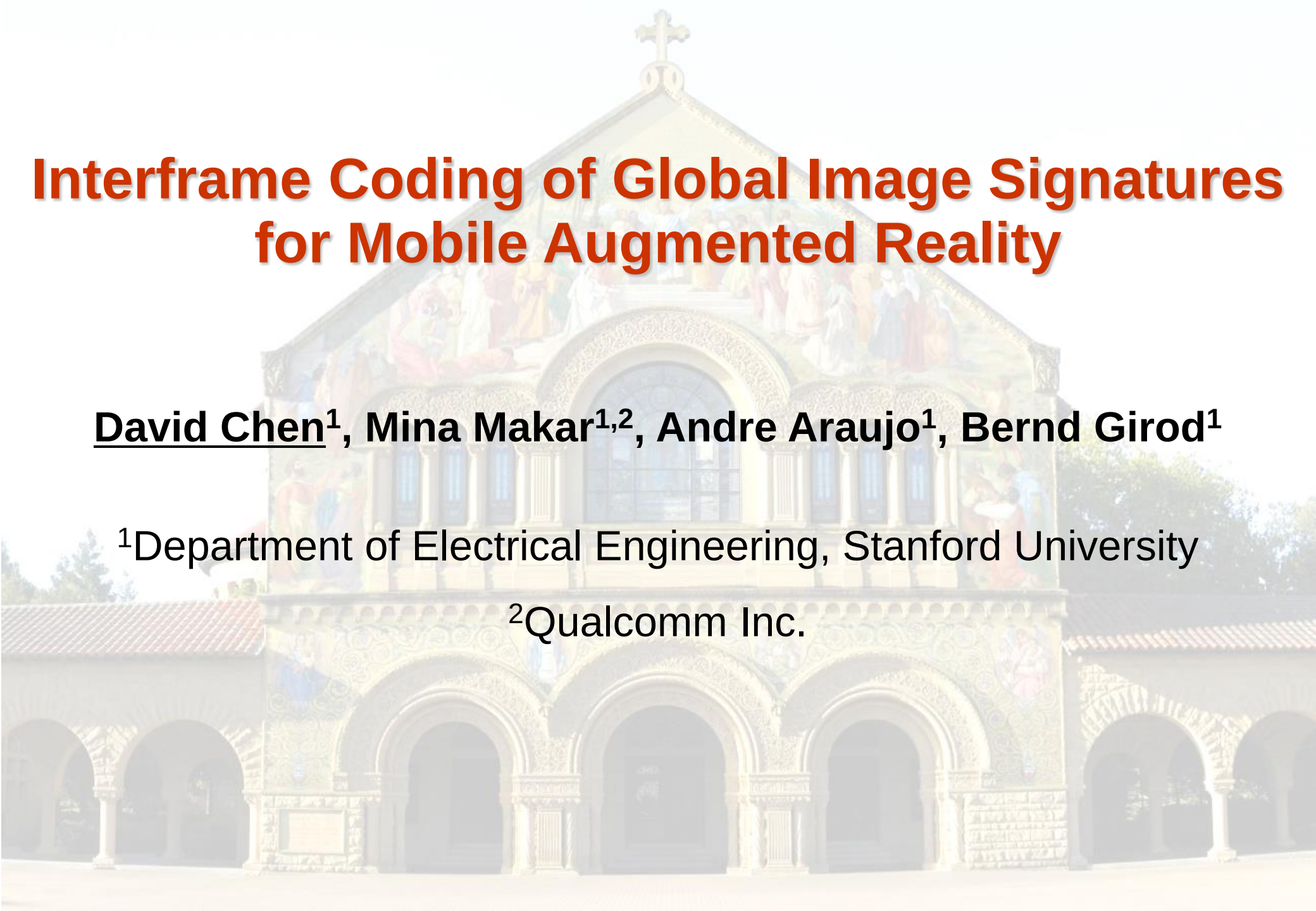


The background of the slide is a faded image of a large, ornate building, likely a Stanford University building. It features a prominent central archway and a large mural on the upper facade depicting various figures. The building has a red-tiled roof and is surrounded by a colonnade of arches.

Interframe Coding of Global Image Signatures for Mobile Augmented Reality

David Chen¹, Mina Makar^{1,2}, Andre Araujo¹, Bernd Girod¹

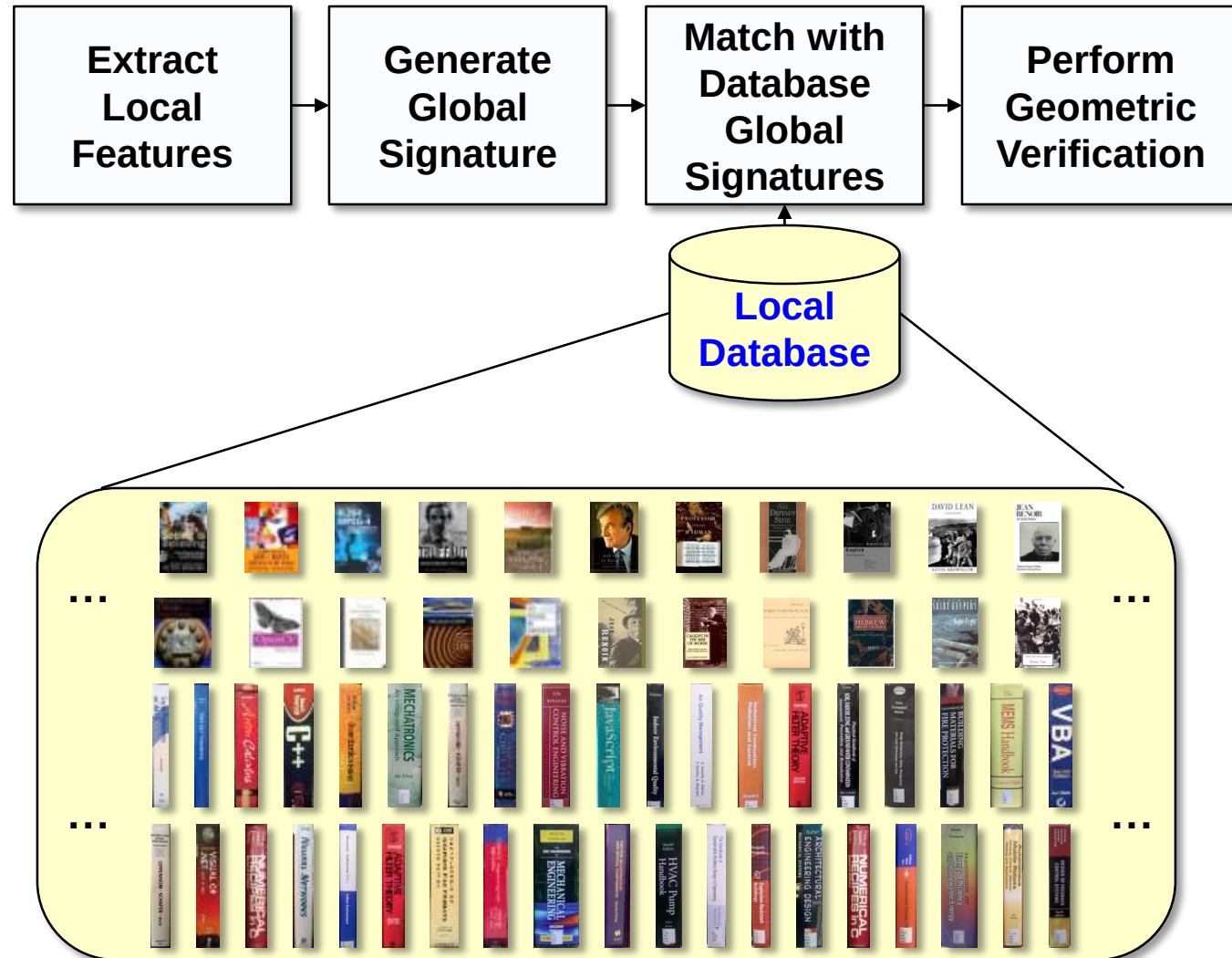
¹Department of Electrical Engineering, Stanford University

²Qualcomm Inc.



Samsung Galaxy S3 Smartphone

On-Device Image Matching System



Hybrid Image Matching System



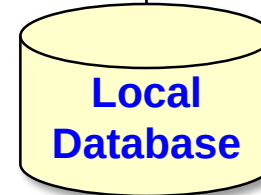
Extract
Local
Features

Generate
Global
Signature

Match with
Database
Global
Signatures

Perform
Geometric
Verification

*Send compact interframe
coded stream of global
signatures in uplink*



Match with
Database
Global
Signatures

*Send labels, local features, and
global signatures for top-ranked
database candidates in downlink*

0.62



0.51



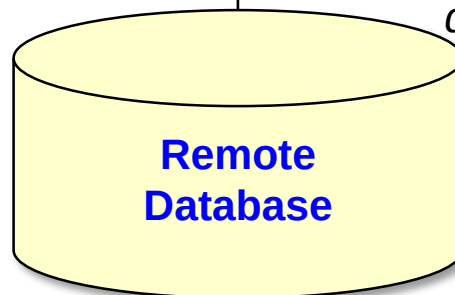
0.50



0.49



⋮



Outline

- Related work on feature compression
- Interframe coding of global signatures
 - Selective codeword propagation
 - Selective frame propagation
 - Selective frame propagation + local search
 - Global signature embedding
- Coding and retrieval results



Related Work on Feature Compression

Local Image Features

Random projections [Yeo, 2008]

Transform coding [Chandrasekhar, 2009]

Location histogram coding [Tsai, 2009]

CHoG [Chandrasekhar, 2009]

Bag of hash bits [He, 2011]

Temporally coherent keypoint detector [Makar, 2012]

Descriptor and location predictive coding [Makar, 2012] [Baroffio, 2014]

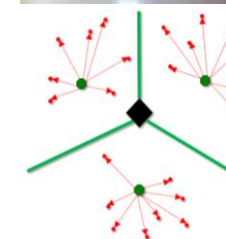
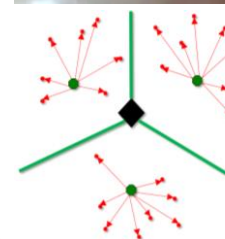
Global Image Signatures

Trace transform [Brasnett, 2007]

Tree histogram coding [Chen, 2009]

Residual vectors [Perronnin, 2010]
[Jegou, 2010] [Chen, 2011]

Our Current Work



Intraframe Coding

Interframe Coding



Temporally Coherent Keypoint Detection

Makar et al., 2012



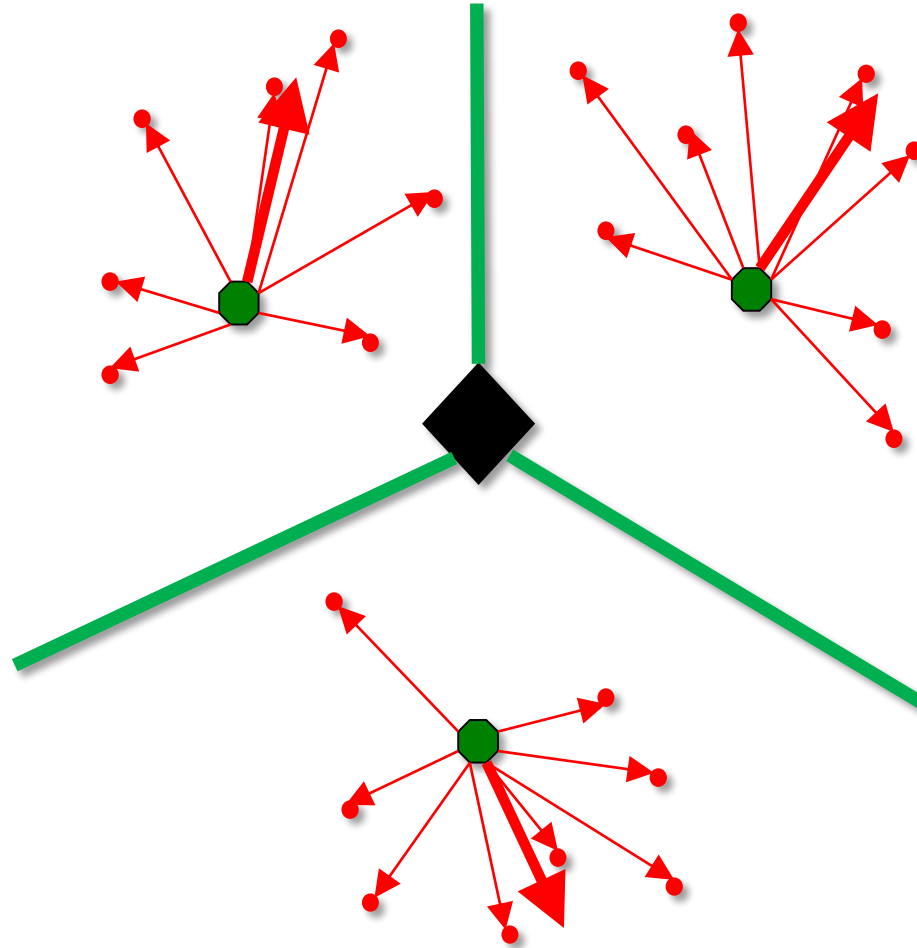
Detection (D) Frame



Forward Propagation (FP) Frame



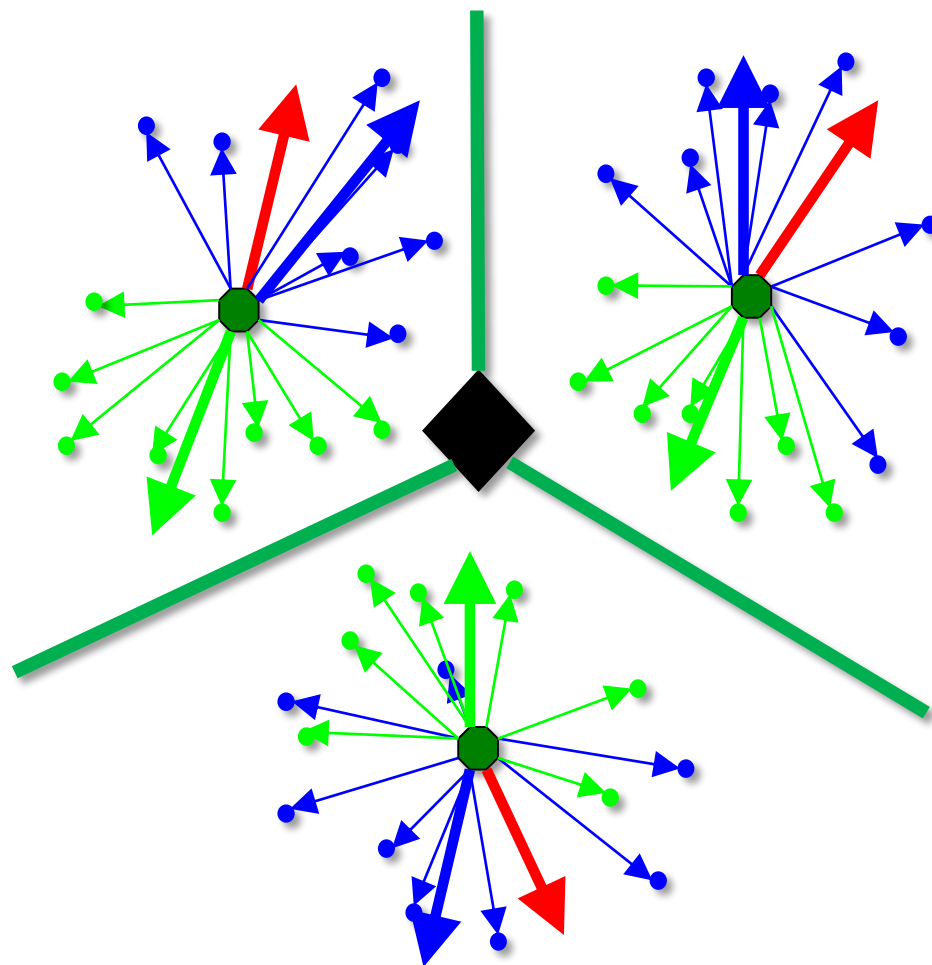
Generating Feature Residuals



$k = 3$



Generating Feature Residuals

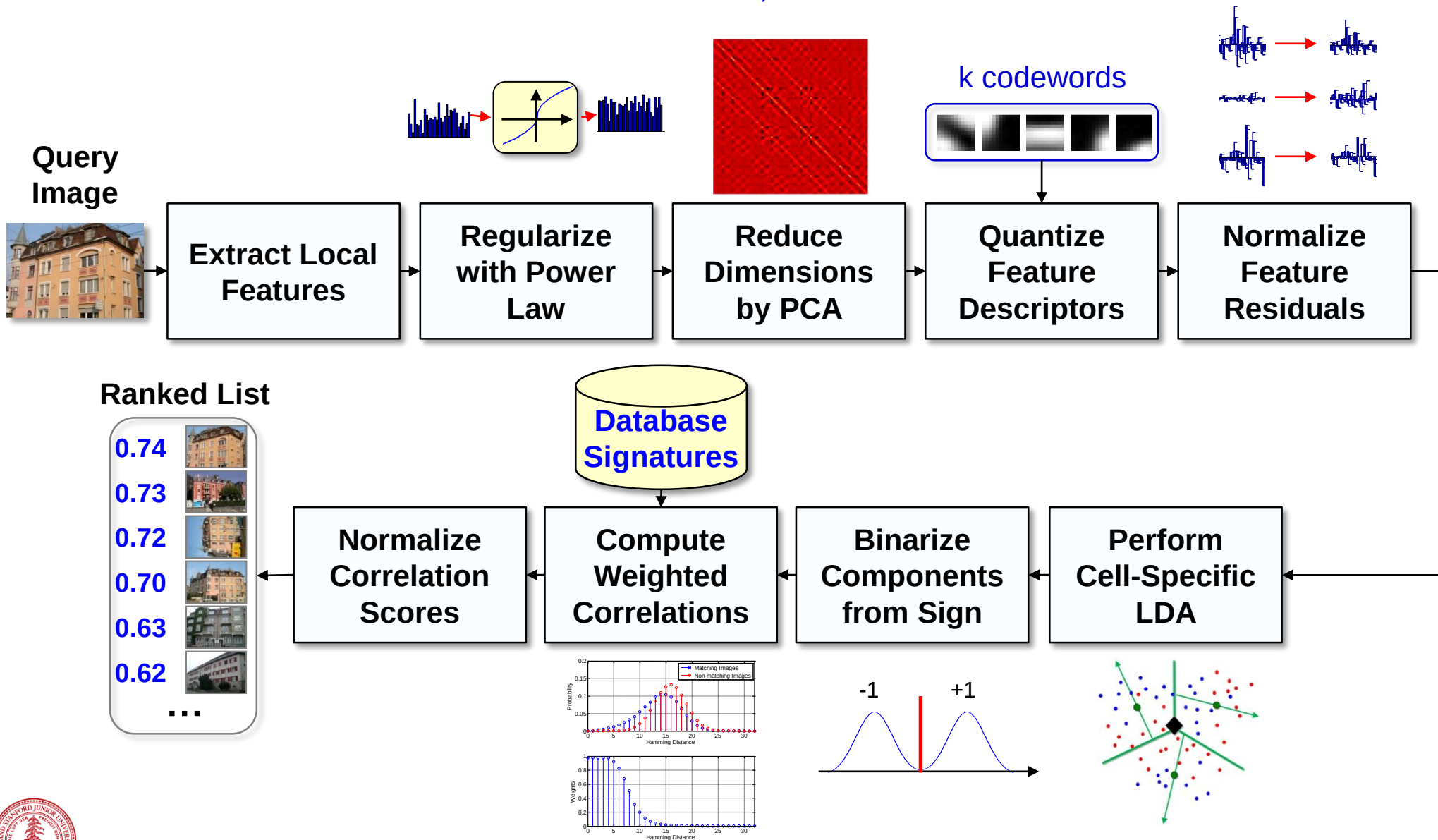


$k = 3$

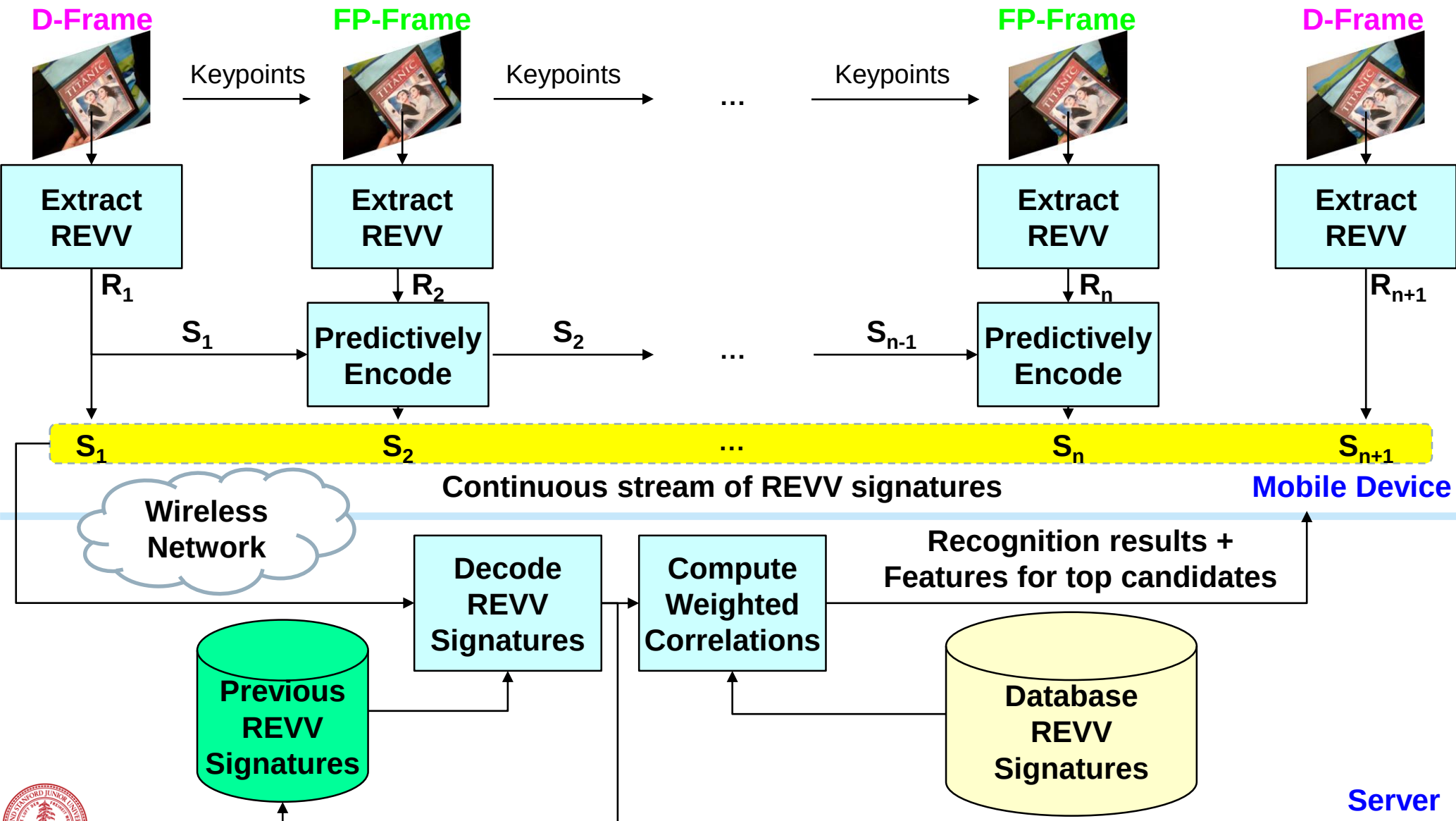


Residual Enhanced Visual Vector (REVV)

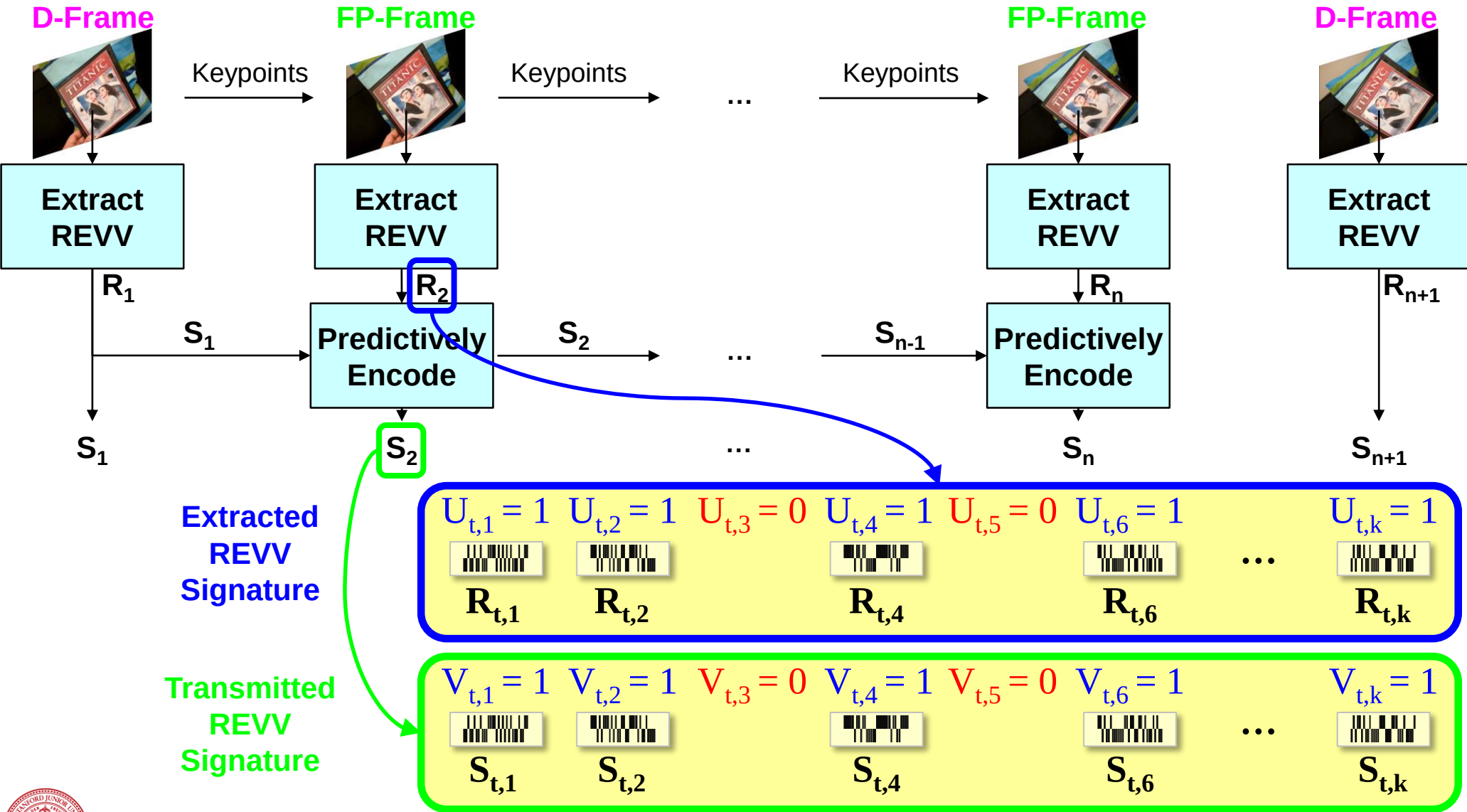
Chen et al., 2011



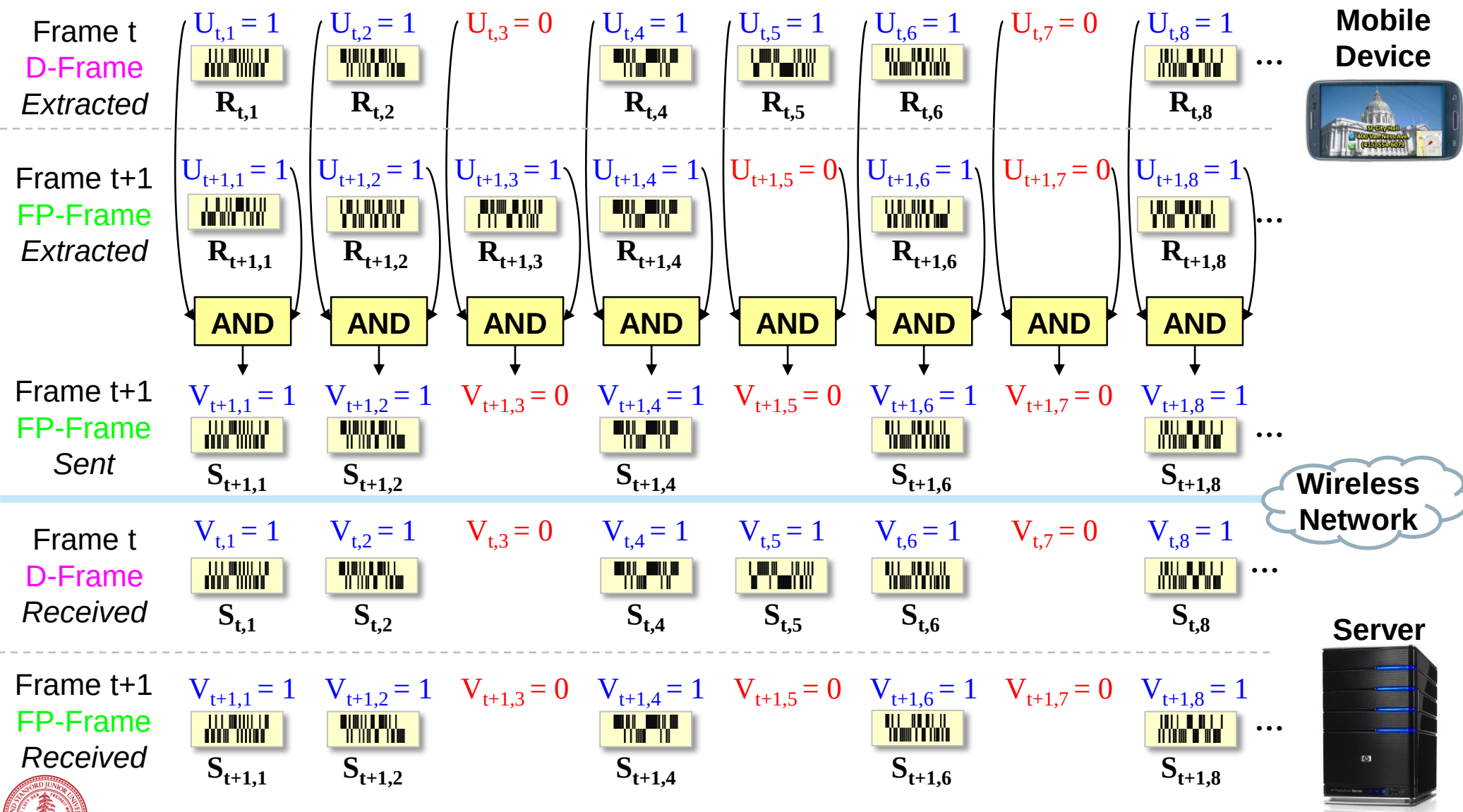
Interframe Coding of REVV



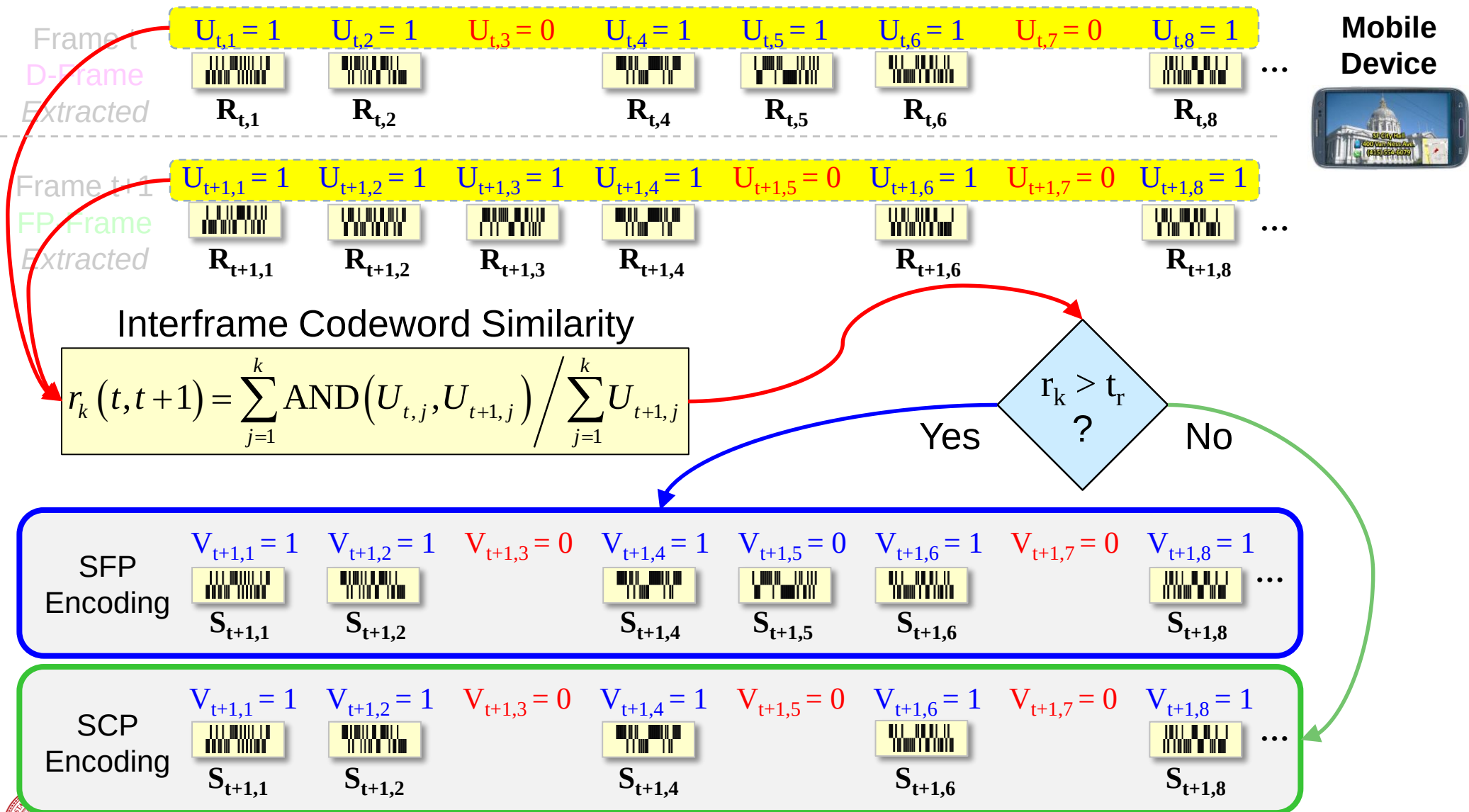
Interframe Coding of REVV



Selective Codeword Propagation (SCP)



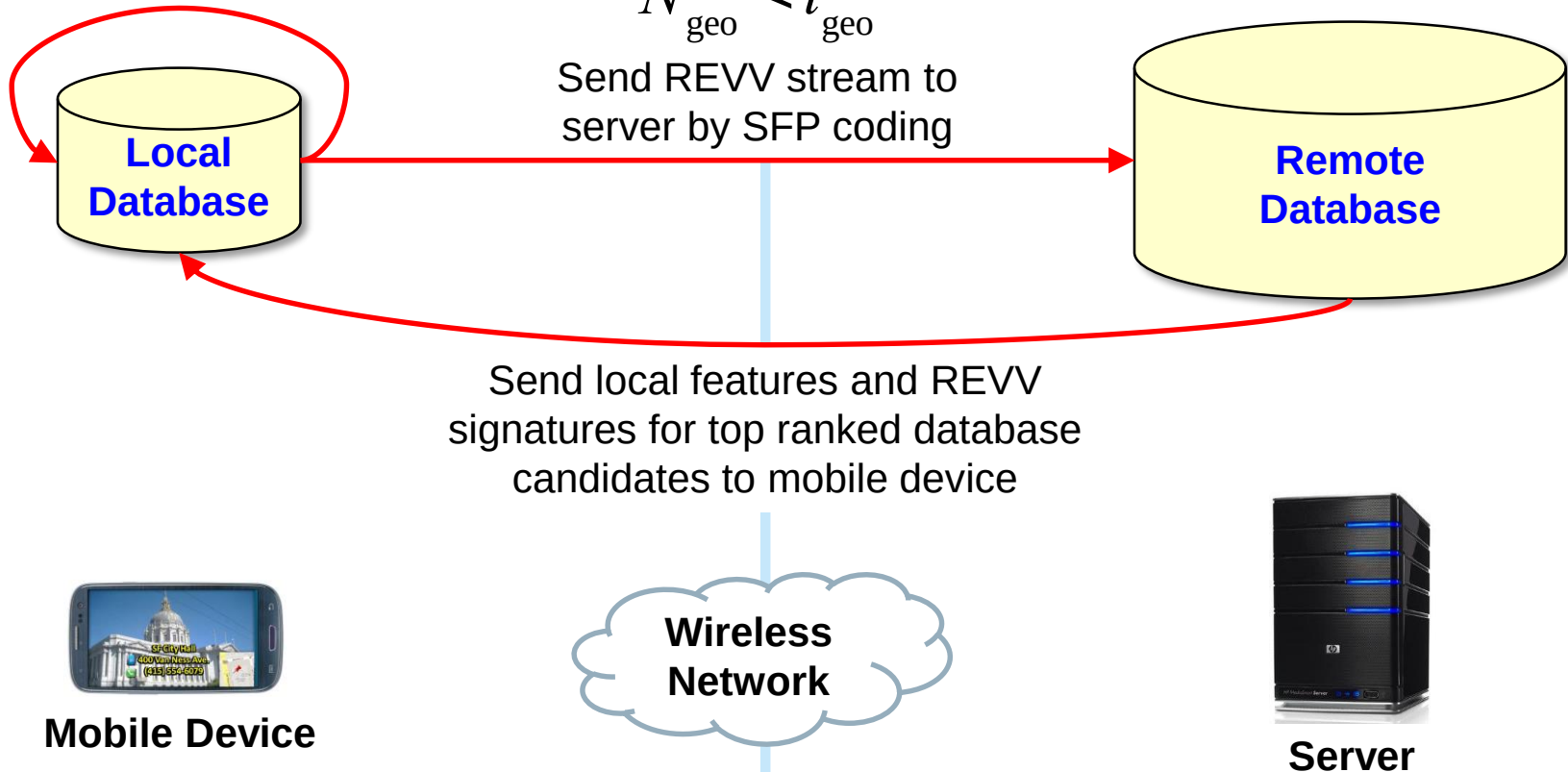
Selective Frame Propagation (SFP)



SFP + Local Search (SFP + LS)

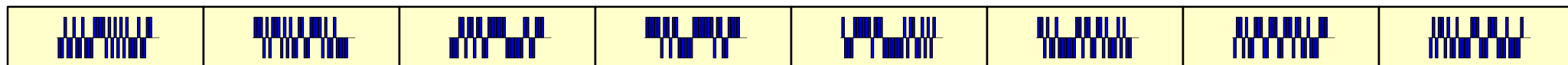
Number of inliers in
geometric verification

$N_{\text{geo}} \geq 3 t_{\text{geo}}$
Terminate query locally
on mobile device

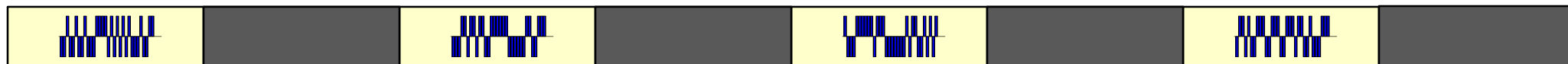


Embedded Global Signatures

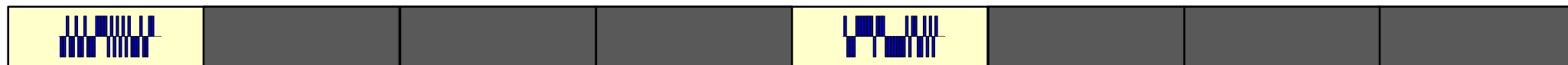
Codeword 1 Codeword 2 Codeword 3 Codeword 4 Codeword 5 Codeword 6 Codeword 7 Codeword 8



Level 1 → Highest Bitrate



Level 2 → Medium Bitrate



Level 3 → Lowest Bitrate

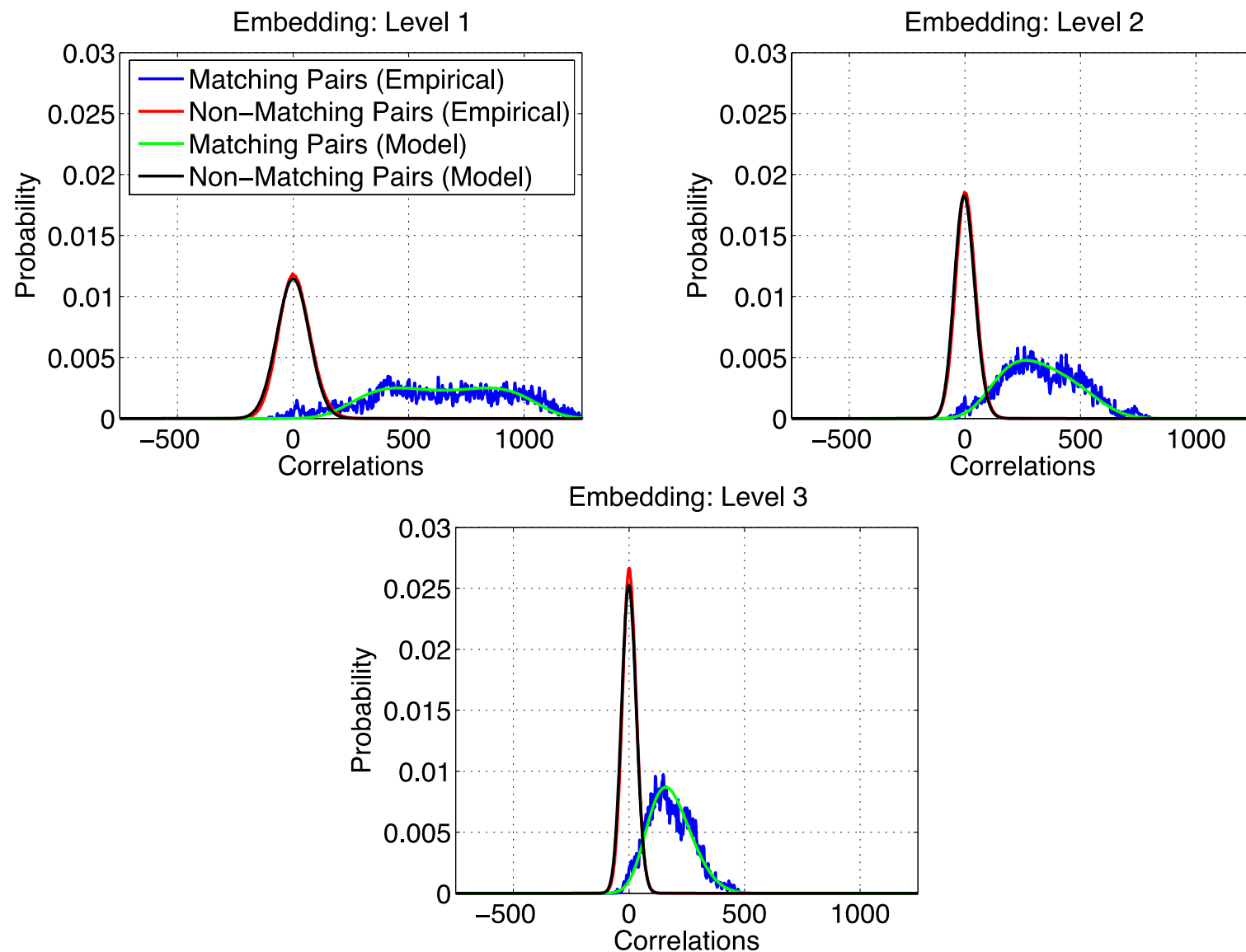


Outline

- Related work on feature compression
- Interframe coding of global signatures
 - Selective codeword propagation
 - Selective frame propagation
 - Selective frame propagation + local search
 - Global signature embedding
- Coding and retrieval results



Analysis of Retrieval Performance



Stanford Streaming MAR Dataset

Mobile Augmented Reality



32 VGA-resolution query videos recorded with a camera phone
Database of 23 labeled objects + 1M distractor images

[Makar et al., 2013]

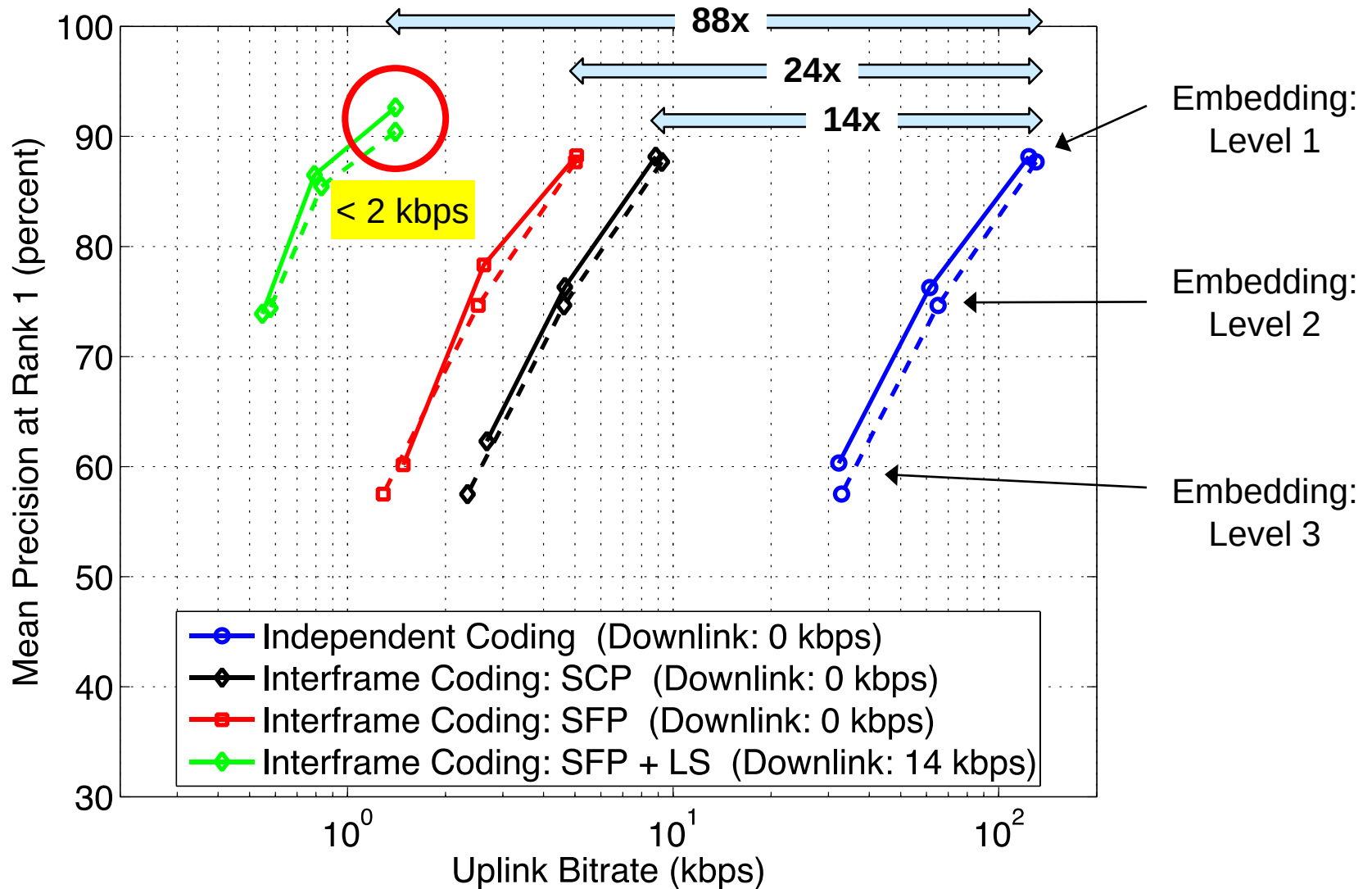


Experimental Setup

- Interframe coding parameters
 - $N_{D-Frames} = 1$ and $N_{FP-Frames} = 29$ for frame rate of 30 fps
 - Interframe codeword similarity threshold: $t_r = 0.9$
 - RANSAC threshold: $t_{geo} = 25$ feature matches
- REVV signature parameters
 - 250 SIFT features extracted for every D-Frame
 - Dimensionality reduction to $d_{LDA} = 32$
 - Codebook of $k = 190$ codewords
- Retrieval accuracy vs. uplink bitrate comparison



Retrieval Accuracy vs. Uplink Bitrate



Conclusions

- Developed efficient methods for interframe coding of a continuous stream of global REVV signatures
 - Adapts bitrate in response to current viewfinder contents
 - Combines advantages of local and remote database search
 - Performs local database update at same time as query expansion
- Analyzed how retrieval accuracy varies with bitrate
 - Models distributions of REVV correlation scores
 - Quantifies bitrate savings for all coding methods
- Achieved substantial bitrate savings
 - Reduces uplink bitrate by almost 2 orders of magnitude
 - Requires only a small downlink bitrate



The background of the slide is a photograph of the Main Quad building at Stanford University. The building is a large, light-colored stone structure with a prominent central gable. The gable is filled with a large, colorful mosaic depicting a religious scene, likely the Resurrection, with figures in various poses and colors. Above the mosaic is a large arched window. Below the window is a row of smaller arched windows. The building has a series of arches along its base, and a cross is visible on the roofline. The overall scene is bright and clear.

Thank You

dmchen@stanford.edu