A MULTI-RESOLUTION APPROACH TO DEPTH FIELD ESTIMATION IN DENSE IMAGE ARRAYS

F. Battisti, M. Brizzi, M. Carli, A. Neri

Università degli Studi Roma TRE,

Roma, Italy

2nd Workshop on Light Fields for Computer Vision (LF4CV), CVPR 2017, Honolulu, Hawaii, USA

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Roadmap

- Proposed method
- Strengths - Drawbacks
- Conclusions

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Proposed method

- Depth field estimation in dense image arrays.
- Based on local correspondence measure based on the maximization of a smoothed version of the Likelihood functional.
- Exploit a subset of available images (cross and diagonal)
- To cope with flat surface regions, while preserving bandwidth in correspondence of edges, a multi-resolution scheme is adopted.

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Likelihood functional in single image-pair

- The relation between the image components of $\mathbf{L}^{(0,0)}(\mathbf{x})$ and those of the corresponding pixels in $\mathbf{L}^{(p,q)}(\mathbf{x})$ is:

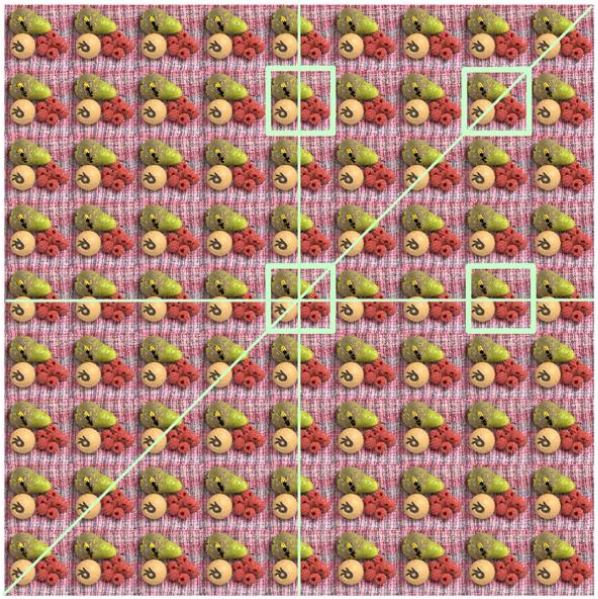
$$\mathbf{L}^{(0,0)}(\mathbf{x}) = \mathbf{L}^{(p,q)}\left[\mathbf{x} - \mathbf{b}^{(p,q)}\frac{\gamma}{z(\mathbf{x})}\right] + \mathbf{n}^{(p,q)}(\mathbf{x});\ \ \forall \mathbf{x} \in V^{(p,q)}$$

- The log-likelihood functional of the depth field, given the pair $\{\mathbf{L}^{(0,0)}(\mathbf{x}), \mathbf{L}^{(p,q)}(\mathbf{x})\}$:

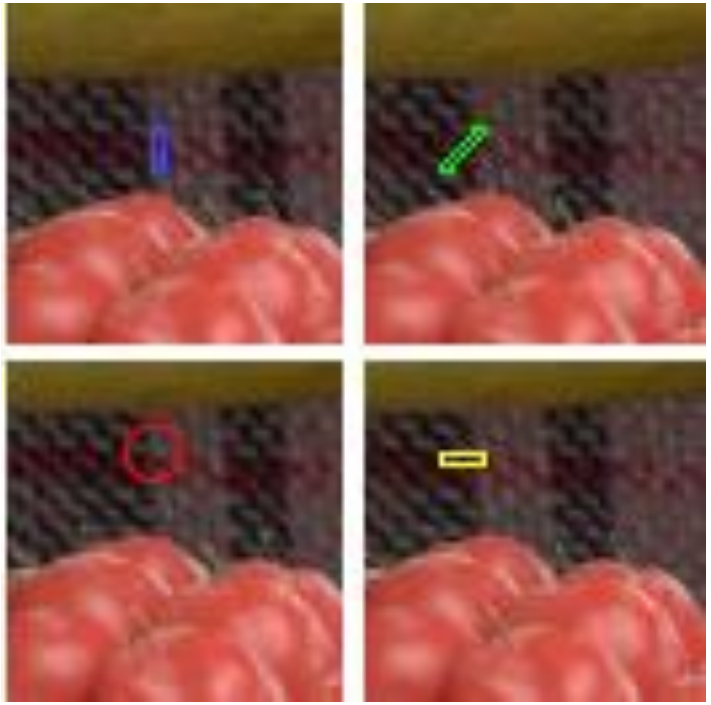
$$\ln \Lambda\left[\mathbf{L}^{(0,0)}, \mathbf{L}^{(p,q)}, Z\right] = -\frac{1}{2\sigma_N^2}\iint_{V^{(p,q)}} \left\| \mathbf{L}^{(0,0)}(\mathbf{x}) - \mathbf{L}^{(p,q)}\left[\mathbf{x} - \mathbf{b}^{(p,q)}\frac{\gamma}{z(\mathbf{x})}\right] \right\|^2 d\mathbf{x}$$

Multiple images

- Redundant information
- Epipolar lines are not only horizontal lines
  - Slope can be easily estimated



Maximum Likelihood Depth Estimation



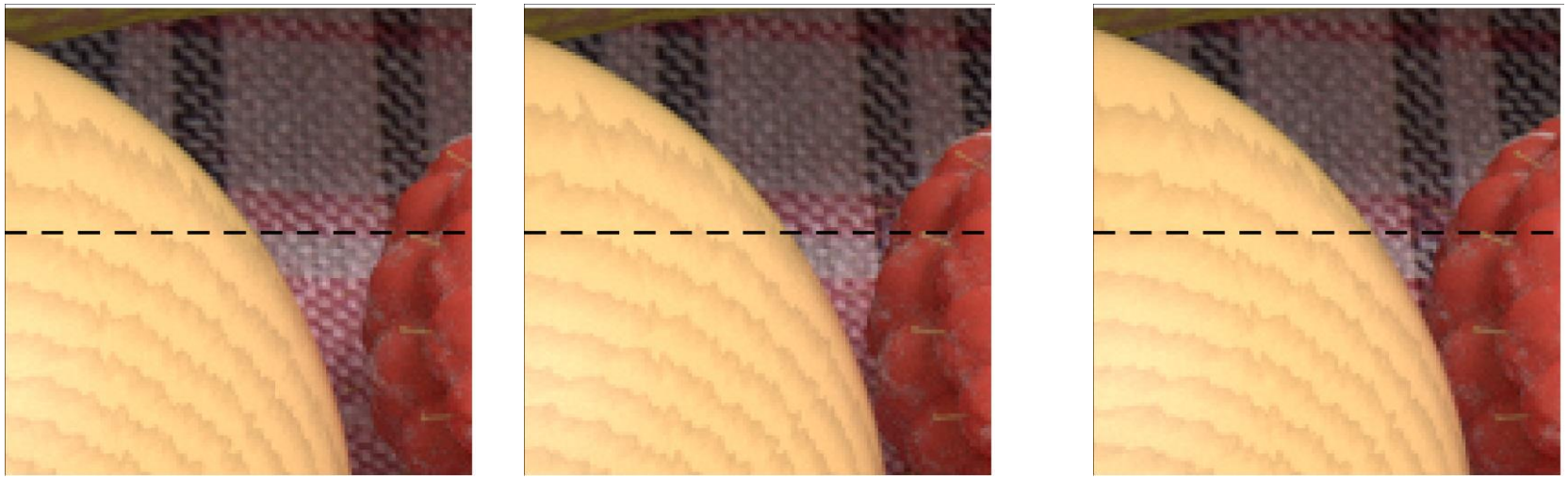
- *Aim*: To reduce the global optimization computational complexity.
- *How to*: local estimation of the depth field by maximizing a smoothed version of the Likelihood functional
  - The norm of the difference between the image components of the central image and the image $\mathbf{L}^{(p,q)}(\mathbf{x})$ is averaged on a neighbor of the current point by means of the moving window along the epipolar line:

$$E_w^{(p,q)}(\mathbf{x};z)=\sum_{m=-N_W}^{N_W} w(m)\left\|\mathbf{L}^{(0,0)}\left[\mathbf{x}-m\frac{\mathbf{b}^{(p,q)}}{\left|\mathbf{b}^{(p,q)}\right|}\right]-\mathbf{L}^{(p,q)}\left[\mathbf{x}-m\frac{\mathbf{b}^{(p,q)}}{\left|\mathbf{b}^{(p,q)}\right|}-\mathbf{b}^{(p,q)}\frac{\gamma}{z(\mathbf{x})}\right]\right\|^2$$

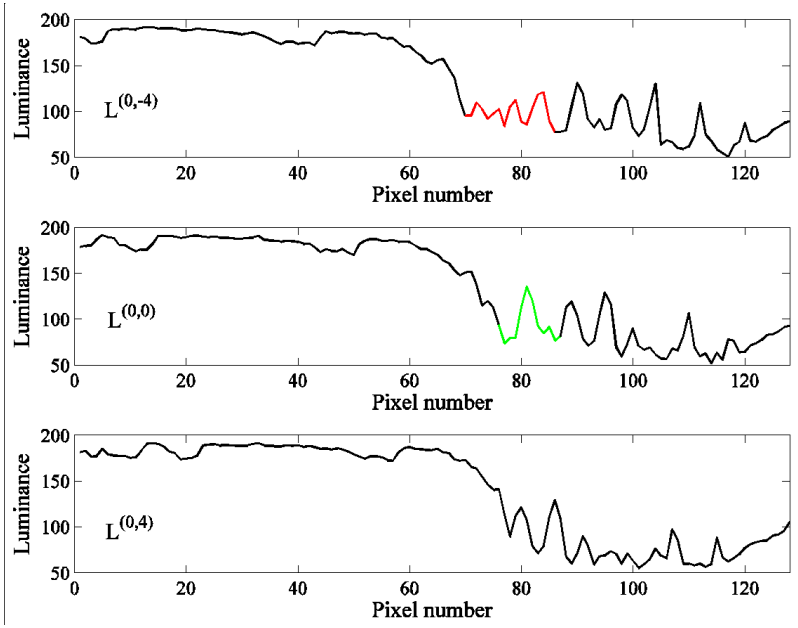
The local estimate of the depth field is computed as follows:

$$\hat{z}_{local}^{\tilde{U}}(\mathbf{x})=Arg\left\{\min_{z(\mathbf{x})}\sum_{(p,q)\in\tilde{U}}E_w^{(p,q)}(\mathbf{x};z)\right\}$$

Occlusions



- Only points visible in both images should be included in the local estimate.
- The local estimator may produce wrong estimates in presence of occlusions.



ROMA TRE
UNIVERSITÀ DEGLI STUDI

Occlusions

- To (partially) account for occlusions:
  - two different averages exploiting one-sided exponential window are computed
  - only the *best match* is used.

- Window size selection requires a trade-off between resolution and accuracy of the estimate:
  - an adaptive logic based on thresholding with a low-pass filtered the squared L^2 norm of the gradient magnitude map is adopted

Coarse-to-fine estimator

1. Quantize depth range into M_D discrete values:

$$Z_Q = \{z_k, k \in [1, M_D]\}$$

2. for each pixel and for each quantized depth z_k :
    1. Compute the functional

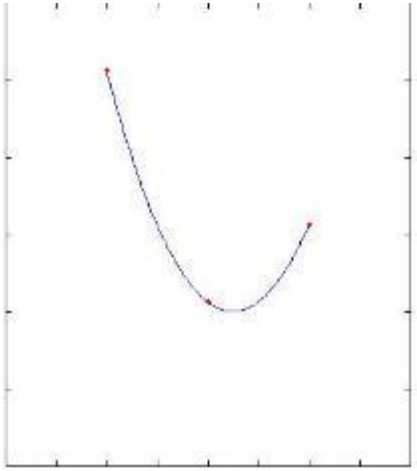
$$\mu^{\tilde{U}}(\mathbf{x}; z) = \sum_{(p,q) \in \tilde{U} \cap U'} \min\left[\mu_{\Delta}^{(p,q)}(\mathbf{x}; z), \mu_{\Delta}^{(-p,-q)}(\mathbf{x}; z)\right]$$

    1. retain as coarse estimate the depth corresponding to smallest value of $\mu^{\tilde{U}}(\mathbf{x}; z)$

$$\hat{z}_{coarse}(\mathbf{x}) = z_{\hat{k}_{coarse}(\mathbf{x})}$$

$$\hat{k}_{coarse}(\mathbf{x}) = Arg\left\{\min_{k \in \{[1, M_D]\}} \mu^{\tilde{U}}(\mathbf{x}; z_k)\right\}$$

    1. refined by computing the depth corresponding to the maximum of the parabola fitting $\mu^{\tilde{U}}(\mathbf{x}; z)$

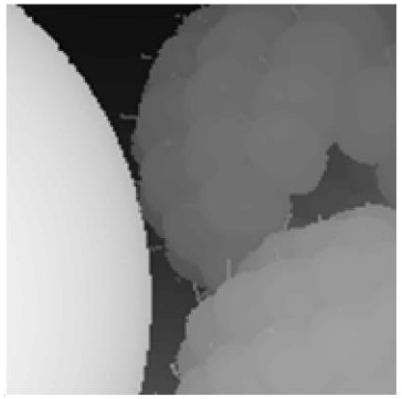


Adaptive Window Size

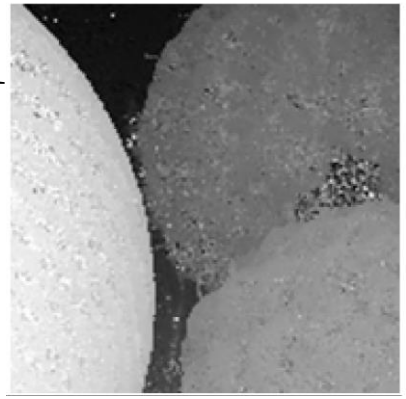
- Estimating depth from a single stereo pair, to avoid estimate ambiguities, **large** window size should be used.
  - Over-smoothing the reconstructed depth field.
    - *This effect represents a major weakness of the local minimization with respect to global methods.*

- Estimating depth from a dense array of images, windows of **small** size can be used.
    - *Tuning of the window size requires a trade-off between resolution of the depth map and accuracy of the estimate.*

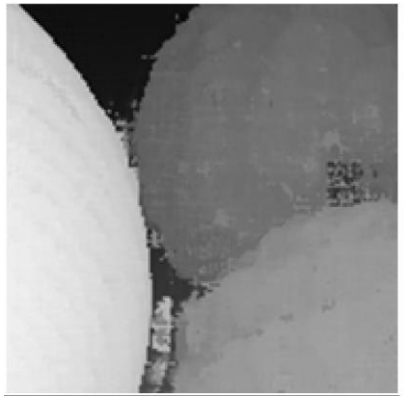
Ground truth



N_w=1

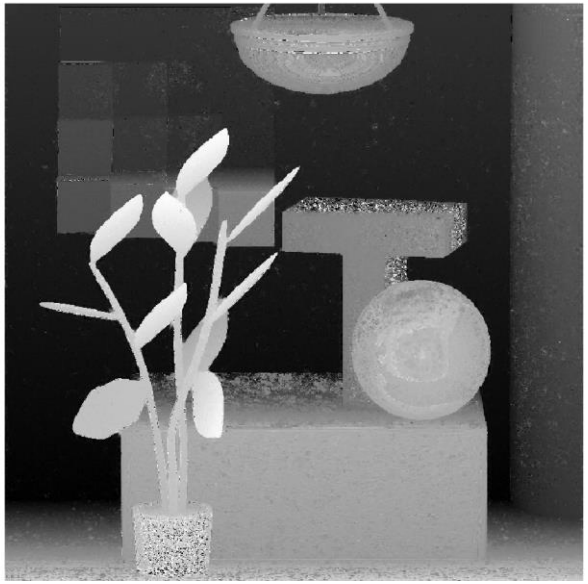


N_w=9



Adaptive Window Size

N_w=1



N_w=4

- If the squared magnitude of the image gradient is large enough, select a small window.
- Otherwise a larger one is used

$$e_{\nabla}(\mathbf{x}) = \sum_{m=1}^{M} \left\| \nabla_{\mathbf{x}} L_m^{(0,0)}(\mathbf{x}) \right\|^2 * w_{\nabla}(\mathbf{x}),$$

$$N_W(\mathbf{x}) = \begin{cases} N_{W_1} & \text{if } e_{\nabla}(\mathbf{x}) > \gamma \\ N_{W_2} & \text{otherwise.} \end{cases}$$

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Multiresolution approach

- In flat uniform regions, the Likelihood functional may be so spread that a large ambiguity may remain even when a large window is adopted.
  - Possible solution: to apply the depth map estimator to the image array at lower resolution.

- Starting from the highest resolution, for each site and for the current resolution the ill-conditioning of the functional to be minimized is tested.
  - Resolution is reduced by a factor 2 until either the well conditioning is met or the lowest allowed resolution is reached.

ROMA TRE
UNIVERSITÀ DEGLI STUDI

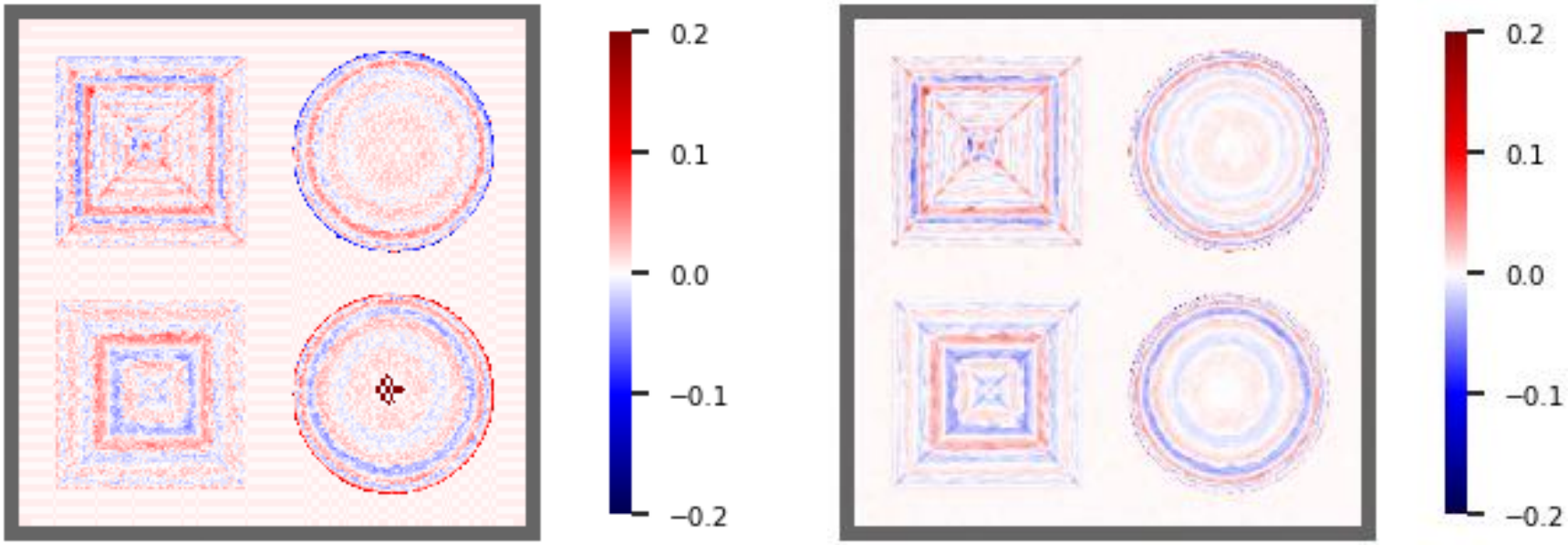
Comments

- Accuracy of the ML depth map estimate is proportional to the magnitude of the image gradient.
  - Thus, very poor performance can be expected on flat regions without textures.

- Fact: local optimization produces noisy depth maps.
  - Solution: a denoising is performed by means of 2D joint-histogram *weighted median filter* to avoid resolution losses associated to linear smoothing.

Weighted Median Filter

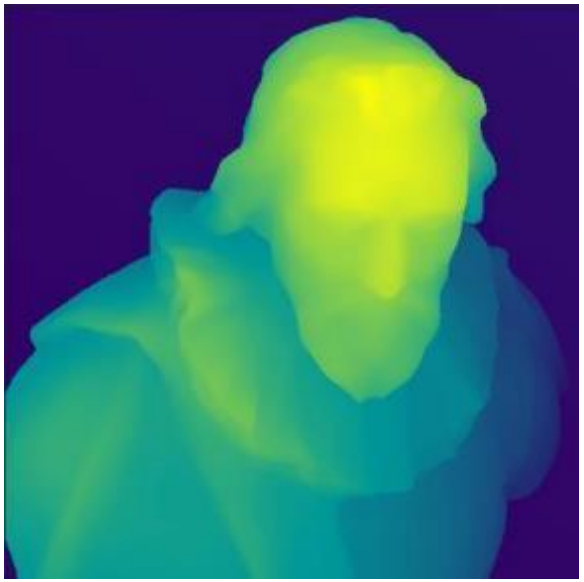
- Denoising is performed by means of 2D joint-histogram weighted median filter

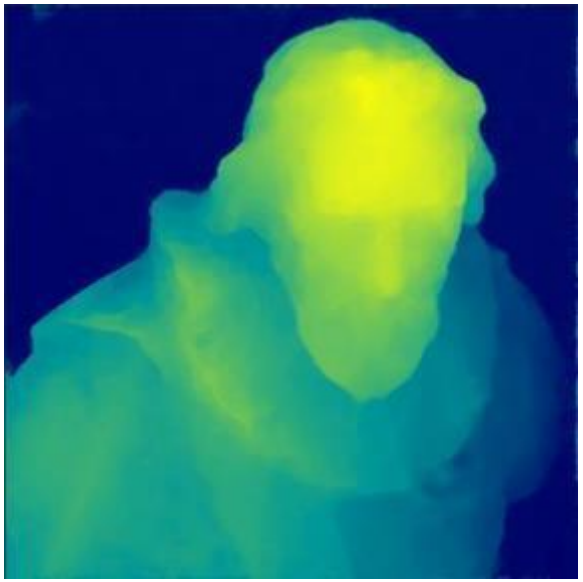


- The depth map was initially quantized on 256 levels before applying the median filter, which led to the presence of a regular pattern in the final results.
  - Using the appropriate number of levels greatly improved performances.

Results

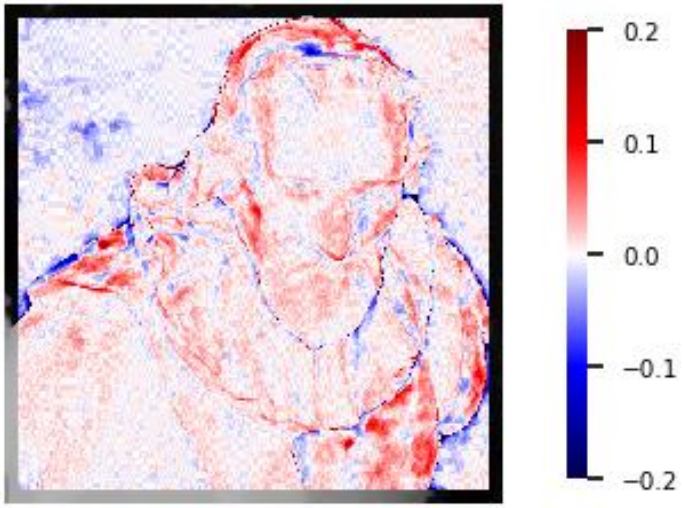
Ground truth

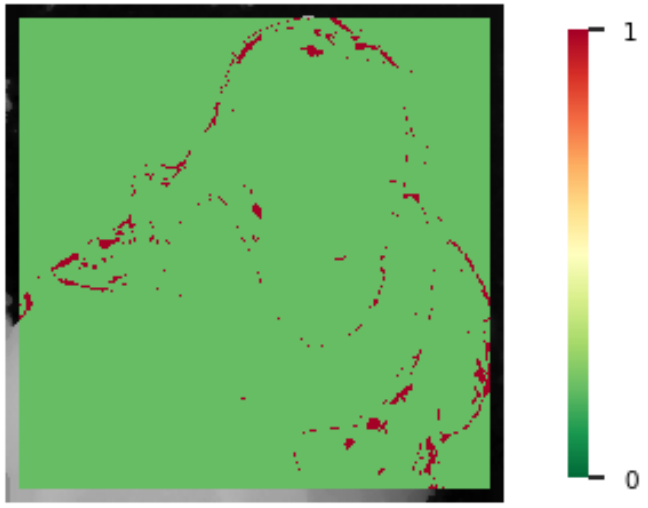




Estimated depth

MSE

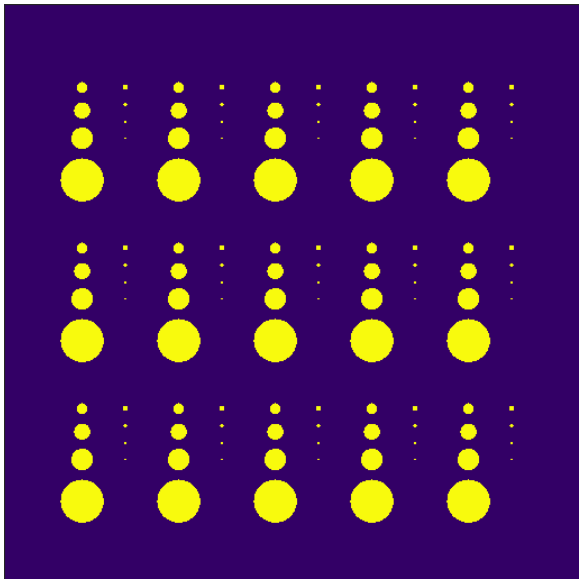


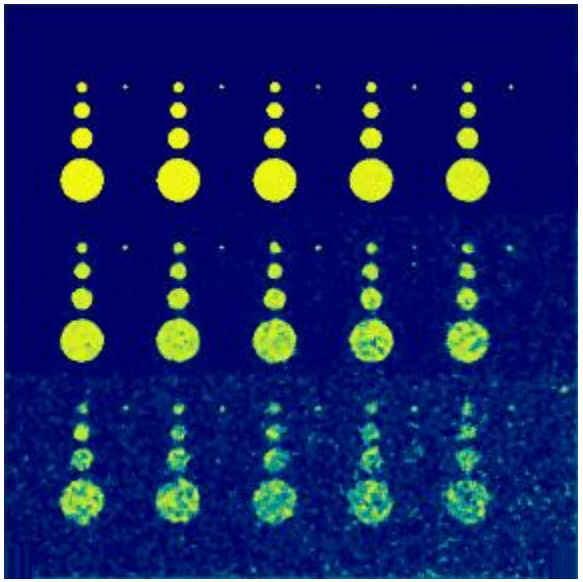


BadPix

Results

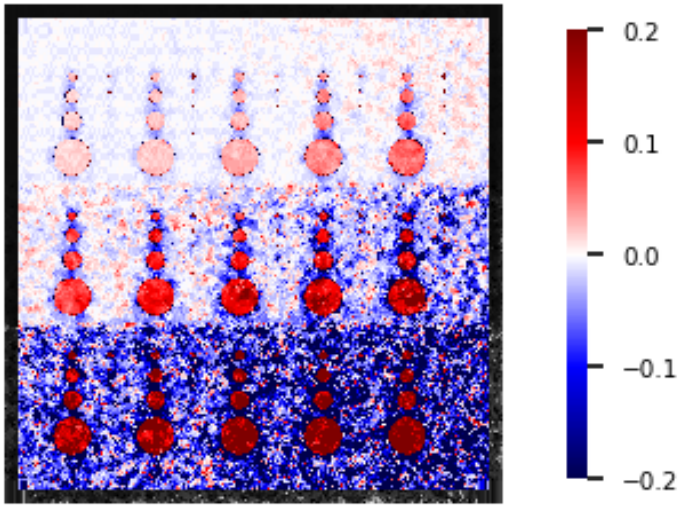
Ground truth

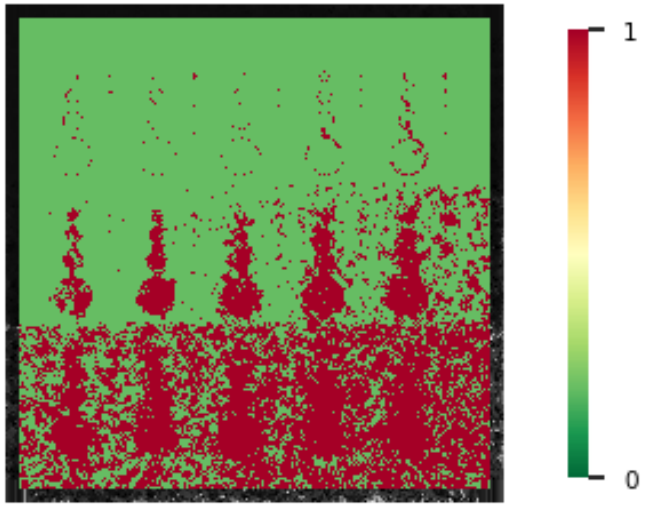




Estimated depth

MSE





BadPix

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Conclusions

- Extension to dense arrays of block matching techniques can take advantage from the high redundancy associated to multiple views
  - this redundancy has been employed to face artifacts originated by occlusions and to increase the depth estimation accuracy.

- For a given site, each element of the array carries an additive amount of Fisher's information about the depth
  - proportional to the magnitude of the spatial derivative along the epipolar direction times the length of the corresponding baseline with respect to the common view.

Conclusions

- The joint use of multiple pairs corresponding to several epipolar directions, allows to collect all the components of the gradient of the image.

- The experimental results indicate that, for the tested dataset, the subset of epipolar directions $\{0, \pi/4, \pi, 3\pi/4\}$ is a good trade-off between computational complexity and performance.
- Nevertheless, ambiguity in the selection of the maxima of the likelihood functional in flat, uniform areas still persists.
  - This fact suggested the adaptive control of the spatial bandwidth as a mitigation of the effects induced by the lack of gradient energy.

- Depth quantization may affect the method performances

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Thanks!