Gyro-based Camera-Motion Detection in User-Generated Videos

Sophia Bano, Andrea Cavallaro, Xavier Parra
{s.bano,a.cavallaro}@qmul.ac.uk, xavier.parra@upc.edu

1. Introduction

Camera-Motion Detection (CMD)

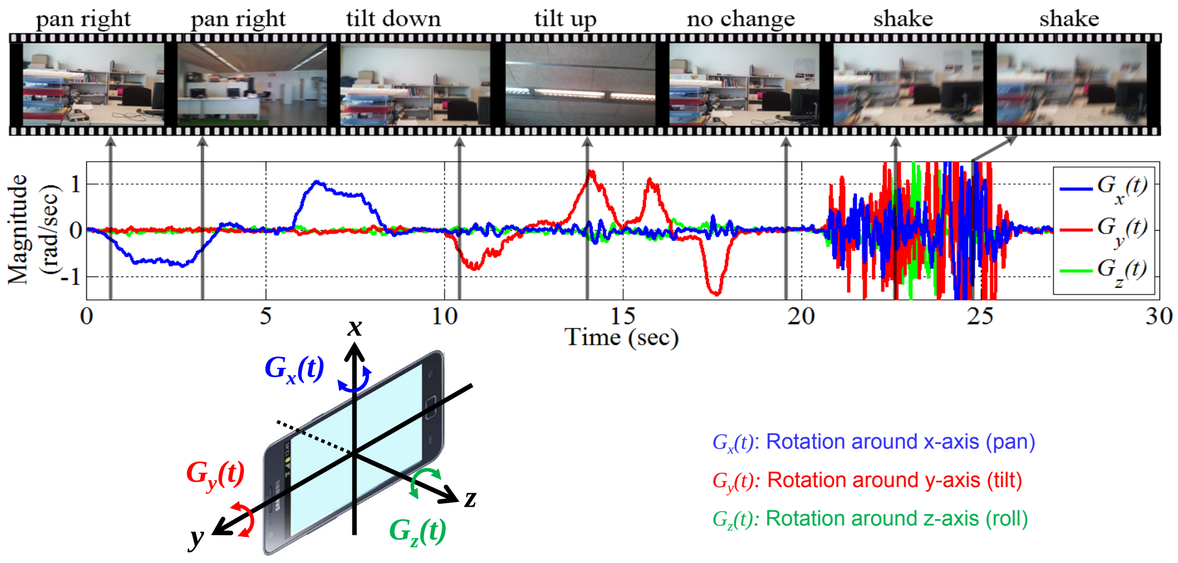
- To detect pan, tilt and shake motions in videos
- Applications: summarization, composition and shot detection

Gyroscope

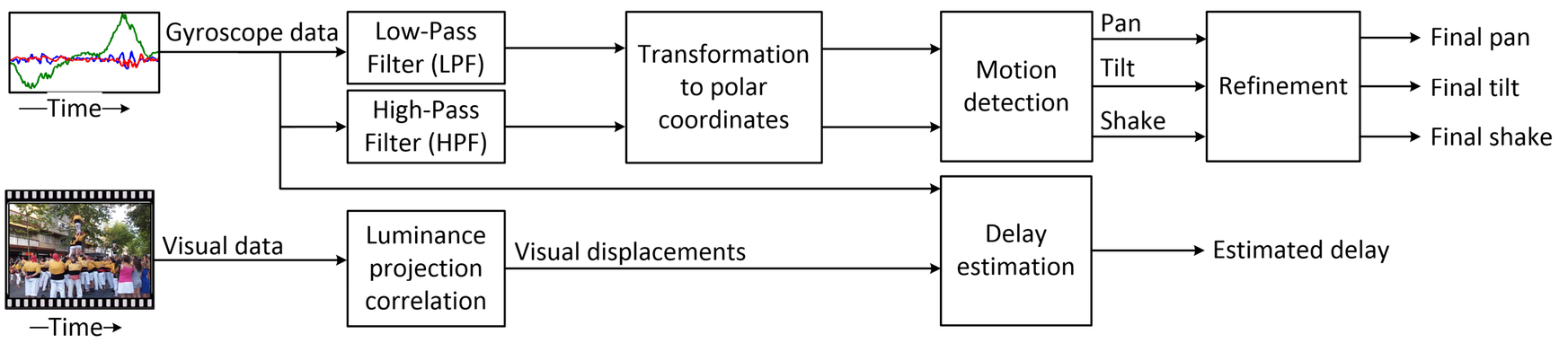
- Measures angular velocities around the device's axes
- Can replace or complement the camera

Challenge

- To estimate the delay between gyroscope and visual data



2. Camera-Motion Detection using Gyroscope (CMDG)

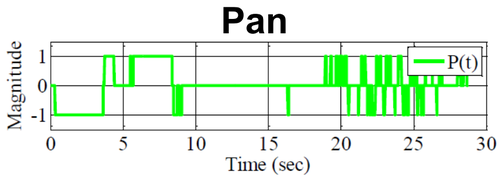


$L_x(t)$: Horizontal displacement
$R_x(t)$: Cross-correlation of horizontal gyro-visual data
$G_x^L(t)$, $G_y^L(t)$: LPF gyroscope x and y axes data
$G_x^H(t)$, $G_y^H(t)$: HPF gyroscope x and y axes data
$P(t)$, $T(t)$, $S(t)$: Pan, tilt and shake detections
$P_d(t)$, $T_d(t)$, $S_d(t)$: Final pan, tilt and shake detections

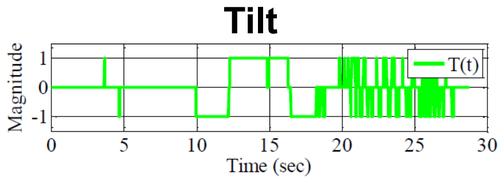
Motion detection

- Analyze LPF and HPF gyroscope data in polar coordinate

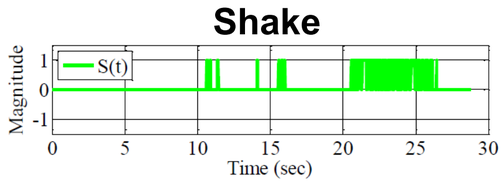
Pan



Tilt

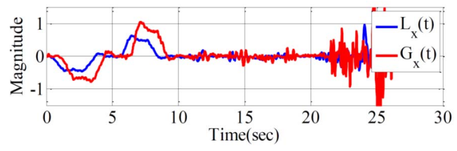


Shake



Luminance projection correlation

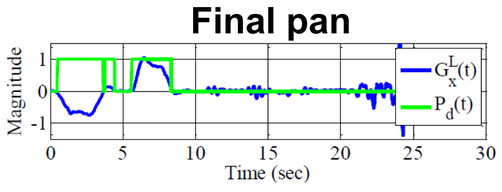
- Gyroscope and visual data



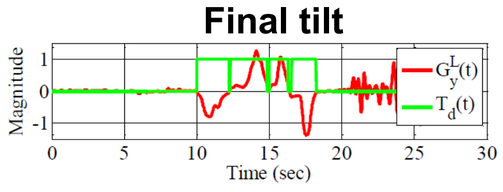
Refinement

- Apply morphological operations to remove outliers and connect disconnected segments

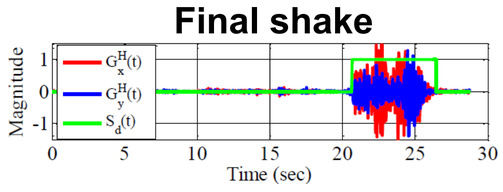
Final pan



Final tilt

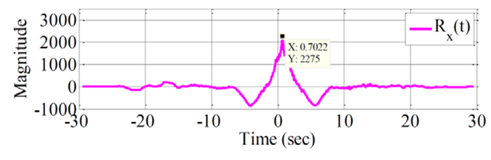


Final shake



Delay estimation

- Cross-correlation of gyro-visual data



3. Experimental evaluation

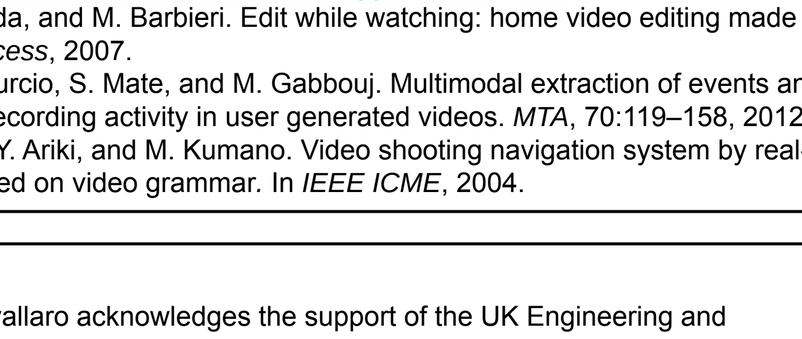
Multimodal (video, inertial sensor) dataset[1]

- 24 User-Generated Videos captured using Cellbot Data Logger [1]

Comparison of the CMDG with

- **VISUAL [2,3]:** detects pan from horizontal, tilt from vertical and shake from HPF visual displacements
- **ISENSOR [4]:** detects pan from LPF compass data, tilt from unfiltered accelerometer data and shake from HPF accelerometer data

Percentage of synchronized data *w.r.t* the absolute delay error. The overlap between the gyroscope and visual data is varied to test the robustness.



Camera-motion detection evaluation. (HB: high brightness recordings; LB: low brightness recordings; P: Precision; R: Recall; F_1: F_1-score $= 2\frac{P \cdot R}{P + R}$)

		Pan			Tilt			Shake		
Method	Type	P	R	F_1	P	R	F_1	P	R	F_1
CMDG	HB	0.96	0.93	**0.94**	0.83	0.81	**0.82**	0.74	0.97	**0.83**
VISUAL	HB	0.77	0.74	0.75	0.31	0.53	0.39	0.69	0.67	0.68
ISENSOR	HB	0.77	0.60	0.67	0.23	0.39	0.29	0.67	0.48	0.56
CMDG	LB	0.91	0.95	**0.93**	0.85	0.84	**0.85**	0.86	0.86	**0.86**
VISUAL	LB	0.41	0.25	0.31	0.17	0.20	0.19	0.24	0.74	0.37
ISENSOR	LB	0.52	0.36	0.43	0.49	0.47	0.48	0.62	0.78	0.69

4. Conclusion

- Aligned multimodal data by estimating the synchronization delay
- Utilized tri-axial gyroscope data for CMD
- Achieved an overall CMD accuracy (F_1-score) of 89%

References

[1] https://cellbots.googlecode.com/files/CellbotsDataLogger_v1.1.0_full.apk (Last accessed: 19/10/2015).
[2] M. Campanella, H. Weda, and M. Barbieri. Edit while watching: home video editing made easy. In *SPIE Multimedia Content Access*, 2007.
[3] F. Cricri, K. Dabov, I. Curcio, S. Mate, and M. Gabbouj. Multimodal extraction of events and of information about the recording activity in user generated videos. *MTA*, 70:119–158, 2012.
[4] K. Uehara, M. Amano, Y. Ariki, and M. Kumano. Video shooting navigation system by real-time useful shot discrimination based on video grammar. In *IEEE ICME*, 2004.

Acknowledgement

S. Bano was supported by the EACEA Agency of the European Commission under EMJD ICE FPA no 2010-0012. A. Cavallaro acknowledges the support of the UK Engineering and Physical Sciences Research Council (EPSRC), under grant EP/K007491/1.

[1]**Dataset available:** http://www.eecs.qmul.ac.uk/~andrea/cmdg

www.youtube.com/smartcameras

@smartcameras