PROVISION

Perceptually Optimised Video Compression

A Marie Skłodowska-Curie Initial Training Network

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Naturalistic measurement of perceived video quality by measuring disruptions in free-viewing gaze patterns :

## **Case study of Interactive streaming**

**Yashas Rai, Gene Cheung, and Patrick Le Callet, in *Image Processing, 2016. ICIP 16. Proceedings. (submitted)*]**

# Agenda

- Why good quality estimation?
- Measuring quality naturalistically
- Modelling the Interactive Gaze System
- The GMM based HMM model
- Prediction Performance of the model
- Content dependence

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Need for good quality metrics

- Compress multimedia in such a way that distortions are cleverly embedded in the invisible areas
  - Reduced size implies less storage space on cloud -> Better profits!
  - More storage space for user-content on handheld devices!
- Efficient representation for faster transmission (broadcast and streaming)
  - Channels now approach Shannon limit whereas computing power is still increasing in a Moore Scale

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Ground Truth quality of a video

PROVISION
Perceptually Optimised Video Compression

- Quality assessment typically involves the subject examining a video sequence of several seconds and providing a quality score.
  - He is often searching for distortions in a very intricate manner.

- Attention Modulation in V4  (Response gain model):
  - Response to unattended stimulus reduces even though its within the receptive field – Neuronal firing is 51% greater with attention [1]
  - Contrast detection thresholds 20%,  Contrast Discrimination 40-50%, Orientation Discrimination 70% lower [1]
  - Attention influences motion processing by enhancing processing of relevant and suppressing the processing or irrelevant objects [2]
  - Tracking in the fovea task combined with a Number recognition in periphery – More is the attention to the  foveal task, more is the drop in performance of the peripheral task [3]

- **Do we examine videos in a similar manner when we freely watch videos at home?**


[1] J. Moran and R. Desimone, "Selective attention gates visual processing in the extrastriate cortex," *Science*, vol. 229, no. 4715, pp. 782–784, 1985.
[2] J. Braun and B. Julesz, "Withdrawing attention at little or no cost: detection
[3] B. Khurana and E. Kowler, "Shared attentional control of smooth eye movement and perception," *Vision research*, vol. 27, no. 9, pp. 1603–1618, 1987.



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Subjective Experiment : Gaze contingency

PROVISION
Perceptually Optimised Video Compression

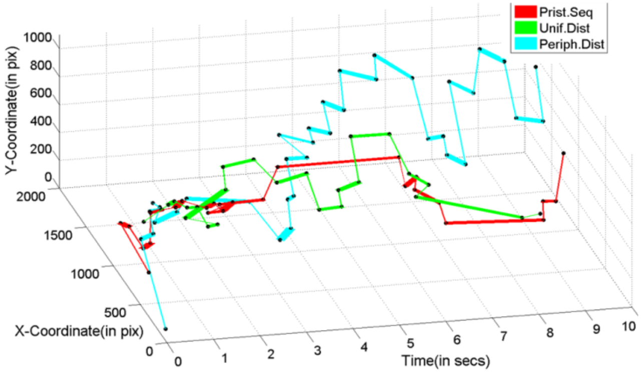
- Alter the quality in the visual peripery in real-time in accordance to gaze position.
- HEVC distortions
  - 4 different Quantizations
  - Para and Peri-Fovea (5.09 and 10.9)

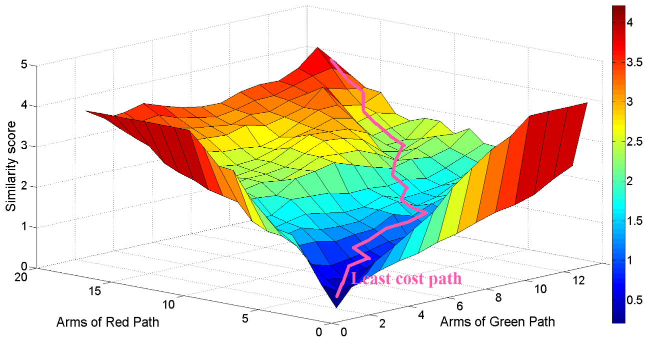
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

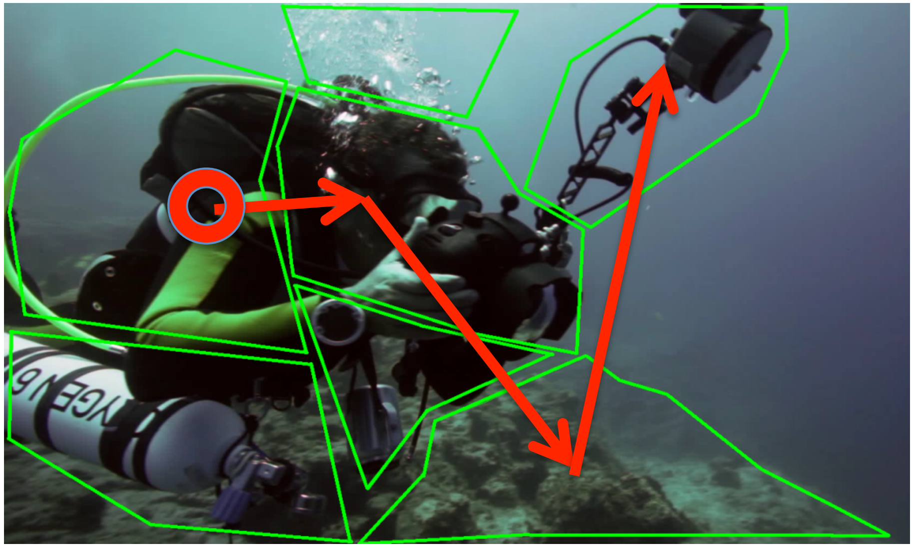
IRCCyN

# Measuring Disruptions

- Vector Similarity: measuring the difference in scan-path trajectories
- Comparing object transitions : **D-B-B-C-C-C-A**, **D-D-C-B-A** followed by Levenshtein similarity of string patterns.



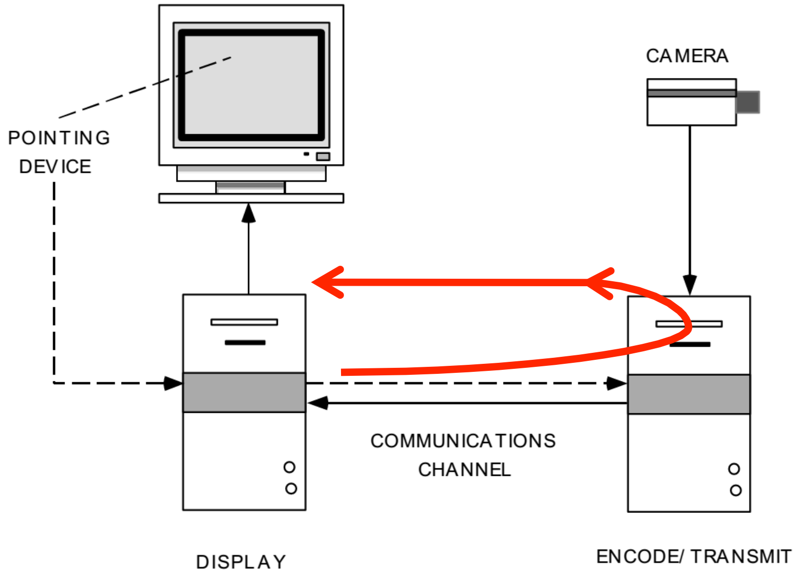



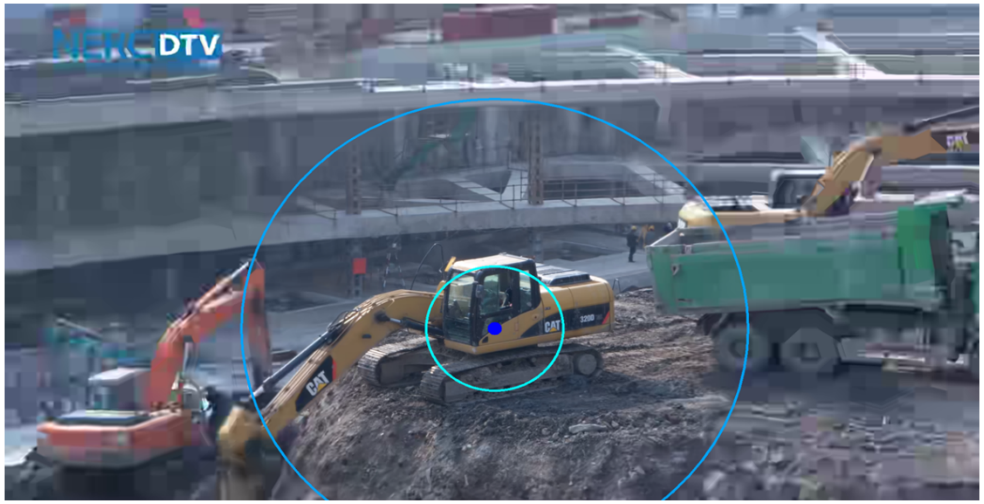



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Interactive Video Streaming

- A reverse channel that transmits gaze data from the client end
- Virtual Reality, Panoramic Displays
- Practical considerations
  - Availability of a camera at the client end
  - Network delay
  - Real-time encoding

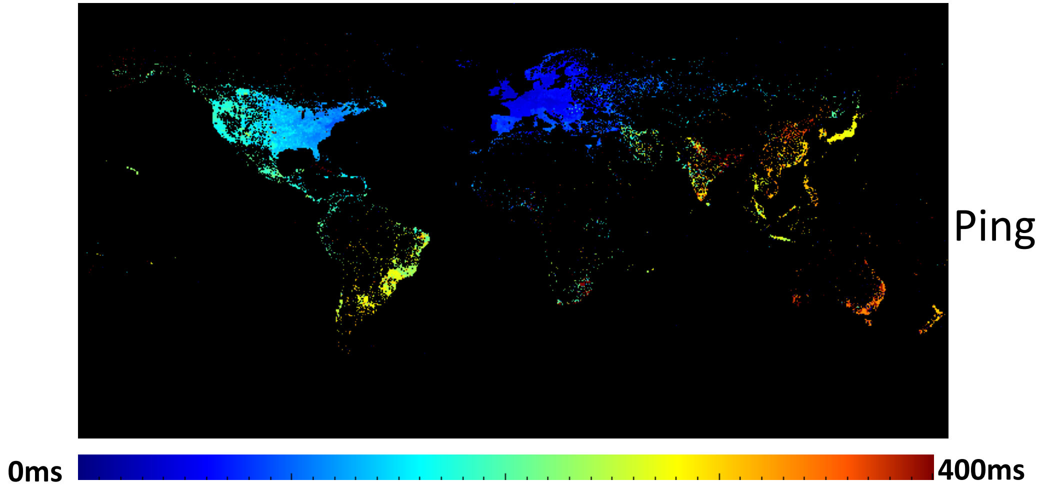



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# RTT delay

- Round Trip Time: The time from which the user makes a head movement till the time the actual update happens on the display
  - Eye Tracker Latency [<2 ms]
  - Network Propagation Delay (x2) [180-190 ms]
  - Encoding Delay [10 ms]
  - Display Refresh Rate [8.3 ms]
- Considering wide geographies, we may consider the latency to be ~200 ms

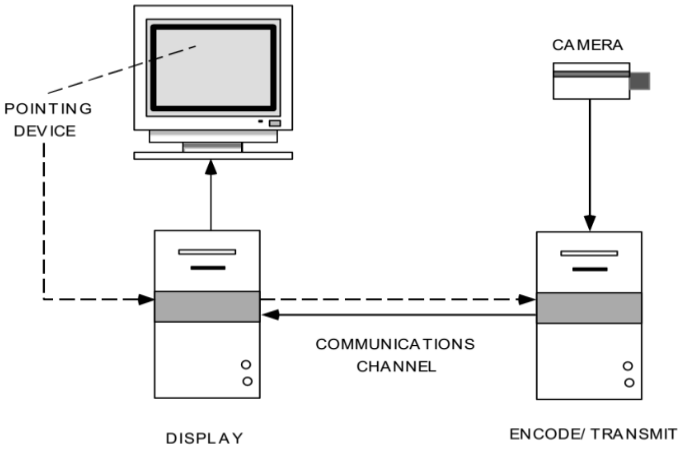



Ping times from Oxford [folk.uio.no]

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Goals of the research

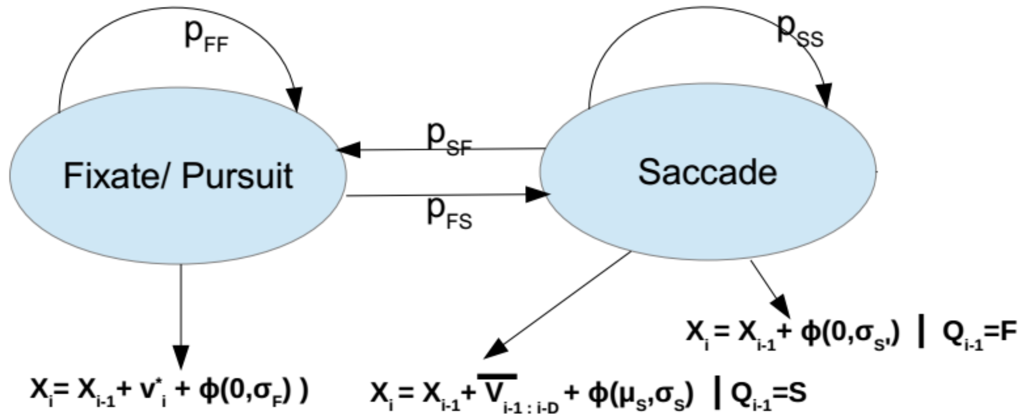
- Does the presence of peripheral HEVC artefacts/ uniform HEVC artefacts alter this prediction accuracy?
    - **Same gaze model designed for the pristine sequence works here, as well?**
    - **HEVC artefacts introduce new Bottom-up effects?**




This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Predicting future gaze locations

PROVISION
Perceptually Optimised Video Compression

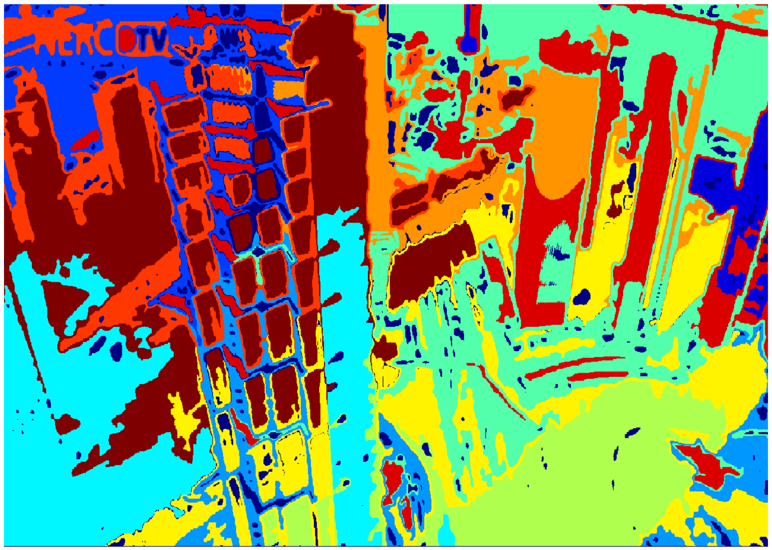


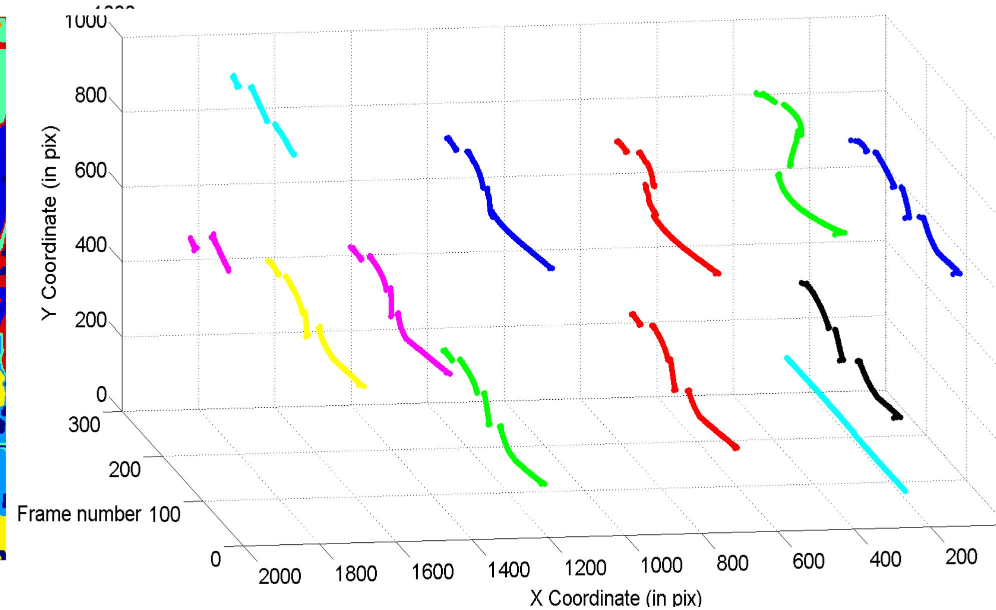

- Types of Eye-movements
  - Saccade – Fast movement of the eye to attend to a different region
  - Fixate – Continue to gaze at the same position for detailed examination
  - Pursuit – Follow the trajectory of a moving object with the eyes
- Terminologies:
  - $X_i$ and $X_{i-1}$ are the (x,y) spatial locations
  - $Q_i$ and $Q_{i-1}$ are the hidden states
  - $V_i$ is determined by the motion of the maximum likelihood object
  - $V_{i-1:i-D}$ indicates the average of the previous saccade speeds
  - $\Phi()$ is the normal noise

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

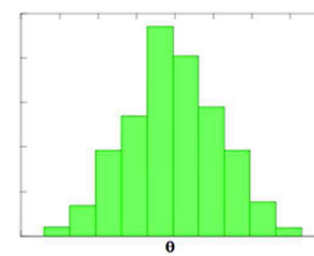
IRCCyN

# Gaze prediction: Tracking objects under motion






- Segmenting the image into discrete super-pixels using texture and color based super-pixelation.
- Motion consistency analysis in each region




[Anil K Jain and Farshid Farrokhnia, "Unsupervised texture segmentation using gabor filters," in *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*. IEEE, 1990, pp. 14–19.]
[M.A. Hasan, Min Xu, Xiangjian He, and Changsheng Xu, "CAMHID: Camera motion histogram descriptor and its application to cinematographic shot classification," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 10, pp. 1682–1695, Oct 2014]

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

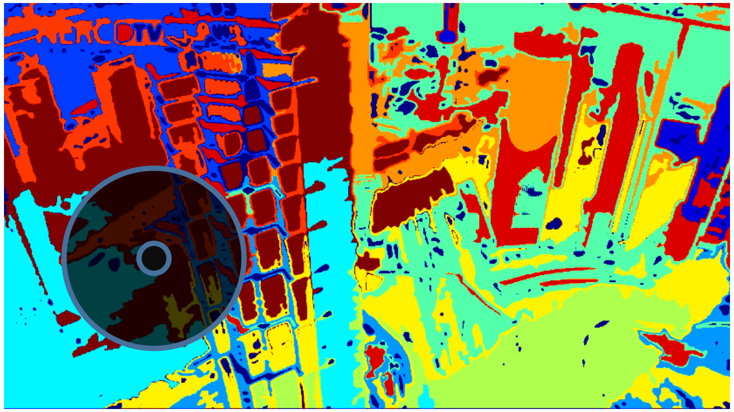
PROVISION
Perceptually Optimised Video Compression

# Determining the velocity $v_i$* of pursuit

- Due to the noise in the gaze data, we do not really know which object the user is actually tracking
- Consider all the objects within a radius of 2 degrees and take the one that maximizes the posterior probability

$$P_Q(X_{i+1}|X_i) = \max_{v_i \in M(X_i)} f_{\sigma_{F^2}}(X_{i+1} - X_i - v_i)$$

$$v_i^* = arg \max_{v_i \in M(X_i)} f_{\sigma_{F^2}}(X_{i+1} - X_i - v_i)$$



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Parameter estimation by training with real gaze data

PROVISION
Perceptually Optimised Video Compression

- Gaze data from 30 observers
  - Estimate transition probabilities $[p_{FF}, p_{FS}, p_{SF}, p_{SS}]$
  - Estimate the Gaussian processes $[\sigma_F, \sigma_S, \sigma_{S'}]$
  - Estimate the state probabilities $[p_F, p_S]$
- Offline Learning: Baum Welch Forward-Backward propagation
- Online Learning: Forward only propagation

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

IRCCyN

PROVISION
Perceptually Optimised Video Compression

# LDS based prediction of future locations

$$\mathbf{Y}_n = \underbrace{\begin{bmatrix} 1 & (1-\beta) \\ 0 & (1-\beta) \end{bmatrix}}_{\mathbf{F}_n} \mathbf{Y}_{n-1} + \underbrace{\begin{bmatrix} \beta c_1 \ldots \beta c_Z \\ \beta c_1 \ldots \beta c_Z \end{bmatrix}}_{\mathbf{B}_n} \underbrace{\begin{bmatrix} v_{n-1}^1 \\ \vdots \\ v_{n-1}^Z \end{bmatrix}}_{\mathbf{u}_n}$$

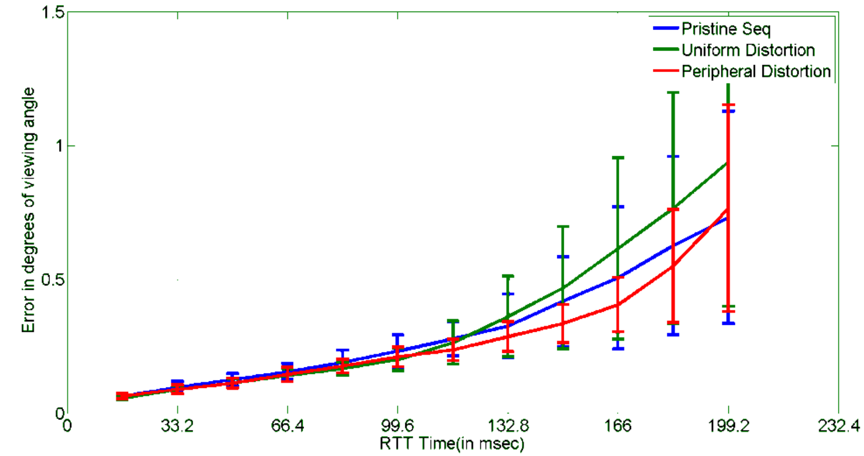
- Perform prediction only if in current as well as RTT seconds in future we have high chances of remaining in Fixation state
- $Y_n$ contains the position and velocity vectors.
- $V^Z_{n-1}$ are the motion vector of surrounding pixels, each weighted by $c_z$
- This is performed repeatedly, RTT samples into the future

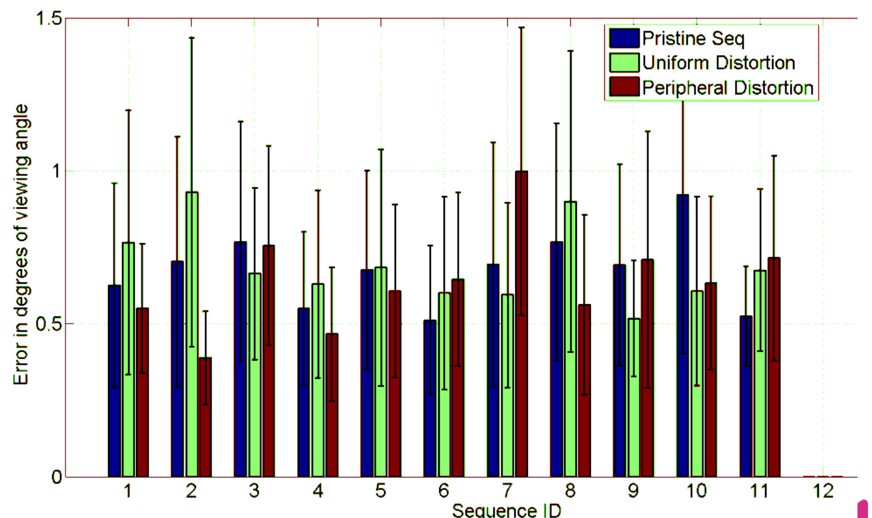
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

IRCCyN

# Prediction performance

- Peripheral or uniform distortions have no special effect on gaze predictability!
- The model trained for the pristine case could also predict these cases well!
- Can perform the prediction to within 1.5 degrees of viewing angle error (200ms)

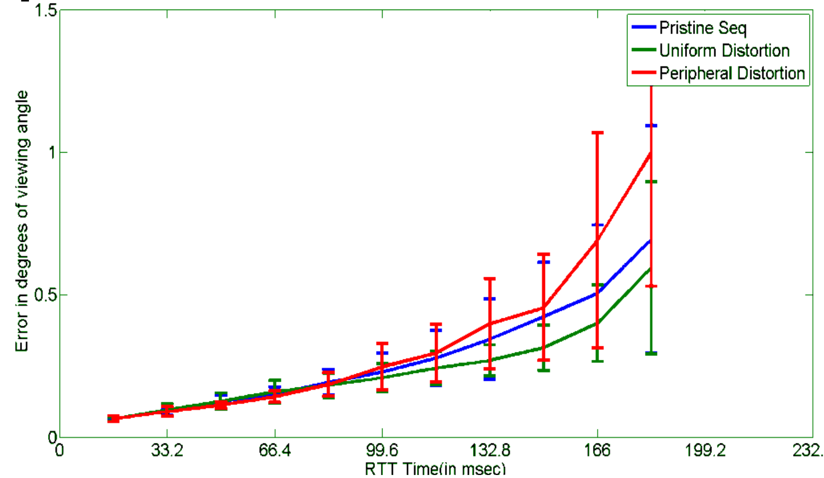


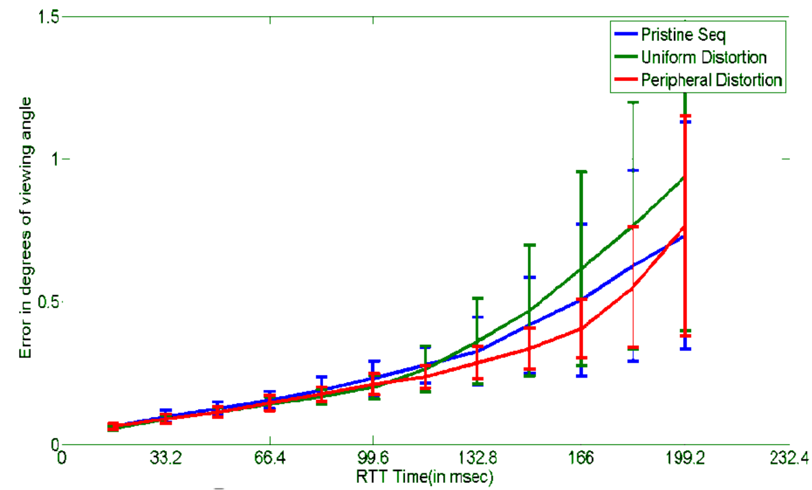

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Prediction by sequence type

- Measuring temporal activity with MC-Saliency maps
- High temporal activity -> High prediction error for peripheral distortions
- Low temporal activity -> High prediction error for uniform distortions




Diver (MC-Saliency=$216 \times 10^{-4}$)




Construction-Field (MC-Saliency=$1.88 \times 10^{-4}$)

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Statistical test for verification

- Check how well the Kalman predictor predicts the gaze locations (Euclidean error)
- Is this error significantly higher, while predicting gaze data in case of peripheral / uniform distortion?

**Table II**

| | | $RTT = 100ms$ | | $RTT = 200ms$ | |
|---|---|---|---|---|---|
| Seq.ID | | Pristine vs Unif.Dist | Pristine vs Periph.Dist | Pristine vs Unif.Dist | Pristine vs Periph.Dist |
| 1 | Construction Field | 0.23 | 0.69 | 0.78 | 0.86 |
| 2 | Fountains | 0.26 | 0.25 | 0.44 | 0.10 |
| 3 | Library | 0.48 | 0.53 | 0.69 | 0.26 |
| 4 | Resid. Building | 0.42 | 0.59 | 0.50 | 0.61 |
| 5 | Tall Buildings | 0.61 | 1.00 | 0.71 | 0.62 |
| 6 | NTIA-Redgold | 0.83 | 0.42 | 0.83 | 0.39 |
| 7 | Diver | 0.64 | 0.27 | 0.42 | 0.84 |
| 8 | Lobsters | 0.42 | 0.50 | 0.18 | 0.67 |
| 9 | Evening Walk | 0.53 | 0.83 | 0.33 | 0.54 |
| 10 | Traffic Building | 0.07 | 0.09 | 0.10 | 0.13 |
| 11 | NTIA-Purple | 0.17 | 0.16 | 0.81 | 0.43 |
| 12 | Tree Shade | - | - | 0.51 | 0.99 |

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Key take-aways

- Peripheral as well as uniform distortions do not affect the prediction capabilities of the model
- Such a model maybe safely used for gaze prediction in interactive streaming systems
- Gaze disruption measurement can be used to test the annoyance of video distortions in a naturalistic manner!

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231

# Thank-You / Questions?

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608231