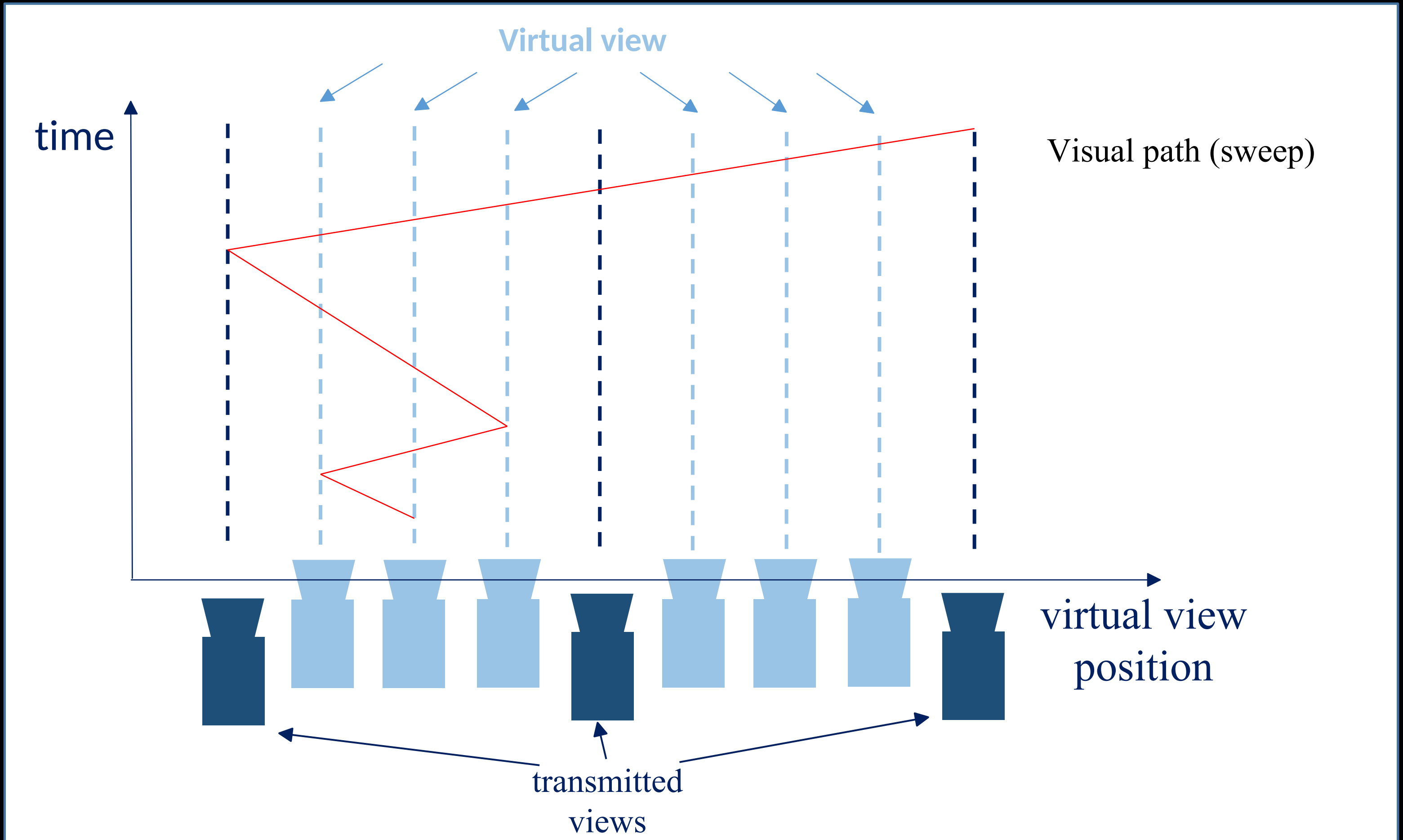
QUALITY ASSESSMENT IN THE CONTEXT OF FREE VIEW POINT NAVIGATION

Patrick Le Callet
Université de Nantes

Free viewpoint television: framework that allows viewing of a 3D world by freely changing the viewpoint

Free viewpoint video: can be understood as the functionality to freely navigate within real world visual scenes

Depth-Image-Based Rendering: process of synthesizing virtual views of the scene from captured color images or videos with associated depth information



Quality metrics for assessing the performances of multi-view videos compression, depth map compression and view synthesis techniques ?

-

New artifacts: Object shifting , Geometric distortion, …

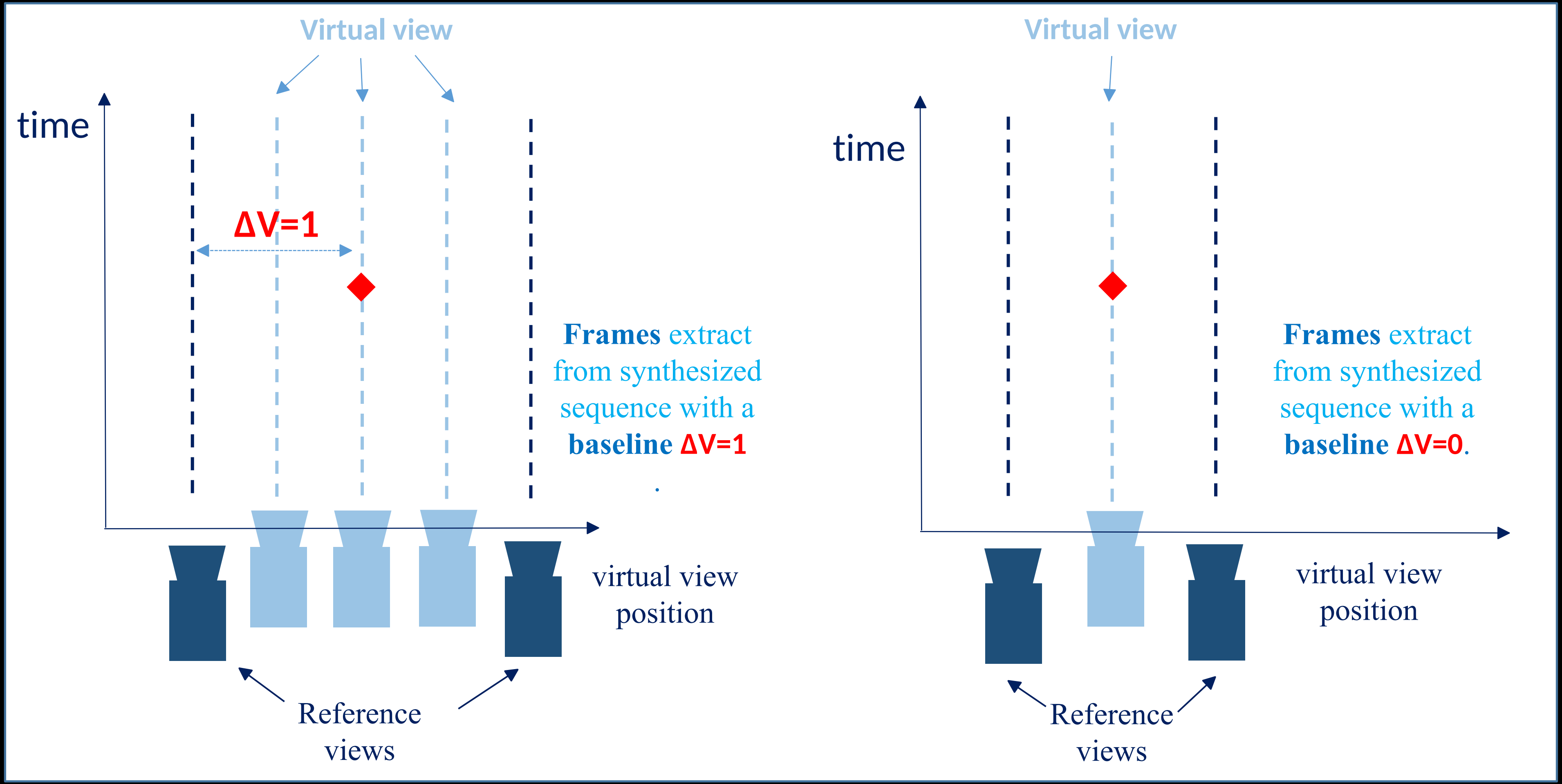
Non uniformly distributed: artifacts mainly locate around disoccluded areas.

DIBR Images database (synthesis only)

3 MPD videos (1024x768) x 7 DIBR algorithms x 4 new viewpoints -> 96 videos -> 96 images

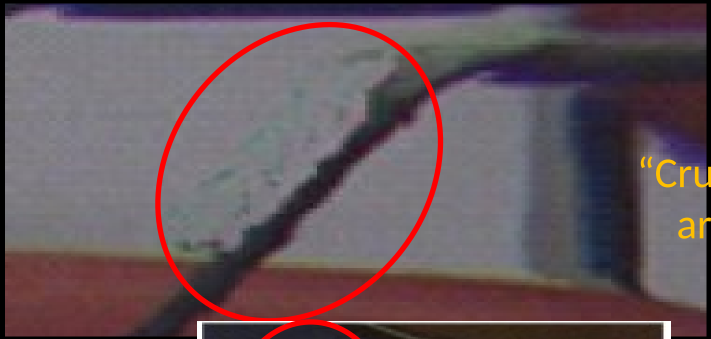
Spatial artifacts

- Different DIBR algorithms
- Different baselines



E. Bosc et al. "Towards a new quality metric for 3-D synthesized view assessment." *IEEE STSP 2011*

Performance of usual quality metrics



“Crumbling” artifact

Shifting artifact

Metric	All conten
	PCC
PSNR	0.2671
SSIM	0.0000*
MS-SSIM	0.0105
VIF	0.0584
VIFP	0.0798
UQI	0.0000*
IFC	0.1289

Usual Image quality metrics:

=> Cannot fully capture global consistent stretched/bent shapes or local geometric distortion

Beyond usual quality metrics

Metric	Name
VSQA [1]	Objective view synthesis quality assessment
3DswIM[2]	Objective image quality assessment of 3d synthesized views
MW-PSNR[3]	Morphological Wavelet Peak Signalto-Noise Ratio metric
ST-SIQA[4]	**Sketch-Token based synthesized image quality assessment**

Common ground: quantify the change of contours as a proxy for a semantic level annoyance

[1]Conze et al.. "Objective view synthesis quality assessment." Electronic Imaging 2012.
[2]Battisti er al. "Objective image quality assessment of 3D synthesized views." Signal Processing: Image Communication 2015
[3]Sandić-Stanković, et al.. DIBR synthesized image quality assessment based on morphological wavelets. Qomex 2015
[4]Ling, and al. " Image quality assessment for free viewpoint video based on mid-level contours", ICME 2017

Beyond usual quality metrics

Metric	PCC	RMSE
VSQA	0.61	0.49
3DswIM	0.69	0.48
MP-PSNR	0.67	0.49
ST-SIAQ	**0.82**	**0.39**

Common ground: quantify the change of contours as a proxy for a semantic level annoyance

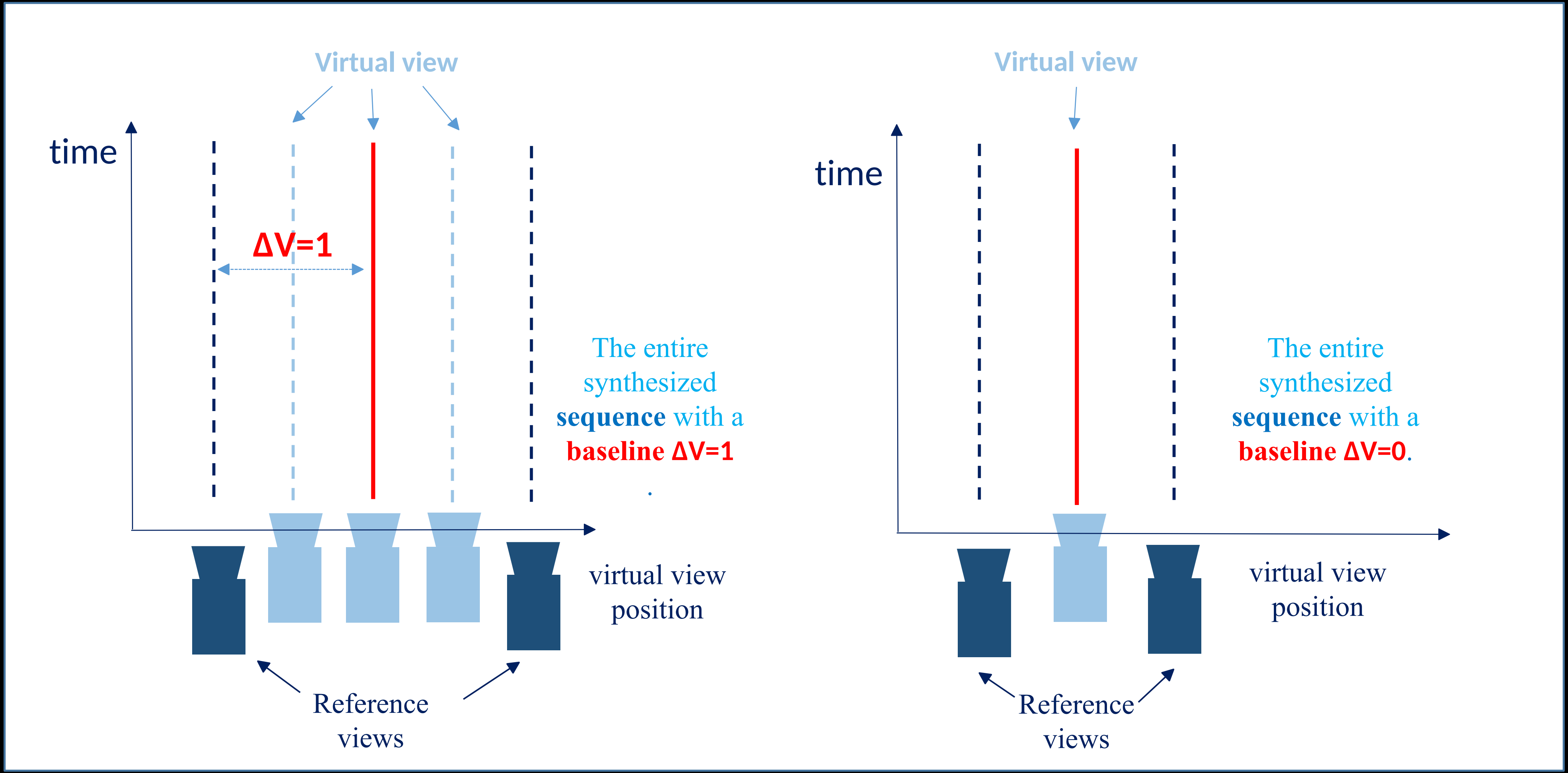
[1]Conze et al.. "Objective view synthesis quality assessment." Electronic Imaging 2012.
[2]Battisti er al. "Objective image quality assessment of 3D synthesized views." Signal Processing: Image Communication 2015
[3]Sandić-Stanković, et al.. DIBR synthesized image quality assessment based on morphological wavelets. Qomex 2015
[4]Ling, and al. " Image quality assessment for free viewpoint video based on mid-level contours", ICME 2017

DIBR Videos database synthesis + compression

3 MPD videos (1024x768) x 7 DIBR algorithms x 4 new viewpoints + 3 bitrates -> 102 videos

Spatial + Temporal artifacts

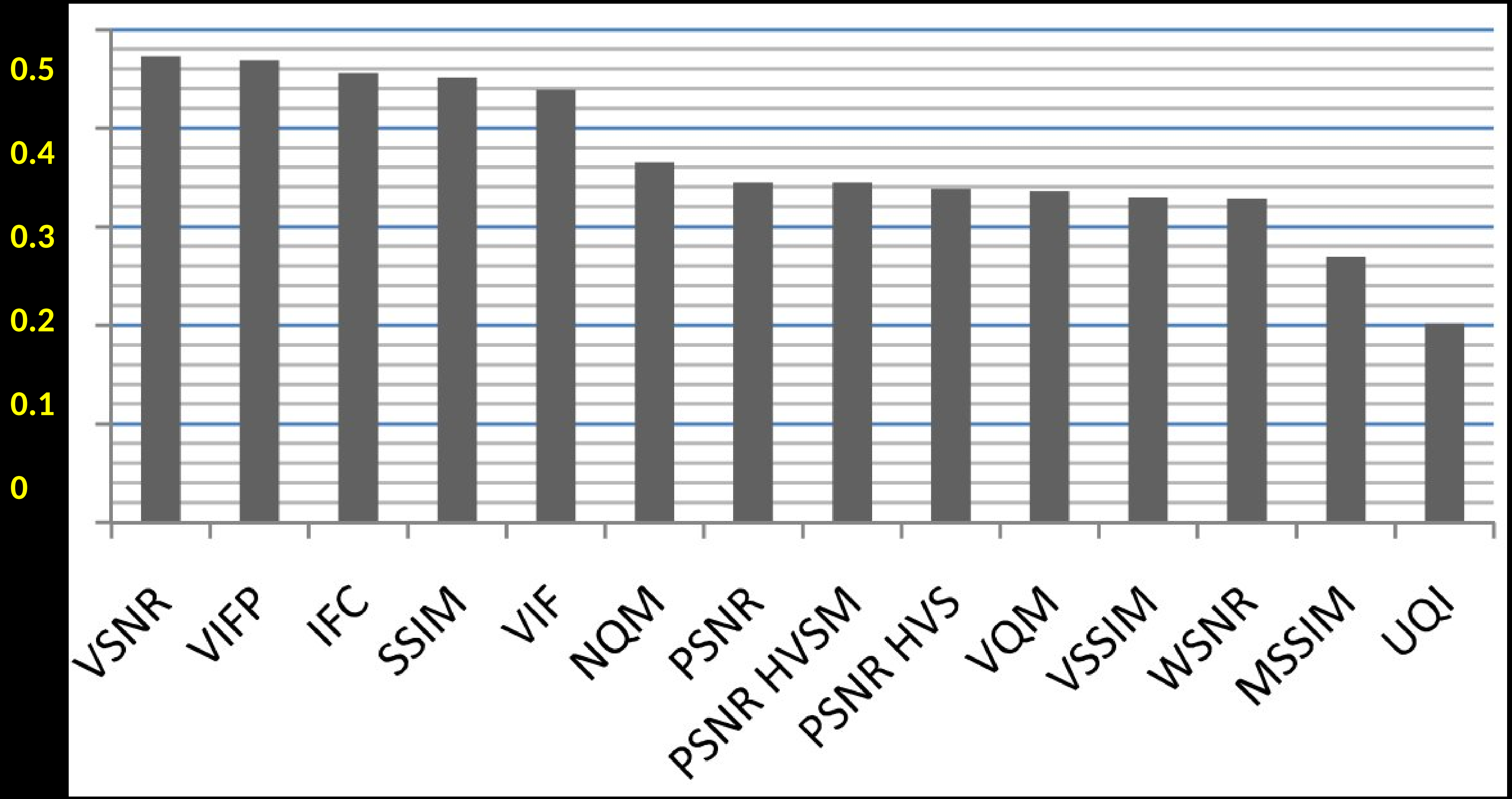
- Different DIBR algorithms
- Different baseline



E. Bosc et al. "Visual quality assessment of synthesized views in the context of 3D-TV." In *3D-TV system with depth-image-based rendering*, 2013.

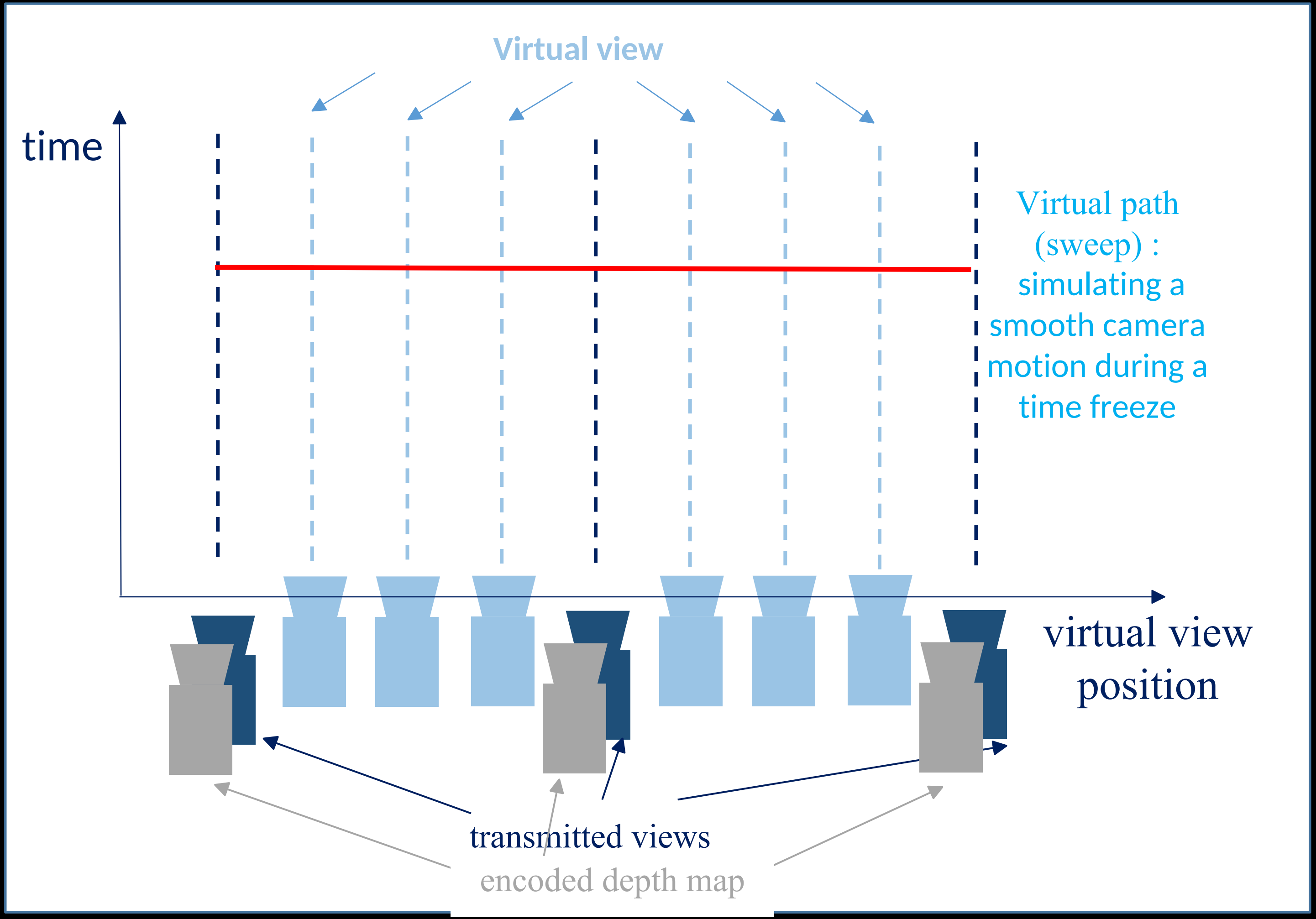

HHI

Performance of commonly used metrics



Free-Viewpoint Synthesized Videos database synthesis + compression on depth map

6 MPD videos (1024x768 / 1920x1080) x 7 depth coding algorithms x 3 bitrates -> 264 videos



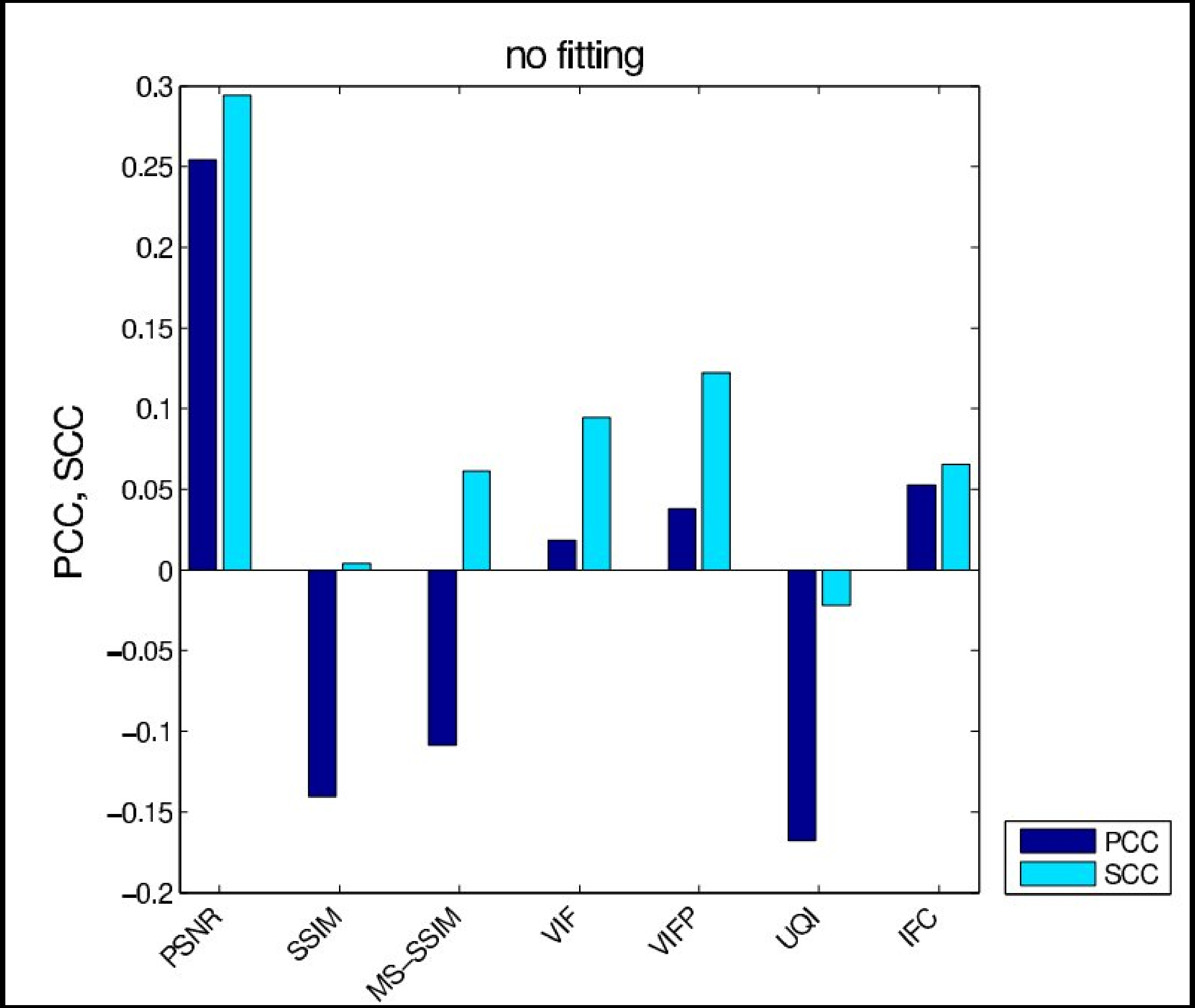
View navigation + Compression on Depth map

- View-sweep effect
- Effect compression on depth map

codecs	HRC descriptions
C1	3D-HEVC Test Model, 3D-HTM 0.4
C2	Multiview Video Coding (MVC), JM 18.4.
C3	HEVC Test Model, HM 6.1.
C4	JPEG2000, Kakadu implementation.
C5	A lossless-edge depth map coding
C6	Correlation is exploited with color frames.
C7	Z-LAR-RP, a region-based algorithm.
Original	Use the real depth maps without any degradation

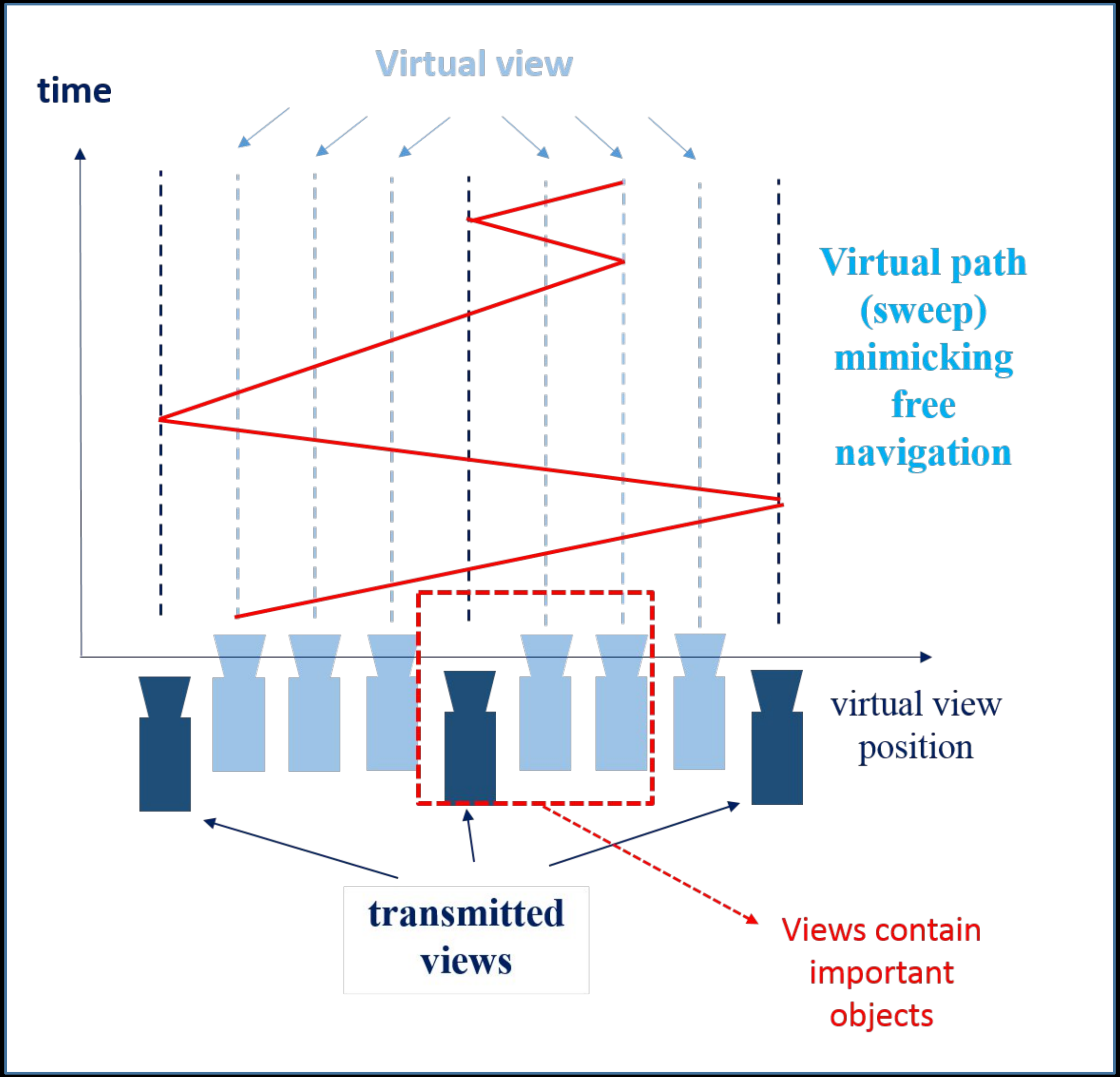
E. Bosc, "A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data." *QoMEX 2013.*

Performance of commonly used metrics



P. Hanhart , Emilie Bosc, Patrick Le Callet, and Touradj Ebrahimi. "Free-viewpoint video sequences: A new challenge for objective quality metrics." MMSP 2014.

Free navigation database



Factors :

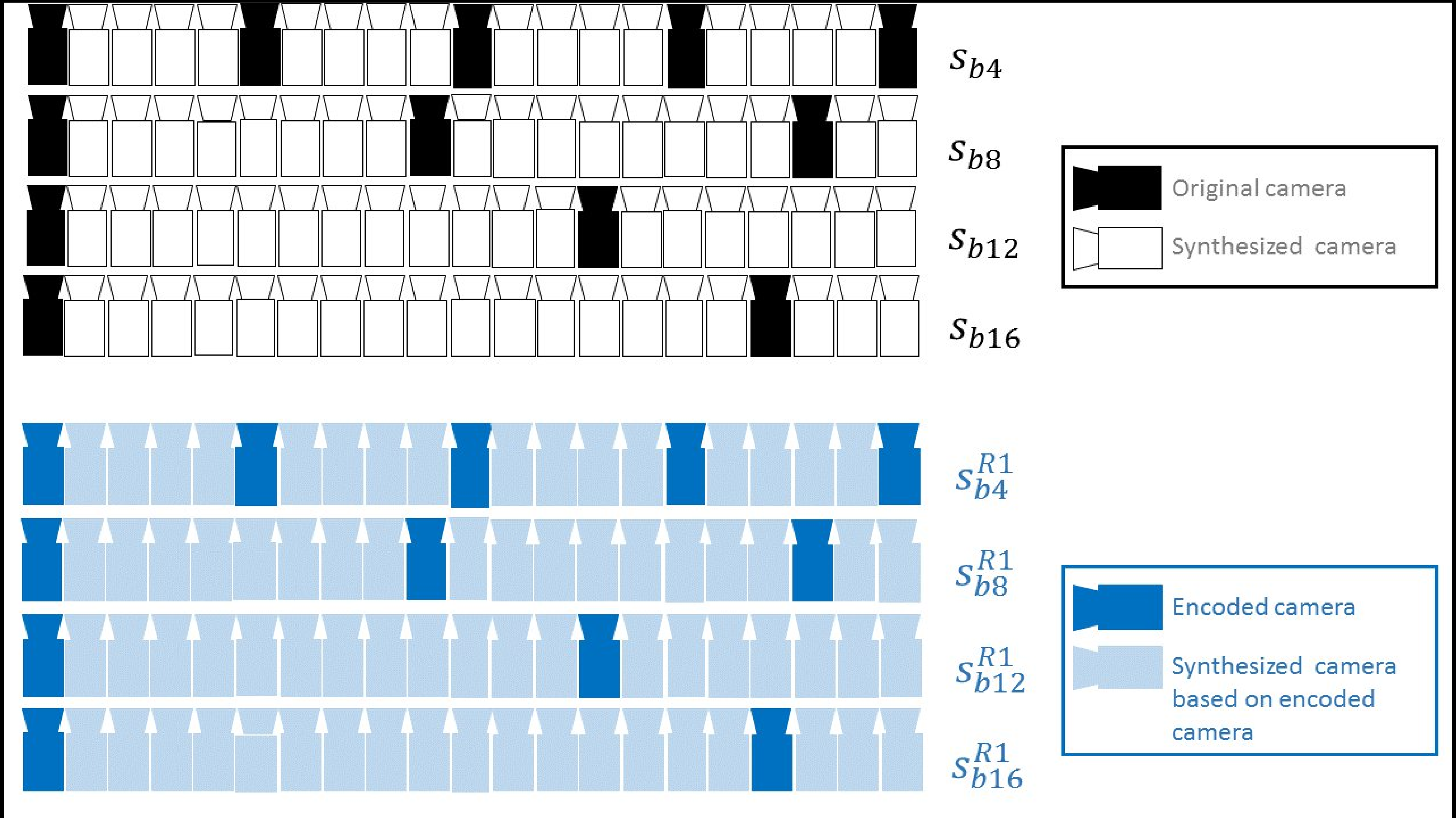
1. Rate-points (RP, bitrate)
2. Baseline distance (V)
3. Virtual path/Sweep/Trajectory (T)

Purpose:

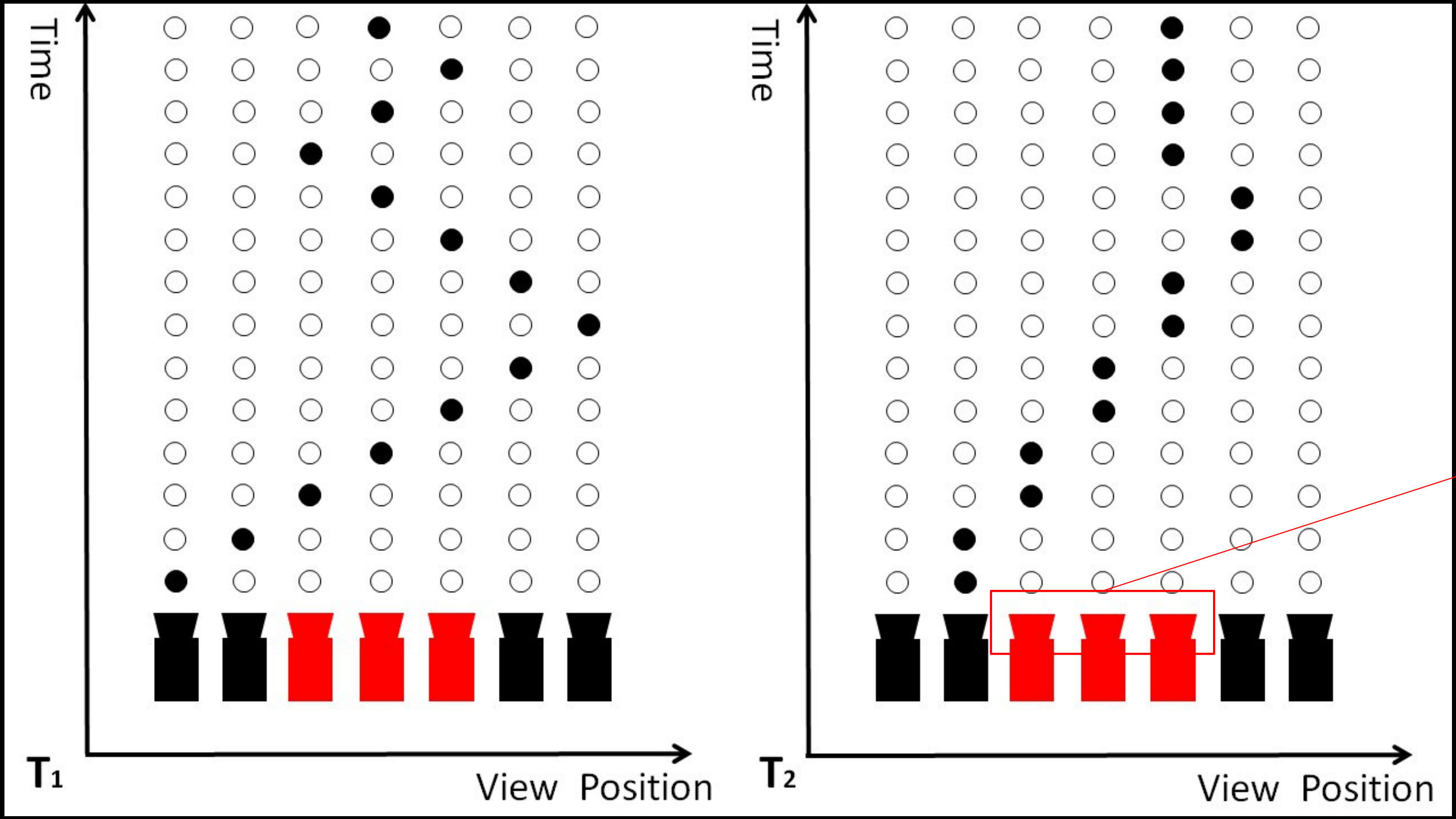
Explore whether how observer navigate among views affect how they judge the quality of the sequence

Stress the system with most challenging configurations (RP, V, T)

Camera arrangements



Generating trajectories



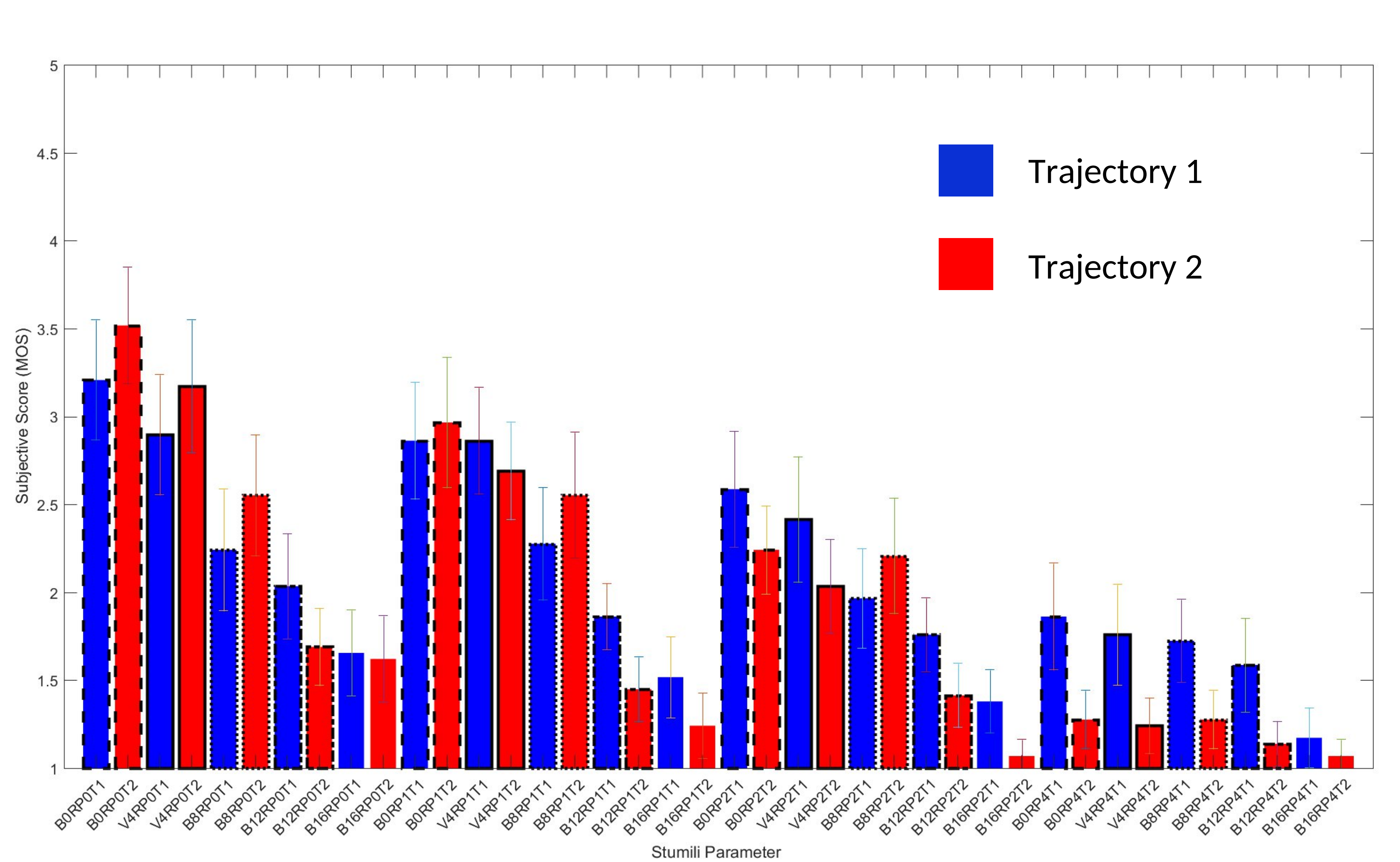
Viewpoints that contain moving objects

Red cameras indicate views contain important objects

T1 (left): Sweeps were constructed at a speed of one frame per view (as what is done in MPEG)

T2 (right): Sweeps were constructed at a speed of two frames per view.

Subjective score of sequence ‘Champagne’



higher quality :
MOS T2 > MOS T1

lower quality :
MOS T1 > MOS T2

Does Trajectories matter ?

three-way analysis of variance (ANOVA):
Influence of Baseline (B), Rate-Points (RP) and virtual Trajectorie (T)

=> The three considered factors have significant impact on the perceived quality ($p=0$ for B and RP, and $p=0.038$ for T).

Ling et al. «Influence of Navigation Scan-path on Perceived Quality of Free-Viewpoint Videos» , IEEE JETCAS 2019

Trajectories matters: proposal for subjective test design

Open questions:

- **Benchmarking codec : which trajectory?**
  - **How to reproduce/control interactivity?**
- **How to identify critical trajectories ?**

Concept of HRT: Hypothetical Rendering Trajectories
Subj. test: SRC, HRC and HRT

Ling et al. «Influence of Navigation Scan-path on Perceived Quality of Free-Viewpoint Videos» , IEEE JETCAS 2019

The lack of proper metrics for assessing the performances of multi-view videos compression, depth map compression and view synthesis techniques

.

Can we do better? => representing and tracking the geometric distortions

White box approach (& Full reference)

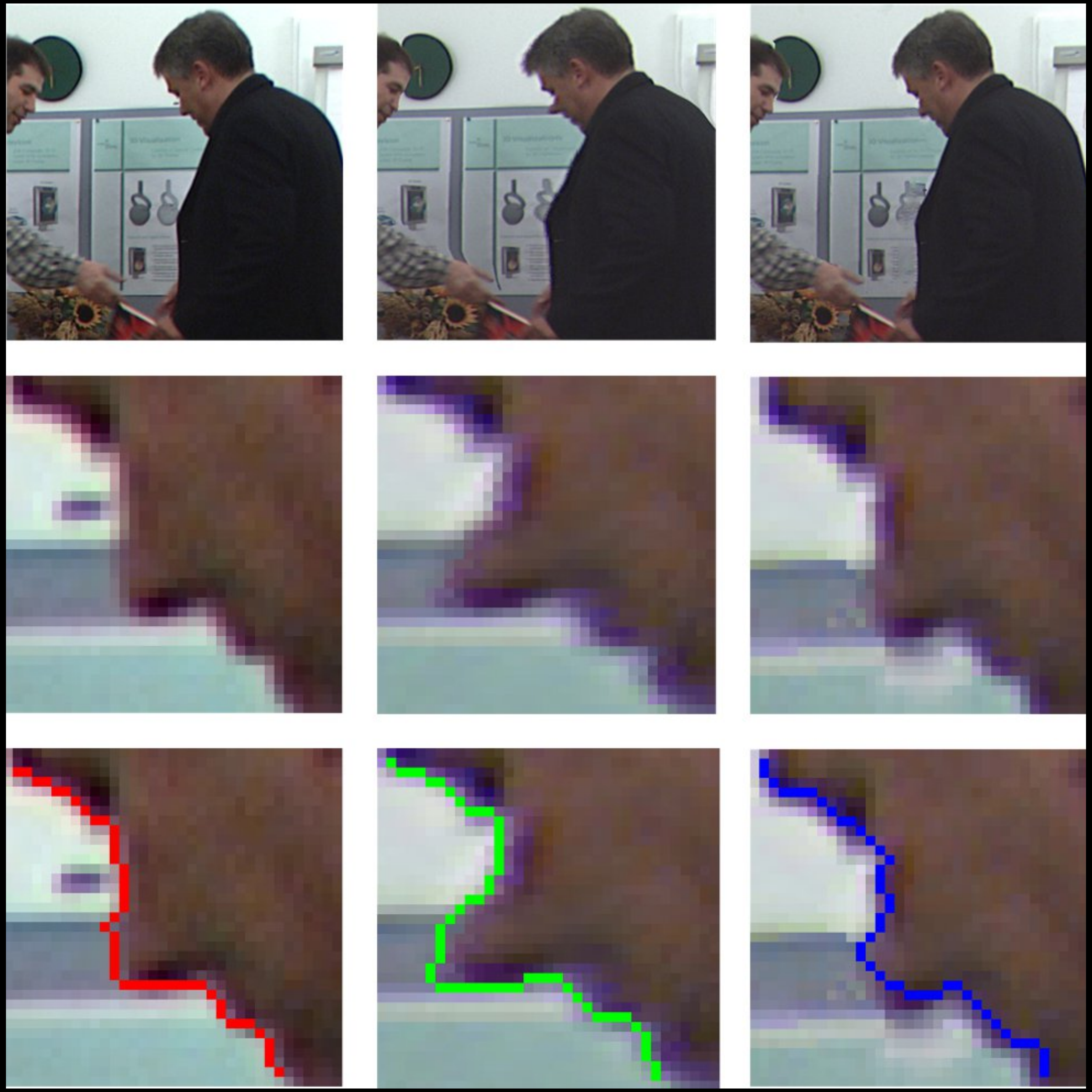
S. Ling, P. Le Callet Image quality assessment for DIBR synthesized views using elastic metric. *ACM MM2017*

Machine Learning approach (& Full reference /Non reference)

S. Ling, P. Le Callet. Image quality assessment for free viewpoint video based on mid-level contours feature. *ICME 2017*

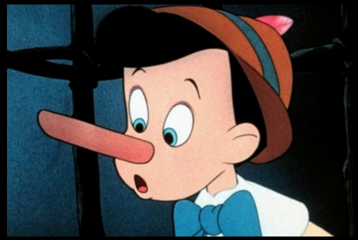
S. Ling, P. Le Callet. How to learn the effect of non-uniform distortion on Perceived Visual Quality ? Case study using Convolutional Sparse Coding for quality assessment of synthesized views. ICIP 2018

Elastic Metric [1,2] : measuring the difference in stretching or bending between two curves



Reference (L) Twisted nose (M) Shifted nose (R)

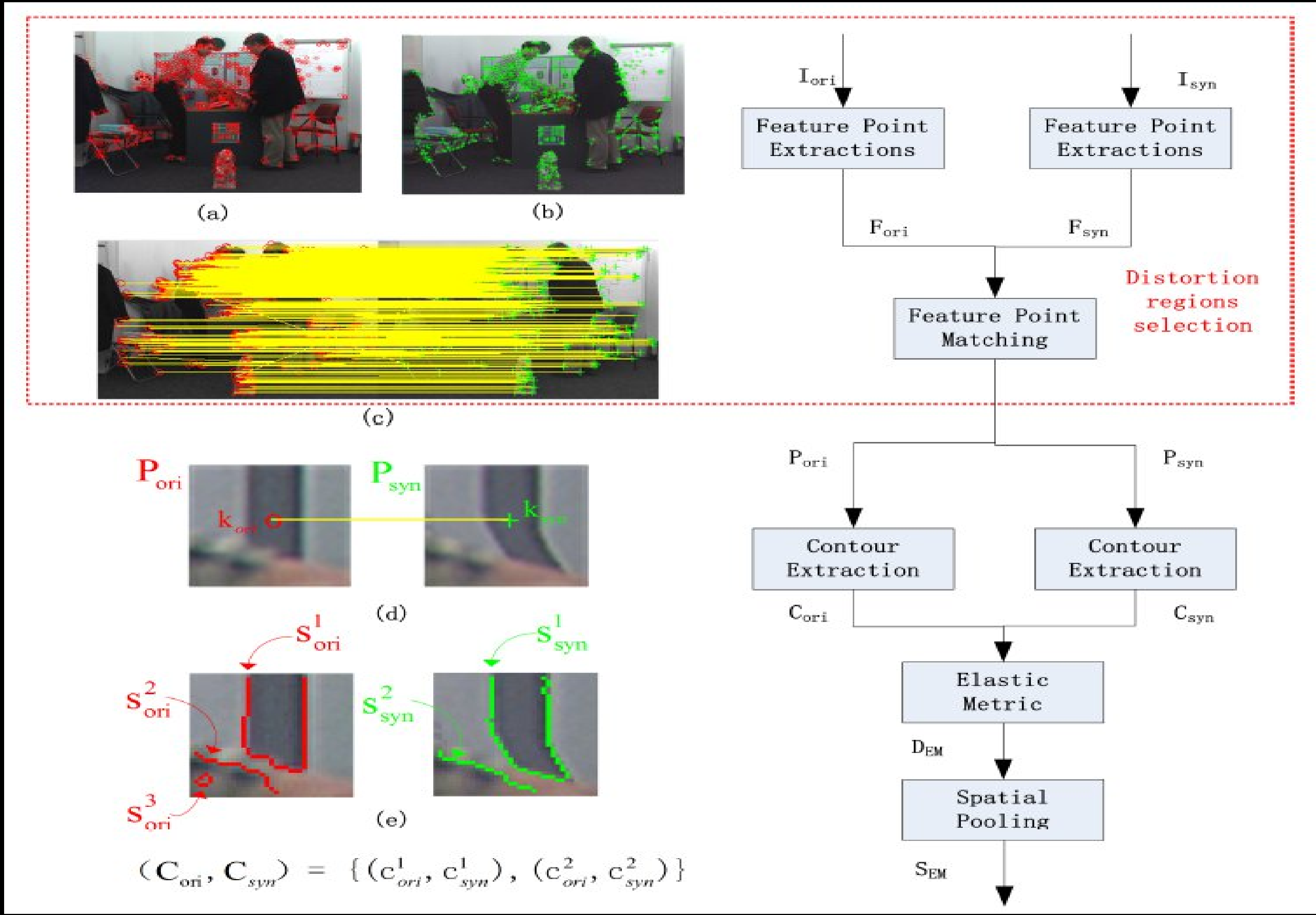
Example:
PSNR(L, M) = 20.2854 db, S_{EM} (L,M) = 0.1926
PSNR(L, R) = 18.6616 db, S_{EM} (L,R) = 0.1781



'What happen to my nose?'

[1]Mio, Washington, Anuj Srivastava, and Shantanu Joshi. "On shape of plane elastic curves." *International Journal of Computer Vision* 73.3 (2007): 307-324.
[2]Srivastava, Anuj, et al. "Shape analysis of elastic curves in euclidean spaces." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.7 (2011): 1415-1428.

Framework of Elastic Metric-IQA



Performance on
video DIBR dataset

Metric	PCC
SoA	0.61
EM-IQA	0.84

X. Liu et al. "Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion." IEEE IP 2015.

Performance on
FVV dataset

Metric	PCC
SoA	0.57
EM-IQA	0.80

S. Ling, P. Le Callet. Image quality assessment for DIBR synthesized views using elastic metric. *ACM MM 2017*

Machine Learning approach?

Algo mostly generate non natural structure …non uniformly

Local & non-uniform

ghosting and structure inconsistency

Global MOS is not a direct proxy for Local "patch" Quality

Natural Scene Statistic (NSS) based models are global & generic (not only structure)

=> Underestimate of local specific structure distortion

Natural Scene Statistics (NSS) based models

Natural Scene Statistic (NSS) based models are global & generic (not only structure)

=> Underestimate of local specific structure distortion

Performance of NSS based models on IRCCyN/IVC DIBR image database [1]

	PCC	SROCC	RMSE
NIQE [2]	0.4022	0.3673	0.6096
BIQI [3]	0.5273	0.3555	0.5657
BliindSII [4]	0.5331	0.1800	0.5633

[1] E. Bosc et al., "Towards a new quality metric for 3-D synthesized view assessment," IEEE J. Sel. Topics Signal Process, Nov. 2011.
[2] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, Mar. 2013
[3] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett*, May 2010.
[4] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, Aug. 2012.

Database : IRCCyN/IVC DIBR image database[2] , 84 images sythesied with 7 algorthms with MOS

Criteria :

Pearson correlation coefficient (PCC)

Spearmans rank order correlation coefficient(SROCC)

Root mean squared error (RMSE)

Machine Learning approach?

Algo mostly generate non natural structure …non uniformly

Local & non-uniform

ghosting and structure inconsistency

Global MOS is not a direct proxy for Local "patch" Quality

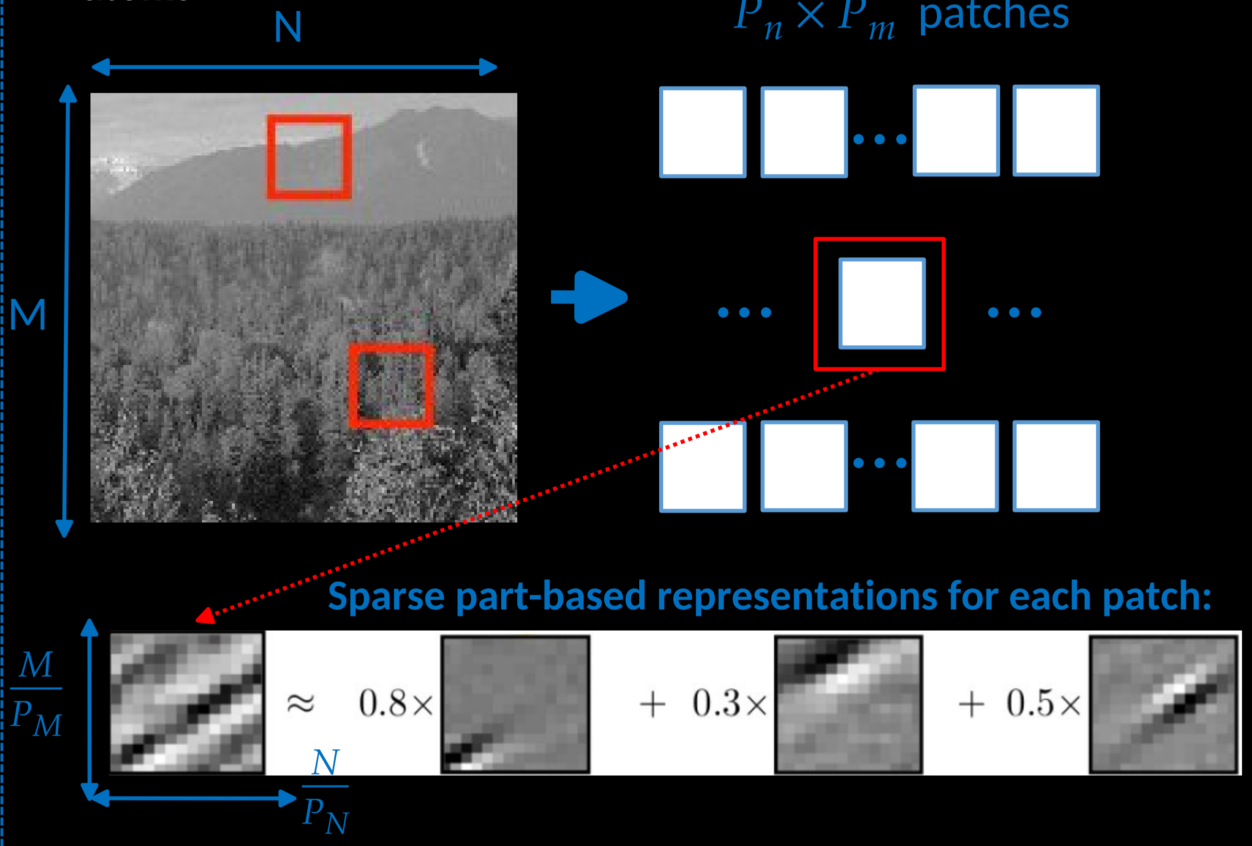
Natural Scene Statistic (NSS) based models are global & generic (not only structure)

=> Underestimate of local specific structure distortion

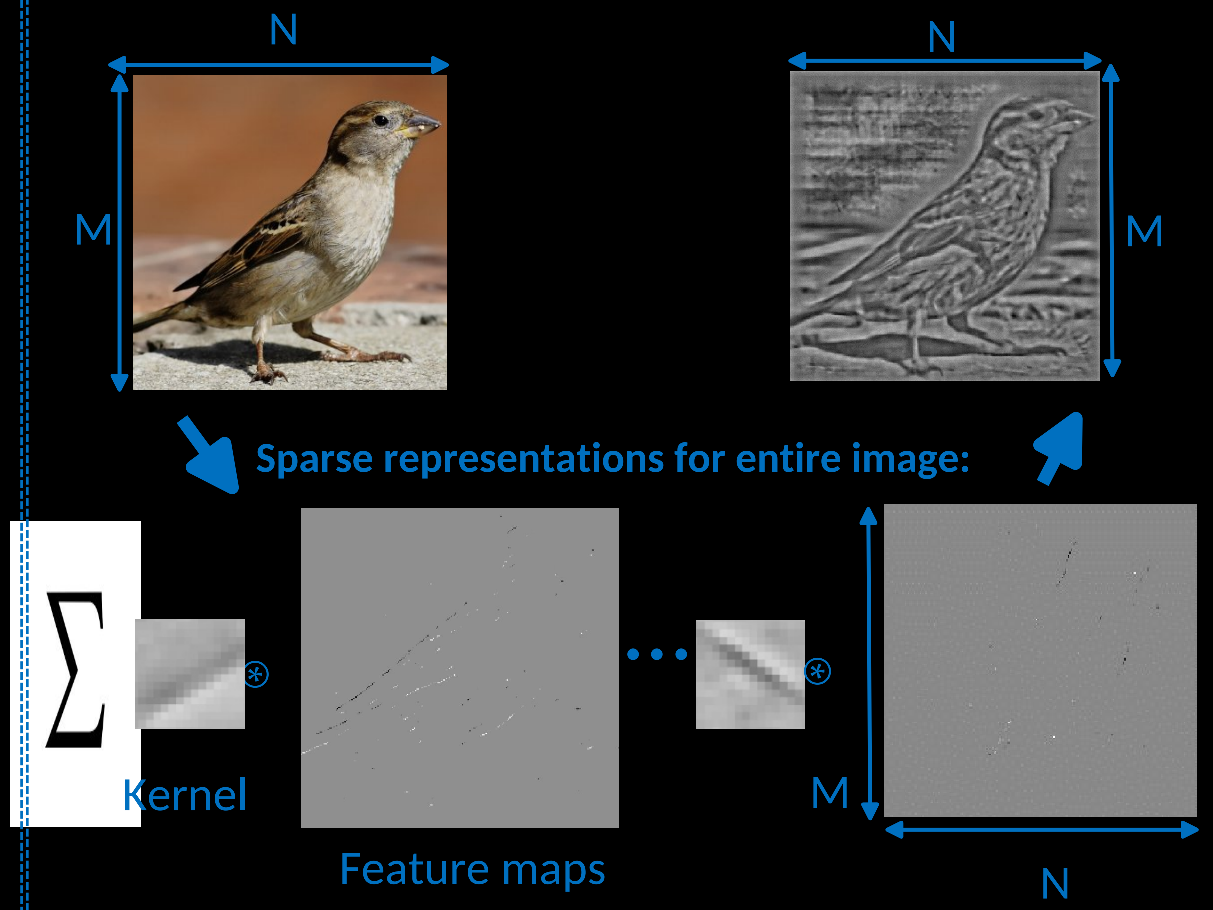
=> Learn structure representation instead of scene statistics

SC & CSC could be a good candidates for learning/coding local NNS

Traditional Sparse Coding:
Represent patches (instead of image) with a set of learned atoms



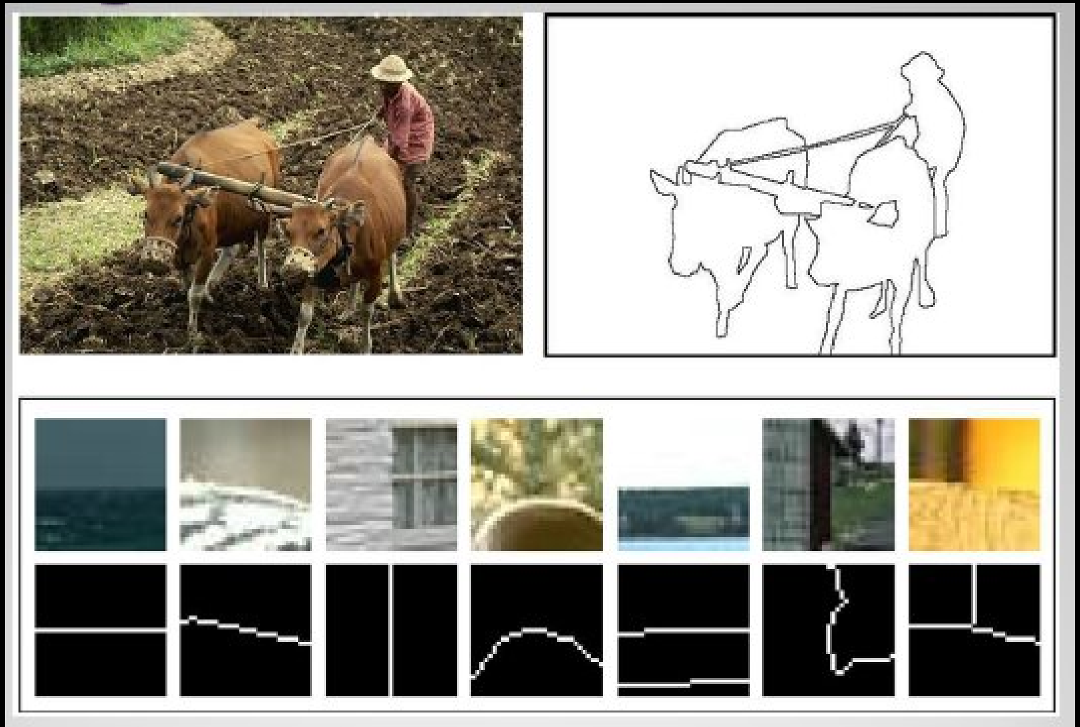
Convolutional Sparse Coding:



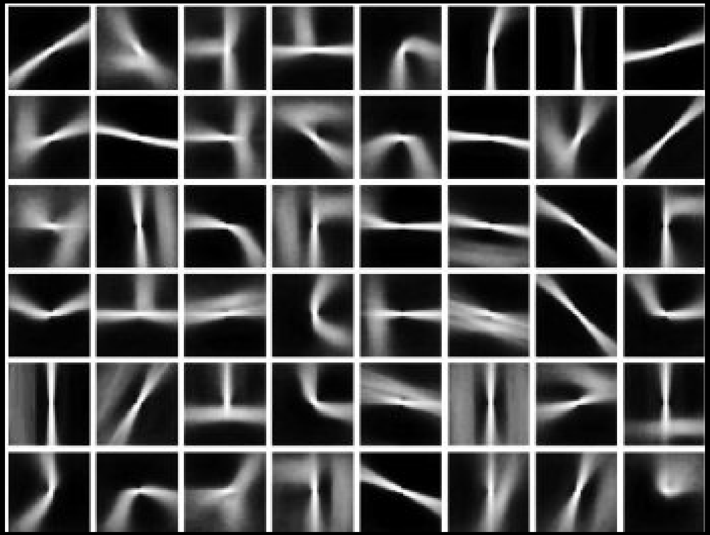
Mairal, Julien, et al. "Online dictionary learning for sparse coding." *ICML*. 2009.

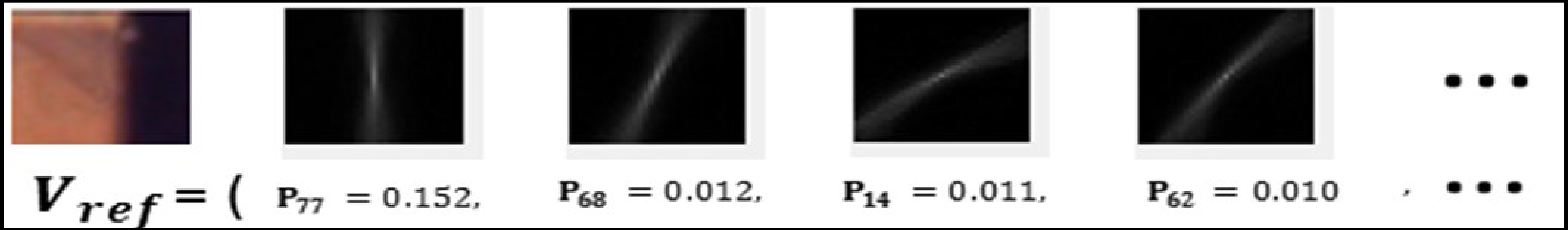
Sketch Token: a learned mid level representation for contour

Sketch token classes: wide variety of local edge structures



Pattern codebook learnt once for all

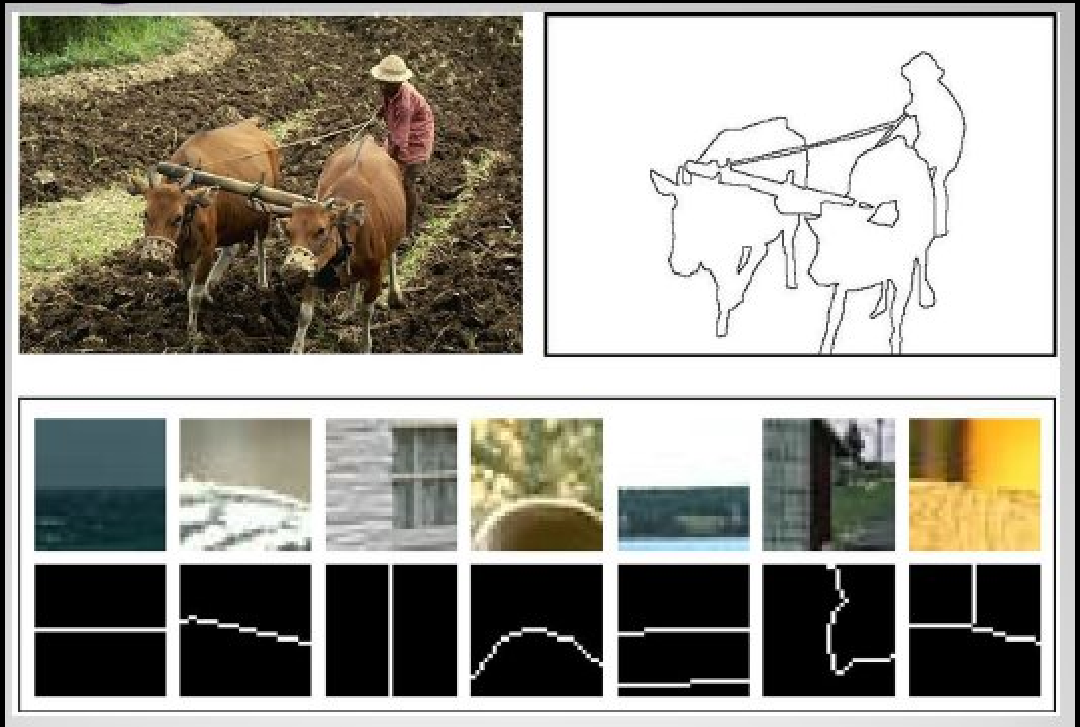




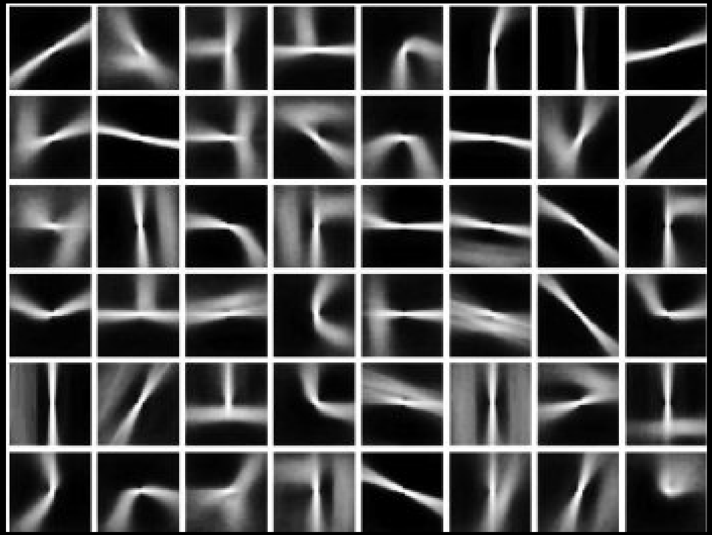
J. Lim and al. "Sketch tokens: A learned mid-level representation for contour and object detection." CVPR 2013

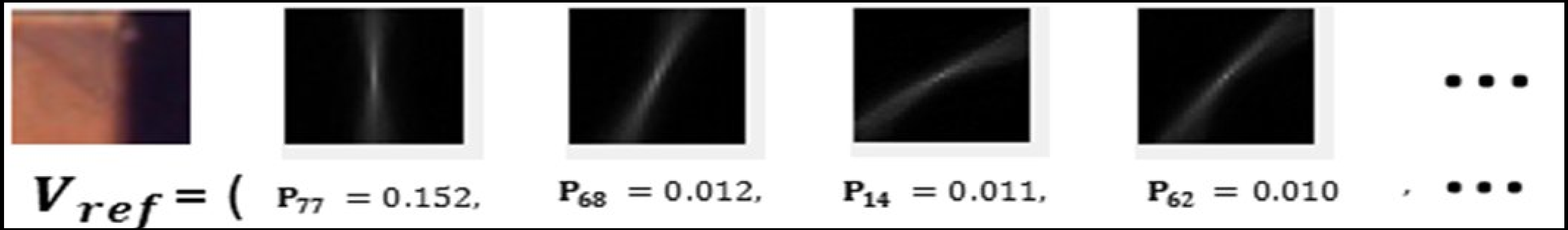
Sketch Token: a learned mid-level representation for contour

Sketch token classes: wide variety of local edge structures



Pattern codebook learnt once for all



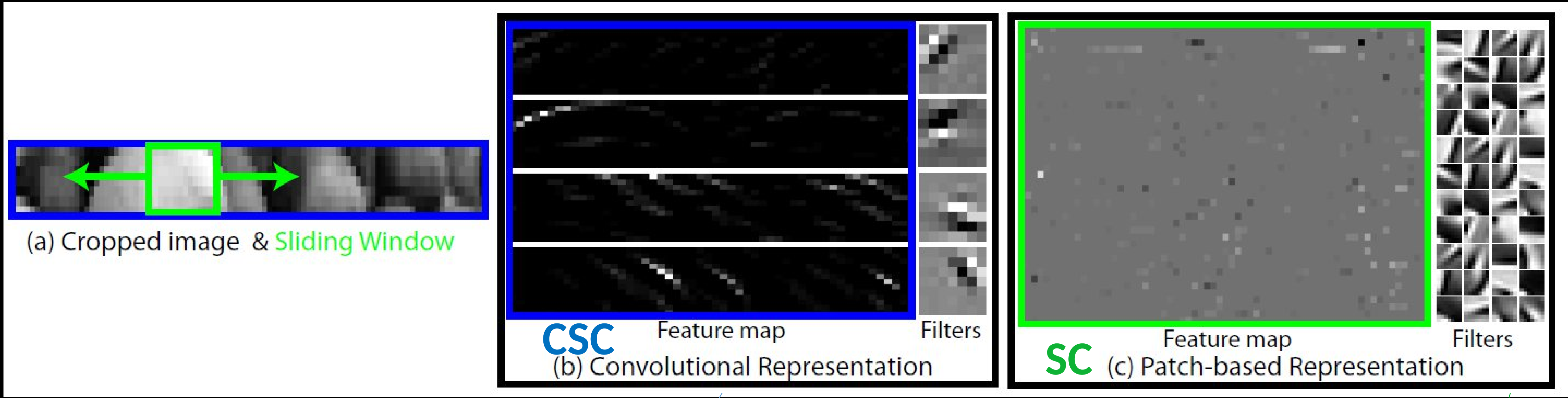


J. Lim and al. "Sketch tokens: A learned mid-level representation for contour and object detection." CVPR 2013

SC vs. CSC

Image from : Zeiler, Matthew D., et al. "Deconvolutional networks." (2010): 2528-2535.

A comparison of convolutional (CSC) and patch-based (SC) sparse representations for a crop from a natural image (a).

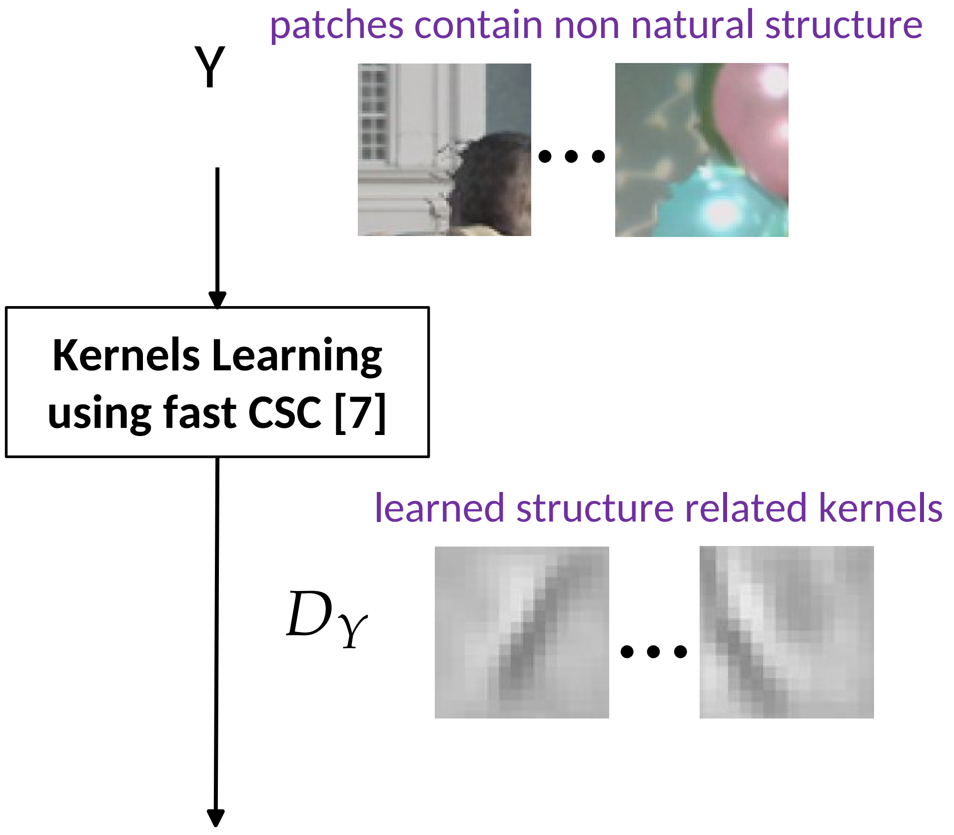


b): Sparse convolutional feature map Z of (a). The smoothly varying feature maps **preserves better spatial locality.**

(c): Patch-based feature map Z of (a) using a sliding window (green). Each column in the feature map corresponds to the sparse vector over the filters for a given x-location of the sliding window.
As the sliding window moves the **latent representation is highly unstable, changing rapidly across edges.**

The Proposed CSC based No Reference Metric (CSC-NRM)

Framework



Y: CSC training set

X: test image

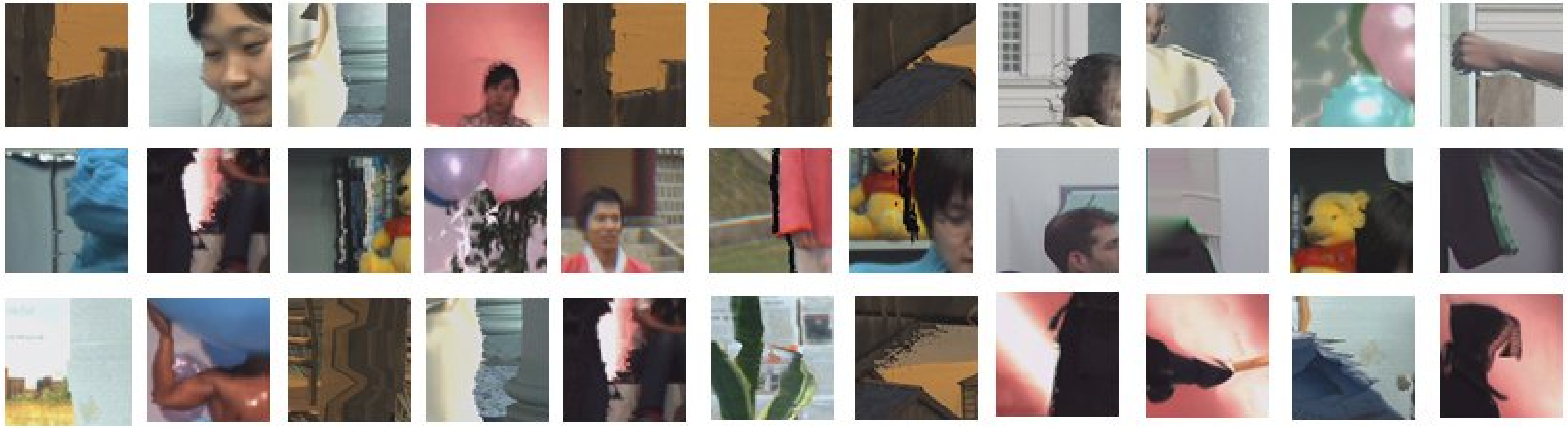
D_Y : Learned Dictionary

v_{csc} : CSC based feature vector

s_{scs}: predicted quality score

Michal Sorel and Filip Sroubek, “Fast convolutional sparse coding using matrix inversion lemma,” Digital Signal Processing, vol. 55, pp. 44–51, 2016.

Collecting Data contain NNS



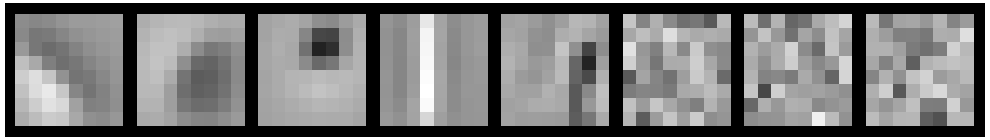
MCL-3D
DIBR-IVC Videos
Free Viewpoint synthesized videos

Rui Song, Hyunsuk Ko, and CC Kuo, “Mcl-3d: A database for stereoscopic image quality assessment using 2d-image-plus-depth source,” arXiv preprint arXiv:1405.1403, 2014.

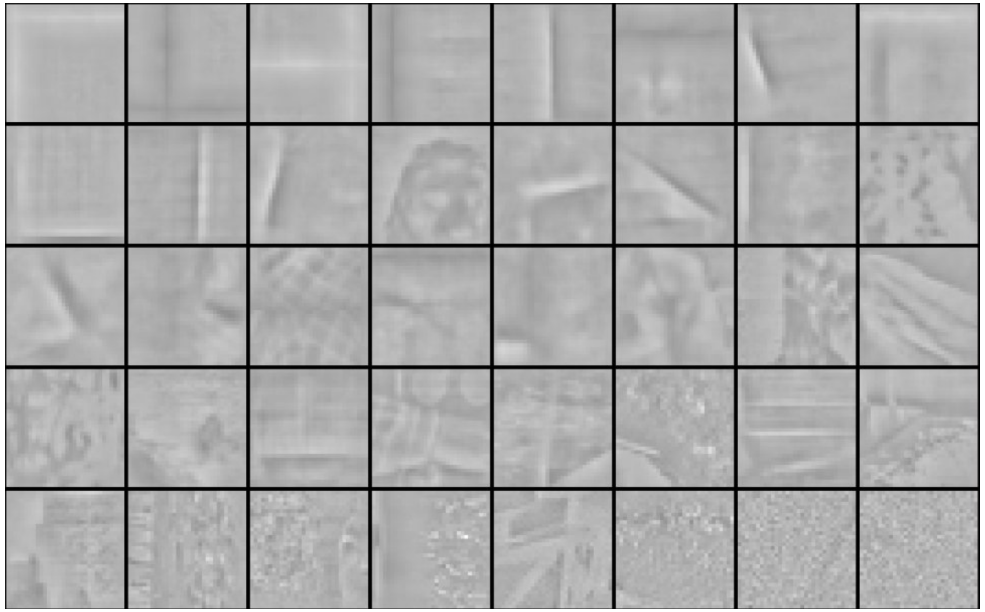
Emilie Bosc, Philippe Hanhart, Patrick Le Callet, and Touradj Ebrahimi, “A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data,” in Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on. IEEE, 2013.

[Emilie Bosc, Philippe Hanhart, Patrick Le Callet, and Touradj Ebrahimi, “A quality assessment protocol for free-viewpoint video sequences synthesized from decompressed depth data, QoMEX .

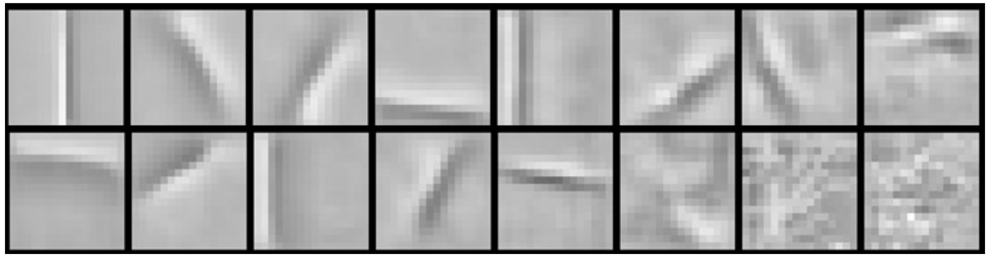
Our 88 Kernels



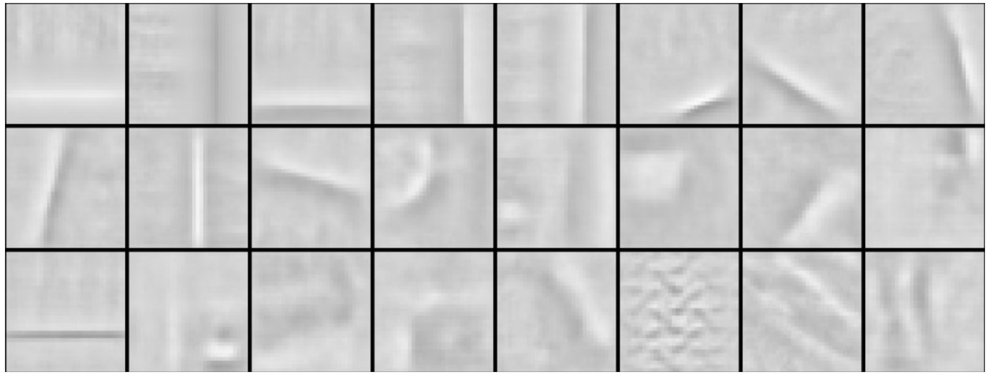
(a) 8 kernels of size 8×8



(b) 16 kernels of size 32 × 32 of lower energy



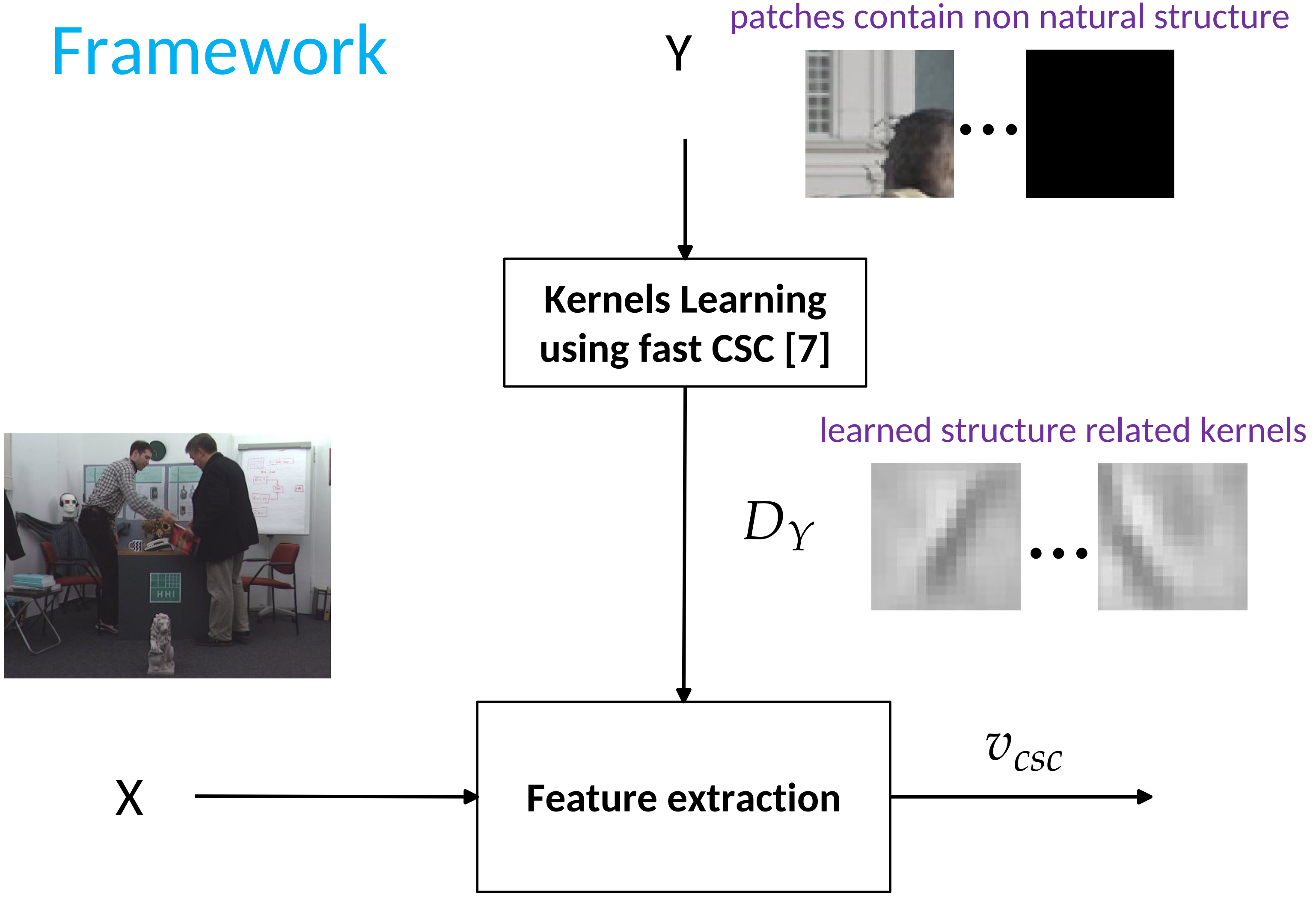
(c) 16 kernels of size 16×16



(d) 16 kernels of size 32 × 32 of higher energy

The Proposed CSC based No Reference Metric (CSC-NRM)

Framework

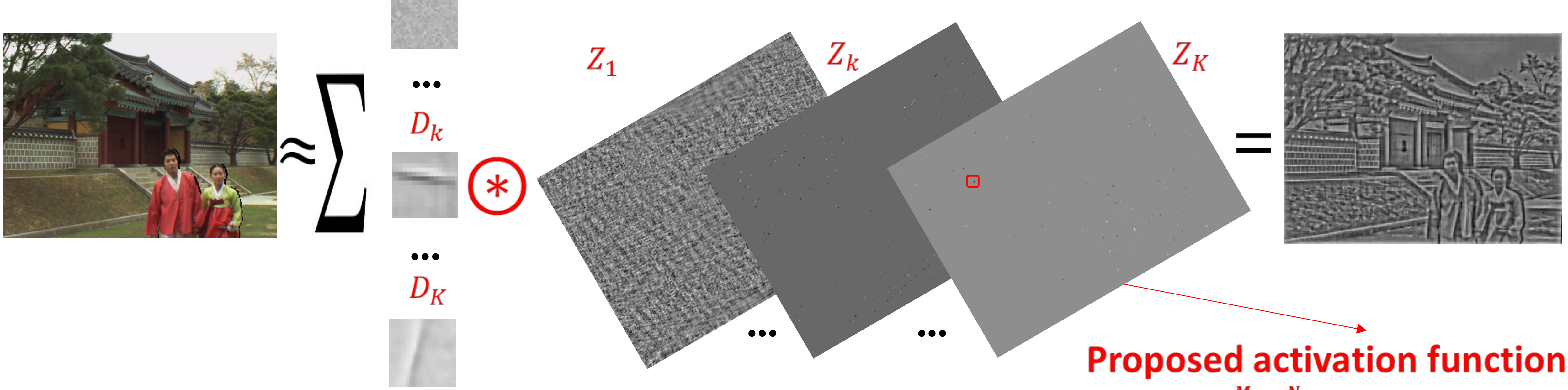


Y: CSC training set
X: test image
D_Y : Learned Dictionary
v_{csc} : CSC based feature vector
s_{scs}: predicted quality score

Activation Function

CSC based Feature Extraction:

For a $M \times N$ test image I, its sparse representation will be $Z_I = [Z_1, .., Z_K]$

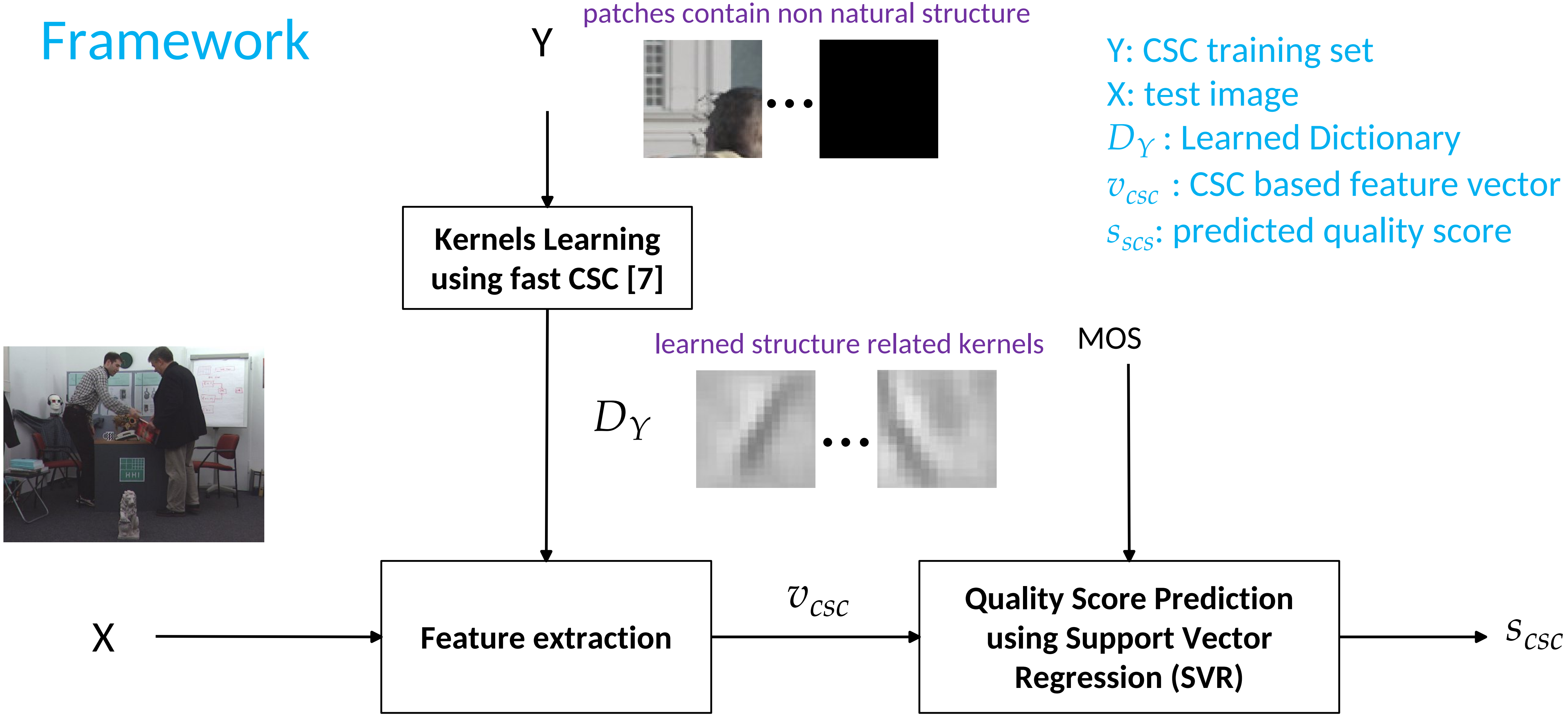


Proposed activation function:

$$f_{act}(Z_k) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} \mathbf{1}\,(Z_k(i,j)>\varepsilon)}{M \times N}$$

$$v_{CSC} = (f_1, .., f_K) = (f_{act}(Z_1) \quad ,...., \quad f_{act}(Z_k) \quad ,...., \quad f_{act}(Z_K))$$

The Proposed CSC based No Reference Metric (CSC-NRM)



Performance of CSC-NRM

Performance of existing NSS based metrics and the proposed metric on IRCCyN/IVC DIBR image dataset

No Reference Metrics	PCC	SROCC	RMSE
Natural Scene Statistics (NSS) based models			
NIQE (NSS) [1]	0.4022	0.3673	0.6096
BIQI (NSS) [2]	0.5273	0.3555	0.5657
BliindSII (NSS) [3]	0.5331	0.1800	0.5633
Non Natural Structure based Model			
CSC-NRM	**0.8302**	**0.7827**	**0.3233**

[1] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, Mar. 2013

[2] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett*, May 2010.

[3] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE TIP.*, Aug. 2012.

[4] E. Bosc et al., "Towards a new quality metric for 3-D synthesized view assessment," IEEE J. Sel. Topics Signal Process, Nov. 2011.

Performance of CSC-NRM (comparison with DIBR specialized NR Metric

	PCC	SROCC	RMSE
Non Learning based No Reference Metrics			
NIQSV [1]	0.6346	0.5146	0.6167
NIQSV+ [2]	0.7114	0.6668	0.4679
APT [3]	0.7307	0.7140	0.4622
Learning based No Reference Metrics			
CSC-NRM	**0.8302**	**0.7827**	**0.3233**

[1] Shishun Tian, Lu Zhang, Luce Morin, and Olivier Deforges, "Niqsv: A no reference image quality assessment metric for 3d synthesized views," in Acoustics, Speech and Signal Processing (ICASSP), 2017.

[2] Tian, Shishun, et al. "NIQSV+: A No-Reference Synthesized View Quality Assessment Metric." IEEE Transactions on Image Processing 27.4 (2018): 1652-1664.

[3] Ke Gu, Vinit Jakhetiya, Jun-Fei Qiao, Xiaoli Li, Weisi Lin, and Daniel Thalmann, "Model-based referenceless quality metric of 3d synthesized images using local image description," IEEE TIP, 2017.

[4] E. Bosc et al., "Towards a new quality metric for 3-D synthesized view assessment," IEEE J. Sel. Topics Signal Process, Nov. 2011.

CONVOLUTION SPARSE CODING WORKS ALSO FOR STICHTING / SEAM ARTEFACTS IN 360

Stitching Algo. Generate also Non Natural Structure ...non uniformly distributed

manually selected patches with stitched/seam artefacts

Metric	PCC
Solh (FR)	0.95
BLIINDS(NR)	0.11
DIVINE (NR)	0.25
Conv. Sparse Coding (NR)	0,85

training dataset available ftp:\\ftp.ivc.polytech.univ-nantes.fr\LS2N_IPI_Stitched_Patches_Database

S. Ling, G. Cheung, P. Le Callet. No-Reference Quality Assessment for Stitched Panoramic Images Using Convolutional Sparse Coding and Compound Feature Selection. *ICME 2018*