Camera Motion Identification in the Rough Indexing Paradigm

Petra KRÄMER and Jenny BENOIS-PINEAU

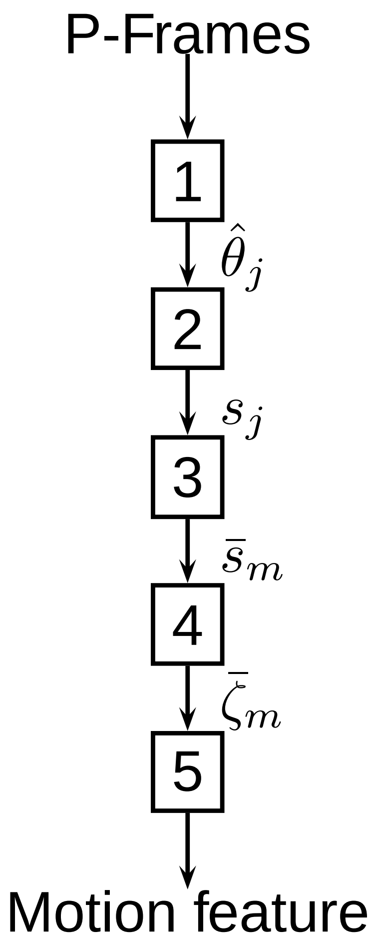
LaBRI – University Bordeaux I, France

{petra.kraemer,jenny.benois}@labri.fr

Introduction

- Task:
  - Given the shot boundary reference
  - Identify the shots in which a certain camera motion (pan, tilt, zoom) is present
- Rough Indexing Paradigm:
  Work on a lower spatial and temporal resolution i.e. P-Frames
- Aim:
  Reuse motion low-level descriptors from the compressed stream
- Main challenge in TRECVID 2005:
  Jitter camera motion due to hand-carried cameras

Overview



1. Global Motion Estimation
2. Significance Value Computation
3. Motion Segmentation
4. Thresholding
5. Classification

j related to frames, m related to segments of homogeneous motion

Overview

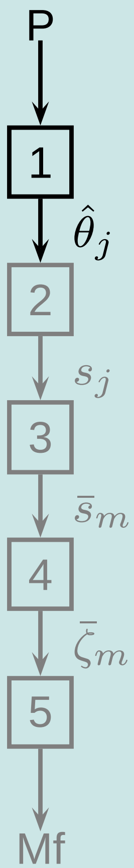
P-Frames

1. Global Motion Estimation — $\hat{\theta}_j$
2. Signifi cance Value Computation — s_j
3. Motion Segmentation — $\bar{s}_m$
4. Thresholding — $\bar{\zeta}_m$
5. Classifi cation

Motion feature

j related to frames, m related to segments of homogeneous motion

Global Motion Estimation



Robust global motion estimator for P-Frames [DBP01]:

- Estimation of the affine 2D motion model:

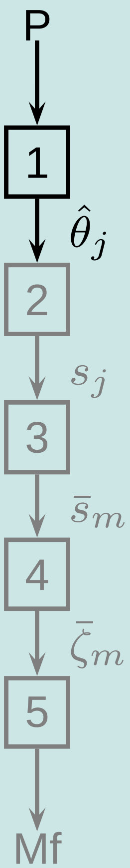
$$\begin{pmatrix} dx_i \\ dy_i \end{pmatrix} = \begin{pmatrix} a_1 \\ a_4 \end{pmatrix} + \begin{pmatrix} a_2 & a_3 \\ a_5 & a_6 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

- Based on the weighted least squares method:

$$\hat{\theta} = (H^T W H)^{-1} H^T W Z$$

$\hat{\theta}$	$= (a_1, a_2, a_3, a_4, a_5, a_6)^T$
Z	MPEG motion compensation vectors
H	macroblock centers
W	weights defined by the derivative of the Tukey function

Global Motion Estimation



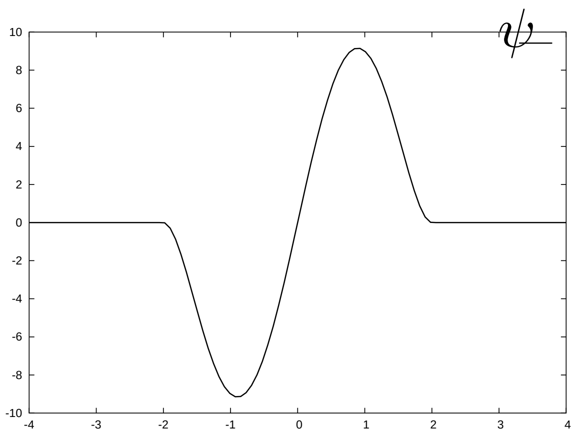
- The derivative of the Tukey function:

$$\psi(r, \lambda_r) = \begin{cases} r(r^2 - \lambda_r^2)^2 & \text{if } |r| < \lambda_r \\ 0 & \text{otherwise} \end{cases}$$

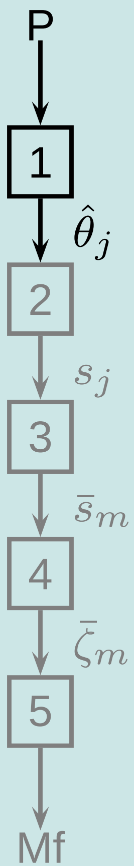
- The weights are [OB95]:

$$w_i = \frac{\psi(r_i)}{r_i}$$

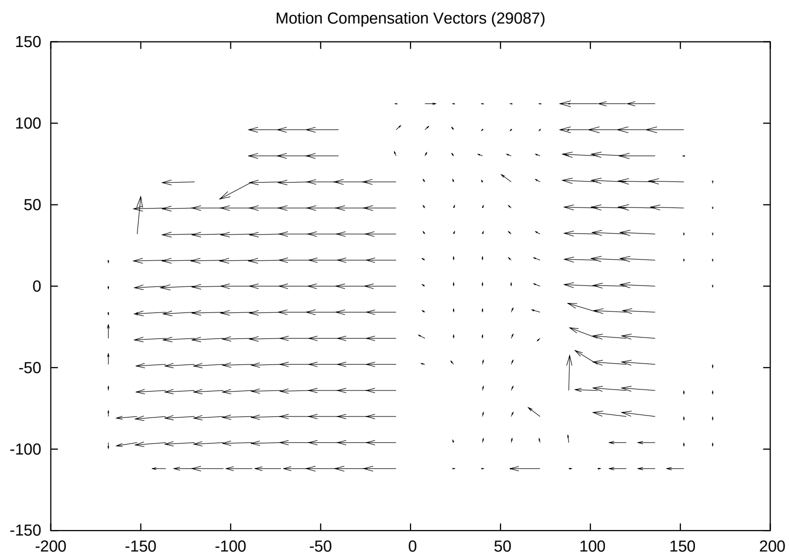
λ_r	threshold
$r_i = z_i - \hat{z}_i$	residuals
z_i	i-th MPEG motion vector
$\hat{z}_i$	estimation of z_i



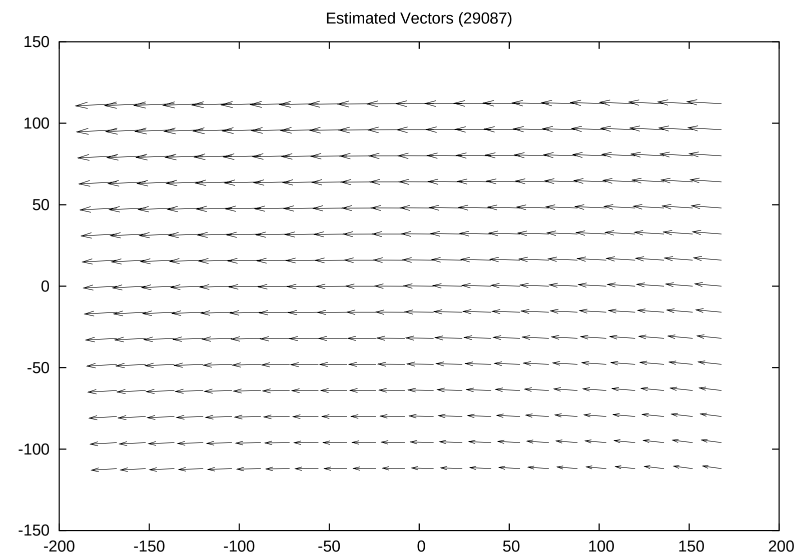
Global Motion Estimation



a)



b)



c)

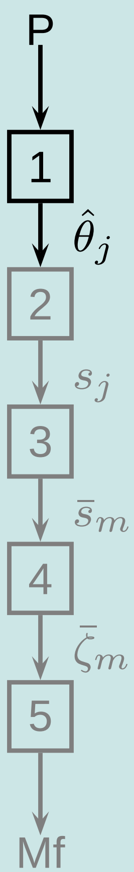
a) P-Frame motion vectors

b) Estimated vectors

c) Macroblocks:

- Outliers
- Dominant estimation support D $(w_i > 0)$

Global Motion Estimation



Problem:

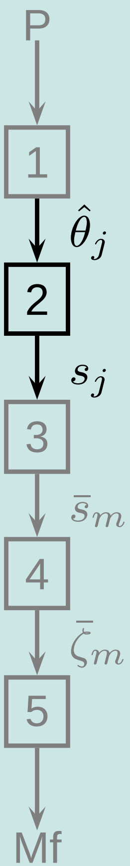
- The global motion parameters are noisy due to jitter motions.
- The global motion parameters have different meanings.

Solution:

- Signifi cance test of the motion parameters:

  Thresholding of likelihood values

Significance Value Computation



Based on [BGG99]:

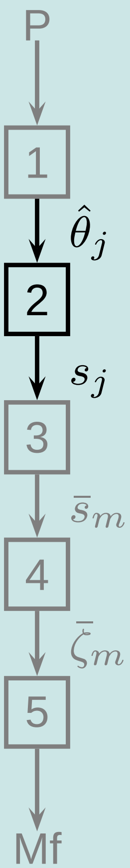
- Change to another basis of elementary motion-subfields:

$$\phi = (pan, tilt, zoom, rot, hyp1, hyp2) \quad \text{with}$$

$$zoom = \tfrac{1}{2}(a_2 + a_6) \quad rot = \tfrac{1}{2}(a_5 - a_3)$$
$$hyp1 = \tfrac{1}{2}(a_2 - a_6) \quad hyp2 = \tfrac{1}{2}(a_3 + a_5)$$

- Consider two hypotheses H_0 and H_1
  - H_0: the considered component of ϕ is significant with $\hat{\phi}_0$ as the corresponding motion model
  - H_1: the considered component of ϕ is not significant $(= 0)$ with $\hat{\phi}_1$ as the corresponding motion model

Significance Value Computation



- Likelihood function associated to each hypothesis:

$$f(\hat{\phi}_l) = \prod_{i \in D} \left(\frac{1}{2\pi \sqrt{\det(\Sigma_l)}} \exp\left(-\frac{1}{2} (r_i^T \Sigma_l^{-1} r_i) \right) \right)$$

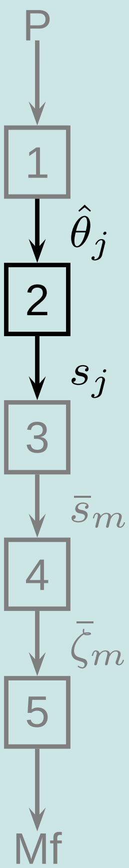
$$= \frac{1}{(2\pi \sigma_{x,l} \sigma_{y,l})^{||D||}} \exp\left(-||D||\right), \; l = 0, 1$$

Assumption:

$$\Sigma_l = \begin{pmatrix} \sigma_{x,l}^2 & 0 \\ 0 & \sigma_{y,l}^2 \end{pmatrix}$$

Σ	covariance matrix
σ_x, σ_y	variances for x and y
D	dominant estimation support

Significance Value Computation



- The significance value s is:

$$s = \ln\left(\frac{f(\hat{\phi}_1)}{f(\hat{\phi}_0)}\right) = ||D||\left(\ln(\sigma_{x,0}\sigma_{y,0}) - \ln(\sigma_{x,1}\sigma_{y,1})\right)$$

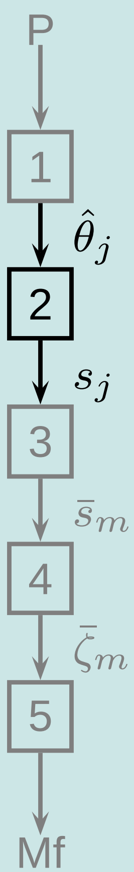
$$=^* ||D||\left(\ln(\sigma_0^2) - \ln(\sigma_1^2)\right)$$

$*$ assuming that $\sigma_x = \sigma_y$

- Aim: Use s to test the significance
- Idea:

If a motion feature (pan, zoom, tilt) is present in a shot, its corresponding motion parameter is significant during a sufficient number of frames.

Significance Value Computation



Problem:

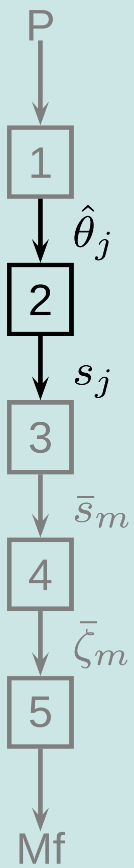
- The signifi cance values can be noisy due to jitter motions.
- The motion models $\hat{\theta}$ can be inaccurate.

Solution:

- Smooth the signifi cance value along the time and take decision on the temporal mean value.

  –> Segment shots into subshots of homogeneous motion
- Introduce confi dence measures in order to reject frames with an inaccurate motion model.

Significance Value Computation



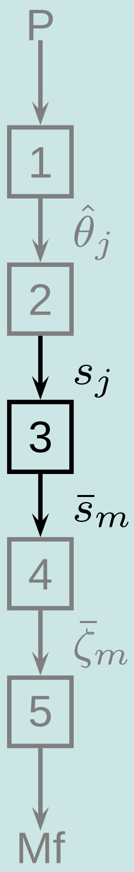
Two reasons for inaccurate motion models:

- Failure of the MPEG encoder
  –> Confidence measure $c_D \approx ||D||$
- Failure of the global motion estimation algorithm
  –> Confidence measure $c_\sigma \approx \sigma_0^2$

Reject of the frame if: $c_D < \lambda_D \;||\; c_\sigma > \lambda_\sigma$

λ_D	threshold
λ_σ	threshold

Motion Segmentation



Hinkley test to detect changes on the temporal mean value $\bar{s}(t)$:

- Downward jump:

$$U_k = \sum_{t=0}^{k} \left(s_t - \bar{s} + \frac{\delta_{min}}{2} \right) \quad (k \geq 0)$$

$$M_k = \max_{0 \leq i \leq k} U_i; \text{ detection if } M_k - U_k > \lambda_H$$

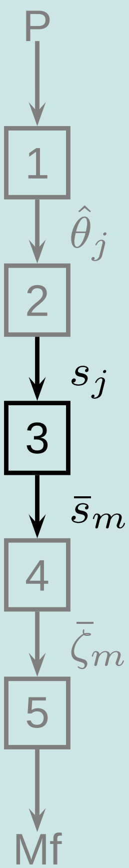
- Upward jump:

$$V_k = \sum_{t=0}^{k} \left(s_t - \bar{s} - \frac{\delta_{min}}{2} \right) \quad (k \geq 0)$$

$$N_k = \min_{0 \leq i \leq k} V_i; \text{ detection if } V_k - N_k > \lambda_H$$

$\bar{s}$	temporal mean value
δ_{min}	minimal jump magnitude
λ_H	predefi ned threshold

Motion Segmentation



Principle of the Hinkely test:

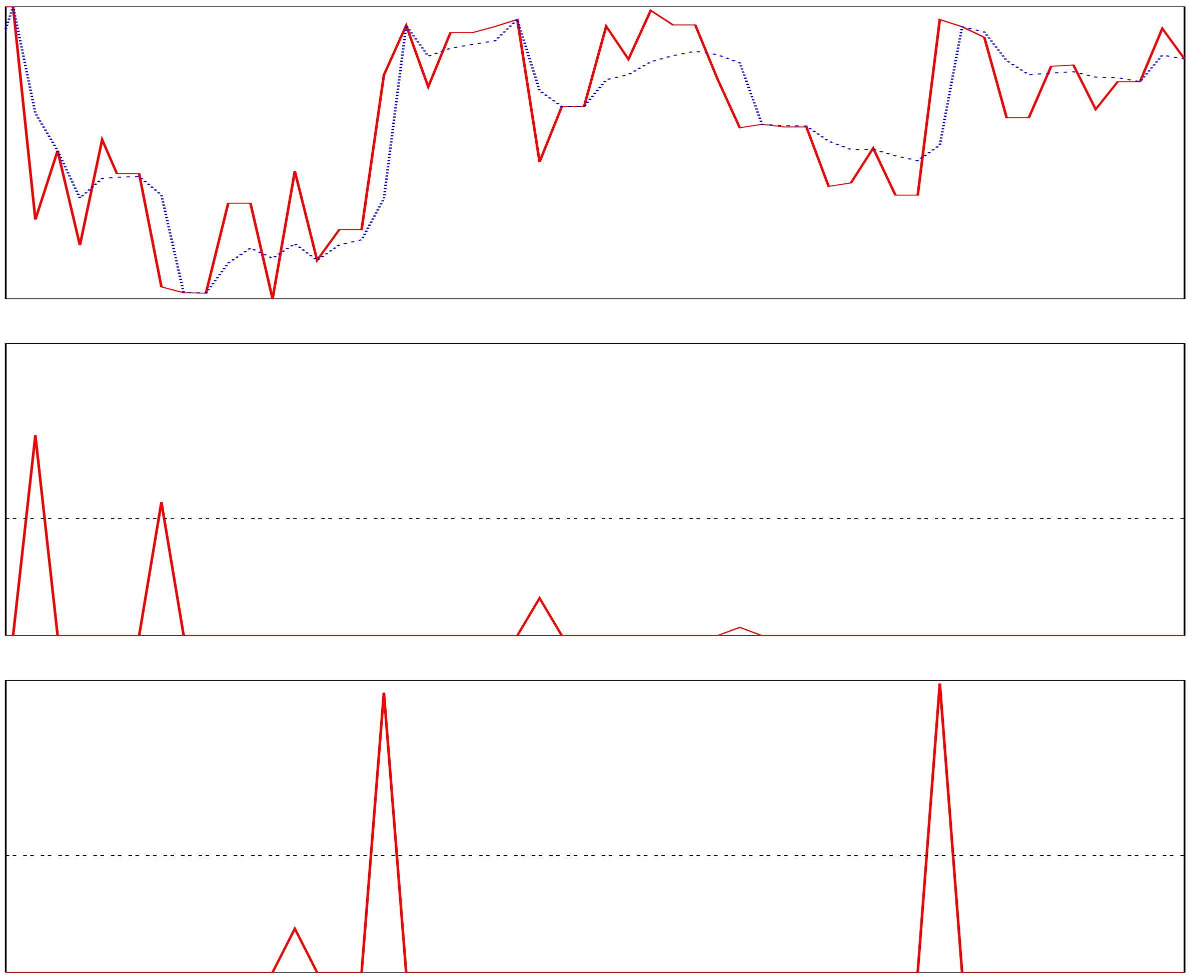
s and $\bar{s}$

Down

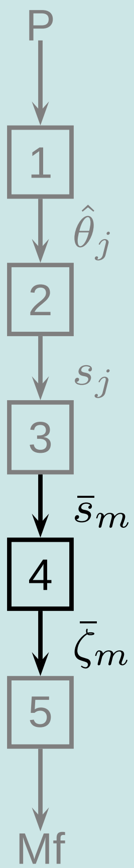
$M_k - U_k$

Up

$V_k - N_k$



Thresholding



- Selection of the hypothesis:

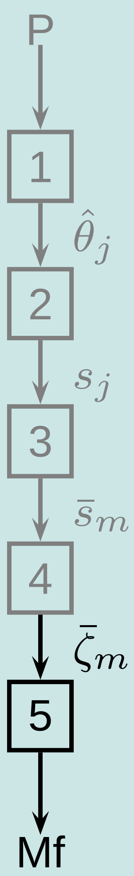
$$\bar{s}(t) = \frac{1}{T - t_0} \sum_{t=t_0}^{t=T} s(t) \underset{H_1}{\overset{H_0}{\gtrless}} \lambda_s$$

- And relative thresholding to determine the dominant motion:

$$\bar{\zeta}(t) = \begin{cases} \bar{s}(t) & \text{if } \bar{s}(t) < \alpha \cdot \min\{\bar{s}_{pan}, \bar{s}_{tilt}, \bar{s}_{zoom}, \bar{s}_{rot}, \bar{s}_{hyp1}, \bar{s}_{hyp2}\} \\ 0 & \text{otherwise} \end{cases}$$

$T - t_0$	segment of homogeneous motion
λ_s	threshold
α	constant

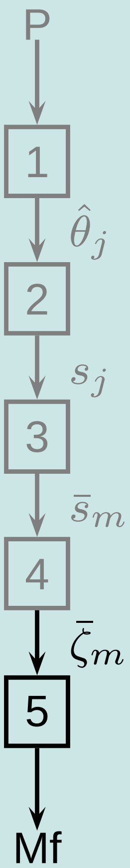
Classification



The following classifi cation scheme is applied to the thresholded mean signifi cance values $\bar{\zeta} = (\bar{\zeta}_{pan}, \bar{\zeta}_{tilt}, \bar{\zeta}_{zoom}, \bar{\zeta}_{rot}, \bar{\zeta}_{hyp1}, \bar{\zeta}_{hyp2})$:

	$\bar{\zeta}$	motion feature
1	$(0, 0, 0, 0, 0, 0)$	static camera/ no signifi cant motion
2	$(\bar{\zeta}_{pan}, 0, 0, 0, 0, 0)$	pan
3	$(0, \bar{\zeta}_{tilt}, 0, 0, 0, 0)$	tilt
4	$(\bar{\zeta}_{pan}, \bar{\zeta}_{tilt}, \bar{\zeta}_{zoom}, 0, 0, 0)$	zoom
5	others	complex camera motion

Classification



Postprocessing:

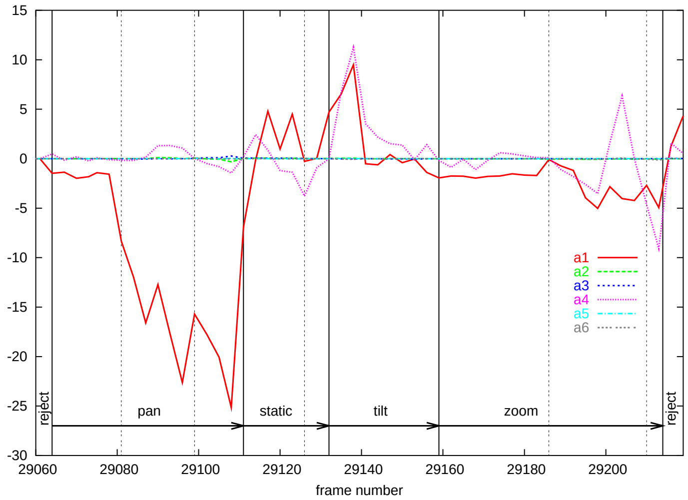
- Join neighbored segments with the same motion feature
- Reject segments with a duration shorter than t_{min} frames

If a motion feature is still present:

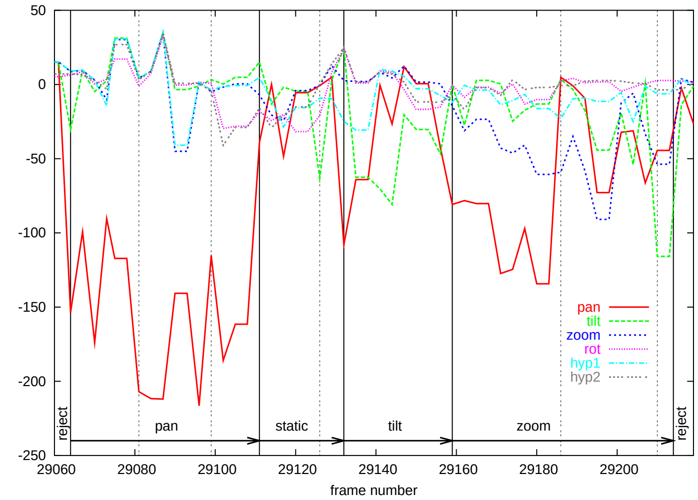
The shot is identified to contain the motion feature.

Results

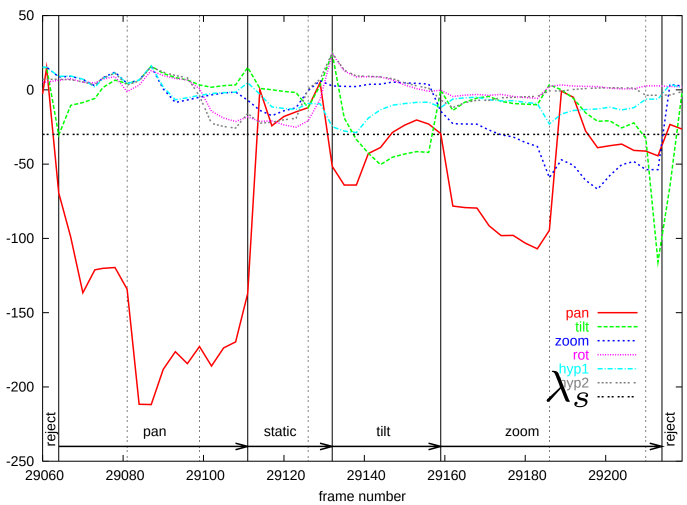
Results for the shot “shot106_136”:



a)



b)

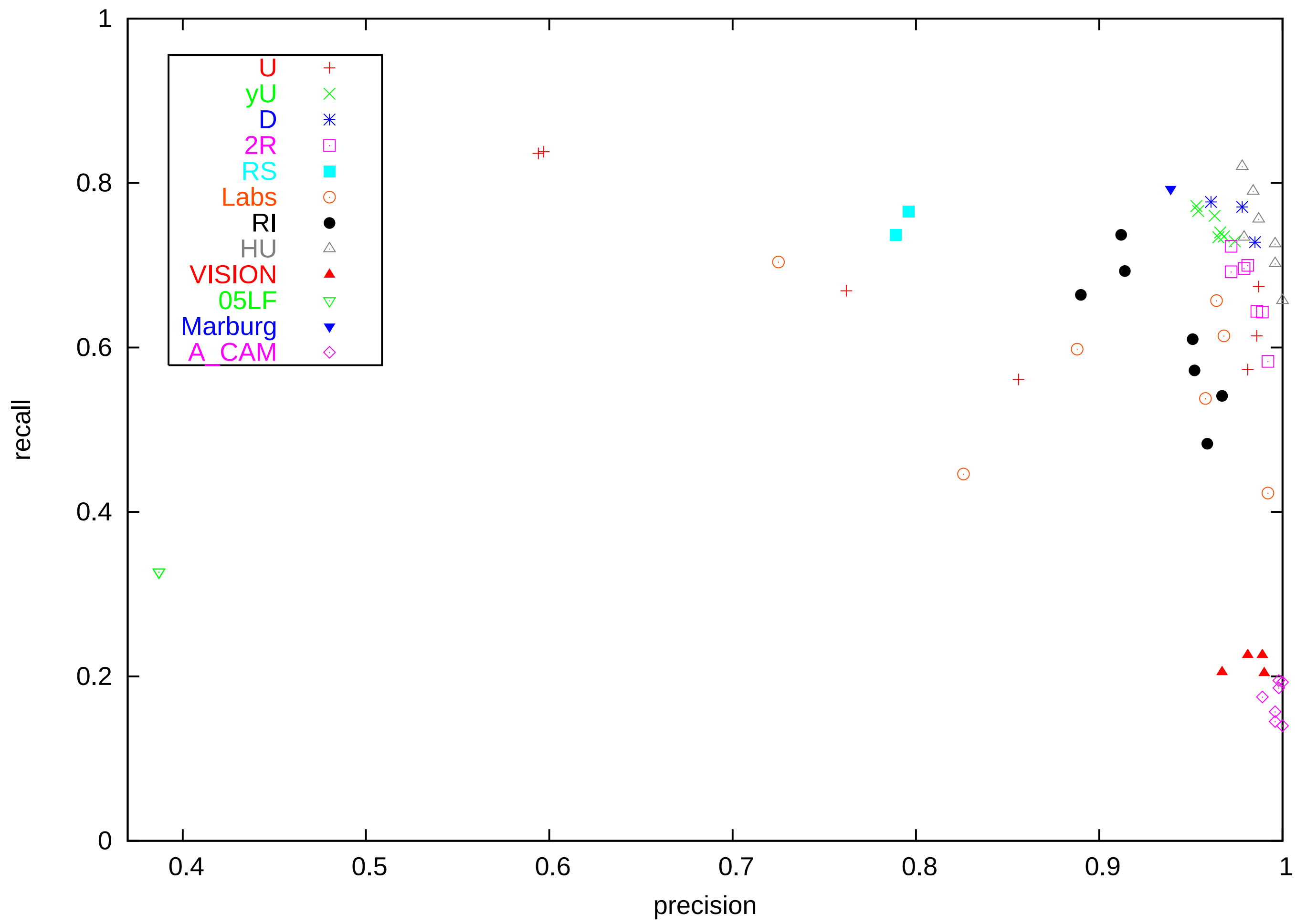


c)

a) Global motion parameters $\hat{\theta}$
b) Signifi cance values s
c) Online mean values $\bar{s}$

Results

Precision and recall for all submissions:



RI –> LaBRI
Precision 0.912
Recall 0.737

Conclusion and Perspectives

Conclusion:

- Proposition of a method based on global motion estimation and signifi cance test.
- The proposed method can handle moving objects and jitter motions.
- No decoding of the compressed stream.
- Performance 3-4 times faster than real time.
- Since no ground truth available, diffi culties to determine the best parameter set.

Future work:

Focus mainly on the correction of motion models if the encoder block-matching algorithm fails.

References

[BGG99] P. Bouthemy, M. Gelgon, and F. Ganansia. A unifi ed approach to shot change detection and camera motion characterization. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(7):1030–1044, October 1999.

[DBP01] M. Durik and J. Benois-Pineau. Robust motion characterisation for video indexing based on MPEG2 optical flow. In *International Workshop on Content-Based Multimedia Indexing, CBMI'01*, pages 57–64, 2001.

[OB95] J.M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, 1995.