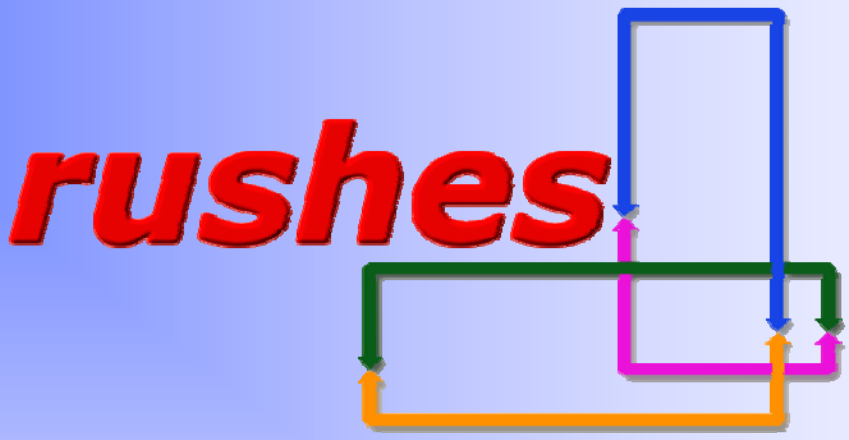


3D Scene Structure Analysis for Semantic Annotation and Retrieval of Unedited Video

Chorus Final Conference, Brussels, 26 May 2009

Dr. Oliver Schreer

Ingo Feldmann, Wolfgang Waizenegger

Fraunhofer Heinrich-Hertz-Institute

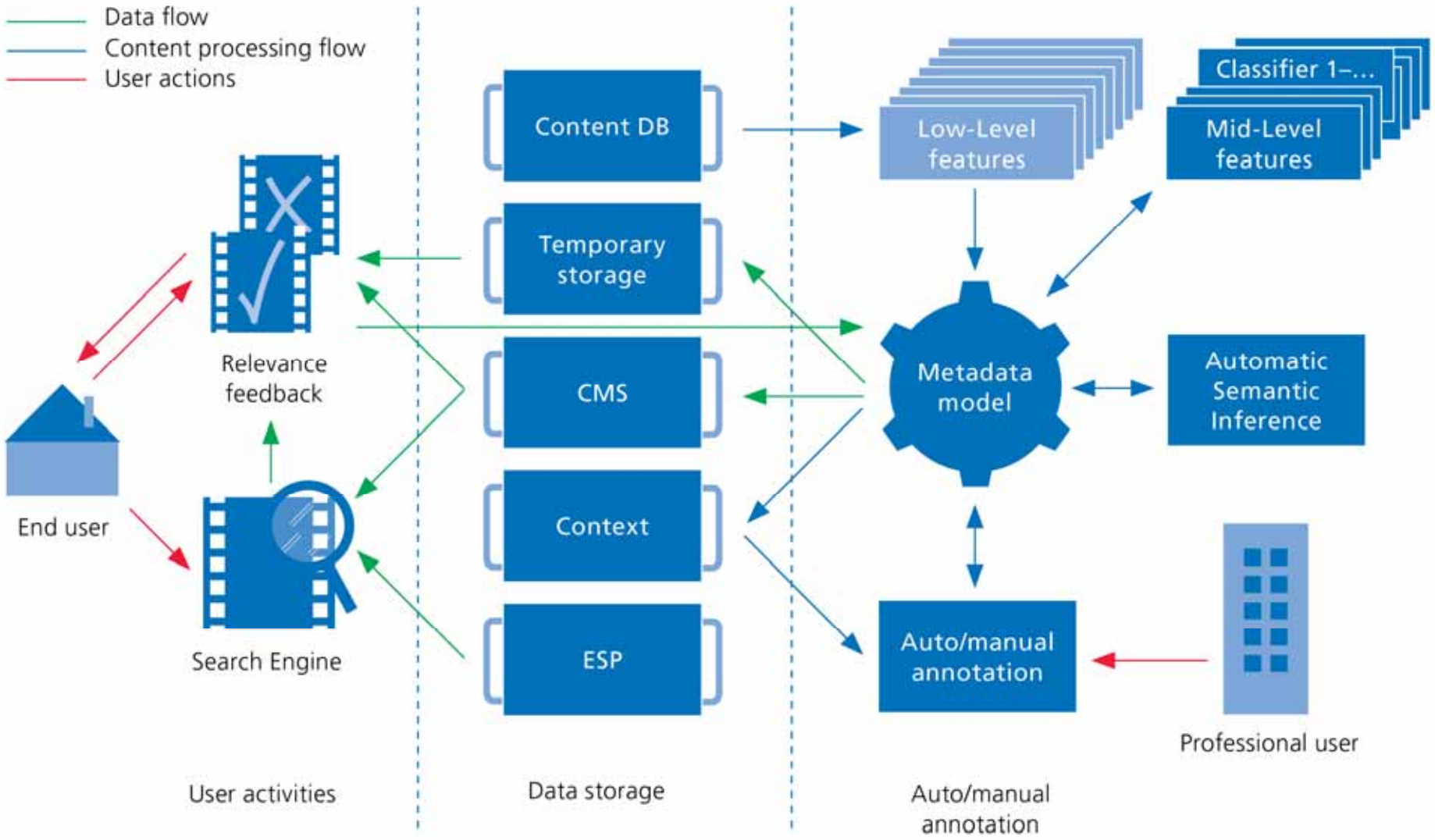
Outline

- RUSHES flashlight
- Motivation and state-of-the-art
- General approach
  - Low level analysis
  - Mid level analysis
  - High level semantic annotation
- Benchmarking and experimental results
- Summary

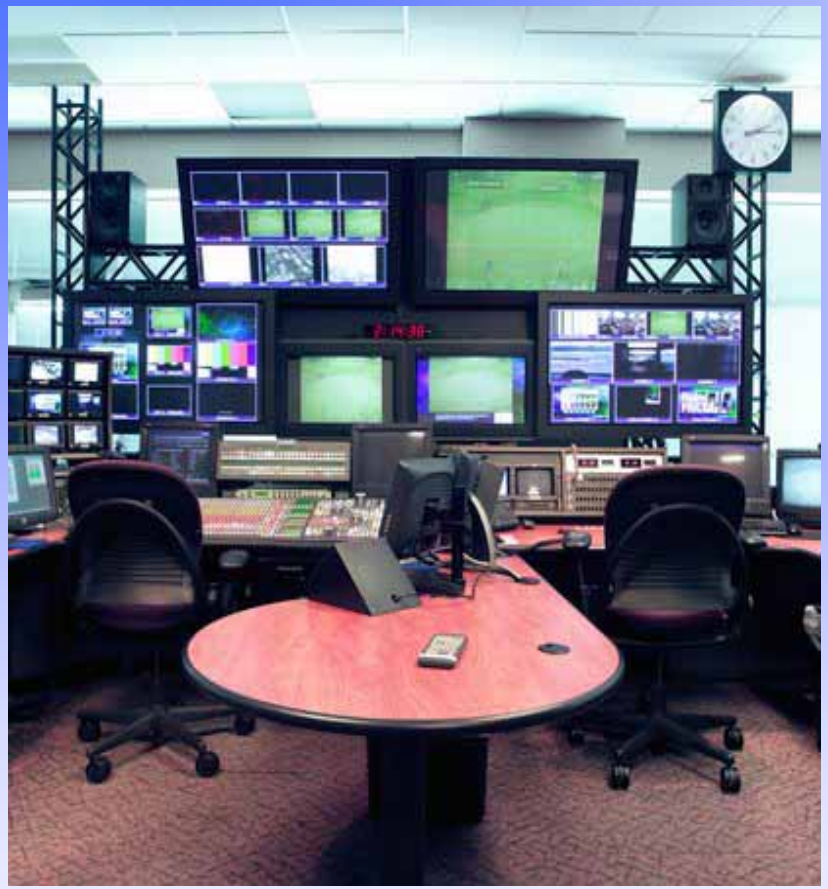
The RUSHES project

- European FP6 funded research project
- 9 partners among 6 countries
- 3 universities, 2 research institutes, 3 companies, 1 SME
- Duration 2.5 years, end of project in July 2009

Components of the RUSHES system



Motivation



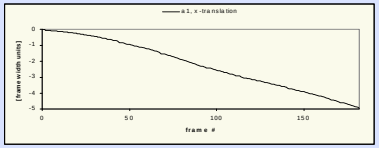
video editing and post processing room of a TV broadcaster

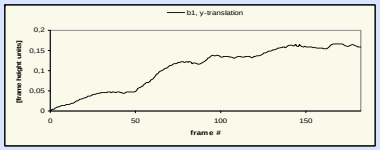
- Target application: Video archives of TV broadcasters and movie producers
- Large amount of unedited raw videos
- Annotation, search and retrieval is in many cases still done manually and text based
- High demand for robust and automatic content-based semantic analysis tools

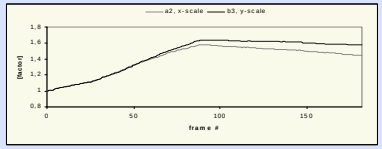
State of the Art

Camera motion analysis

- Global motion estimation
- Motion vector analysis
- 3D camera motion descriptor (defined in MPEG7)







Limitations/drawbacks of global motion estimation

- Motion is usually predicted in 2D
- Scene is modeled by a plane (affine, parabolic models)
- 3D camera motion in relation to the 3D scene is not considered

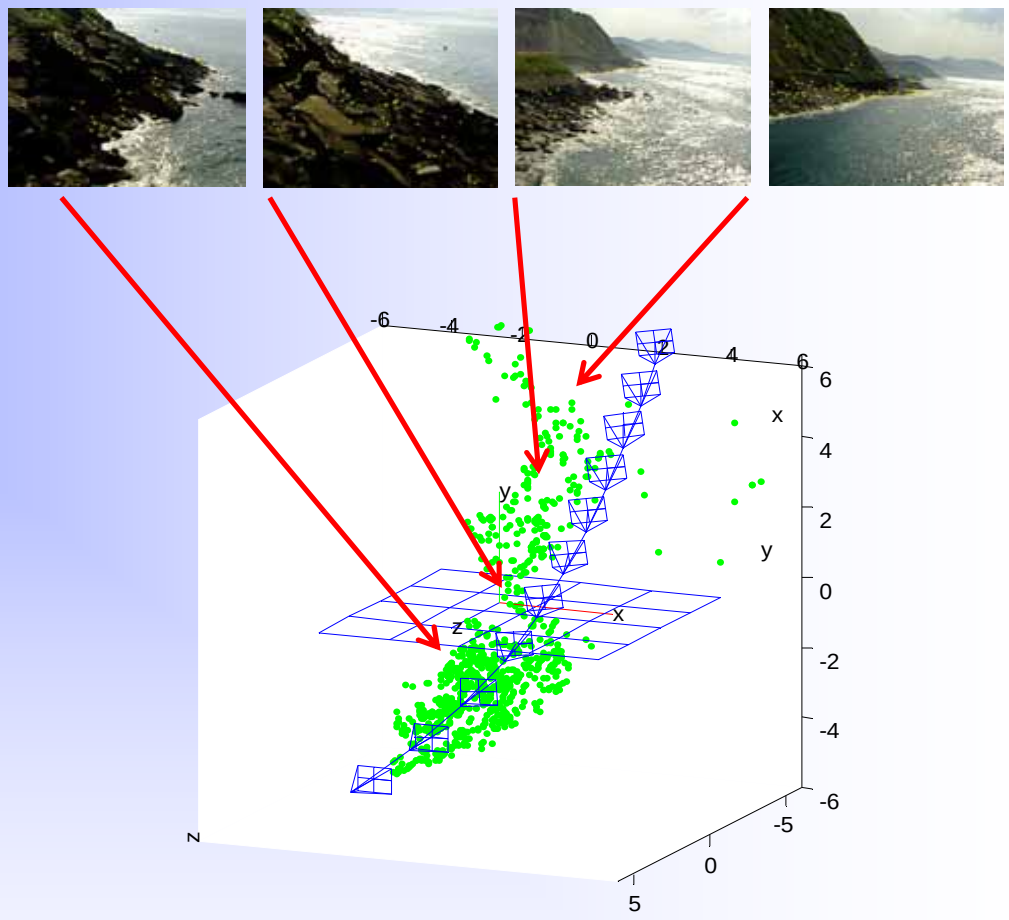
Our approach

Our approach

- Apply state of the art 3D camera self calibration techniques
- Estimate 3D camera motion within a reconstructed 3D scene
- Reconstruct sparse set of robust 3D feature points
- Model and simplify the scene
- Provide interpretation schemes at various semantic levels

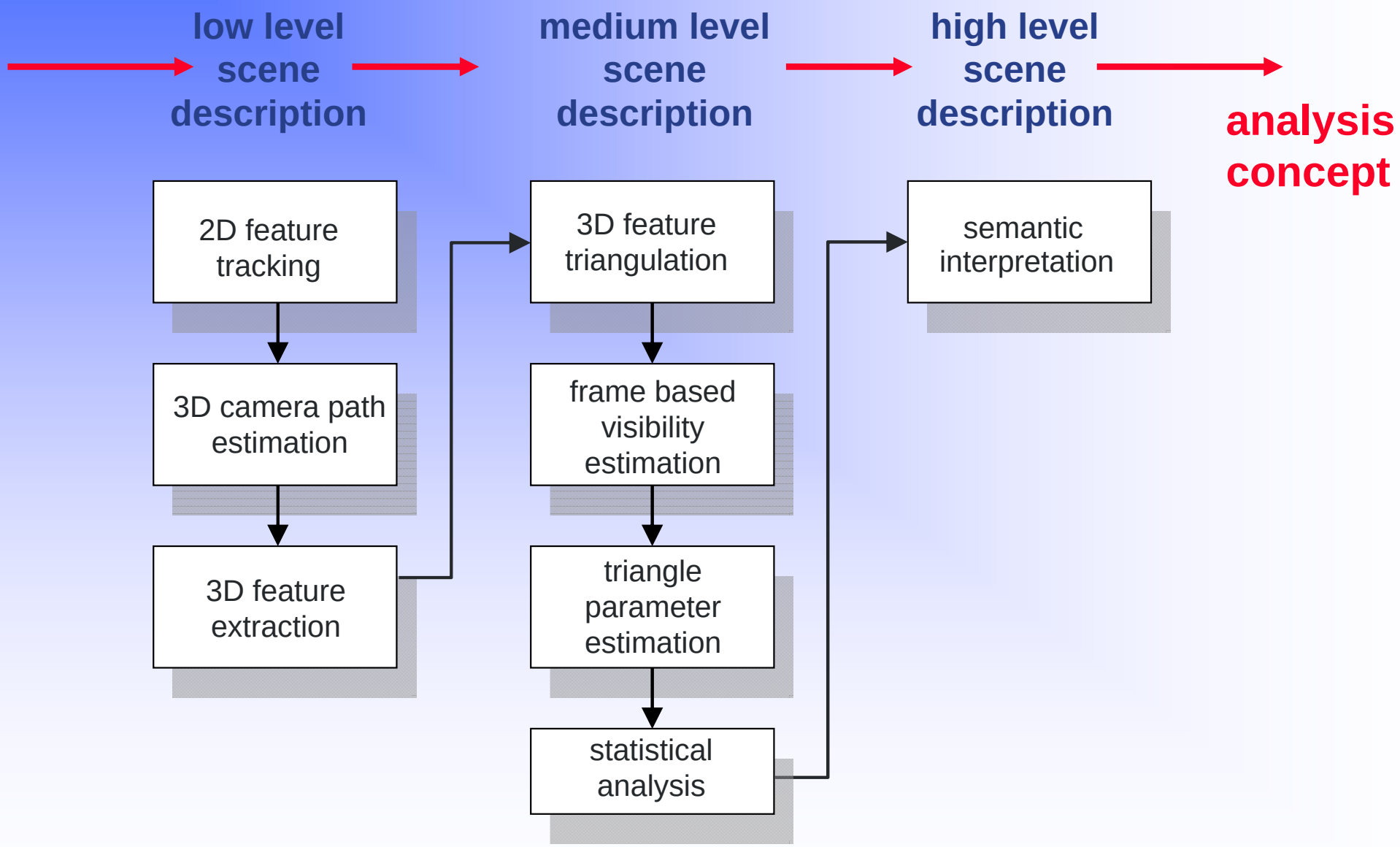
Advantage

- Exploit 3D scene structure and 3D camera movement within the scene



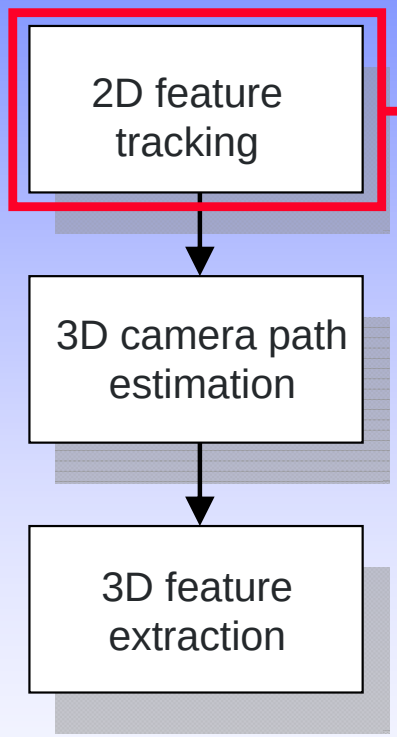
3D camera motion in the reconstructed sparse 3D Scene

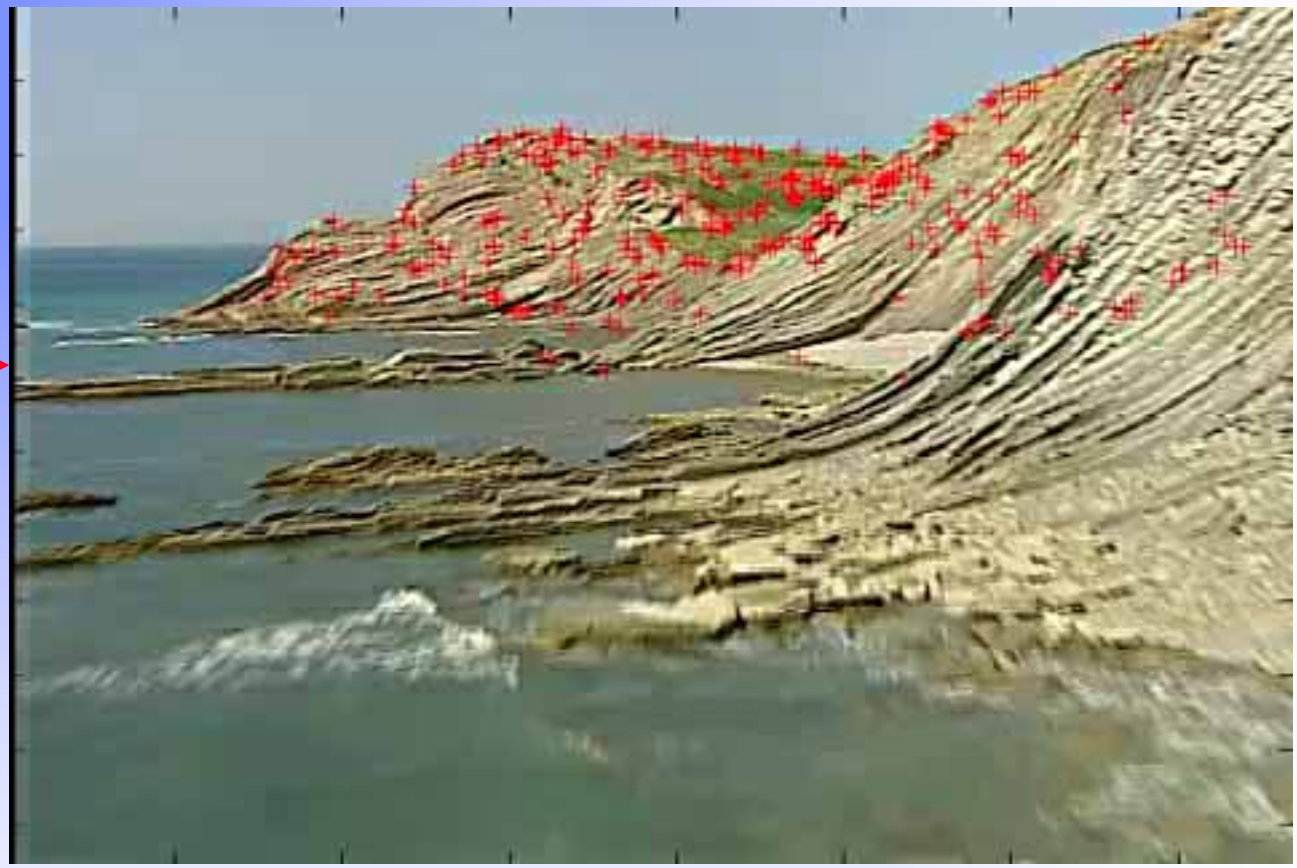
General module overview



Low level metadata analysis

low level scene description

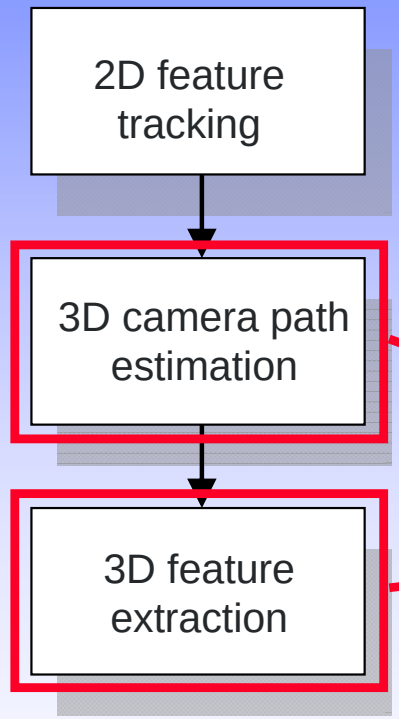


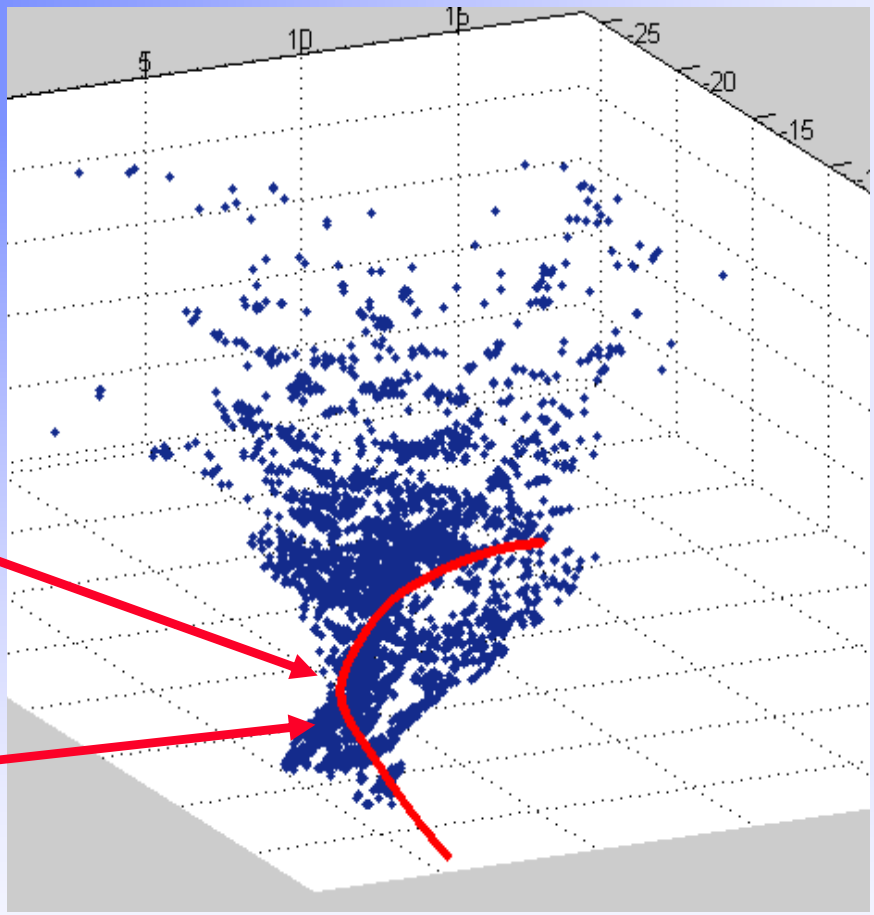


2D feature tracking
- KLT
- SIFT
- boujou + KLT refinement

Low level analysis

low level scene description





reconstructed 3D camera motion and sparse 3D scene model

[Hartley, Zisserman]

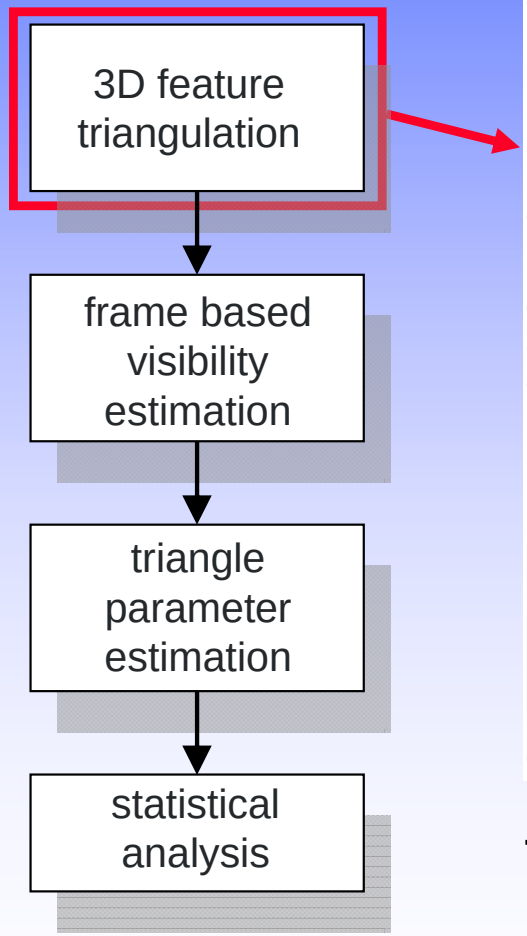
- camera self calibration
- robust 3D feature point extraction

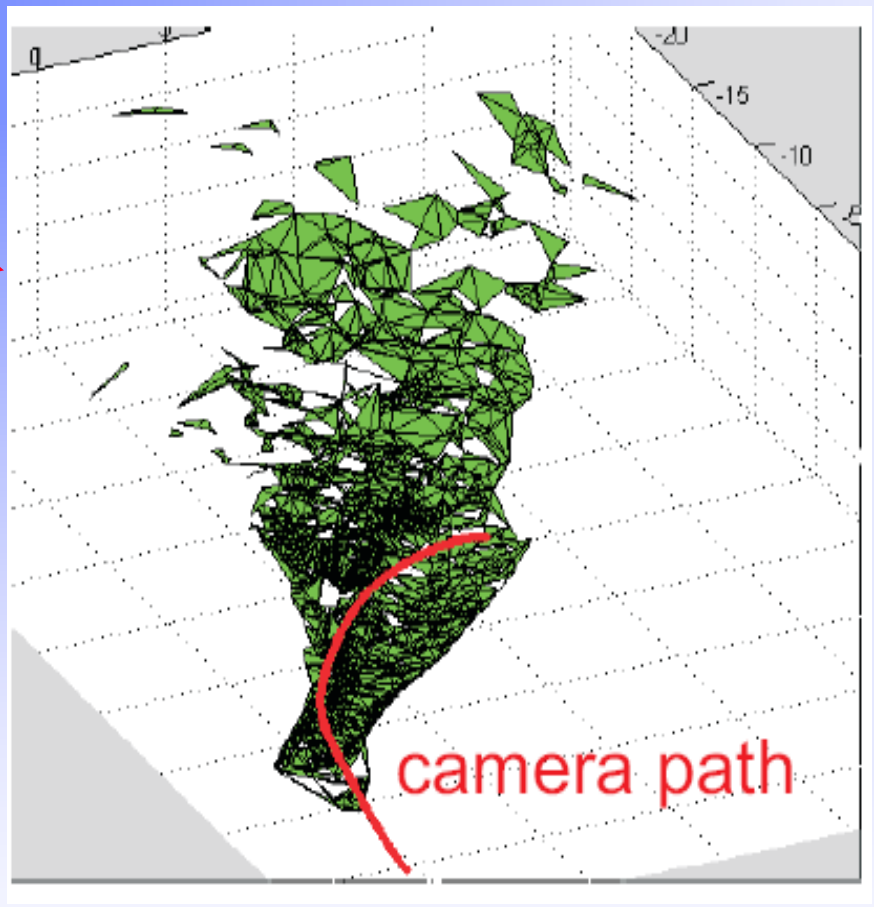
- boujou
- ICARUS

- split + merge of large sequences (25.000 frames)

Medium level analysis

medium level scene description



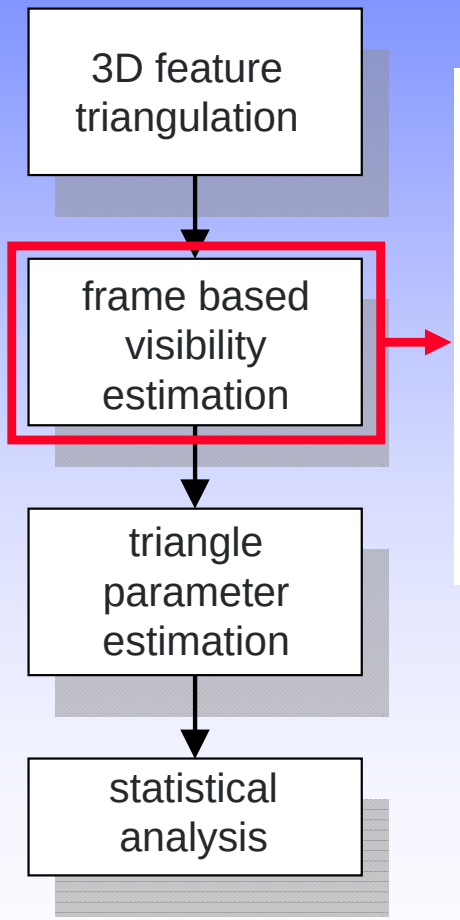


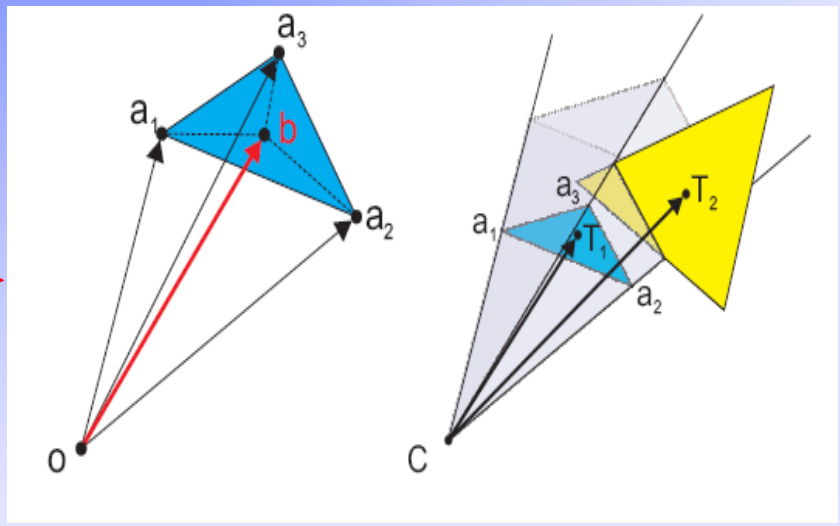
triangulated rough 3D scene model

- PCA of the 3D point dataset
- Subsequent projection along the direction with least variance
- Delaunay triangulation

Medium level analysis

medium level scene description



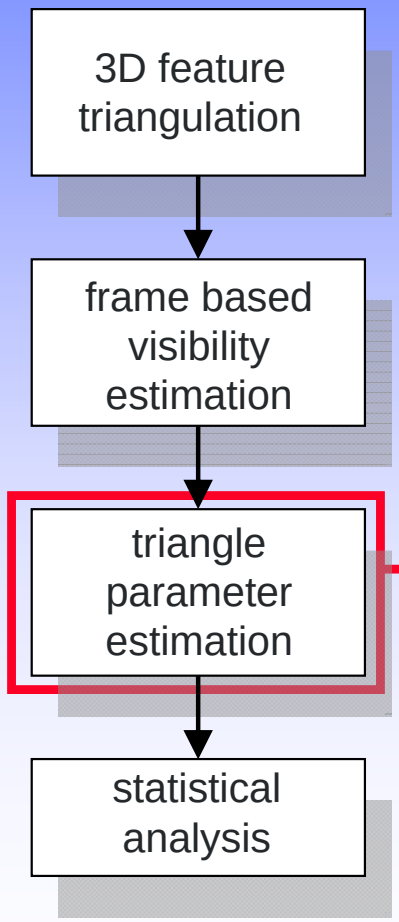


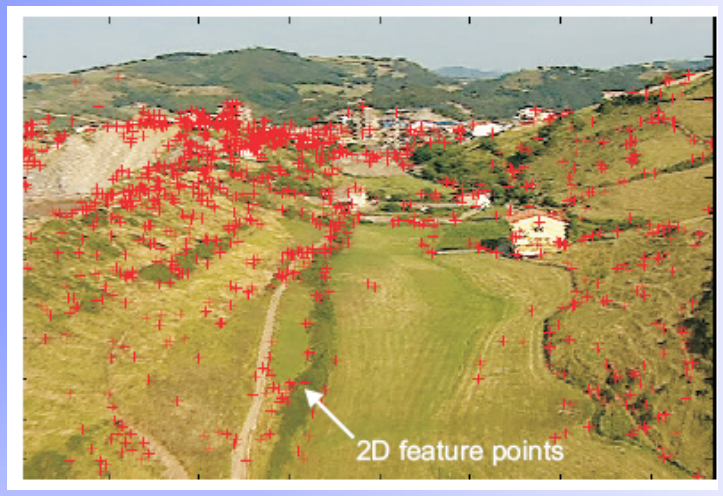
Removal of triangles

- behind the camera
- with less than two visible corners
- occluded triangles by simple ray tracing
- outliers

Medium level analysis

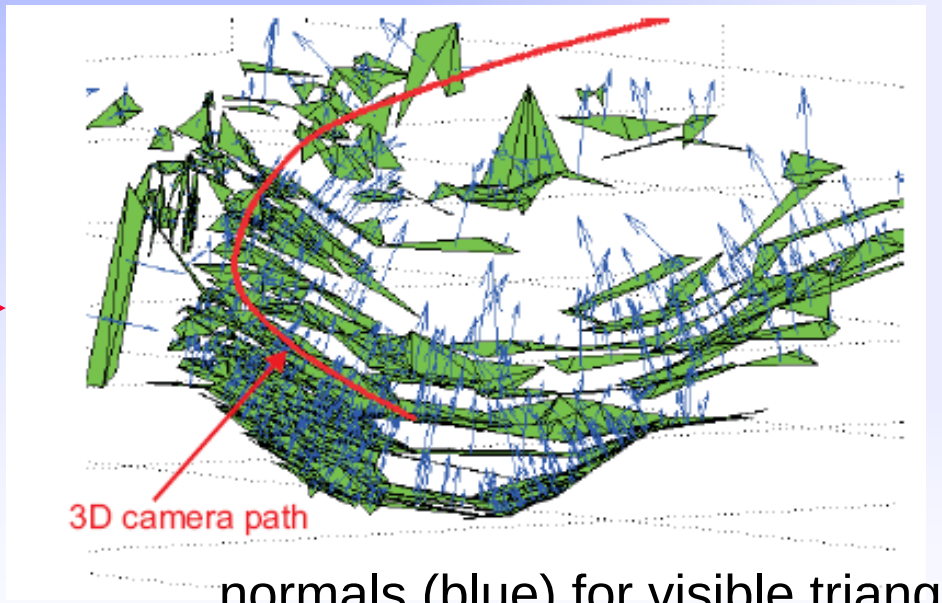
medium level scene description





triangle parameters

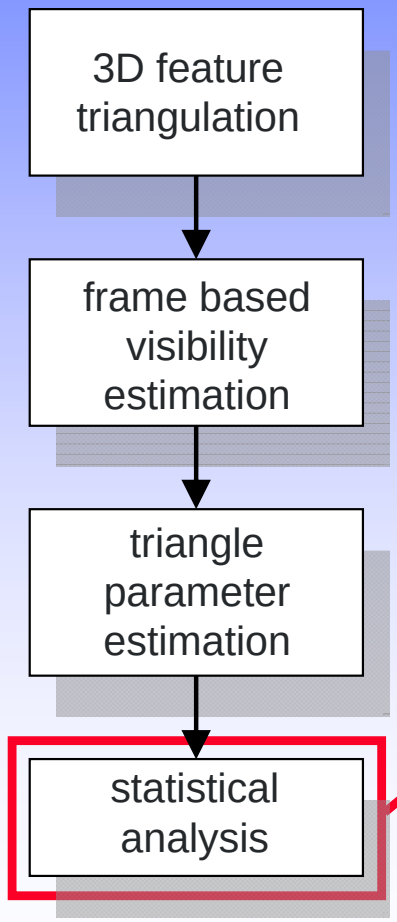
- orientation (normal)
- size (area)
- distance to camera
- neighbors

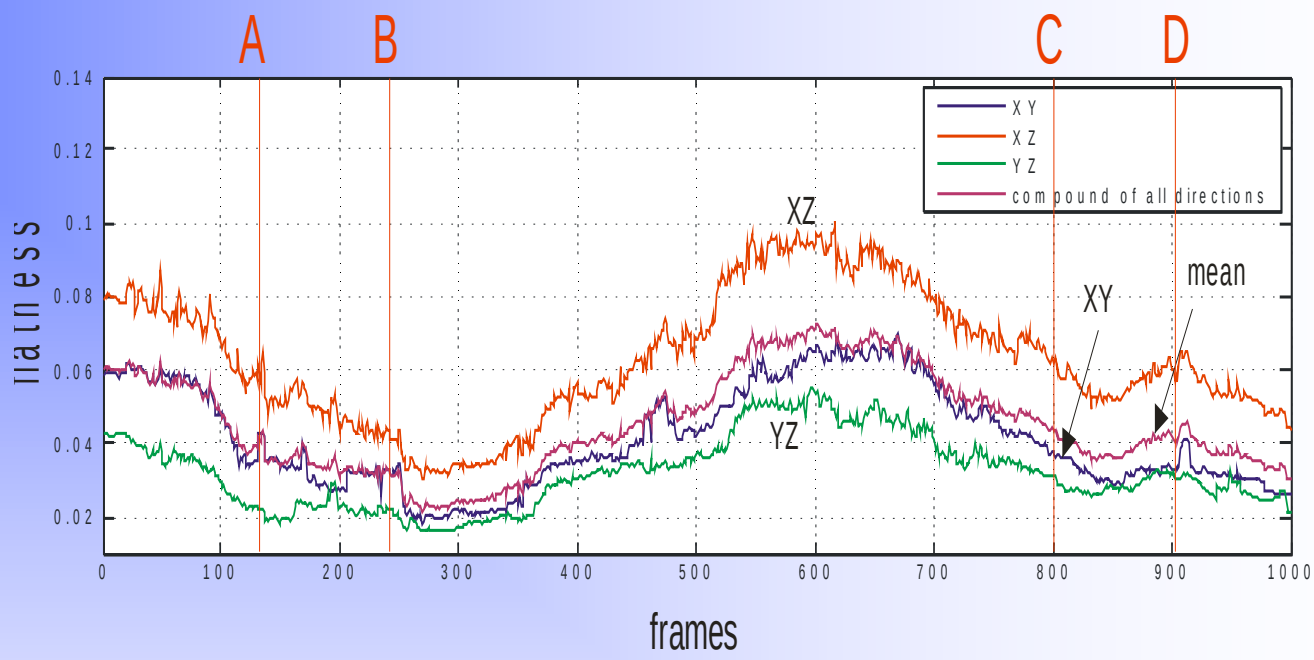


normals (blue) for visible triangles

Medium level analysis

medium level scene description





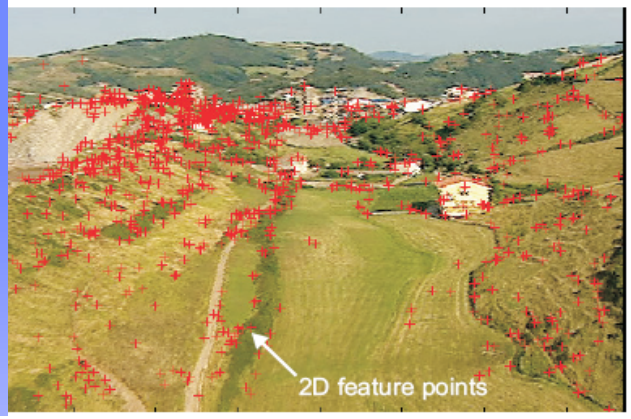
variance of angles of the triangle normals
(angles are weighted by triangle size)

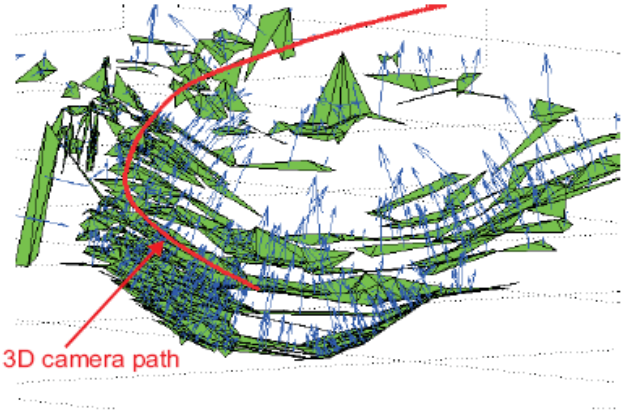
straightforward scene flatness classification

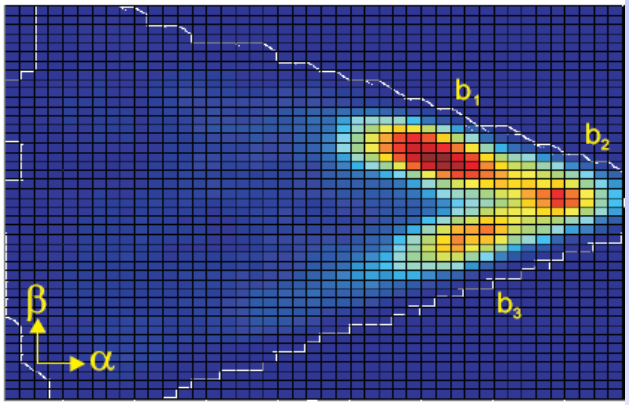
Medium level analysis

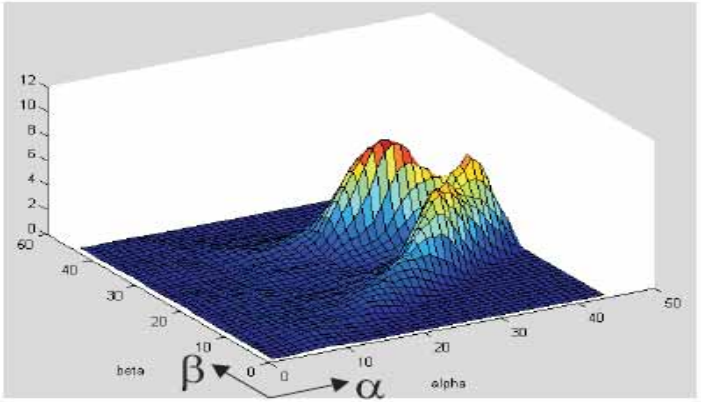
medium level scene description

- 3D feature triangulation
- frame based visibility estimation
- triangle parameter estimation
- statistical analysis









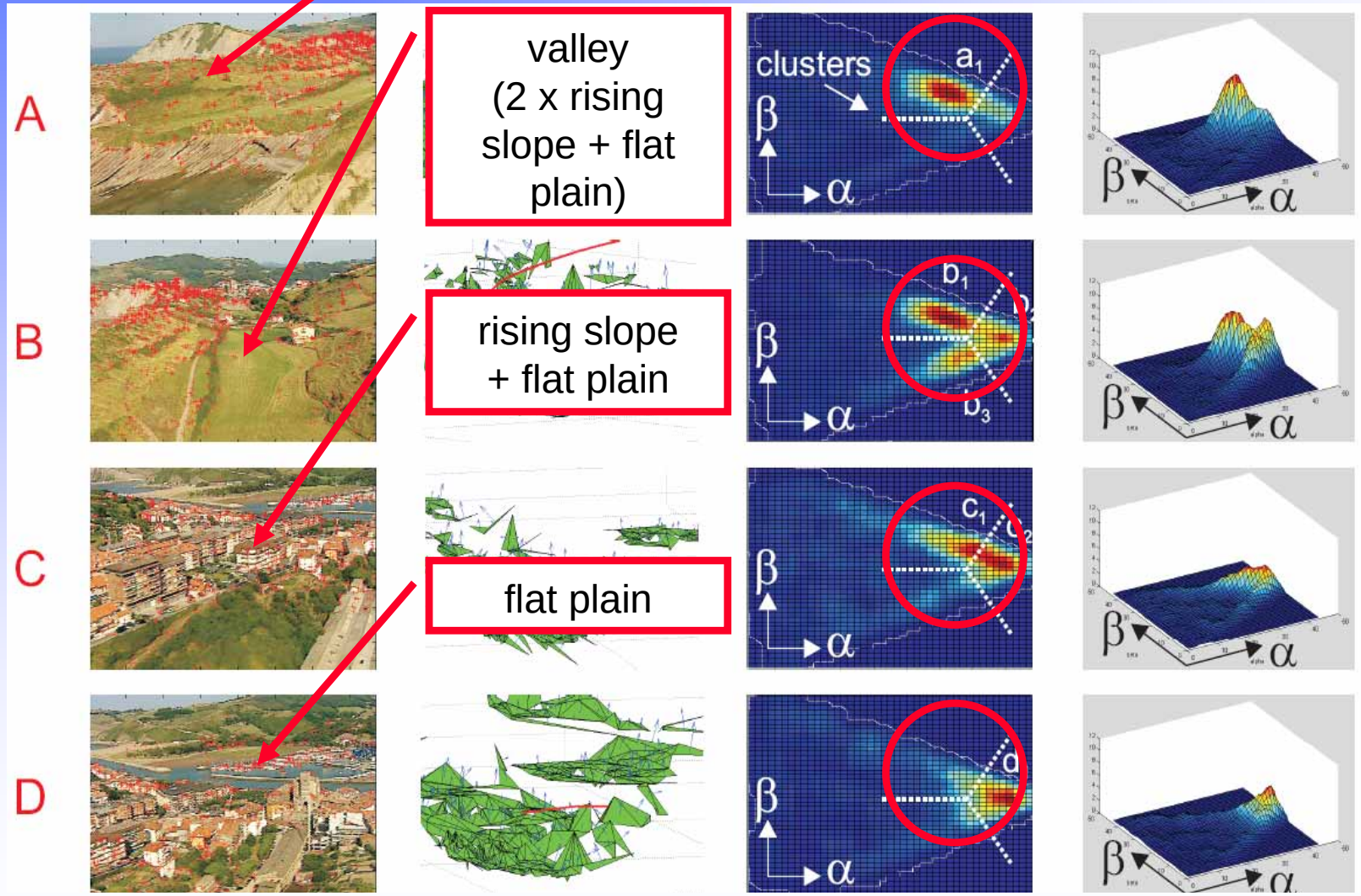
2D histogram based clustering of angles of the triangle normals

straightforward 3D scene content classification

High level semantic annotation

3D scen sification: hill, valley,…

ased clustering



Benchmarking

Database

- Approx. 90 min unedited video (mostly helicopter flights)
- Typical search scenarios for journalists

Evaluation

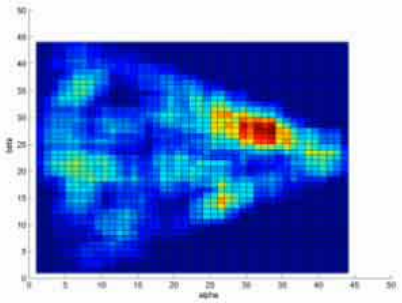
- Manual annotation of the considered database as ground truth
- Precision and recall evaluation for all clips

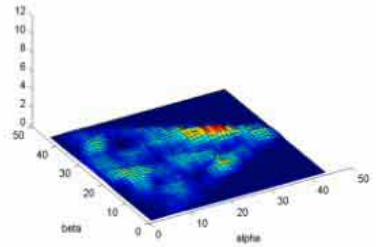
– Micro average: frame based

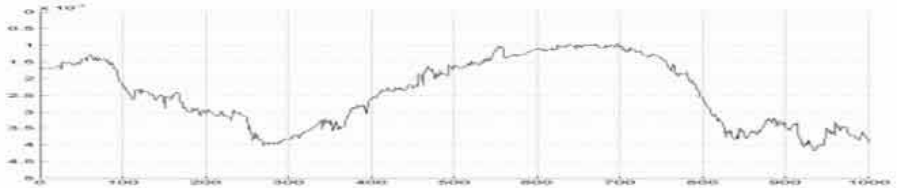
– Macro average: clip based

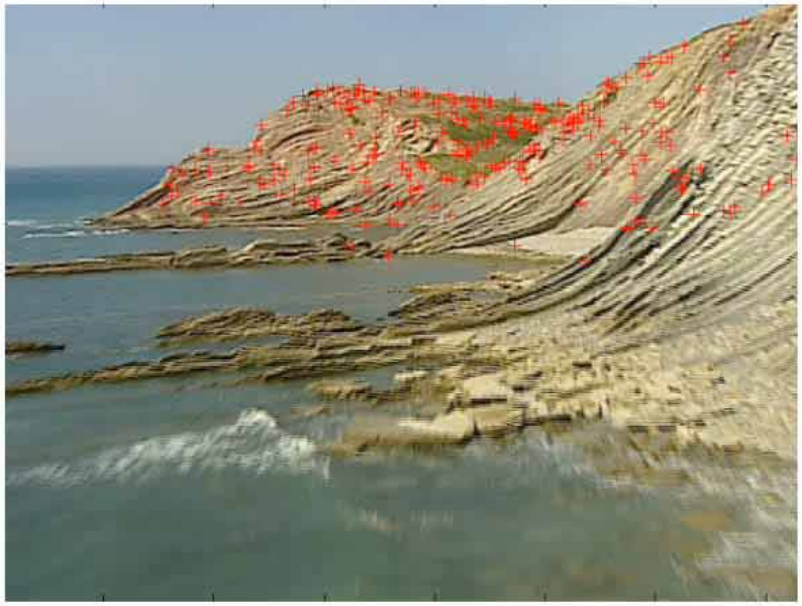
	macro average		micro average	
	precision in %	recall in %	precision in %	recall in %
Reliable motion	99.28	72.17	99.89	84.47
zoom in	99.05	92.57	98.40	89.87
zoom out	99.27	91.51	98.62	91.46
rotation	96.88	67.42	95.99	61.89
flatness	66.41	56.78	77.61	73.46

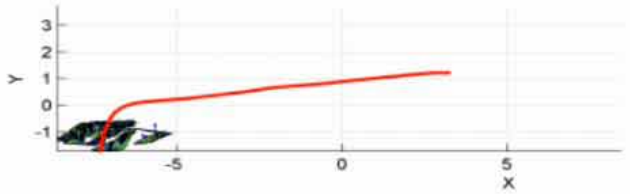
Demo: Low + mid level analysis





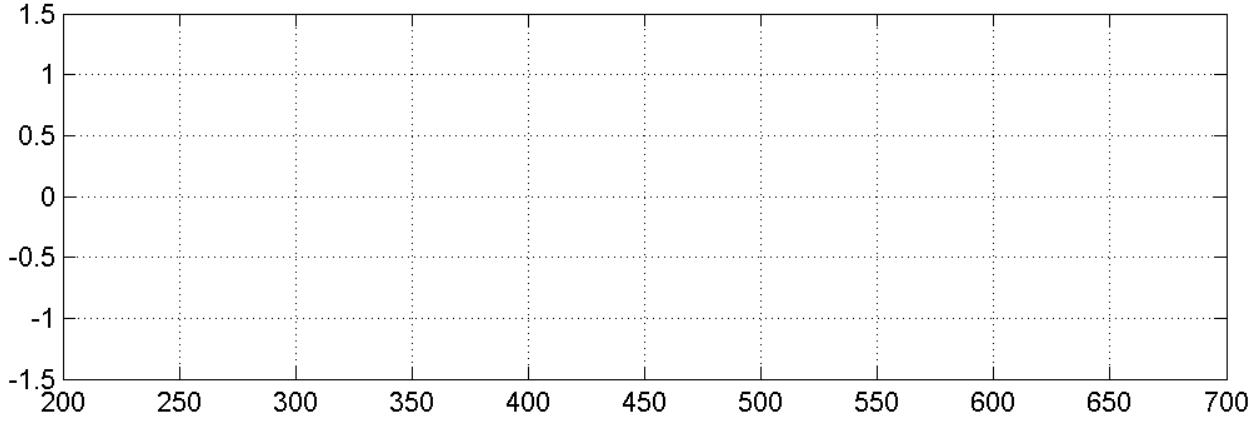




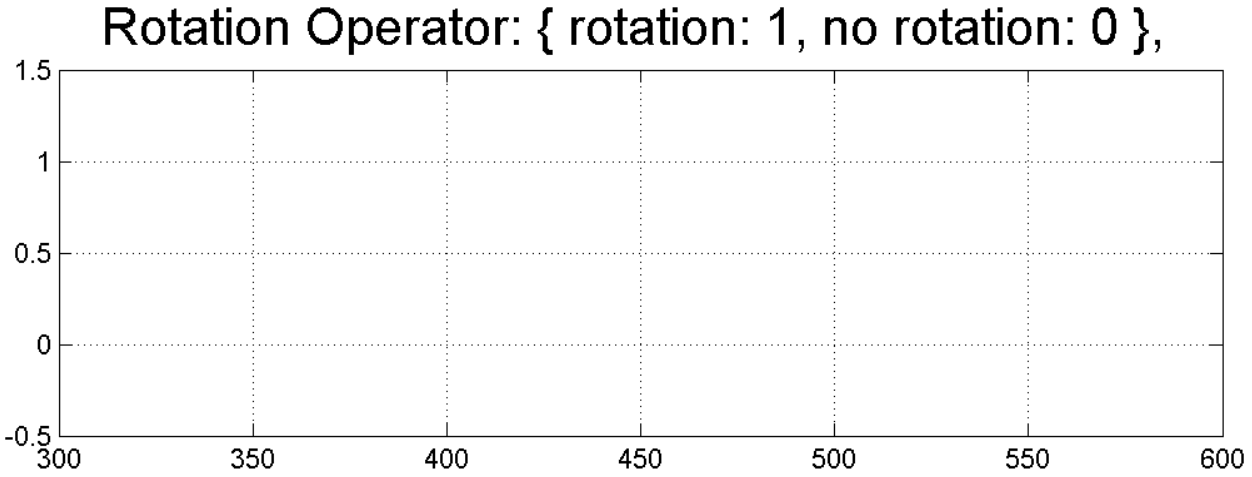


Classification examples: Zoom

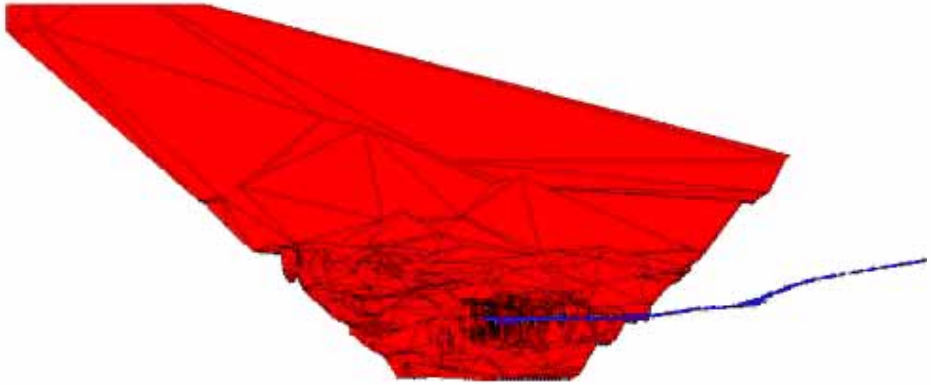
Zoom Operator: { zoom in: 1, no zoom: 0, zoom out: -1 },

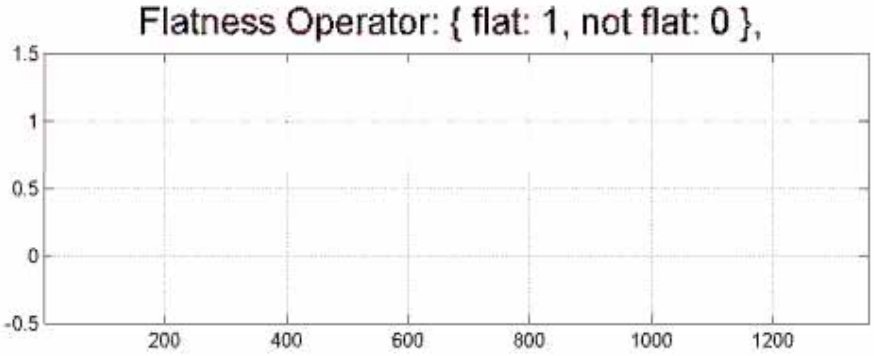


Classification examples: Rotation

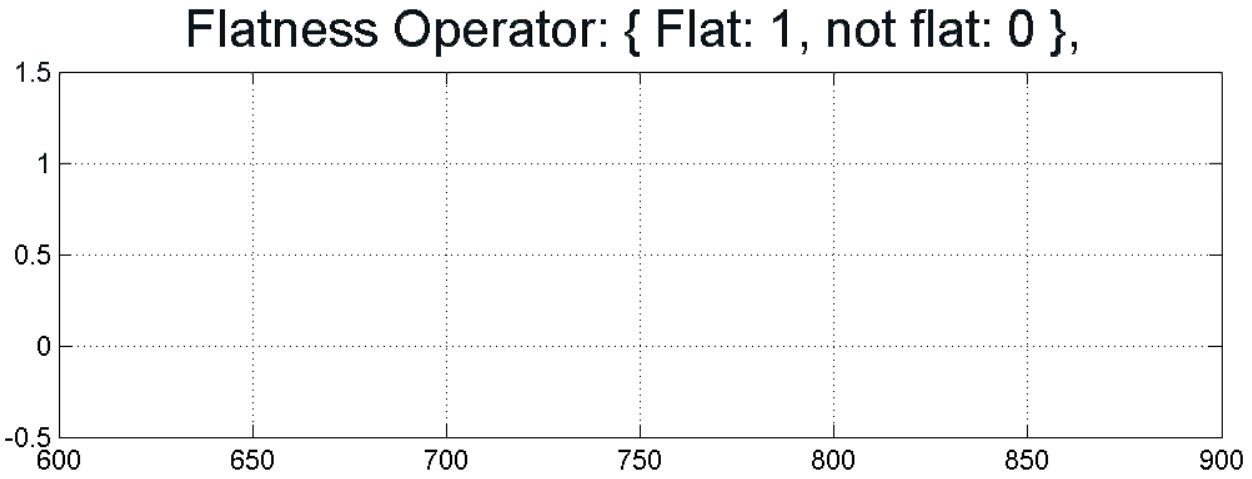


Classification examples: Flatness





Classification examples: Flatness



Summary

Low and mid level analysis

- Fully automatic
- Robust to noisy data sets and distortions

General approach

- Powerful approach for semantic 3D scene content annotation
- Many possibilities for extensions and further research

Visit the exhibit outside!

Thank you !!!

Project coordinator

Fraunhofer Heinrich-Hertz-Institute

Dr. Oliver Schreer

Email: Oliver.Schreer@hhi.fraunhofer.de