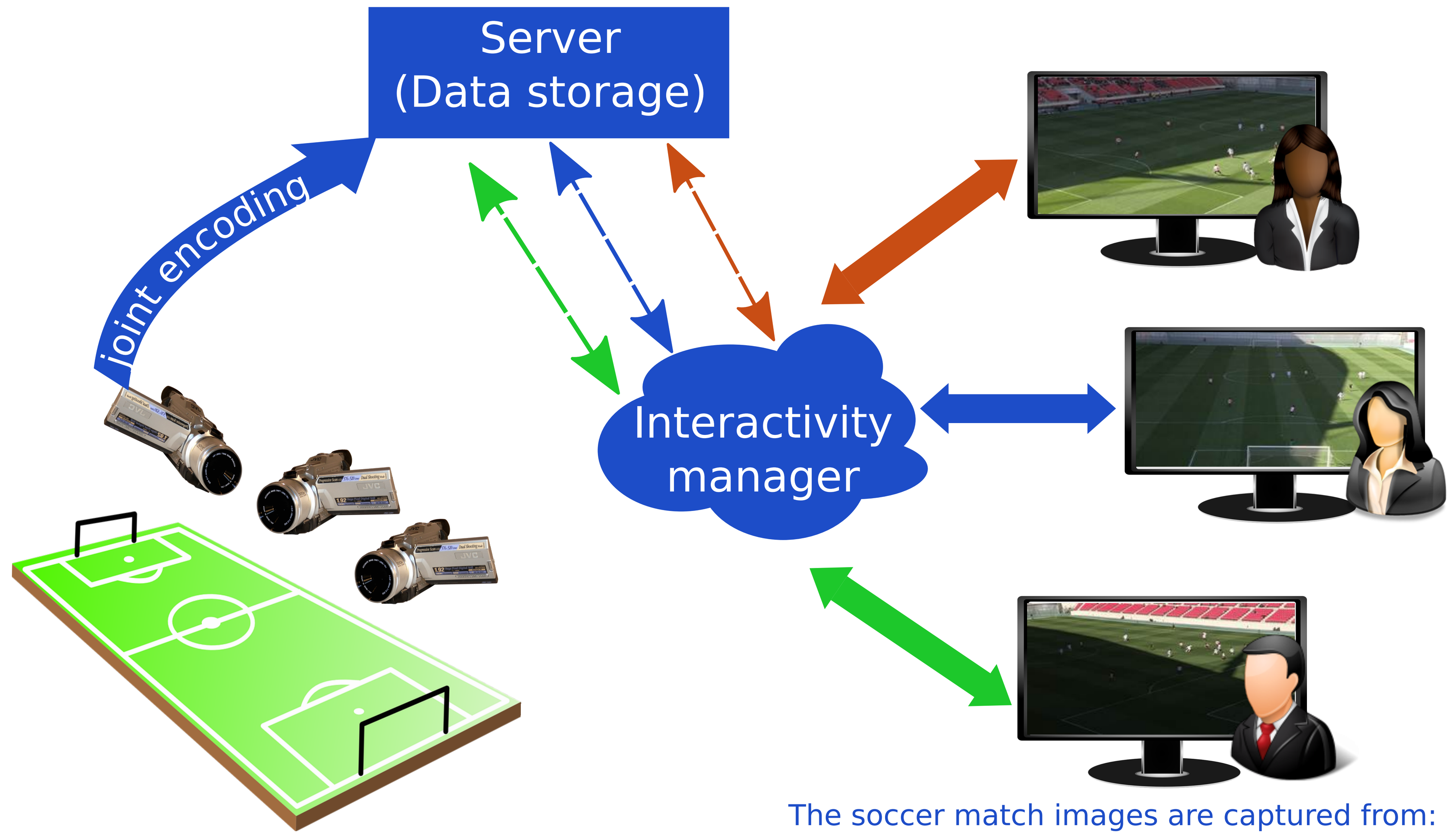
Correlation Model Selection for Interactive video Communication

Navid MAHMOUDIAN BIDGOLI, Thomas MAUGEY, Aline ROUMY
INRIA Rennes Bretagne-Atlantique

Introduction

Interactivity: user can switch within views to see the scene from different viewpoints.

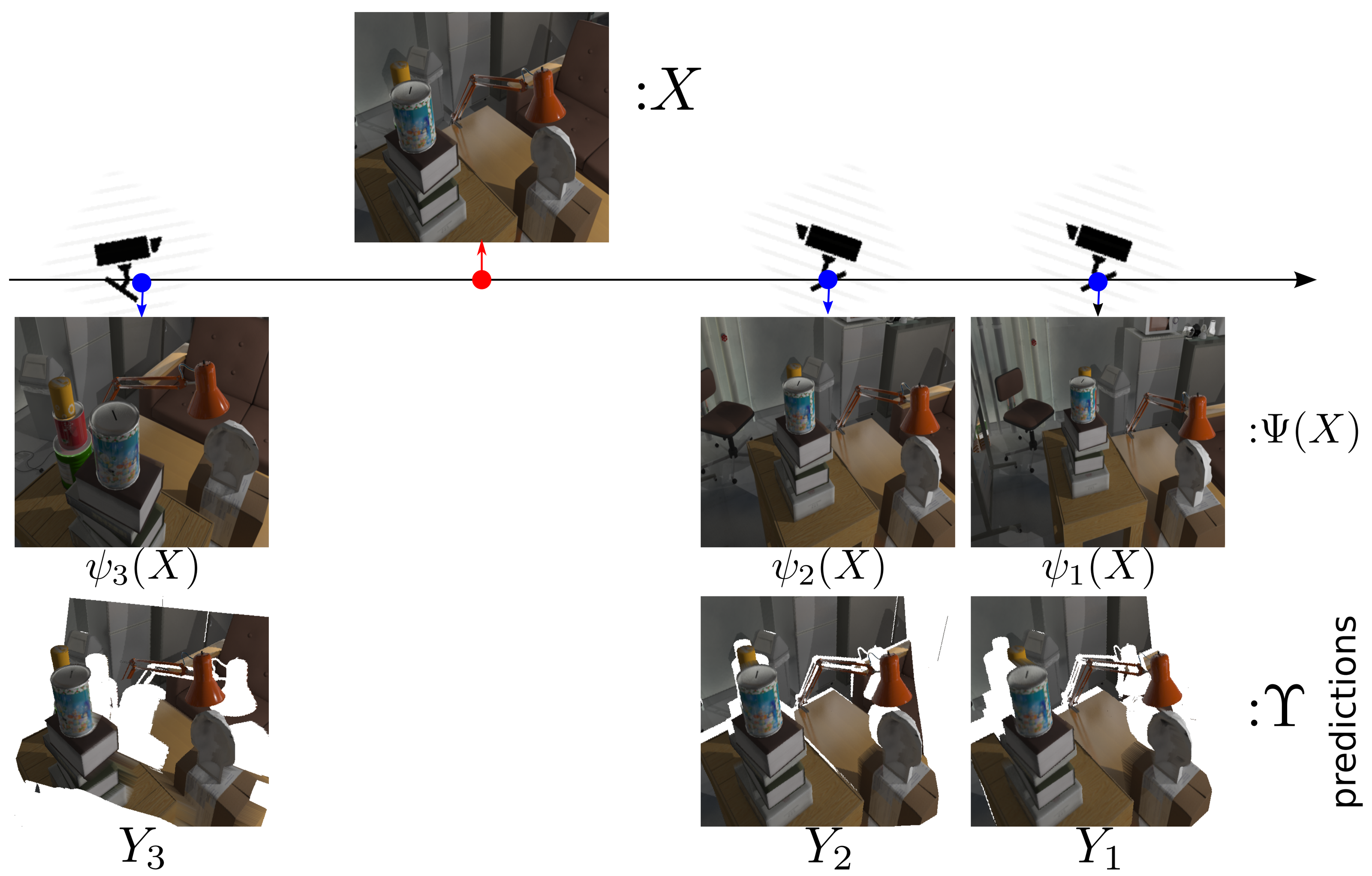


The soccer match images are captured from: www.fujii.nuee.nagoya-u.ac.jp/multiview-data

compress All videos
- Storing the videos on the server as compact as possible.
- Transmiting at minimum rate between server & the user.
- Low complexity at server (Massive Random Access)

Interactive Video Coding (IVC)

The views are correlated and one should exploit this correlation through prediction:



Model based coding

Classic (predictive) coding:
- Encode residual between prediction and frame, i.e. X-Y

In IVC reference is different for each user.
Model based coding:
- Use Slepian-Wolf (SW) coder to perform compression.
- No need for the exact residual. Only its distribution is needed
- Model the correlation betwen current frame and set of predictions coming from possible previous frames.

Objective:
- What is the impact if we use an approximate model?
- What is a good model to select?

Optimal storage and transmission rates

Let (X,Y) be a static source determined by $\{P(X,Y_i),p_i\}_i$
The achievable rate-storage pair in lossless coding is:

no loss (better than expected) ← $R \geq H(X|Y_{i^*})$

loss in storage only (as expected) ← $S \geq \max_{j\in\{1..i..K\}} H(X|Y_j)$

A. Roumy and T. Maugey, "Universal lossless coding with random user access: The cost of interactivity," ICIP 2015

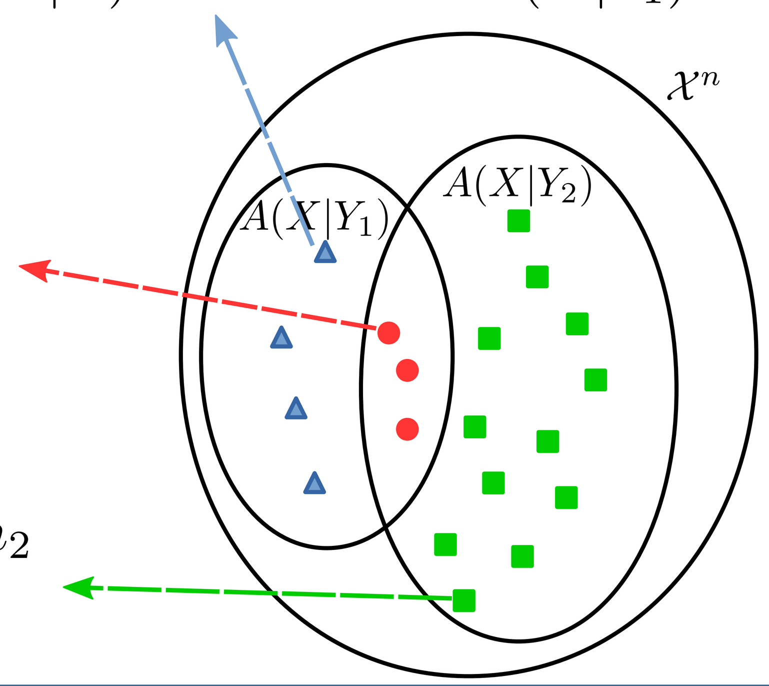
where Y_{i^*} is the prediction available at the deoder.

Fixed Length source coding in IVC

$A(X|Y)$:= set of typical sequences of $X^n = (X_1, X_2, ..., X_n)$ given Y^n. We need $nH(X|Y)$ bits on average to encode X^n given Y^n

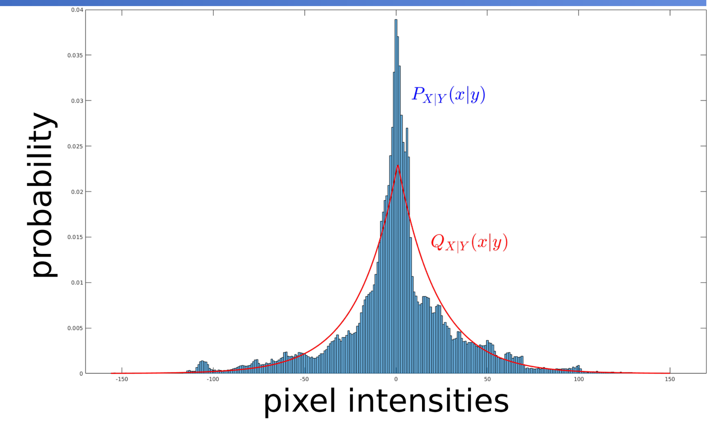
Example with 2 predictions Y_1, Y_2:

- Assign codeword of length $nH(X|Y)$ to each $x^n \in A(X|Y_1)$
- upon encoding $x^n \in A(X|Y_2)$:
  - $x^n \in A(X|Y_1) \cap A(X|Y_2)$: same length-$n_1$ as used in $A(X|Y_1)$ padded with $n_2 - n_1$ zeros
  - use the remaining length-n_2 codewords for the rest



Cost of using approximate distribution

Idea: send θ_i parameters of approximate Distribution $Q_{X|Y}$ instead of whole $P_{X|Y}$.



Vanishing error probability:

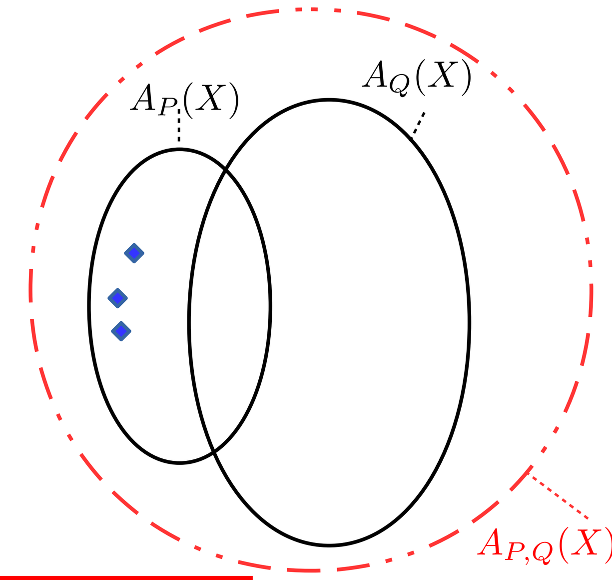
$A_P(X)$:= set of typical sequences of true distribution
$A_Q(X)$:= set of typical sequences of approximate distribution

- Enlarge the set $A_Q(X)$ to a larger set $A_{P,Q}(X)$ which contains $A_P(X)$
- Extra transmission rate:

$$\Delta R_i = D_{KL}(P(X,Y_i), Q(X,Y_i))$$

- Extra storage cost:

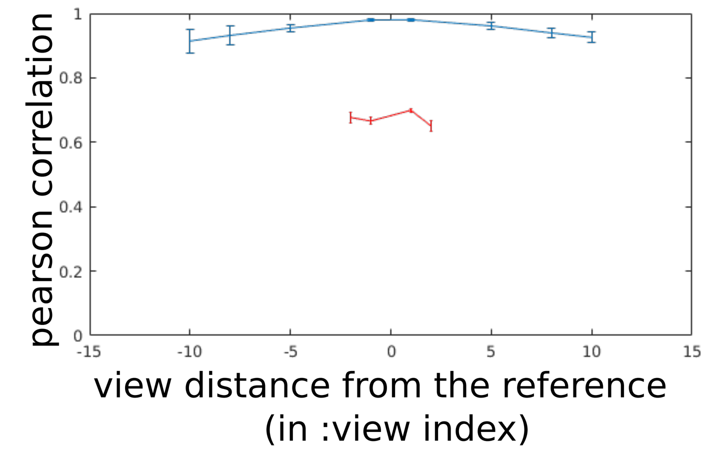
$$\Delta S_X = \max_j[H(X|Y_j) + \Delta R_j] - \max_j H(X|Y_j)$$



Experimental results

Predictions are provided by Depth-Image Based Rendering

Validity of additive model:
- Linear dependency evaluated
- X = Y+Z



Model selection:

	Tsukuba		Ballet	
	ΔS	ΔR	ΔS	ΔR
Laplace	0.634	0.610	0.552	0.3
Normal	1.43	1.401	0.996	0.555
Gassian Mixture	0.211	0.2	0.364	0.19
Erasure channel	Inf	Inf	Inf	Inf

Extra rate and extra storage cost in bit per pixel. Inf means infinity

Conclusion

- A criterion is proposed for model selection in IVC.
- Several distribution evaluated. It is not possible to code a view losslessly by completely relying on DIBR prediction.

This work was partially supported by the Cominlabs excellence laboratory with funding from the French National Research Agency (ANR-10-LABX-07-01) and by the Brittany Region (Grant No. ARED 9582 InterCOR).