Complexity measurement and characterization of 360-degree content

VQEG Meeting, March 2019, Berlin

Francesca De Simone

Jesús Gutiérrez
Patrick Le Callet

Distributed & Interactive Systems Group
Centrum Wiskunde & Informatica
Amsterdam, the Netherlands

Image, Interaction, Perception Group (IPI)
Laboratoire des Sciences du Numérique de Nantes (LS2N)
Ecole Polytechnique de l'Université de Nantes

CWI Distributed and Interactive Systems

Paper at IS&T Electronic Imaging 2019 Symposium

UNIVERSITÉ DE NANTES

Content characterization

Appropriate content selection is crucial to:

- **Validate and benchmark** image/video compression/streaming/processing/etc. algorithms
- Design “useful” **subjective quality** assessment tests
- Analyze the results from **objective quality** metrics

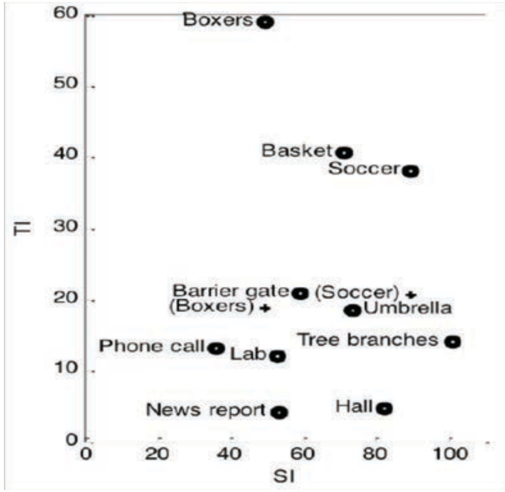
“Subjective video quality assessment is a well-understood field, yet **scene selection is often driven by convenience or content availability**.” from “Selecting scenes for 2D and 3D subjective video quality tests,” M. Pinson *et al*., *EURASIP J. Image Video Process.,* 2013

Content selection should be done according to:

- Application under study
- Purpose of the task (perceptual evaluation, system benchmarking, etc.)
- **Representative content characteristics !**

Related work on characterization of **2D content**:

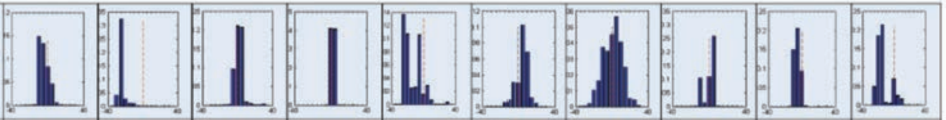
- Type of content:
  - Natural vs. animation
  - Realistic application (e.g., movies, sports, etc.) vs. system-performance evaluation (e.g., coding development and standardization)
- Duration, resolution, framerate…
- Complexity properties:
  - Spatial complexity: Spatial indicator (SI) [ITU-T P.910]
  - Temporal complexity: Temporal indicator (TI) [ITU-T P.910]
  - Colorfulness, contrast, etc.



“Selecting scenes for 2D and 3D subjective video quality tests,” M. Pinson *et al*., *EURASIP J. Image Video Process.*, 2013.

Related work on characterization of Stereoscopic **3D content**:

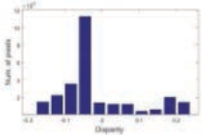
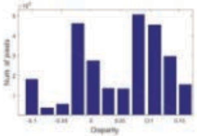
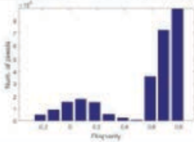
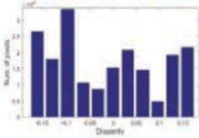
- Horizontal disparities.
- Depth budget.
- Motion in depth plane.
- Disparity changes (e.g., DTI).
- Occlusions.
- Visual comfort and fatigue.

Sequence	Barrier gate	Basket	Boxers	Hall	Lab	News report	Phone call	Soccer	Tree branches	Umbrella
Duration	16s	16s	16s	16s	16s	16s	16s	16s	16s	13s
Scenes	1	1	2	1	1	1	1	2	1	1
Condition	Outdoor	Indoor	Indoor	Indoor	Indoor	Indoor	Indoor	Indoor	Outdoor	Outdoor
Description	Car and barrier gate	Basket ball training	1. Boxer warms up 2. Boxing training	Persons meeting in a hall.	Two lab assistants working	News report mimic	Phone call in an office	1. 2 players score 2. Goal keeping	Tree leaves and wind	Person playing with an umbrella
Shooting conditions	Medium	Long distance	Close distance	Long distance, high angle	Close distance	Close distance	Close distance	Long distance	Medium distance, low angle	Close distance
C	73	n. a.	n. a.	99	n. a.	50	36	n. a.	61	60
Z	0	n. a.	n. a.	0	n. a.	0	n. a.	n. a.	0	0
Source compression	No	Panasonic H.264 HP	Panasonic H.264 HP	No	Panasonic H.264 HP	No	No	No	No	No
SI	59	71	50 (50)	82	53	53	36	89 (89)	101	74
TI	21	41	56 (19)	5	12	4	13	38 (21)	14	19
DSI	20.42	11.72	24.40	17.02	17.77	21.58	21.56	24.70	23.02	17.02
DTI	15.43	9.69	18.03	6.97	10.04	8.71	11.78	18.08	13.63	15.24
Coding α	7.89	11.66	11.36	8.20	8.94	4.68	6.63	10.73	8.52	6.57
Coding β	9.88	2.97	3.01	7.94	9.34	25.99	19.98	3.76	3.09	9.38
Disparity histogram										
D+	6	-14	9	7	6	16	17	7	3	5
D-	9	26	3	-3	22	6	15	10	9	17

"Efficient measurement of stereoscopic 3D video content issues", S. Winkler, EI 2014.
"NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences", M. Urvoy, *et al*. QoMEX 2012.
perceived depth in natural images and study of its relation with monocular and binocular depth cues, P. Lebreton *et al*. EI 2014

Related work on characterization of **Light Field content and Free Viewpoint TV**:

- Occluded Pixels
- Refocusing range

Preview				
Dataset	Own	Own	Own	EPFL
Main application	Viewpoint changing	Refocusing	Viewpoint changing & Refocusing	Refocusing
Spatial Indicator	34.10	24.83	55.21	36.14
Colorfulness	10.12	10.37	38.97	45.85
Contrast	0.07	0.11	0.59	0.14
Refocusing Range	[-0.4, 0.2]	[-0.3, 0.1]	[-1.4, 0.1]	[-0.3, 0.4]
Occluded pixels	930	1091	171	813
Disparity range	[-0.16, 0.22]	[-0.1, 0.16]	[-0.22, 0.79]	[-0.16, 0.16]
Depth Distribution				

Characterization and selection of light field content for perceptual assessment, Paudyal *et al*. *QoMEX 2017*.

- Critical Trajectories: Hypothetical Rendering Trajectories

A Study on the Impact of Visualization Techniques on Light Field Perception, Battisti *et al*. EUSIPCO 2018
Prediction of the Influence of Navigation Scan-path on Perceived Quality of Free-Viewpoint Videos , Li *et al*. IEEE JETCAS.

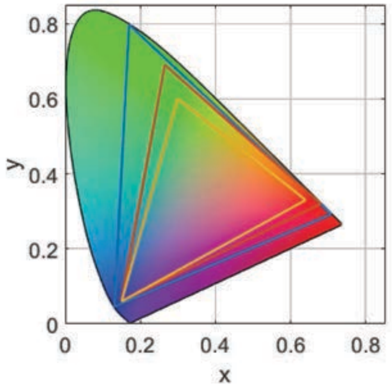
Related work on characterization of **HDR and WCG content**:

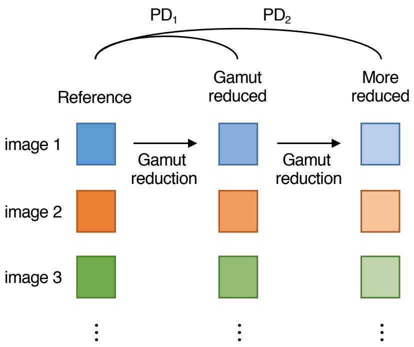
- « perceptual difference » due to Tone Mapping Operator

  An objective method for High Dynamic Range source content selection , Narwaria *et al.* *QoMEX 2014*.

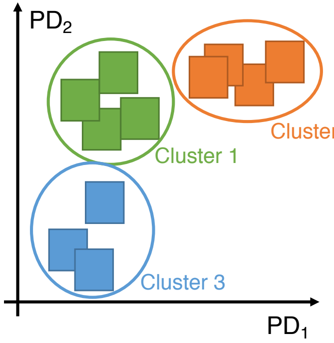
- « perceptual difference » due to gamut reduction

  A Perception-Based Framework for Wide Color Gamut Content Selection. , Lee *et al.* *ICIP 2017*.

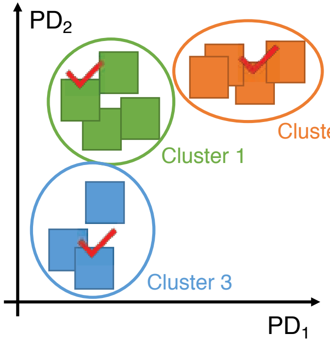




1. Measure perceptual differences (PDs) while successive gamut reduction



2. Cluster images by PDs

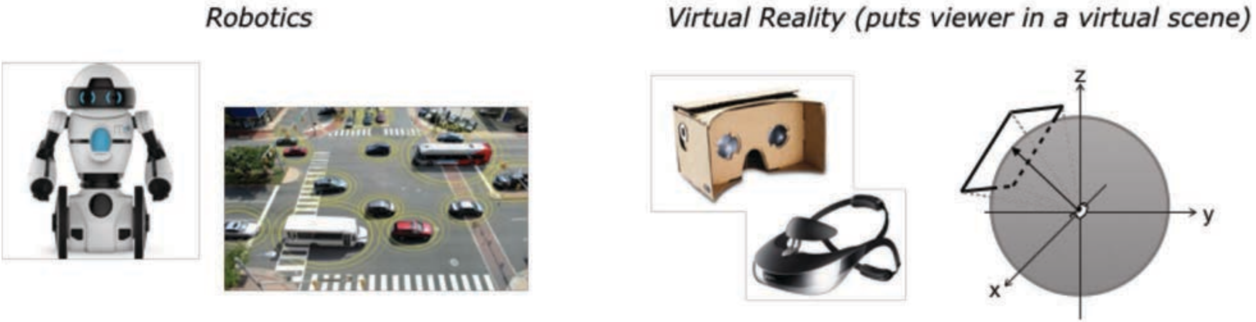


3. Select images from each cluster

Characterization of 360-degree content

- Omnidirectional signal acquisition:

- Applications:



Characterization of 360-degree content

New technical and perceptual factors must be considered

- Geometrical domain of the signal
- Interactive navigation by users

Previous work in the SoA (service provider viewpoint) :

- Afzal *et al*. “Characterization of 360-degree Videos”, VR/AR Network 2017: characterization of 360-degree YouTube videos from content provider perspective based on statistics on video duration, resolution, bitrate, motion vectors.

Open questions:

- Content characterization to be performed in which domain?
- Can content be characterized based on visual attention and navigation patterns?

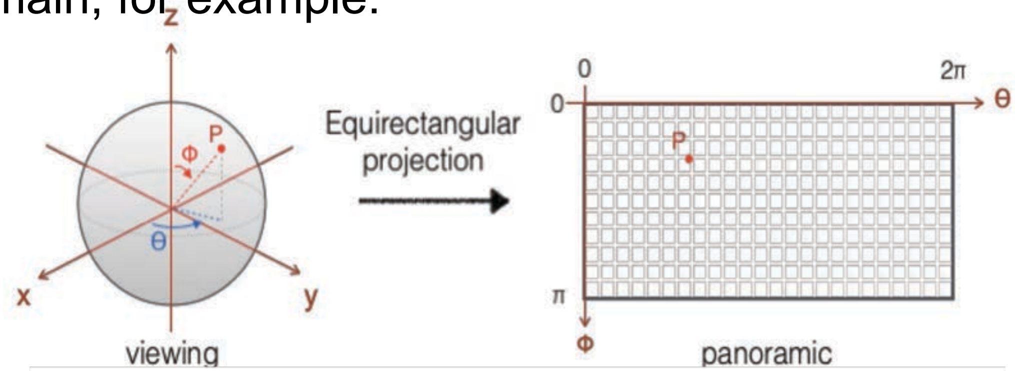
Characterization of 360-degree content

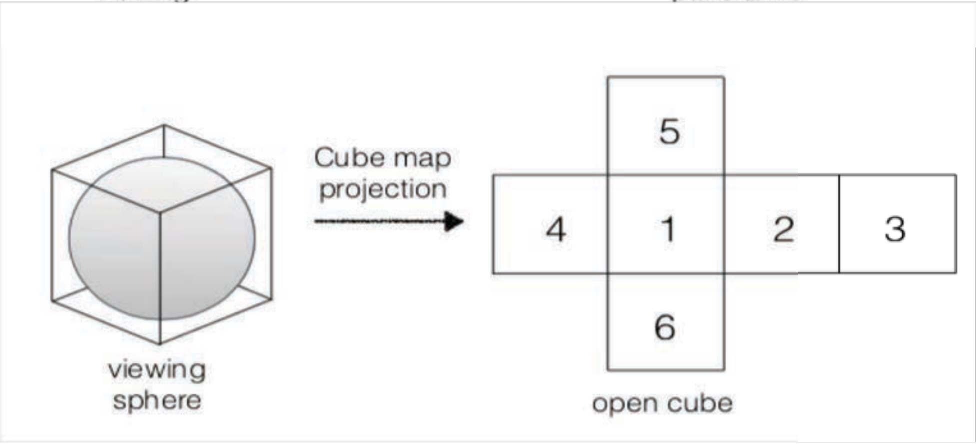
Open questions:

- **Content characterization to be performed in which domain?**
- Can content be characterized based on visual attention and navigation patterns?

Spherical versus planar representation

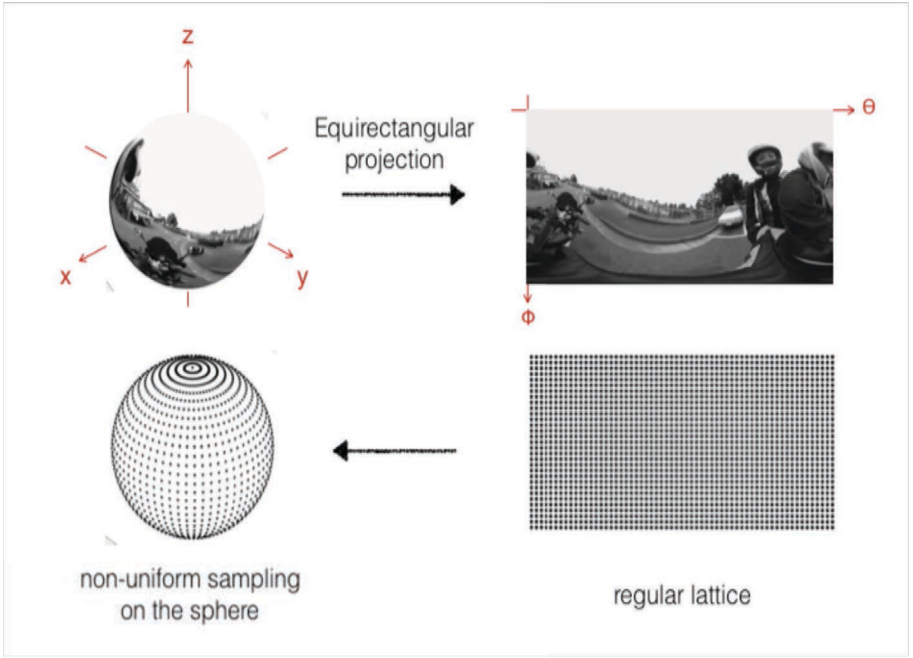
- To use existing algorithms it is convenient to map the spherical signal in planar domain, for example:

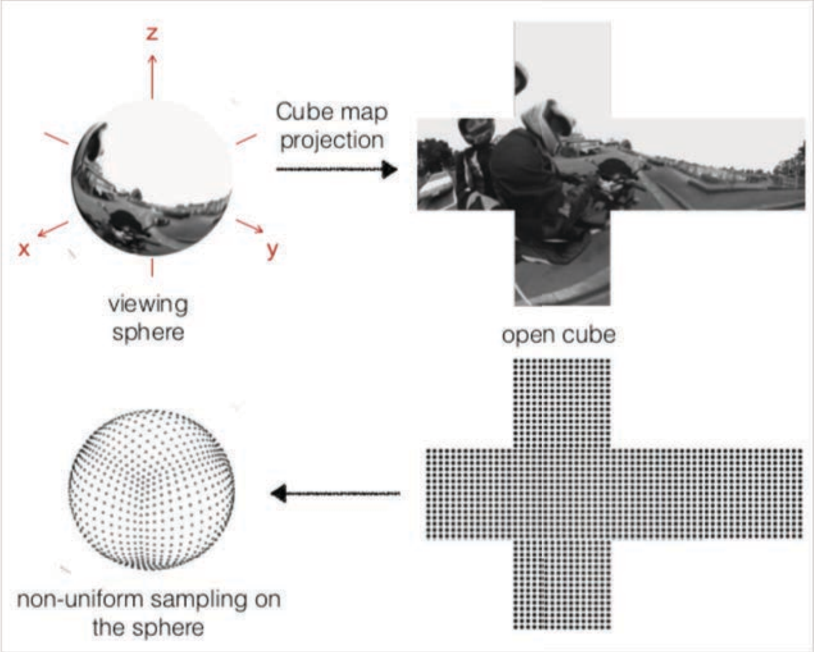




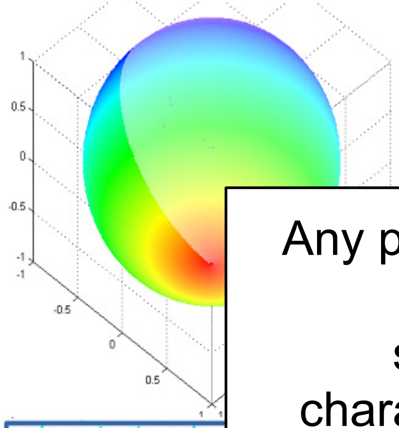
Spherical versus planar representation

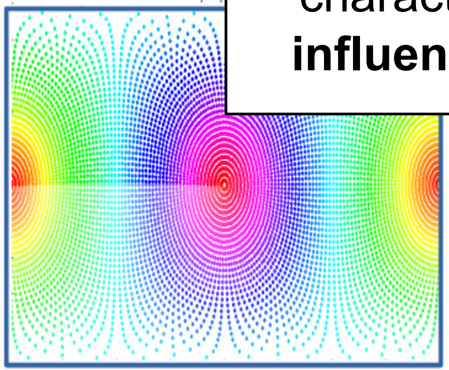
Sampling?

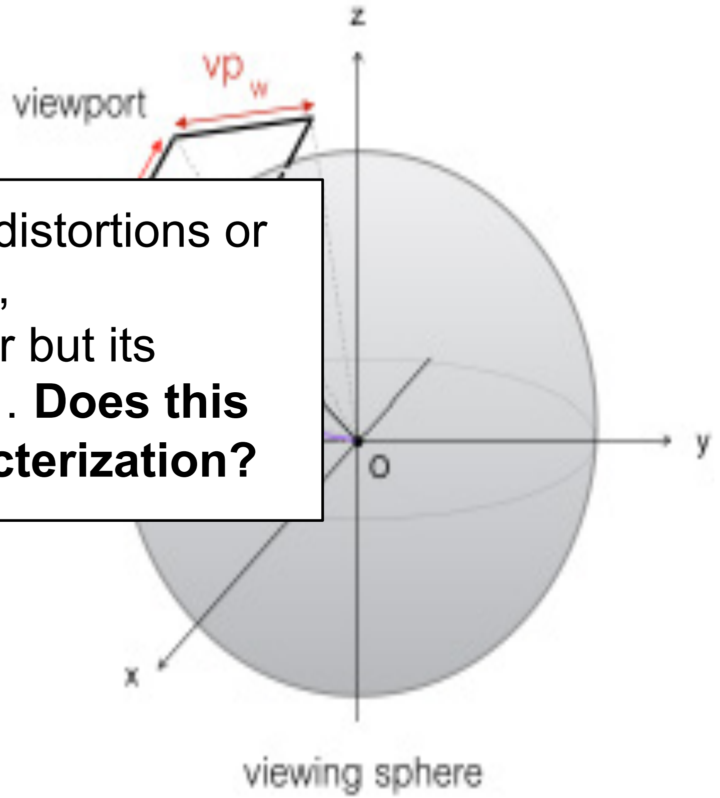




Spherical versus planar representation







Any projection introduces distortions or discontinuities,
so the signal is planar but its characteristics may vary… **Does this influence content characterization?**

Our case study

Salient360! Dataset

- 85 equirectangular images
- Processed data from **head and eye movements**:
  - Head saliency maps.
  - Head-Eye saliency maps.
  - Head-Eye scanpaths
  - Head scanpaths/trajectories

- To access it: https://salient360.ls2n.fr
- Used for the ICME Grand Challenges 2017 and 2018.

A dataset of head and eye movements for 360 degree images", Y. Rai, and al. , *ACM MMSys2017*.

Same measure, different domains & dependency on viewing direction: example

- Spatial perceptual information: `SI = std[Sobel(image_Y)]`

ITU-T P.910 "Subjective video quality assessment methods for multimedia applications"

Equirectangular

Overall SI = 15.8

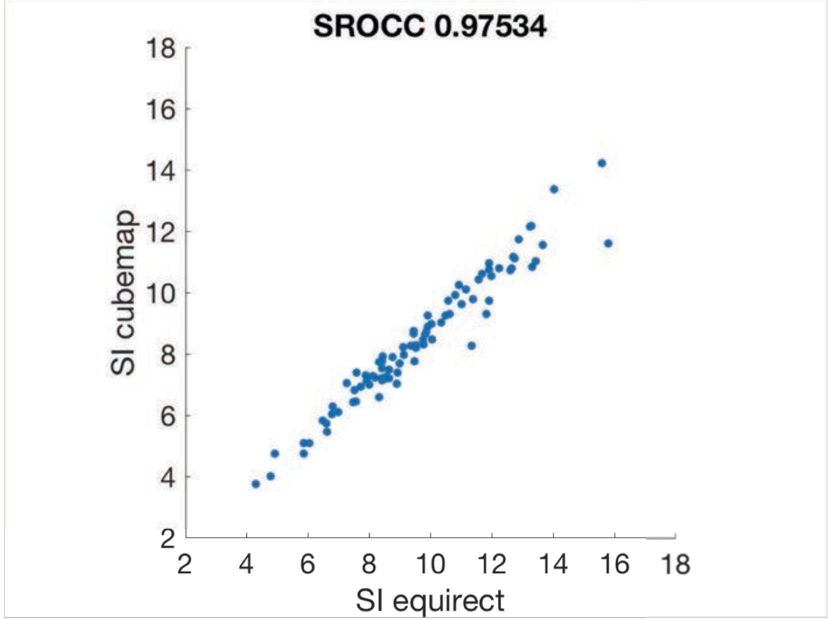
Cubemap

Overall SI = 11.6

UpFace SI = 10.3
FrontFaceSI = 10.1
LeftFaceSI= 9.9
RightFaceSI= 10.4
BackFaceSI= 7.6
BottomFaceSI=21.2

Same measure, different domains

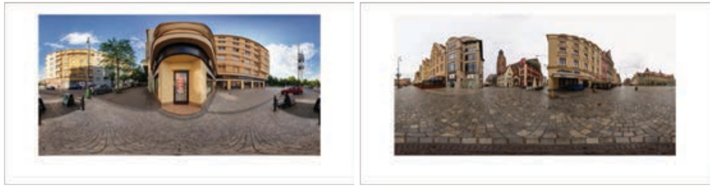
Scatter plot of SI value computed on equirectangular vs cubemap



Example of inconsistent ranking

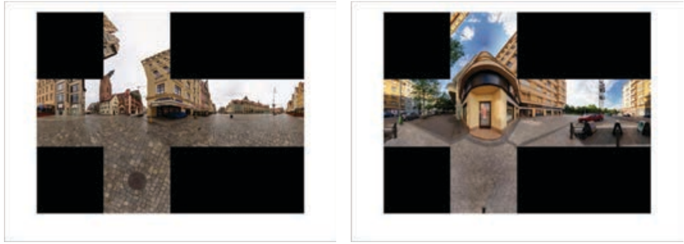
- Equirectangular SI:

Lower complexity - Higher complexity



- Cubemap SI:

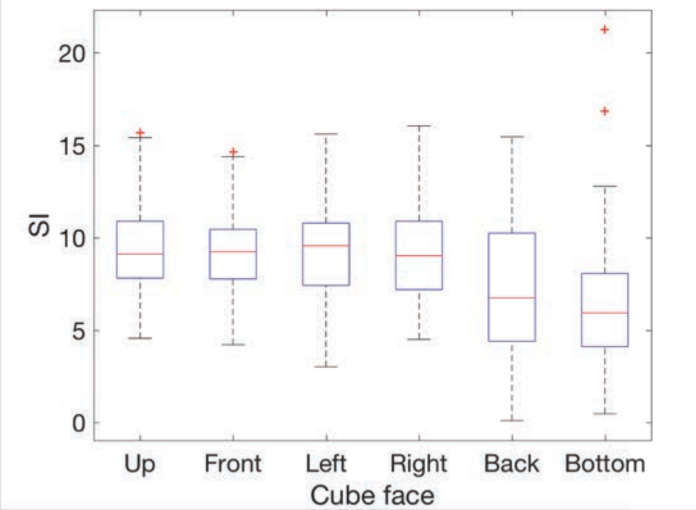
Lower complexity - Higher complexity



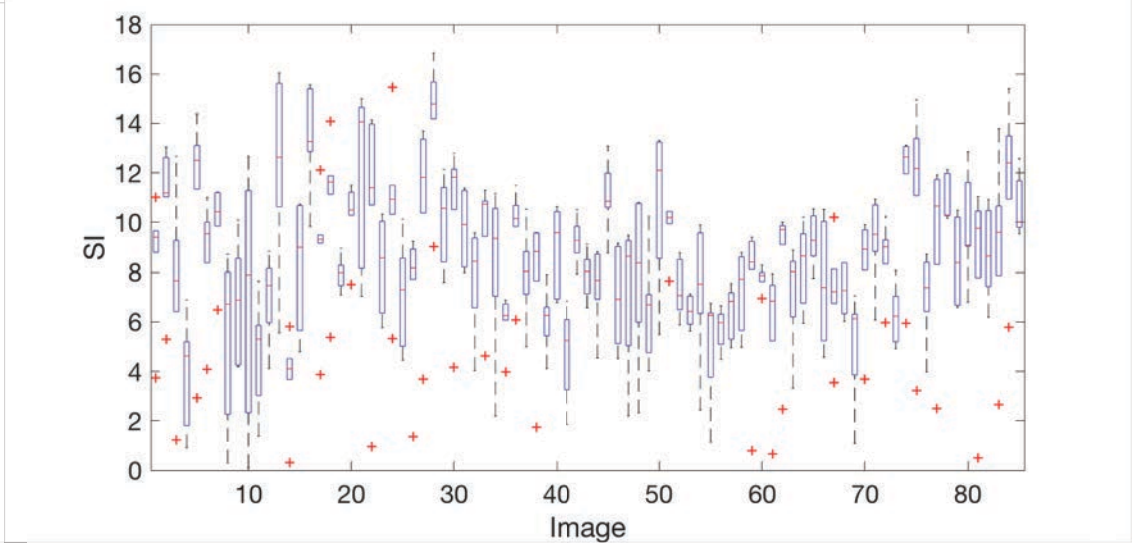
Dependency with viewing direction

- One single value might not be very informative to characterize the entire 360-images

Boxplot of SI values per cube face on entire set of images



Boxplot of SI values per image on all cube faces



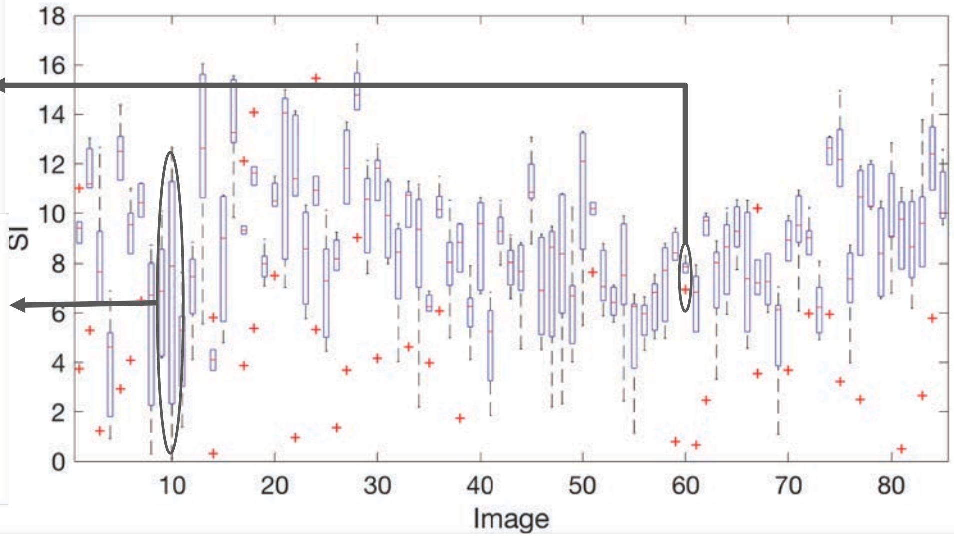
Dependency with viewing direction

- One single value might not be very informative to characterize the entire 360-images

Image in the set with lowest SI variability across faces

Image in the set with highest SI variability across faces

Boxplot of SI values per image on all cube faces



Characterization of 360-degree content

UNIVERSITÉ DE NANTES

CWI Distributed and Interactive Systems

Open questions:

- Content characterization to be performed in which domain?
- **Can content be characterized based on visual attention and navigation patterns?**

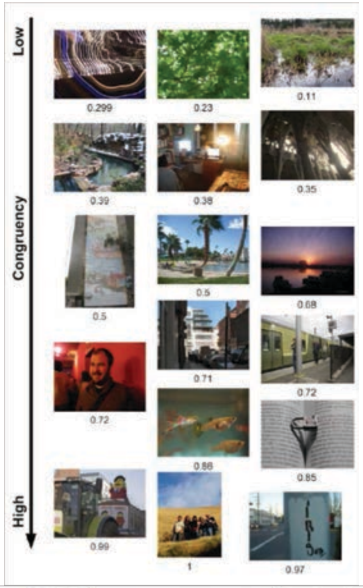
Related work on characterization of **2D content**:

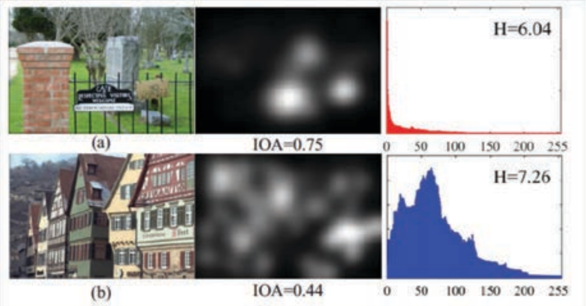
Visual attention complexity:

- **Inter-Observer Congruency/Agreement (IOC):**
  "Prediction of the inter-observer visual congruency (IOVC) and application to image ranking," O. Le Meur amd al. *ACM MM* 2011.
    - Requires fixations per observer!
    - ↓**IOC**, nothing stands out → **Exploratory** images
    - ↑**IOC**, things attracting attention → **Focused** images

- **Entropy of the saliency maps:** "A Saliency Dispersion Measure for Improving Saliency-Based Image Quality Metrics," , W Zhang and al. *IEEE Trans. CSVT* 2018.
    - ↓**Entropy**, saliency is concentrated → **Focused** images
    - ↑**Entropy**, saliency is diffused → **Exploratory** images

- Generally good correlation between both measures



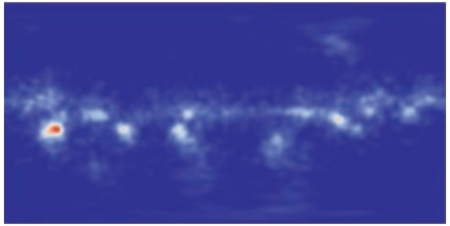


Using ground-truth eye+head tracking data

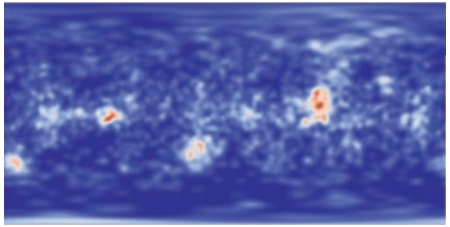
UNIVERSITÉ DE NANTES

CWI
Distributed and Interactive Systems

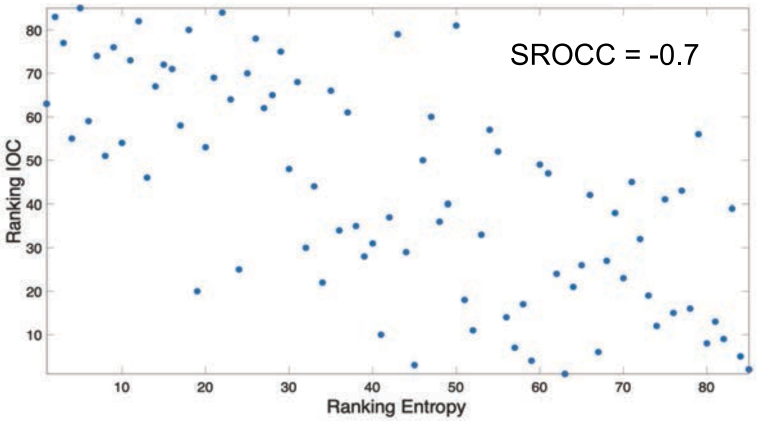
IOC on scanpaths and entropy on the head+eye saliency maps (sphere sampling)



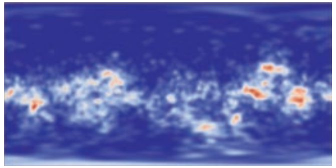
2nd most **focused** image by Entropy, 3rd by IOC.



Most **exploratory** image by Entropy, 2nd by IOC.



But some inconsistencies...

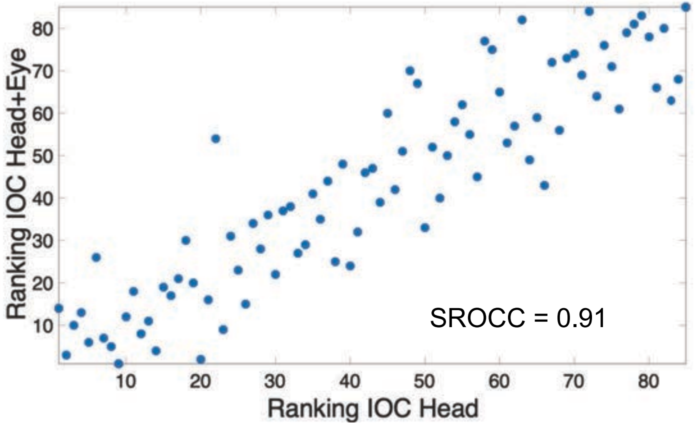


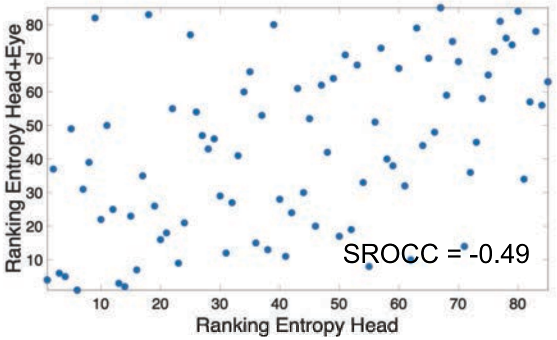
Ranking difference of 34 between Entropy and IOC

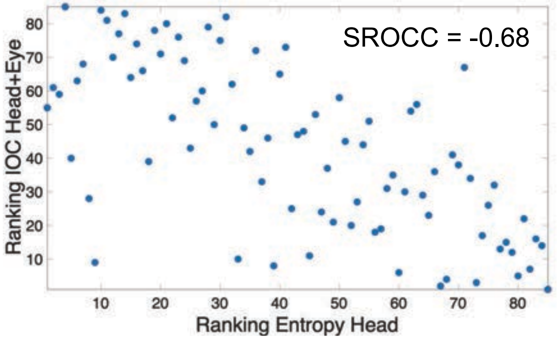
Using ground-truth ~~eye+~~head tracking data

IOC on head trajectories and entropy on the head saliency maps (sphere sampling):

- Good correlation between rankings from Head+Eye and Head **with IOC**
- Worse performance for entropy







Using saliency models

Can we characterize **without the ground-truth** tracking data (i.e., without previous exploration tests)?

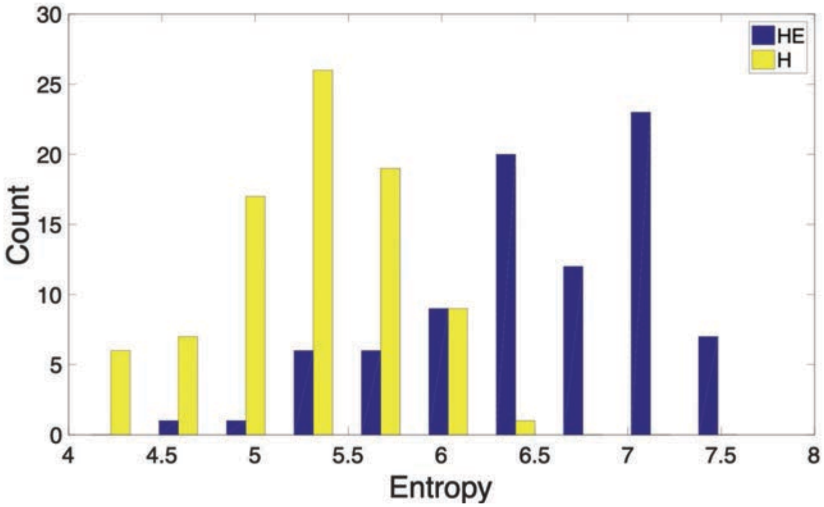
- Is entropy of the predicted saliency maps a good estimator?
- Considered models: Best models in GC **Salient360!** 2017/18* with online available code:
  - Prediction of **Head+Eye Saliency**:
    - M. Startsev, M. Dorr, “360-aware saliency estimation with conventional image saliency predictors”, Signal Processing: Image Communication, 2018. Code: http://www.michaeldorr.de/salient360/
  - Prediction of **Head Saliency**:
    - P. Lebreton, A. Raake, "GBVS360, BMS360, ProSal: Extending existing saliency prediction models from 2D to omnidirectional images", Signal Processing: Image Communication, 2018. Code:https://github.com/Telecommunication-Telemedia-Assessment/GBVS360-BMS360-ProSal

*https://salient360.ls2n.fr

Using saliency models

- Models can differentiate between clearly “focused” / “exploratory” images
  - Head+eye saliency predictions are more sensitive.
- Predictions of saliency models still improvable: limited use for direct ranking
  - See performance values in Salient360! website: https://salient360.ls2n.fr

Focused image:
- Entropy HE saliency model: 4.8
- Entropy H saliency model: 5.4



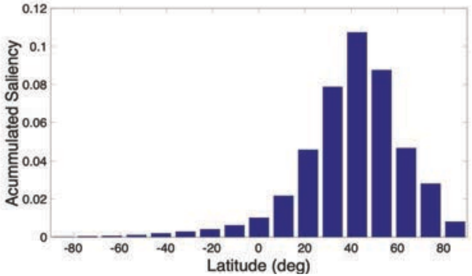
Exploratory image:
- Entropy HE saliency model: 7
- Entropy H saliency model: 6.3

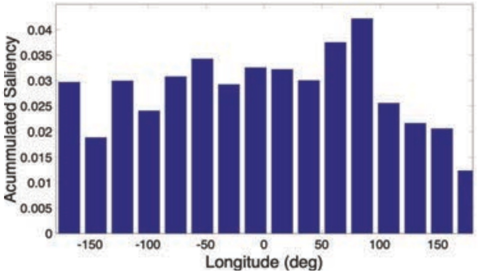
UNIVERSITÉ DE NANTES

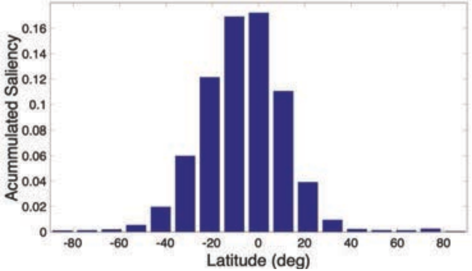
CWI Distributed and Interactive Systems

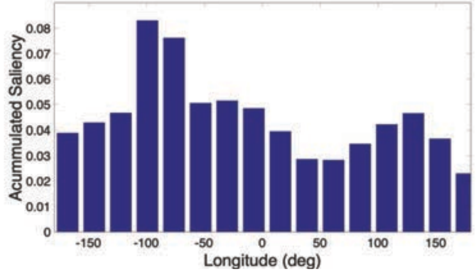
Using saliency models

- Prediction **saliency distribution** in **latitude** and **longitude**: Saliency models are **useful**!









Using saliency maps from Lebreton *et al.* model

Conclusions

UNIVERSITÉ DE NANTES

CWI Distributed and Interactive Systems

- **Geometrical domain** of the signal and **interactive navigation** by users must be considered when characterizing 360-degree content

- **Visual attention complexity** in addition to other characteristics

- **On-going work:** clustering based on Visual attention, characterization of 360-degree videos

Complexity measurement and characterization of 360-degree content

VQEG Meeting, March 2019, Berlin

Francesca De Simone

Distributed & Interactive Systems Group
Centrum Wiskunde & Informatica
Amsterdam, the Netherlands

Jesús Gutiérrez
Patrick Le Callet

Image, Interaction, Perception Group (IPI)
Laboratoire des Sciences du Numérique de Nantes (LS2N)
Ecole Polytechnique de l'Université de Nantes

Paper at IS&T Electronic Imaging 2019 Symposium