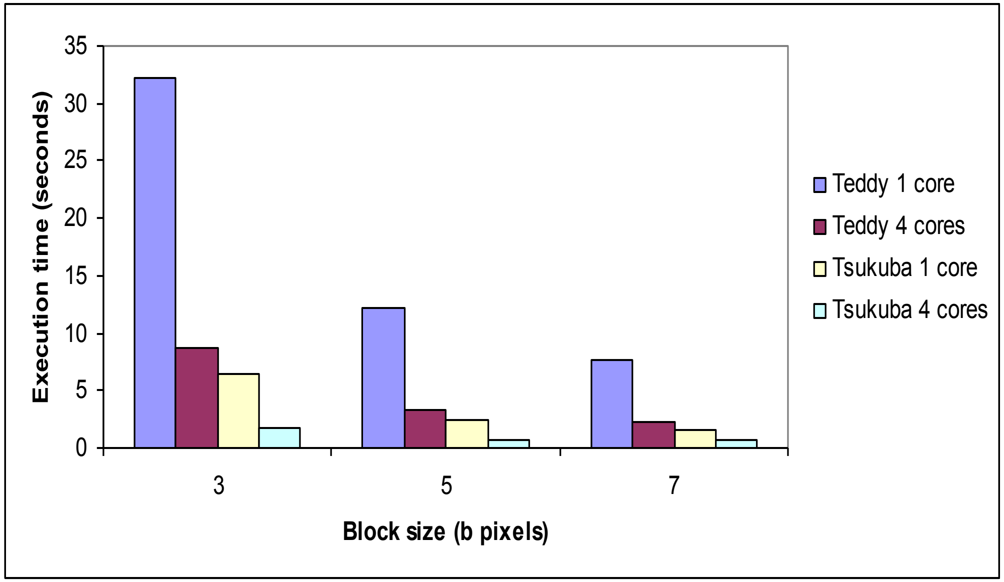
S. Mattoccia, M. Viti, F. Ries,. Near real-time Fast Bilateral Stereo on the GPU, 7th IEEE Workshop on Embedded Computer Vision (ECVW20011), CVPR Workshop, June 20, 2011, Colorado Springs (CO), USA

Fast Segmentation-driven (FSD)



- **Applies the SS strategy on a block basis**
- **Results equivalent to SS much more quickly (comparable to FBS)**
- **Compared to AW and FBS is effective also with greyscale images**

S. Mattoccia, L. De-Maeztu, A fast segmentation-driven algorithm for fast and accurate stereo correspondence, IC3D 2011



35
30
25
20
15
10
5
0
Execution time (seconds)
3
5
7
Block size (b pixels)
Teddy 1 core
Teddy 4 cores
Tsukuba 1 core
Tsukuba 4 cores

b=3

b=5

b=7

Locally Consistent (LC) stereo [66]

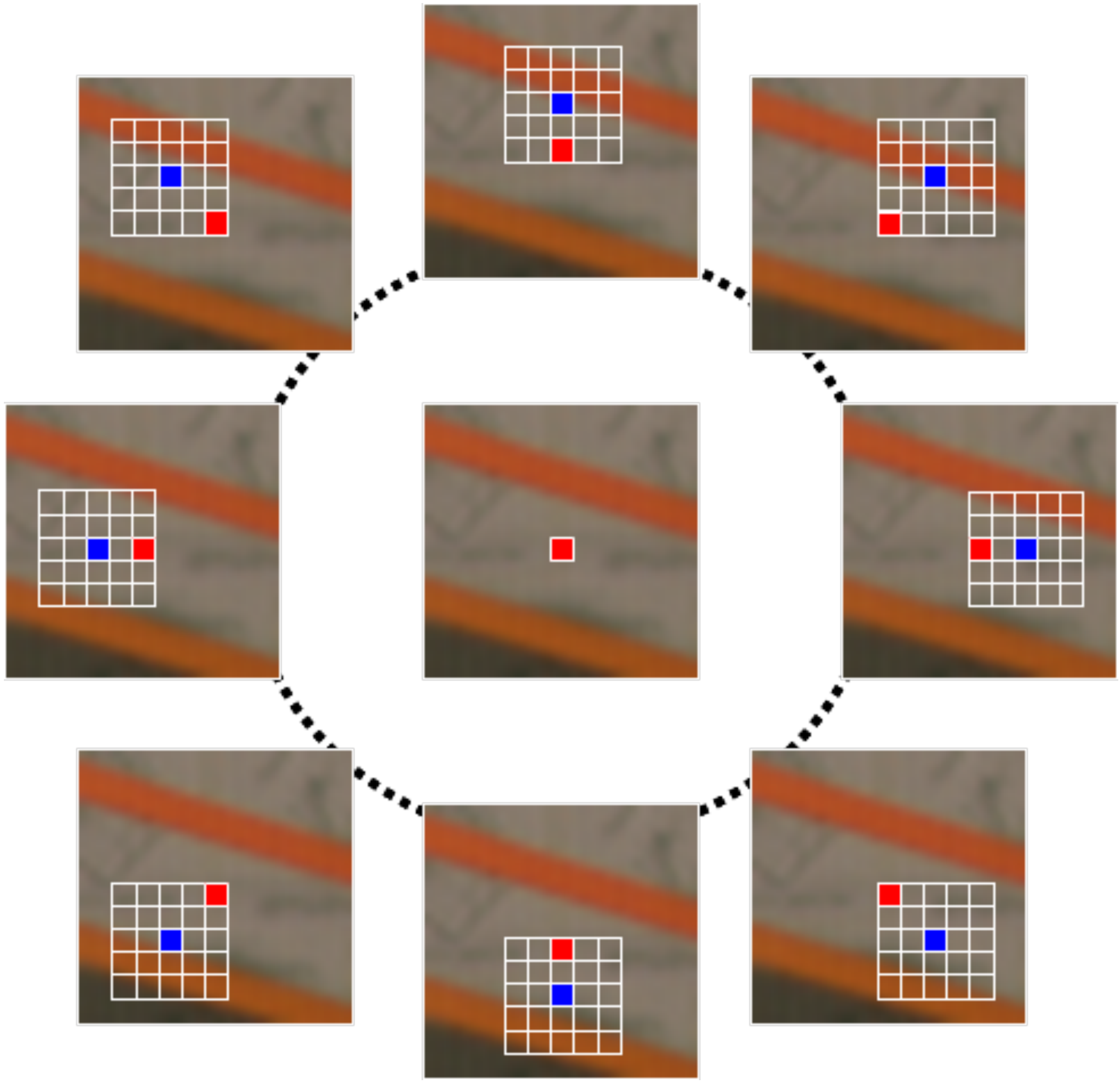
- Exploits the mutual relationships among neighboring pixels by explicitly modeling the continuity constraints
- Very accurate (significant improvements near depth discontinuities and low textured regions)
- Notable improvements compared to state-of-the-art approaches
- Fast 37 sec* on Teddy (unoptimized code) deploying the disparity hypotheses provided by Fast Bilateral Stereo
- Fast: 15 sec* on Teddy (unoptimized code) deploying the disparity hypotheses provided by Fixed Window

* significantly reduced (see next slides/ECVW 2010 paper [68])

S. Mattoccia, A locally global approach to stereo correspondence, 3D Digital Imaging and Modeling (3DIM2009)

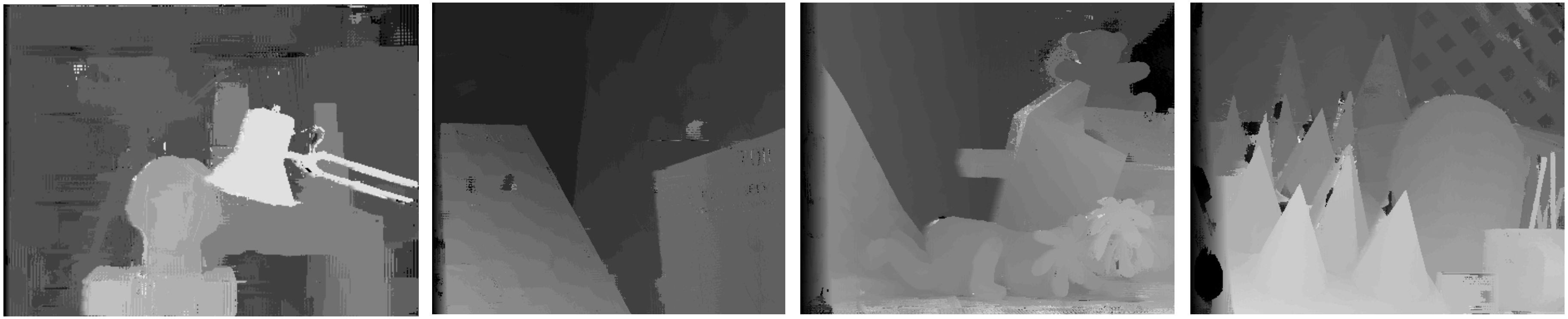
www.vision.deis.unibo.it/smatt/lc stereo.htm

Measurements performed on a 2.49GHz Intel Core Duo processor

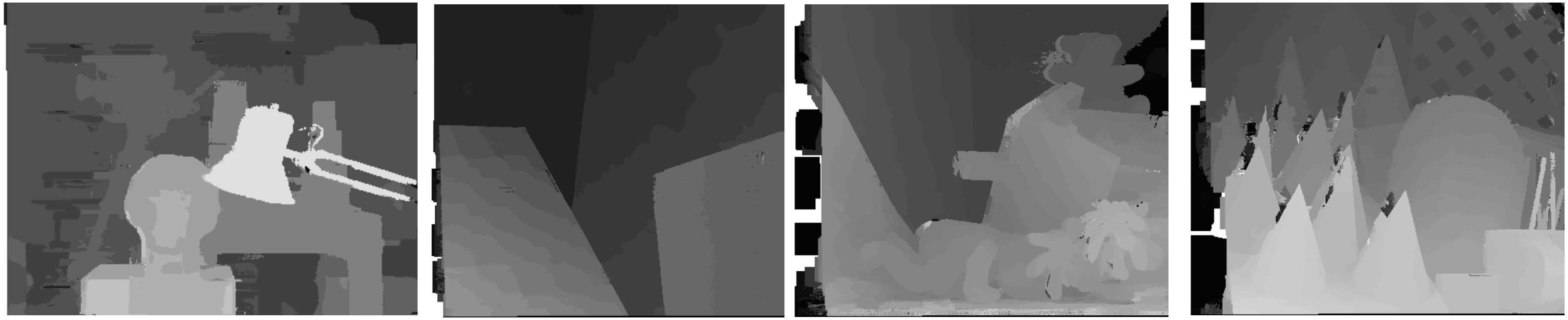


Locally Consistent stereo: results with FBS

Before ($FBS_{19(3)}$)



LC_{19}

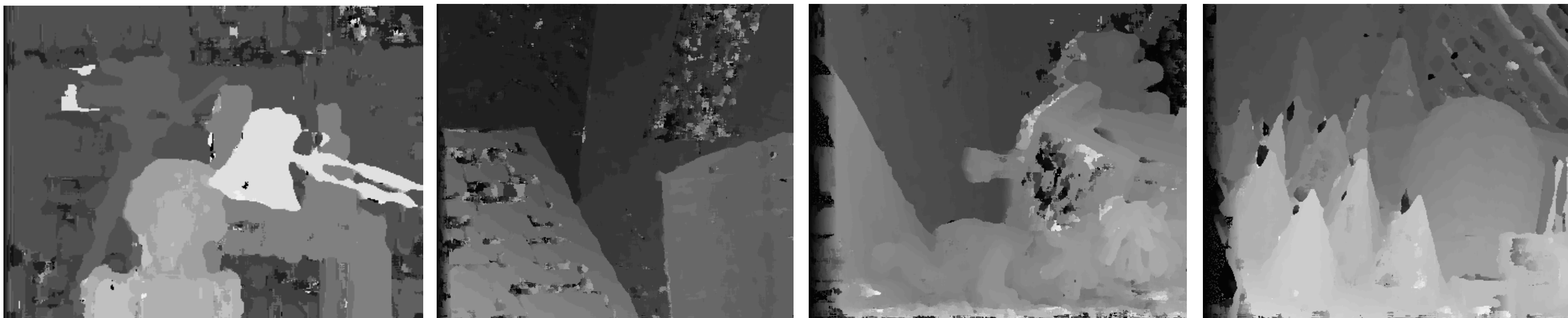


After LC_{19} (+ FBS_{19})

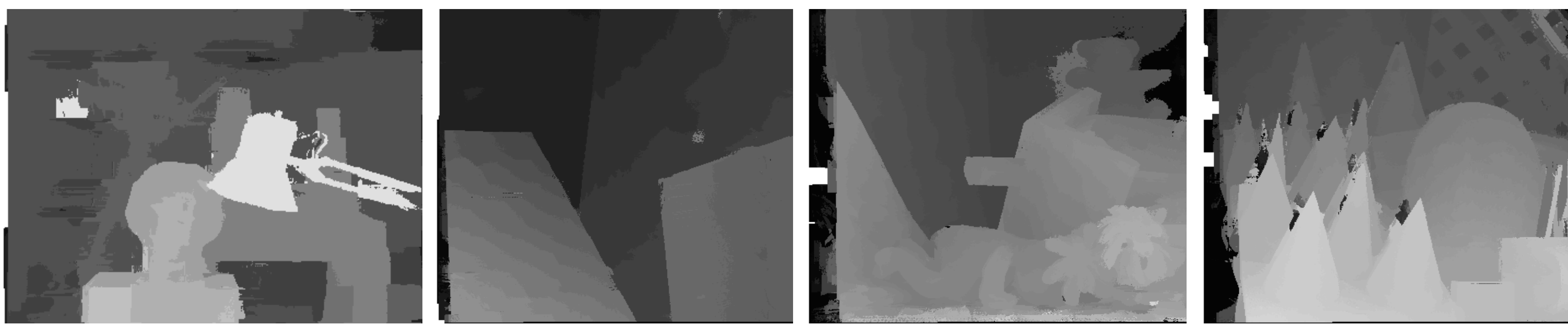
Stefano Mattoccia

Locally Consistent stereo: results with FW

Before (FW_4)



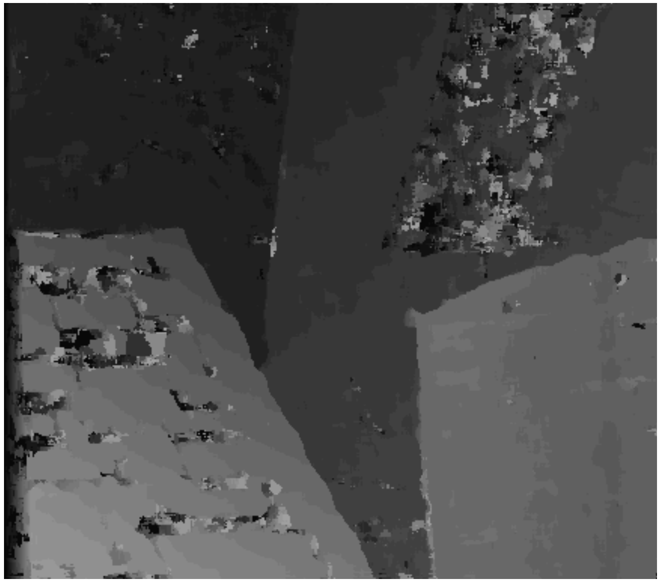
LC_{19}

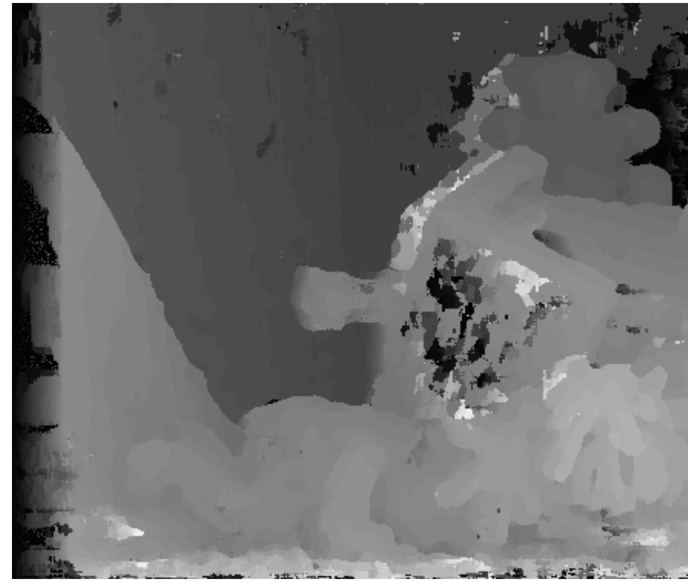


After LC_{19} (+ FW_4)

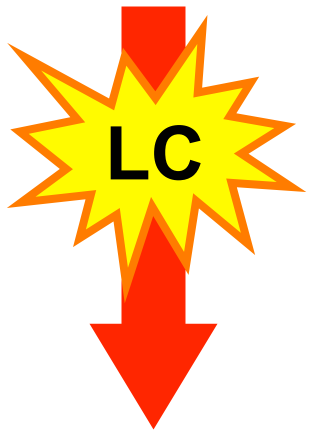
Stefano Mattoccia

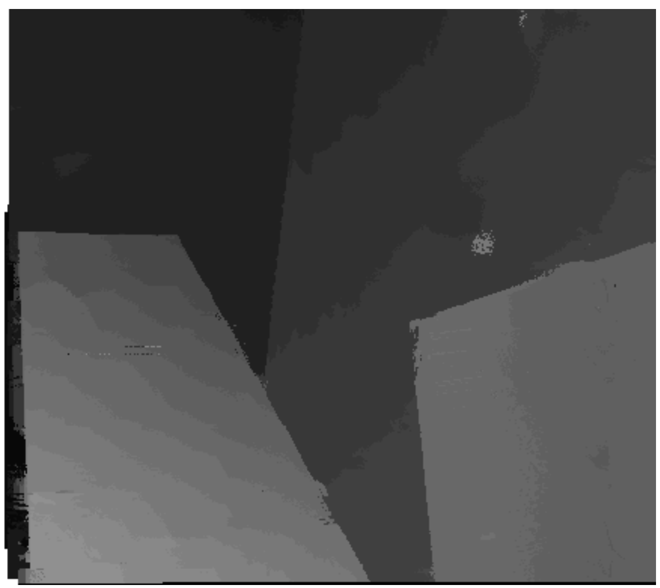
Locally consistent (LC) stereo vs FW: details





FW_4





After

LC_{19} (+ FW_4)

Stefano Mattoccia

- The next slide provides an updated quantitative evaluation of the approaches described so far (yellow) according to the methodology described in [1]
- The updated evaluation is available online at:

http://www.vision.deis.unibo.it/spe/SPEresults.aspx

- According to this evaluation the Locally Consistent approach combined with the disparity hypotheses provided by the Fast Bilateral Stereo (FBS) algorithm outperforms the other approaches
- The FBS ranks second and provides a good trade-off between accuracy and execution time (see the results in the table with different parameters of the FBS algorithm)
- In the successive slides will be described novel approaches that rely on the LC technique (see papers [67],[68],[69])

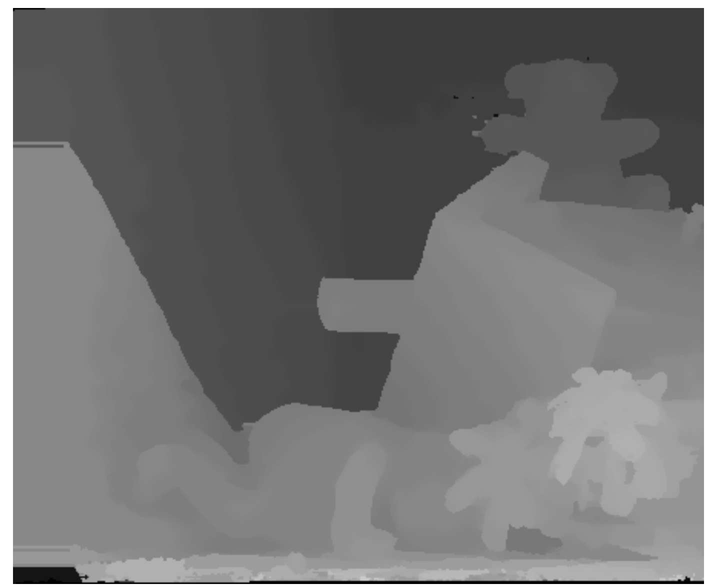
(Updated) Quantitative evaluation [1] (TAD)

Algorithm	Rank Acc.	Tsukuba nonocc	Tsukuba disc	Venus nonocc	Venus disc	Teddy nonocc	Teddy disc	Cones nonocc	Cones disc	Time Teddy hh:mm:ss
LocallyConsist(FBS 39(3))	**1**	1.77	5.92	0.27	1.77	9.3	17.9	4.75	10.5	**00:00:37**
FBS 39(3)	**3.13**	2.95	8.69	1.15	6.64	10.7	20.8	5.23	11.4	**00:00:28**
Segment support	**3.25**	2.15	7.22	1.38	6.27	10.5	21.2	5.83	11.8	**00:39:30**
LocallyConsist(FW)	**3.5**	3.07	9.63	0.66	5.11	10.6	21.8	5.3	11.6	**00:00:15**
FBS 45(5)	**5.75**	3.34	9.99	2.11	6.72	11.5	21.8	6.81	13.8	**00:00:14**
Segmentation based	**6.75**	2.25	8.87	1.37	9.4	12.7	24.8	11.1	20.1	**00:05:14**
Adaptive Weight	**6.88**	4.66	8.25	4.61	13.3	12.7	22.4	5.5	11.9	**00:20:35**
FBS 49(7)	**7**	3.99	12.3	3.01	8.42	12.3	23	7.5	15.1	**00:00:09**
FBS 45(9)	**8.75**	4.6	13.7	5.42	10.6	13.9	24.8	9.47	17.7	**00:00:05**
Variable Windows	**11.13**	3.12	12.4	2.42	13.3	17.7	25.5	21.2	27.3	**00:00:26**
Reliability	**11.13**	5.08	17.9	3.92	13.9	18.9	29.9	11.3	18.3	**00:13:39**
Multiple windows* (25W)	**14.5**	7.57	22.7	3.91	21.1	20.9	33.2	13.7	26.9	**00:00:13**
Multiple windows (9W)	**14.88**	7.6	25.7	7.02	33	16	36.9	10.6	26.9	**00:00:04**
Multiple windows (25W)	**15.13**	7.28	25.9	6.18	29	18	35.6	11.8	27.1	**00:00:14**
Gradient guided	**15.25**	7.41	16.2	12.9	32.3	20.1	32.8	13.5	24.9	**00:00:16**
Multiple windows* (9W)	**15.63**	9.18	22.6	6.23	28.1	21.4	34.5	13.2	26.7	**00:00:04**
Recursive adaptive	**16.38**	9.66	29.8	5.94	29.8	20.1	34.6	11.7	25.3	**00:20:20**
Shiftable windows	**16.75**	9.58	14.4	9.66	16.5	23.6	31.2	24.4	33.6	**00:00:05**
Multiple windows (5W)	**16.88**	7.62	27.2	7.55	37.2	17.4	39.7	11	27.8	**00:00:02**
Multiple adaptive	**17**	11.7	27.3	11.9	13.7	20.4	31.8	15.8	25.3	**02:08:17**
Multiple windows* (5W)	**18.25**	9.61	25.1	9.36	38.3	22.2	38	12.1	27.5	**00:00:02**
Max connected	**21**	11.8	26.4	42.5	50.9	34.5	41	17.7	22.7	**01:59:09**
Fixed Window (FW)	**21.13**	9.58	27.1	10.6	42.5	25.1	42.4	19.7	36	**< 1 s**
Oriented rod*	**22.25**	18.6	31.1	20.3	26.6	30.7	41.8	37.8	47.3	**00:17:19**
Oriented rod	**22.5**	14.2	25.8	21.9	29.8	37.5	48.6	48.5	55.5	**00:17:00**
Radial adaptive	**23**	14.8	21.8	22.4	40.4	49.6	50.1	50.2	53.6	**01:06:21**

Table available at: http://www.vision.deis.unibo.it/spe/SPEresults.aspx

O(1) adaptive cost aggregation

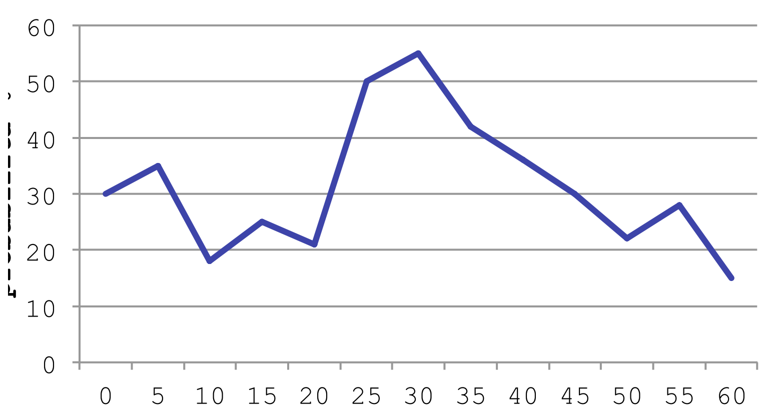
- Symmetric cost aggregation inspired by guided filter
- Aggregation independent of the window size
- Can be applied to color images (differently by integral histogram-based methods)
- Results comparable to state of the art



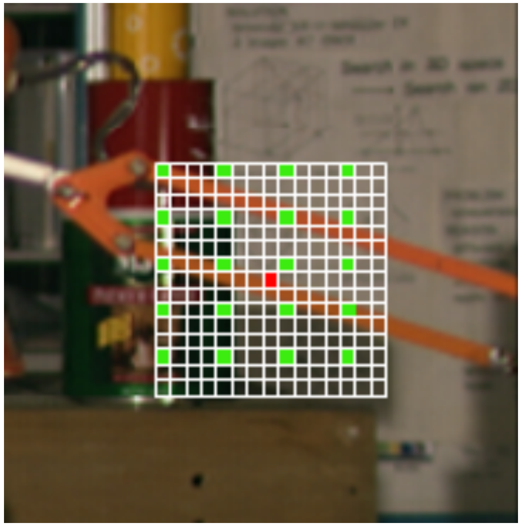
L. De-Maeztu, S. Mattoccia, A. Villanueva, R. Cabeza, "Linear stereo matching", ICCV 2011

Fast/simplified adaptive cost aggregation

- Asymmetric cost aggregation
- Cost computed on a selected number of points (determined by means of FW (5x5))
- Matching cost computed on a subset of (fixed) points



+



D. Min, J. Lu, and M. Do, "A revisit to cost aggregation in stereo matching: how far can we reduce its computational redundancy?", ICCV 2011

Disparity computation/optimization (3)

This step aims at finding the best disparity assignment (e.g. the best path/surface within the DSI) that minimizes a cost function over the whole* stereo pair.

In many cases the energy function has two terms:

$$E(d) = E_{data}(d) + E_{smooth}(d)$$

- **The data term E_{data} measure how well the assignment fits to the stereo pair (in terms of overall matching cost). Several approaches rely on simple pixel-based cost functions but effective support aggregation strategies have been successfully adopted**

- **The smoothness/regularization E_{smooth} term explicitly enforces piecewise assumptions (continuity) about the scene. This term penalizes disparity variations and large variation are allowed only at (unknown) depth borders. Plausibility of depth border is often related to edges.**

*** subset of the stereo pair**

Since finding the best assignment that minimizes the energy function a NP-hard problem, approximated but effective energy minimization strategies have been proposed.

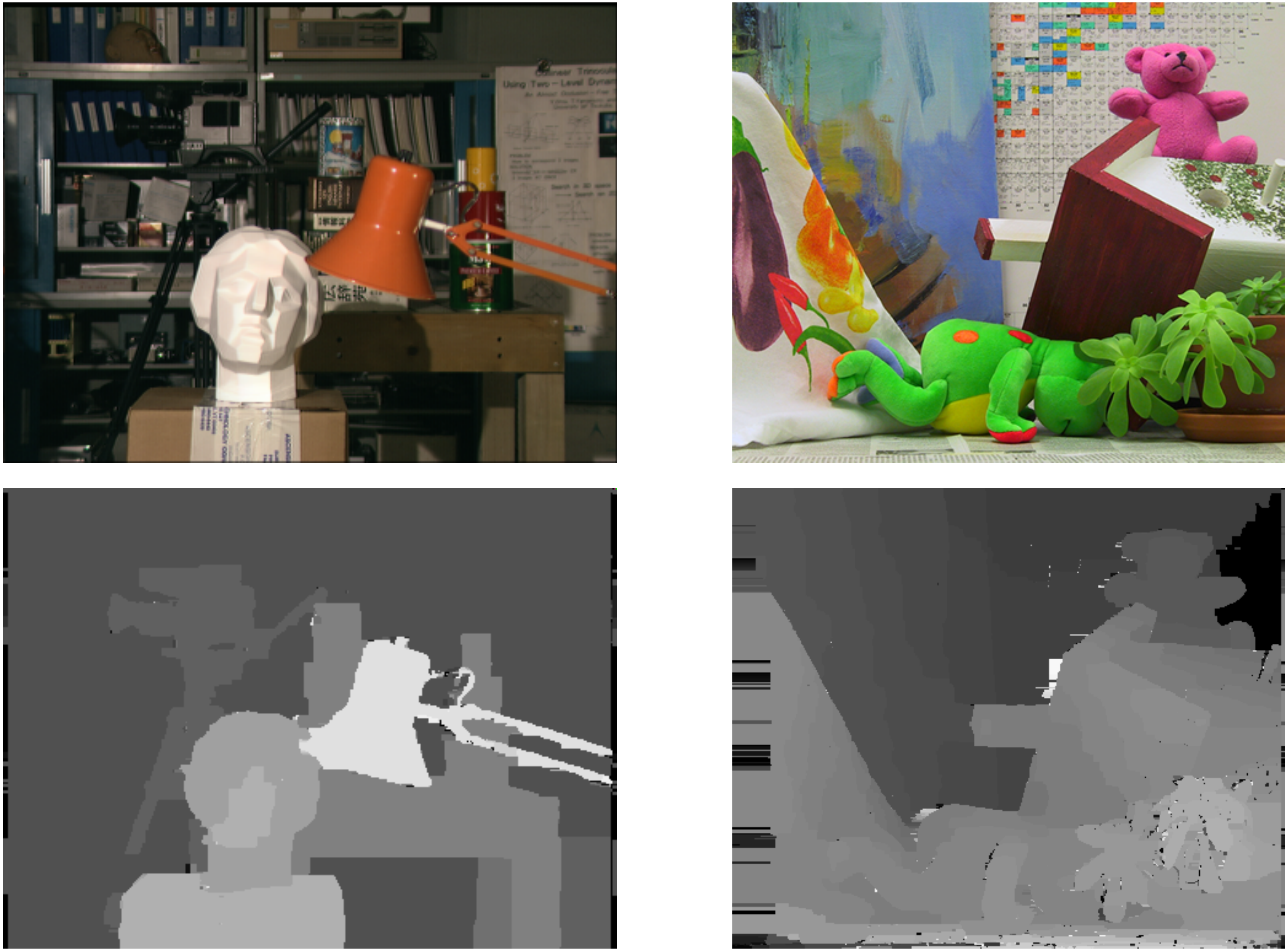
Relevant approaches are:

- **Graph Cuts [52]**
- **Belief Propagation [53]**
- **Cooperative optimization [54]**

A detailed comparison of relevant energy minimization methods can be found in [63].

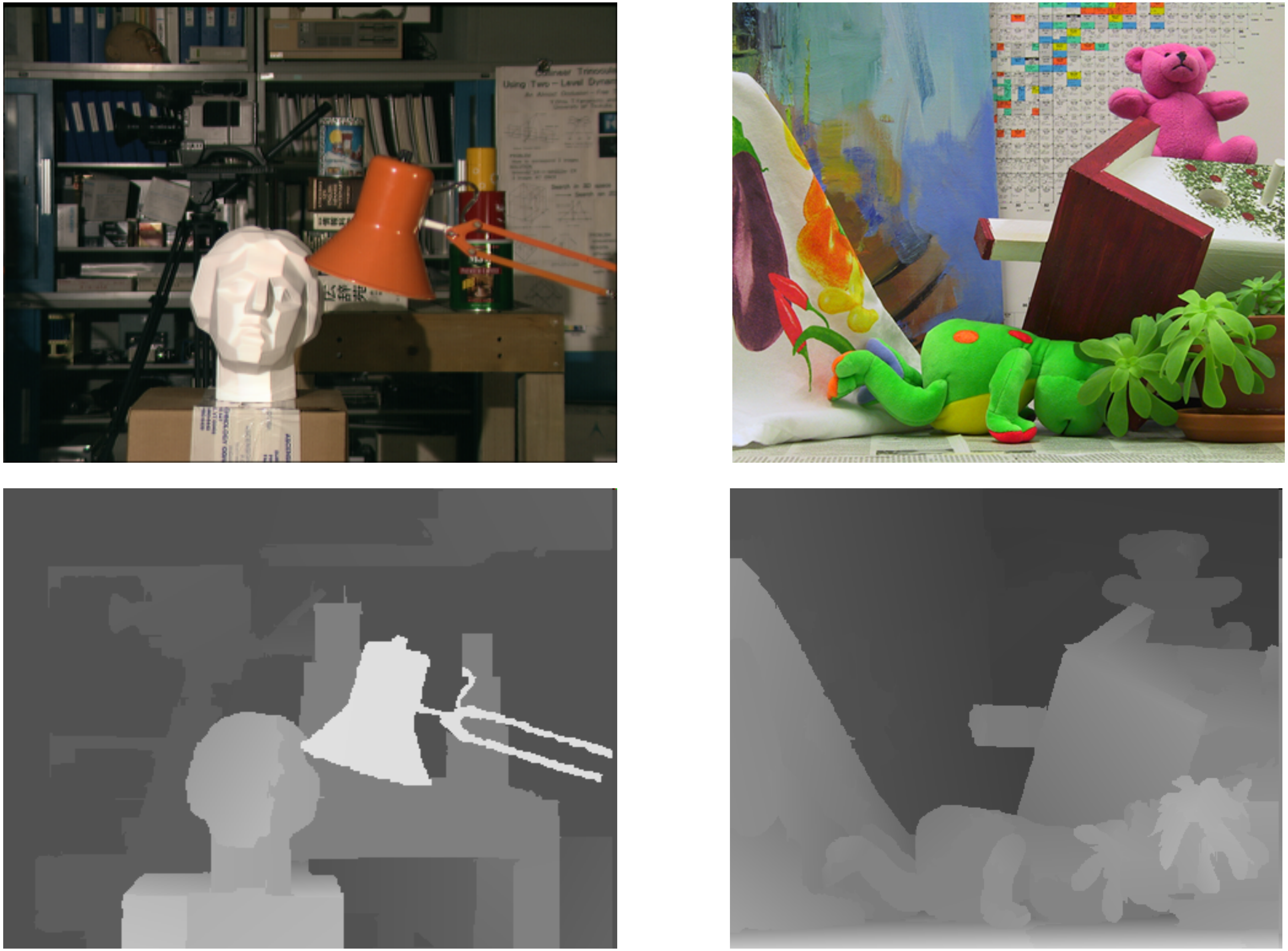
An further and interesting class of approximated approaches minimizes the energy function on a subset of points of the stereo pair (typically along scanlines). In these cases the energy minimization problem is efficiently solved by means of Dynamic Programming (DP) or Scanline Optimization (SO) techniques.

Graph Cuts



V. Kolmogorov and R. Zabih, Computing visual correspondence with occlusions using graph cuts, ICCV 2001

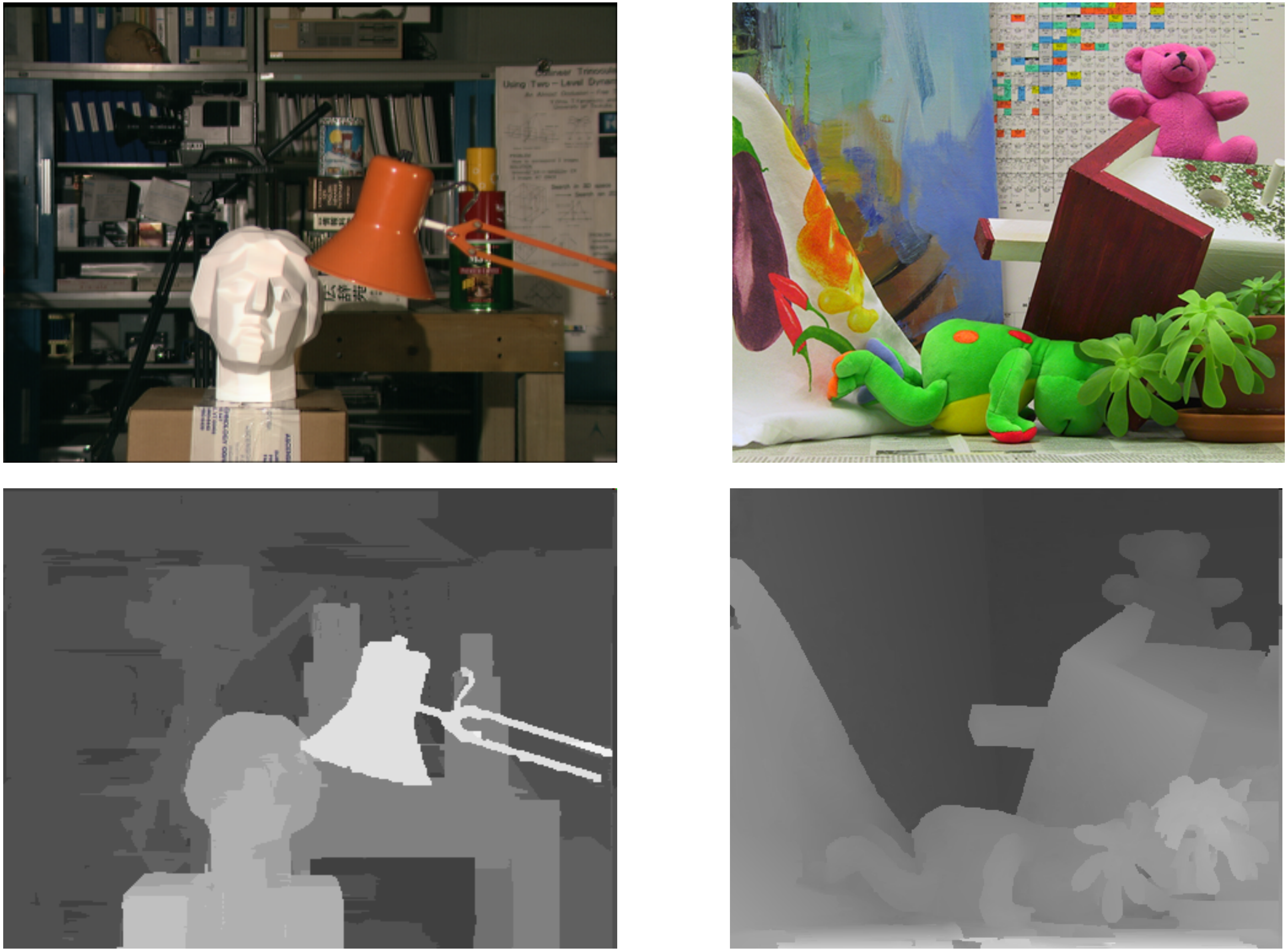
BP + segmentation



A. Klaus, M. Sormann and K. Karner, Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. ICPR 2006

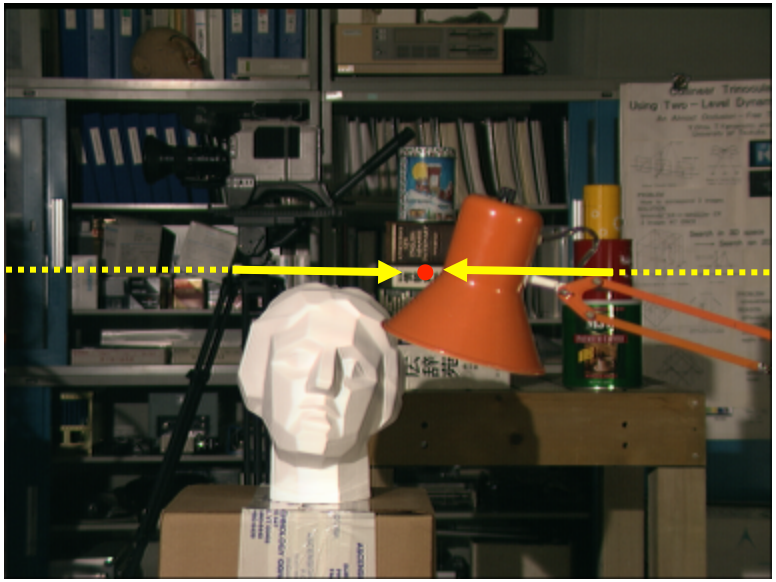
Stefano Mattoccia

Cooperative + segmentation



Z. Wang and Z. Zheng, A region based stereo matching algorithm using cooperative optimization, CVPR 2008

Dynamic Programming (DP)



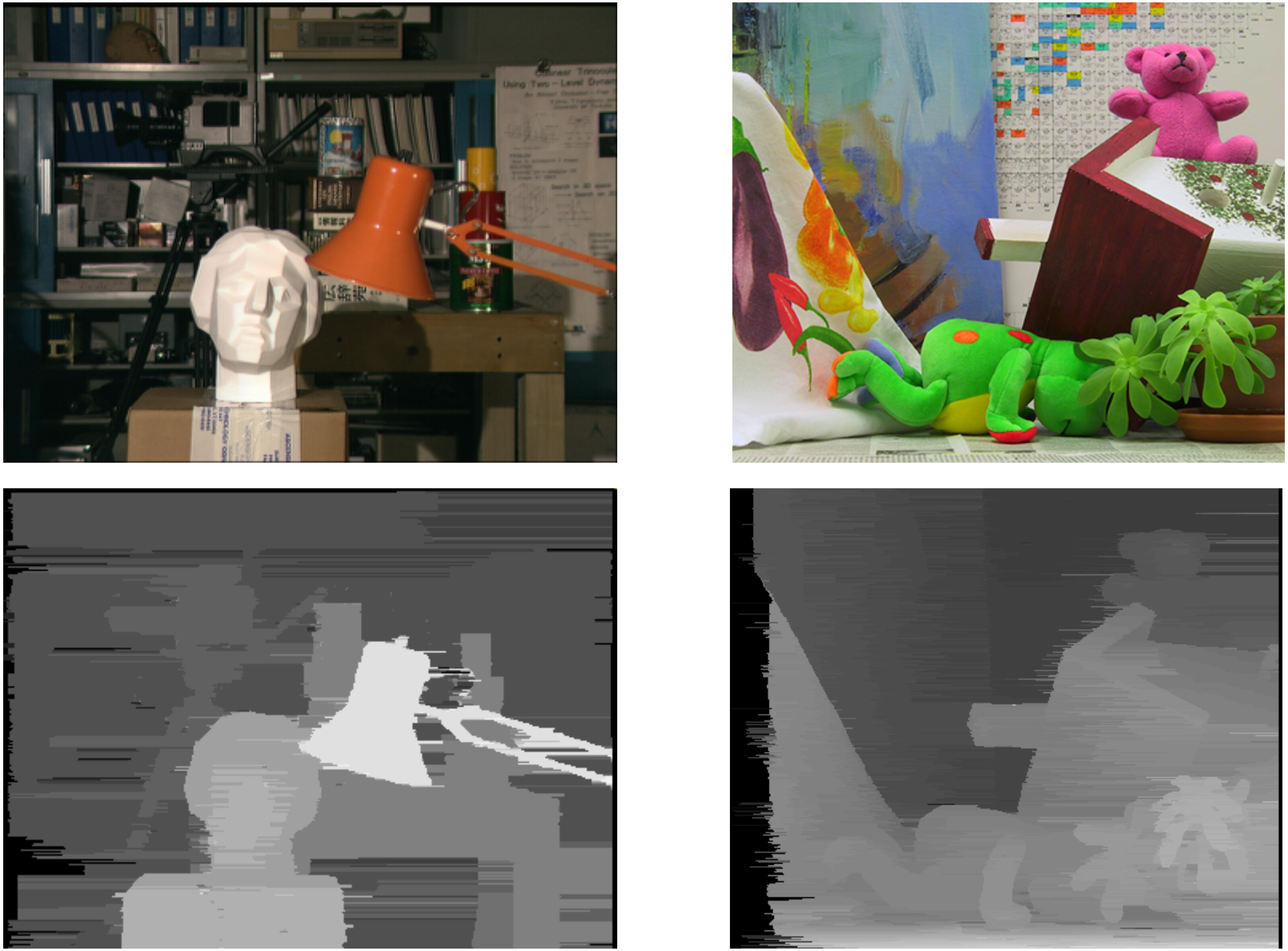
- efficient (polynomial time) ≈ 1 sec

- enforces the ordering constraint

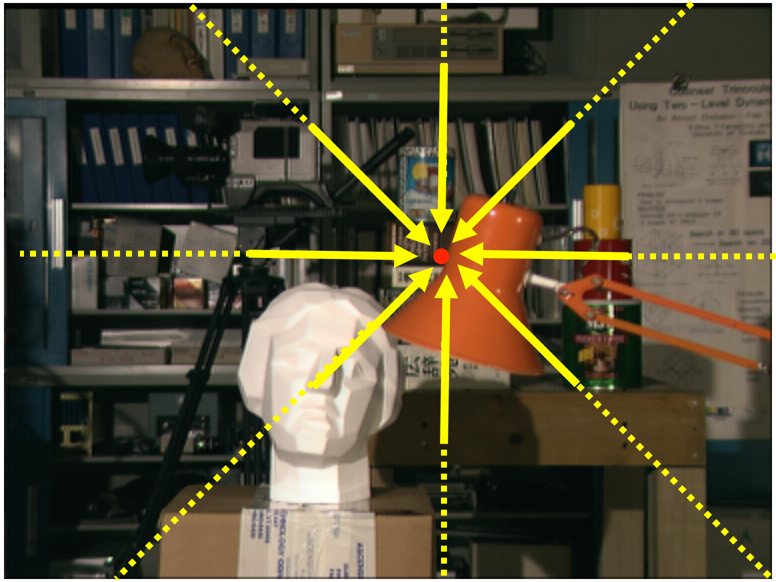
- accurate at depth borders and uniform regions

- streaking effect (see next slide)

DP [11]



Scanline Optimization (SO)

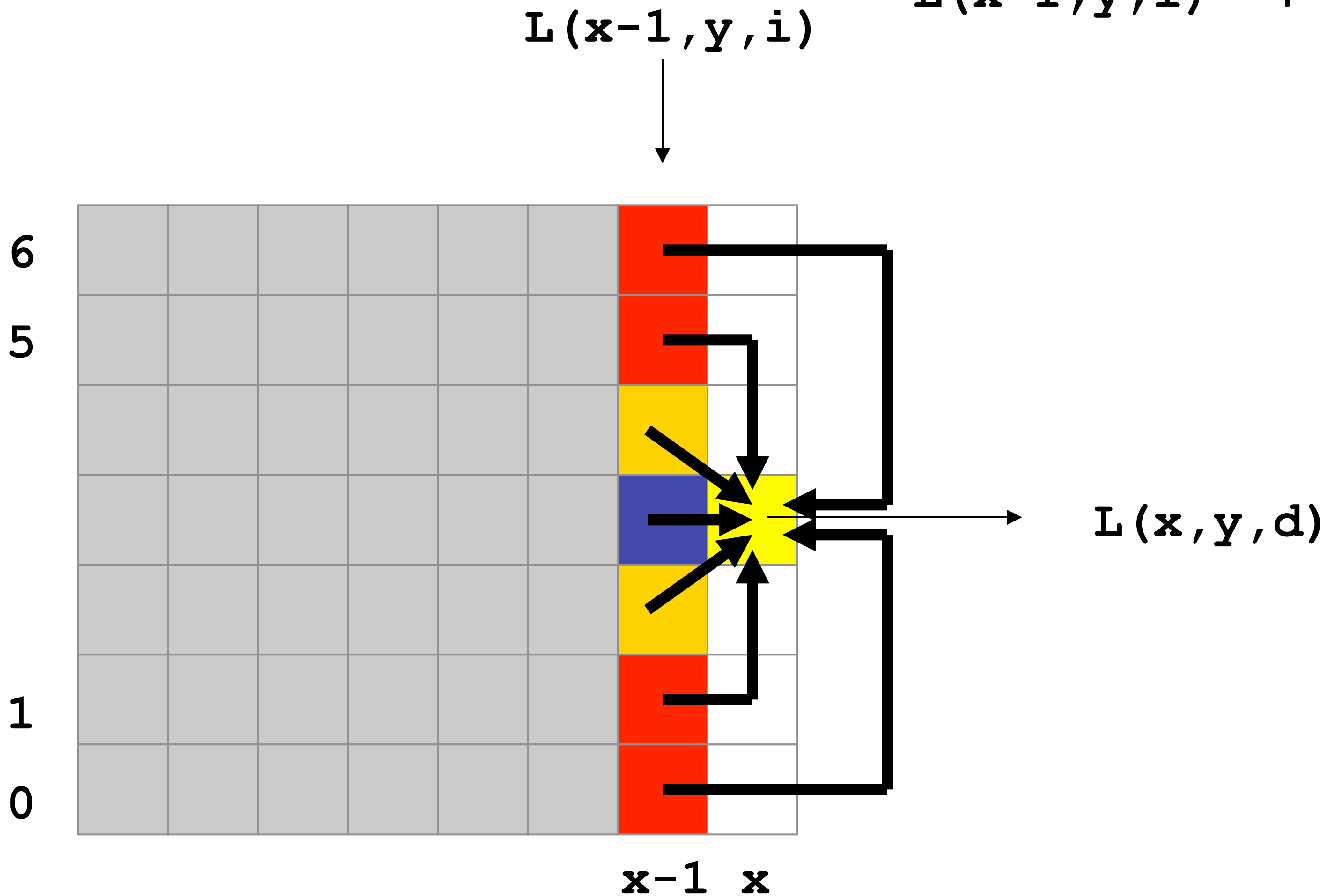


- Efficient (polynomial time) ≈ few seconds
- Cannot enforce the ordering constraint
- accurate at depth borders and uniform regions
- overcomes the streaking effect problem (see next slide)
- high memory requirement

H. Hirschmüller. Stereo vision in structured environments by consistent semi-global matching. CVPR 2006, PAMI 30(2):328-341, 2008

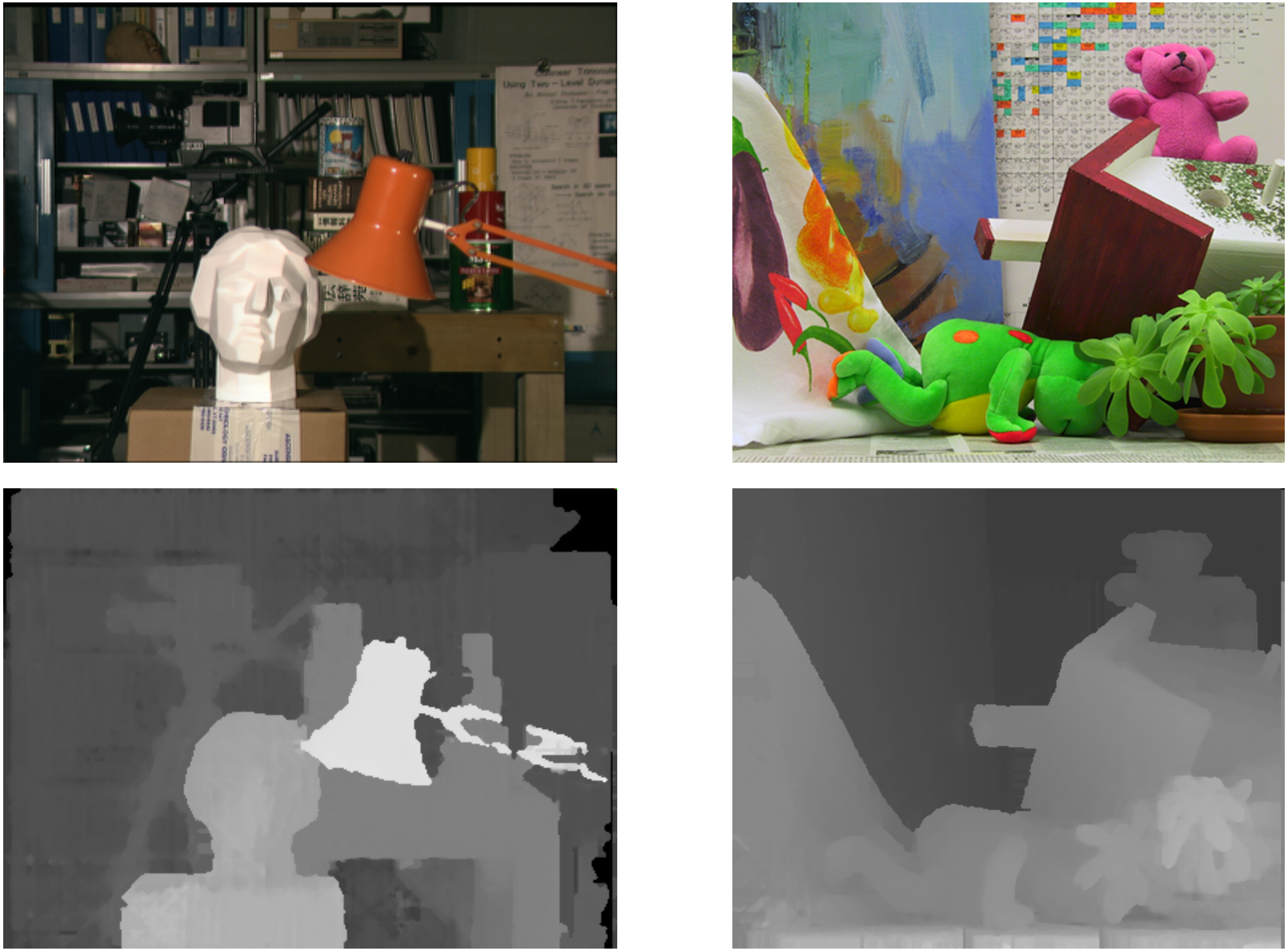
In SO, the cost is defined as:

$$L(x,y,d) = C(x,y,d) + \min \begin{cases} L(x-1,y,d), \\ L(x-1,y,d-1) + P1, \\ L(x-1,y,d+1) + P1, \\ L(x-1,y,i) + P2 \end{cases}$$

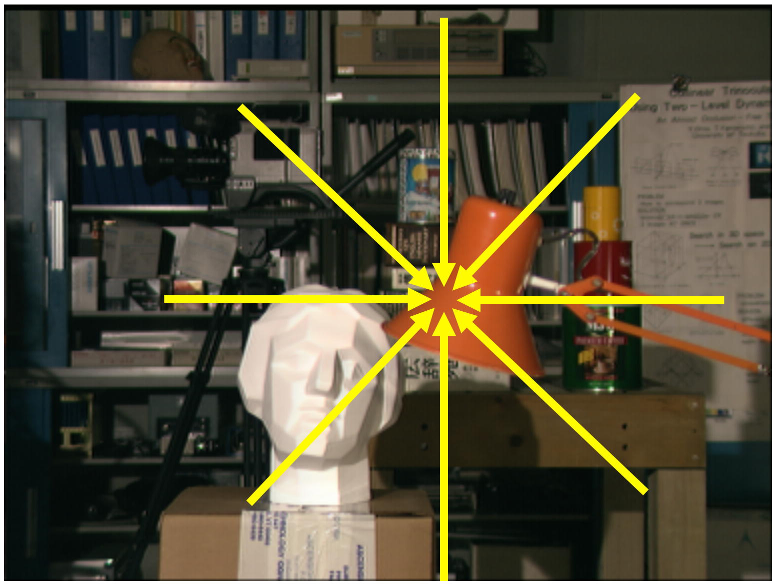


$$L(x,y,4) = C(x,y,4) + \min \begin{pmatrix} L(x-1,y,4) \\ L(x-1,y,5) + P_1 \\ L(x-1,y,3) + P_1 \\ L(x-1,y,7) + P_2 \\ L(x-1,y,6) + P_2 \\ L(x-1,y,2) + P_2 \\ L(x-1,y,1) + P_2 \\ L(x-1,y,0) + P_2 \end{pmatrix} - \min L(x-1,y,k)$$

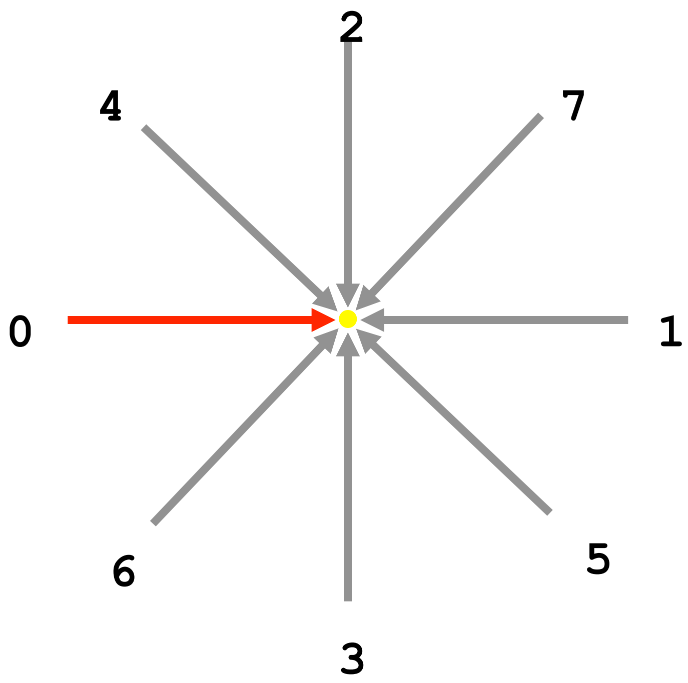
Scanline Optimization [30]



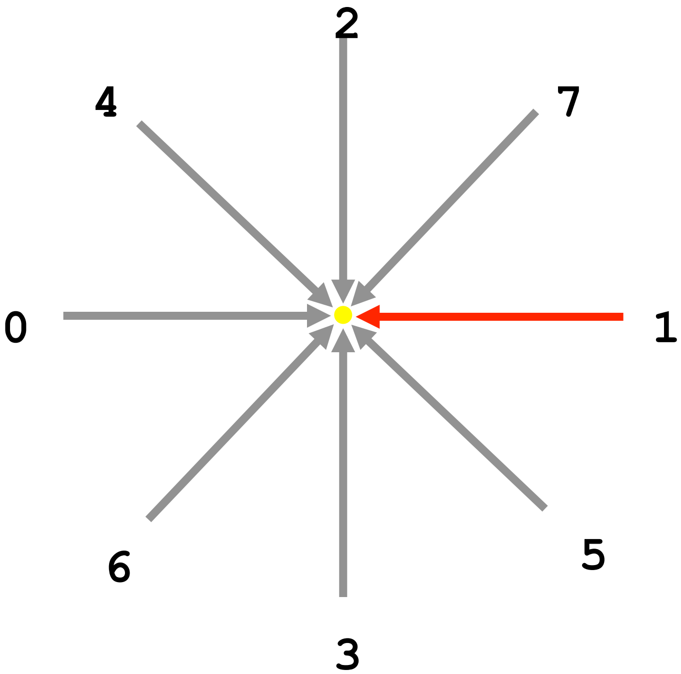
Scanline Optimization: details



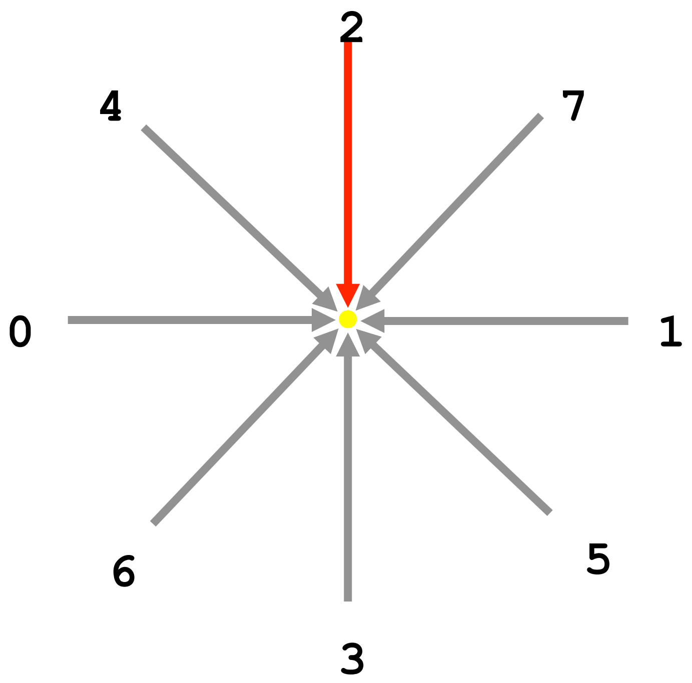
Scanline 0



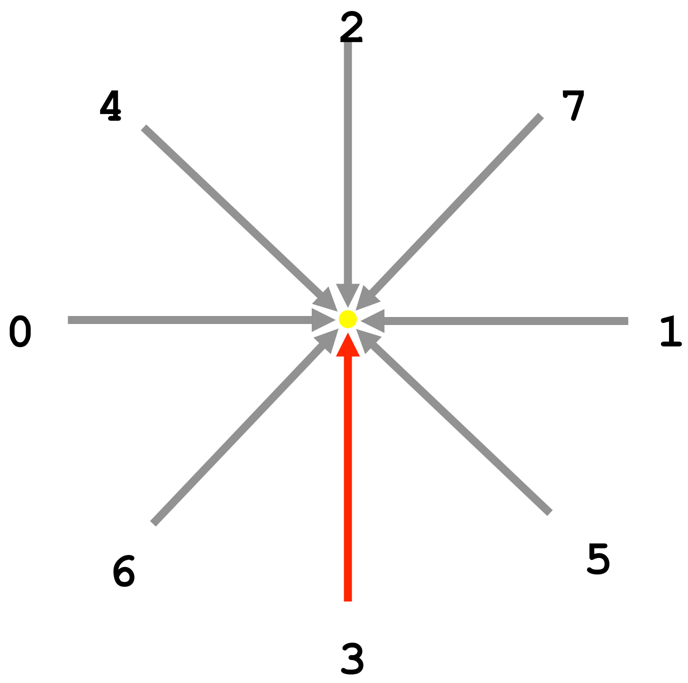
Scanline 1

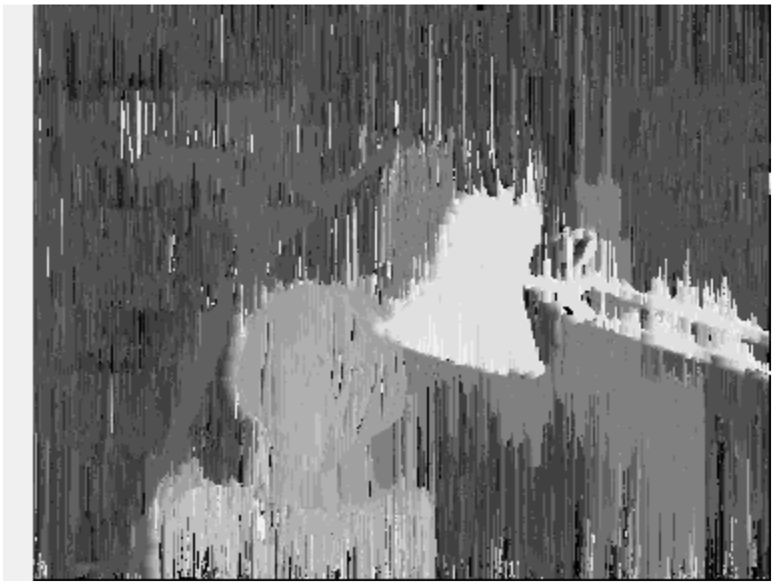


Scanline 2

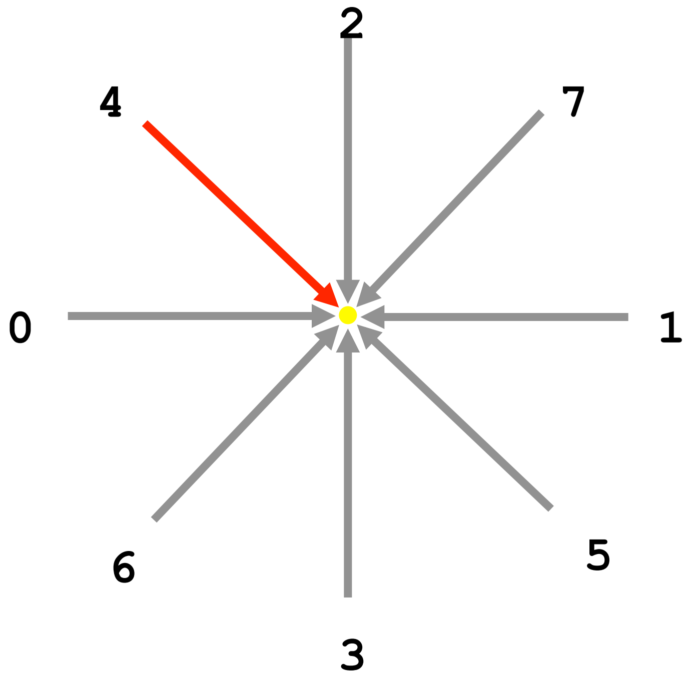


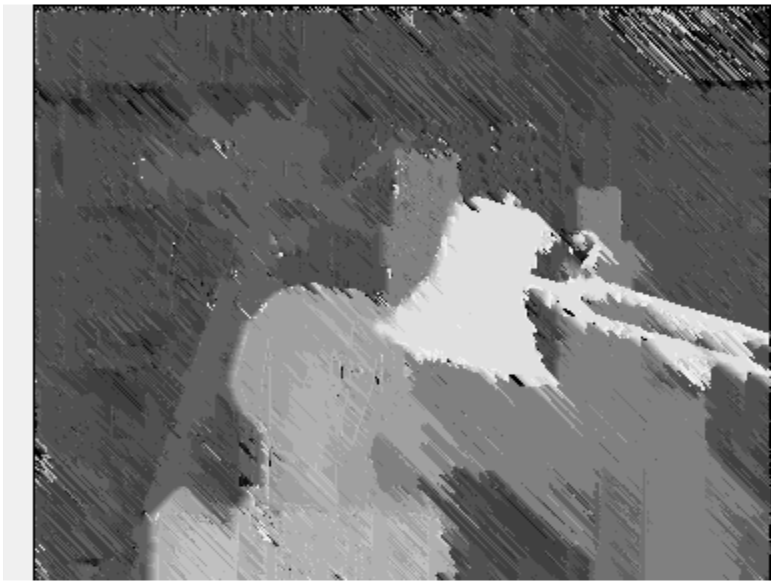
Scanline 3



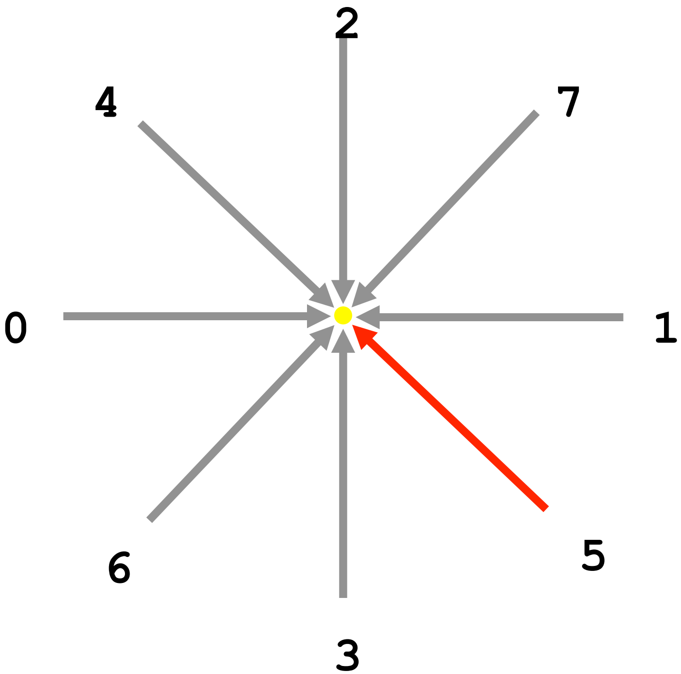


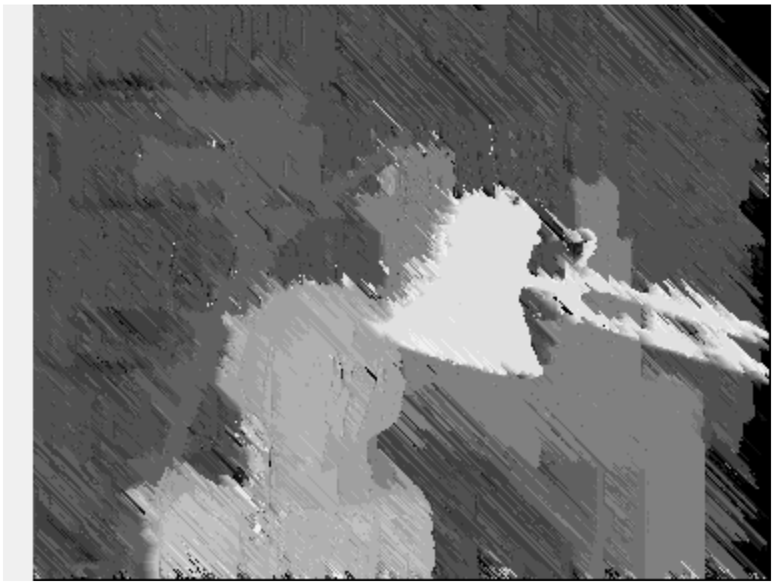
Scanline 4



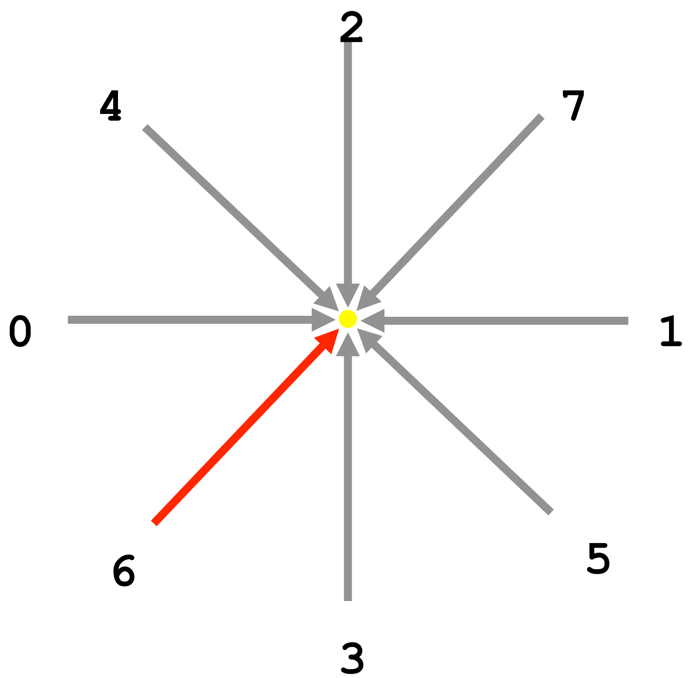


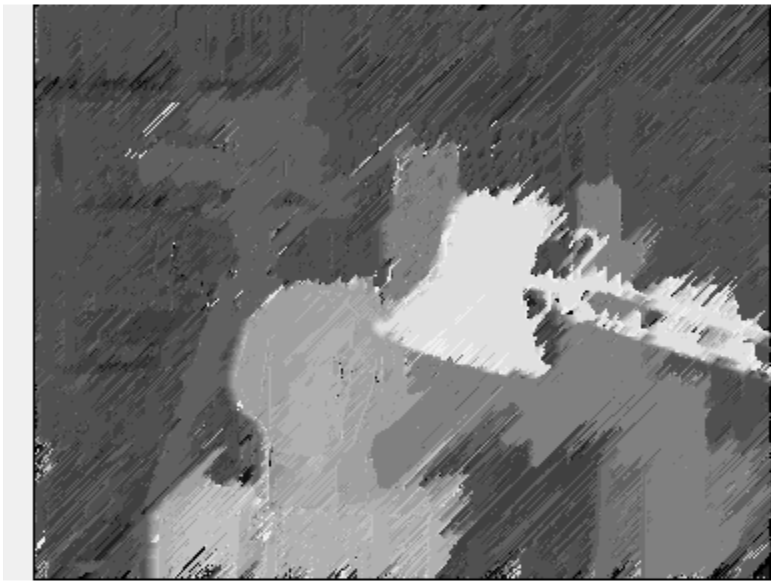
Scanline 5



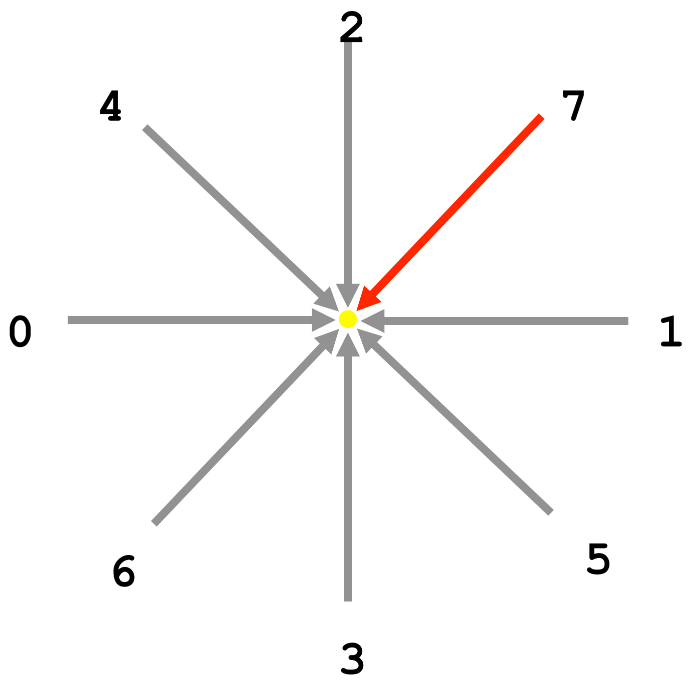


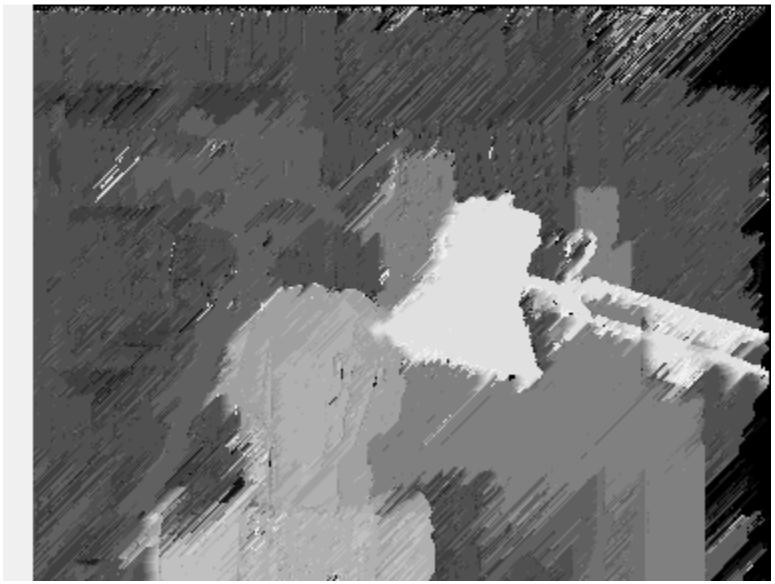
Scanline 6



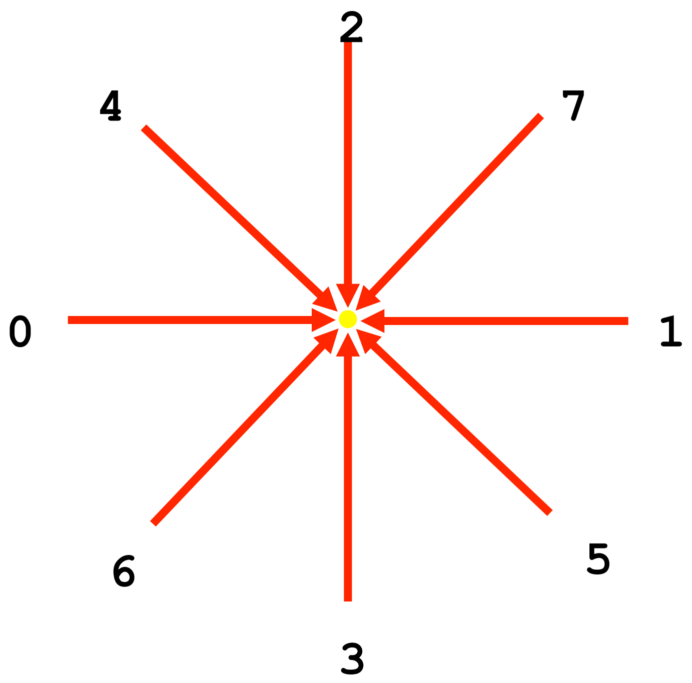


Scanline 7





Full SGM (8 scanlines, TAD color)



SO + support aggregation

This method combines an effective cost aggregation strategy with a SO based disparity computation framework.

- costs are computed by means of an effective strategy cost aggregation strategy (Segment Support)

- disparity computation relies on SO

- uses only 4 directions

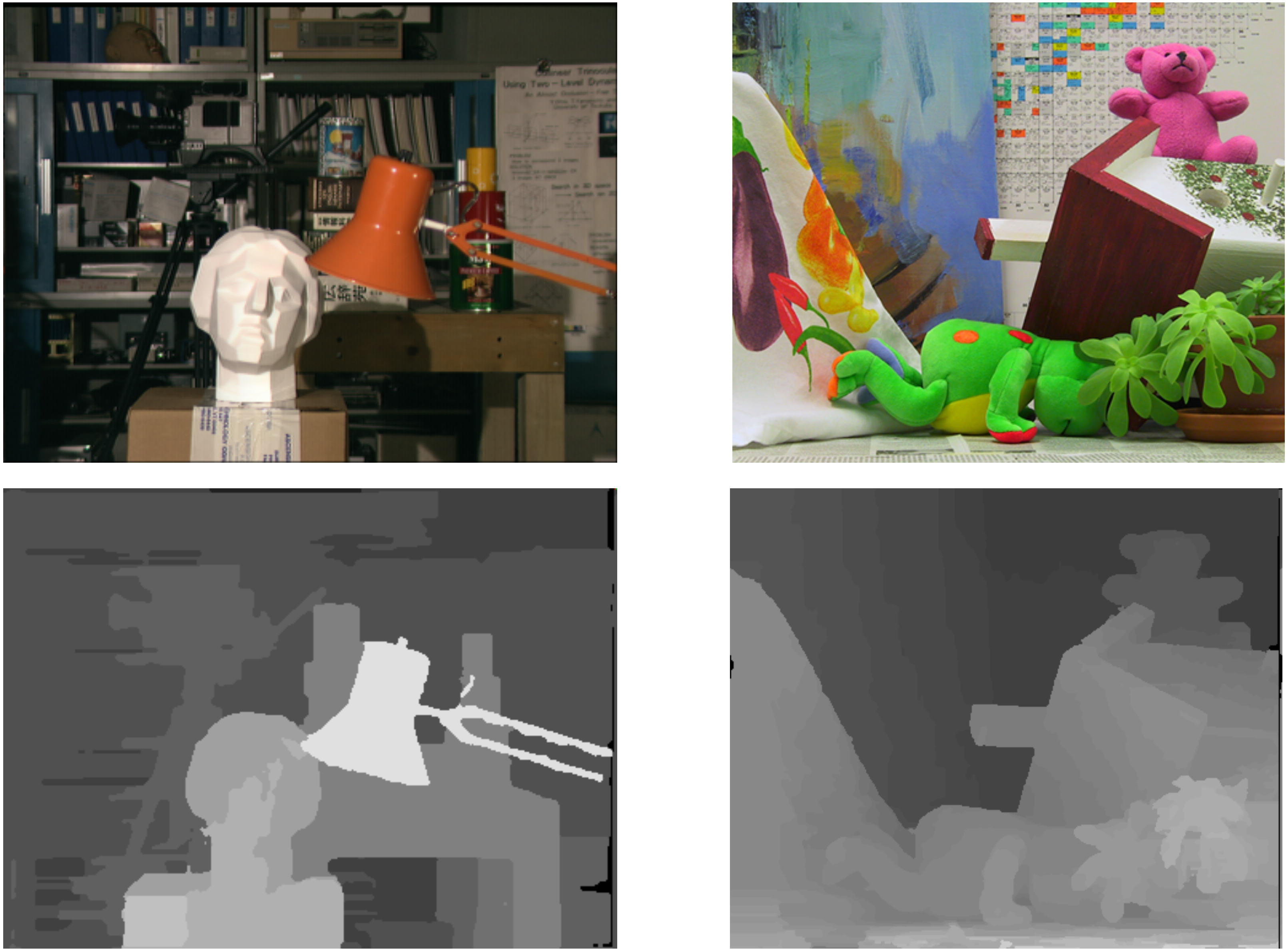
- excellent results

- very slow (due to cost aggregation strategy)

Using effective cost aggregation strategy within accurate disparity computation frameworks is an interesting trend successfully deployed also by other researchers [,].

S. Mattoccia, F. Tombari, and L. Di Stefano, Stereo vision enabling precise border localization within a scanline optimization framework, ACCV 2007

SO + support aggregation [29]



Stefano Mattoccia

Enforcing local consistency of disparity fields in fast SO/DP based algorithms [67]

This method aims at improving the accuracy of fast SO/DP based algorithms by enforcing the local consistency [66] of an initial disparity hypothesis.

- evaluated deploying the initial disparity hypotheses of C-Semiglobal [30] and RealTimeGPU [70]

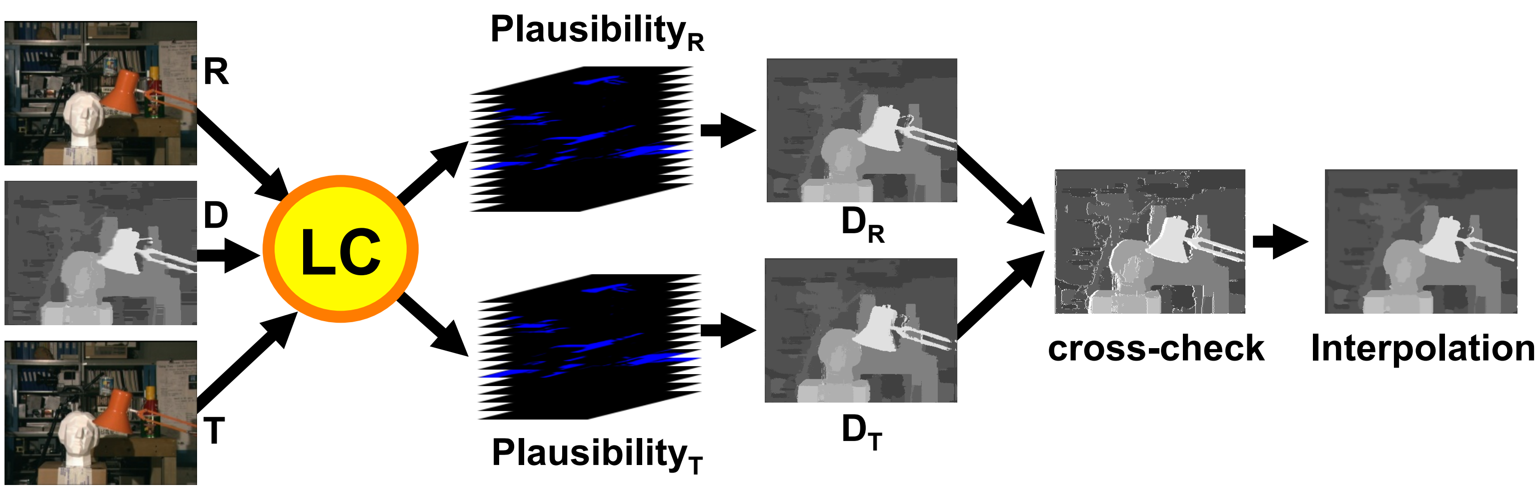
- dramatically improves the initial disparity field

- relatively fast, about 15 seconds on a standard PC with a single core

- computational optimizations/simplifications [68] enable us to obtain almost equivalent results in less than 2 seconds on a standard multicore PC (see next slides concerned with paper[68])

S. Mattoccia, Improving the accuracy of fast dense stereo correspondence algorithms by enforcing local consistency of disparity fields, 3DPVT2010

www.vision.deis.unibo.it/smatt//3DPVT2010.htm



This method:

- deploys the initial dense disparity hypotheses provided by a dense stereo algorithm (tested with fast and SO and DP algorithms [30] and [70])

- enforces local consistency by means of the LC technique [66] obtaining two *independent* disparity fields D_R and D_T

- detects and interpolates uncertain disparity assignments according to D_R and D_T

Experimental results deploying the initial disparity hypotheses of C-Semiglobal [30] available on the Middlebury web site

Error Threshold = 1

Error Threshold...

Sort by nonocc | Sort by all | Sort by disc

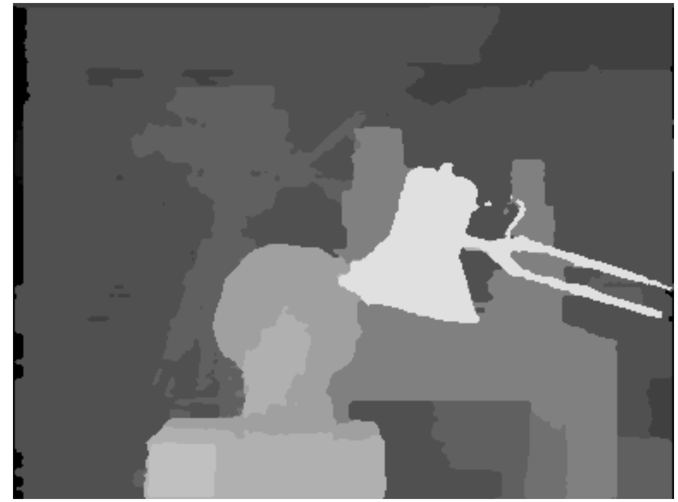
Algorithm	Avg. Rank	Tsukuba (ground truth) nonocc	Tsukuba all	Tsukuba disc	Venus (ground truth) nonocc	Venus all	Venus disc	Teddy (ground truth) nonocc	Teddy all	Teddy disc	Cones (ground truth) nonocc	Cones all	Cones disc	Average Percent Bad Pixels
CoopRegion [41]	4.8	0.87 1	1.16 1	4.61 1	0.11 2	0.21 2	1.54 4	5.16 11	8.31 8	13.0 8	2.79 6	7.18 4	8.01 10	4.41
AdaptingBP [17]	4.9	1.11 10	1.37 5	5.79 12	0.10 1	0.21 3	1.44 2	4.22 4	7.06 5	11.8 5	2.48 2	7.92 7	7.32 3	4.23
DoubleBP [35]	6.8	0.88 3	1.29 2	4.76 3	0.13 5	0.45 14	1.87 8	3.53 3	8.30 7	9.63 2	2.90 9	8.78 18	7.79 7	4.19
OutlierConf [42]	7.6	0.88 2	1.43 7	4.74 2	0.18 12	0.26 7	2.40 15	5.01 7	9.12 11	12.8 7	2.78 5	8.57 14	6.99 2	4.60
YOUR METHOD	9.3	1.03 8	1.54 8	5.56 8	0.14 7	0.28 9	1.95 11	5.44 12	11.0 15	13.6 11	2.87 7	8.35 11	7.47 5	4.93
SubPixDoubleBP [30]	10.5	1.24 17	1.76 19	5.98 13	0.12 4	0.46 15	1.74 7	3.45 2	8.38 9	10.0 3	2.93 11	8.73 17	7.91 9	4.39
SurfaceStereo [79]	11.2	1.28 22	1.65 13	6.78 24	0.19 14	0.28 8	2.61 21	3.12 1	5.10 1	8.65 1	2.89 8	7.95 8	8.26 14	4.06
WarpMat [55]	12.4	1.16 11	1.35 4	6.04 14	0.18 13	0.24 5	2.44 16	5.02 8	9.30 12	13.0 10	3.49 19	8.47 13	9.01 24	4.98
Undr+OvrSeg [48]	16.3	1.89 40	2.22 36	7.22 32	0.11 3	0.22 4	1.34 1	6.51 19	9.98 13	16.4 21	2.92 10	8.00 9	7.90 8	5.39
GC+SegmBorder [57]	17.6	1.47 31	1.82 21	7.86 37	0.19 15	0.31 10	2.44 16	4.25 5	5.55 2	10.9 4	4.99 50	5.78 1	8.66 19	4.52
AdaptOvrSegBP [33]	18.3	1.69 34	2.04 31	5.64 10	0.14 6	0.20 1	1.47 3	7.04 30	11.1 17	16.4 23	3.60 23	8.96 21	8.84 21	5.59
GeoSup [64]	19.8	1.45 30	1.83 23	7.71 36	0.14 8	0.26 6	1.90 9	6.88 27	13.2 32	16.1 18	2.94 12	8.89 20	8.32 16	5.80
PlaneFitBP [32]	20.1	0.97 7	1.83 22	5.26 7	0.17 11	0.51 17	1.71 6	6.65 22	12.1 26	14.7 12	4.17 38	10.7 39	10.6 34	5.78
SymBP+occ [7]	20.8	0.97 6	1.75 18	5.09 6	0.16 9	0.33 12	2.19 13	6.47 18	10.7 14	17.0 30	4.79 45	10.7 41	10.9 38	5.92
AdaptDispCalib [36]	22.7	1.19 14	1.42 6	6.15 16	0.23 18	0.34 13	2.50 19	7.80 36	13.6 36	17.3 36	3.62 24	9.33 26	9.72 28	6.10
Segm+visib [4]	23.1	1.30 24	1.57 9	6.92 30	0.79 42	1.06 37	6.76 47	5.00 6	6.54 3	12.3 6	3.72 26	8.62 16	10.2 31	5.40
C-SemiGlob [19]	23.1	2.61 53	3.29 44	9.89 49	0.25 21	0.57 19	3.24 26	5.14 10	11.8 20	13.0 8	2.77 4	8.35 12	8.20 11	5.76

-12

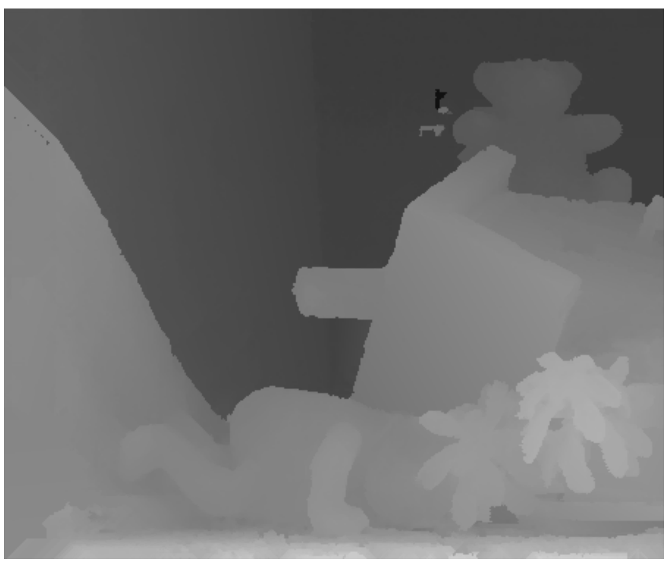
Experimental results according to the automatic evaluation procedure available at:
http://vision.middlebury.edu/stereo/

Stefano Mattoccia

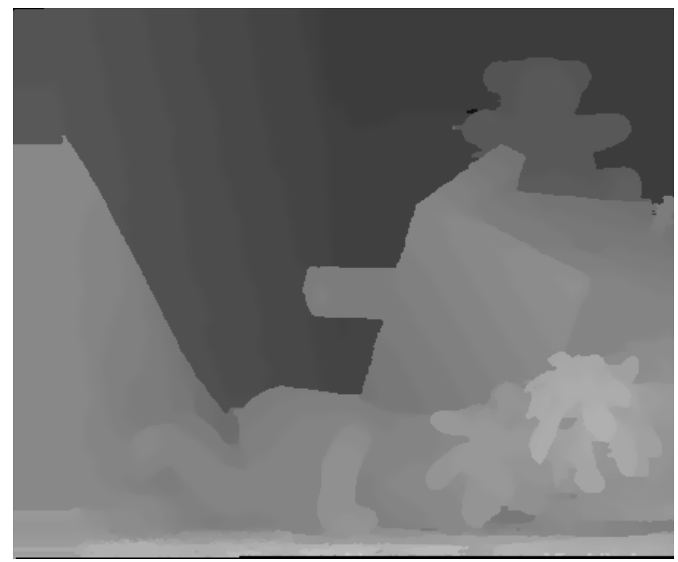
C-Semiglobal [30]



LC(C-Semiglobal)[67]



C-Semiglobal [30]



LC(C-Semiglobal) [67]

Stefano Mattoccia

Experimental results deploying the initial disparity hypotheses of RealTimeGPU [70] available on the Middlebury web site

	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc	
CoopRegion [41]	4.8	**0.87** 1	**1.16** 1	**4.61** 1	0.11 2	0.21 2	1.54 4	5.16 11	8.31 8	13.0 8	2.79 6	7.18 4	8.01 9	4.41
AdaptingBP [17]	4.9	1.11 10	1.37 5	5.79 12	**0.10** 1	0.21 3	1.44 2	4.22 4	7.06 5	11.8 5	2.48 2	7.92 7	7.32 3	4.23
DoubleBP [35]	6.4	0.88 3	1.29 2	4.76 3	0.13 5	0.45 13	1.87 8	3.53 3	8.30 7	9.63 2	2.90 8	8.78 17	7.79 6	4.19
OutlierConf [42]	7.3	0.88 2	1.43 7	4.74 2	0.18 11	0.26 7	2.40 14	5.01 7	9.12 11	12.8 7	2.78 5	8.57 13	6.99 2	4.60
SubPixDoubleBP [30]	10.2	1.24 17	1.76 19	5.98 13	0.12 4	0.46 14	1.74 7	3.45 2	8.38 9	10.0 3	2.93 10	8.73 16	7.91 8	4.39
SurfaceStereo [79]	10.9	1.28 22	1.65 13	6.78 24	0.19 13	0.28 8	2.61 20	**3.12** 1	**5.10** 1	**8.65** 1	2.89 7	7.95 8	8.26 13	4.06
WarpMat [55]	12.0	1.16 11	1.35 4	6.04 14	0.18 12	0.24 5	2.44 15	5.02 8	9.30 12	13.0 10	3.49 18	8.47 12	9.01 23	4.98
Undr+OvrSeg [48]	16.2	1.89 40	2.22 36	7.22 32	0.11 3	0.22 4	**1.34** 1	6.51 19	9.98 13	16.4 21	2.92 9	8.00 9	7.90 7	5.39
GC+SegmBorder [57]	17.2	1.47 31	1.82 21	7.86 37	0.19 14	0.31 9	2.44 15	4.25 5	5.55 2	10.9 4	4.99 50	**5.78** 1	8.66 18	4.52
AdaptOvrSegBP [33]	18.0	1.69 34	2.04 31	5.64 10	0.14 6	**0.20** 1	1.47 3	7.04 30	11.1 16	16.4 23	3.60 22	8.96 20	8.84 20	5.59
GeoSup [64]	19.4	1.45 30	1.83 23	7.71 36	0.14 7	0.26 6	1.90 9	6.88 27	13.2 32	16.1 18	2.94 11	8.89 19	8.32 15	5.80
PlaneFitBP [32]	20.0	0.97 8	1.83 22	5.26 8	0.17 10	0.51 16	1.71 6	6.65 22	12.1 26	14.7 12	4.17 37	10.7 39	10.6 34	5.78
SymBP+occ [7]	20.7	0.97 7	1.75 18	5.09 6	0.16 8	0.33 11	2.19 12	6.47 18	10.7 14	17.0 30	4.79 45	10.7 41	10.9 38	5.92
YOUR METHOD	21.2	0.96 6	1.63 11	5.19 7	0.32 23	0.64 22	3.23 25	6.29 16	12.1 25	14.2 11	4.20 39	10.1 36	10.5 33	5.78
AdaptDispCalib [36]	22.2	1.19 14	1.42 6	6.15 16	0.23 17	0.34 12	2.50 18	7.80 36	13.6 36	17.3 36	3.62 23	9.33 25	9.72 27	6.10
C-SemiGlob [19]	22.7	2.61 53	3.29 44	9.89 49	0.25 20	0.57 18	3.24 26	5.14 10	11.8 19	13.0 8	2.77 4	8.35 11	8.20 10	5.76
ConvexTV [46]	51.2	3.61 63	5.72 69	18.0 73	1.16 53	2.50 60	12.4 60	6.10 15	15.7 59	16.8 27	3.88 29	14.4 64	11.5 43	9.30
GenModel [20]	53.4	2.57 52	4.74 62	13.0 60	1.72 62	3.08 62	16.9 67	6.86 26	15.0 51	19.2 48	4.64 44	14.9 66	11.4 41	9.50
RTCensus [50]	54.8	5.08 78	6.25 75	19.2 76	1.58 60	2.42 58	14.2 64	7.96 40	13.8 39	20.3 57	4.10 35	9.54 27	12.2 49	9.73
TensorVoting [9]	54.9	3.79 65	4.79 63	8.86 41	1.23 54	1.88 52	11.5 56	9.76 57	17.0 64	24.0 69	4.38 41	11.4 47	12.2 50	9.25
RealTimeGPU [14]	55.8	2.05 46	4.22 57	10.6 54	1.92 65	2.98 61	20.3 72	7.23 33	14.4 49	17.6 37	6.41 65	13.7 62	16.5 69	9.82
ReliabilityDP [13]	58.3	1.36 25	3.39 45	7.25 33	2.35 67	3.48 69	12.2 59	9.82 59	16.9 63	19.5 50	12.9 80	19.9 79	19.7 71	10.7
CostRelax [11]	58.6	4.76 75	6.08 74	20.3 79	1.41 58	2.48 59	18.5 69	8.18 46	15.9 61	23.8 67	3.91 32	10.2 37	11.8 46	10.6
TreeDP [8]	62.0	1.99 44	2.84 42	9.96 50	1.41 57	2.10 55	7.74 50	15.9 76	23.9 76	27.1 75	10.0 75	18.3 74	18.9 70	11.7

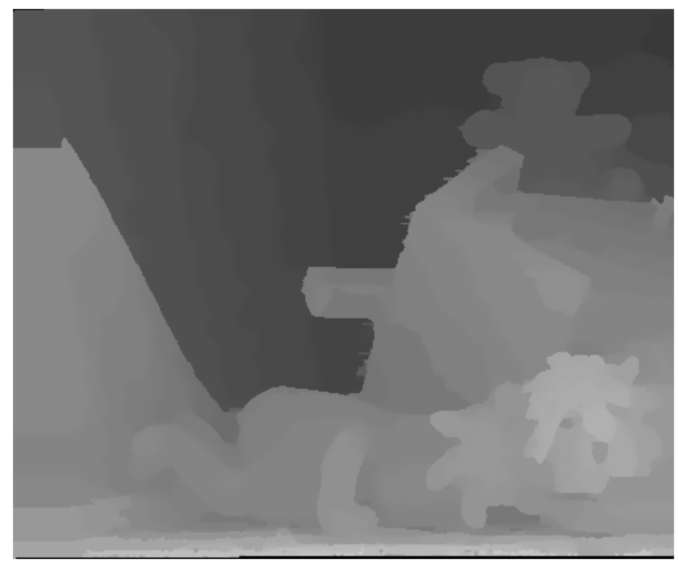
-49

Experimental results according to the automatic evaluation procedure available at:
http://vision.middlebury.edu/stereo/

Stefano Mattoccia

RealTimeGPU [70]

LC(RealTimeGPU) [67]



RealTimeGPU [70]

LC(RealTimeGPU) [67]

Fast dense stereo on multicore deploying a relaxed local consistency constraint [68]

The execution time of previously described method [67], can be dramatically reduced according to the methodologies proposed in [68].

Deploying the same initial disparity hypotheses (that is, C-Semiglobal and RealTimeGPU), this method enables us to obtain almost equivalent results (see [67] in previous page) in less than 2 seconds on a Core2 Quad CPU @ 2.49 GHz.

This methods:

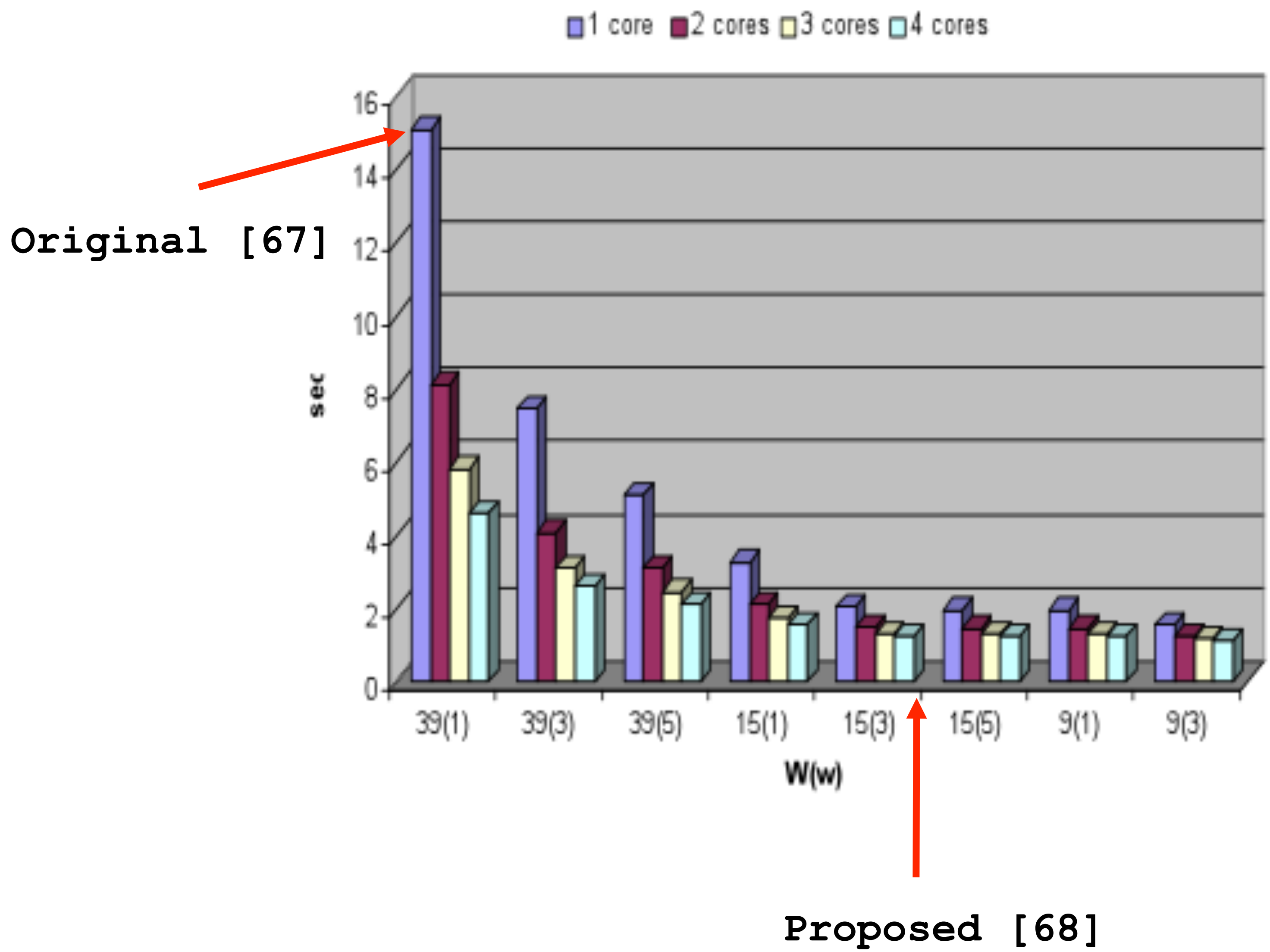
- relies on a relaxed local consistency constraint

- takes advantage of coarse-grained thread-level paralellism

S. Mattoccia, Fast locally consistent dense stereo on multicore, Sixth IEEE Embedded Computer Vision Workshop (ECVW2010), CVPR workshop, June 13, 2010, San Francisco, USA

www.vision.deis.unibo.it/smatt/RLC stereo.htm

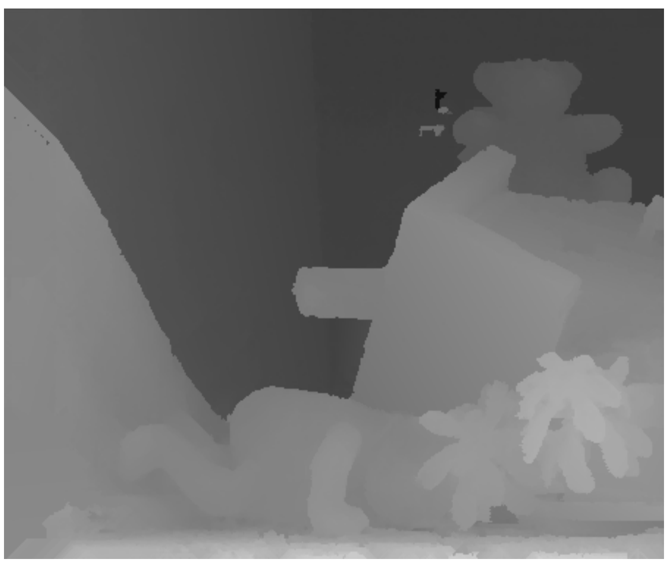
Measured speed-ups on a Core2 Quad CPU @ 2.49 GHz



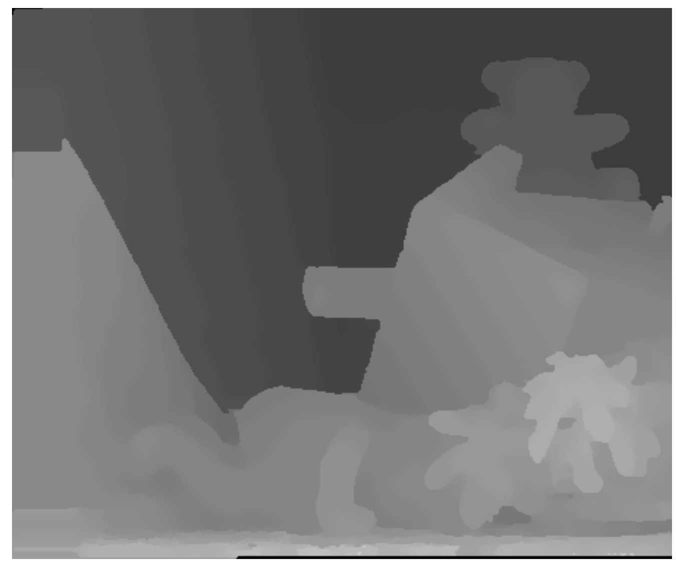
Measurements concerned with the Teddy stereo pair

C-Semiglobal [30]

RLC(C-Semiglobal)[68]



C-Semiglobal [30]



RLC(C-Semiglobal) [68]

Constraining local consistency on superpixels [69]

The effectiveness of the locally consistent technique [66] can be further improved by constraining its behavior on superpixels obtained by means of segmentation [50].

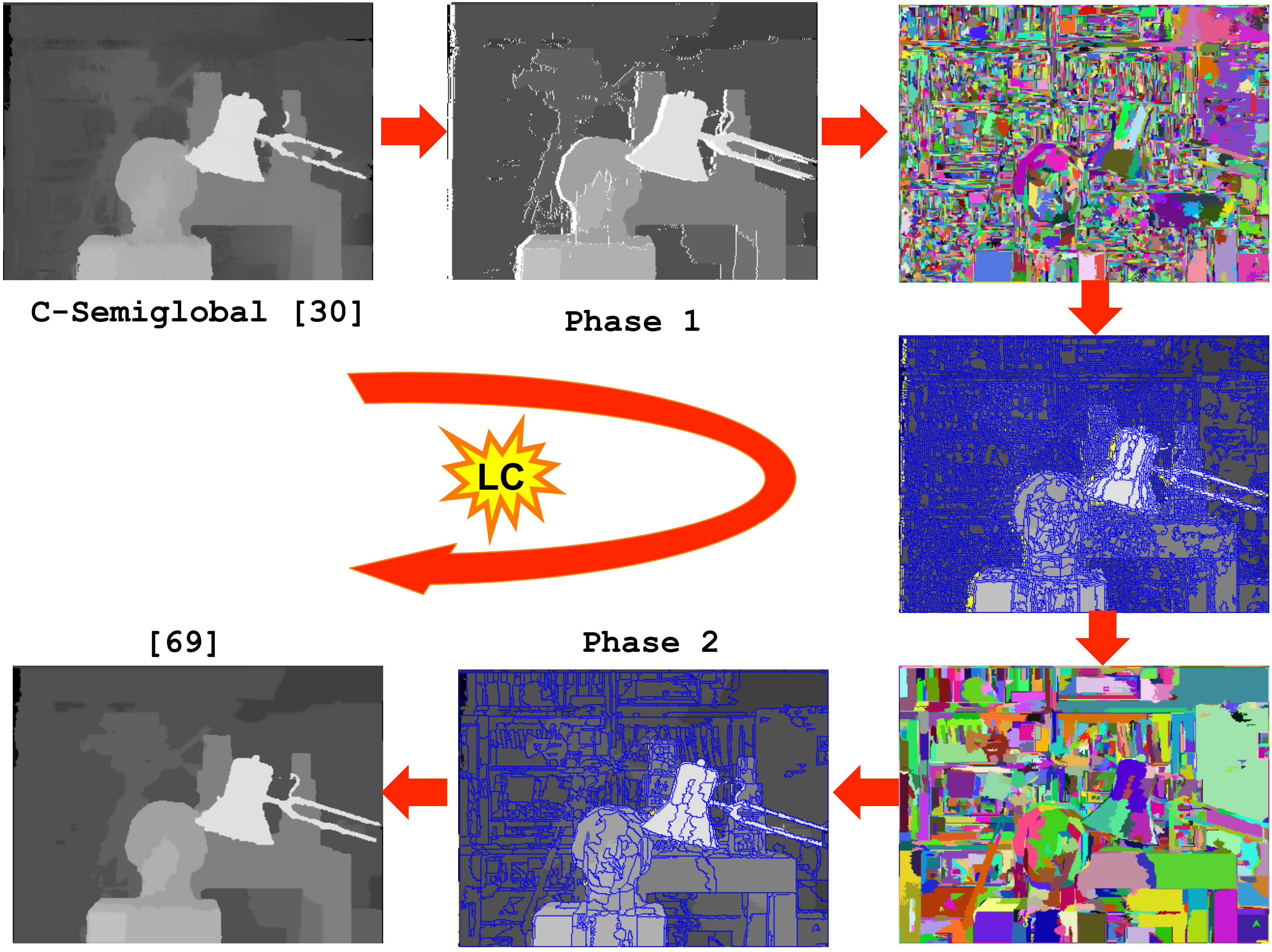
This method deploys a two stage strategy to constraint Local Consistency [66] on superpixels.
During the first phase, we over-segment the reference image:

- to detect uncertain disparity measurements
- to regularize disparity within superpixels

During the second phase we relax the segmentation constraint in order to propagate the regularized disparity assumptions.

As for previous approaches, we start with an initial disparity hypothesis (C-Semiglobal algorithms [30] available on [15])

S. Mattoccia, Accurate dense stereo by constraining local consistency on superpixels, 20th International Conference on Pattern Recognition (ICPR2010), August 23-26, 2010, Istanbul, Turkey



C-Semiglobal [30]
Phase 1
LC
[69]
Phase 2

Experimental results for [69] deploying the initial disparity hypotheses of C-Semiglobal [30] available on the Middlebury evaluation site

Error Threshold = 1 | Error Threshold... | Sort by nonocc | Sort by all | Sort by disc

Algorithm	Avg. Rank	Tsukuba (ground truth) nonocc	all	disc	Venus (ground truth) nonocc	all	disc	Teddy (ground truth) nonocc	all	disc	Cones (ground truth) nonocc	all	disc	Average Percent Bad Pixels
CoopRegion [41]	5.2	0.87 2	**1.16** 1	**4.61** 1	0.11 3	0.21 2	1.54 5	5.16 11	8.31 8	13.0 8	2.79 7	7.18 4	8.01 10	4.41
AdaptingBP [17]	5.3	1.11 10	1.37 6	5.79 12	0.10 2	0.21 3	1.44 3	4.22 4	7.06 5	11.8 5	2.48 3	7.92 7	7.32 4	4.23
YOUR METHOD	5.8	0.87 1	1.31 3	4.69 2	**0.09** 1	0.29 9	**1.29** 1	5.44 12	11.0 15	13.6 11	2.48 2	8.16 10	6.97 2	4.69
DoubleBP [35]	7.1	0.88 4	1.29 2	4.76 4	0.13 6	0.45 14	1.87 9	3.53 3	8.30 7	9.63 2	2.90 9	8.78 18	7.79 7	4.19
OutlierConf [42]	8.0	0.88 3	1.43 8	4.74 3	0.18 12	0.26 7	2.40 15	5.01 7	9.12 11	12.8 7	2.78 6	8.57 14	6.99 3	4.60
SubPixDoubleBP [30]	10.7	1.24 17	1.76 19	5.98 13	0.12 5	0.46 15	1.74 8	3.45 2	8.38 9	10.0 3	2.93 11	8.73 17	7.91 9	4.39
SurfaceStereo [79]	11.2	1.28 22	1.65 13	6.78 24	0.19 14	0.28 8	2.61 21	**3.12** 1	**5.10** 1	**8.65** 1	2.89 8	7.95 8	8.26 14	4.06
WarpMat [55]	12.5	1.16 11	1.35 5	6.04 14	0.18 13	0.24 5	2.44 16	5.02 8	9.30 12	13.0 10	3.49 19	8.47 13	9.01 24	4.98
Undr+OvrSeg [48]	16.5	1.89 40	2.22 36	7.22 32	0.11 4	0.22 4	1.34 2	6.51 19	9.98 13	16.4 21	2.92 10	8.00 9	7.90 8	5.39
GC+SegmBorder [57]	17.6	1.47 31	1.82 21	7.86 37	0.19 15	0.31 10	2.44 16	4.25 5	5.55 2	10.9 4	4.99 50	**5.78** 1	8.66 19	4.52
AdaptOvrSegBP [33]	18.5	1.69 34	2.04 31	5.64 10	0.14 7	**0.20** 1	1.47 4	7.04 30	11.1 17	16.4 23	3.60 23	8.96 21	8.84 21	5.59
GeoSup [64]	19.8	1.45 30	1.83 23	7.71 36	0.14 8	0.26 6	1.90 10	6.88 27	13.2 32	16.1 18	2.94 12	8.89 20	8.32 16	5.80
PlaneFitBP [32]	20.3	0.97 8	1.83 22	5.26 8	0.17 11	0.51 17	1.71 7	6.65 22	12.1 26	14.7 12	4.17 38	10.7 39	10.6 34	5.78
SymBP+occ [7]	21.0	0.97 7	1.75 18	5.09 7	0.16 9	0.33 12	2.19 13	6.47 18	10.7 14	17.0 30	4.79 45	10.7 41	10.9 38	5.92
AdaptDispCalib [36]	22.8	1.19 14	1.42 7	6.15 16	0.23 18	0.34 13	2.50 19	7.80 36	13.6 36	17.3 36	3.62 24	9.33 26	9.72 28	6.10
Segm+visib [4]	23.1	1.30 24	1.57 9	6.92 30	0.79 42	1.06 37	6.76 47	5.00 6	6.54 3	12.3 6	3.72 26	8.62 16	10.2 31	5.40
C-SemiGlob [19]	23.2	2.61 53	3.29 44	9.89 49	0.25 21	0.57 19	3.24 26	5.14 10	11.8 20	13.0 8	2.77 5	8.35 12	8.20 11	5.76

-14

Experimental results according to the automatic evaluation procedure available at:
http://vision.middlebury.edu/stereo/

Disparity refinement (4)

- Raw disparity maps computed by correspondence algorithms contain outliers that must be identified and corrected

- Moreover, since the disparity maps are typically computed at discrete pixel level more accurate disparity assignments would b desirable

- Several approaches aimed at improving the raw disparity maps computed by stereo correspondence algorithms have been proposed

- In the next slides is provided a description of some (not mutually exclusive) relevant approaches

Disparity refinement (4)

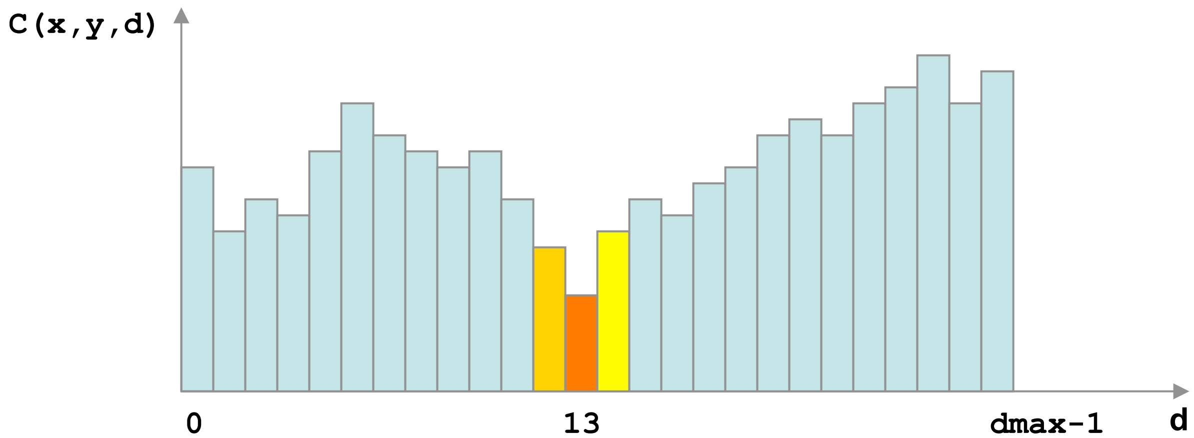
- Raw disparity maps computed by correspondence algorithms contain outliers that must be identified and corrected

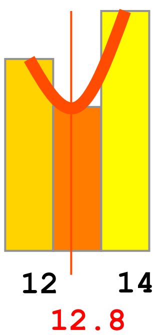
- Moreover, since the disparity maps are typically computed at discrete pixel level more accurate disparity assignments would b desirable

- Several approaches aimed at improving the raw disparity maps computed by stereo correspondence algorithms have been proposed

- A description of some (not mutually exclusive) relevant approaches is provided in the next slides

Sub-pixel interpolation





- (Typically) sub-pixel disparity is obtained interpolating the three matching costs with a second degree function (parabola)
- Computationally inexpensive and reasonably accurate
- In [55] proposed a floating-point free approach
- More accurate (and computational expensive) approaches perform directly matching cost computation on sub-pixel basis

L. Di Stefano, S. Mattoccia, Real-time stereo within the VIDET project Real-Time Imaging, 8(5), pp. 439-453, Oct. 2002

Image filtering techniques

Sometime the disparity maps are simply refined by means of image filtering techniques without (explicitly) enforcing any constraint about the underlining disparity maps.

Common image filtering operators are:

- **Median filtering**
- **Morphological operators**
- **Bilateral filtering [51]**

Bidirectional Matching*

Bidirectional matching (BM) is a widely used technique for detecting outliers [56] in stereo (local and global).

The correspondence problem is solved two times

- assuming left image as reference ($d_{LR}(x,y)$)
- assuming right image as reference ($d_{RL}(x,y)$)

and the disparity values that are not consistent between the two maps are classified as outliers enforcing

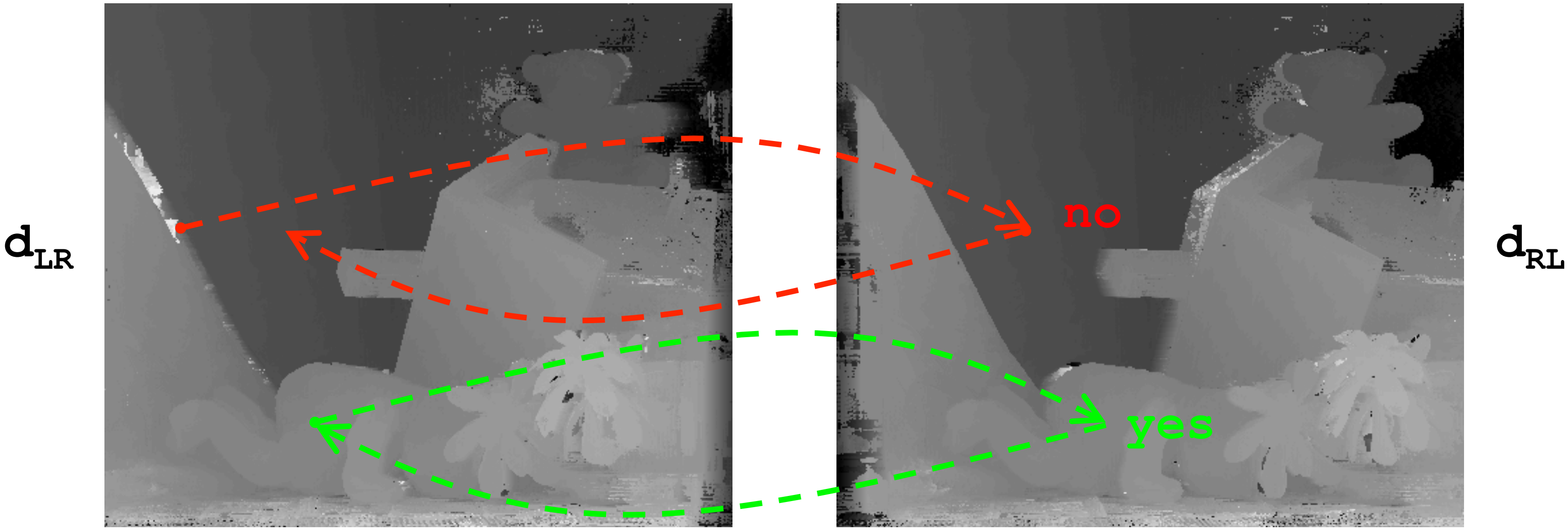
$$|d_{LR}(x,y) - d_{RL}(x+d_{LR}(x,y),y)| < T$$

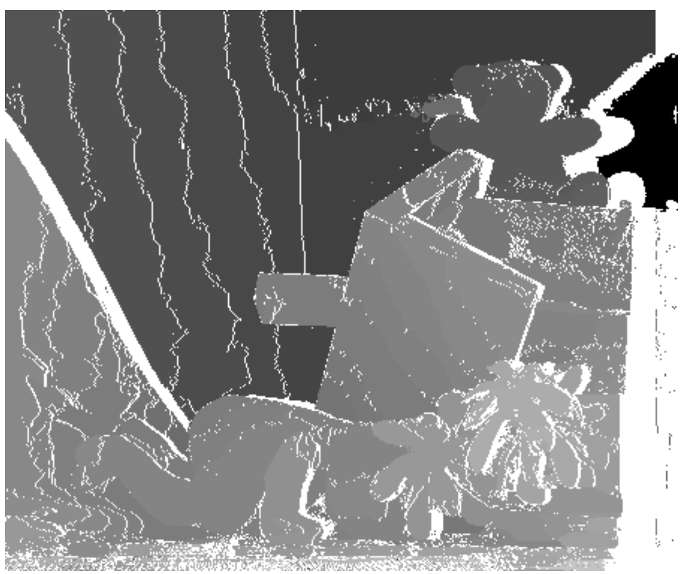
with threshold T typically set to 1

* aka Left-Right (consistency) check

P. Fua, Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities 12th. Int. Joint Conf. on Artificial Intelligence, pp 1292–1298, 1993

$$|d_{LR}(x,y) - d_{RL}(x+d_{LR}(x,y),y)| < T \ ?$$





Outliers detected by BM are encoded in white

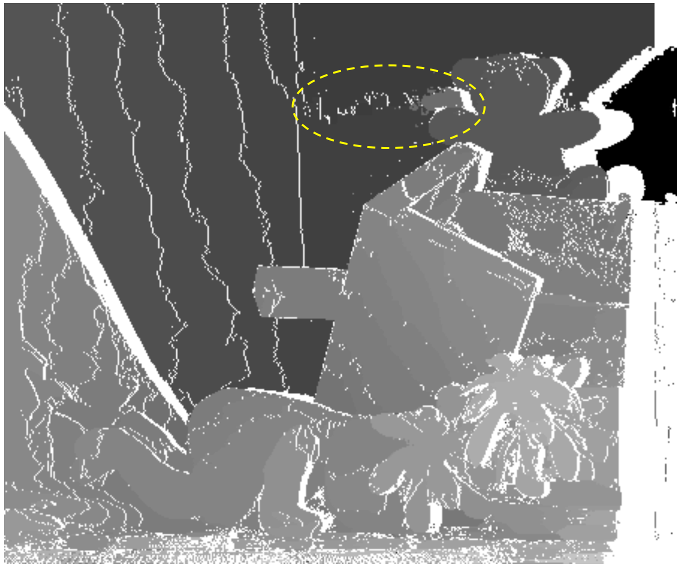
-useful for detecting occlusions

- preserves depth discontinuities

- (partially) effective for detecting outliers in ambiguous regions (see figure)

- two matching phases

- implicitly enforces the uniqueness constraint



Single Matching Phase (SMP) - Uniqueness+

The Single Matching Phase (SMP) approach [48] aims at detecting unreliable disparity assignments using a more computationally efficient technique.

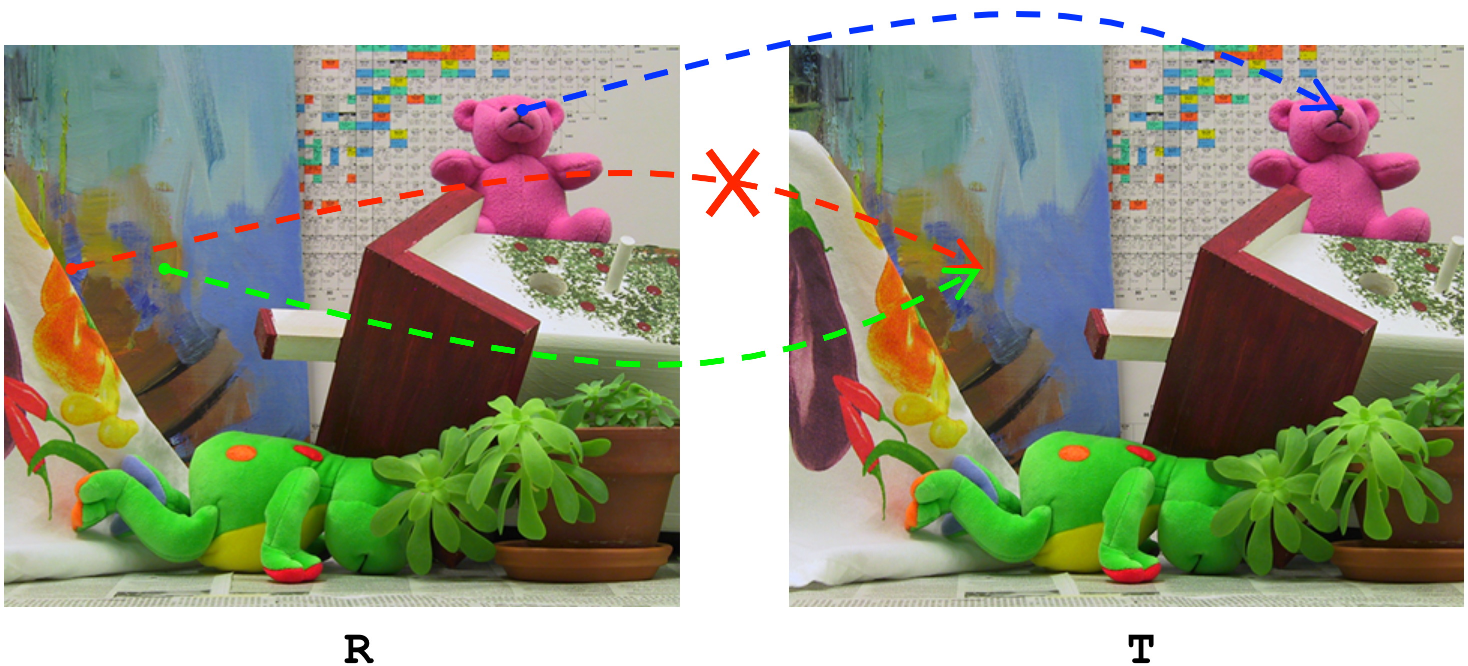
- **uses a single matching phase (1/2 vs BM)**
- **explicitly enforces the uniqueness constraint***
- **dynamically updates the disparity map when the uniqueness constraint is violated**
- **strengthened by additional constraints (next slides)**
- **effectiveness comparable to BM []**
- **suitable for efficient SIMD implementation**

*** Sometime violated (e.g. foreshortening)**

The correspondences are dynamically evaluated and corrected within a single matching phase ($d_{RT}(x,y)$).

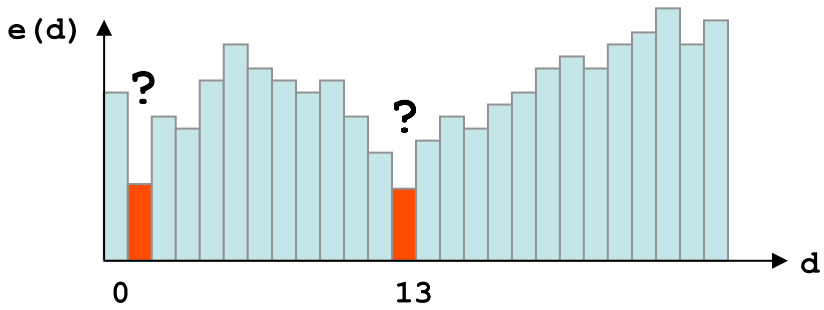
When two correspondences fall in the same point of the target image:

- the correspondence with the best score is kept
- the other correspondence is discarded



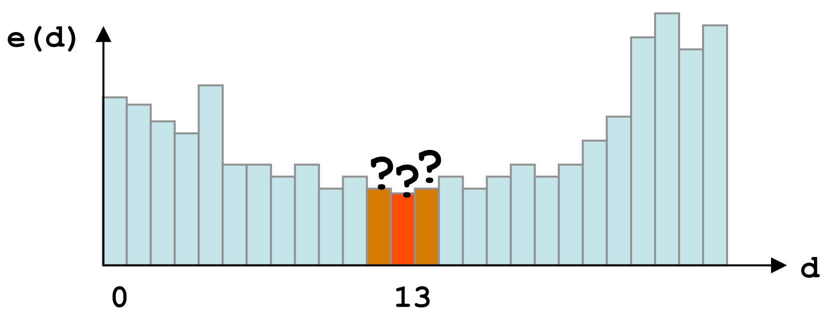
The basic SMP approach can be strengthened by means of two additional constraints:

a) Distinctiveness

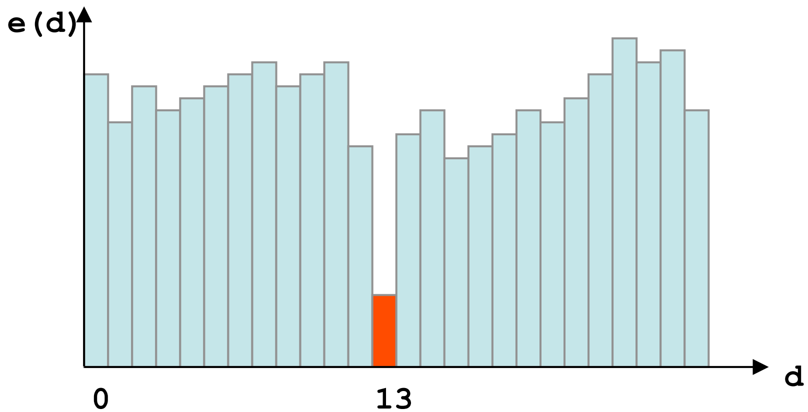


Example:
repetitive pattern

b) Sharpness

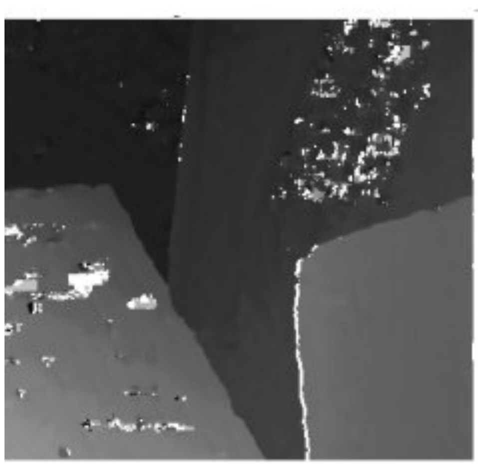


Example:
uniform region

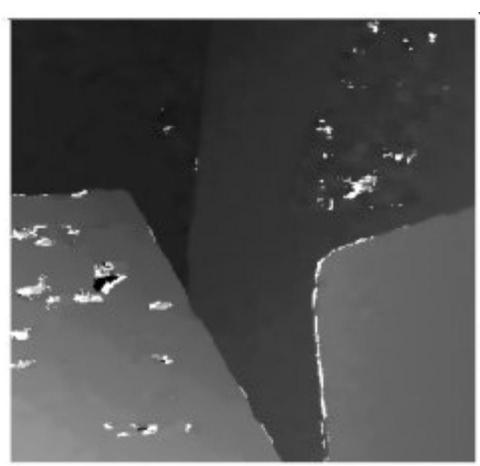


Example of reliable correspondence

An exhaustive comparison between DM and SMP on stereo pairs with groundtruth can be found in [48].



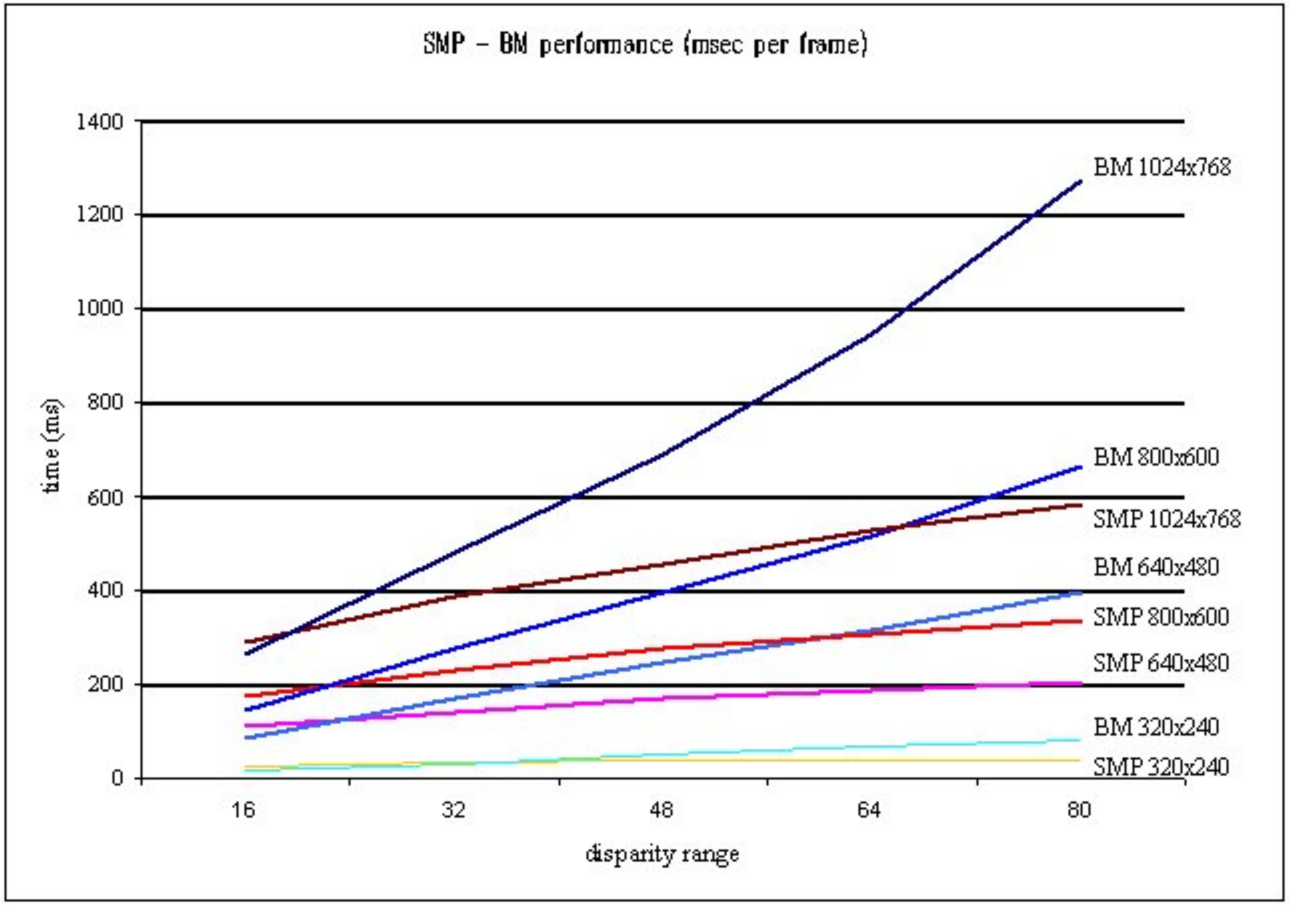
SMP



BM

Outliers are encoded in white

Performance evaluation [48]: SMP vs BM (PIII 800 MHz)



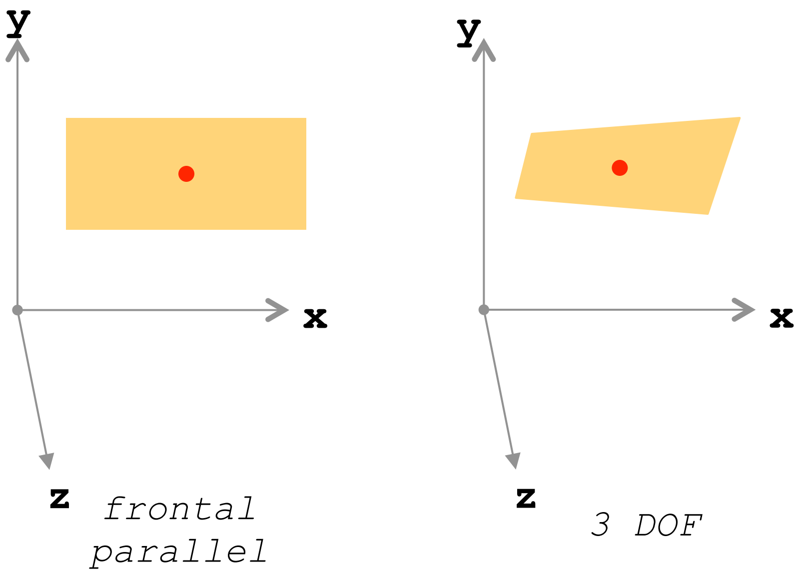
Segmentation based outliers identification and replacement

Two fundamental assumptions:

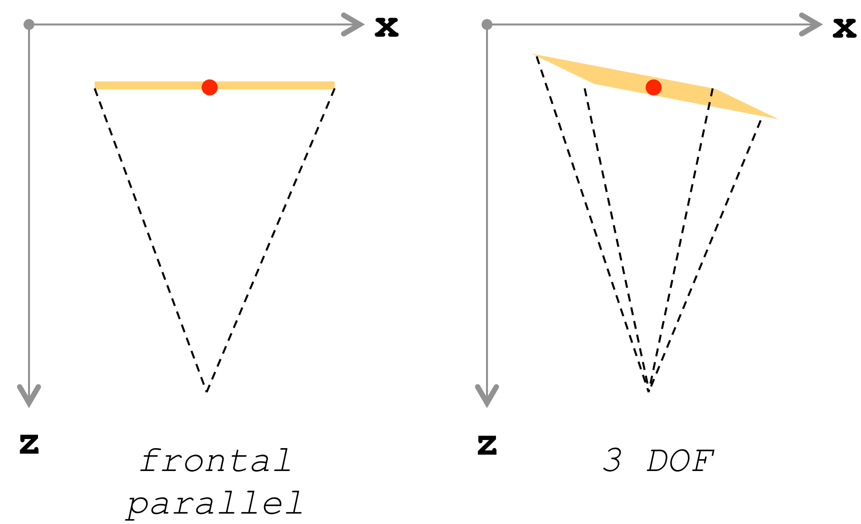
1) disparity within each segment varies smoothly

2) each segment can be approximated with a plane

Sometime 2) is not verified (below) ⇒ over-segmentation



3D view



Top view

Stefano Mattoccia

Each segment is modelled with a plane in the 3D space (3 DOF):

$$d(x,y) = \alpha \cdot x + \beta \cdot y + \gamma$$

Robust plane fitting of disparity measurements:

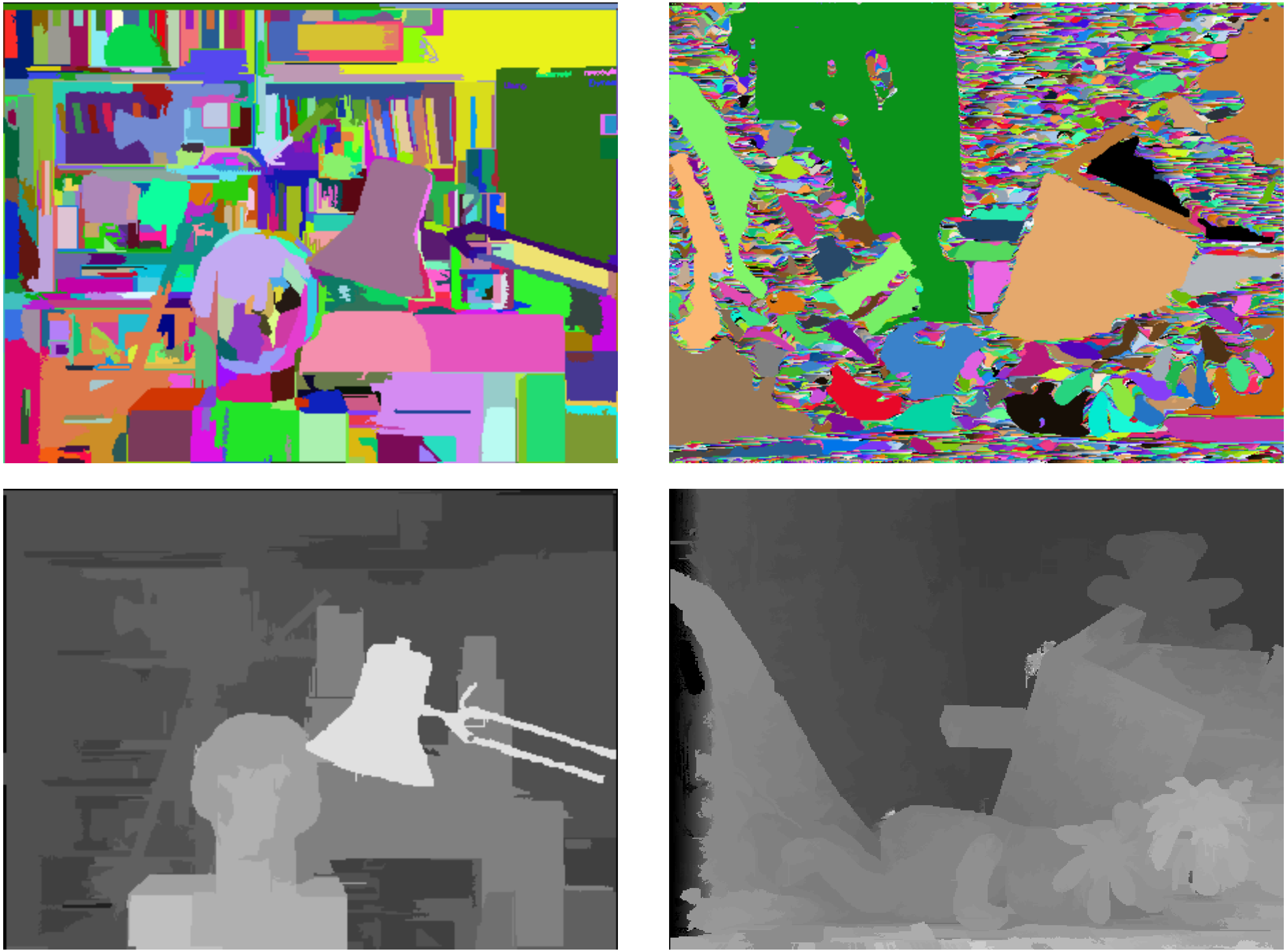
- **RANSAC [25] (iterative)**
- **Histogram Voting [54] (non iterative)**

The best performing algorithms on the Middlebury dataset cast robust plane fitting within a global energy minimization framework.

The next slide shows robust plane fitting of disparity measurements computed by a local approach (WTA + BM + Histogram Voting).

Interesting research activity: replacing planes with more complex surfaces

Example of robust plane fitting

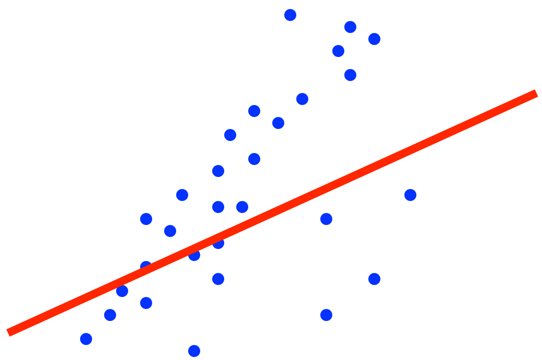


Local approach (FBS) + WTA + BM + robust plane fitting

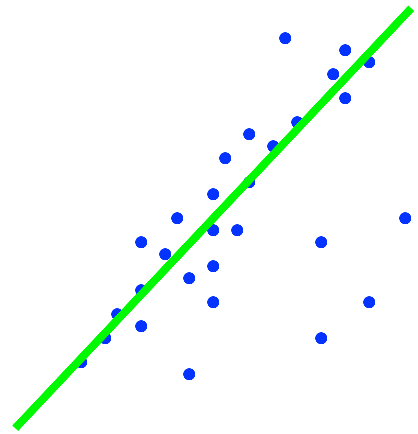
Stefano Mattoccia

Robust interpolation of noisy measurements

- **Disparity maps always contain outliers**
- **Reliable fitting with planes requires interpolation techniques robust to outliers**



Traditional approach (Least Square (LS))



Robust interpolation

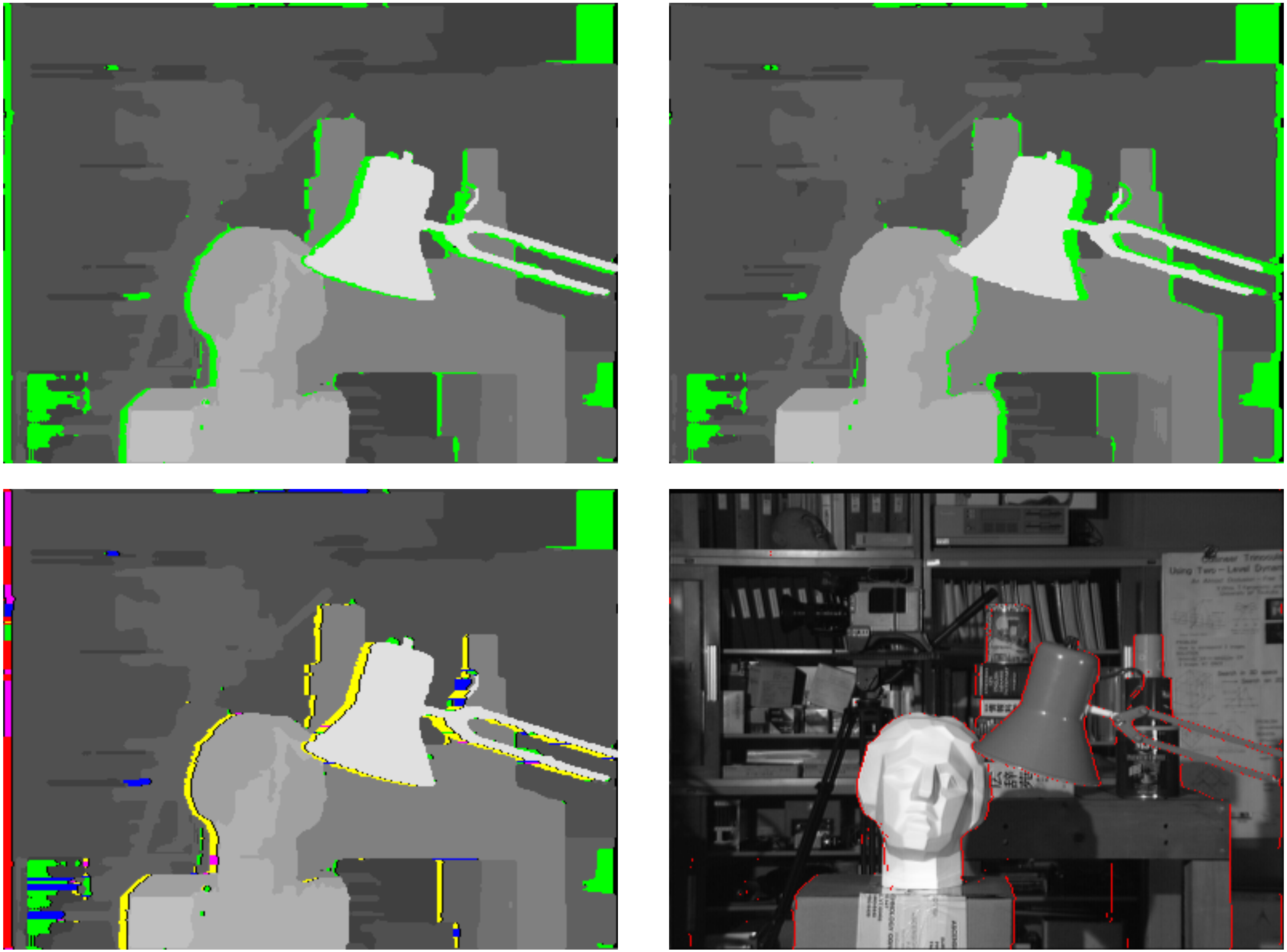
RANSAC and Histogram Voting are two techniques used in stereo for robust interpolation of noisy disparity measurements

Accurate localization of borders and occlusions

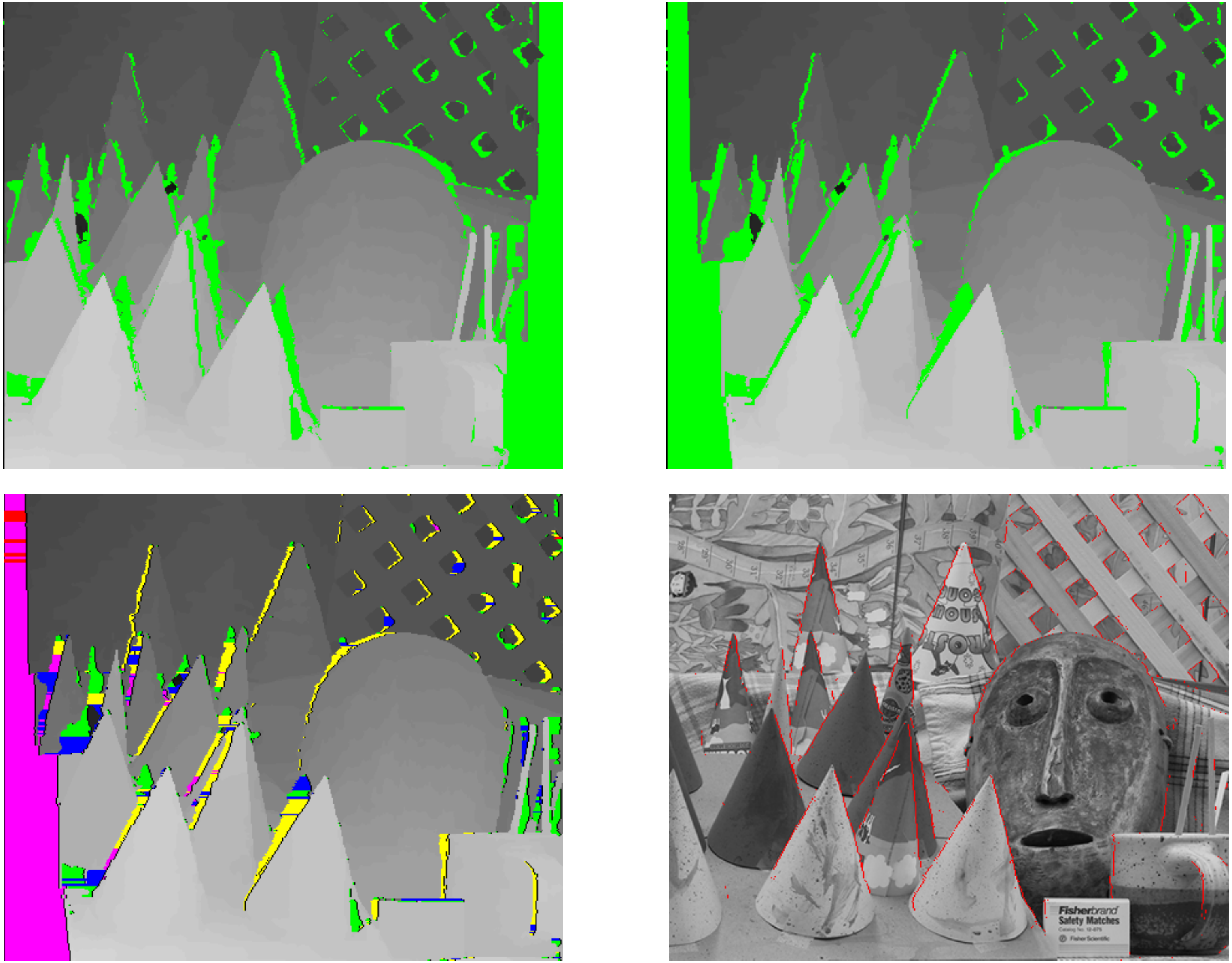
In [29] was proposed a method for accurate detection of depth borders and occlusions.

- **This method uses the disparity maps (d_{LR} and d_{RL}) computed by a (local or global) stereo correspondence algorithm**
- **Borders and occlusions are detected (without global energy minimization frameworks) enforcing, along scanlines, constraints between occlusions (in one image) and discontinuities (in the other image)**
- **Accurate results (see the next slides)**
- **Evaluated with the disparity maps provided by the algorithm described in [29] (SO + SegmentSupport)**

[29] S. Mattoccia, F. Tombari, and L. Di Stefano, Stereo vision enabling precise border localization within a scanline optimization framework, ACCV 2007



Occlusions (yellow)   Borders (red)



Occlusions (yellow) **Borders (red)**

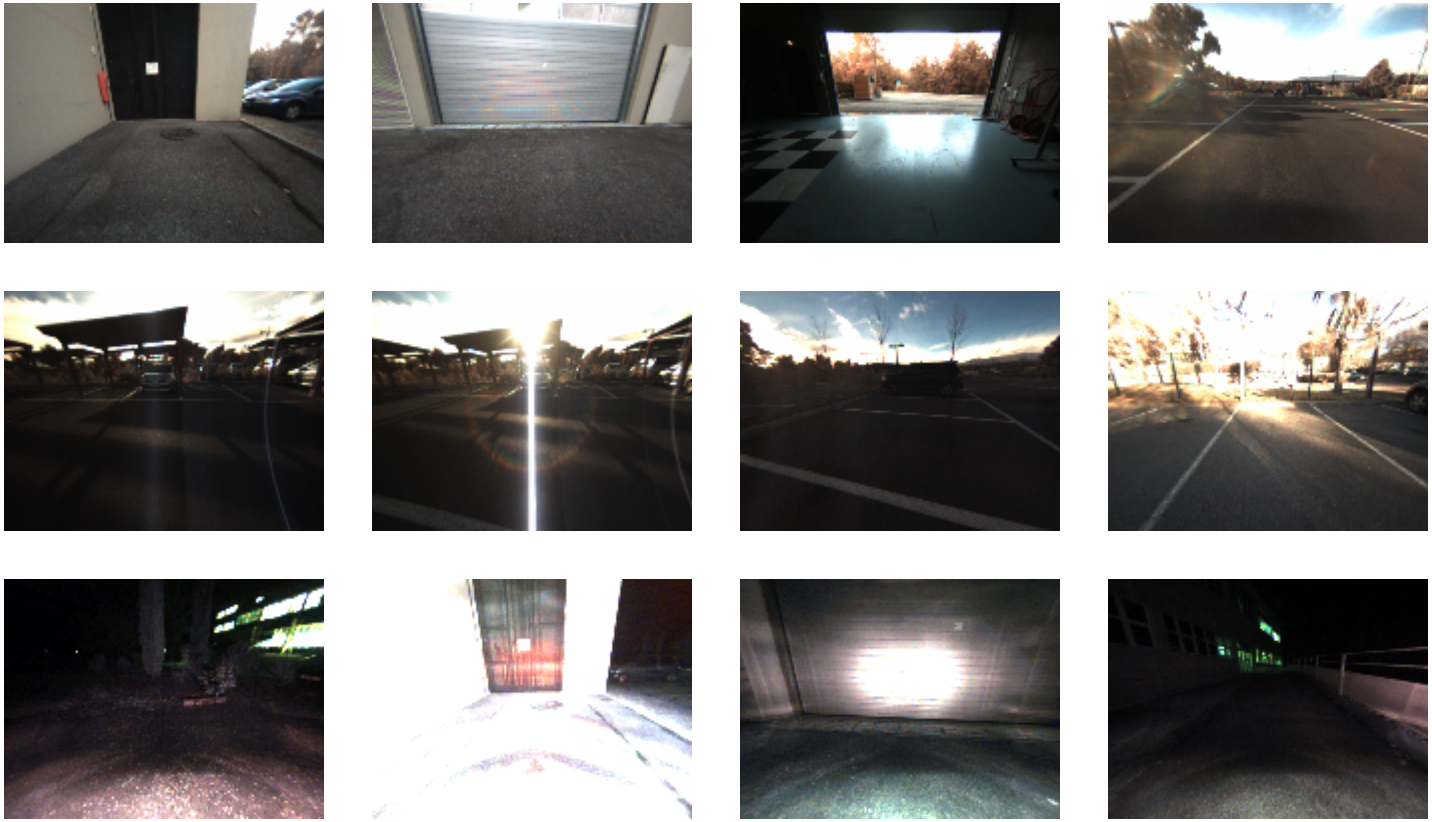
Iterative approaches

L. De-Maeztu, S. Mattoccia, A. Villanueva, R. Cabeza, "Efficient aggregation via iterative block-based adapting support weight", IC3D 2011

Computational Optimizations

Hardware implementation

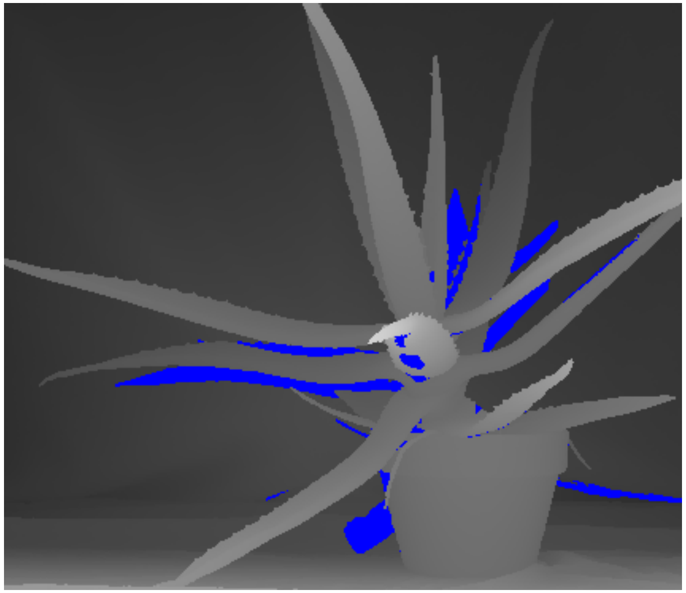
Open problem: radiometric variations



Courtesy of IMRA Europe, Sophia Antipolis (FR)

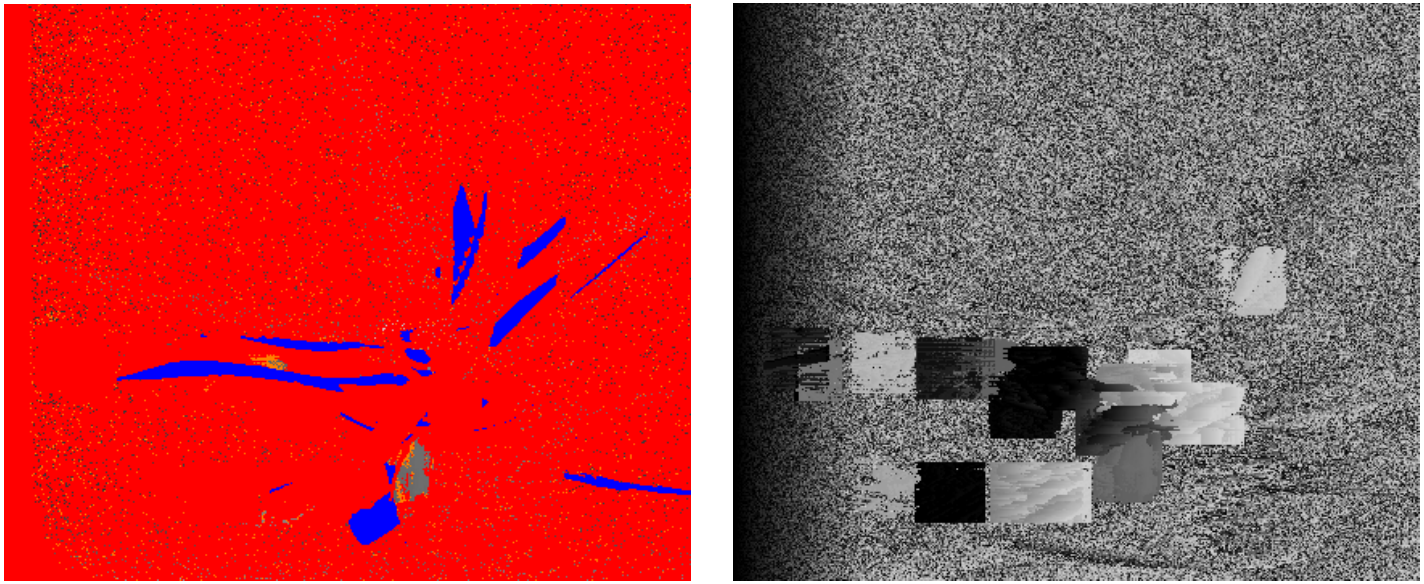
Left ILL(1)-EXP(0)

Right ILL(3)-EXP(2)

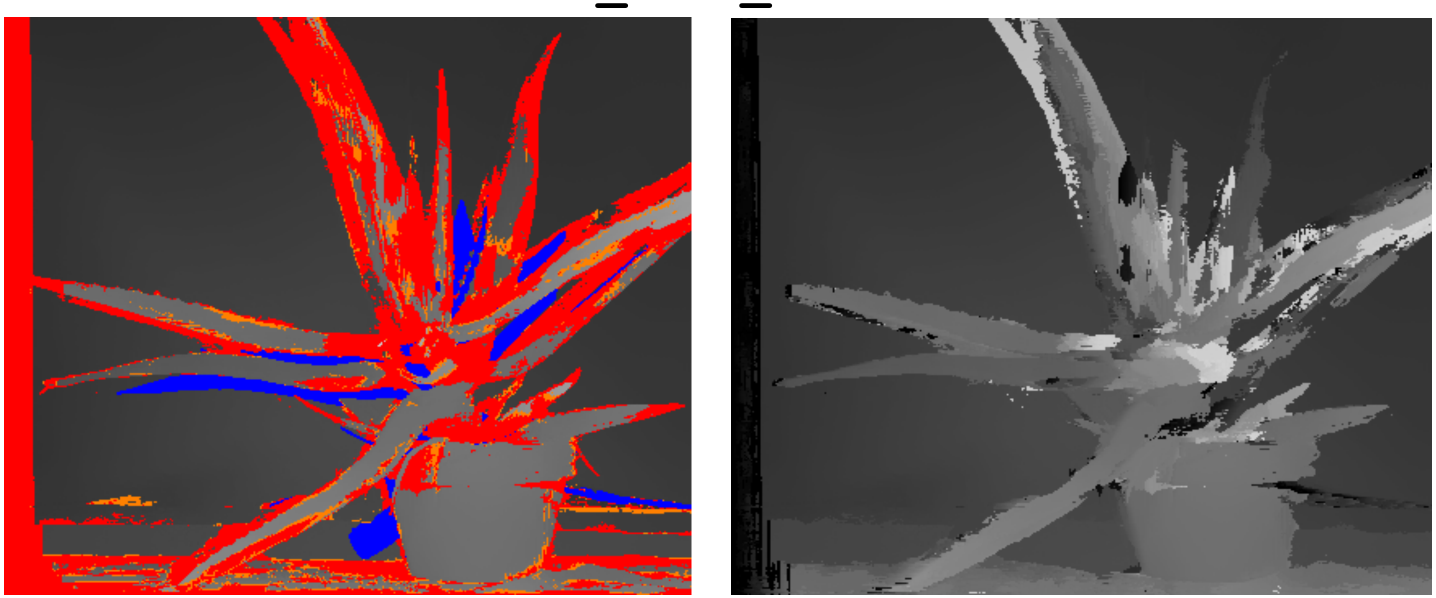


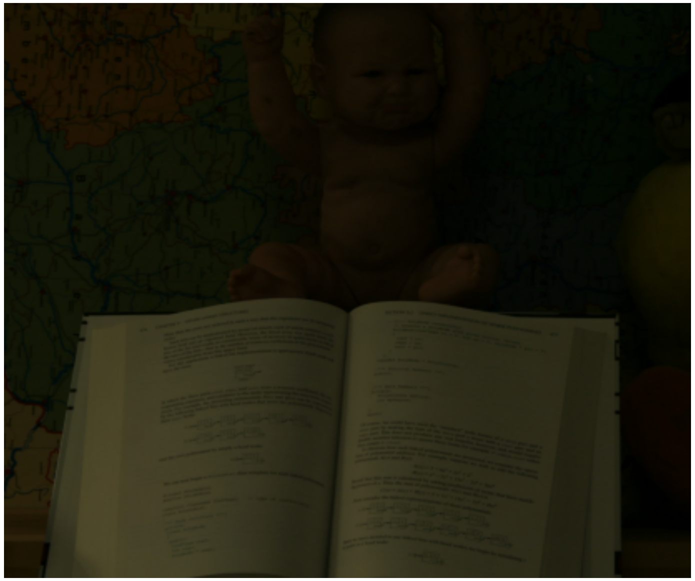
Groundtruth

TAD

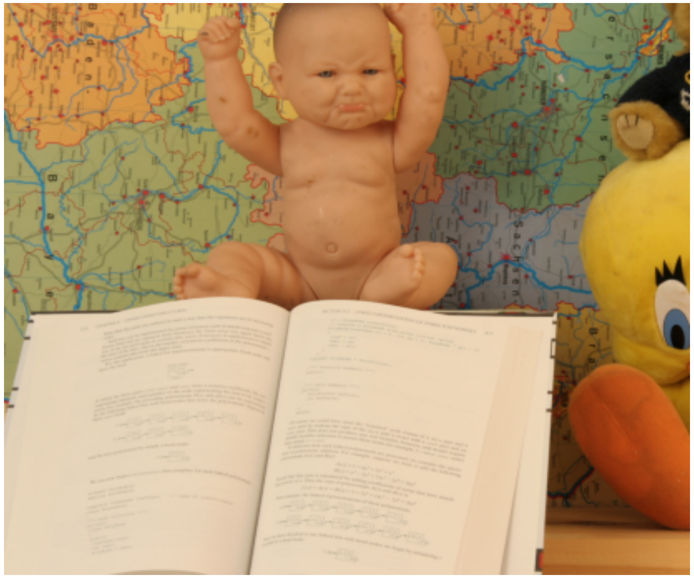


ROBUST_COST_FUNCTION

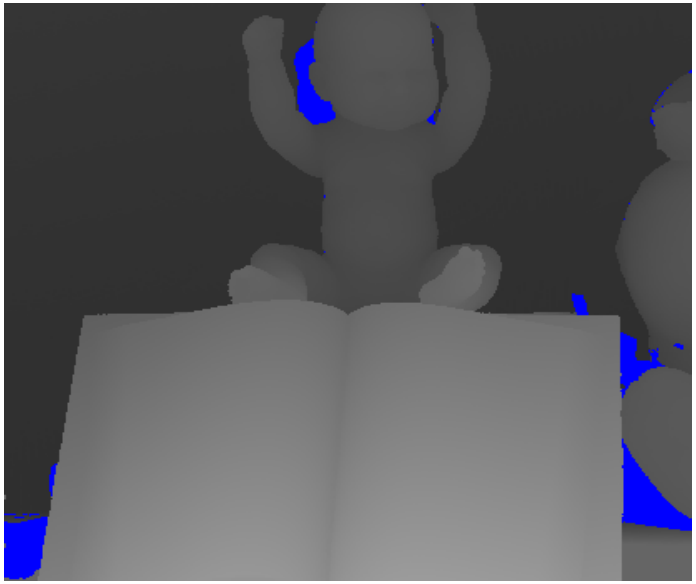




Left ILL(1)-EXP(0)

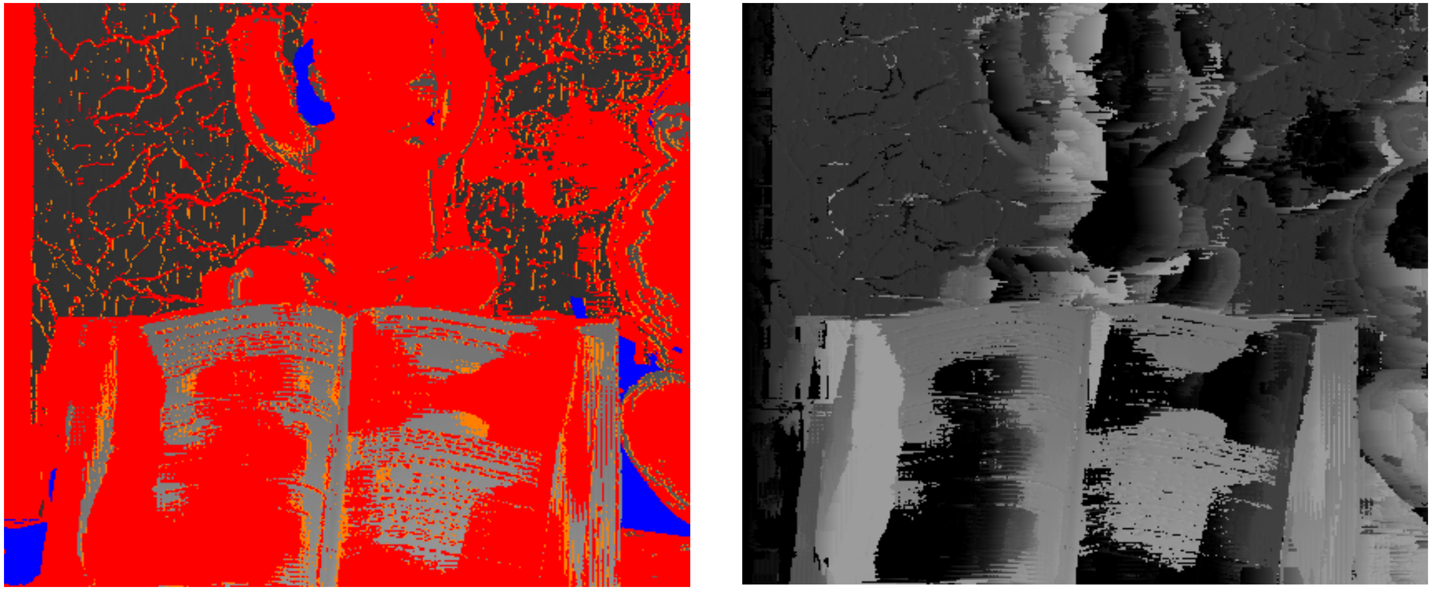


Right ILL(3)-EXP(2)

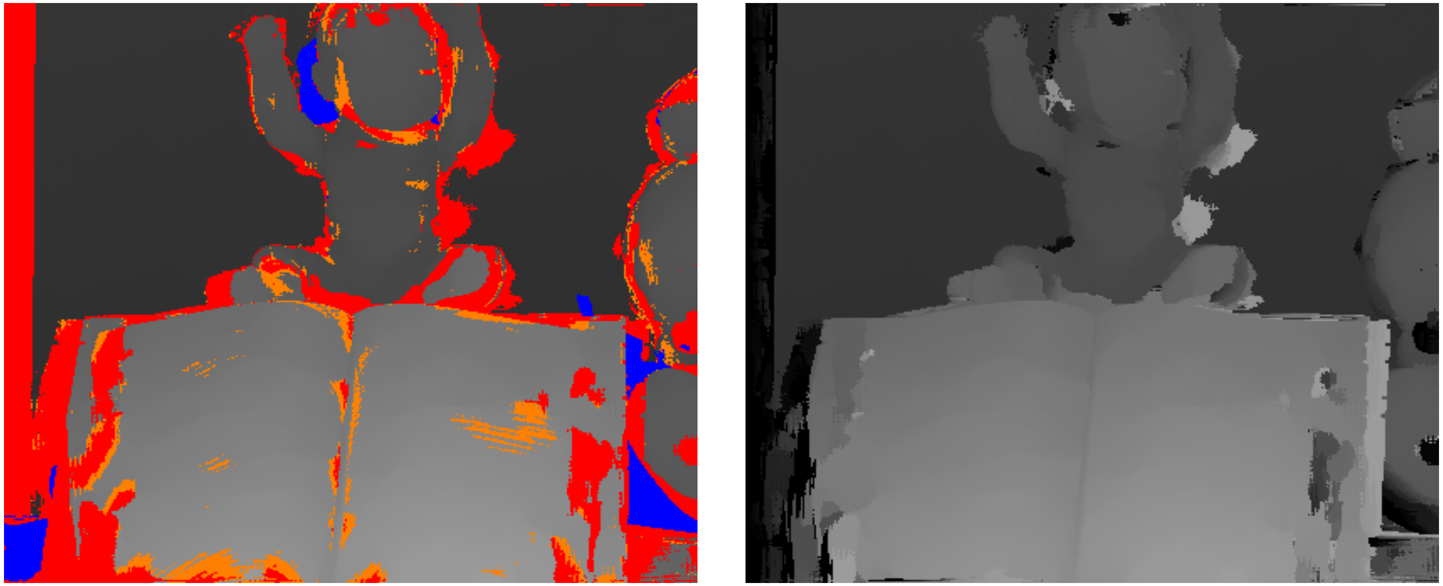


Groundtruth

NCC

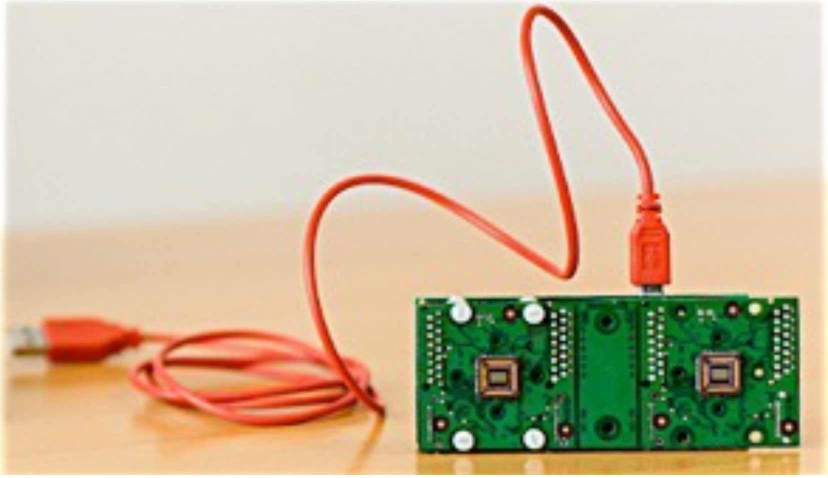


ROBUST_COST_FUNCTION



Real-time applications based on our embedded 3D camera

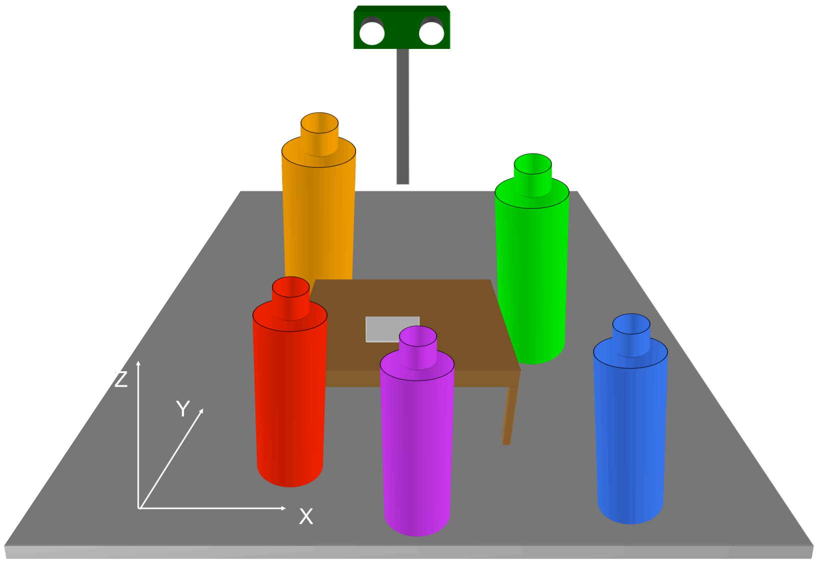
- 3D tracking
- SLAM
- Autonomous robot navigation
- Mobility aid for visually impaired

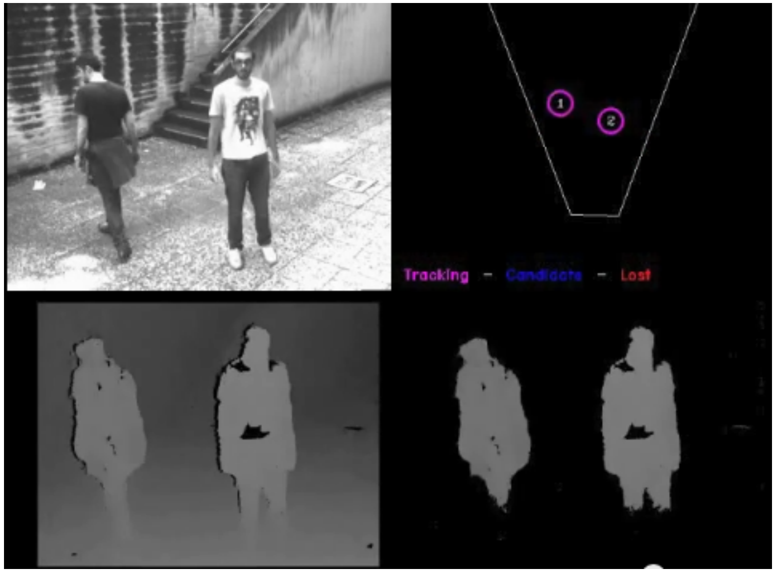


3D Tracking 1/2

Applications:

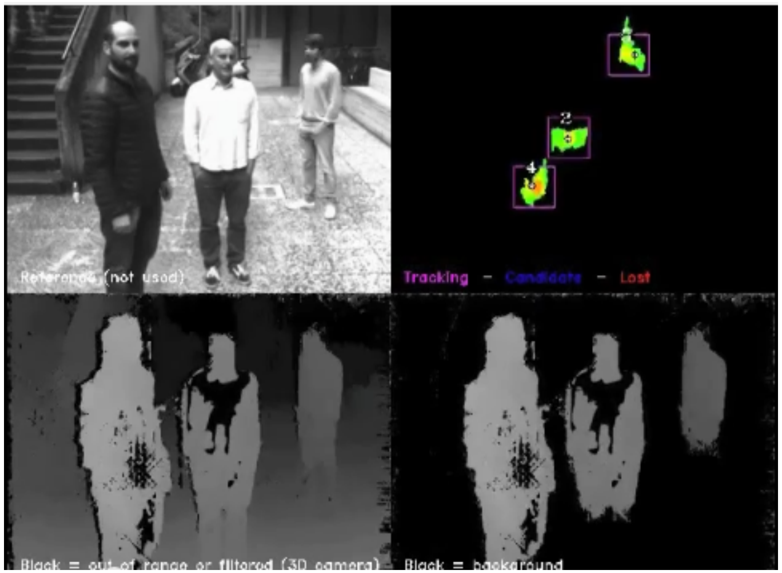
- people counting (building, bus, train)
- monitoring trajectories (shopping, sport)
- safety
- surveillance and security





3D Tracking 2/2

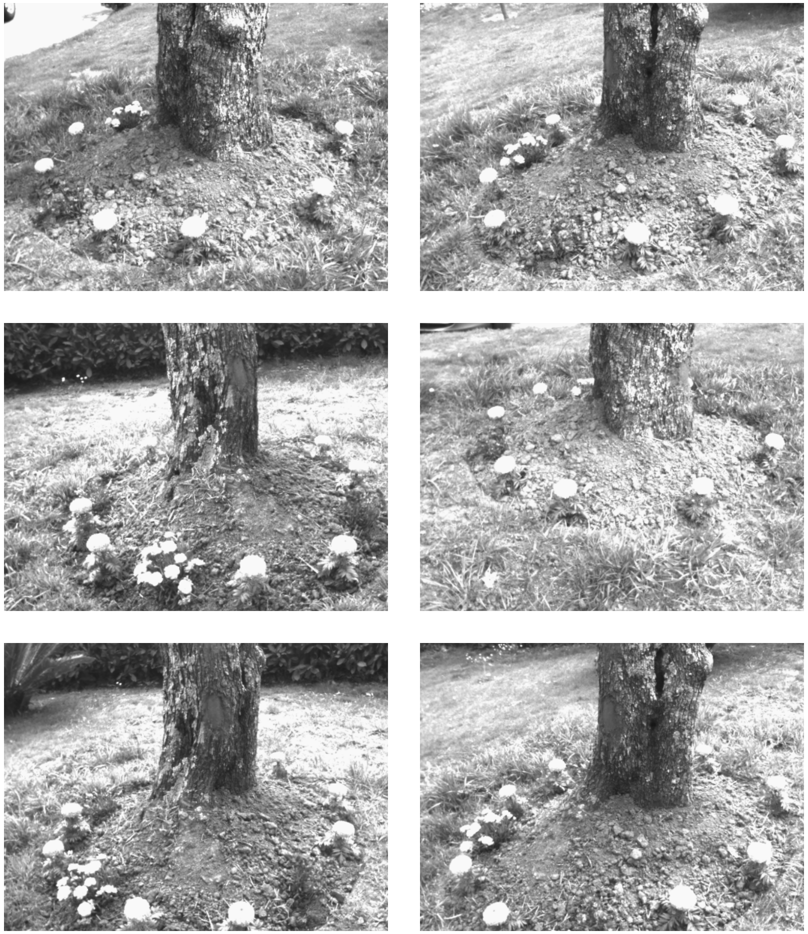
- Embedded computer + FPGA stereso camera
- 20+ fps

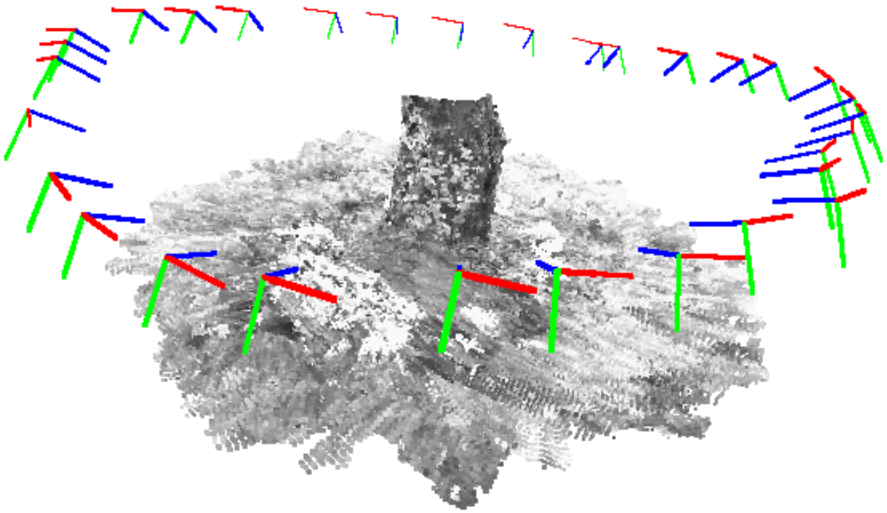


https://www.youtube.com/watch?v=2vorrRhBssQ

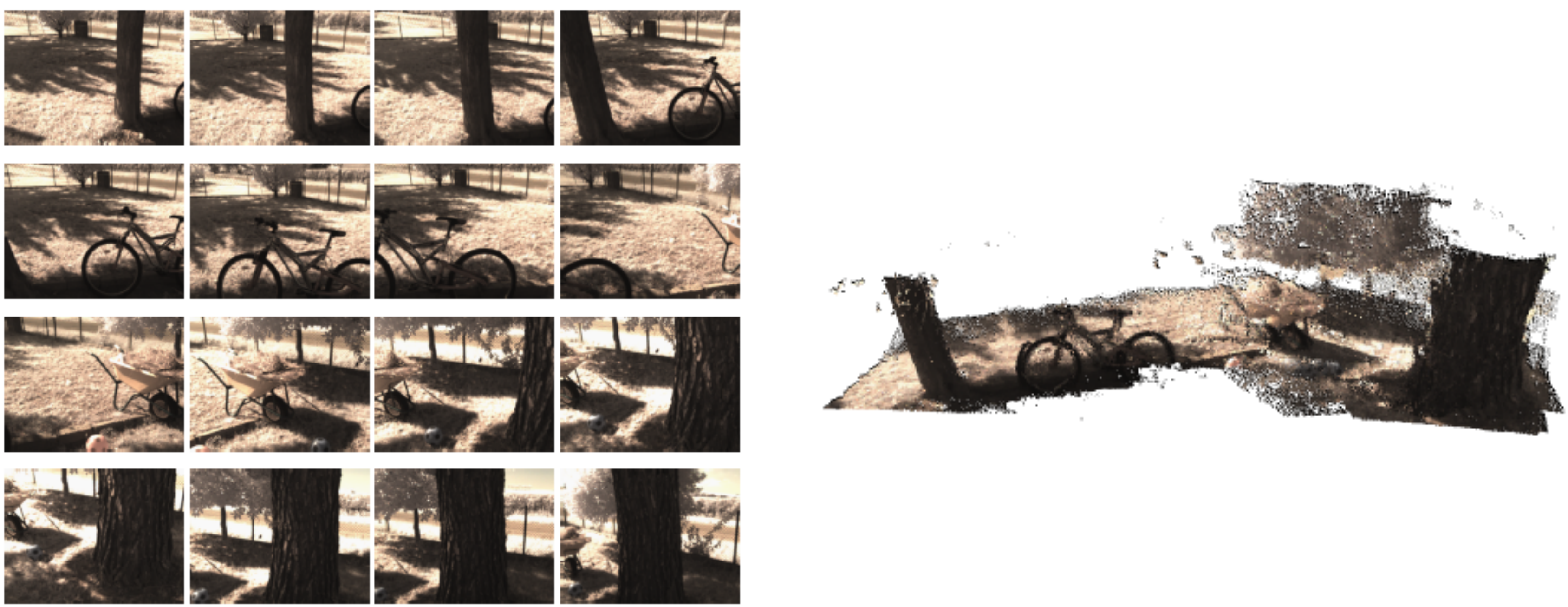
SLAM 1/2

- 3D scanning at 5+ fps (with bundle adjustment)



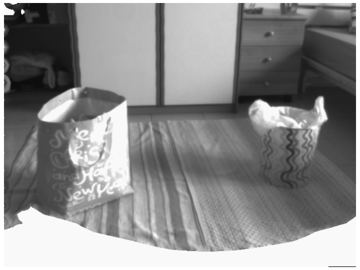


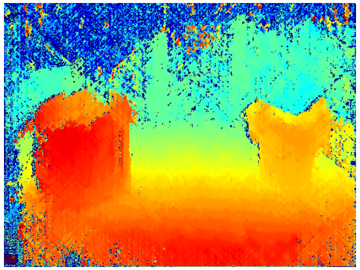
SLAM 2/2

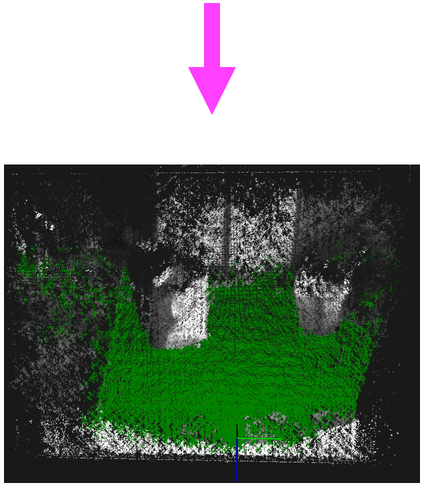


Autonomous robot navigation

- Real-time and reliable obstacle detection with the 3D camera and an embedded computer at 20+ fps
- Battery powered



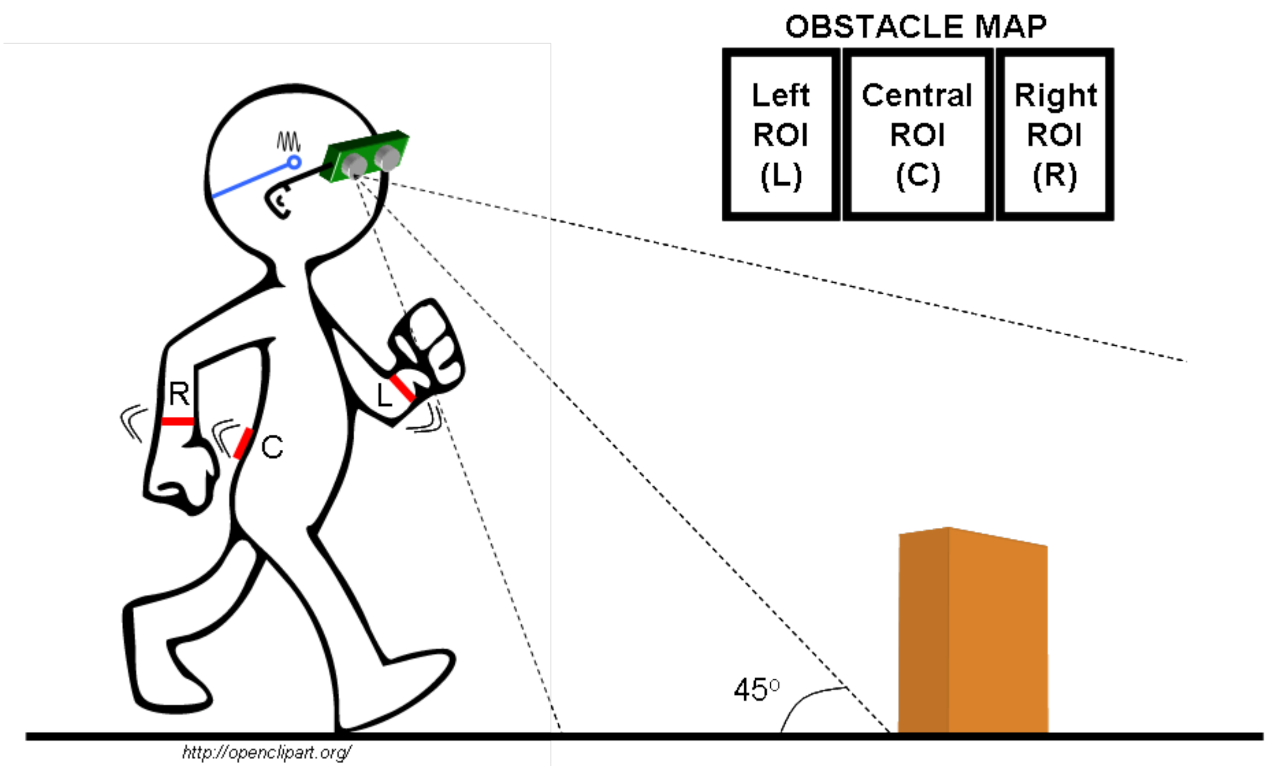


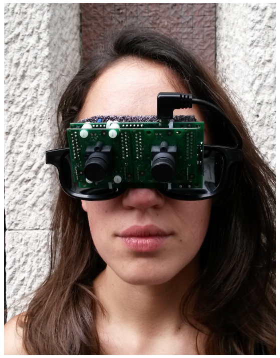


www.youtube.com/watch?v=7rieq3wfGDo

Mobility aid for visually impaired 1/4

- Wearable and lightweight (3D camera + computing platform about 150 g) system for autonomous nav.
- Feedback: vibrotactile and audio (by means of bone conductive headset)
- Enables hours of autonomous navigation with a small battery (3200 mAh) at 15+ fps

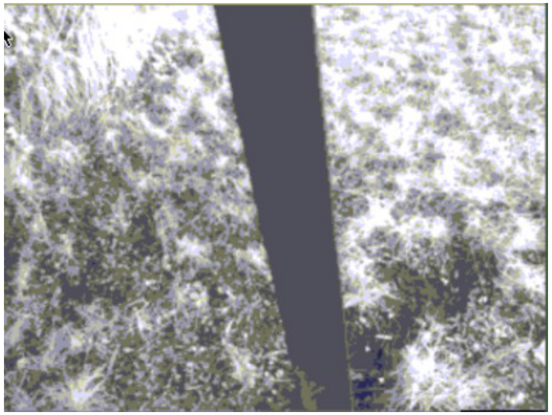


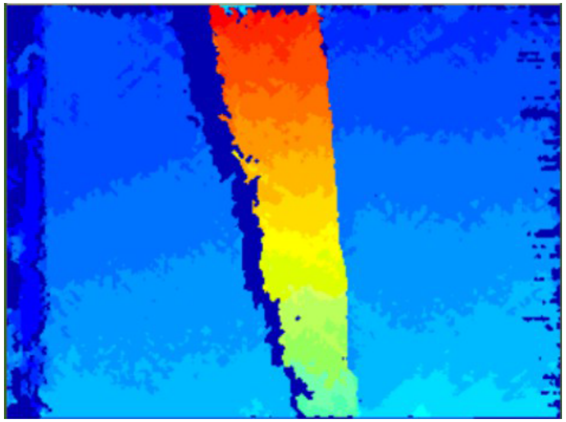


Mobility aid for visually impaired 2/4

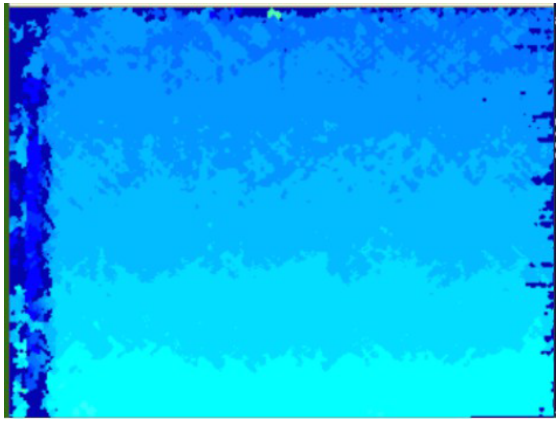
Mobility aid for visually impaired 3/4

- Real-time navigation example with obstacle





- Real-time navigation example without obstacles



White: detected plane
Black: obstacles

Mobility aid for visually impaired 4/4

Current prototype (in the news):

http://www.rai.tv/dl/RaiTV/programmi/media/ContentItem-fbb80bea-9d96-44ea-ae62-1fa3b5e572a5-tgr.html?refresh_ce#p=0

https://www.youtube.com/watch?v=DQ7x3PtFkJw#t=1346

http://www.corriere.it/salute/disabilita/14_novembre_28/dal-video-telefono-sordociechi-all-app-che-aiuta-badante-393744a6-7701-11e4-90d4-0eff89180b47.shtml

First prototype:

www.youtube.com/watch?v=G1UIUXUu2wY

References

[1] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Classification and evaluation of cost aggregation methods for stereo correspondence, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2008)*

[2] Y. Boykov, O. Veksler, and R. Zabih, A variable window approach to early vision
IEEE Trans. PAMI, 20(12):1283–1294, 1998

[3] S. Chan, Y. Wong, and J. Daniel, Dense stereo correspondence based on recursive adaptive size multi-windowing
In Proc. Image and Vision Computing New Zealand (IVCNZ'03), volume 1, pages 256–260, 2003

[4] C. Demoulin and M. Van Droogenbroeck, A method based on multiple adaptive windows to improve the determination of disparity maps. In Proc. IEEE Workshop on Circuit, Systems and Signal Processing, pages 615–618, 2005

[5] M. Gerrits and P. Bekaert. Local Stereo Matching with Segmentation-based Outlier Rejection
In Proc. Canadian Conf. on Computer and Robot Vision (CRV 2006), pages 66-66, 2006

[6 M. Gong and R. Yang. Image-gradient-guided real-time stereo on graphics hardware
In Proc. Int. Conf. 3D Digital Imaging and Modeling (3DIM), pages 548–555, 2005

[7] H. Hirschmuller, P. Innocent, and J. Garibaldi, Real-time correlation-based stereo vision with reduced border errors
Int. Journ. of Computer Vision, 47:1–3, 2002

[8] S. Kang, R. Szeliski, and J. Chai, Handling occlusions in dense multi-view stereo
In Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2001), pages 103–110, 2001

[9] J. Kim, K. Lee, B. Choi, and S. Lee. A dense stereo matching using two-pass dynamic programming with generalized ground control points, In Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2005), pages 1075–1082, 2005

[10] F. Tombari, S. Mattoccia, and L. Di Stefano, Segmentation-based adaptive support for accurate stereo correspondence
PSIVT 2007

[11] D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms
Int. Jour. Computer Vision, 47(1/2/3):7–42, 2002

[12] O. Veksler. Fast variable window for stereo correspondence using integral images, In Proc. Conf. on Computer Vision and Pattern Recognition (CVPR 2003), pages 556–561, 2003

[13] Y. Xu, D. Wang, T. Feng, and H. Shum, Stereo computation using radial adaptive windows, In Proc. Int. Conf. on Pattern Recognition (ICPR 2002), volume 3, pages 595– 598, 2002

[14] K. Yoon and I. Kweon, Adaptive support-weight approach for correspondence search, IEEE Trans. PAMI, 28(4):650–656, 2006

[15] D. Scharstein and R. Szeliski, **http://vision.middlebury.edu/stereo/eval/**

[16] A. Ansar, A. Castano, L. Matthies, Enhanced real-time stereo using bilateral filtering
IEEE Conference on Computer Vision and Pattern Recognition 2004

[17] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light"
In IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2003), volume 1, pages 195-202

[18] D. Scharstein and C. Pal. Learning conditional random fields for stereo.
In IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2007)

[19] H. Hirschmüller and D. Scharstein. Evaluation of cost functions for stereo matching.
In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)

[20] E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, 1998

[21] R.I.Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000

[22] G. Bradsky, A. Kaehler, Learning Opencv, O'Reilly, 2008

[23] OpenCV Computer Vision Library, http://sourceforge.net/projects/opencvlibrary/

[24] Jean-Yves Bouguet , Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/

[25] M. A. Fischler and R. C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Comm. of the ACM* **24**: 381–395, June 1981

[26] Z. Wang and Z. Zheng, A region based stereo matching algorithm using cooperative optimization
IEEE CVPR 2008

[27] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(4):401-406, April 1998

[28] J. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence. European Conf. on Computer Vision, Stockholm, Sweden, 151–158

[29] S. Mattoccia, F. Tombari, and L. Di Stefano, Stereo vision enabling precise border localization within a scanline optimization framework, ACCV 2007

[30 H. Hirschmüller. Stereo vision in structured environments by consistent semi-global matching.
CVPR 2006, PAMI 30(2):328-341, 2008

[31] F. Tombari, S. Mattoccia, L. Di Stefano, F. Tonelli, Detecting motion by means of 2D and 3D information
ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing (ACCV 2007 WS)

[32] P. Azzari, L. Di Stefano, F. Tombari, S. Mattoccia, Markerless augmented reality using image mosaics
International Conference on Image and Signal Processing (ICISP 2008)

[33] Li Zhang, Brian Curless, and Steven M. Seitz Spacetime Stereo: Shape Recovery for Dynamic Scenes
IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), pp. 367-374

[34] J. Davis, D. Nehab, R. Ramamoothi, S. Rusinkiewicz. Spacetime Stereo : A Unifying Framework for Depth from Triangulation, *IEEE Trans. On Pattern Analysis and Machine Intelligence (PAMI),* vol. 27, no. 2, Feb 2005

[35] F. Tombari, L. Di Stefano, S. Mattoccia, A. Zanetti, Graffiti detection using a Time-Of-Flight camera
Advanced Concepts for Intelligent Vision Systems (ACIVS 2008)

[36] L. Di Stefano, F. Tombari, A. Lanza, S. Mattoccia, S. Monti, Graffiti detection using two views
ECCV 2008 - 8th International Workshop on Visual Surveillance (VS 2008)

[37] T. Darrell, D. Demirdijan, N. Checka, P. Felzenszwalb, Plan-view trajectory estimation with dense stereo background models, International Conference on Computer Vision (ICCV 2001), 2001

[38] M. Harville, Stereo person tracking with adaptive plan-view templates of height and occupancy statistics Image and Vision Computing 22(2) pp 127-142, February 2004

[39] OpenCV Computer Vision Library, http://sourceforge.net/projects/opencvlibrary/

[40] Jean-Yves Bouguet , Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/

[41] T. Kanade, H. Kato, S. Kimura, A. Yoshida, and K. Oda, Development of a Video-Rate Stereo Machine International Robotics and Systems Conference (IROS '95), Human Robot Interaction and Cooperative Robots, 1995

[42] O. Faugeras, B. Hotz, H. Mathieu, T. Viville, Z. Zhang, P. Fua, E. Thron, L. Moll, G. Berry, Real-time correlation-based stereo: Algorithm. Implementation and Applications, INRIA TR n. 2013, 1993

[43] F. Crow, Summed-area tables for texture mapping, Computer Graphics, 18(3):207–212, 1984

[44] M. Mc Donnel. Box-filtering techniques, Computer Graphics and Image Processing, 17:65–70, 1981

[45] A. Goshtasby, 2-D and 3-D Image Registration for Medical, Remote Sensing and Industrial Applications New York: Wiley, 2005

[46] B. Zitova and J. Flusser, Image registration methods:A survey, Image Vision Computing, vol. 21, no. 11, pp. 977–1000, 2003

[47] Changming Sun, Recursive Algorithms for Diamond, Hexagon and General Polygonal Shaped Window Operations Pattern Recognition Letters, 27(6):556-566, April 2006

[48] L. Di Stefano, M. Marchionni, S. Mattoccia, A fast area-based stereo matching algorithm, Image and Vision Computing, 22(12), pp 983-1005, October 2004

[49] L. Di Stefano, M. Marchionni, S. Mattoccia, A PC-based real-time stereo vision system, Machine Graphics & Vision, 13(3), pp. 197-220, January 2004

[50] D. Comaniciu and P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24:603–619, 2002

[51] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV98*, pages 839–846, 1998

[52] V. Kolmogorov and R. Zabih, Computing visual correspondence with occlusions using graph cuts, ICCV 2001

[53] A. Klaus, M. Sormann and K. Karner, Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure, ICPR 2006

[54] Z. Wang and Z. Zheng, A region based stereo matching algorithm using cooperative optimization, CVPR 2008

[55] L. Di Stefano, S. Mattoccia, Real-time stereo within the VIDET project Real-Time Imaging, 8(5), pp. 439-453, Oct. 2002

[56] F. Tombari, S. Mattoccia, L. Di Stefano, Full search-equivalent pattern matching with Incremental Dissimilarity Approximations, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1), pp 129-141, January 2009

[57] S. Mattoccia, F. Tombari, L. Di Stefano, Fast full-search equivalent template matching by Enhanced Bounded Correlation, IEEE Transactions on Image Processing, 17(4), pp 528-538, April 2008

[58] L. Di Stefano, S. Mattoccia, F. Tombari, ZNCC-based template matching using Bounded Partial Correlation Pattern Recognition Letters, 16(14), pp 2129-2134, October 2005

[59] F. Tombari, L. Di Stefano, S. Mattoccia, A. Galanti, Performance evaluation of robust matching measures 3rd International Conference on Computer Vision Theory and Applications (VISAPP 2008)

[60] R. Zabih, J John Woodll Non-parametric Local Transforms for Computing Visual Correspondence, ECCV 1994

[61] D. N. Bhat, S. K. Nayar, Ordinal measures for visual correspondence, CVPR 1996

[62] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110

[63] R.Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30, 6, June 2008, pp 1068-1080

[64] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Near real-time stereo based on effective cost aggregation International Conference on Pattern Recognition (ICPR 2008)

[65] S. Mattoccia, S. Giardino,A. Gambini, Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering, Asian Conference on Computer Vision (ACCV 2009), September 23-27 2009, Xiang, China

[66] S. Mattoccia, A locally global approach to stereo correspondence, 3D Digital Imaging and Modeling (3DIM 2009), pp 1763-1770, October 3-4, 2009, Kyoto, Japan

[67] S. Mattoccia, Improving the accuracy of fast dense stereo correspondence algorithms by enforcing local consistency of disparity fields, 3D Data Processing, Visualization, and Transmission (3DPVT 2010), 17-20 May 2010, Paris, France

[68] S. Mattoccia, Fast locally consistent dense stereo on multicore, Sixth IEEE Embedded Computer Vision Workshop (ECVW2010), CVPR workshop, June 13, 2010, San Francisco, USA

[69] S. Mattoccia, Accurate dense stereo by constraining local consistency on superpixels, 20th International Conference on Pattern Recognition (ICPR2010), August 23-26, 2010, Istanbul, Turkey

[70] L. Wang, M. Liao, M. Gong, R. Yang, and D. Nistér. High-quality real-time stereo using adaptive cost aggregation and dynamic programming. 3DPVT 2006

[71] S. Mattoccia, M. Viti, F. Ries,. Near real-time Fast Bilateral Stereo on the GPU, 7th IEEE Workshop on Embedded Computer Vision (ECVW20011), CVPR Workshop, June 20, 2011, Colorado Springs (CO), USA

[72] S. Mattoccia, L. De-Maeztu, "A fast segmentation-driven algorithm for stereo correspondence", International Conference on 3D (IC3D 2011), December 7-8, 2011, Liege, Belgium

[73] L. De-Maeztu, S. Mattoccia, A. Villanueva, R. Cabeza, "Efficient aggregation via iterative block-based adapting support weight",International Conference on 3D (IC3D 2011), December 7-8, 2011, Liege, Belgium

[74] D. Min, J. Lu, and M. Do, “A revisit to cost aggregation in stereo matching: how far can we reduce its computational redundancy?”, ICCV 2011

[75] L. De-Maeztu, S. Mattoccia, A. Villanueva, R. Cabeza, "Linear stereo matching", International Conference on Computer Vision (ICCV 2011), November 6-13, 2011, Barcelona, Spain

Stefano Mattoccia

mail: stefano.mattoccia@unibo.it
web: www.vision.deis.unibo.it/smatt

Tel: +39 051 2093860
Fax: +39 051 2093073

Address:

Department of Computer Science and Engineering (DISI)
University of Bologna
Viale Risorgimento, 2
40136 Bologna, ITALY