The Prefetch Aggressiveness Tradeoff in 360 Video Streaming

Mathias Almquist, *Linköping University*

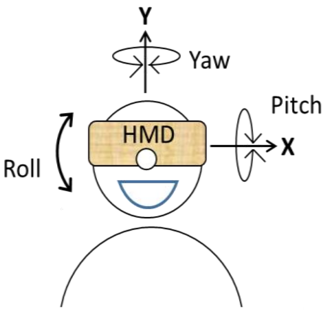
Viktor Almquist, *Linköping University*

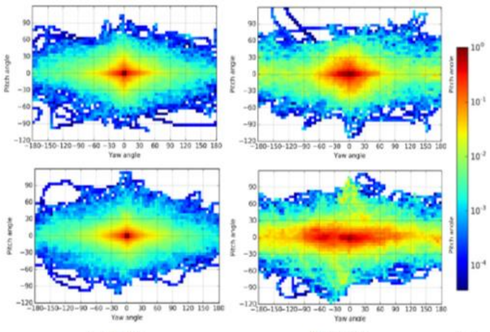
Vengatanathan Krishnamoorthi, *Linköping University*

Niklas Carlsson, *Linköping University*

Derek Eager, *University of Saskatchewan*

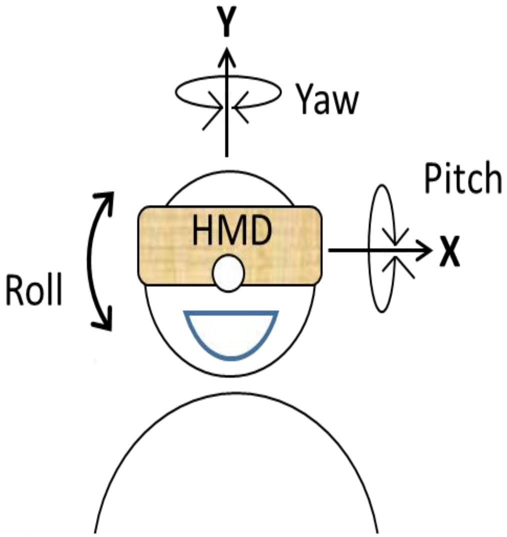
LINKÖPING UNIVERSITY

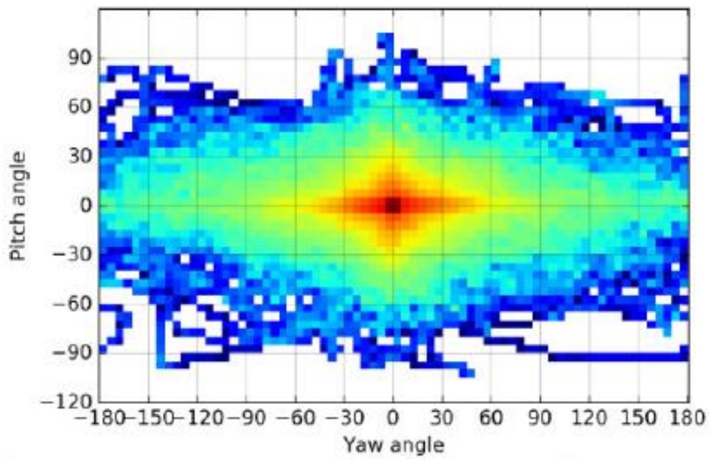


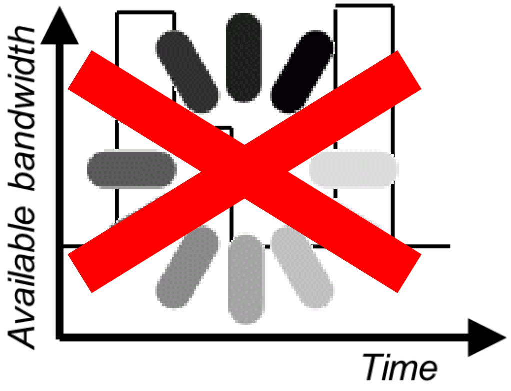


Proc. ACM MMSys, Amsterdam, June 2018

Prefetching during 360 video streaming



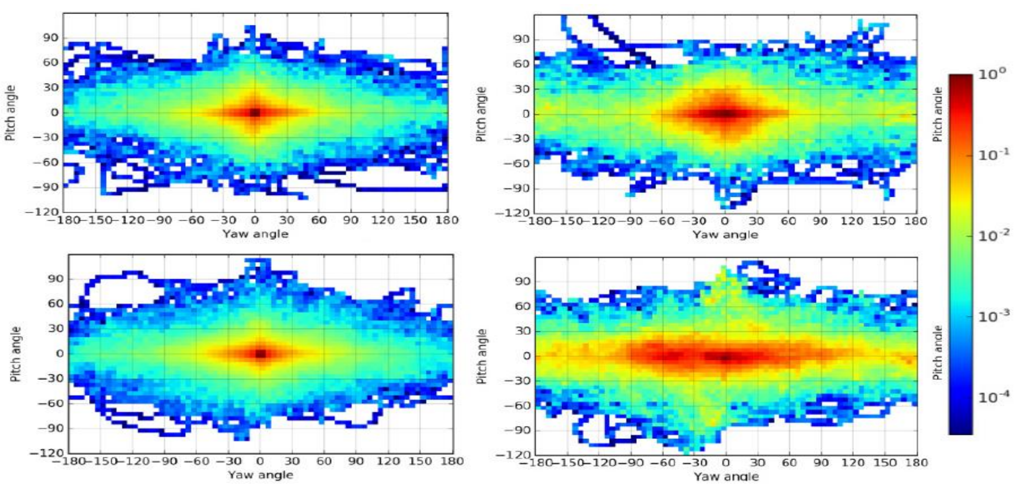




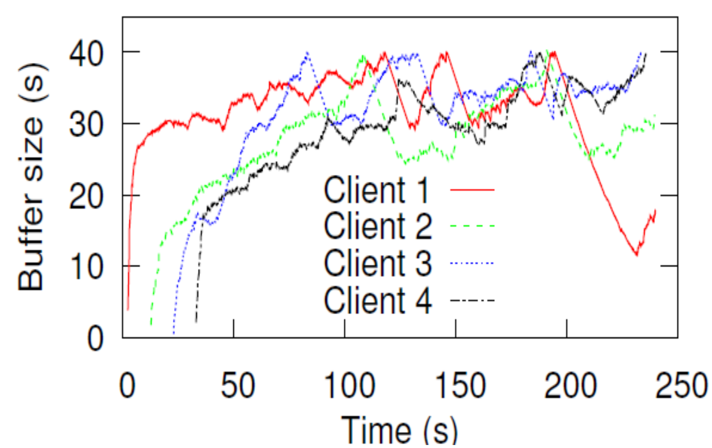
- 360 videos are large and consume lots of bandwidth
- Recently, many papers consider techniques that allow prefetching of alternative video qualities in each viewing direction
- However, neither head-movement prediction nor bandwidth prediction is perfect …

Prefetch aggressiveness tradeoff

1. Uncertainty in the viewer direction
   - Prediction is most accurate when done close to the playback deadline of each frame
2. Uncertainty in the available bandwidth
   - Buffer typically used to protect against stalls caused by (future) bandwidth variations (or instability due to competing players)
   - Larger buffer (as typically used by HAS/DASH clients) provides better protection against stalls and bandwidth variations



(a) T=2s (b) T=20s (c)



Prefetch aggressiveness tradeoff

1. Uncertainty in the viewer direction
   - Prediction is most accurate when done close to the playback deadline of each frame
2. Uncertainty in the available bandwidth
   - Buffer typically used to protect against stalls caused by (future) bandwidth variations (or instability due to competing players)
   - Larger buffer (as typically used by HAS/DASH clients) provides better protection against stalls and bandwidth variations

Addressing both these uncertainties in simultaneously results in a prefetch aggressiveness tradeoff, not addressed by prior works

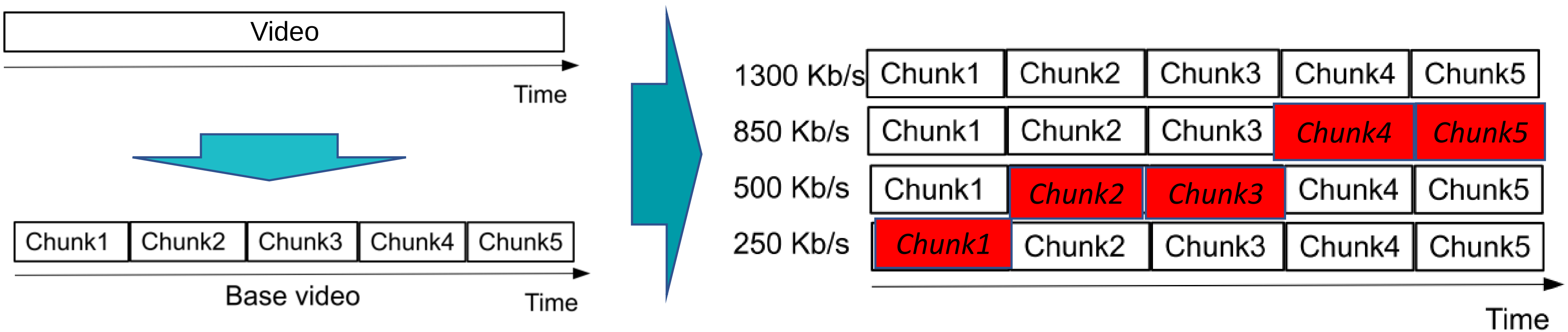
- E.g., how far ahead in time should prefetching be done?
- Important problem with conflicting goals

Contributions

1. Data-driven head-movement characterization
   - Head movements analysis over different time scales and for different categories of 360 video
2. Optimized buffer-quality tradeoffs
   - Optimization framework that captures tradeoff between the goals of prefetching far ahead (to protect against bandwidth variations and stalls) and the expected quality selection for each viewing direction (based on the conditional probabilities of each direction)
   - Use framework and data to derive qualitative and quantitative insights into the best tradeoff
3. Data-driven discussion of further design optimizations
   - Motivated by the observations from the above characterization

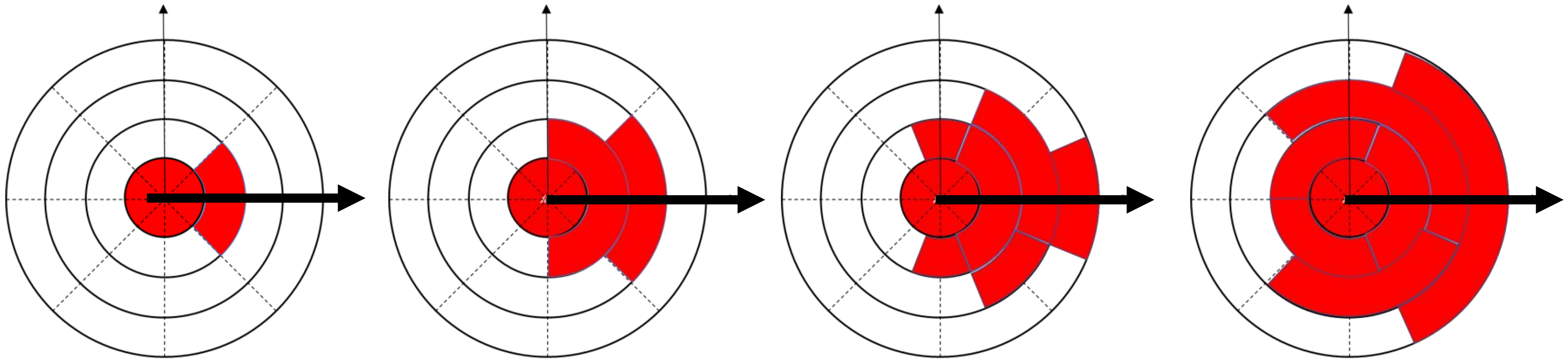
HAS/DASH + Tiling

HTTP-based Adaptive Streaming (HAS)



- HTTP-based streaming
  - Video is split into chunks
  - Support for VoD (Video on Demand) functionalities
- HTTP-based adaptive streaming
  - Each chunk in multiple bitrates (qualities)
  - Clients adapt quality encoding based on buffer/network conditions

360 HAS with tiles

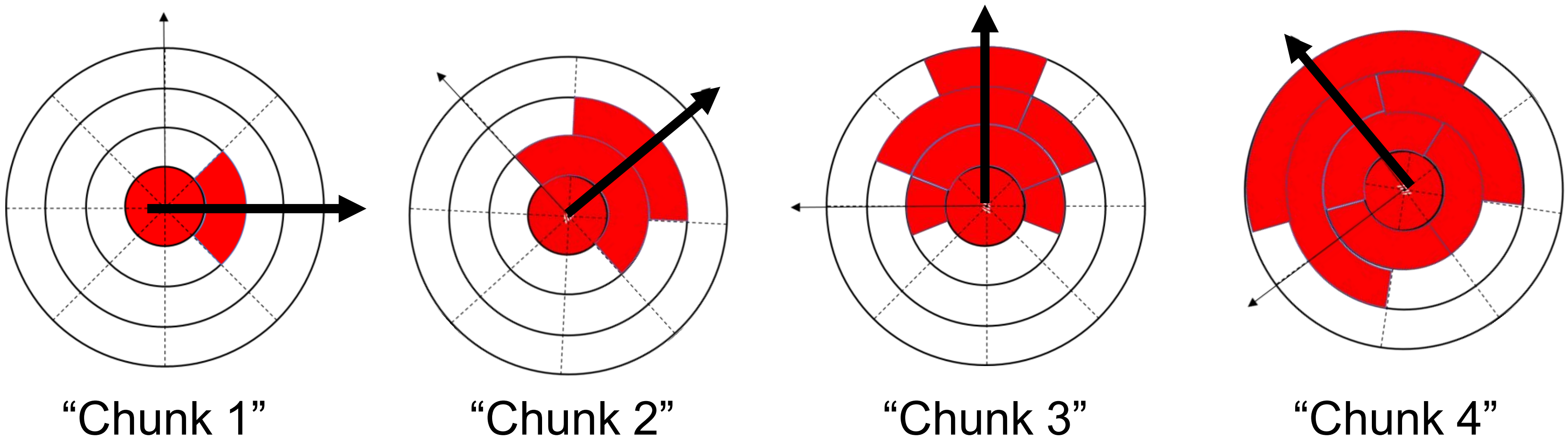


"Chunk 1" "Chunk 2" "Chunk 3" "Chunk 4"

- In addition to chunks, we have
  - Tiles of different quality in each direction

- Clients adapt quality encoding of each chunk and tile based on **both**
  - buffer/network conditions, and
  - expected view field

360 HAS with tiles

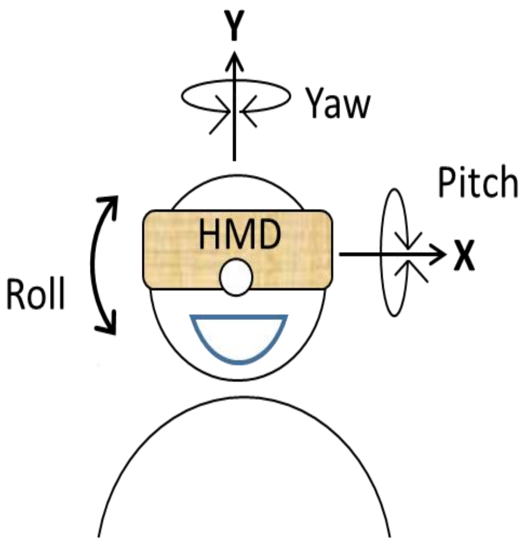


- In addition to chunks, we have
  - Tiles of different quality in each direction

- Clients adapt quality encoding of each chunk and tile based on **both**
  - buffer/network conditions, and
  - expected view field

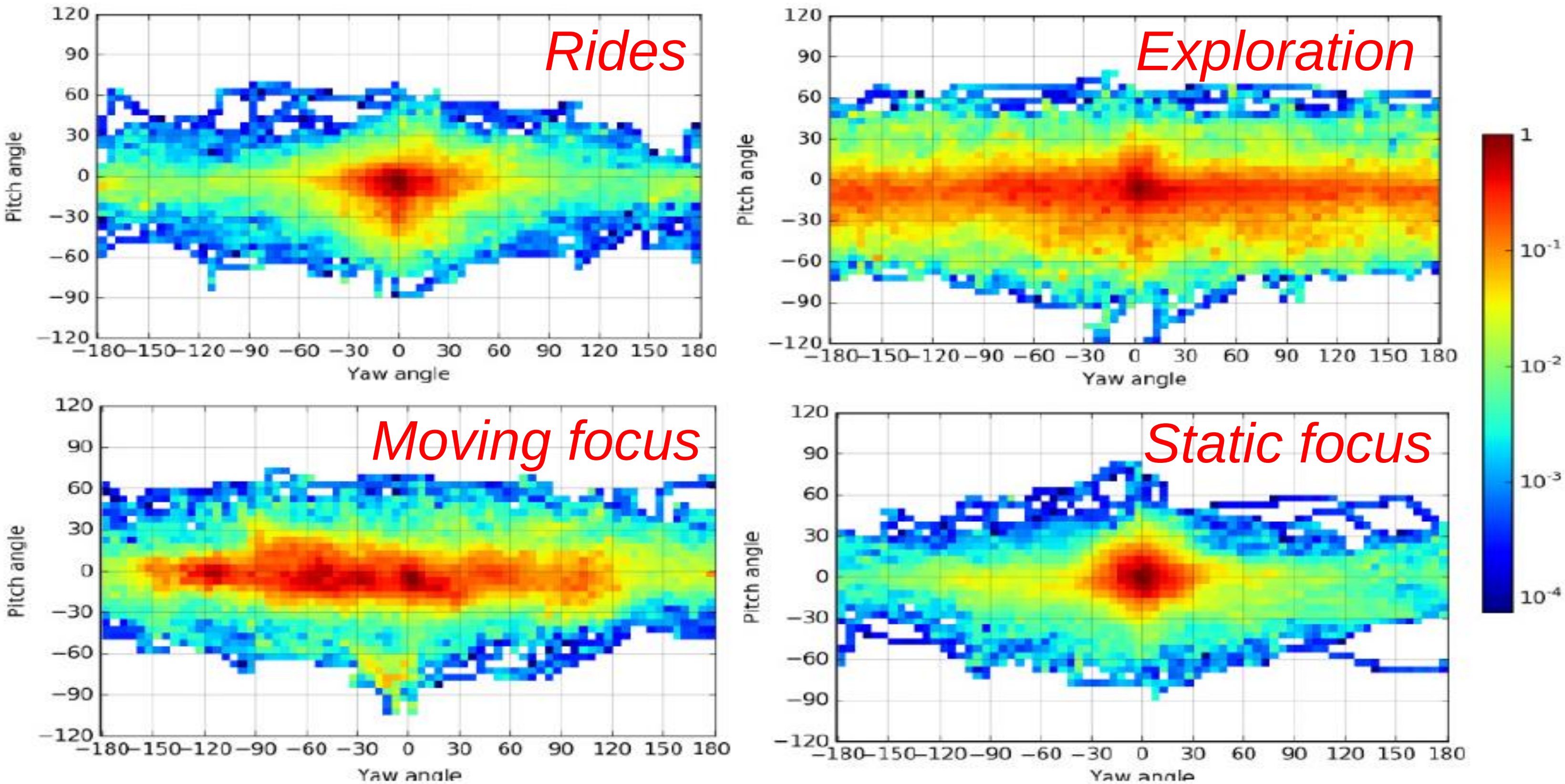
User study and dataset

User study and dataset



- Oculus rift
- 30 YouTube 360 videos with 4K resolution
- Duration 1-5 minute (3 min on average)
- Five categories
  - Rides: “virtual ride ...”
  - Exploration: “no particular focus ...”
  - Static focus: “main focus of attention static ...”
  - Moving focus: “object of attention moves ...”
  - Miscellaneous: “unique feel ...”
- 32 users, 45 x 45 min sessions (439 viewings)
  - Semi-random view order
  - One “representative” video of each category viewed by all 32 users; rest got 8-13 views

Angular utilization across the session

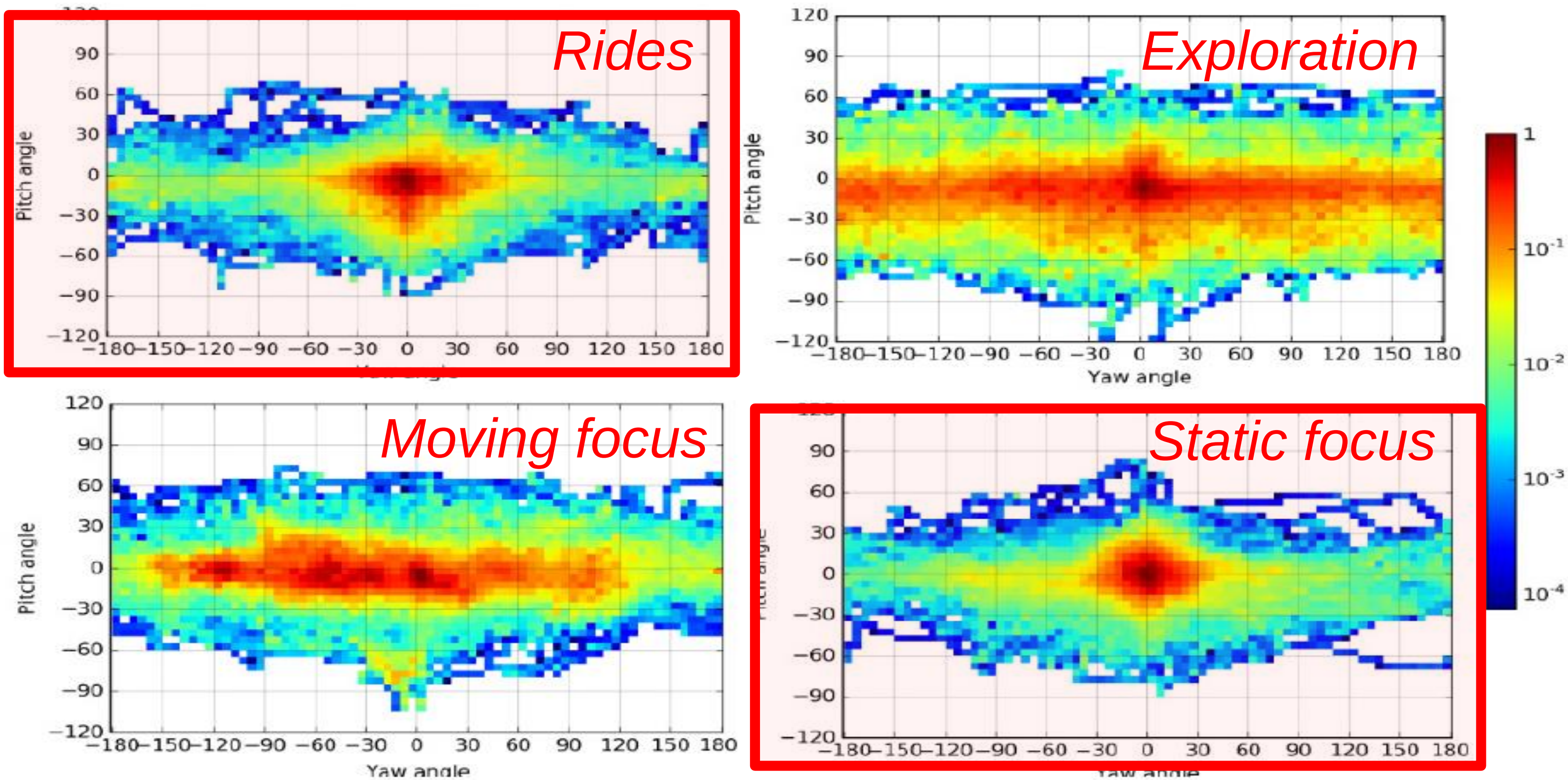


Substantial differences between categories

- Rides and static focus see the least head movements
- Exploration (and moving focus) sees the most variation

Yaw the most dominant orientation movement across the categories

Angular utilization across the session

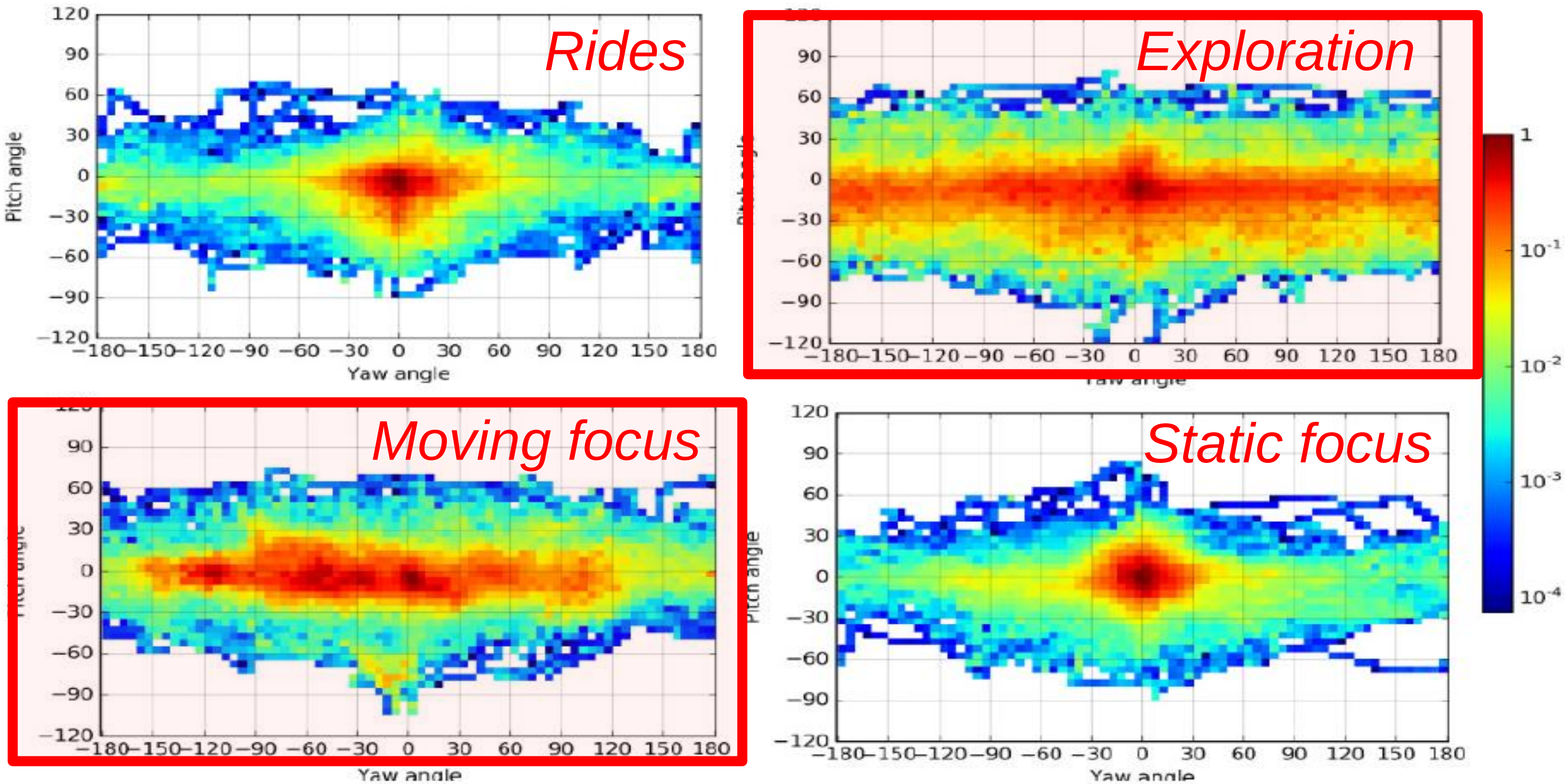


Substantial differences between categories

- Rides and static focus see the least head movements
- Exploration (and moving focus) sees the most variation

Yaw the most dominant orientation movement across the categories

Angular utilization across the session

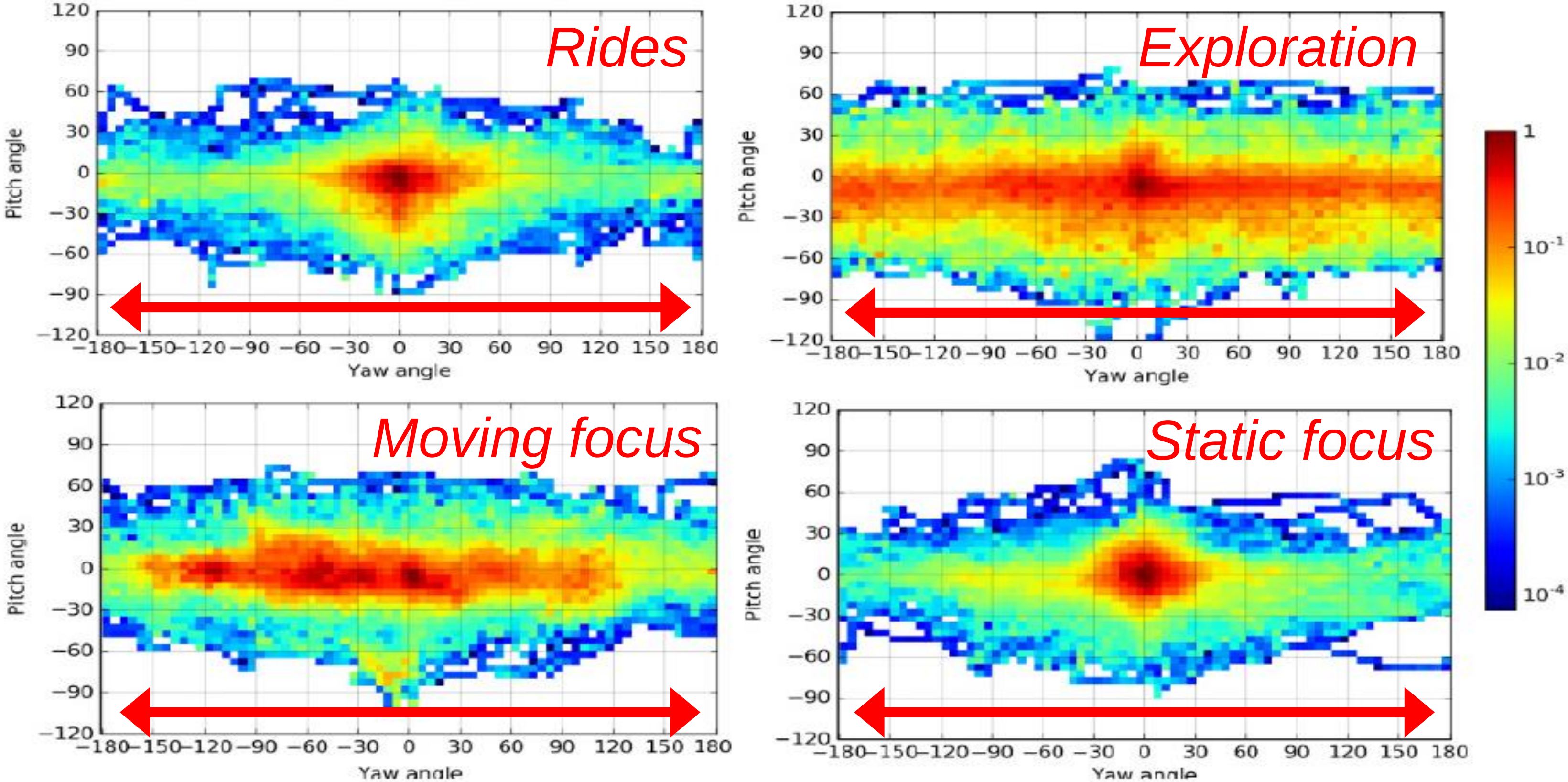


Substantial differences between categories

- Rides and static focus see the least head movements
- Exploration (and moving focus) sees the most variation

Yaw the most dominant orientation movement across the categories

Angular utilization across the session

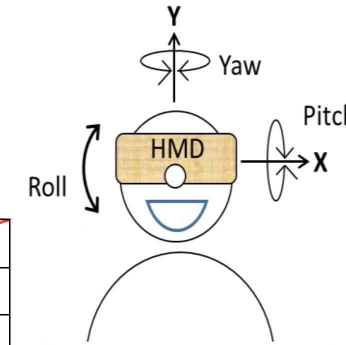


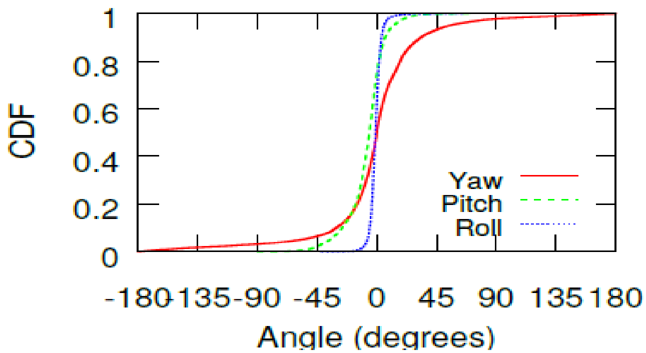
Substantial differences between categories

- Rides and static focus see the least head movements
- Exploration (and moving focus) sees the most variation

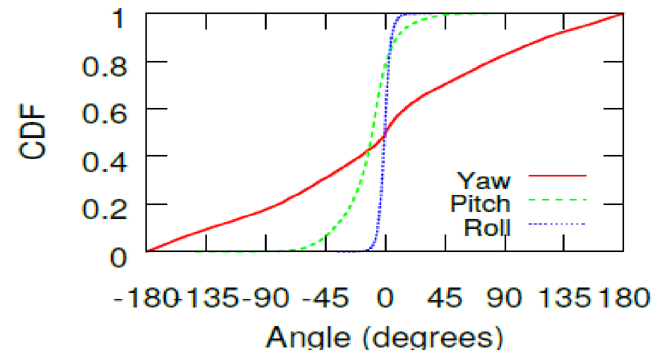
Yaw the most dominant orientation movement across the categories

Angular utilization across the session

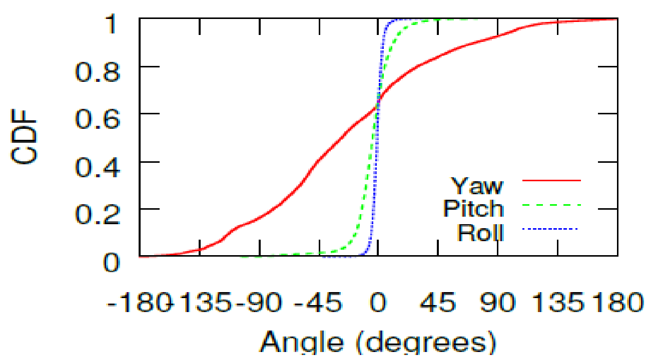




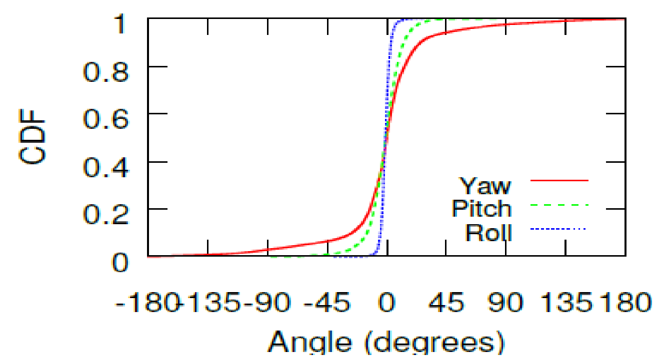
(a) Rides



(b) Exploration



(c) Moving focus



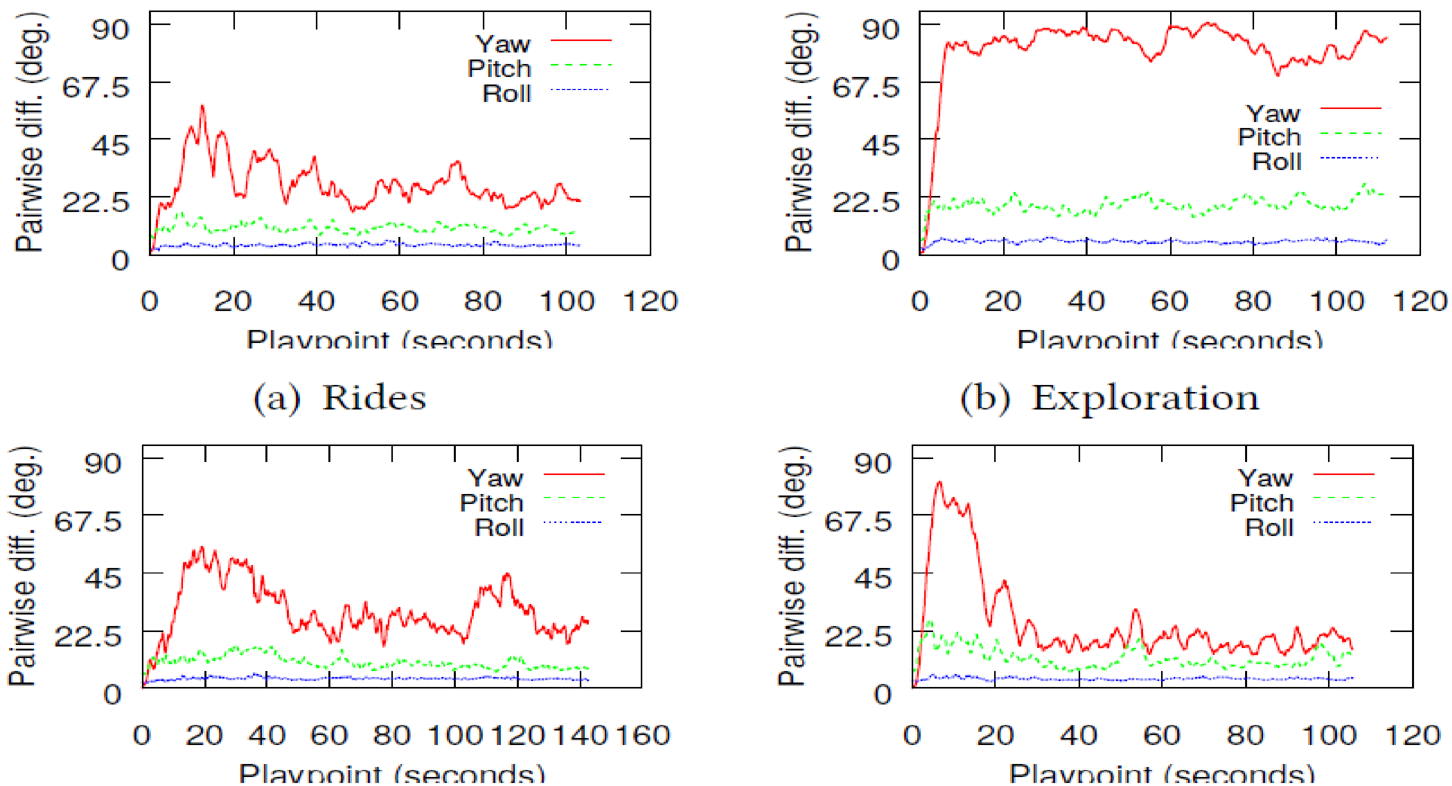
(d) Static focus

Substantial differences between categories

- Rides and static focus see the least head movements
- Exploration (and moving focus) sees the most variation

Yaw the most dominant orientation movement across the categories

Average differences between users watching same video



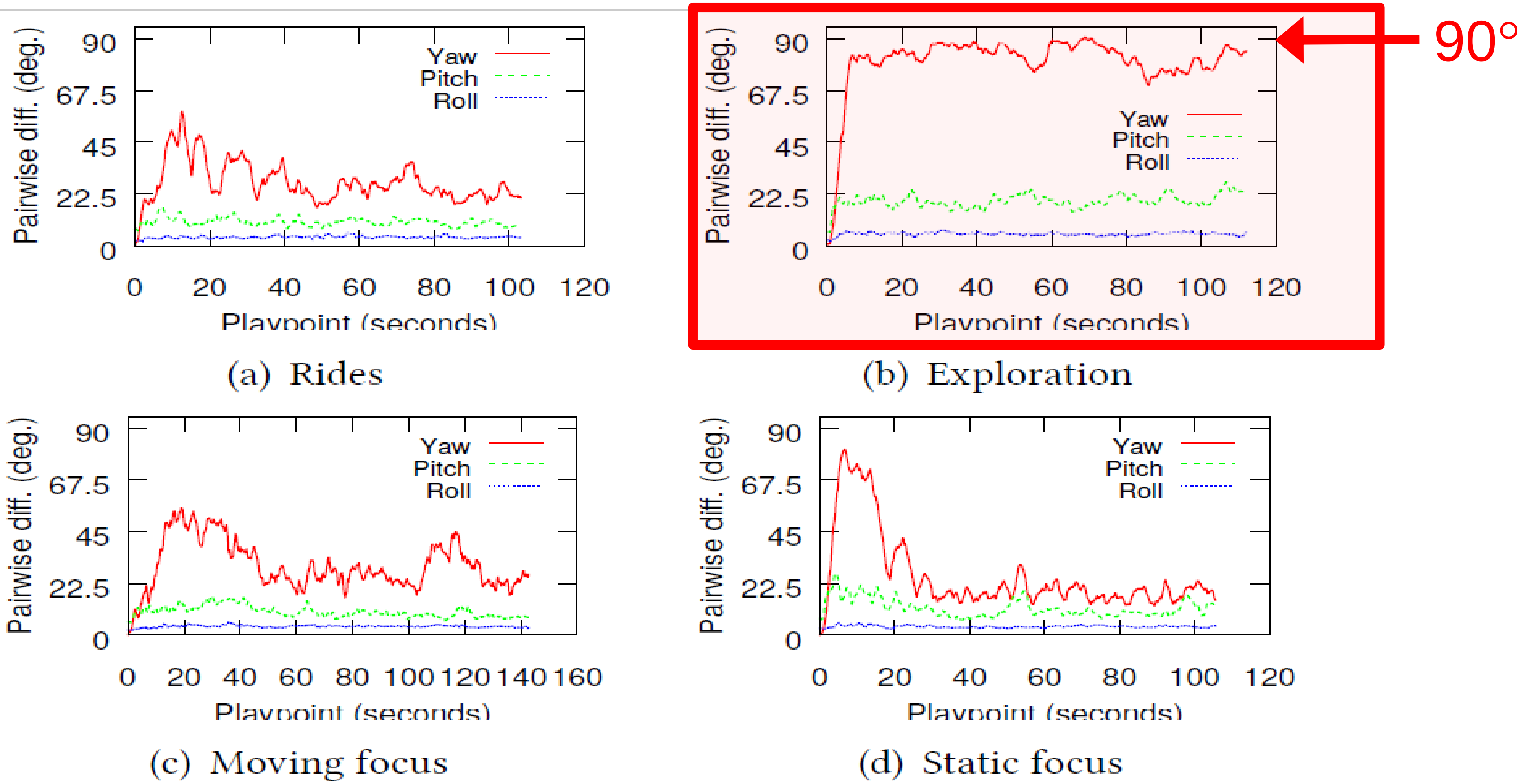
(a) Rides (b) Exploration

(c) Moving focus (d) Static focus

Again, substantial differences between categories

- Exploration close to 90 degrees (fully independent)
- For moving focus, past clients can be good predictor
- Others, especially static focus, has initial exploration phase (or reduced exploration) ... Shows need to look at longer user sessions

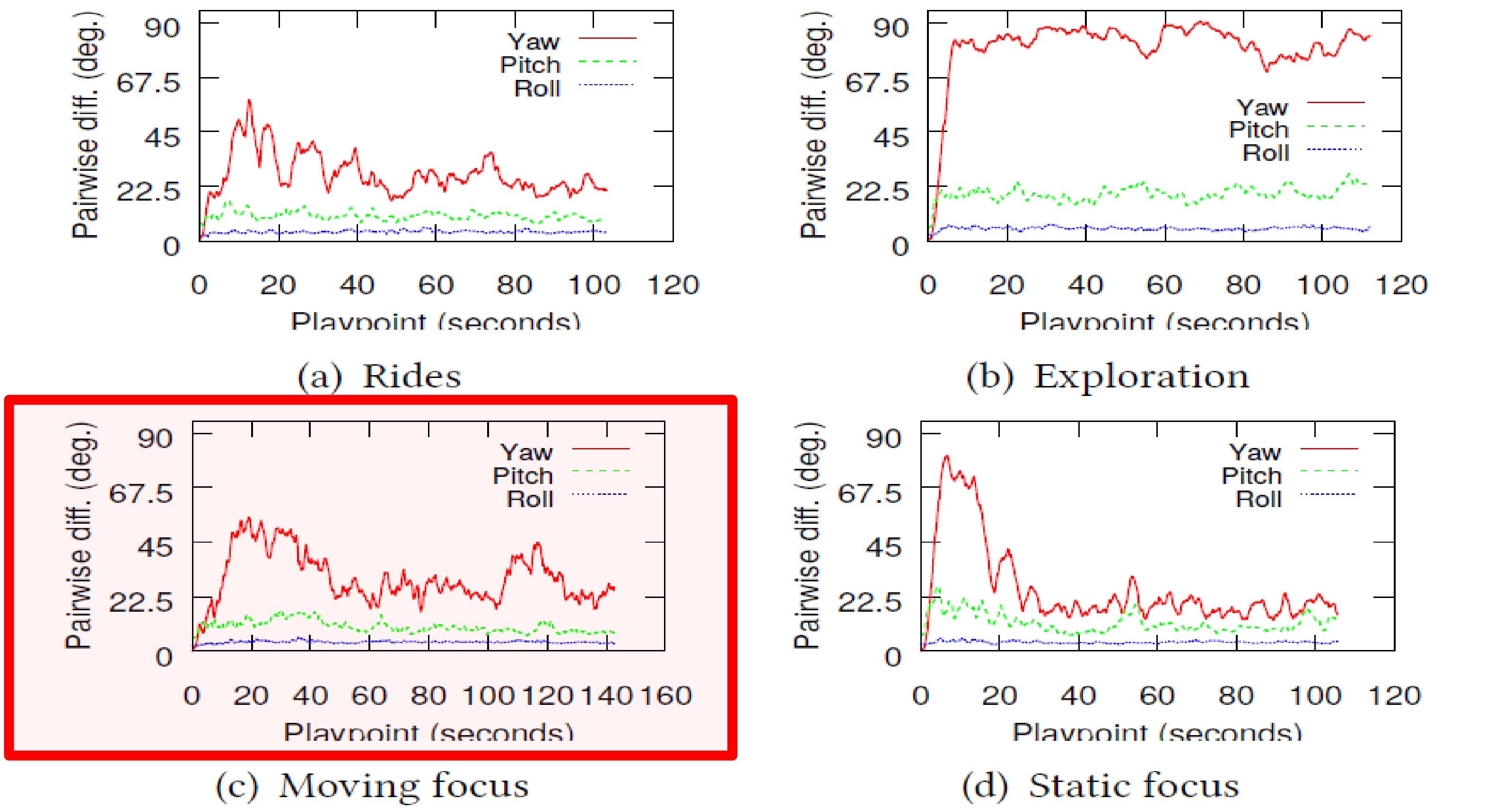
Average differences between users watching same video



Again, substantial differences between categories

- Exploration close to 90 degrees (fully independent)
- For moving focus, past clients can be good predictor
- Others, especially static focus, has initial exploration phase (or reduced exploration) ... Shows need to look at longer user sessions

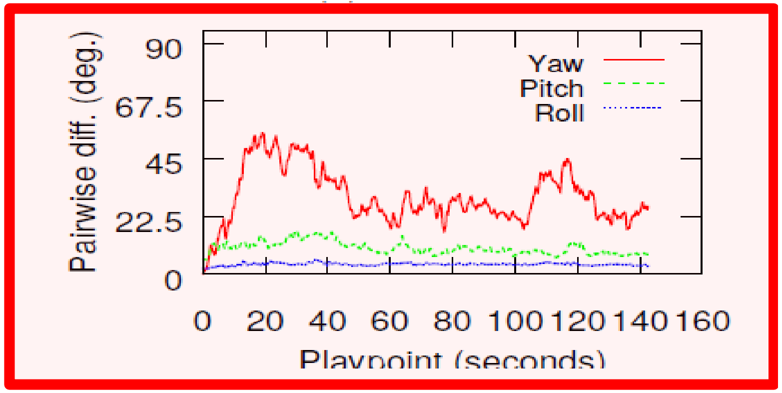
Average differences between users watching same video



Again, substantial differences between categories

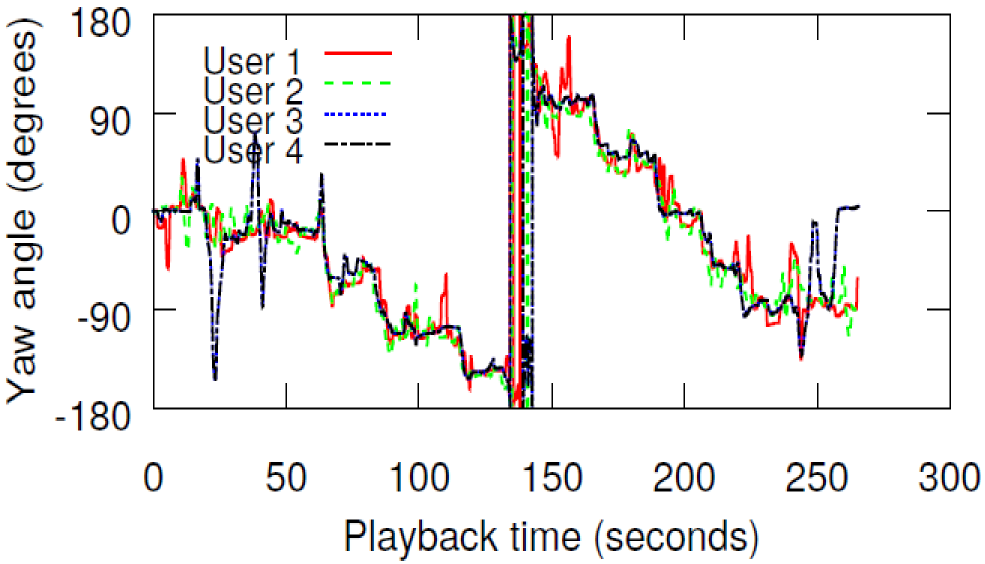
- Exploration close to 90 degrees (fully independent)
- For moving focus, past clients can be good predictor
- Others, especially static focus, has initial exploration phase (or reduced exploration) … Shows need to look at longer user sessions

Average differences between users watching same video



(c) Moving focus

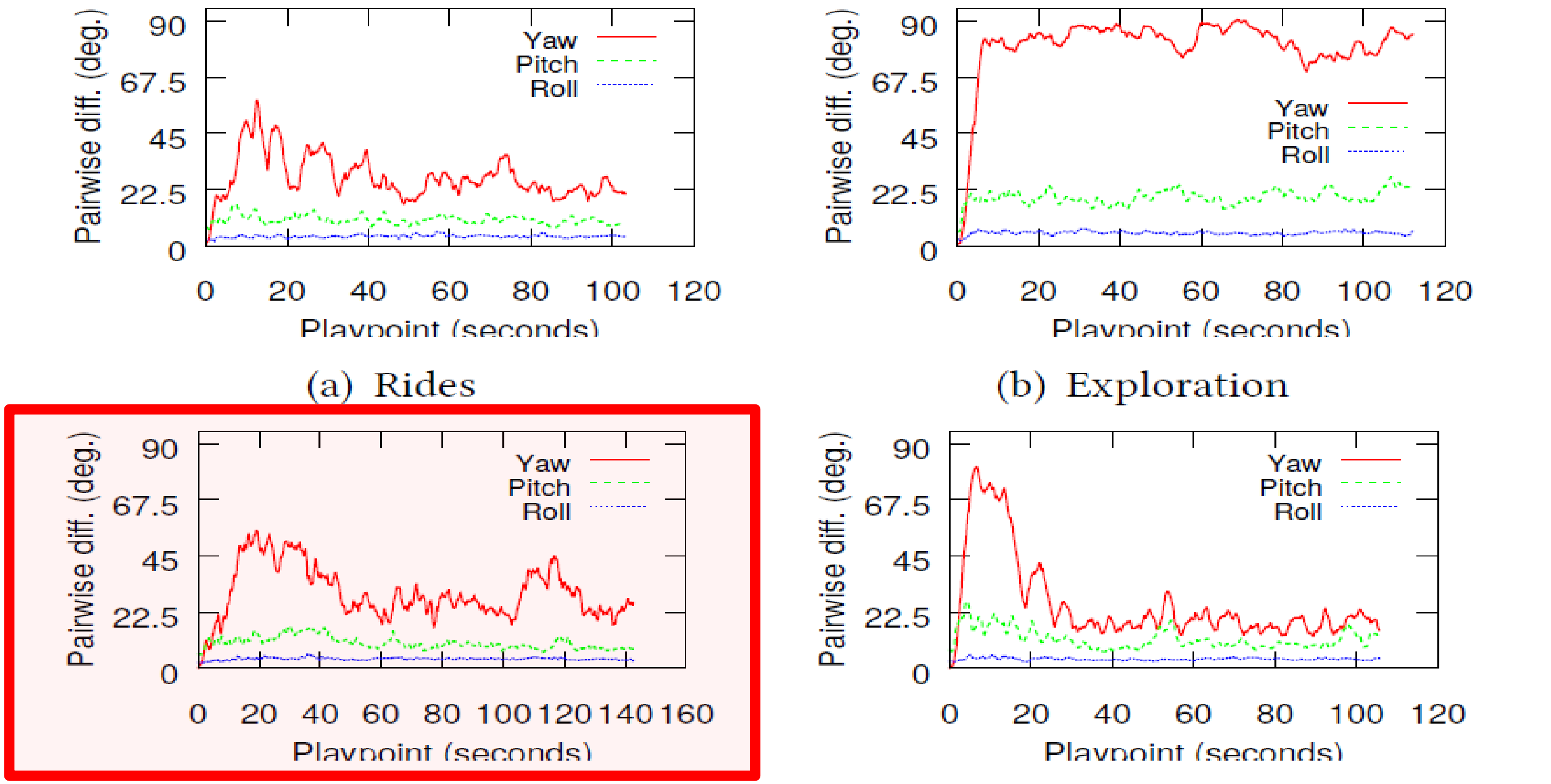
Example users watching a moving focus video



Again, substantial differences between categories

- Exploration close to 90 degrees (fully independent)
- For moving focus, past clients can be good predictor
- Others, especially static focus, has initial exploration phase (or reduced exploration) ... Shows need to look at longer user sessions

Average differences between users watching same video



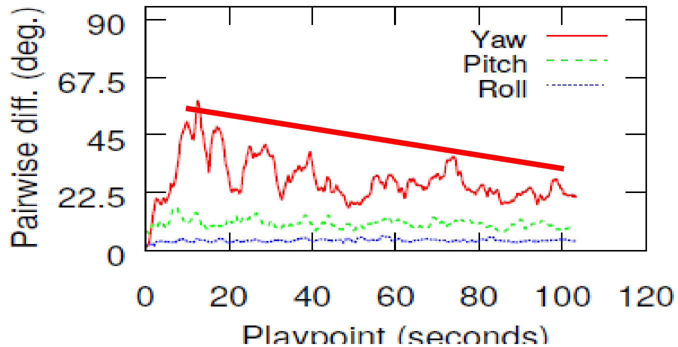
(a) Rides (b) Exploration

(c) Moving focus (d) Static focus

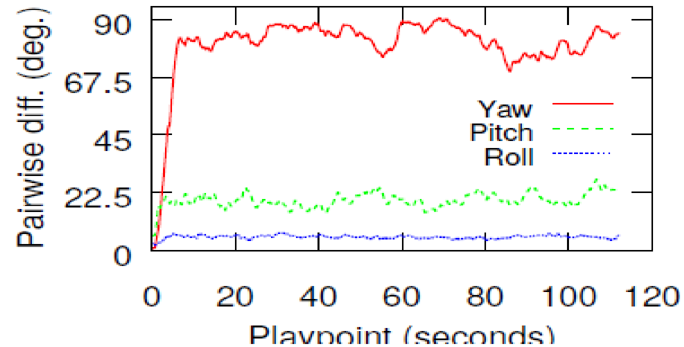
Again, substantial differences between categories

- Exploration close to 90 degrees (fully independent)
- For moving focus, past clients can be good predictor
- Others, especially static focus, has initial exploration phase (or reduced exploration) ... Shows need to look at longer user sessions

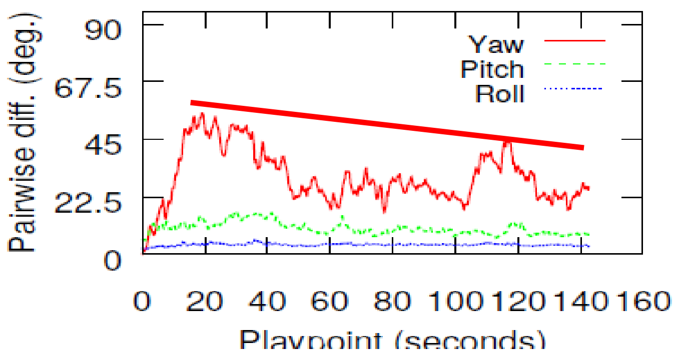
Average differences between users watching same video



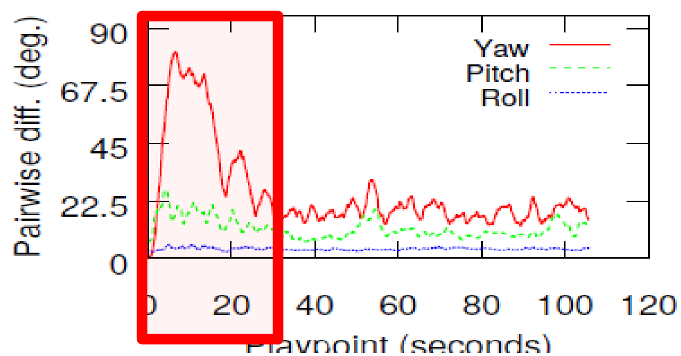
(a) Rides



(b) Exploration



(c) Moving focus



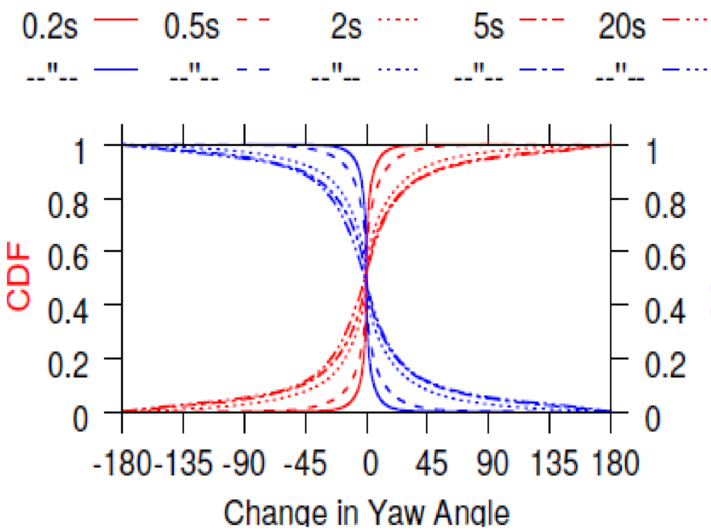
(d) Static focus

Again, substantial differences between categories

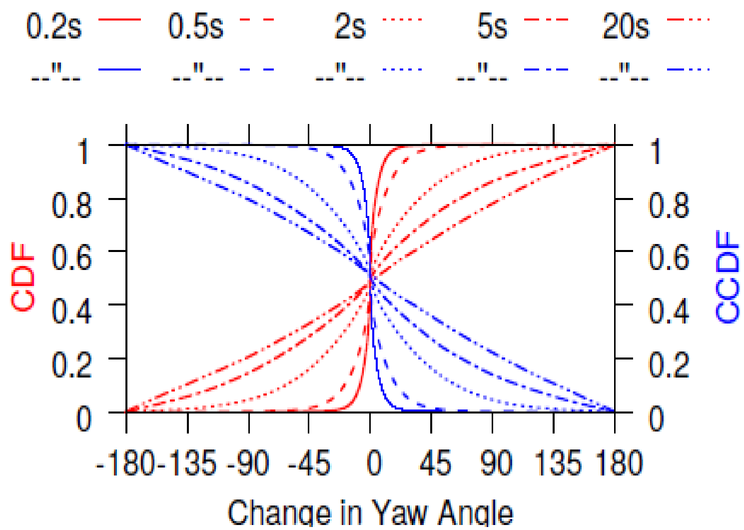
- Exploration close to 90 degrees (fully independent)
- For moving focus, past clients can be good predictor
- Others, especially static focus, has initial exploration phase (or reduced exploration) … Shows need to look at longer user sessions

Change of viewpoint

Change of viewpoint at different time scales



(a) Rides

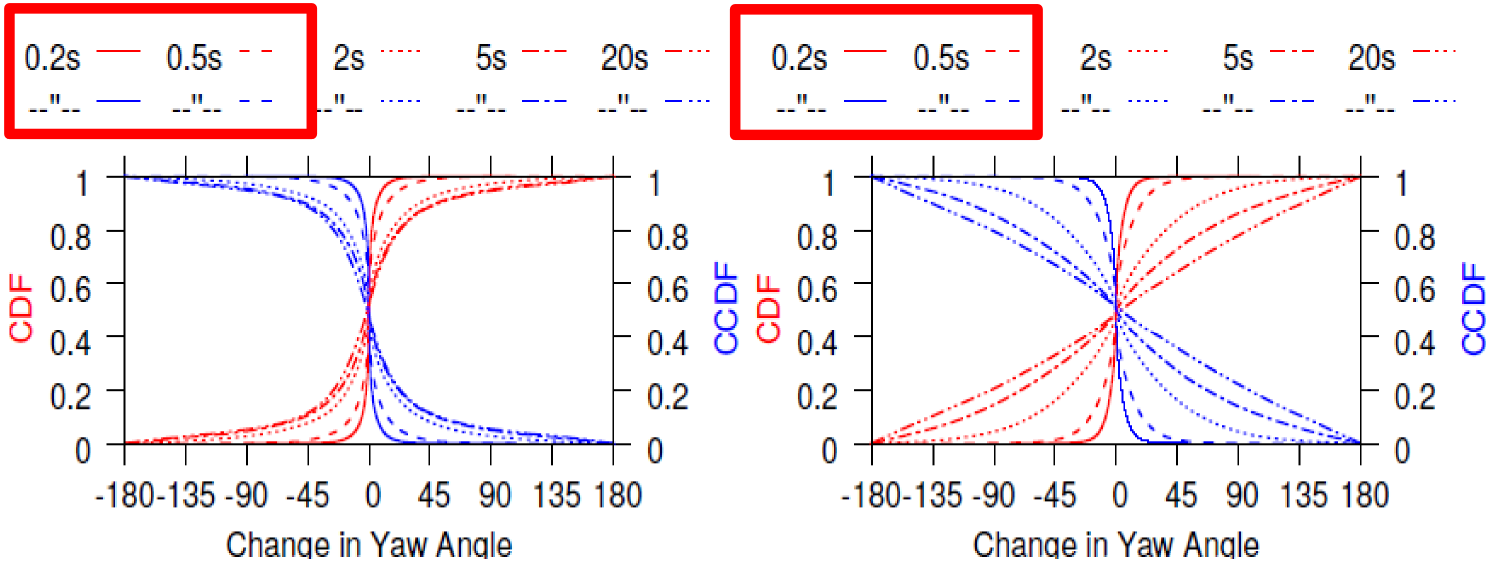


(b) Exploration

Cover range of 360 technologies

- Short time scales (e.g., 0.2-1 second): Low latency scenarios; e.g., edge-based rendering
- Multi-second range (e.g., 2-20 seconds): More applicable to HAS

Change of viewpoint at different time scales

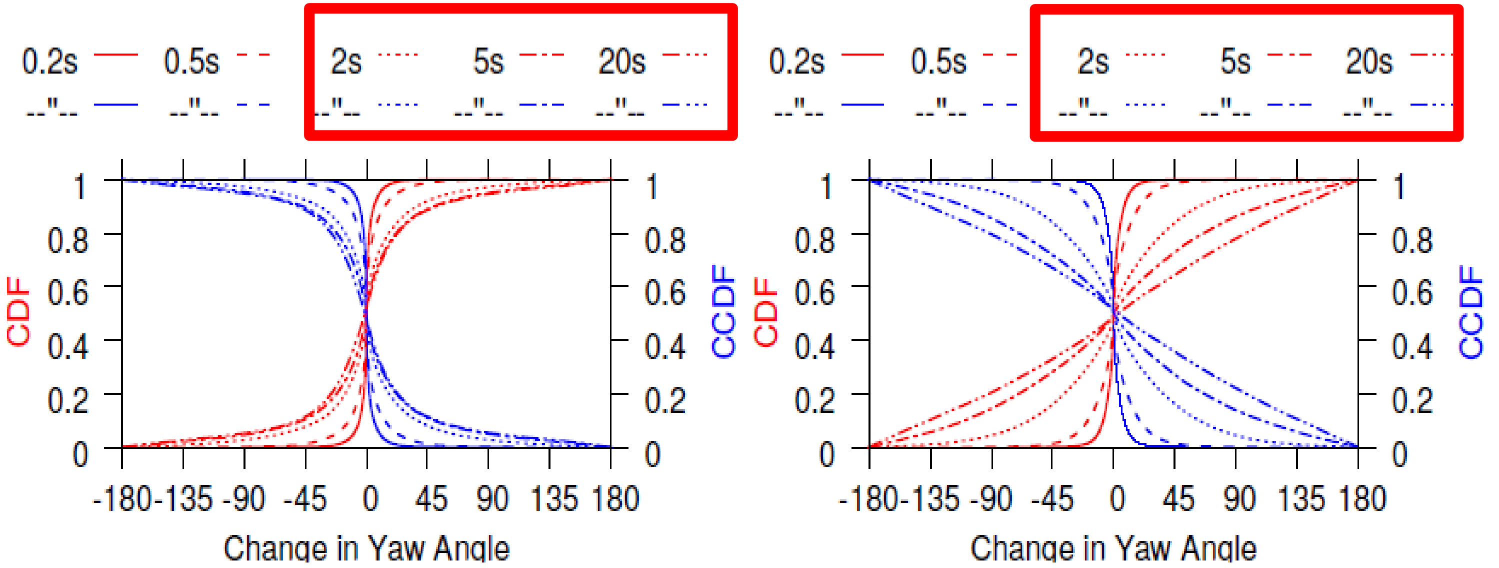


(a) Rides (b) Exploration

Cover range of 360 technologies

- Short time scales (e.g., 0.2-1 second): Low latency scenarios; e.g., edge-based rendering
- Multi-second range (e.g., 2-20 seconds): More applicable to HAS

Change of viewpoint at different time scales

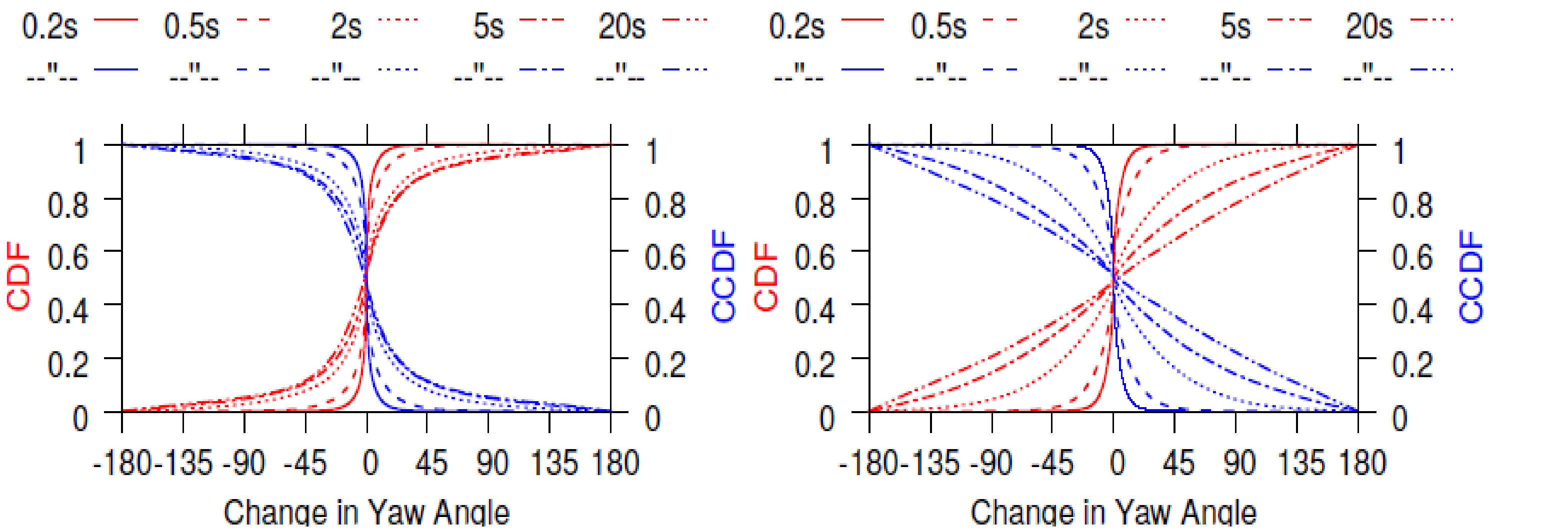


(a) Rides (b) Exploration

Cover range of 360 technologies

- Short time scales (e.g., 0.2-1 second): Low latency scenarios; e.g., edge-based rendering
- Multi-second range (e.g., 2-20 seconds): More applicable to HAS

Change of viewpoint at different time scales



(a) Rides (b) Exploration

Cover range of 360 technologies

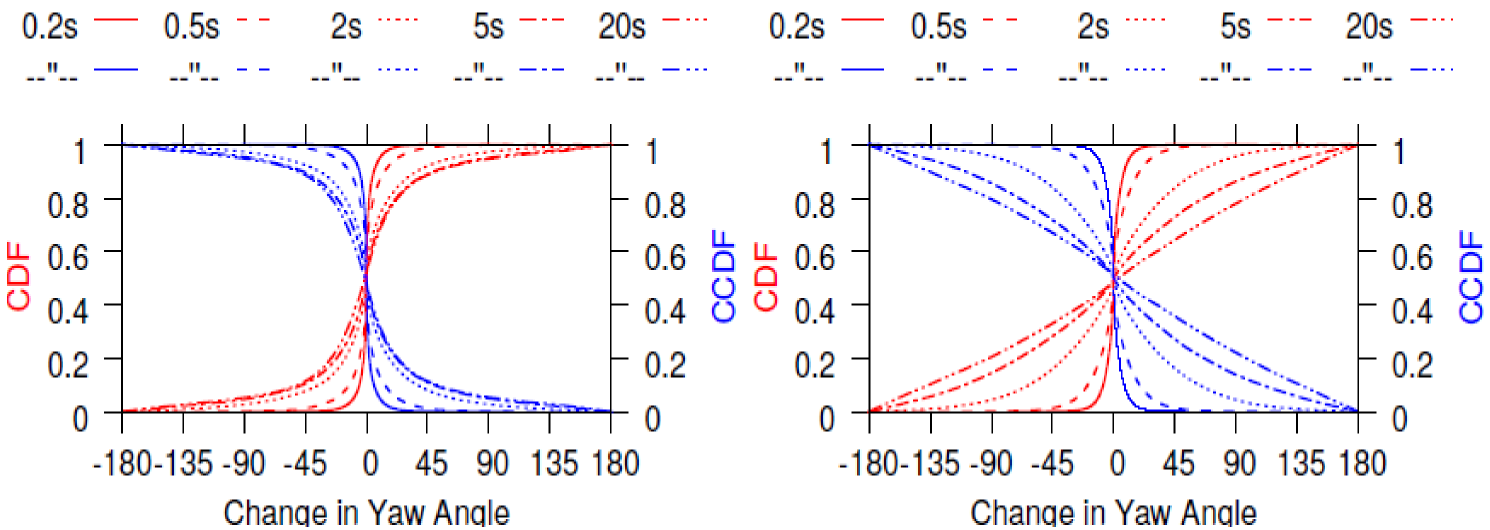
- Short time scales (e.g., 0.2-1 second): Low latency scenarios; e.g., edge-based rendering
- Multi-second range (e.g., 2-20 seconds): More applicable to HAS

Note that prefetching and buffering needed in all cases

- Human tolerance sub-50ms
- Delays caused by modern LTE networks typically are at least 100ms (*)

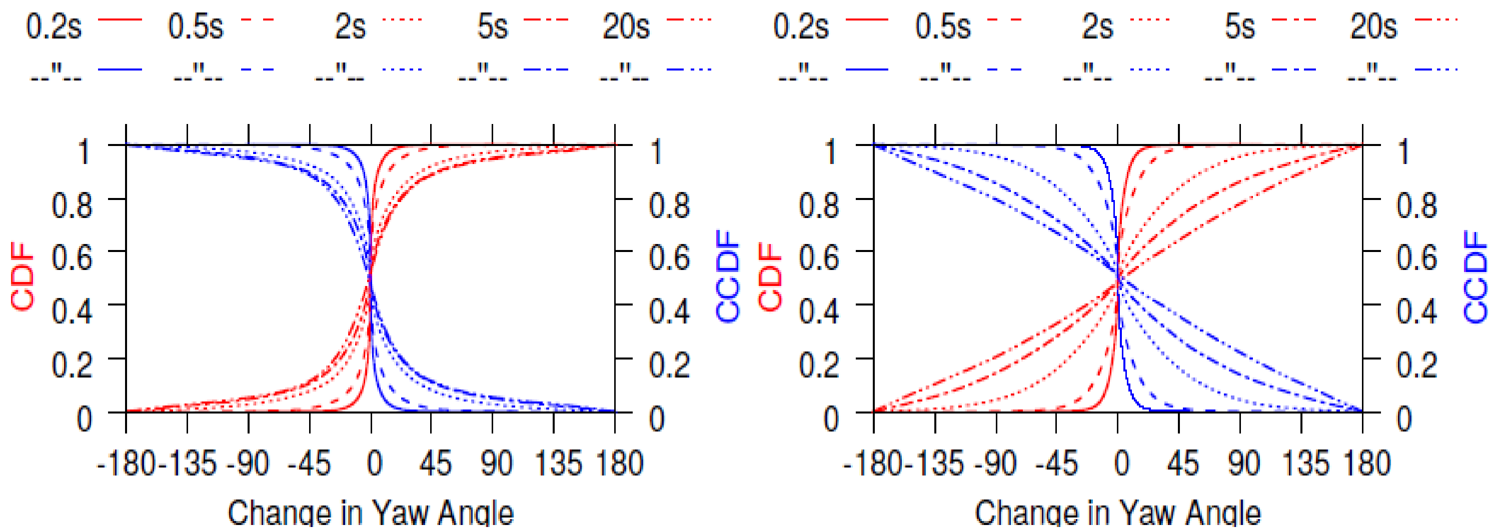
** Tan et al., Enabling Mobile VR in LTE Networks: How Close Are We?. Proc. ACM SIGMETRICS 2018.*

Change of viewpoint at different time scales



(a) Rides (b) Exploration

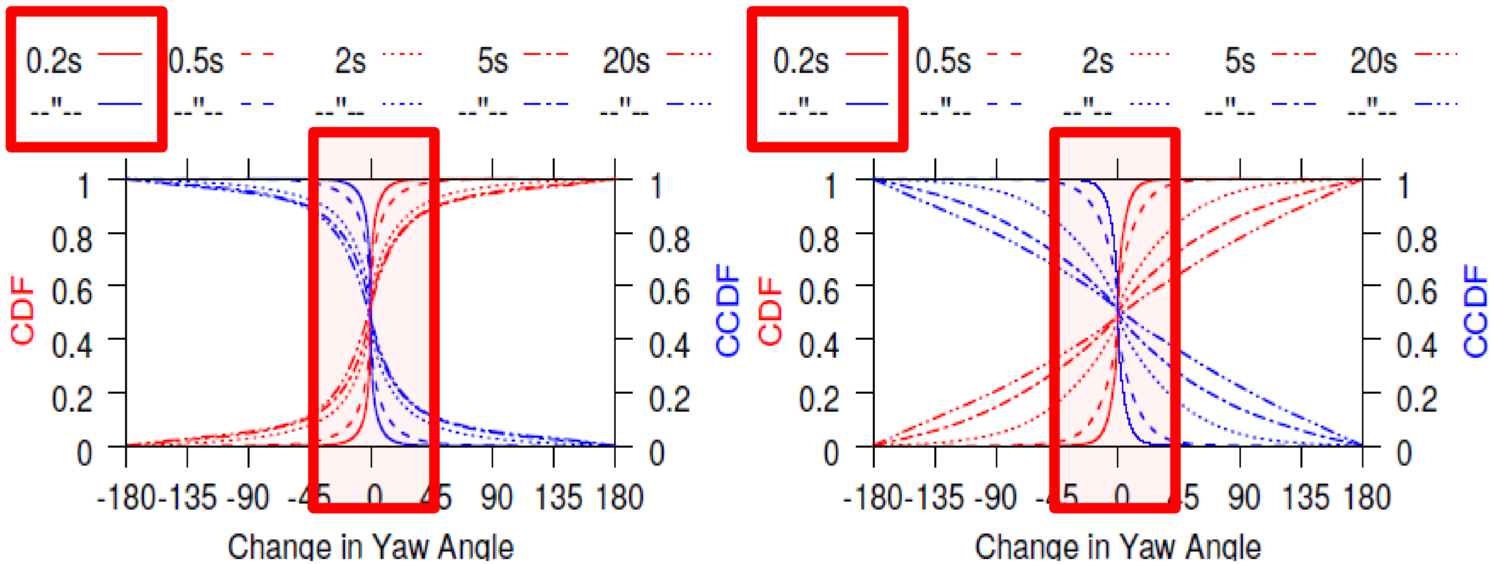
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

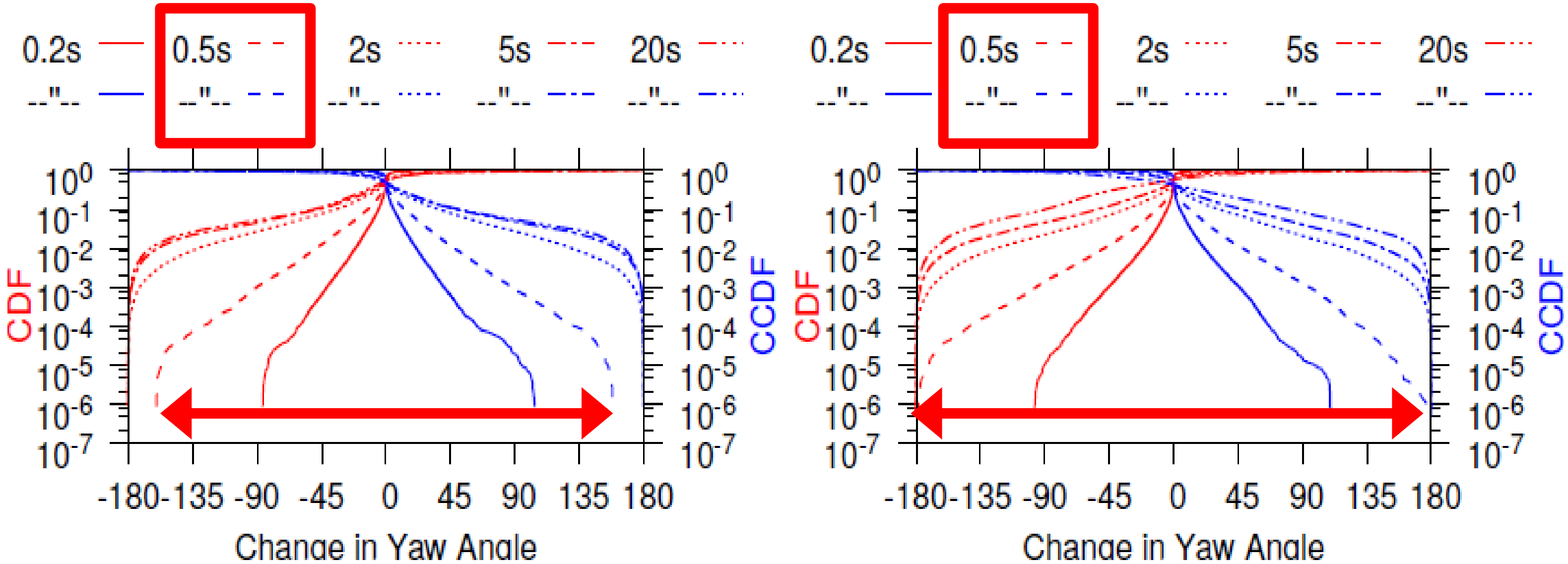
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

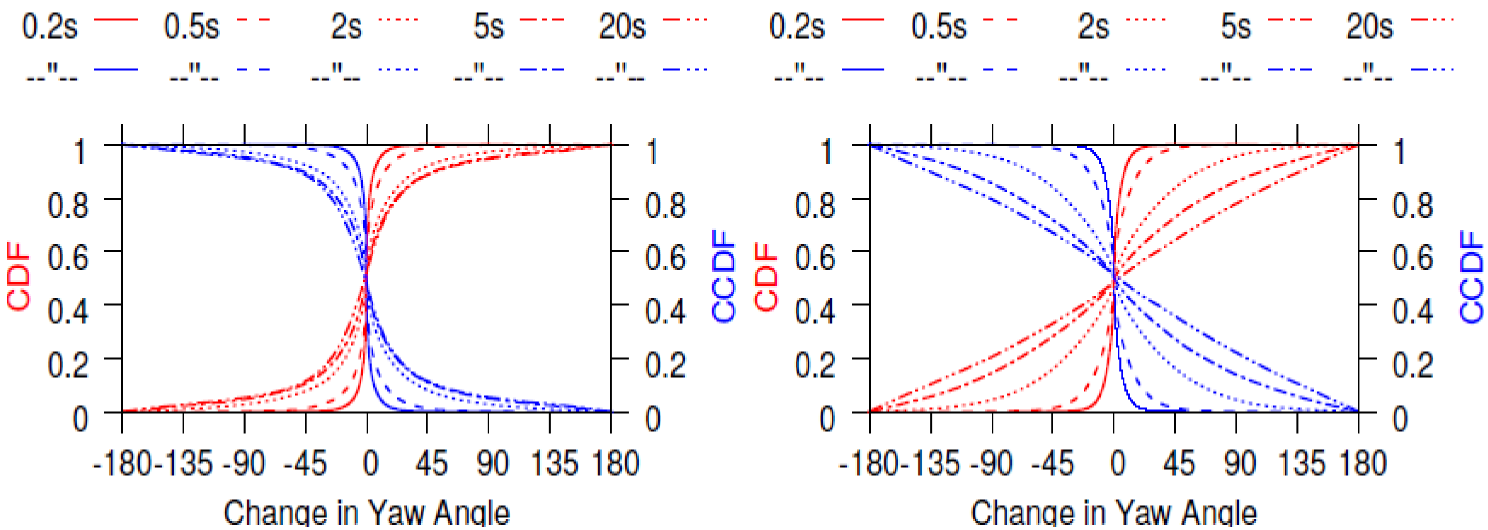
Change of viewpoint at different time scales



(a) Rides (b) Moving focus

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

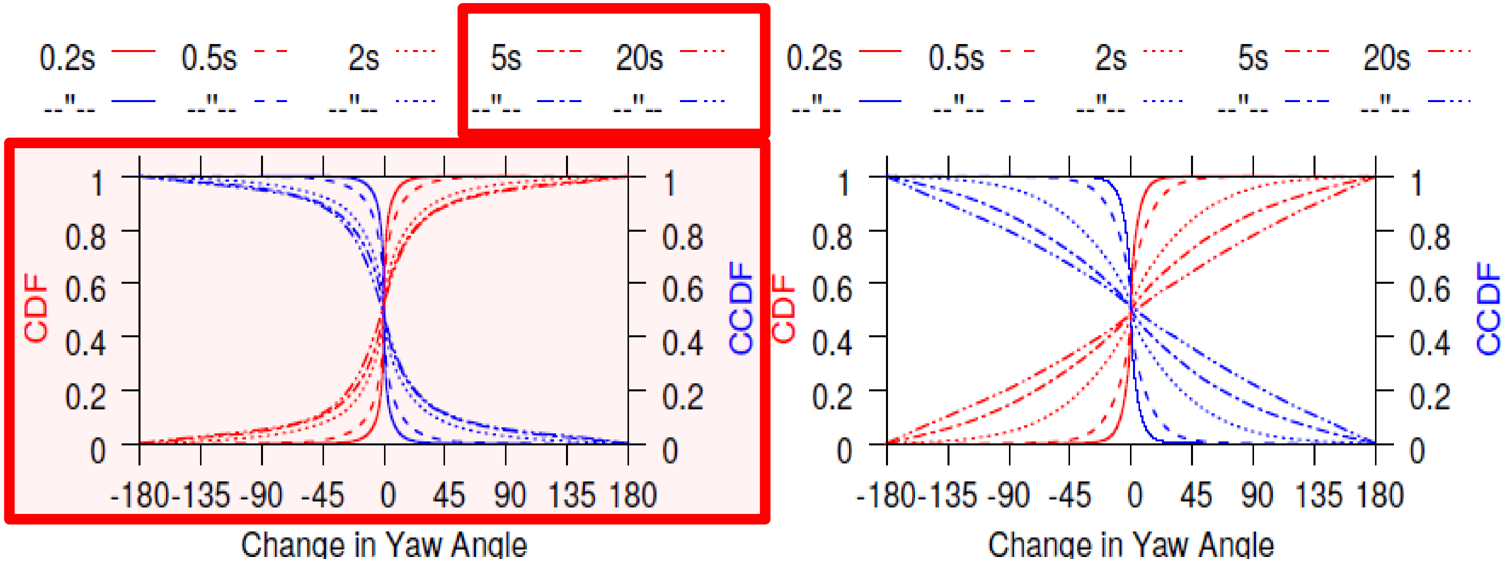
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

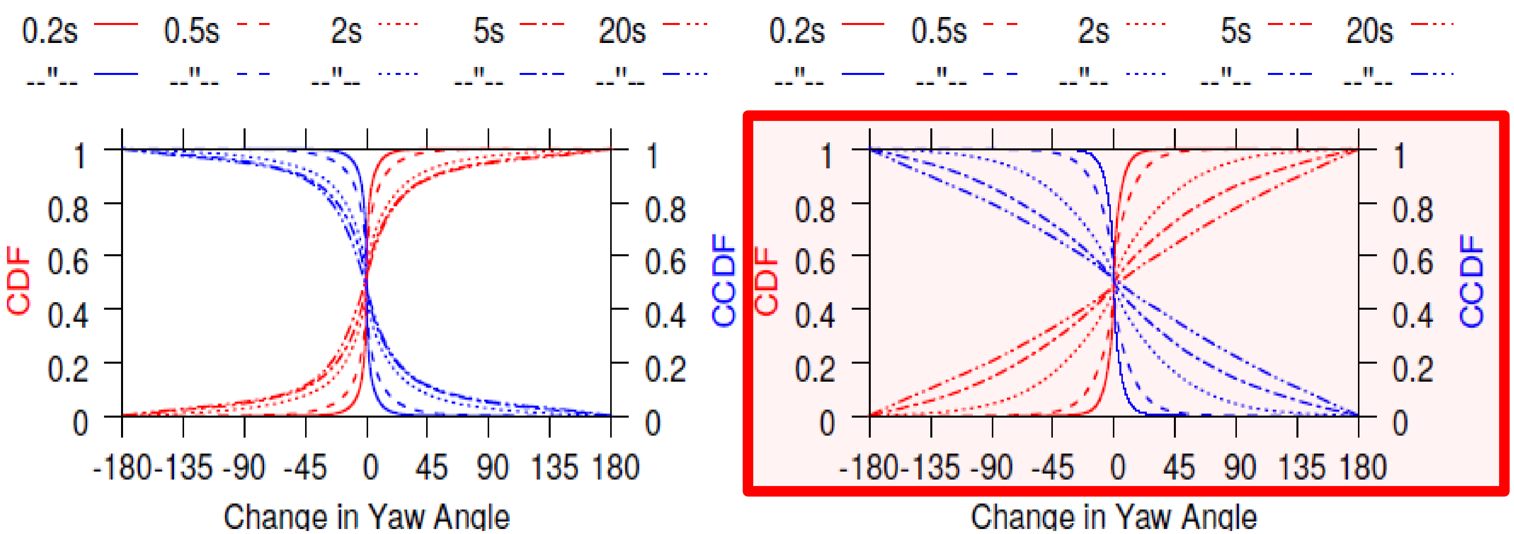
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

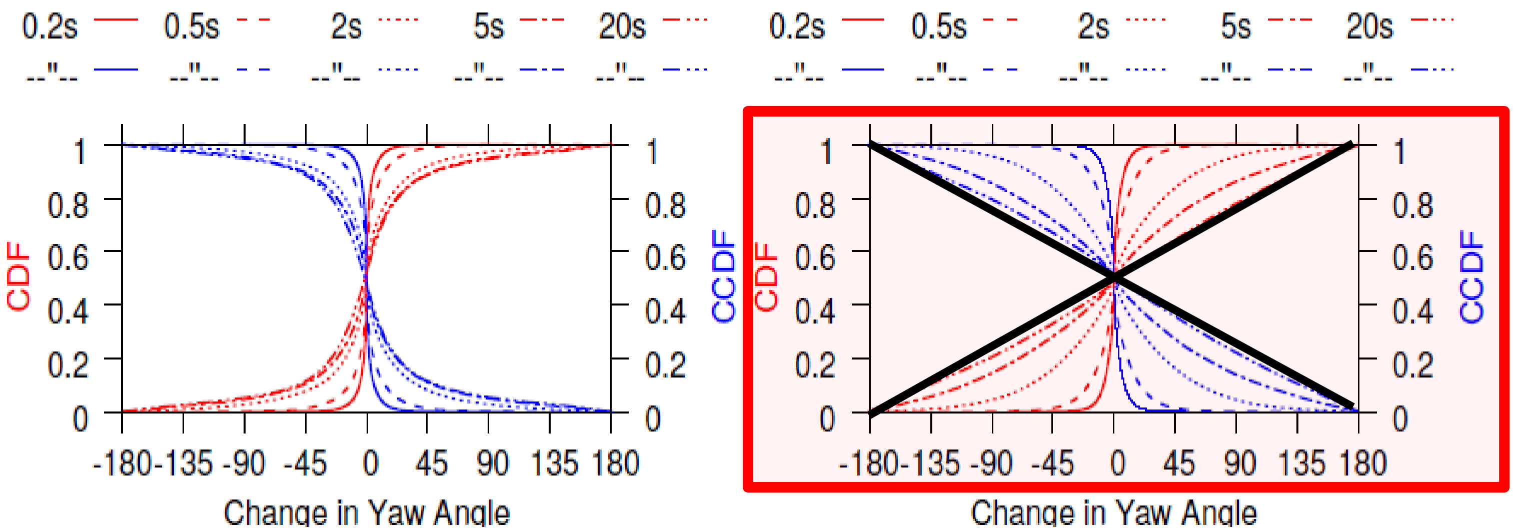
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

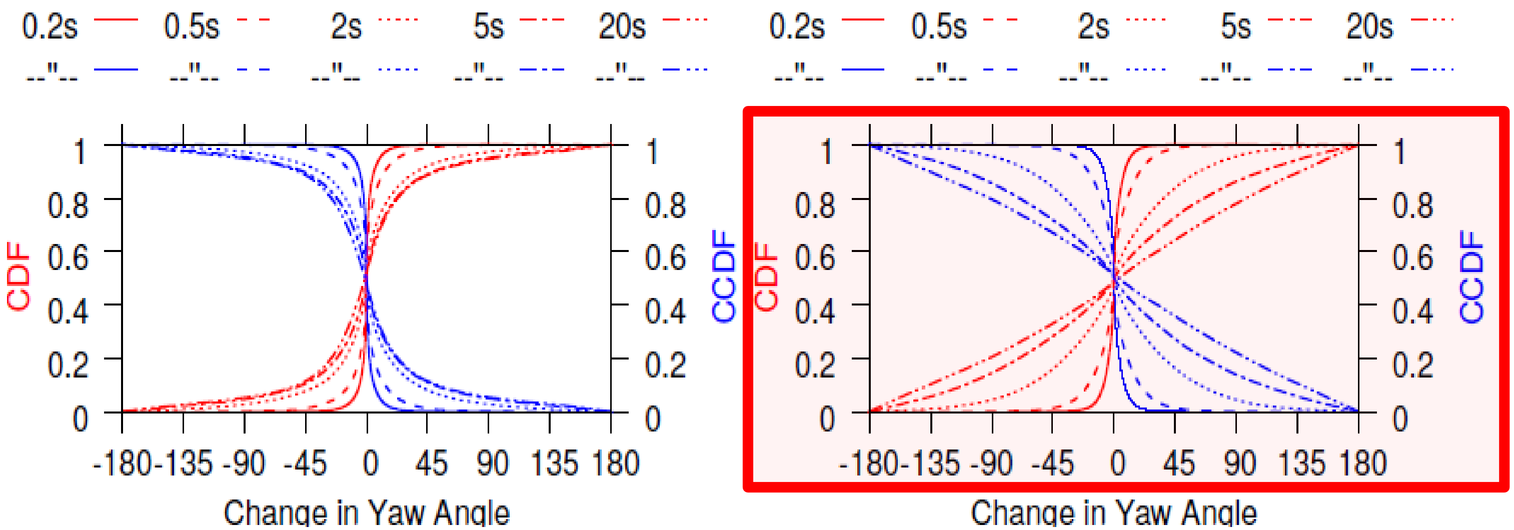
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

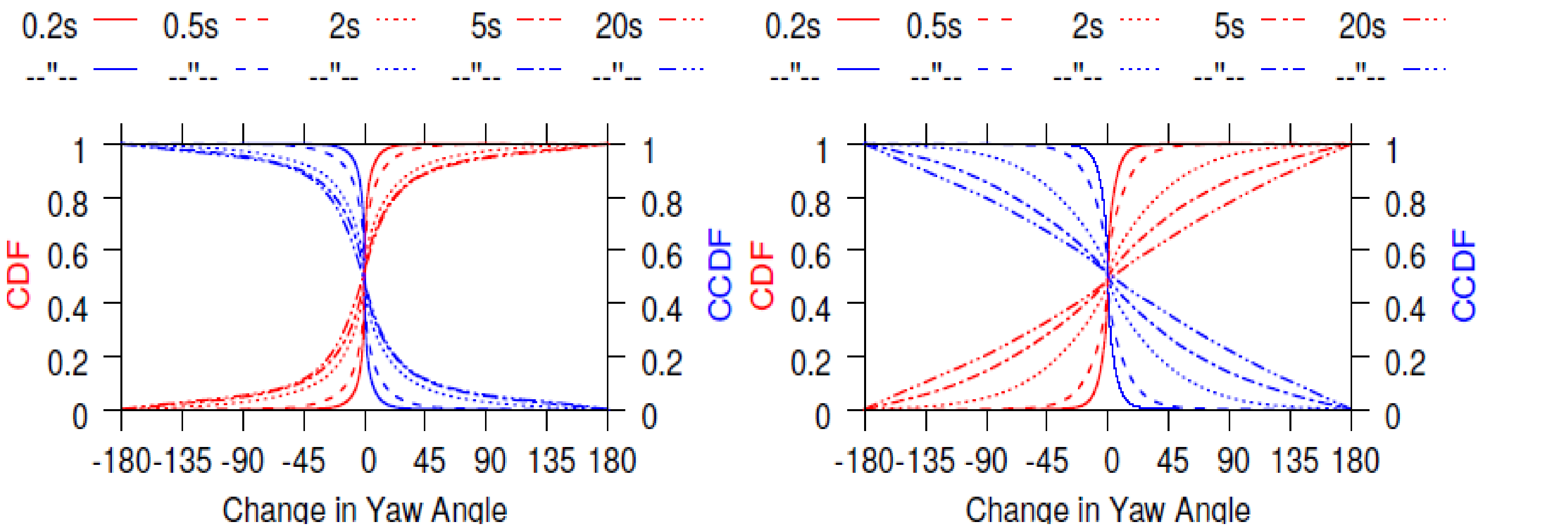
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

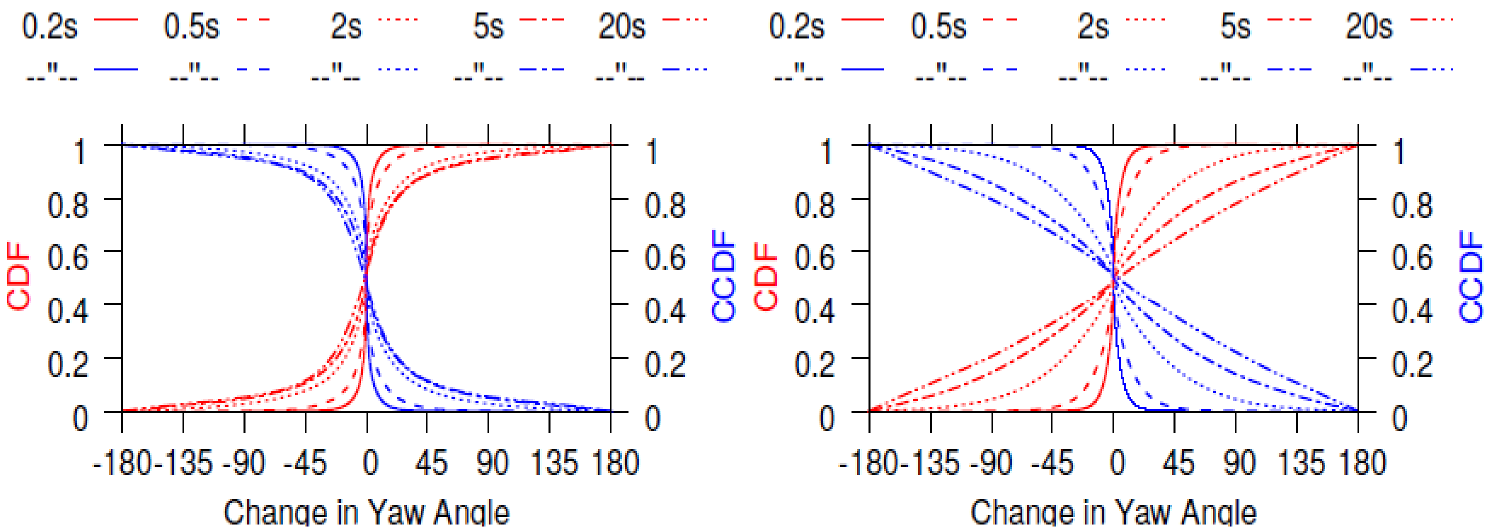
Change of viewpoint at different time scales



(a) Rides (b) Exploration

- @0.2 seconds: Small differences between categories
  - Mostly due to speed of head movements
  - Do not cover full range (could skip data behind users)
- Already at 0.5 seconds rotations cover full range
- Diminishing increase in variations (as T increases)
  - Substantial difference between categories
  - Rides (and static) almost same as lifetime utilization after 5-20 seconds
  - Exploration uniform @ 20 seconds (so zero-degree could be better here)

Change of viewpoint at different time scales



(a) Rides (b) Exploration

We will use these distributions (or the conditional probabilities) when analyzing the best prefetch aggressiveness tradeoff

Optimized prefetching tradeoff

High-level optimization model

Objective function

$$\boxed{E[u|T]} = \sum_{n=0}^{N-1} p_n(T)u(n|q_0, q_1, \ldots, q_{N-1}),$$

Expected utility, conditioned on prefetching *T* ahead
(larger T allows larger buffering)

High-level optimization model

Objective function

$$E[u|T] = \boxed{\sum_{n=0}^{N-1}} p_n(T)u(n|q_0, q_1, \ldots, q_{N-1}),$$

Sum over all viewing directions (granularity of tiles)

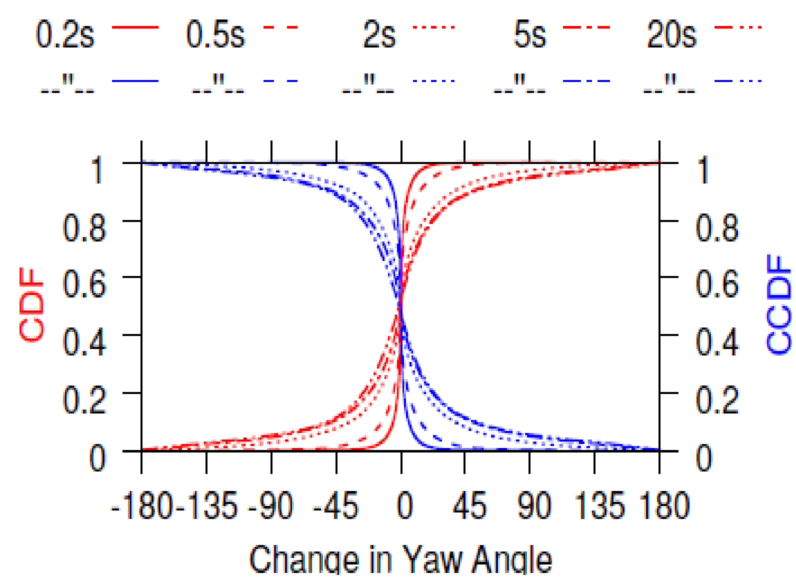
High-level optimization model

Objective function

$$E[u|T] = \sum_{n=0}^{N-1} \boxed{p_n(T)}\, u(n|q_0, q_1, \ldots, q_{N-1}),$$

Conditional probability

$$p_n(T) = \int_{\theta_n}^{\theta_{n+1}} p(\theta|T)\mathrm{d}\theta,$$



High-level optimization model

Objective function

$$E[u|T] = \sum_{n=0}^{N-1} p_n(T)\boxed{u(n|q_0, q_1, \ldots, q_{N-1}),}$$

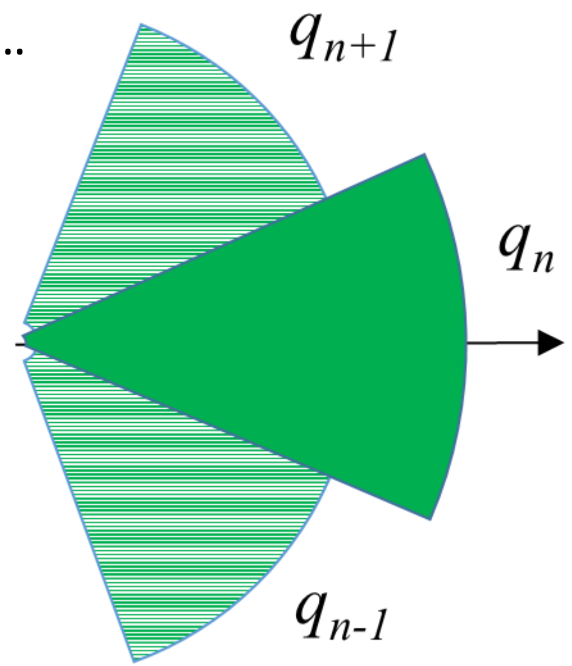
Utility for that direction

High-level optimization model

Objective function

$$E[u|T] = \sum_{n=0}^{N-1} p_n(T) \boxed{u(n|q_0, q_1, \ldots, q_{N-1})},$$

Utility depends on quality of neighboring tiles …



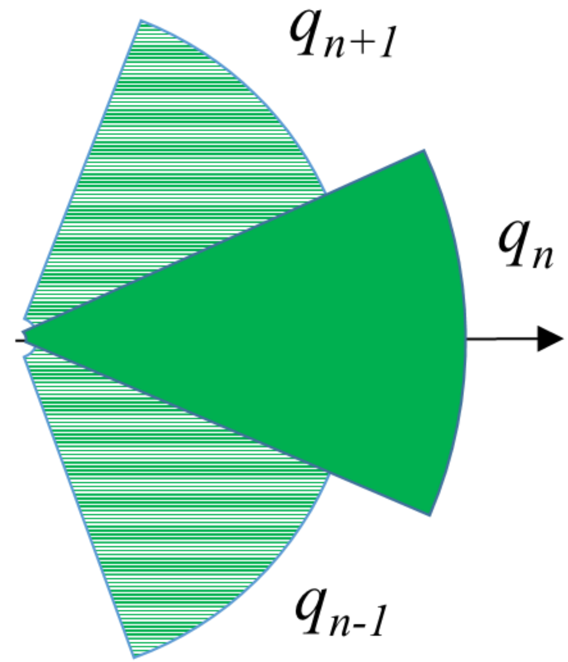
High-level optimization model

The utility $u(n|q_0, q_1, \ldots, q_{N-1})$ combines

$$(1-\alpha-\beta)u(q_n)$$

$$\frac{\alpha}{2}\left(\frac{p_{n-1}}{p_n}u(q_{n-1}) + \frac{p_{n+1}}{p_n}u(q_{n+1})\right)$$

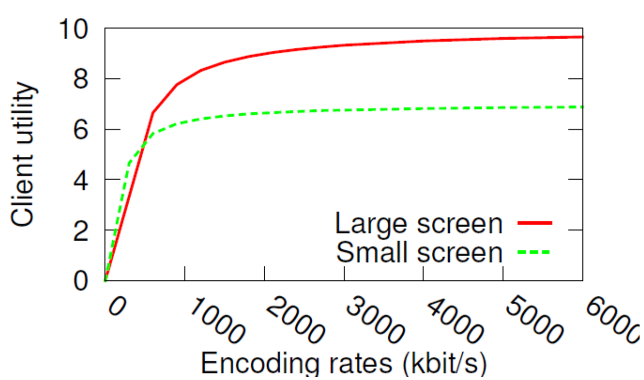
$$-\frac{\beta}{2}\left(|u(q_n)-u(q_{n-1})| + |u(q_n)-u(q_{n+1})|\right)$$

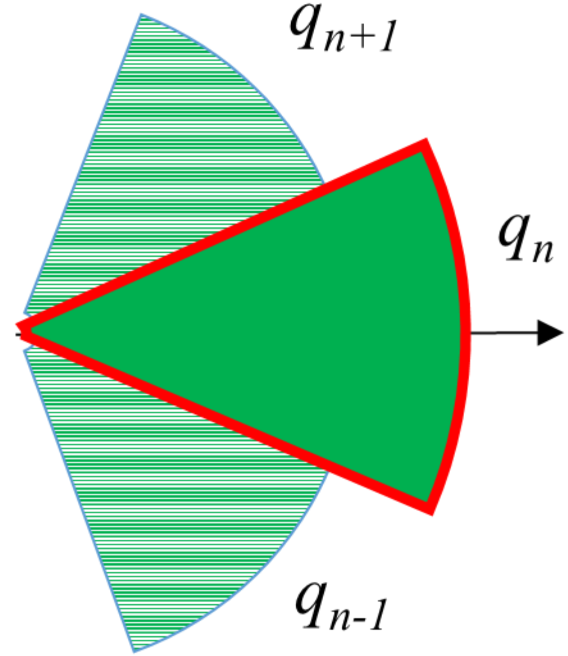


High-level optimization model

The utility $u(n|q_0, q_1, \ldots, q_{N-1})$ combines

$$\boxed{(1 - \alpha - \beta)u(q_n)}$$





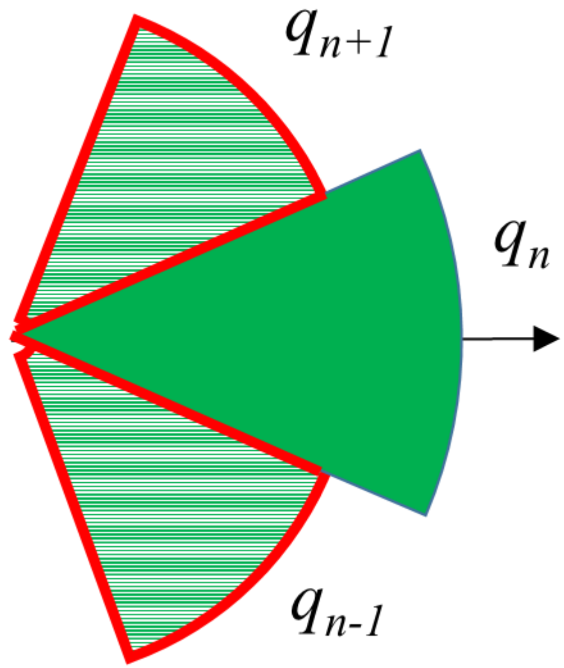
High-level optimization model

The utility $u(n|q_0, q_1, \ldots, q_{N-1})$ combines

$$(1 - \alpha - \beta)u(q_n)$$

$$\boxed{\frac{\alpha}{2}\left(\frac{p_{n-1}}{p_n}u(q_{n-1}) + \frac{p_{n+1}}{p_n}u(q_{n+1})\right)}$$

$$-\frac{\beta}{2}\left(|u(q_n) - u(q_{n-1})| + |u(q_n) - u(q_{n+1})|\right)$$



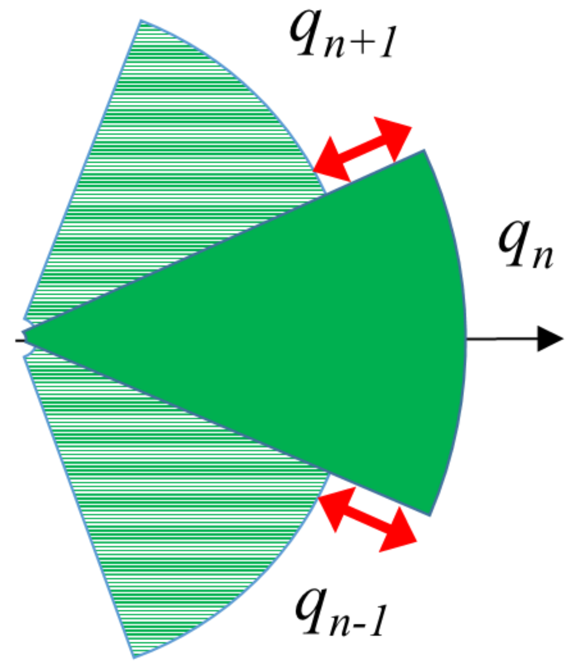
High-level optimization model

The utility $u(n|q_0, q_1, \ldots, q_{N-1})$ combines

$$(1-\alpha-\beta)u(q_n)$$

$$\frac{\alpha}{2}\left(\frac{p_{n-1}}{p_n}u(q_{n-1}) + \frac{p_{n+1}}{p_n}u(q_{n+1})\right)$$

$$\boxed{-\frac{\beta}{2}\left(|u(q_n)-u(q_{n-1})| + |u(q_n)-u(q_{n+1})|\right)}$$



Detailed optimization model

Optimization problem

$$\text{maximize } E[u|T], \tag{4}$$

where

$$E[u|T] = (1-\beta)\left(\sum_{n=0}^{N-1} p_n(T) \sum_{l=0}^{L} x_{n,l} u_{n,l}\right) - \beta\left(\sum_{n=0}^{N-1} \frac{p_n(T) + p_{n+1}(T)}{2} \times \sum_{l=0}^{L} \sum_{l'=0}^{L} x_{n,l} x_{n+1,l'} \left| u_{n,l} - u_{n+1,l'} \right| \right), \tag{5}$$

such that

$$\sum_{l=0}^{L} x_{n,l} = 1, \qquad 0 \le n < N, \tag{6}$$

$$\sum_{n=0}^{N-1} \sum_{l=0}^{L} x_{n,l} b_{n,l} \le D\Delta, \tag{7}$$

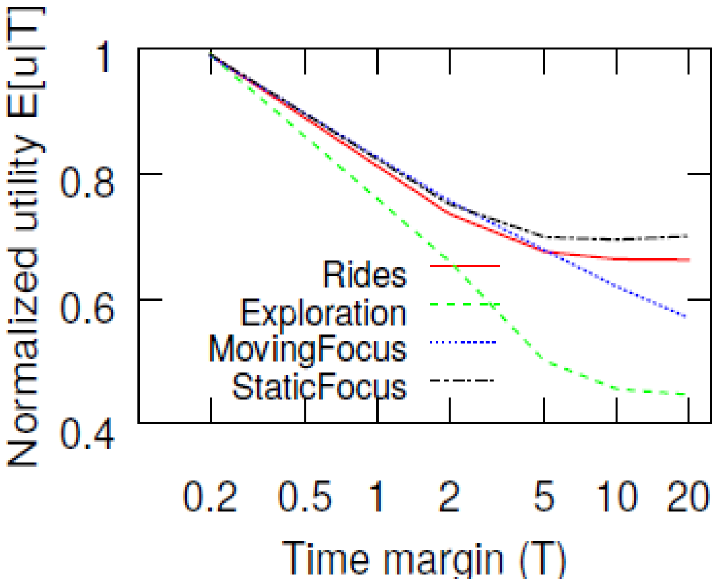
$$x_{n,l} \in \{0,1\}, \quad 0 \le n < N, 0 \le l \le L. \tag{8}$$

Problem formulation

NP hard (special case reduces to 0-1 knapsack, which is NP complete)

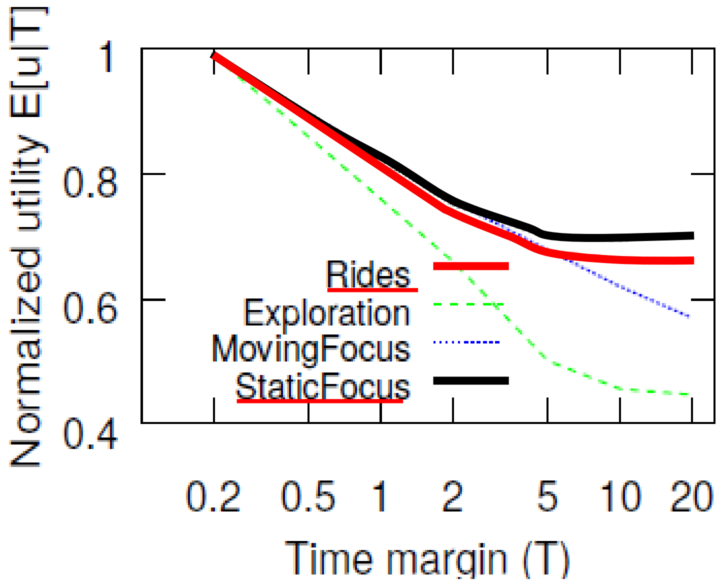
Solve problem using dynamic programming (DP) ...

Example tradeoffs: Different categories



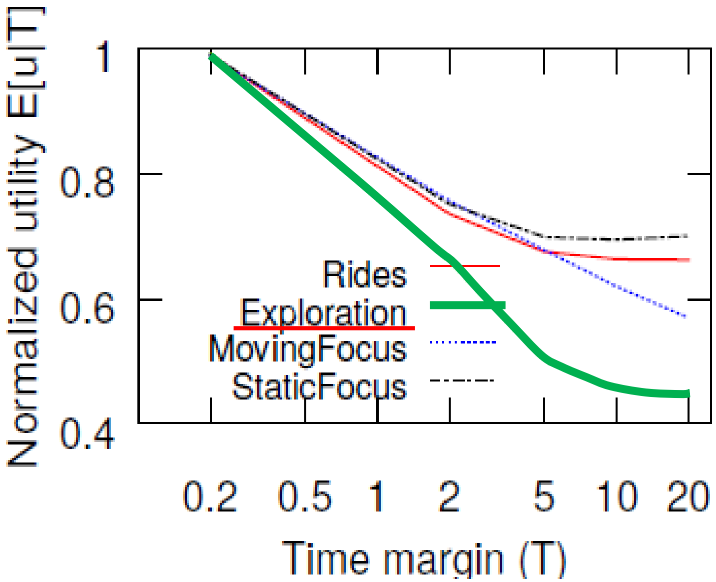
- Static and rides has the best tradeoff curves
- Exploration has the worst tradeoff curve
- Three categories flattens out after 5-10 seconds
  - Opportunity to use larger buffers (e.g., 20s rather than 5s)

Example tradeoffs: Different categories



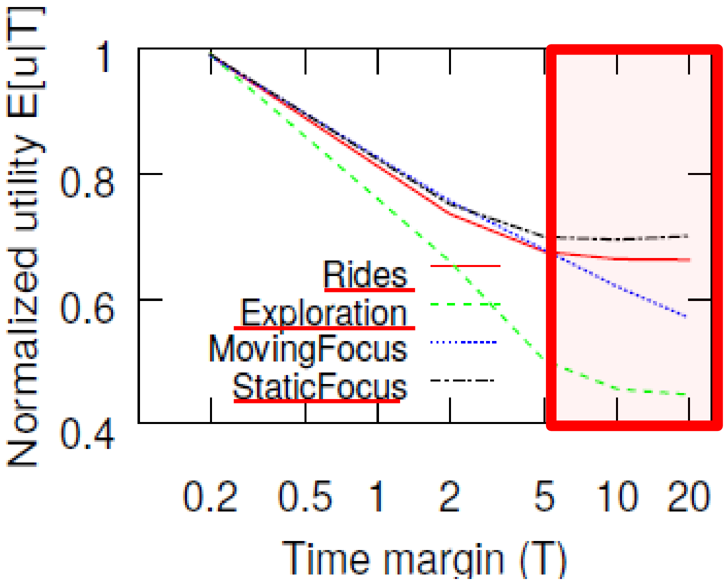
- Static and rides has the best tradeoff curves
- Exploration has the worst tradeoff curve
- Three categories flattens out after 5-10 seconds
  - Opportunity to use larger buffers (e.g., 20s rather than 5s)

Example tradeoffs: Different categories



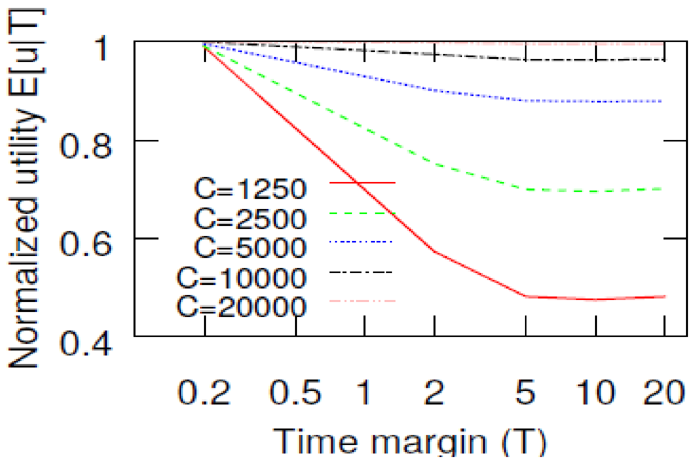
- Static and rides has the best tradeoff curves
- Exploration has the worst tradeoff curve
- Three categories flattens out after 5-10 seconds
  - Opportunity to use larger buffers (e.g., 20s rather than 5s)

Example tradeoffs: Different categories



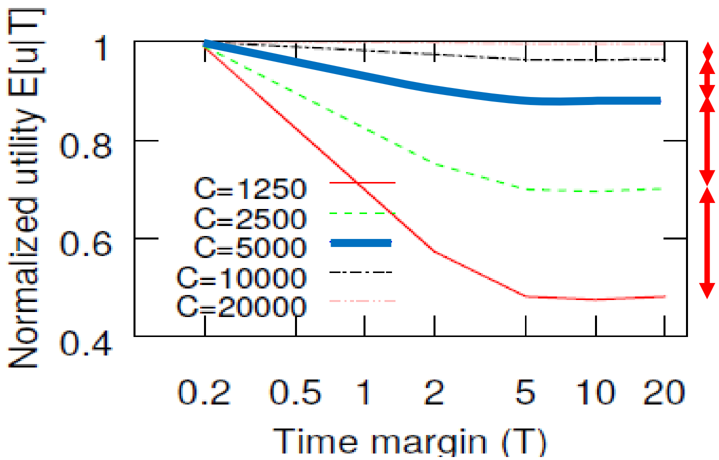
- Static and rides has the best tradeoff curves
- Exploration has the worst tradeoff curve
- Three categories flattens out after 5-10 seconds
  - Opportunity to use larger buffers (e.g., 20s rather than 5s)

Example tradeoffs: Impact of X …



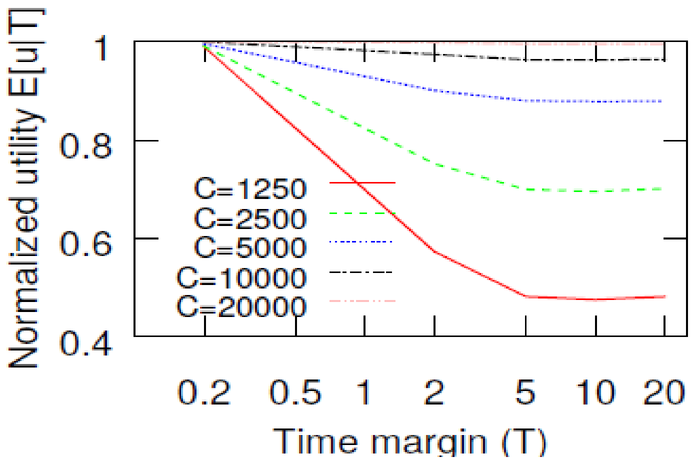
- Diminishing prefetch capacity returns
  - E.g., 0.837 with C=5000 (less than 20% of max 25,188) and *T* = 20s

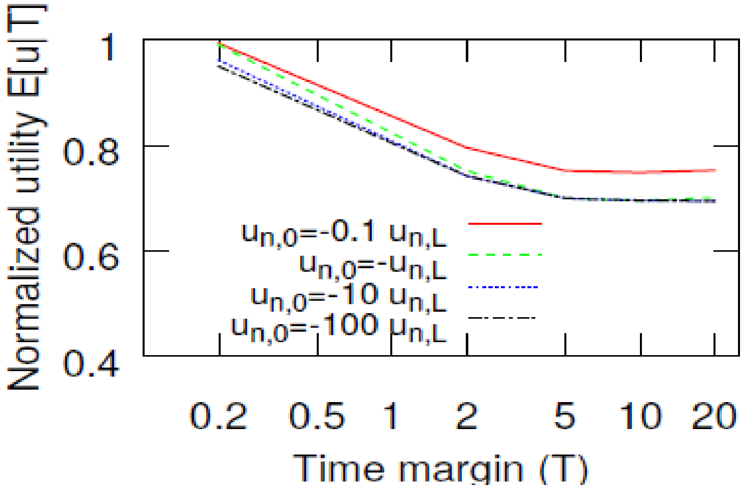
Example tradeoffs: Impact of X …



- Diminishing prefetch capacity returns
  - E.g., 0.837 with C=5000 (less than 20% of max 25,188) and *T* = 20s

Example tradeoffs: Impact of X …

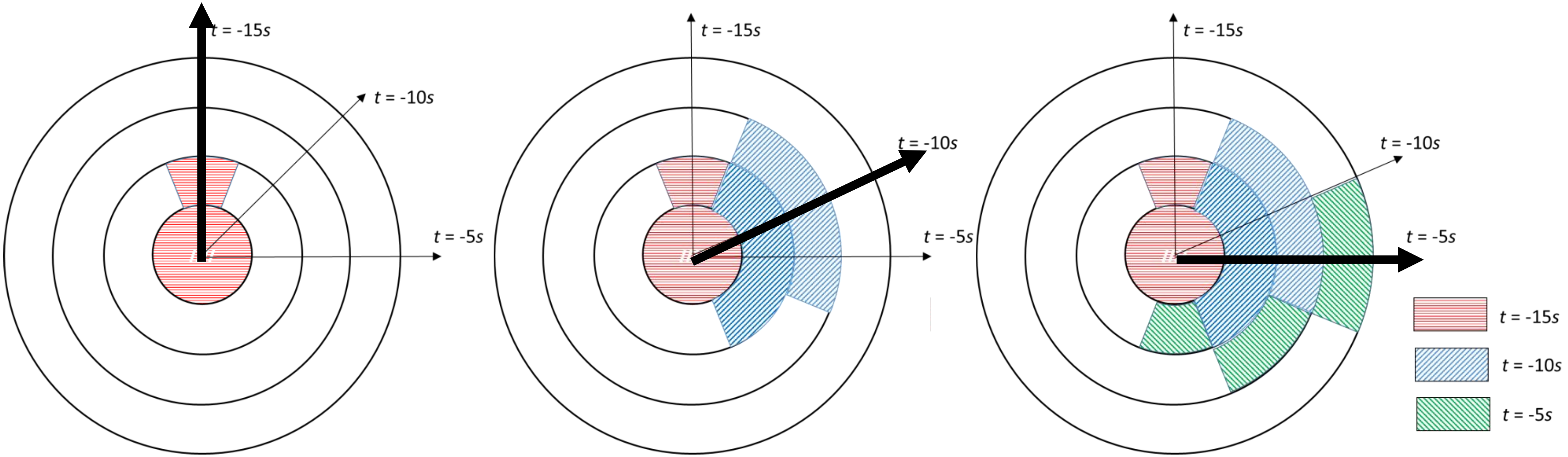




- Diminishing prefetch capacity returns
  - E.g., 0.837 with C=5000 (less than 25% of max 25,188) and *T* = 20s
- Limited impact of stall penalty
  - Only when very limited b/w does stalls play a factor
  - Above result with C=2500 (less than 10% of max)
  - With C=5000 (less than 20% of max) no impact …

Some additional design optimizations

Personalized layers



Initial “layer” Second “layer” Third “layer”

- Key idea is to combine
  - Long term prefetching to protect against bandwidth variations
  - Fine-grained optimized prefetching based on viewing direction (closer to deadline)
- Approach here
  - Personalized “layers” based on view direction and downloaded tiles
  - Example figure above assumes SVC-based tiles and 3 modules

Conclusions

Conclusions

Data-driven characterization of prefetching aggressiveness tradeoff

- Significant differences between four different categories of 360 video: static focus, moving focus, rides, and exploration

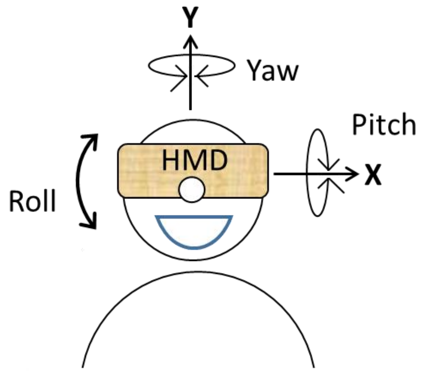
Optimized buffer-quality tradeoffs

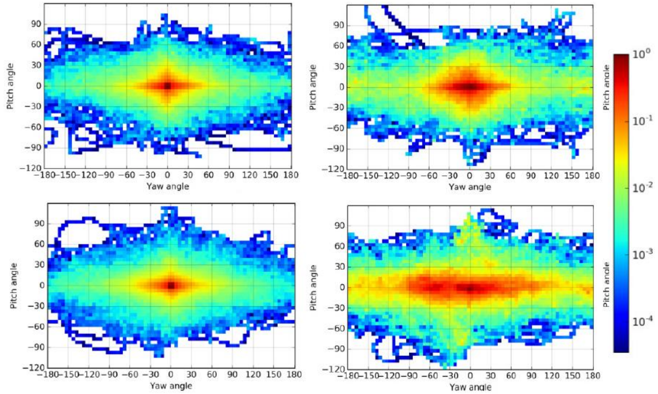
- Optimization framework that captures tradeoff between the goals of prefetching far ahead and the best quality selection for each viewing direction

Data-driven discussion of further design optimizations

- E.g., personalized “layers” that combine long-term prefetching (i.e., larger buffers) and fine-grained prefetching enhancements closer to playout time
- Paper also consider head movement speed, directional changes, and the time within the video (e.g., account of initial exploration) ...

Thanks for listening!





(a) T=2s (b) T=20s (c)

The Prefetch Aggressiveness Tradeoff in 360 Video Streaming

Mathias Almquist, Viktor Almquist, Vengatanathan Krishnamoorthi,
Niklas Carlsson, and Derek Eager

LINKÖPING UNIVERSITY

Niklas Carlsson (niklas.carlsson@liu.se)