Perceptual depth indicator for S-3D content based on binocular and monocular cues

Pierre Lebreton[1,2], Alexander Raake[1], Marcus Barkowsky[2], Patrick Le Callet[2]

[1] Assessment of IP-Based Application, Telekom Innovation Laboratories, TU Berlin
[2] LUNAM Universite, Universite de Nantes, IRCCyN UMR CNRS 6597, Polytech' Nantes

4-7 November 2012

Telekom Innovation Laboratories

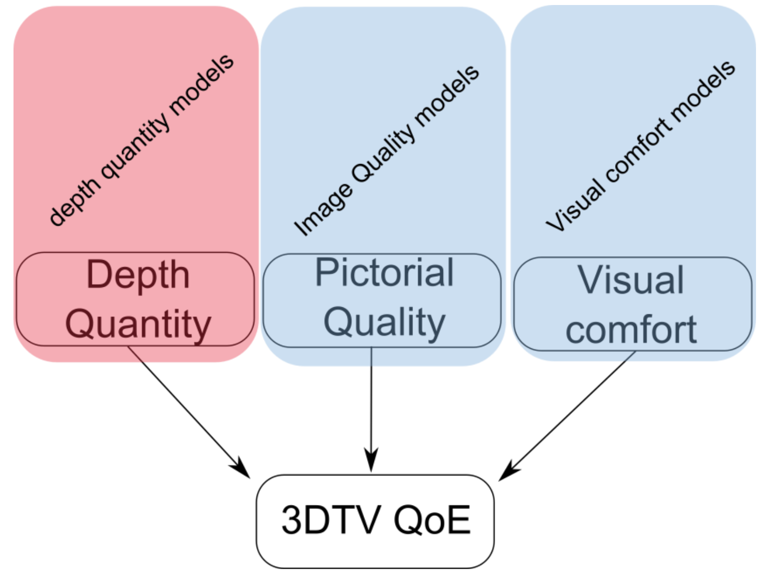
Outline

- **Motivation**
- Evaluation of monocular and binocular depth cues
- Depth cues reliability and pooling
- Conclusion

Evaluation and characterization of 3D video sequences

Quality evaluation purpose

- Need for objective model for evaluating perceived depth evaluation in 3D video sequences.



Classification

- 3D sequences need to be characterized:
  - Amount /quality of depth
  - Visual discomfort
  - Spatial / temporal complexity

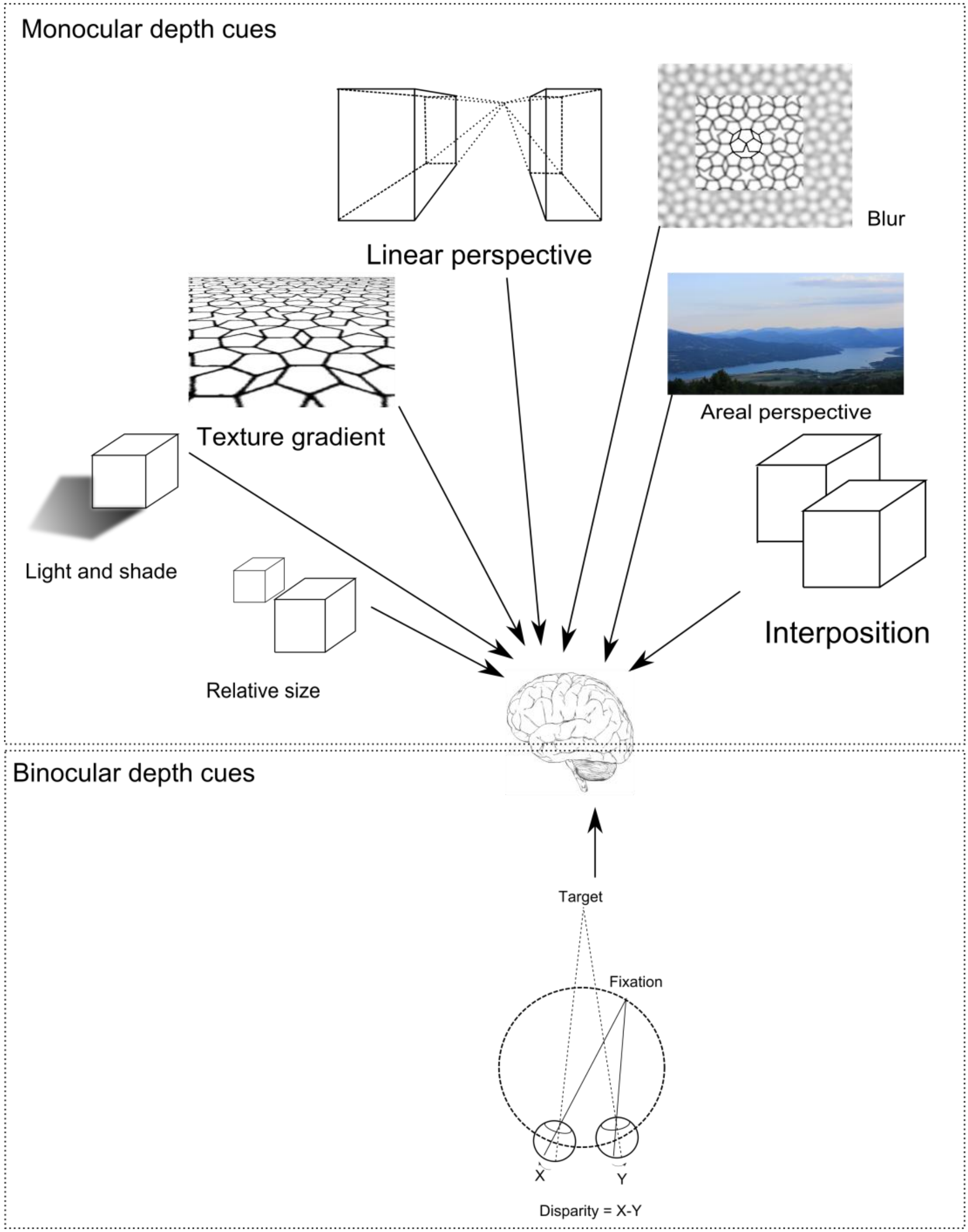
→

- Selection of source sequences in subjective experiments
- Identification of critical sequences
- Describe properties for indexation/search
- ...

Depth perception

- Depth is perceived due to different cues
- Evaluating depth cues provides information on:
  - Depth quantity
  - Depth quality
  - Visual discomfort (case of conflict between depth cues)

➢ Evaluation of the individual depth cues



Outline

- Motivation
- **Evaluation of monocular and binocular depth cues**
- Depth cues reliability and pooling
- Conclusion

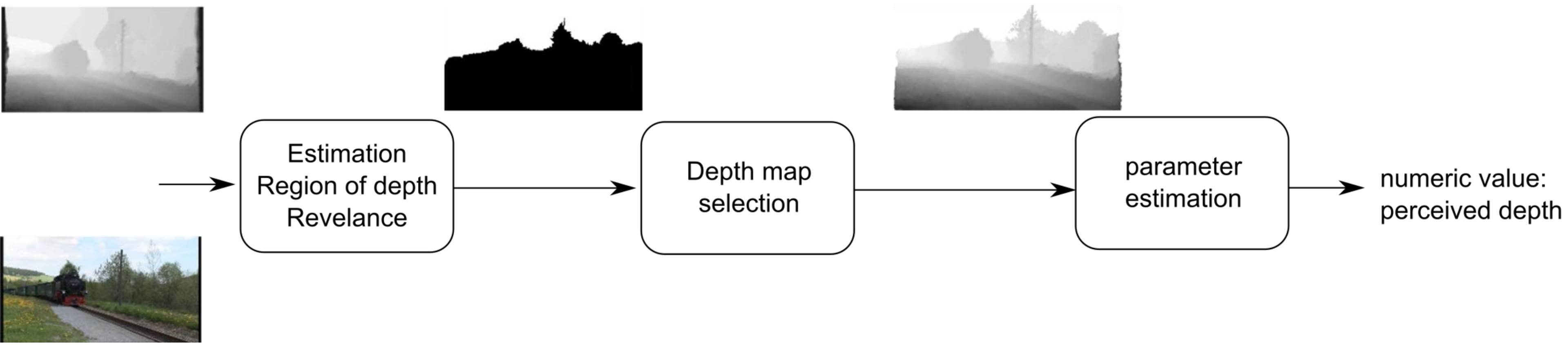
Instrumental evaluation of depth cues

- In this study 5 depth cues are considered:

- Binocular depth cues:
  - Binocular disparities

- Monocular depth cues:
  - Linear perspective
  - Blur from defocus
  - Motion parallax
  - Texture gradient

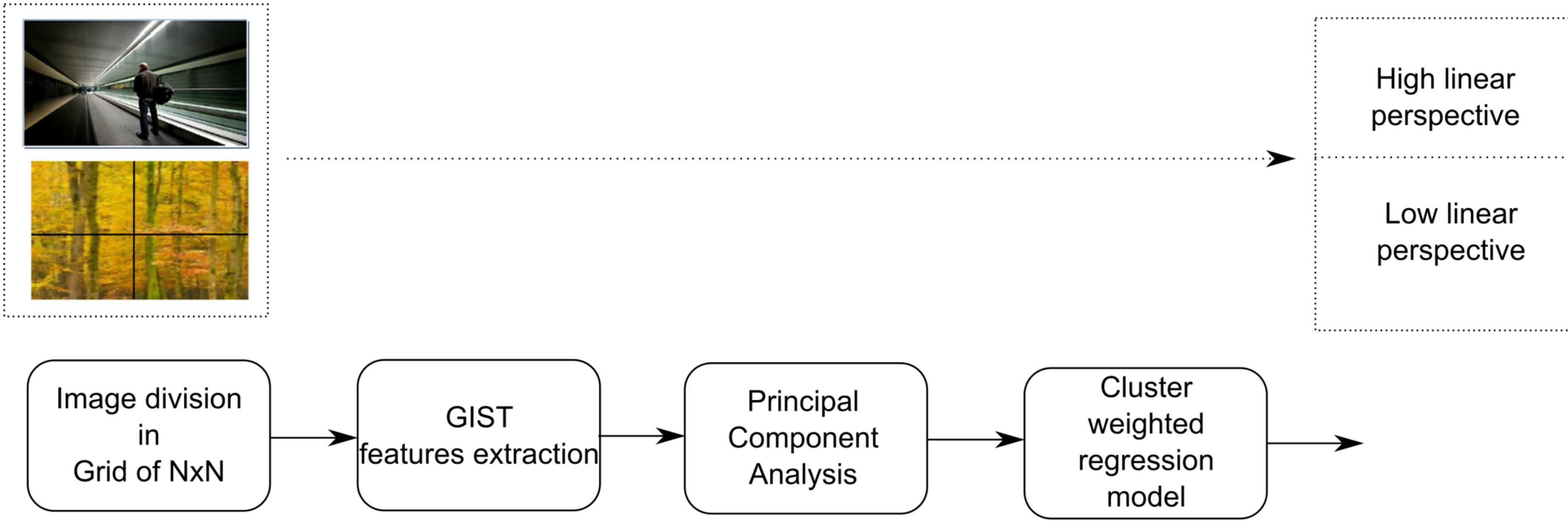
Binocular depth cues – Binocular disparitie



- Estimation of perceived depth from binocular depth cues.
- One numeric value quantifying the amount of binocular depth cue variation

P. Lebreton and A. Raake and M. Barkowsky and P. Le Callet, “Evaluating depth perception of 3D stereoscopic videos”, IEEE Journal of Selected Topics in Signal Processing, vol. 6, pp.710-720, 2012

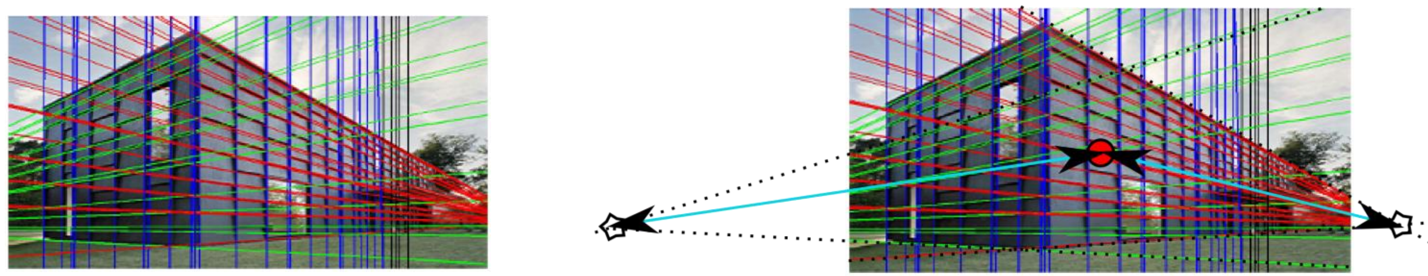
Monocular depth cues – Linear perspective
Global layout property model (GLP)



Antonio Torralba and Aude Oliva, "Depth Estimation from Image Structure", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp.1226-1238, 2002

- Estimation of linear perspective based on the repartition of spatial frequencies within the image.
- One numeric value quantifying the amount of linear perspective

Monocular depth cues – Linear perspective Vanish point model (VPM)



Line segment detection → Line grouping using Hough transform → Vanish point detection using J-Linkage → Heuristic rule based on vanishing points →

Lutz Goldmann and Touradj Ebrahimi and Pierre Lebreton and Alexander Raake, “Towards a descriptive depth index for 3D content : measuring perspective depth cues”, VPQM, 2012

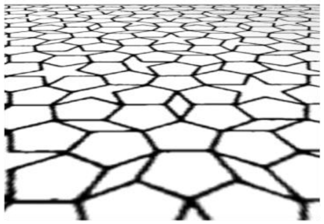
- Estimation of linear perspective based on geometrical properties
  - Linear perspective is defined by:

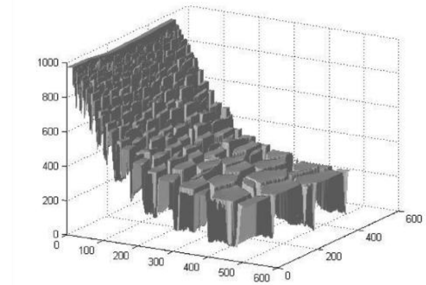
    d, closest vanishing point to the center

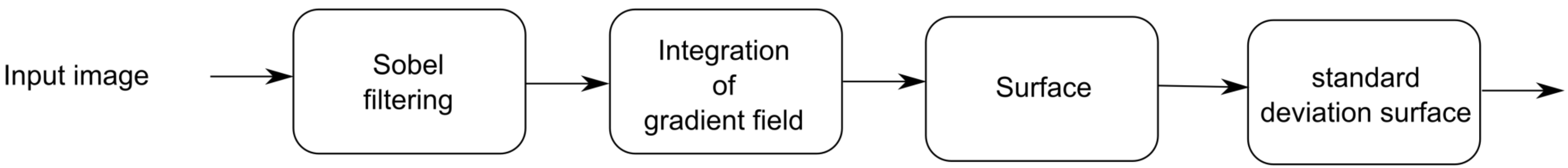
$$L = \frac{1}{1+d}$$

➢ One numeric value quantifying amount of linear perspective

Monocular depth cues – Texture gradient



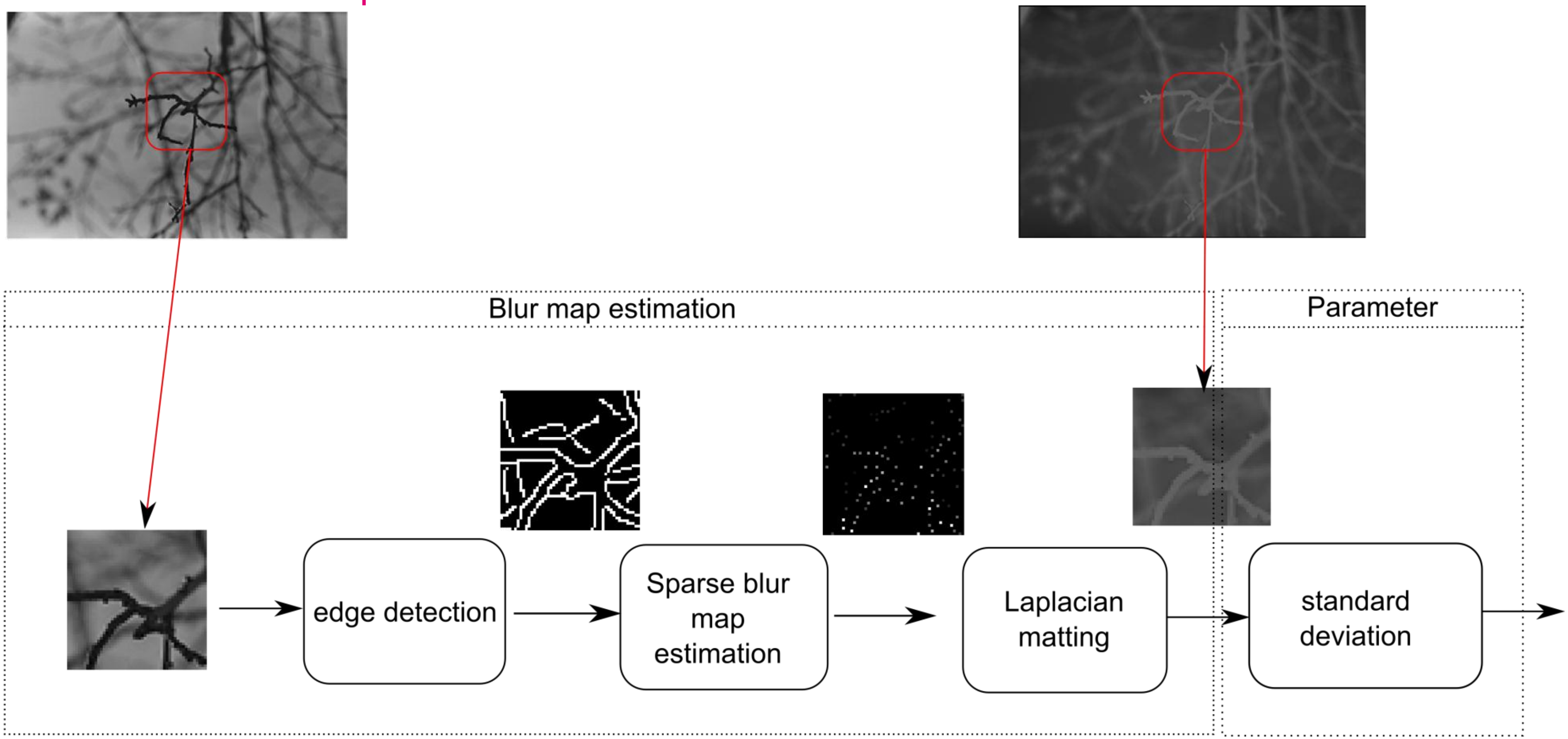




A. Agrawal and R. Chellappa and R. Raskar, “An Algebraic approach to surface reconstructions from gradient fields”, International Conference on Computer Vision (ICCV), 2005

- Estimation of the contribution texture gradient to depth perception
- ➢ One numeric value quantifying the amount of depth variation perceivable from the texture gradient

Monocular depth cues – Defocus blur

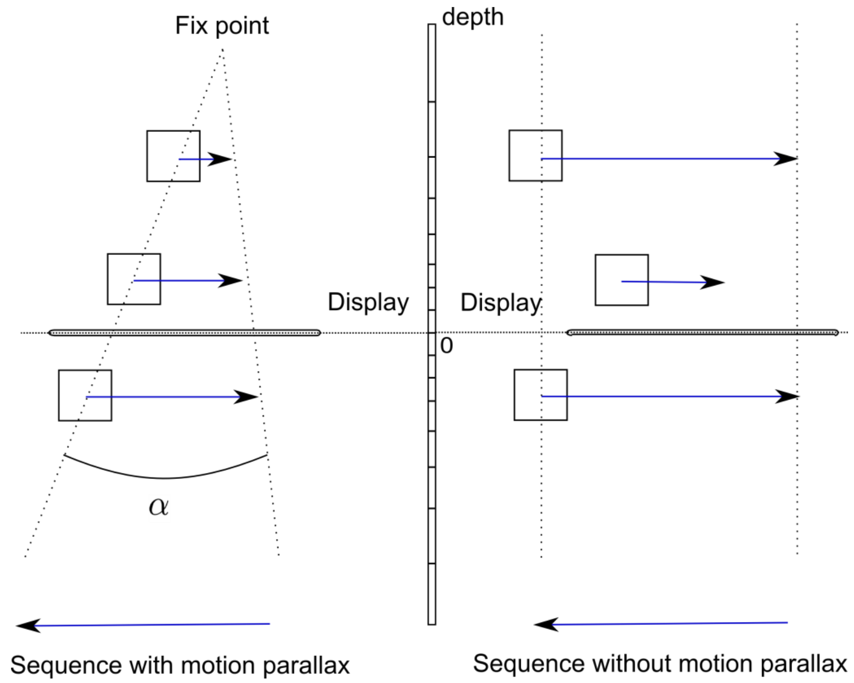


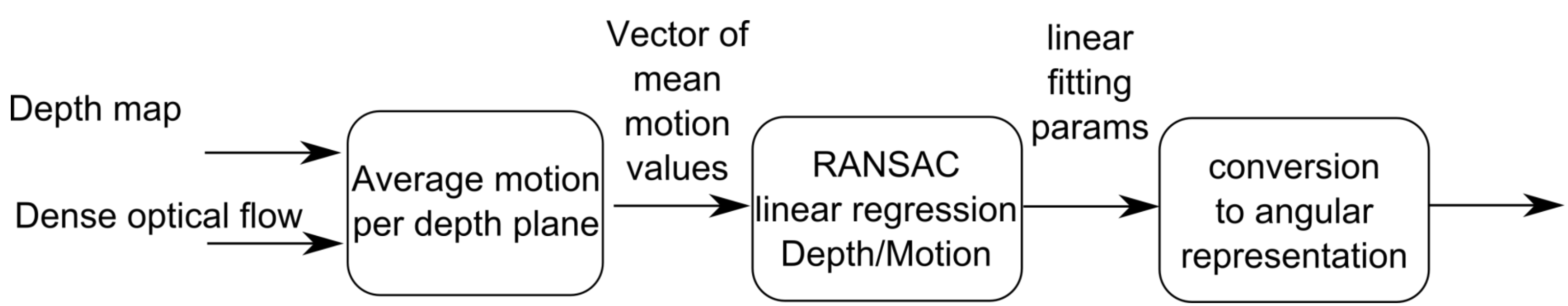
Zhuo, S. and Sim, T., "Defocus map estimation from a single image.", Pattern Recognition, vol. 44, pp.1852-1858, 2011

- ➢ Estimation of one numerical parameter quantifying the amount of blur variation in images

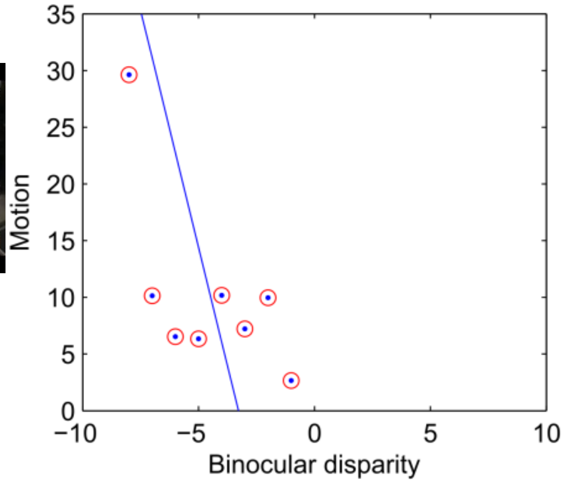
Monocular depth cues – Motion parallax
Proposition of a new metric

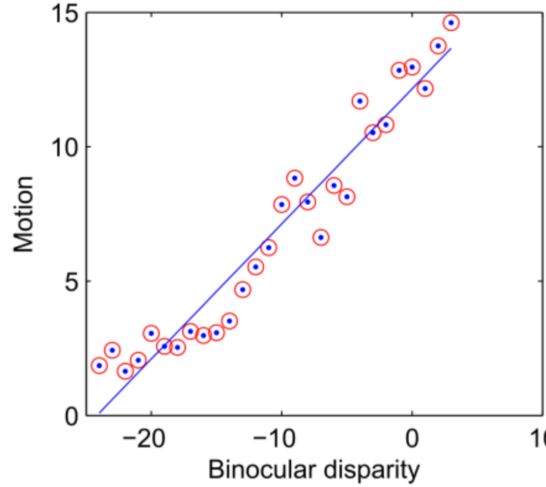
- Estimation of one numeric value quantifying the existence of a gradient of motion in function of depth

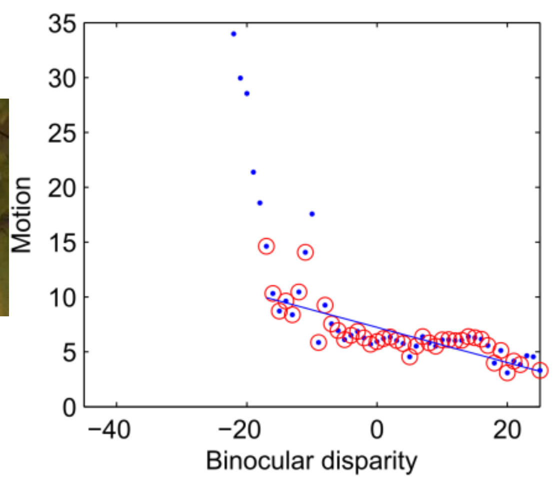


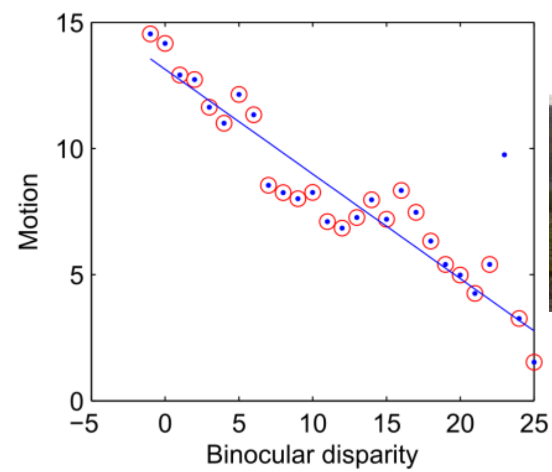


Monocular depth cue – Motion parallax
Results







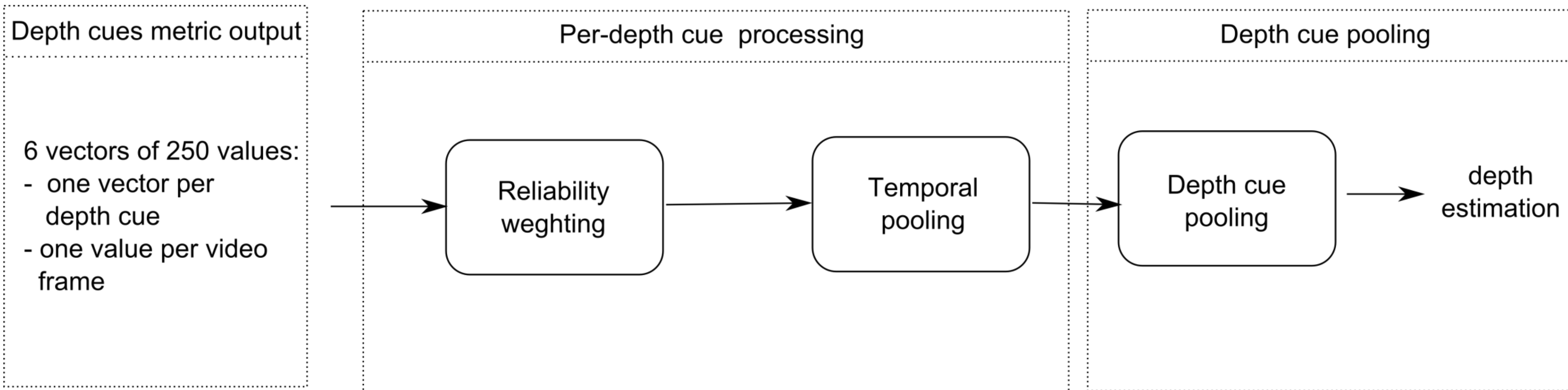


Outline

- Motivation
- Evaluation of monocular and binocular depth cues
- **Depth cues reliability and pooling**
- Conclusion

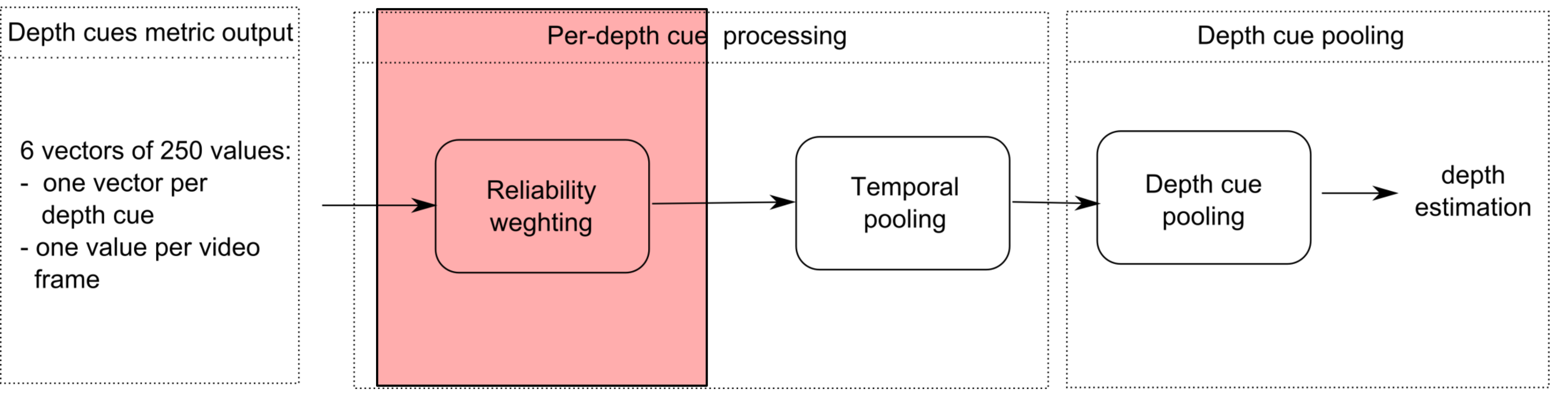
Depth cues reliability and pooling

- To perform depth cue pooling two main aspects must be considered:
  - All depth cues have different relevance
  - All depth cues metric have a specific application scope



- The application scope is critical: monocular and binocular depth cue evaluation is complex and individual metric may fail!
- It is proposed to evaluate the confidence in each metrics.

Step 1: Reliability estimation



Maximum likelihood estimation model

Objective:

- Estimation of the reliability of objective evaluations

Hypothesis:

- It is expected that, at least on a small temporal window, objective metrics should have consistent values.
- Proposition: maximum likelihood estimation

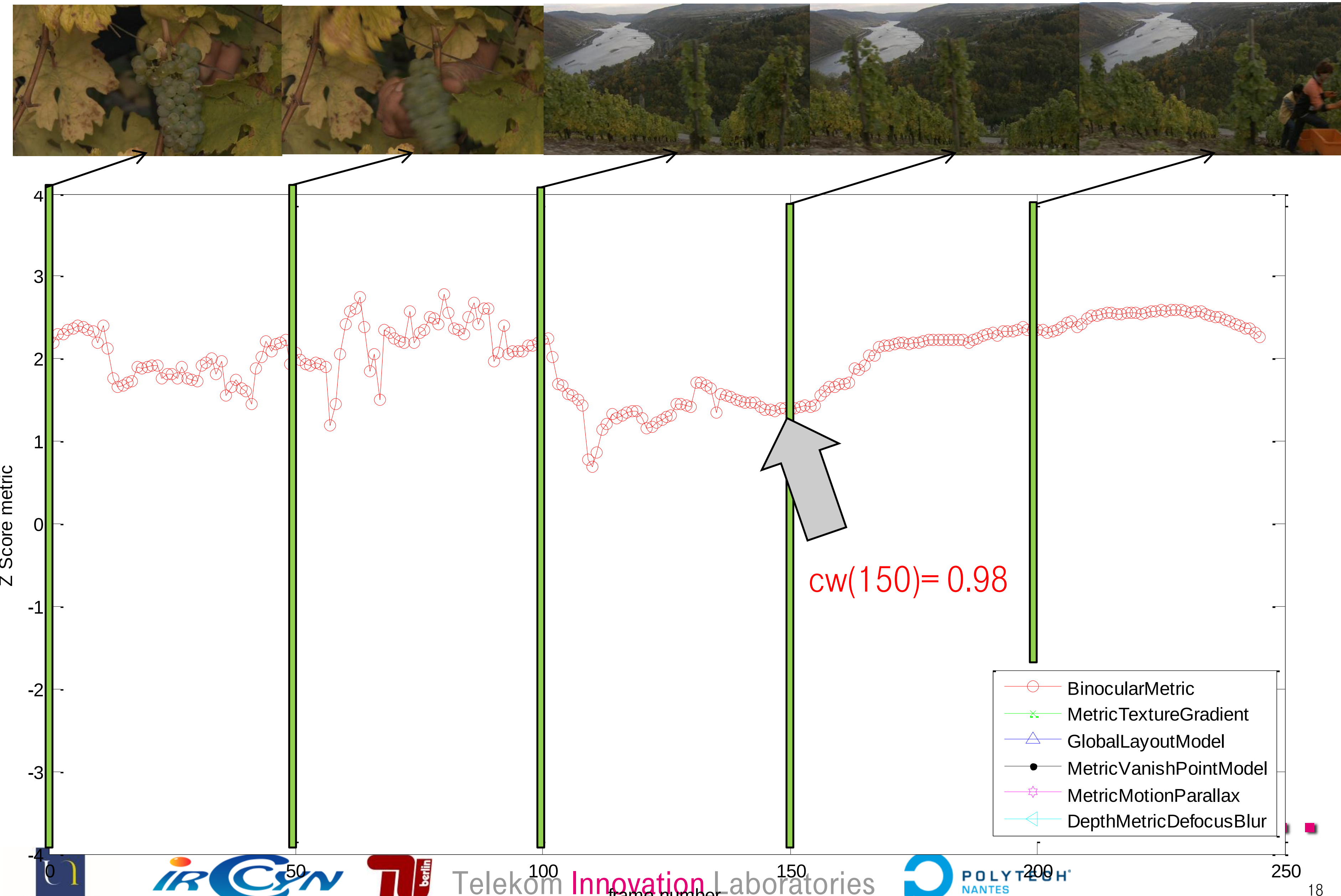
$$GD = \sum_{k=1}^{ND} cw_{k,w} \times DC_k \qquad cw_{k,w} = \frac{\sum_{i=1,i\neq k}^{ND} \sigma^2_{DC_{i,w}}}{\sum_{i=1}^{ND} \sigma^2_{DC_{i,w}}}$$

$\sigma^2_{DC_{i,w}}$ the standard deviation of parameter values on a temporal window **w** around **i**

Limitation:

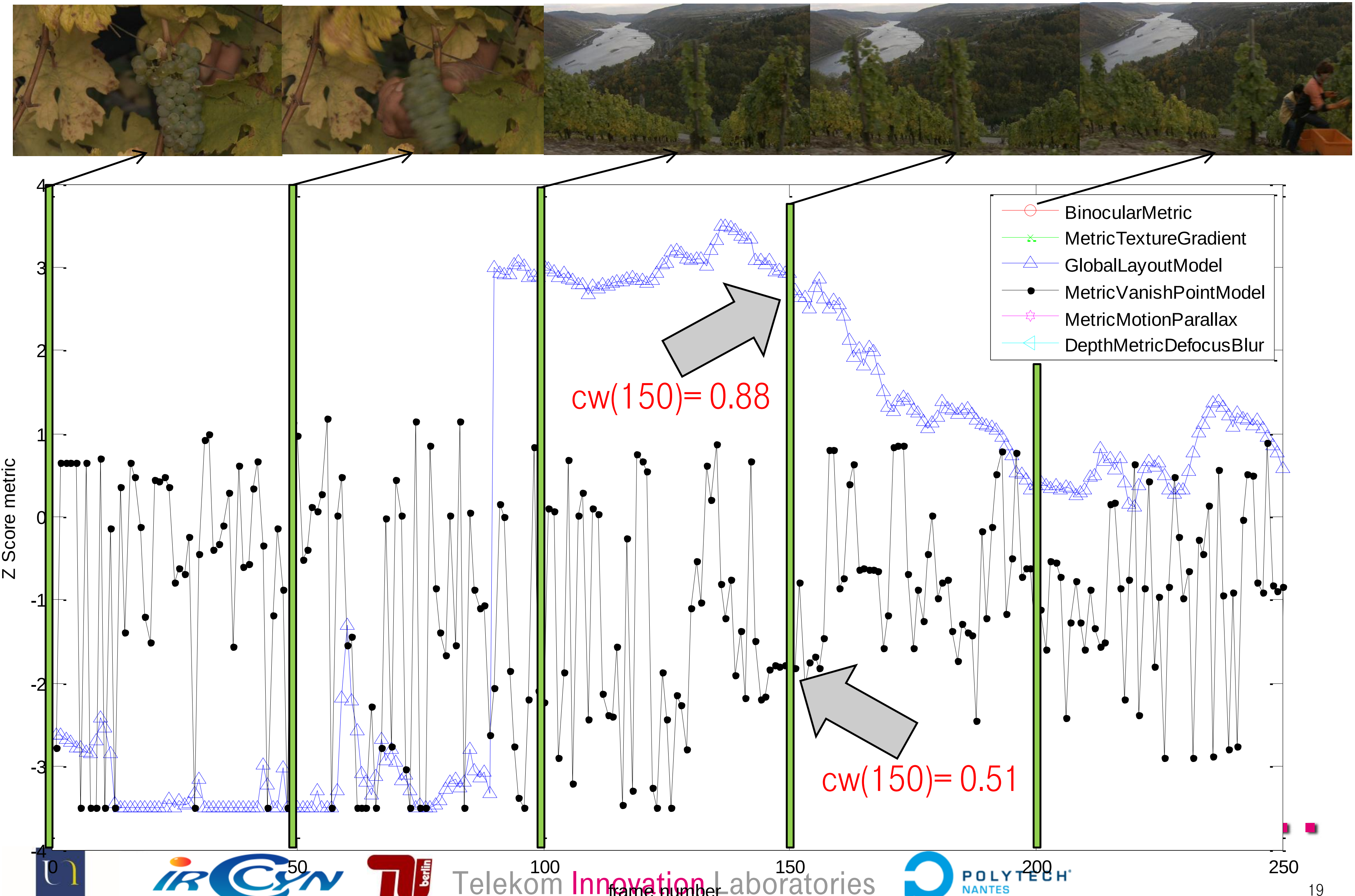
/!\ The consistency of each individual metric is only one factor of the reliability evaluation: stability can be achieved with incorrect depth estimation → metric reliability under study

Verification of hypothesis on one sequence

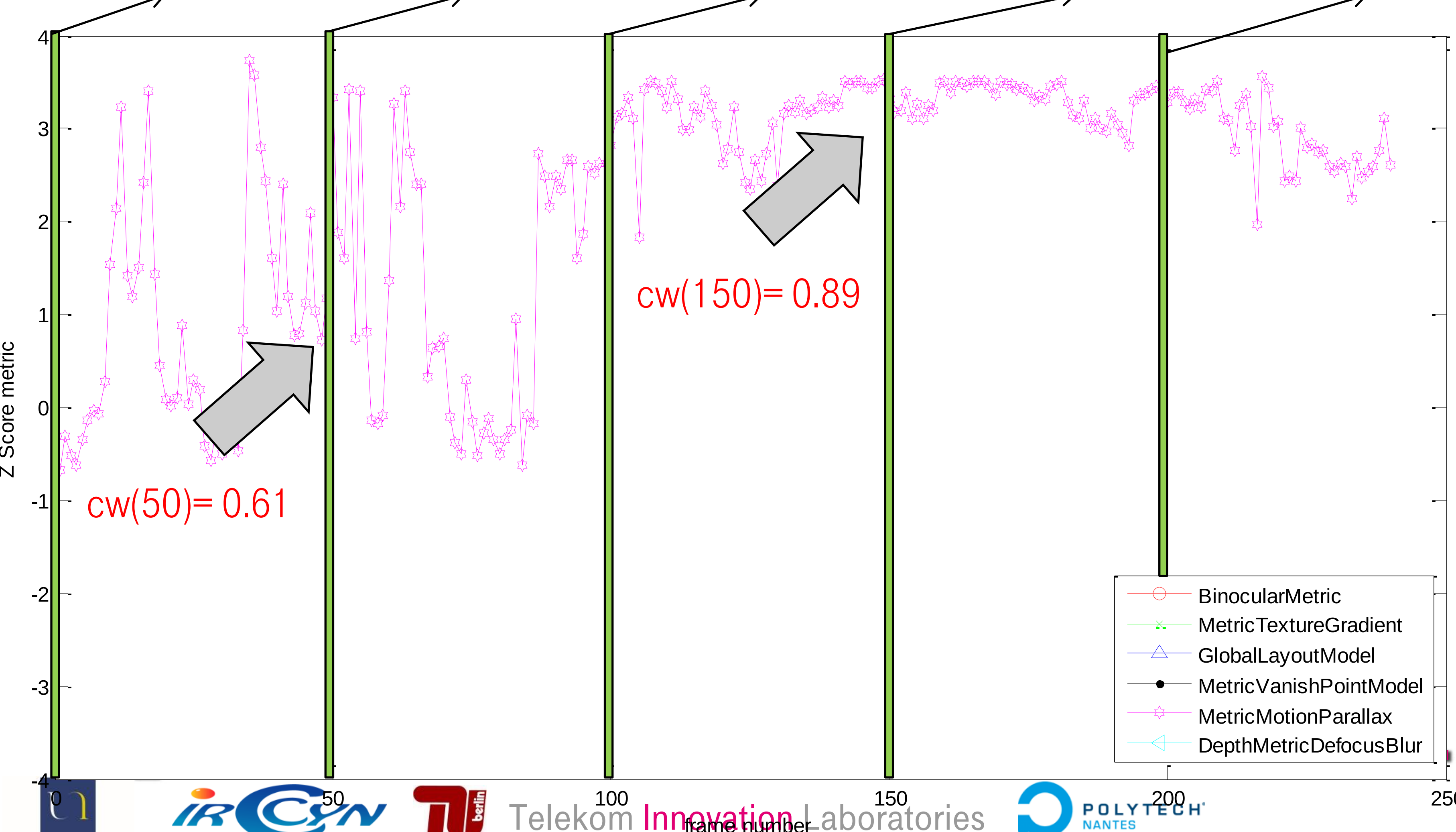


UNIVERSITE DE NANTES IRCCyN Telekom Innovation Laboratories POLYTECH NANTES

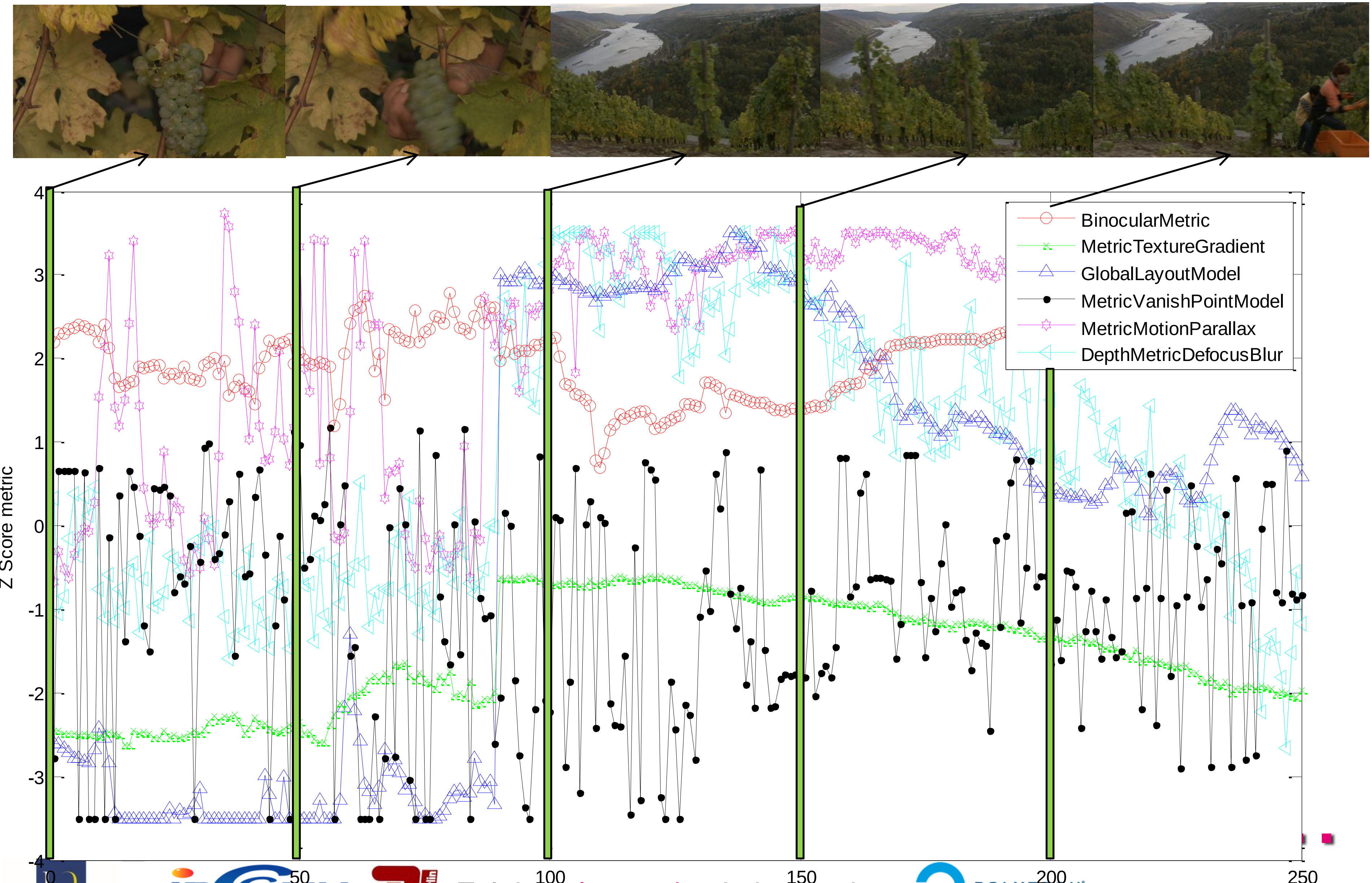
Verification of hypothesis on one sequence



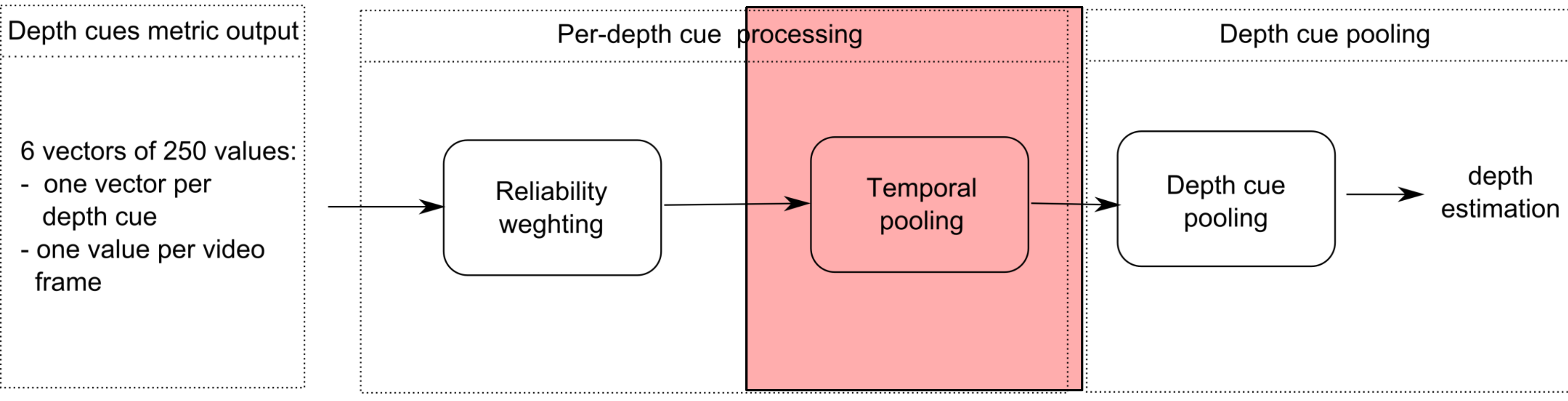
Verification of hypothesis on one sequence



Verification of hypothesis on one sequence



Step 2: temporal pooling



Temporal pooling of depth cues

Hypothesis

- Previous work has shown that temporal pooling should be done with a Minkovsky 4 norm

$$DCV_k = \{v \in DC_k / cw_{k,w}(i) < th\}$$

$$DC_k = \frac{1}{\#DCV_k} \sqrt[4]{\sum_{i=1}^{\#DCV_k} DCV_k(i)^4}$$

- With DC_k the value of the depth cue **k** over the time

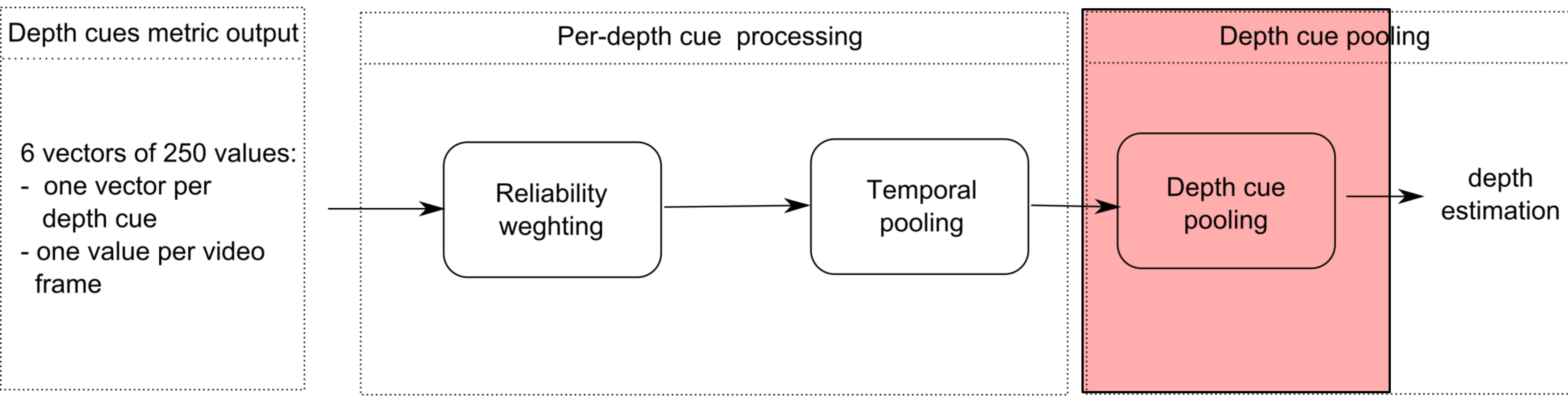
  $cw_{k,w}(i)$ the confidence value of the i^{th} score based on a temporal window **w**
- Depth cues scores being normalized to Z-score. Mean and standard deviation determined based on all scores available.

Subjective database

- video sequences of 10s length (example are shown above)
- 3d synthetic and natural videos
  - Sequences with no visible compression artefact
  - Evaluated by 24 observers with ACR on a 5 grade scale on:
    - Depth (depth in the scene is: very strong, strong, medium, low, very low)
    - Quality of experience (excellent, good, fair, poor, bad)
    - Visual discomfort. (much more comfortable than 2D, more comfortable than 2D, as comfortable as 2D, less comfortable than 2D, much less comfortable than 2D...)
  - Test lab. environment according to ITU-R. BT.500
  - Sequences evaluated on a 24“ PC display with a Native resolution of 1920x1080

➢ **The proposed algorithm attempt modeling the depth score of this experience**

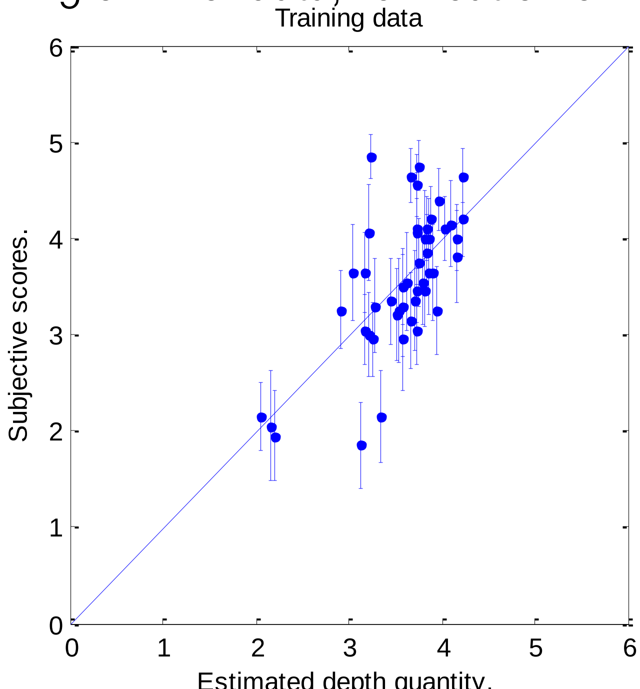
Step 3: depth cue pooling

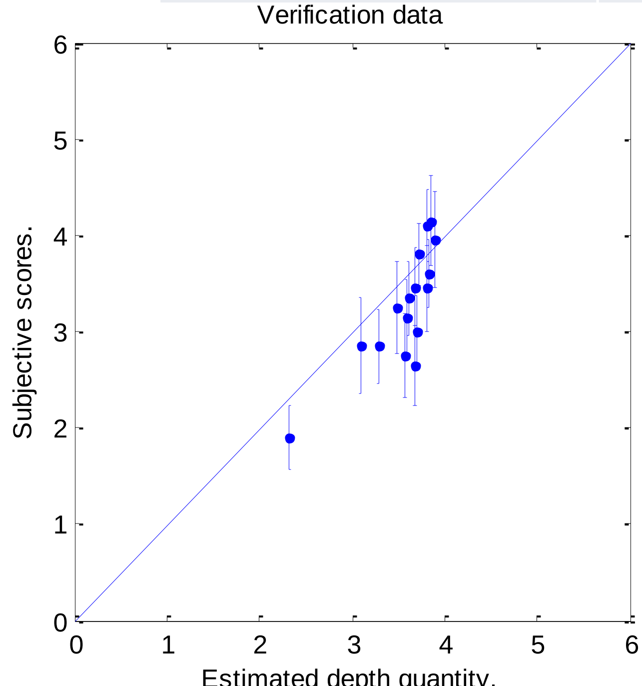


Depth cues pooling / Results

- Final depth cue pooling is done through support vector regression: SVR
  - radial basis function: exp(-gamma*|u-v|^2))
- Results after cross test validation:
  - 4 Training/Verification
  - Training on ¾ of data, verification on ¼

	Ver.	Ver+Tr
Pearson correlation	0.77	0.74
RMSE / 5 Grade scale	0.54	0.47





Outline

- Motivation
- Evaluation of monocular and binocular depth cues
- Depth cues reliability and pooling
- Conclusion

Conclusion

The two parts of our problematic were answered:

- **Depth indicator based on monocular and binocular**
  - A characterization of 3D material according to 5 different depth cues

- **Depth quantity metric based on monocular and binocular depth cues**
  - A weighting based on the reliability of instrumental metric was presented
  - Results were checked based on a subjective database

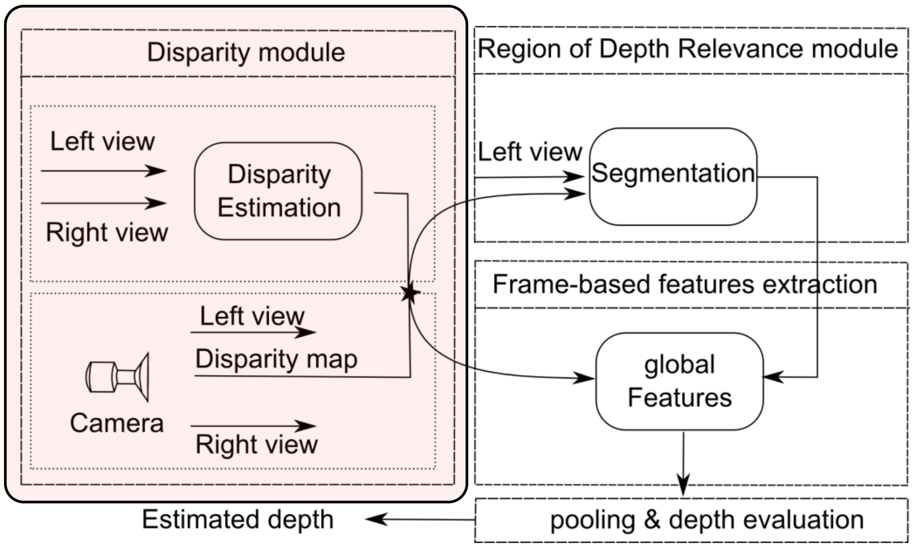
- Limitations:
  - Reliability of depth cues estimation can be found during the evaluation process of depth cues
  - Subjective database is too small to validate the proposed model

Further work

- An in-depth analysis of the metric is needed to better identify case of failure of depth cues
- More subjective testing are required for validating the metric.
- Experiment on images will be done to better identify the depth cue pooling

Backup

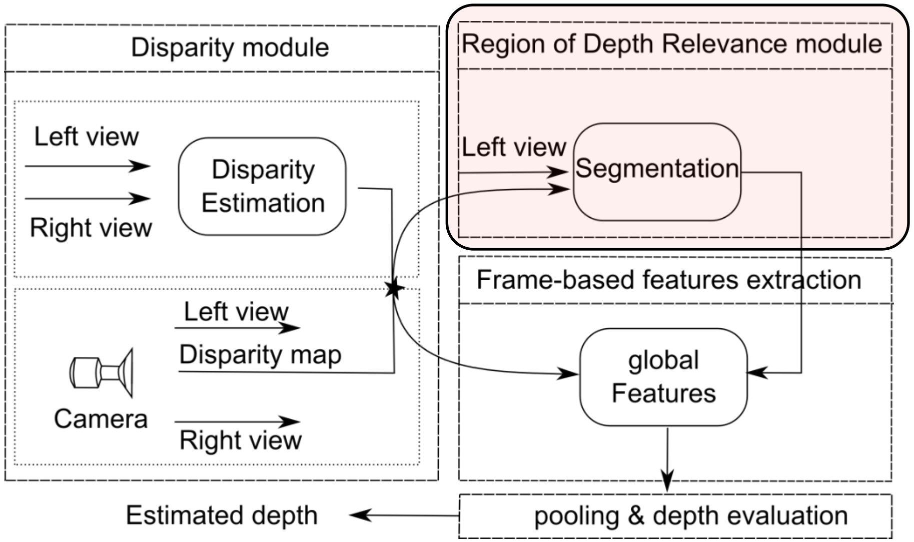
Binocular depth cues – Binocular disparities
General process of the algorithm



P. Lebreton and A. Raake and M. Barkowsky and P. Le Callet, "Evaluating depth perception of 3D stereoscopic videos", IEEE Journal of Selected Topics in Signal Processing, vol. 6, pp.710-720, 2012

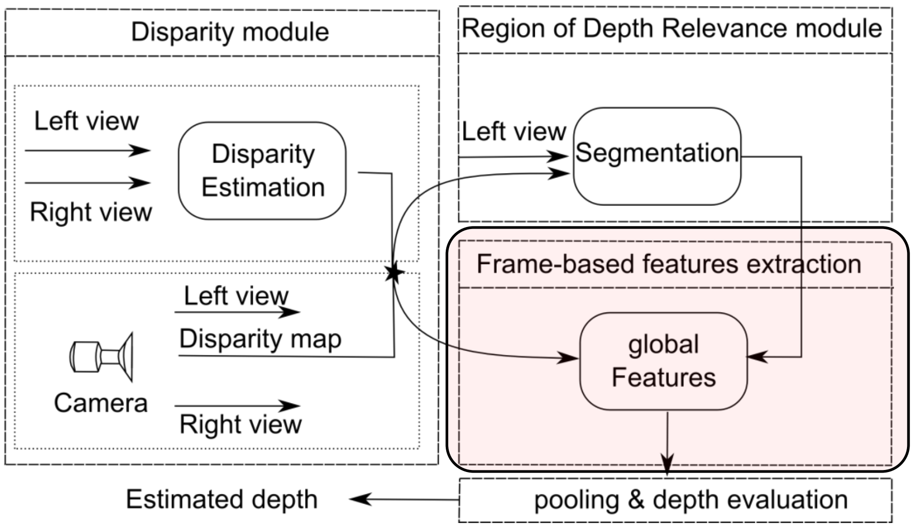
- Estimation of perceived depth from binocular depth cues
- Input:
  - Color image
  - Disparity map (estimated or from camera)

Binocular depth cues – RODR



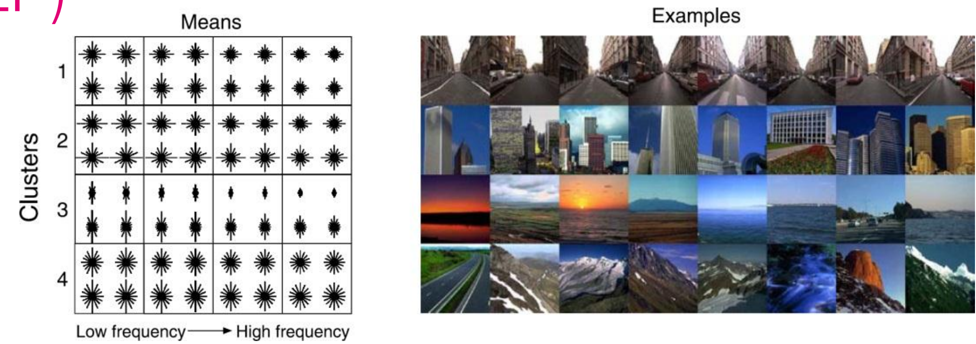
- The image is not entirely useful for evaluating depth
  - ➢ Extraction of the region of depth relevance (RODR)

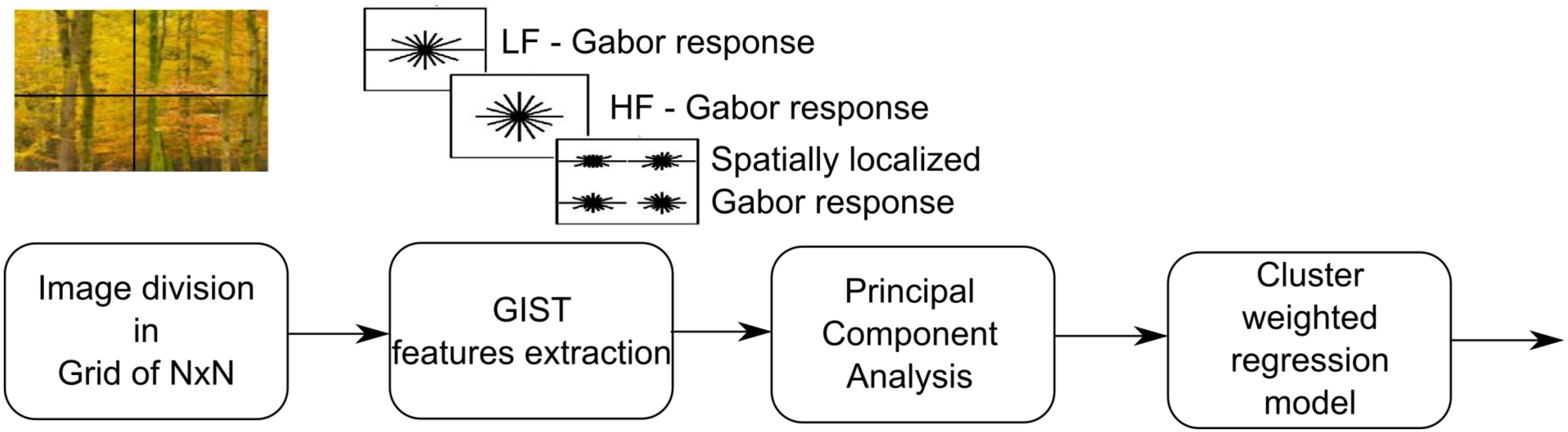
Binocular depth cues – Frame feature extraction



- Estimation of the parameter by frames

Monocular depth cues – Linear perspective
Global layout property model (GLP)

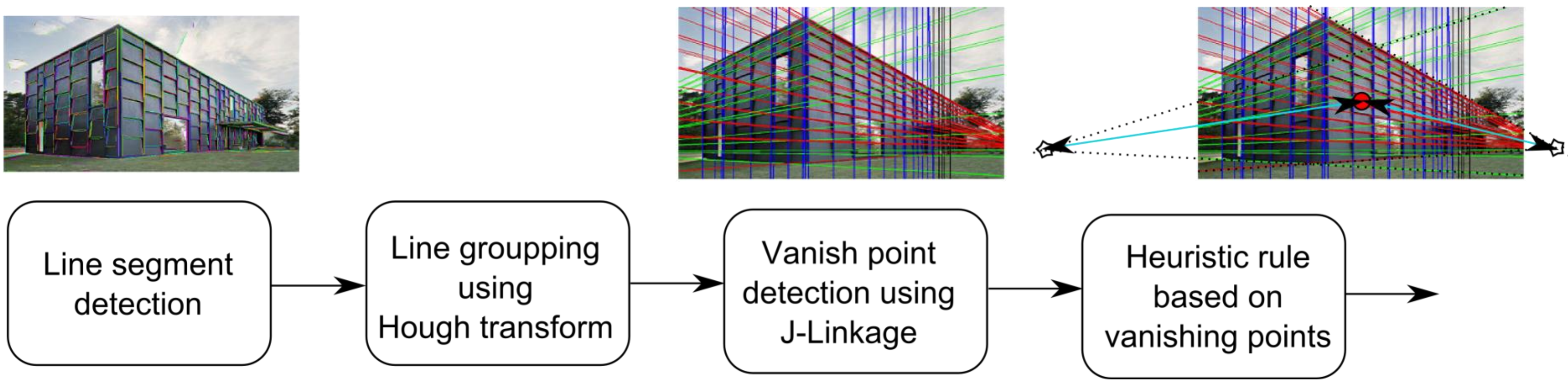




Antonio Torralba and Aude Oliva, "Depth Estimation from Image Structure", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp.1226-1238, 2002

- Estimation of linear perspective based on the repartition of spatial frequencies within the image.

Monocular depth cues – Linear perspective Vanish point model (VPM)



Lutz Goldmann and Touradj Ebrahimi and Pierre Lebreton and Alexander Raake, “Towards a descriptive depth index for 3D content : measuring perspective depth cues”, VPQM, 2012

- Estimation of linear perspective based on geometrical properties
  - Linear perspective is defined by:

    d being the minimum distance to the center of the image

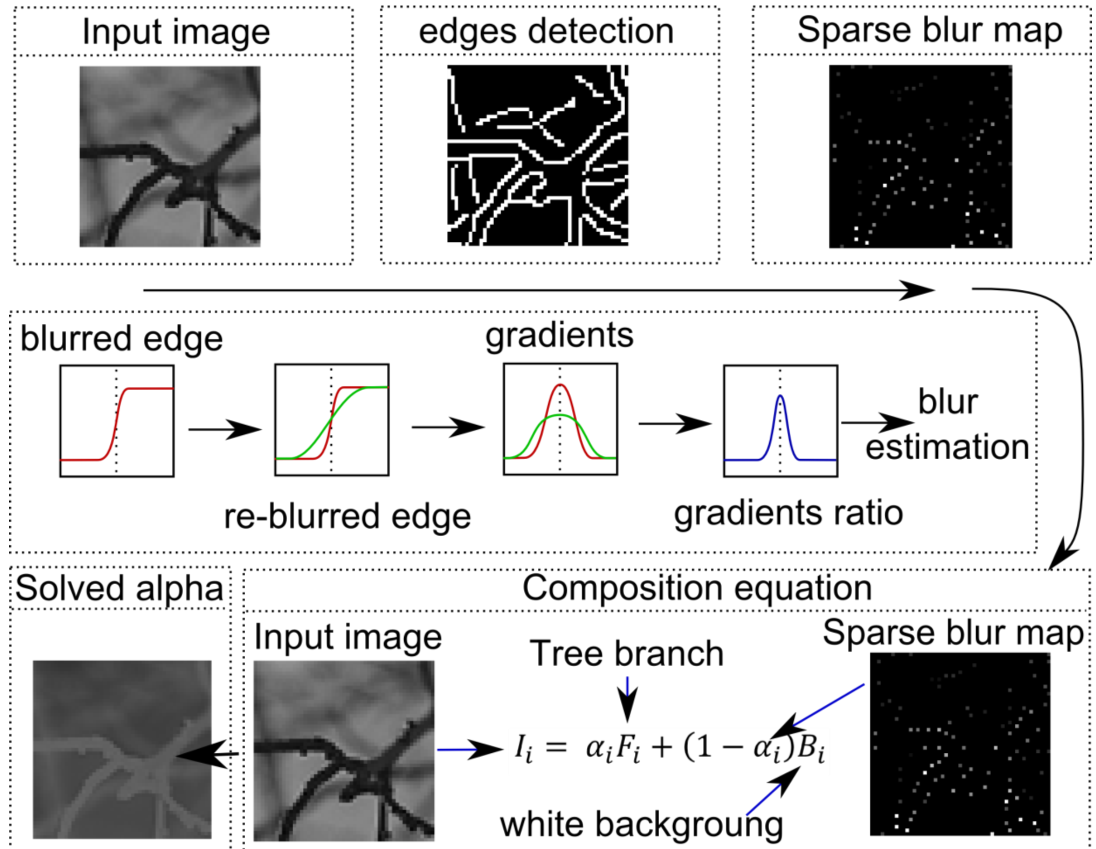
$$L = \frac{1}{1+d}$$

Monocular depth cues – Defocus blur

- Estimation of the contribution of blur to depth in two steps:
  - Blur map estimation
  - Parameter extraction

- metric:

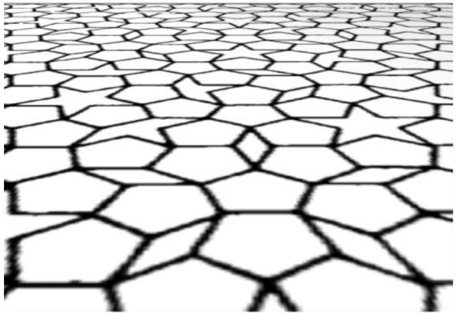
$$Blur = std(blurmap)$$

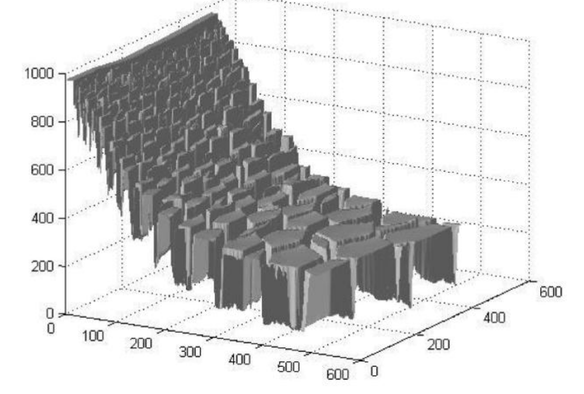


Zhuo, S. and Sim, T., "Defocus map estimation from a single image.", Pattern Recognition, vol. 44, pp.1852-1858, 2011

Monocular depth cues – Texture gradient

A. Agrawal and R. Chellappa and R. Raskar, “An Algebraic approach to surface reconstructions from gradient fields?”, Intenational Conference on Computer Vision (ICCV), 2006





- Let **S(x,y)** be the 2D surface
  - S(x,y) defined on a rectangular grid {x=0,...,W-1 ; y=0,...,H-1}

- Let $p^0 = \frac{\partial s}{\partial x}$; $q^0 = \frac{\partial s}{\partial y}$ be the integrable gradient field of S

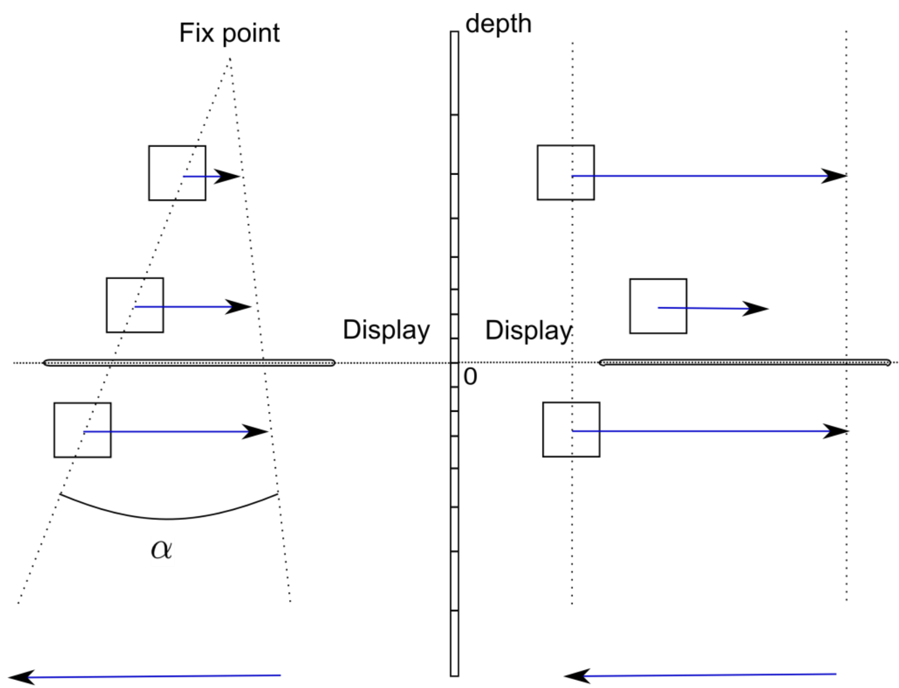
- The estimation of **S**, $\hat{S}$ can be found be found by minimizing the cost function:

$$J(\hat{S}) = (\widehat{S_x} - p)^2 + (\widehat{S_y} - q)^2$$

The Euler-Lagrange equation gives the Poisson equation to solve:

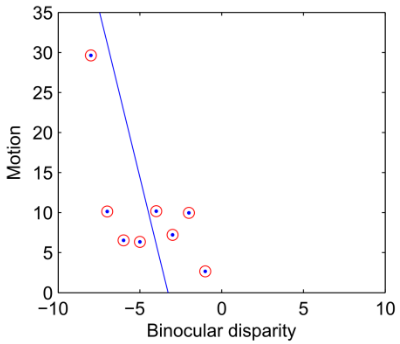
$$\nabla^2\hat{S} = div(p,q)$$

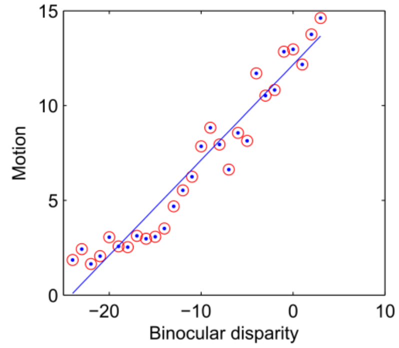
Monocular depth cues – Motion parallax

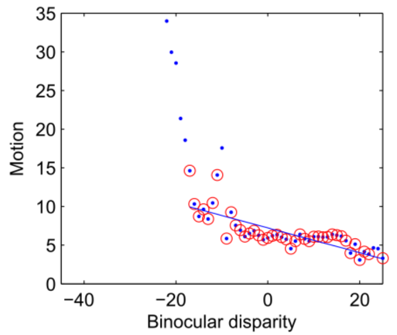


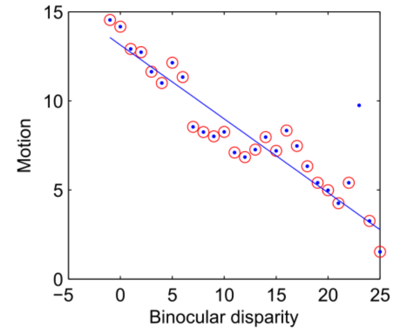
Sequence with motion parallax

Sequence without motion parallax

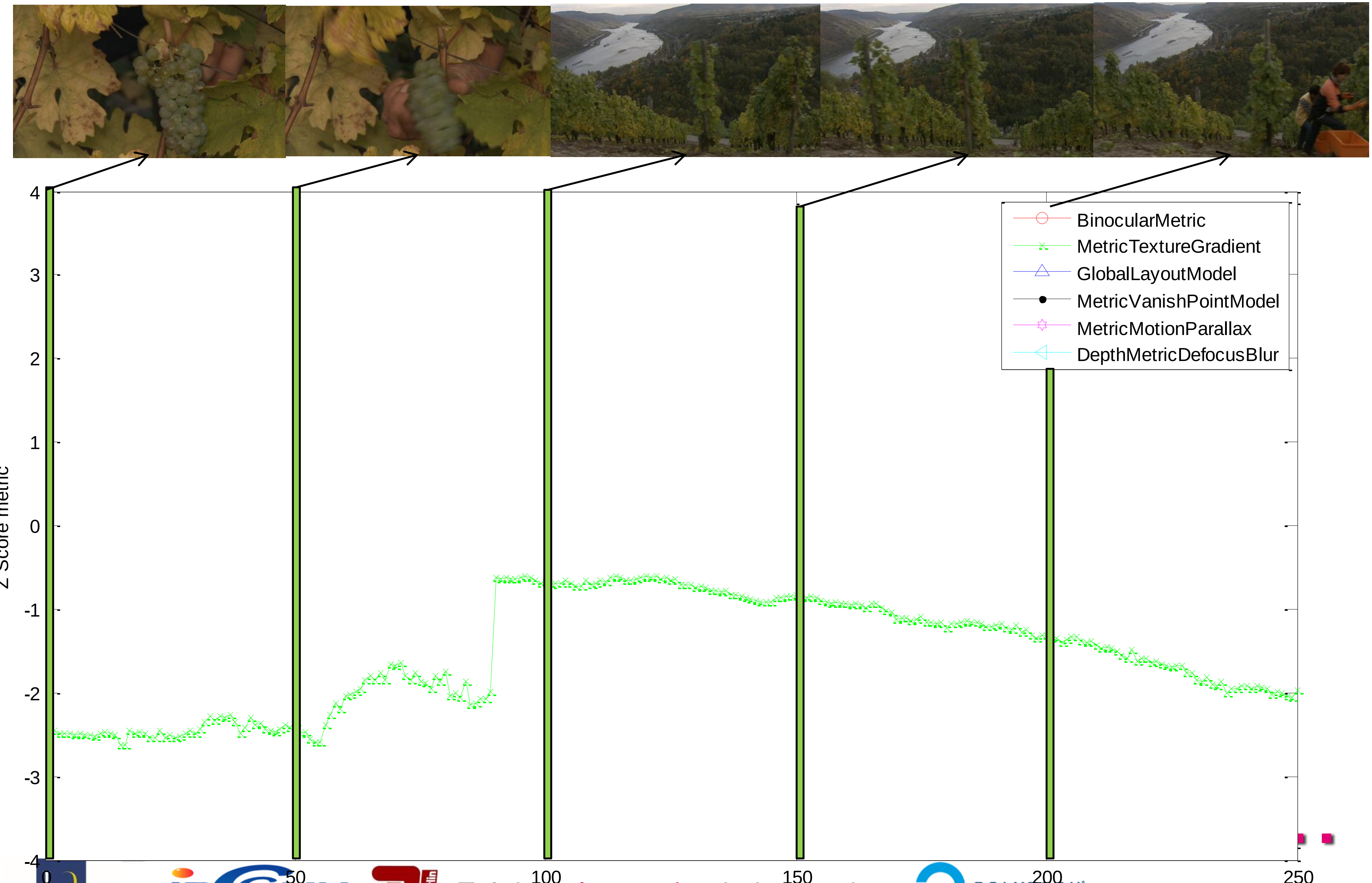




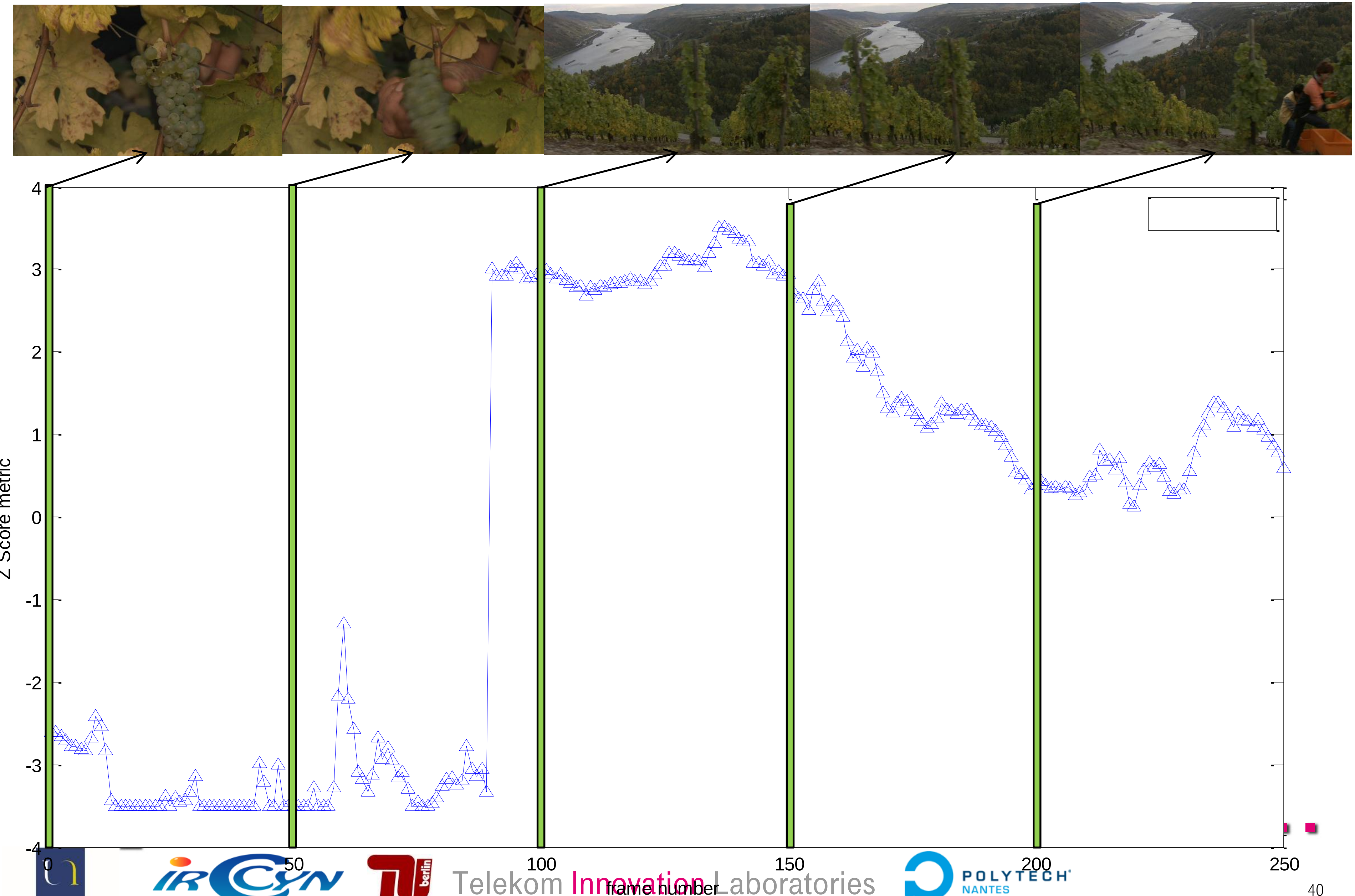




Verification of hypothesis on one sequence



Verification of hypothesis on one sequence



Verification of hypothesis on one sequence