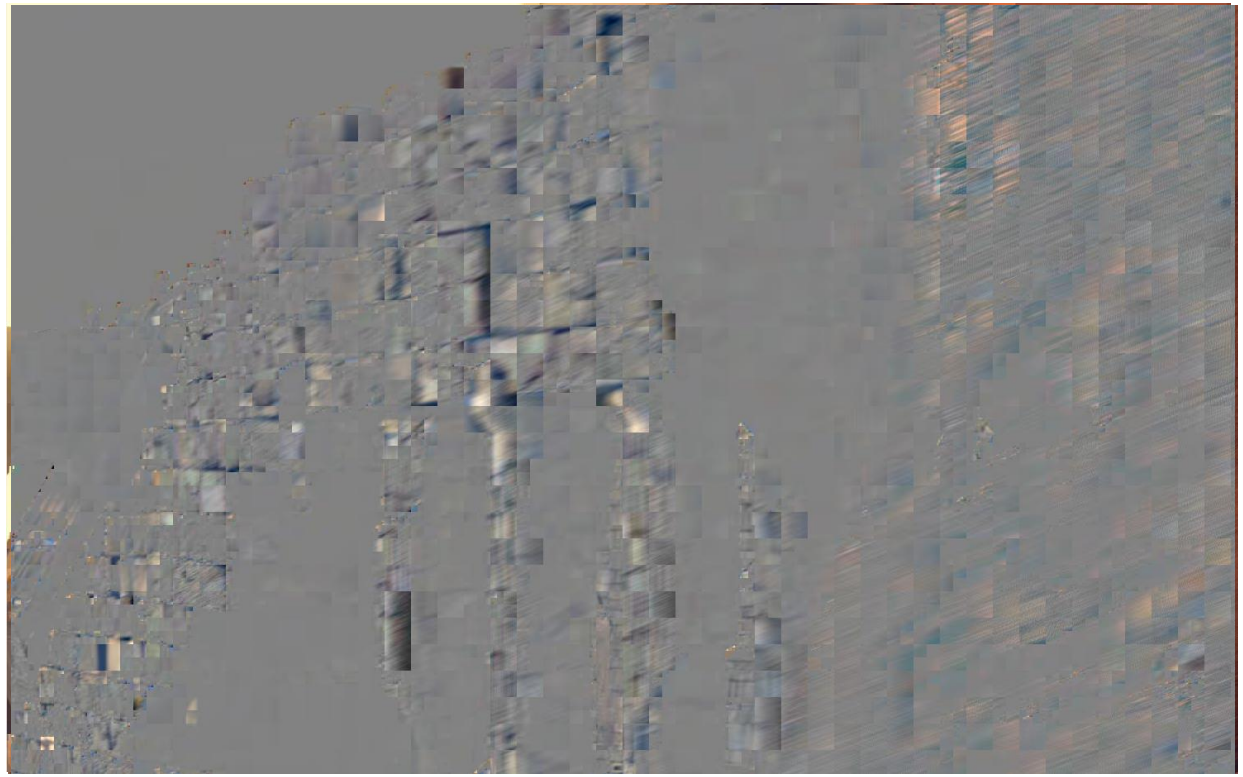
3D Models in Motion Compensation

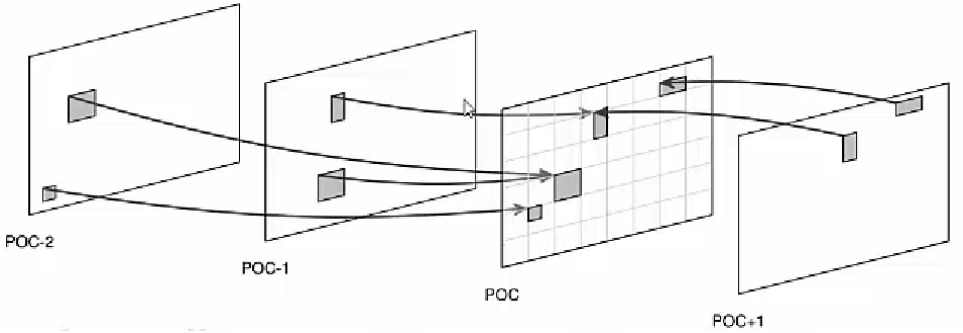
Hossein Golestani
Institut für Nachrichtentechnik
RWTH Aachen University

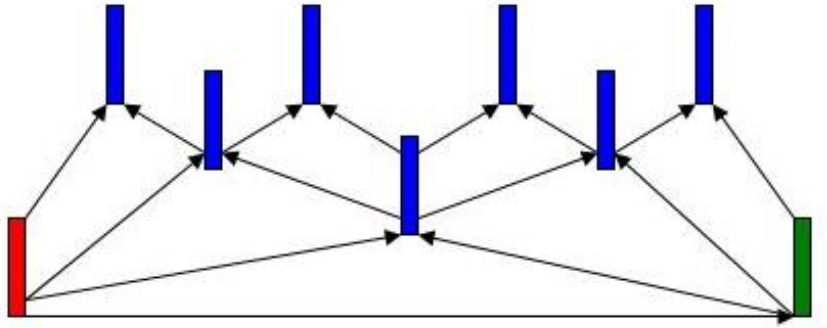
Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- Multi-View Reconstruction (MVR)
- Virtual view synthesis for 2D frame prediction
- Coding results
- Conclusion

Introduction to the main idea







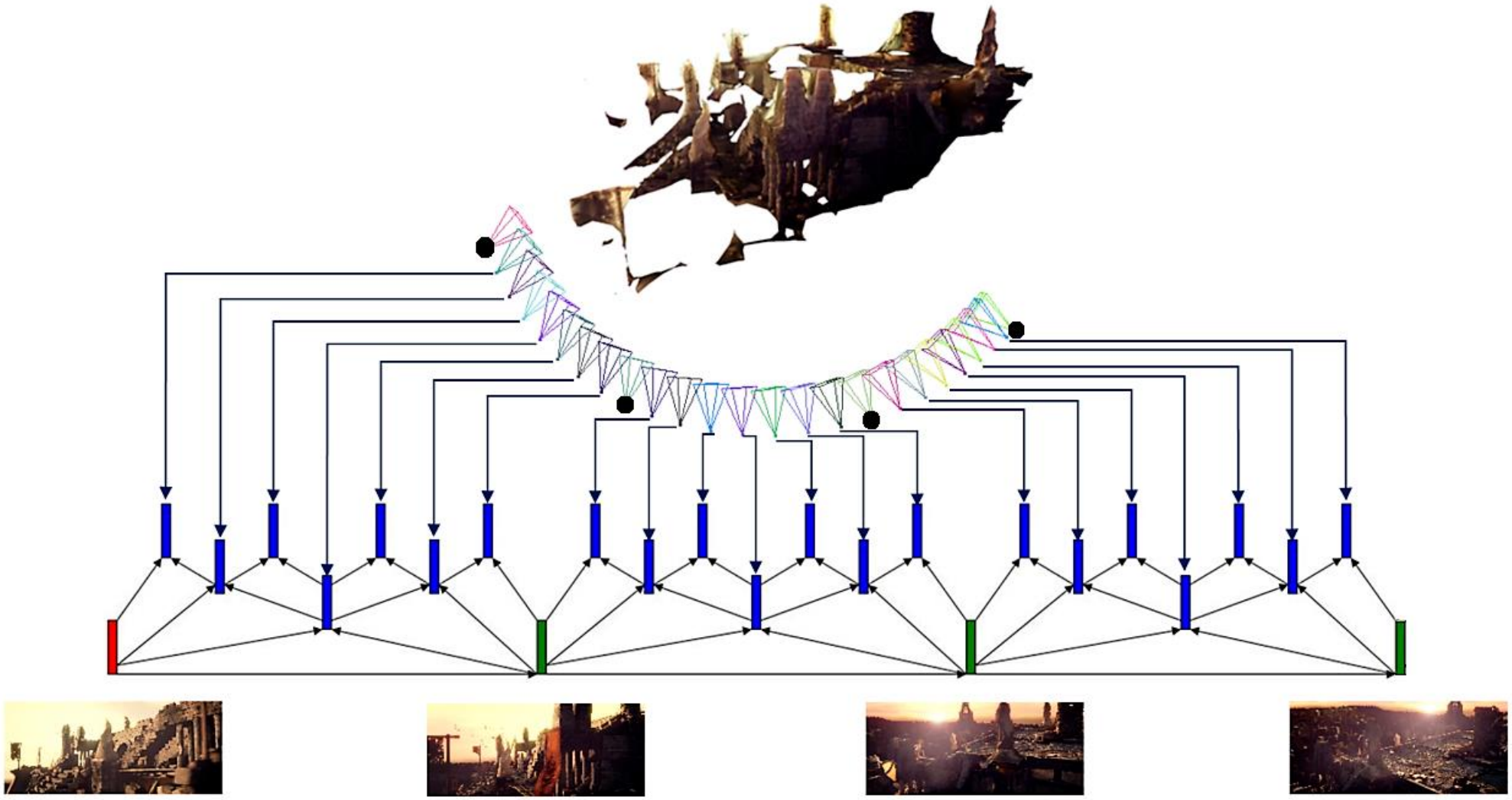
Introduction to the main idea

- Target Sequences (moving camera)

Introduction to the main idea

- 3D model based frame prediction
  - **B-Frame Prediction**
  - Key-Frame Prediction

Introduction to the main idea



Introduction to the main idea

- 3D model based frame prediction
  - B-Frame Prediction
  - **Key-Frame Prediction**

Introduction to the main idea

Outline

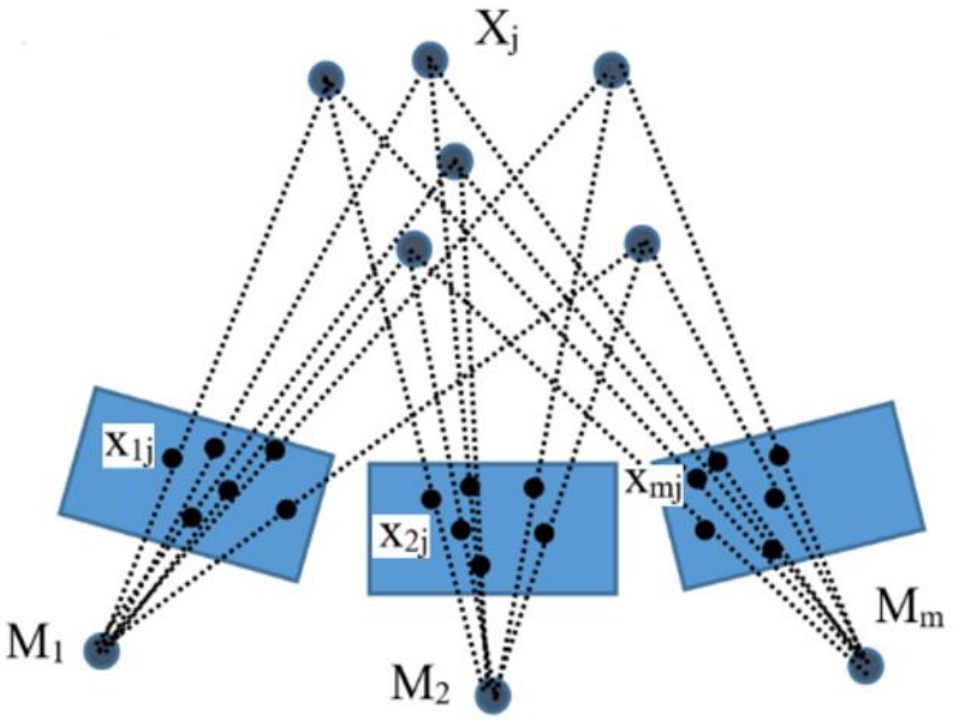
- Introduction to the main idea
- **Structure from Motion (SfM)**
- Multi-View Reconstruction (MVR)
- Virtual view synthesis for 2D frame prediction
- Coding results
- Conclusion

Structure from Motion (SfM)

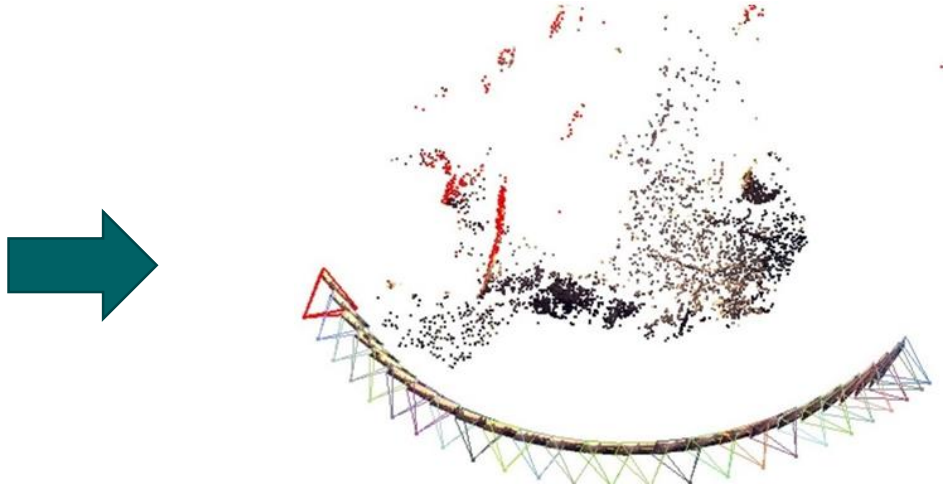
- Structure from Motion (SfM) [1]

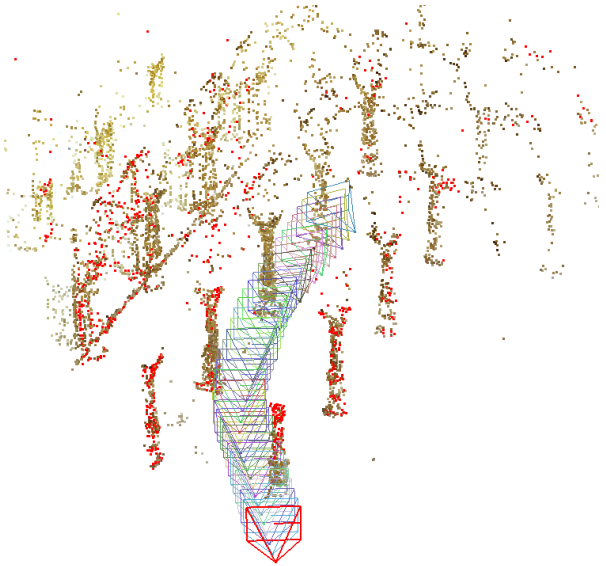
$$x_{ij} = M_i X_j, \qquad i = 1, \dots, m \;\; j = 1, \dots, n$$

- How to solve SfM?
  - Feature extraction & Finding corresponding points
  - Finding Fundamental matrix F (${x'}^T F x = 0$)
  - Estimating projective cameras
  - Triangulating
  - Bundle Adjustment

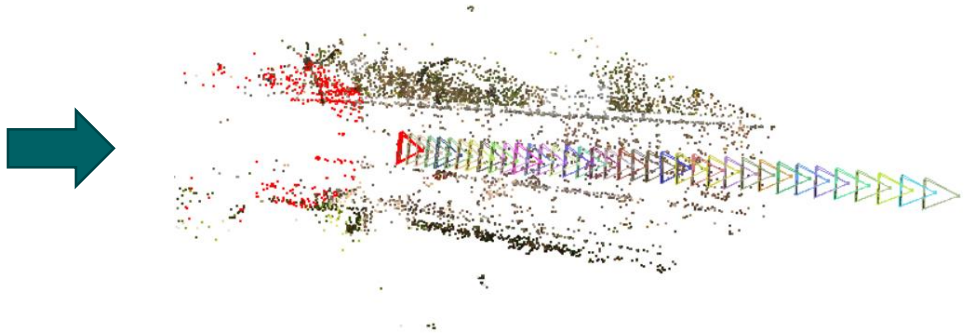


[1] R. Hartley and A. Zisserman, “Multiple view geometry in computer vision“, second edition, Cambridge press, 2013.

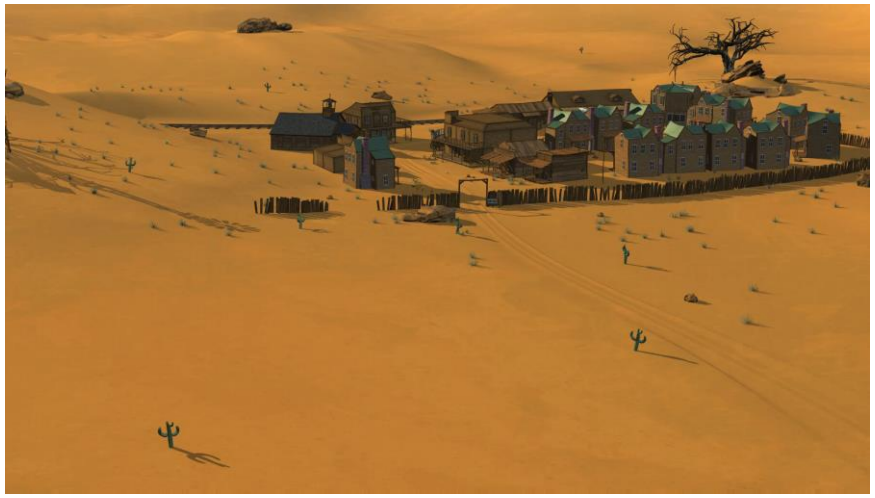


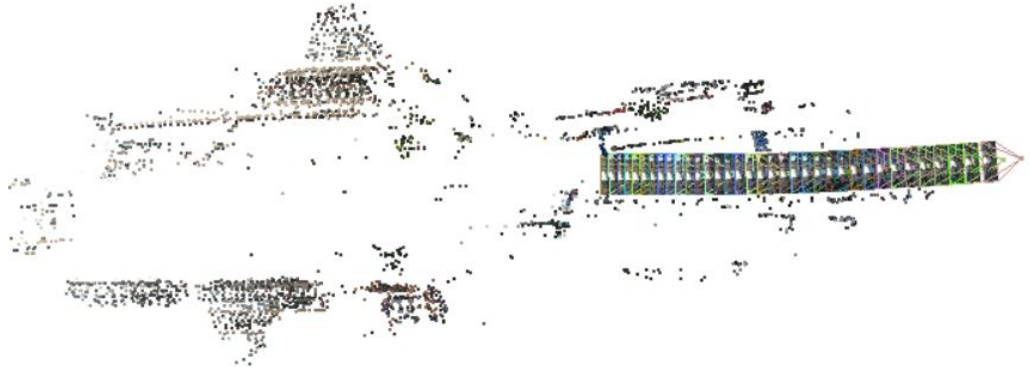


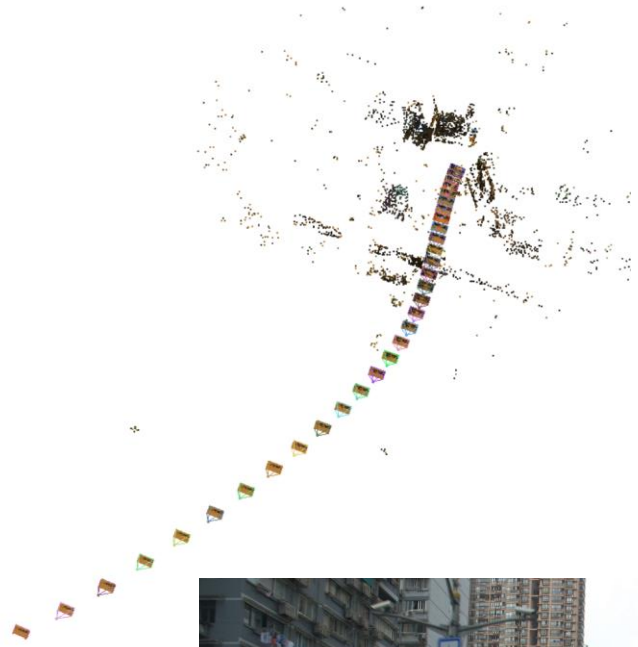
SfM Results

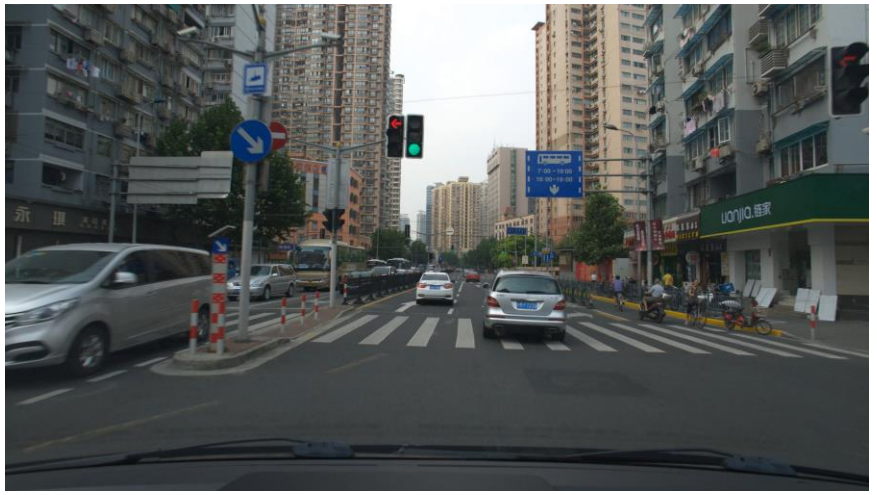


SfM Results









Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- **Multi-View Reconstruction (MVR)**
- Virtual view synthesis for 2D frame prediction
- Coding results
- Conclusion

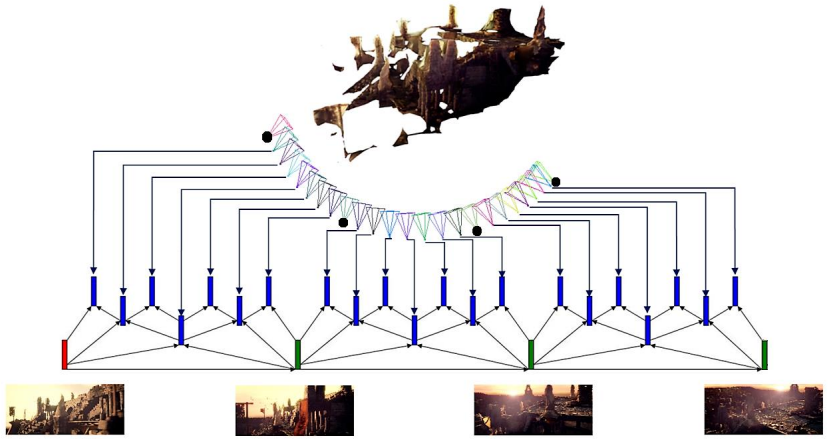
Multi-View Reconstruction (MVR)

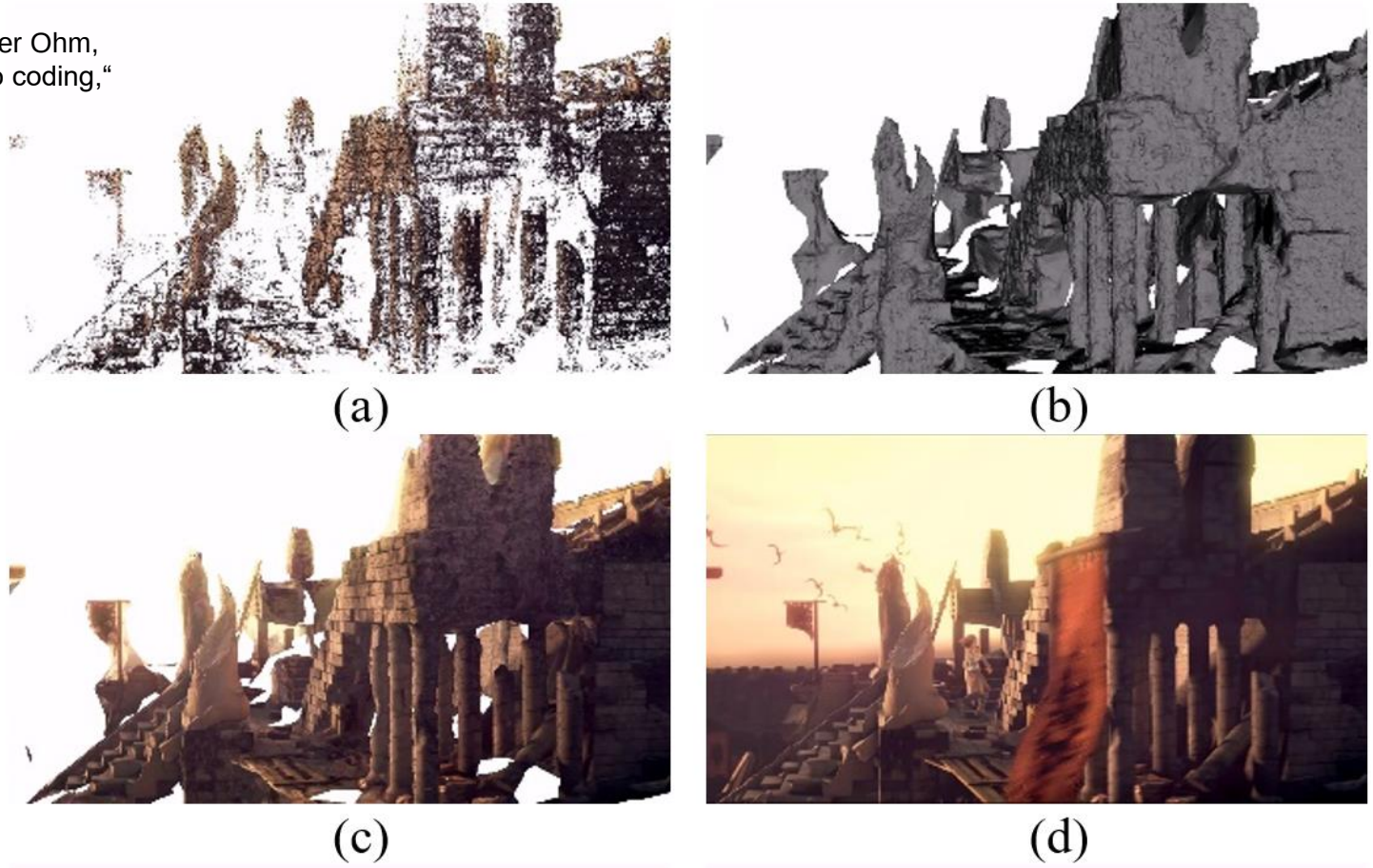
- Generating a quasi-dense point cloud
- Visibility based surface reconstruction
  - Delaunay tetrahedralization
  - Surface optimization with minimum s-t cut [2])
- The variational refinement of mesh to optimize its photo-consistency (Minimizing reprojection error)

[2] H. Vu, P. Labatut, J. Pons and R. keriven, “High accuracy and visibility-consistent dense multiview stereo”, IEEE Transaction on pattern analysis and machine intelligence, vol. 34, no. 5, 2012.

Application in Video Coding?

[3] Hossein B. Golestani, Jens Schneider, Mathias Wien and JensRainer Ohm, „Point cloud estimation for 3D structure based frame prediction in video coding,“ ICME 2017, Hong Kong.





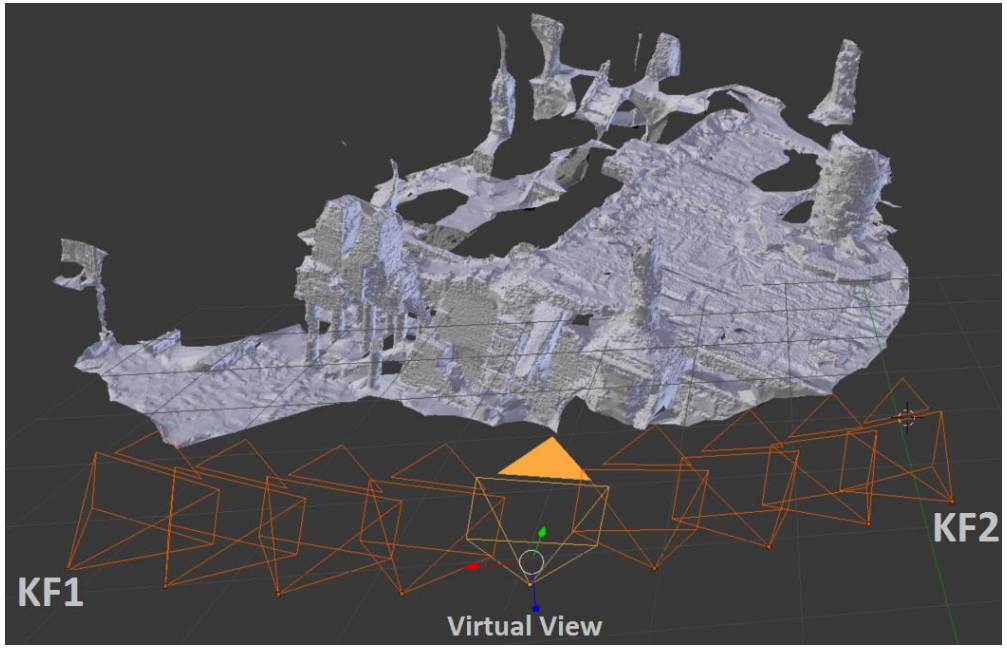
(a) quasi-dense point cloud, (b) shaded surface, (c) textured surface and (d) ground thruth

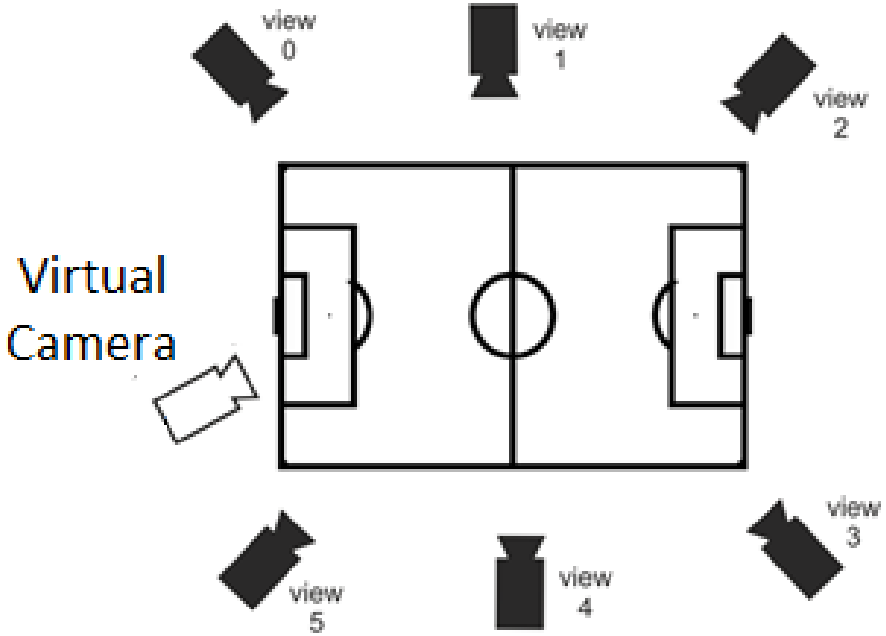
Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- Multi-View Reconstruction (MVR)
- **Virtual view synthesis for 2D frame prediction**
- Coding results
- Conclusion

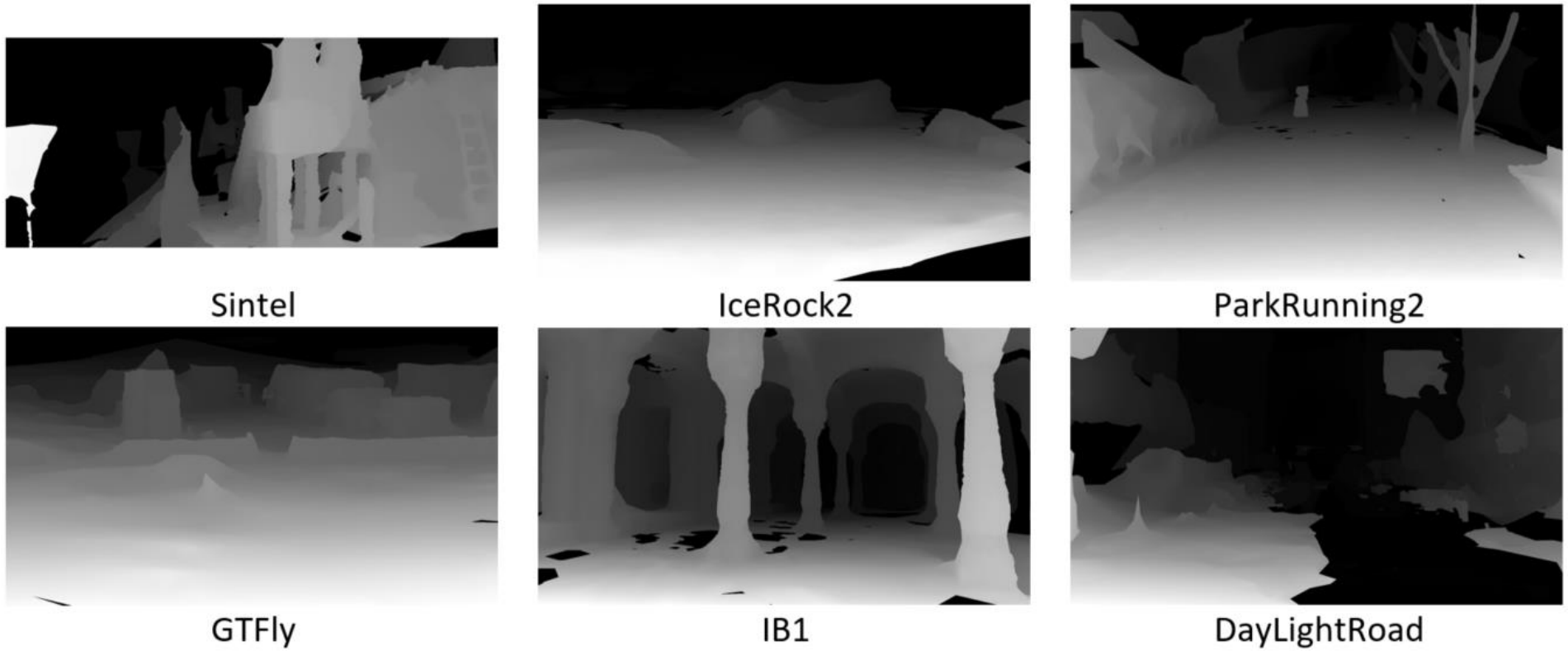
New method for frame prediction: Virtual View Synthesis

- How to improve the quality of predicted frame? Virtual View synthesis
- What does it need?
  - Left and right views
  - Camera parameters for left, right and virtual cameras
  - Depth information of left and right views (depthmaps, Z-min & Z-max)





Estimated Depthmaps



Virtual View Synthesis

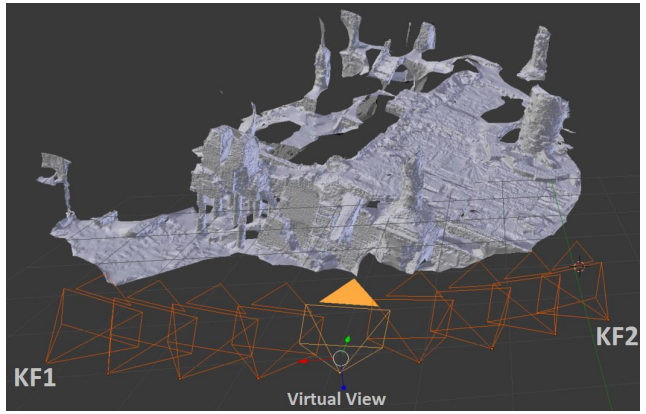
- Steps:
  - Estimating the depthmaps of the virtual view from KFs‘ depthmaps
    - Finding a homography transformation between real and virtual views for each depth level

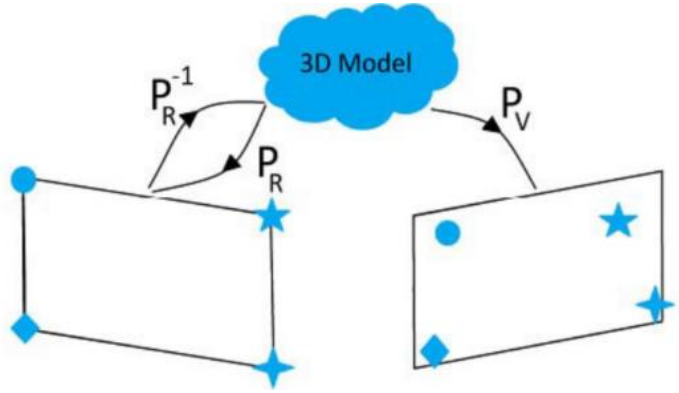
$$X_V[i] = P_V P_R^{-1} X_R[i], \quad i = depth\,levels$$

  - Synthesizing the virtual view by warping textures from left and right key-frames

$$[h, w, 1]^T = H^{-1}[i] \times [u, v, 1]^T, \quad [u, v]: virtual\,view, \quad [h, w]: real\,view$$

  - Blending the synthesized view from left and right

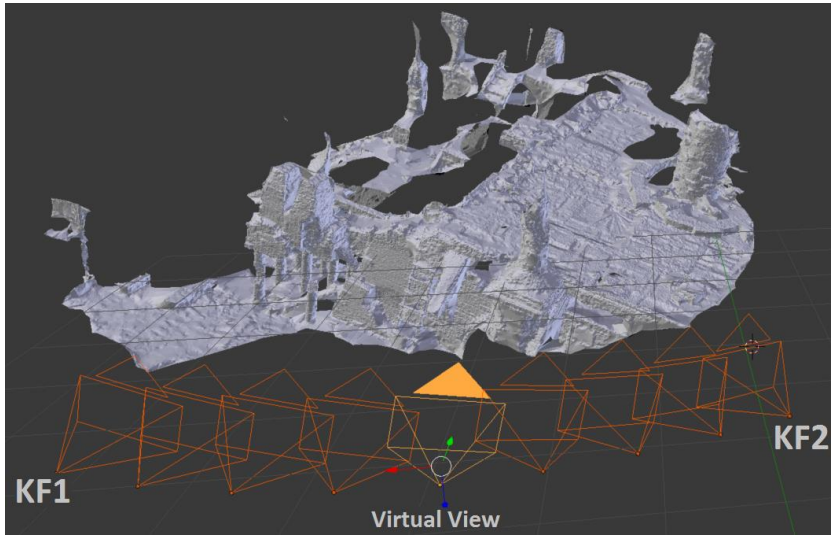




Synthesized frames

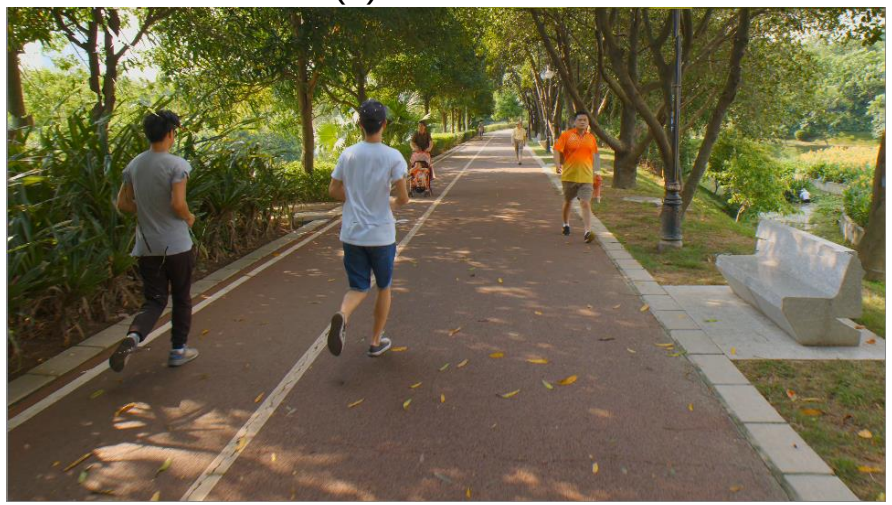
(a) Ground Truth

(b) Synthesized (virtual view)

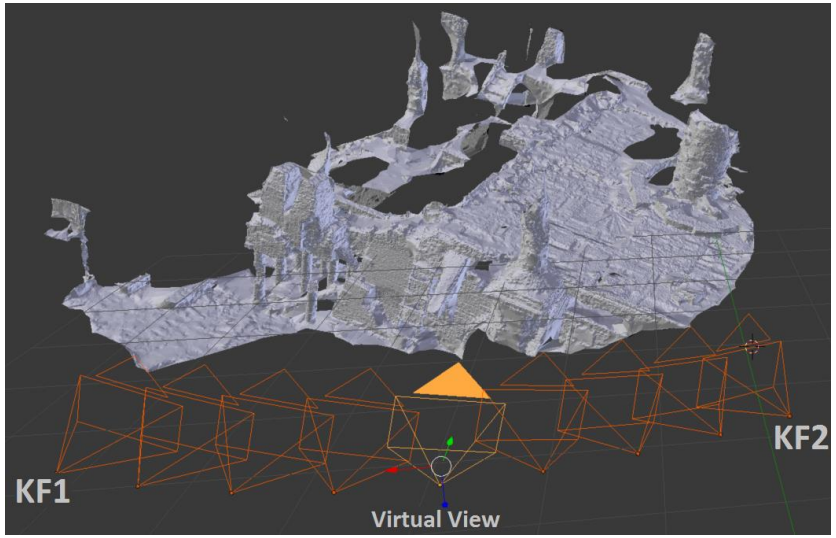


Synthesized frames

(a) Ground Truth



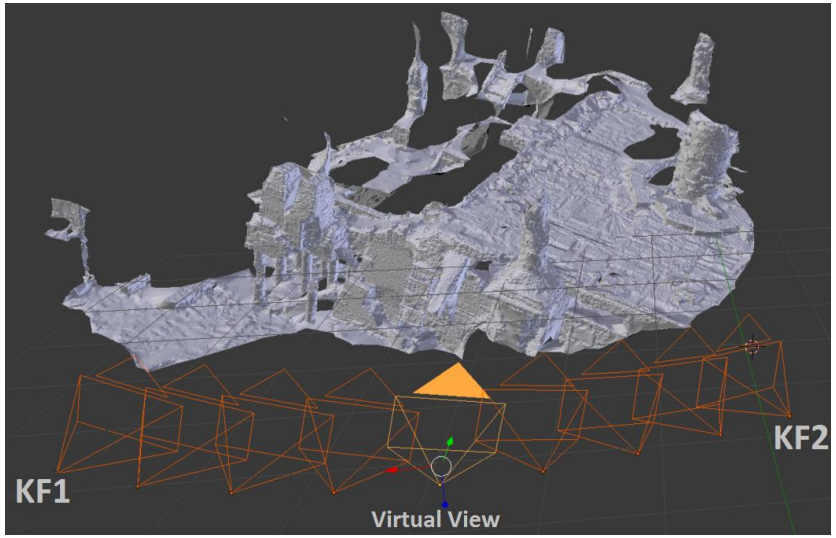
(b) Synthesized (virtual view)



Synthesized frames

(a) Ground Truth

(b) Synthesized (virtual view)

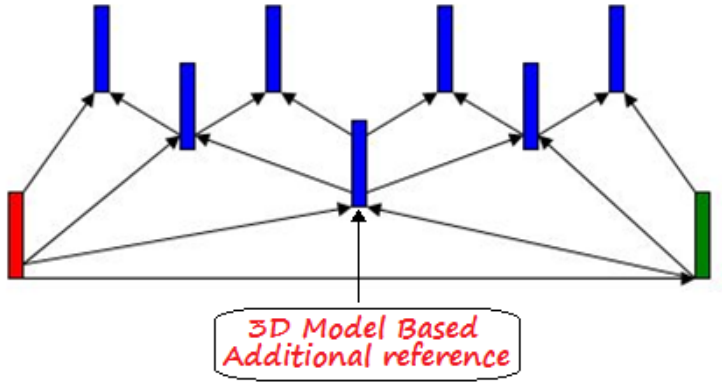


Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- Multi-View Reconstruction (MVR)
- Virtual view synthesis for 2D frame prediction
- **Coding results**
- Conclusion

SfM based frame prediction in Video Coding

- HEVC test Model (HM 16.7)
  - SfM-based prediction is added to L0 and L1
  - The encoder can choose between
    - its built-in reference images
    - The offered SfM-based prediction
  - QP: 25, 29, 33 and 37, Random Acess main profile

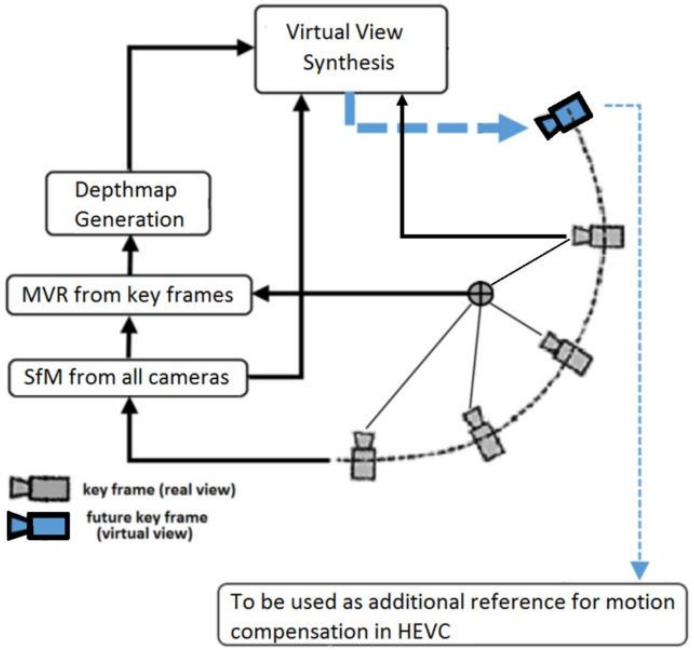


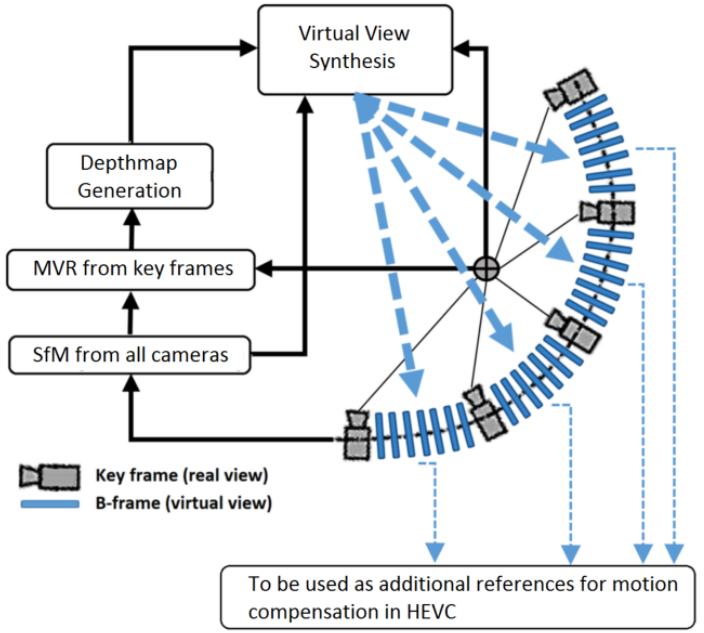

```
POC    24 TId: 0 ( B-SLICE, nQP 38 QP 38 )      48048 bits [Y 39.0652 dB    U 46.8478 dB    V 47.1814 dB] [ET   158 ] [L0 16 8 24(AddRef) ] [L1 16 8 24(AddRef) ]
POC    20 TId: 1 ( B-SLICE, nQP 39 QP 39 )       6552 bits [Y 39.3872 dB    U 46.9528 dB    V 47.2622 dB] [ET   149 ] [L0 16 12 20(AddRef) ] [L1 24 16 20(AddRef) ]
POC    18 TId: 2 ( B-SLICE, nQP 40 QP 40 )       3576 bits [Y 38.7807 dB    U 46.8727 dB    V 47.1718 dB] [ET   161 ] [L0 16 12 18(AddRef) ] [L1 20 24 18(AddRef) ]
POC    17 TId: 3 ( B-SLICE, nQP 41 QP 41 )        992 bits [Y 38.8425 dB    U 46.8946 dB    V 47.2129 dB] [ET   144 ] [L0 16 18 17(AddRef) ] [L1 18 20 17(AddRef) ]
POC    19 TId: 3 ( B-SLICE, nQP 41 QP 41 )       1464 bits [Y 38.5592 dB    U 46.8896 dB    V 47.1888 dB] [ET   158 ] [L0 18 16 19(AddRef) ] [L1 20 24 19(AddRef) ]
POC    22 TId: 2 ( B-SLICE, nQP 40 QP 40 )       3232 bits [Y 39.1846 dB    U 46.9233 dB    V 47.1872 dB] [ET   145 ] [L0 20 16 22(AddRef) ] [L1 24 20 22(AddRef) ]
POC    21 TId: 3 ( B-SLICE, nQP 41 QP 41 )       1576 bits [Y 38.8811 dB    U 46.9153 dB    V 47.1979 dB] [ET   158 ] [L0 20 16 21(AddRef) ] [L1 22 24 21(AddRef) ]
POC    23 TId: 3 ( B-SLICE, nQP 41 QP 41 )       1656 bits [Y 38.7014 dB    U 46.8517 dB    V 47.1399 dB] [ET   144 ] [L0 22 20 23(AddRef) ] [L1 24 22 23(AddRef) ]
POC    32 TId: 0 ( P-SLICE, nQP 37 QP 37 )     383656 bits [Y 38.9582 dB    U 46.6173 dB    V 46.8710 dB] [ET   226 ] [L0 32(AddRef) 32(AddRef) 32(AddRef) ] [L1 ]
```

Seq.	Sintel	IceRock2	GTFly	ParkRunning2	DayLightRoad	IB1
BD-Rate (%)	-2.21	-10.76	-6.64	-2.55	-2.80	-5.91
BD_PSNR (dB)	+0.08	+0.32	+0.19	+0.08	+0.04	+0.07

Conclusion

- An SfM-based frame prediction scheme has been presented.
- SfM for camera calibration
- MVS for surface reconstruction.
- Virtual view synthesis for synthesizing novel views.
- SfM-based prediction is added to reference lists (L0 and L1)
- So far, 2% to 10% bitrate reduction is achieved.
- Estimating better depthsmaps will improve results.





Vielen Dank
für Ihre Aufmerksamkeit