Adaptive Resolution Change using Uncoded Areas and Dictionary Learning-based Super-Resolution in Versatile Video Coding

Lehrstuhl und
Institut für
Nachrichtentechnik

RWTH AACHEN UNIVERSITY

Contents

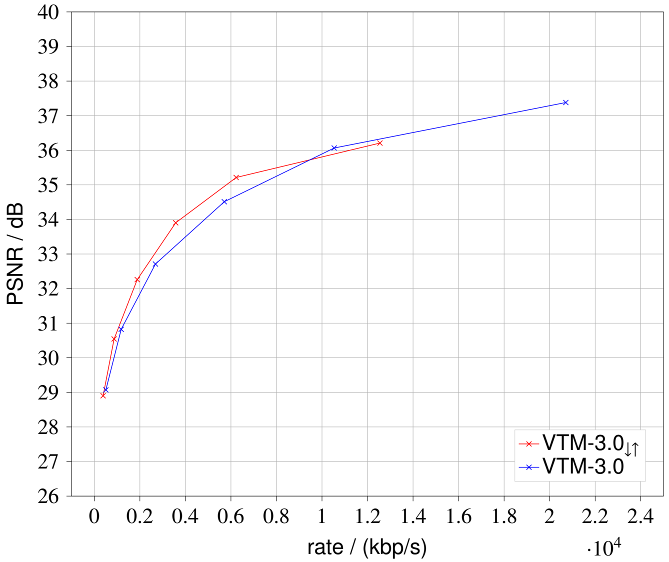
Motivation

- Dictionary Learning-based super-resolution showed promising results when applied to inter-layer prediction in SHVC [1].

Motivation

- Dictionary Learning-based super-resolution showed promising results when applied to inter-layer prediction in SHVC [1].
- The concept of adaptive resolution change is already known from MPEG 4 [2] and raised attention recently [3].

- Dictionary Learning-based super-resolution showed promising results when applied to inter-layer prediction in SHVC [1].
- The concept of adaptive resolution change is already known from MPEG 4 [2] and raised attention recently [3].
- The convex hull of the RD curve can be estimated by downsampling the video before coding [4].



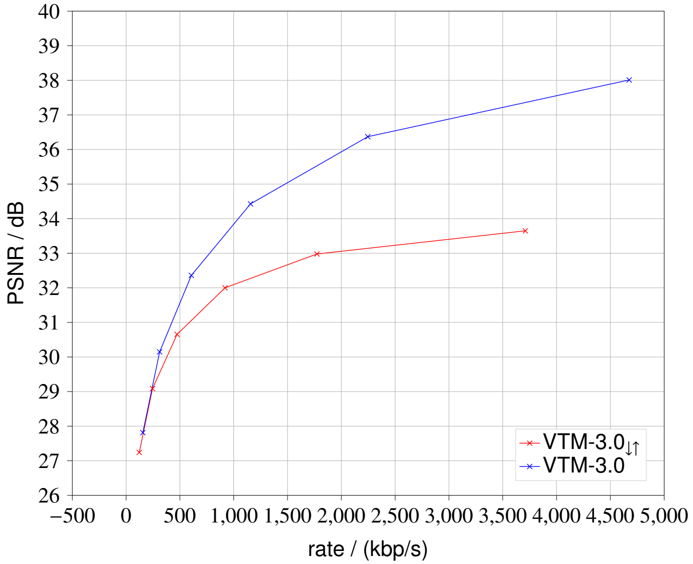


Figure: RD-curves for Campfire sequence (left) and Basketballdrive (right). First 100 frames, RA coding configuration.

Adaptive Resolution Change

- On which level of the encoding scheme should the resolution change happen?

Option	signaling cost	spacial adaptivity	temporal adaptivity	boundary issues
CU level	high	yes	yes	yes
CTU level	moderate	yes	yes	moderate
picture level	low	no	yes	almost none
TID level	low	no	yes	almost none
intra period level	low	no	yes	almost none
sequence level	none	no	no	almost none

Adaptive Resolution Change

- On which level of the encoding scheme should the resolution change happen?

Option	signaling cost	spacial adaptivity	temporal adaptivity	boundary issues
CU level	high	yes	yes	yes
CTU level	moderate	yes	yes	moderate
picture level	low	no	yes	almost none
TID level	low	no	yes	almost none
intra period level	low	no	yes	almost none
sequence level	none	no	no	almost none

➥ SVCP-19 ARC scheme on the CTU level
- tricky regarding the implementation
- limited to intra frames and in terms of coding gains.

Adaptive Resolution Change

- On which level of the encoding scheme should the resolution change happen?

Option	signaling cost	spacial adaptivity	temporal adaptivity	boundary issues
CU level	high	yes	yes	yes
CTU level	moderate	yes	yes	moderate
picture level	low	no	yes	almost none
TID level	low	no	yes	almost none
intra period level	low	no	yes	almost none
sequence level	none	no	no	almost none

➥ SVCP-19 ARC scheme on the CTU level
- tricky regarding the implementation
- limited to intra frames and in terms of coding gains.

➥ SVCP-20/21 ARC scheme on the picture level.
- Code the picture at full and half resolution.
- Upsample or apply SR to downsampled reconstructed pictures.
- Decide based on RD-cost which one is coded into the bitstream.

Contents

Uncoded Areas for ARC

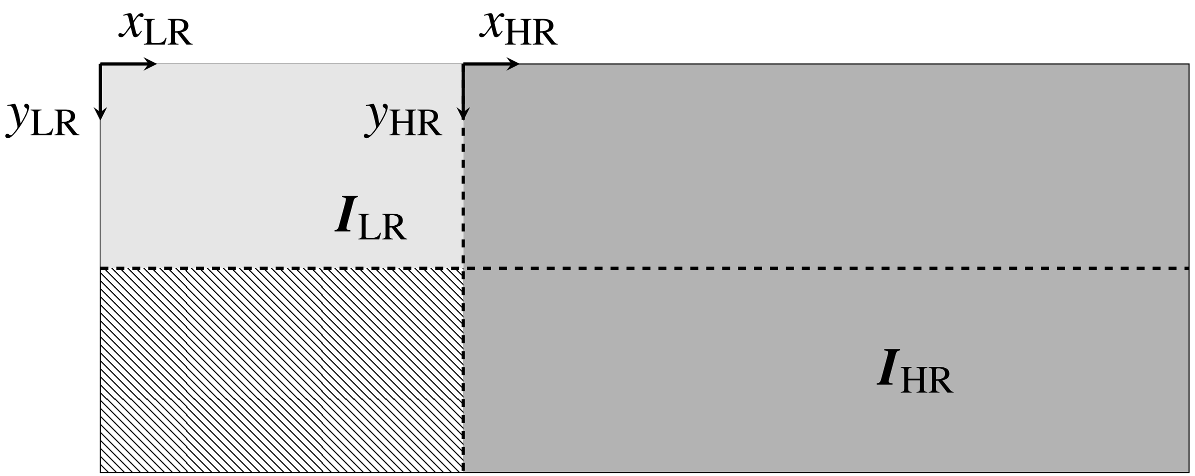


Figure: Uncoded areas for ARC. uncoded, half res. video, full res. video

Uncoded Areas for ARC

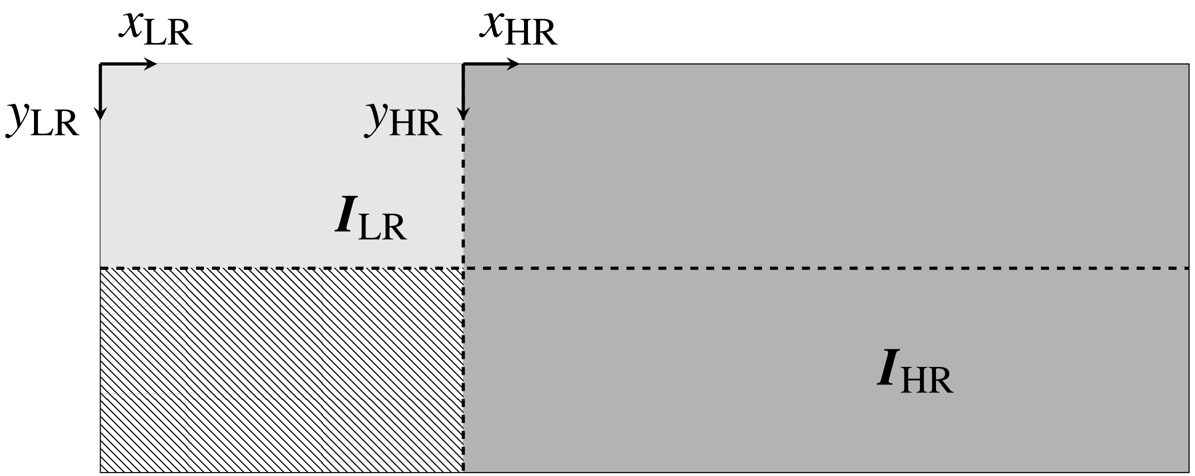


Figure: Uncoded areas for ARC. uncoded, half res. video, full res. video

- Based on the cost function $J = D + \lambda R$ the decision on which area will be coded is made.
- The quantization parameter for the low resolution video is lowered such that $\mathrm{QP}_{\mathrm{LR}} = \mathrm{QP}_{\mathrm{HR}} - 6$ [5].

Uncoded Areas for ARC

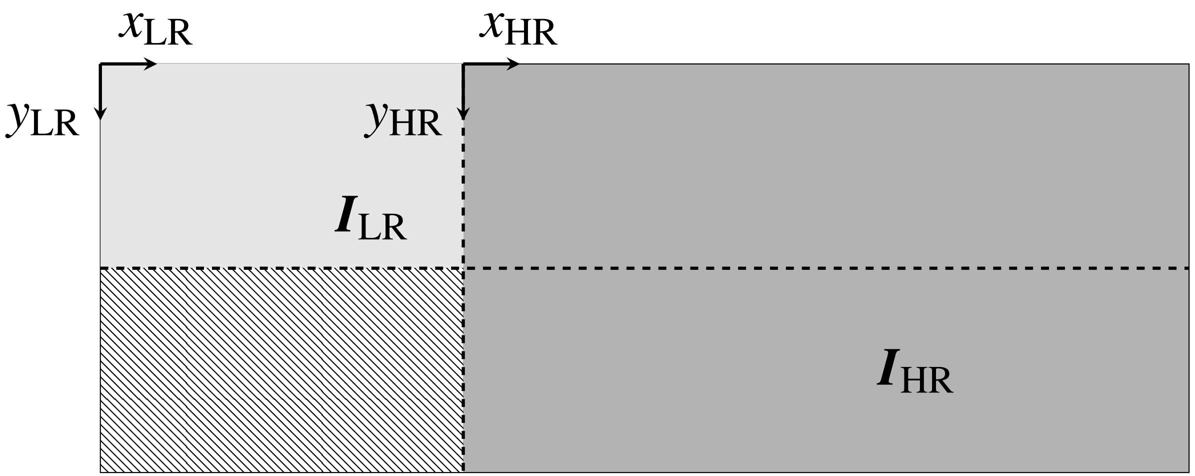


Figure: Uncoded areas for ARC. uncoded, half res. video, full res. video

- Based on the cost function $J = D + \lambda R$ the decision on which area will be coded is made.
- The quantization parameter for the low resolution video is lowered such that $\mathrm{QP}_{\mathrm{LR}} = \mathrm{QP}_{\mathrm{HR}} - 6$ [5].
- Upsampling and downsampling
  – Using the SHVC filters

$$\boldsymbol{h}_{\downarrow} = [2, -3, -9, 6, 39, 58, 39, 6, -9, -3, 2]/128 \qquad \boldsymbol{h}_{\uparrow} = [-1, 0, 4, 0, -11, 0, 40, 64, 40, 0, -11, 0, 4, 0, -1]/64$$

  – or a machine learning-based method?

Contents

Dictionary Learning Fundamentals

- A dictionary is typically trained using vectorized training patches $\boldsymbol{x}_i$ of a size $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$:

$$\boldsymbol{D} \leftarrow \arg\min_{\boldsymbol{D}} \sum_{i=1}^{n} \frac{1}{2} \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$$

Dictionary Learning Fundamentals

- A dictionary is typically trained using vectorized training patches $\boldsymbol{x}_i$ of a size $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$:

$$\boldsymbol{D} \leftarrow \arg\min_{\boldsymbol{D}} \sum_{i=1}^{n} \frac{1}{2} \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$$

- A sparse representation of an image patch is found by sparse encoding the patch $\boldsymbol{x}$ in the dictionary $\boldsymbol{D}$:

$$\boldsymbol{\alpha} \leftarrow \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

$$\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

Dictionary Learning Fundamentals

- A dictionary is typically trained using vectorized training patches $\boldsymbol{x}_i$ of a size $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$:

$$\boldsymbol{D} \leftarrow \arg\min_{\boldsymbol{D}} \sum_{i=1}^{n} \frac{1}{2} \|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$$

- A sparse representation of an image patch is found by sparse encoding the patch $\boldsymbol{x}$ in the dictionary $\boldsymbol{D}$:

$$\boldsymbol{\alpha} \leftarrow \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

$$\boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

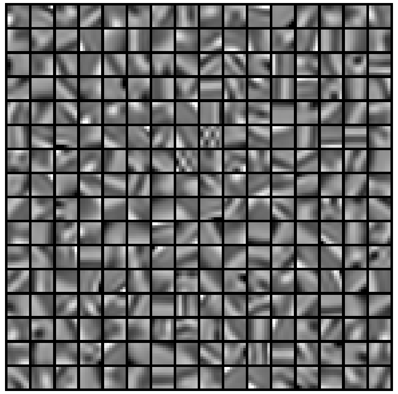
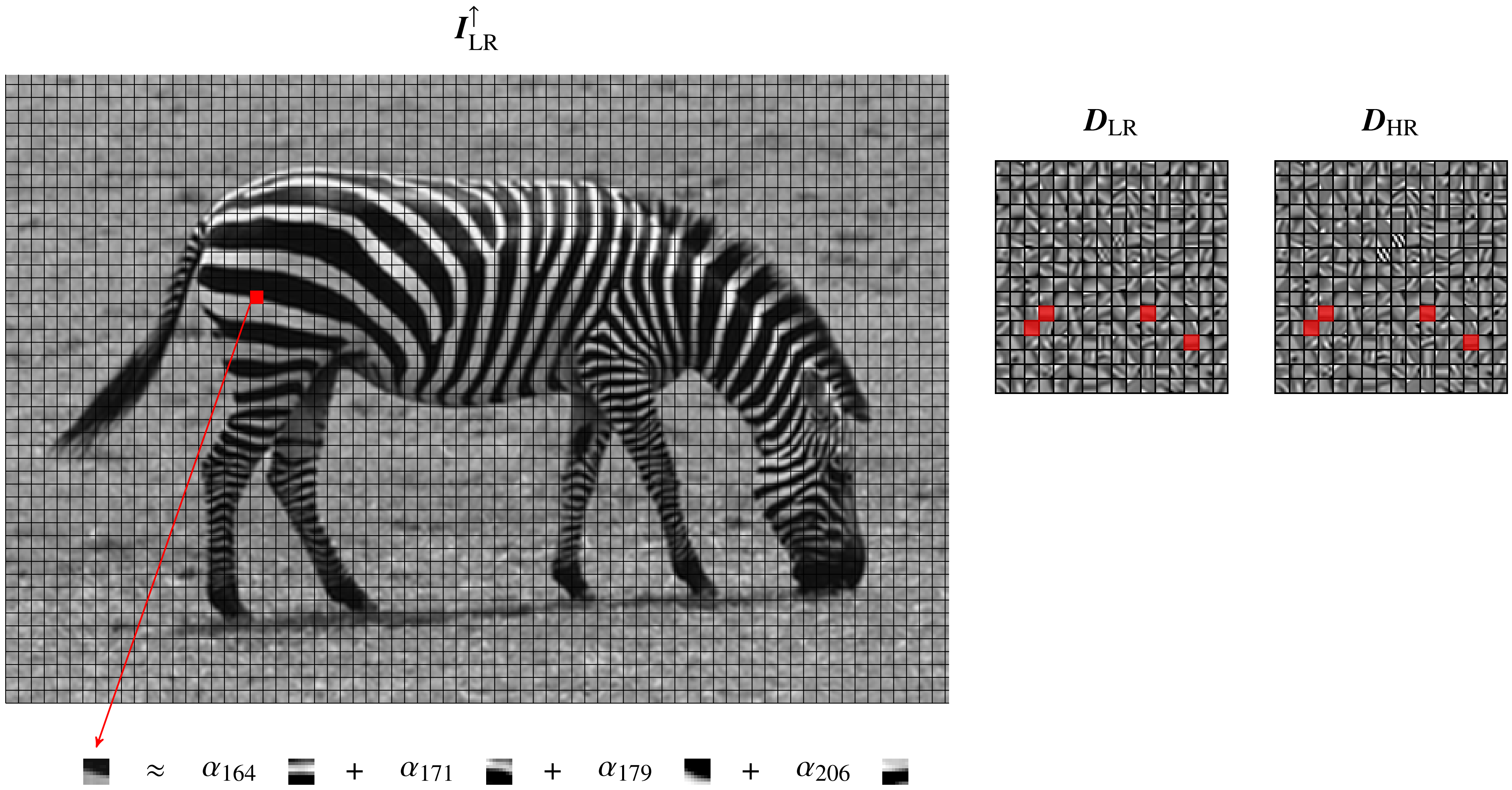


Figure: Example Dictionary

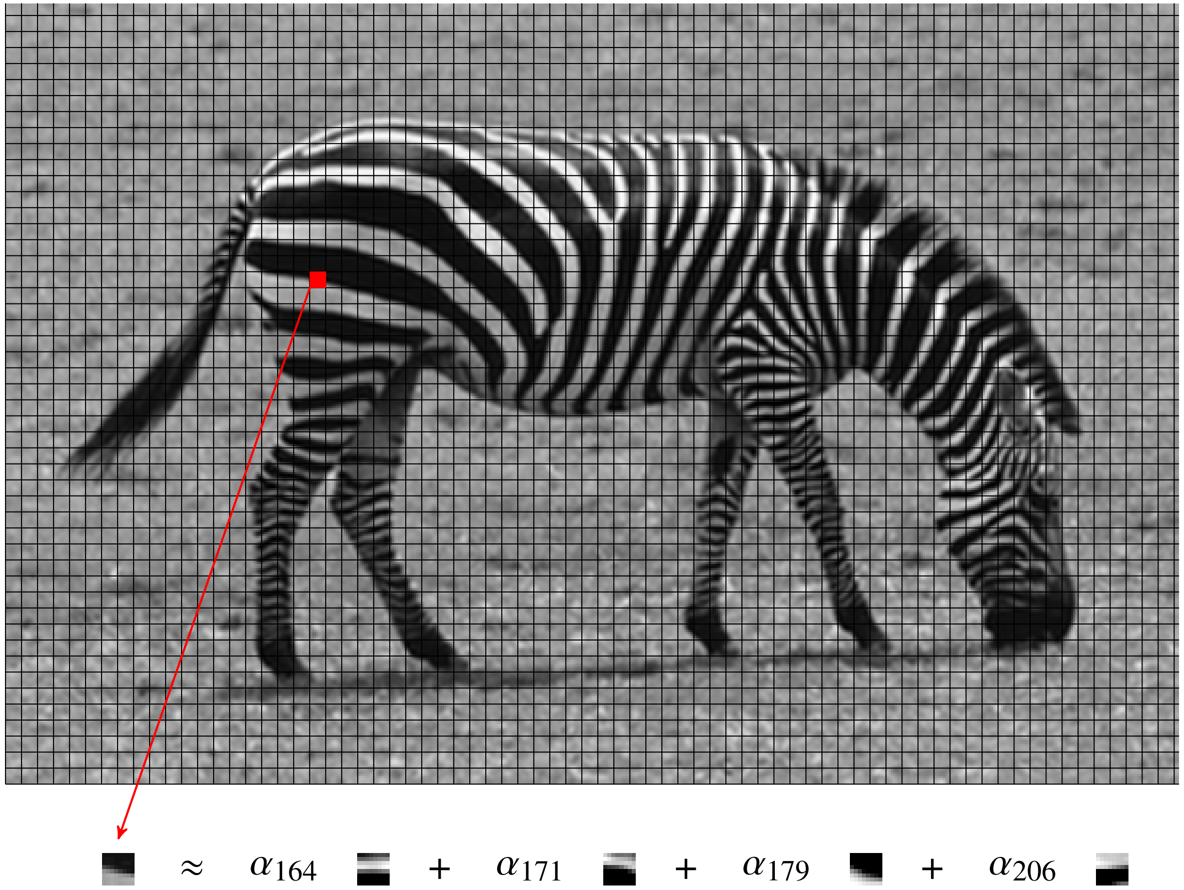
- The concept of dictionary learning can be used for super-resolution by training coupled dictionaries [6].

DLSR: Coupled dictionaries approach

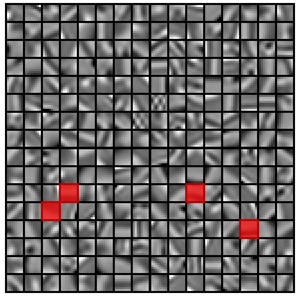


DLSR: Coupled dictionaries approach

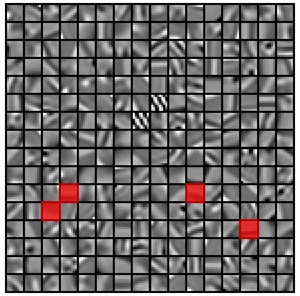
$\boldsymbol{I}_{\mathrm{LR}}^{\uparrow}$



$\boldsymbol{D}_{\mathrm{LR}}$



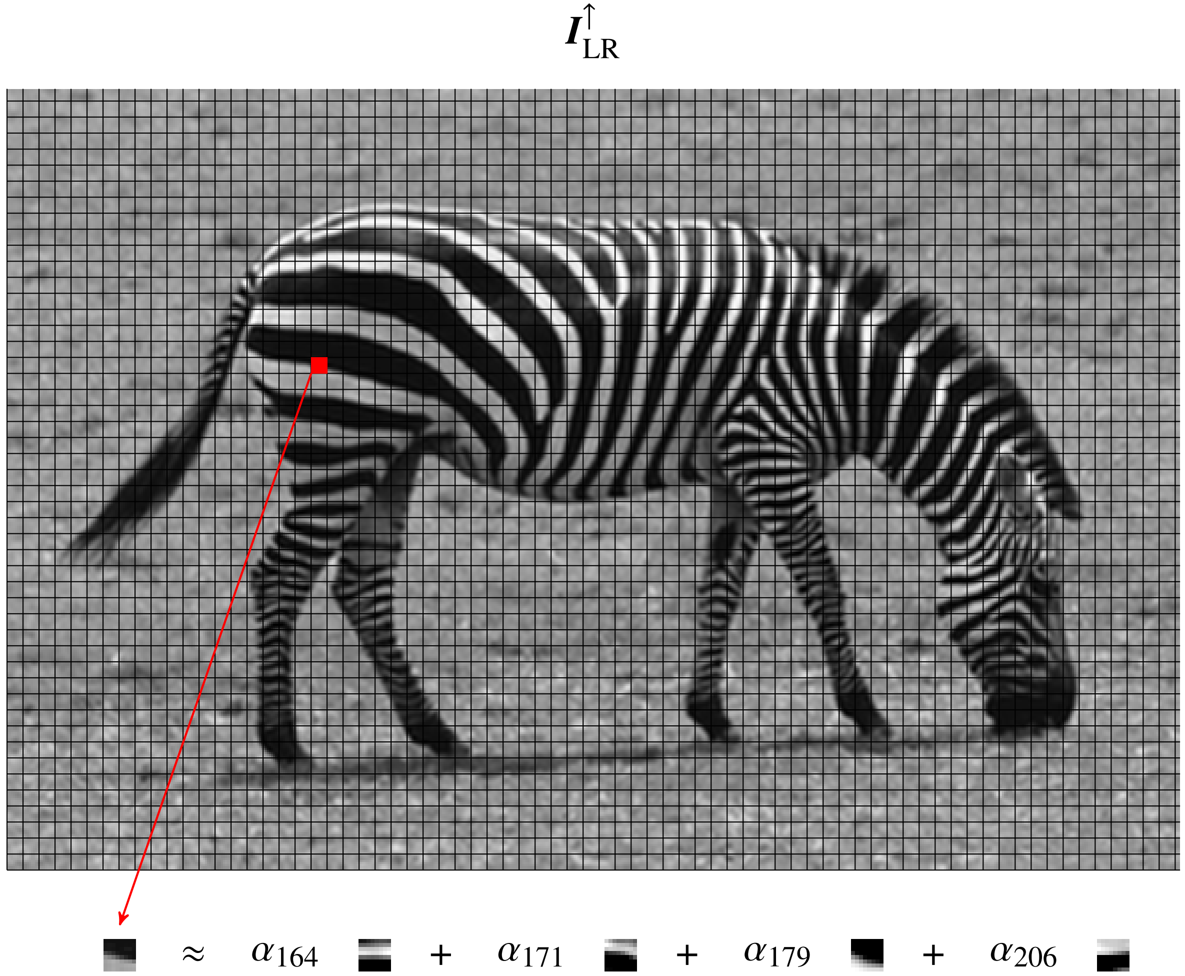
$\boldsymbol{D}_{\mathrm{HR}}$

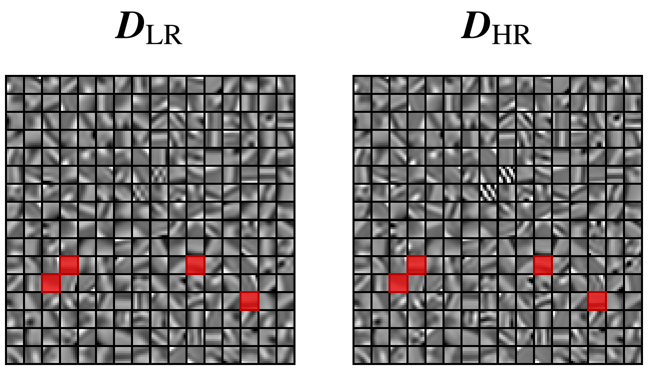


$$\boldsymbol{x}_{\mathrm{LR}} \approx \boldsymbol{D}_{\mathrm{LR}}\boldsymbol{\alpha}$$

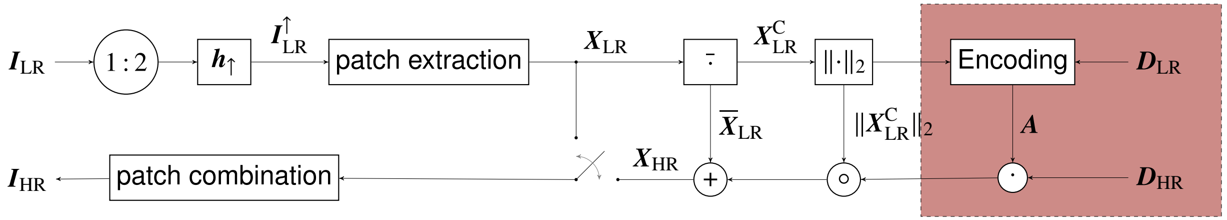
$$\boldsymbol{x}_{\mathrm{HR}} \approx \boldsymbol{D}_{\mathrm{HR}}\boldsymbol{\alpha}$$

DLSR: Coupled dictionaries approach





$$x_{\mathrm{LR}} \approx D_{\mathrm{LR}}\alpha$$
$$x_{\mathrm{HR}} \approx D_{\mathrm{HR}}\alpha$$



DLSR: Mathematical Summary

dictionary learning for LR: $\boldsymbol{D}_{\text{LR}} \leftarrow \arg\min_{\boldsymbol{D}_{\text{LR}}} \sum_{i=1}^{n} \frac{1}{2} \|\boldsymbol{x}_{\text{LR},i} - \boldsymbol{D}_{\text{LR}}\boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$

dictionary learning for HR: $\boldsymbol{D}_{\text{HR}} \leftarrow \arg\min_{\boldsymbol{D}_{\text{HR}}} \|\boldsymbol{X}_{\text{HR}} - \boldsymbol{D}_{\text{HR}}\boldsymbol{A}\|_2^2$

DLSR: Mathematical Summary

dictionary learning for LR: $\boldsymbol{D}_{\text{LR}} \leftarrow \arg\min_{\boldsymbol{D}_{\text{LR}}} \sum_{i=1}^{n} \frac{1}{2} \|\boldsymbol{x}_{\text{LR},i} - \boldsymbol{D}_{\text{LR}} \boldsymbol{\alpha}_i\|_2^2 + \lambda \|\boldsymbol{\alpha}_i\|_1$

dictionary learning for HR: $\boldsymbol{D}_{\text{HR}} \leftarrow \arg\min_{\boldsymbol{D}_{\text{HR}}} \|\boldsymbol{X}_{\text{HR}} - \boldsymbol{D}_{\text{HR}} \boldsymbol{A}\|_2^2$

sparse coding: $\boldsymbol{\alpha} \leftarrow \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{D}_{\text{LR}} \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$

general function approximation: $\boldsymbol{x}_{\text{HR}} \approx \boldsymbol{D}_{\text{HR}} \left(\arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{x}_{\text{LR}} - \boldsymbol{D}_{\text{LR}} \boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \right)$

Contents

Simulation Setup and Results

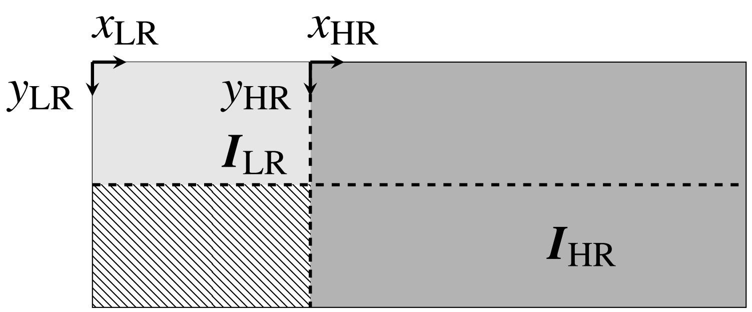
- DLSR setup
  - patchsize $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$
  - number of atoms $K = 512$,
  - sparse coding penalties
    - $\lambda_\mathrm{train} = 0.23$
    - $\lambda_\mathrm{test} = 0.19$

Simulation Setup and Results

- DLSR setup
  - patchsize $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$
  - number of atoms $K = 512$,
  - sparse coding penalties
    - $\lambda_\mathrm{train} = 0.23$
    - $\lambda_\mathrm{test} = 0.19$
- video coding setup
  - 4K sequences from JVET testset
  - anchor VTM-5.0
  - $\mathrm{QP} \in \{37, 42, 47, 52\}$

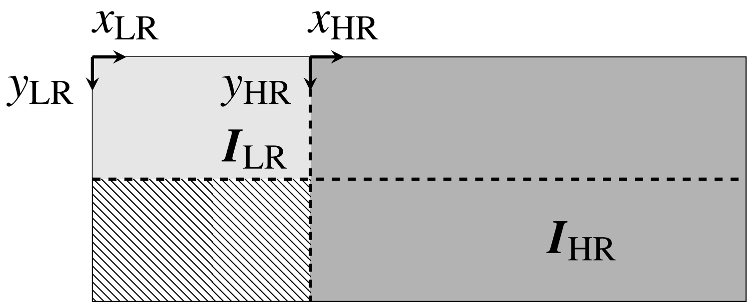
Simulation Setup and Results

- DLSR setup
  - patchsize $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$
  - number of atoms $K = 512$,
  - sparse coding penalties
    - $\lambda_\mathrm{train} = 0.23$
    - $\lambda_\mathrm{test} = 0.19$
- video coding setup
  - 4K sequences from JVET testset
  - anchor VTM-5.0
  - $\mathrm{QP} \in \{37, 42, 47, 52\}$



Simulation Setup and Results

- DLSR setup
  - patchsize $s_\mathrm{p} \times s_\mathrm{p} = 8 \times 8$
  - number of atoms $K = 512$,
  - sparse coding penalties
    - $\lambda_\mathrm{train} = 0.23$
    - $\lambda_\mathrm{test} = 0.19$
- video coding setup
  - 4K sequences from JVET testset
  - anchor VTM-5.0
  - $\mathrm{QP} \in \{37, 42, 47, 52\}$



	All Intra		Random Access	
	$\boldsymbol{h}_\uparrow$	DLSR	$\boldsymbol{h}_\uparrow$	DLSR
Campfire	−19.5 %	−21.0 %	−12.3 %	−13.5 %
CatRobot1	−9.6 %	−10.7 %	−5.9 %	−8.1 %
DaylightRoad2	−7.2 %	−8.0 %	−7.4 %	−8.1 %
FoodMarket4	−8.0 %	−8.2 %	−12.5 %	−12.7 %
ParkRunning3	−16.4 %	−16.8 %	−14.7 %	−15.3 %
Tango2	−10.0 %	−10.1 %	−11.5 %	−11.7 %
AVG	−11.8 %	−12.5 %	−10.7 %	−11.6 %

Table: Bjøntegaard Delta rate savings

Conclusion

- Coding gains with respect to VTM 5.0 can be achieved by performing an adaptive resolution change on the picture level

Conclusion

- Coding gains with respect to VTM 5.0 can be achieved by performing an adaptive resolution change on the picture level
- Dictionary learning based super-resolution leads to an additional coding gain
  - For All Intra additional 0.7% of rate savings are achieved.
  - For Random Access the additional gain is even higher and the rate can be reduced by further 0.9%.

Thank you for your attention!

Any questions?

Jens Schneider schneider@ient.rwth-aachen.de

Institut für Nachrichtentechnik, RWTH Aachen University
www.ient.rwth-aachen.de

References

[1] Jens Schneider, Johannes Sauer, and Mathias Wien. "Dictionary Learning based High Frequency Inter-Layer prediction for Scalable HEVC". In: *Proc. of IEEE Visual Communications and Image Processing VCIP '17*. St. Petersburg, USA: IEEE, Piscataway, Dec. 2017.

[2] F.C.N. Pereira and T. Ebrahimi. *The MPEG-4 Book*. IMSC Press multimedia series. Prentice Hall PTR, 2002.

[3] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang. "Convolutional Neural Network-Based Block Up-Sampling for Intra Frame Coding". In: *IEEE Transactions on Circuits and Systems for Video Technology* 28.9 (Sept. 2018), pages 2316–2330.

[4] A Aaron, Z. Li, M. Manohara, J De Cock, and D. Ronca. *Per-Title Encode Optimization*. `https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2`. [Online; accessed 2-May-2019]. 2015.

[5] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang. "Convolutional Neural Network-Based Block Up-Sampling for Intra Frame Coding". In: *IEEE Transactions on Circuits and Systems for Video Technology* 28.9 (Sept. 2018), pages 2316–2330.

[6] Roman Zeyde, Michael Elad, and Matan Protter. "On single image scale-up using sparse-representations". In: *International conference on curves and surfaces*. Springer. 2010, pages 711–730.

[7] Gisle Bjontegaard. *Calculation of average PSNR differences between RD-curves*. Technical report Doc. VCEG-M33. Austin, USA: ITU-T SG16/Q6 VCEG, 2001.

[8] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. 1st. Springer Publishing Company, Incorporated, 2010.

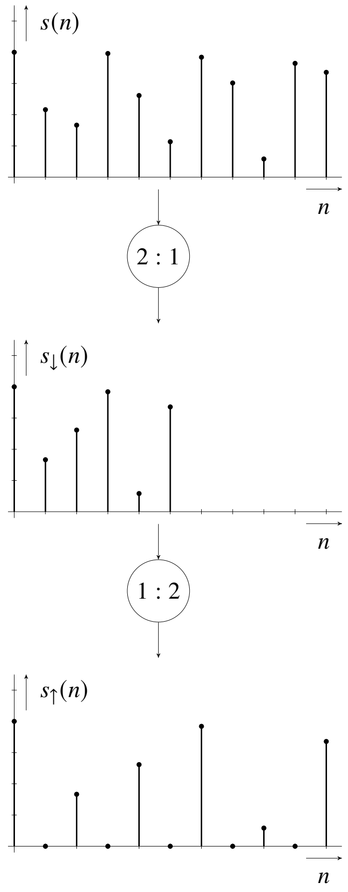
[9] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. "Online dictionary learning for sparse coding". In: *Proceedings of the 26th annual international conference on machine learning*. ACM. 2009, pages 689–696.

[10] M. Wien. *High Efficiency Video Coding*. 1st edition. Springer-Verlag Berlin Heidelberg, 2015.

[11] Radu Timofte, Vincent De Smet, and Luc Van Gool. "A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution". In: volume 9006. Apr. 2015, pages 111–126.

Fundamentals: Downsampling and Upsampling

- Downsampling realized by taking e.g. every second sample
- This introduces alias in general
  - The signal is filtered with a anti-aliasing filter
- Upsampling is realized by inserting zeros
  - The signal is filtered with an interpolation filter
- MATLAB's *imresize* function does not strictly follow this methodology, when using the bicubic kernel
  - samples are shifted when downsampling and shifted back when upsampling



- Bicubic downsampling filter has 8 taps
  - This introduces a phase shift of the downsampled signal
- The downsampling filter used in SHVC has 11 taps
  - no distortion of the phase during downsampling

(a) bicubic downsampling filter

(b) downsampling filter used for SHVC

Figure: different downsampling filters

Lehrstuhl und Institut für Nachrichtentechnik

Fundamentals: Downsampling Filters

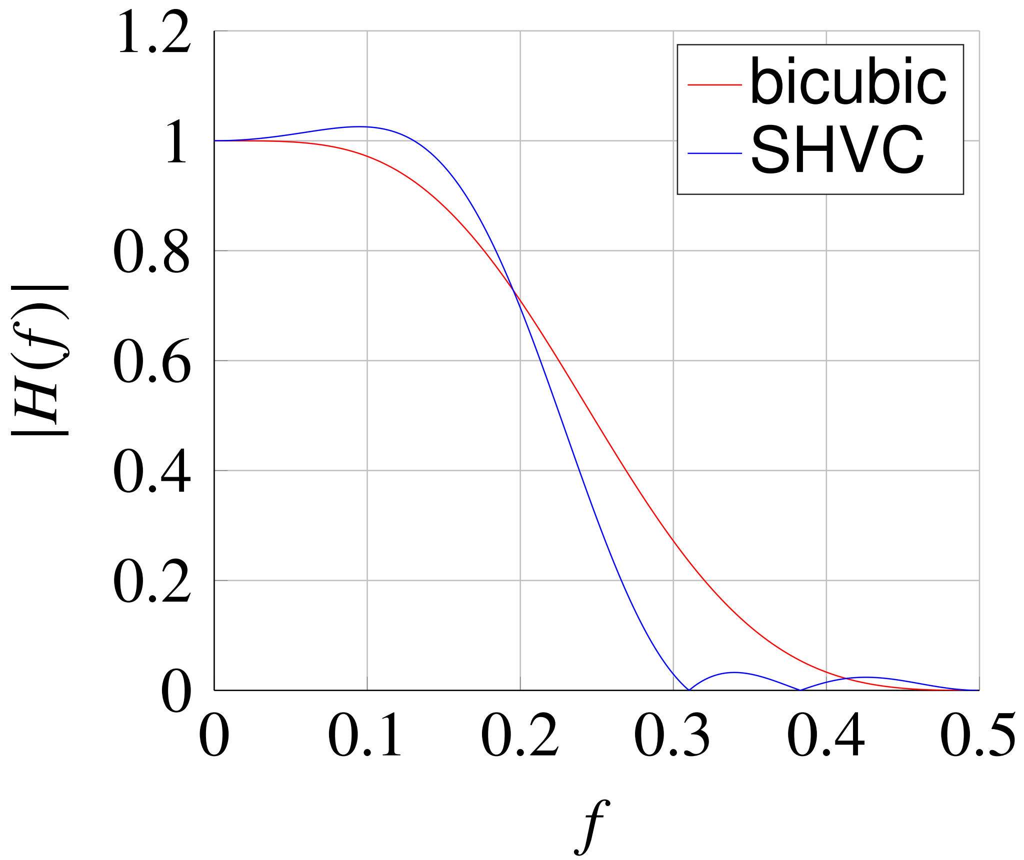
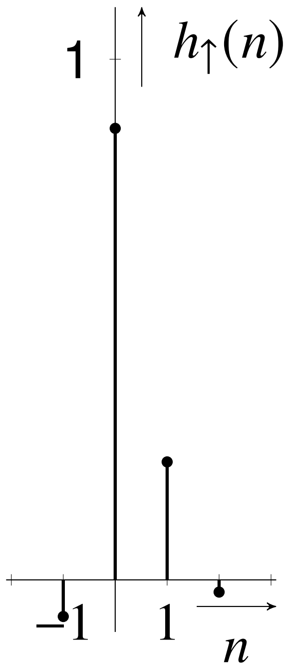


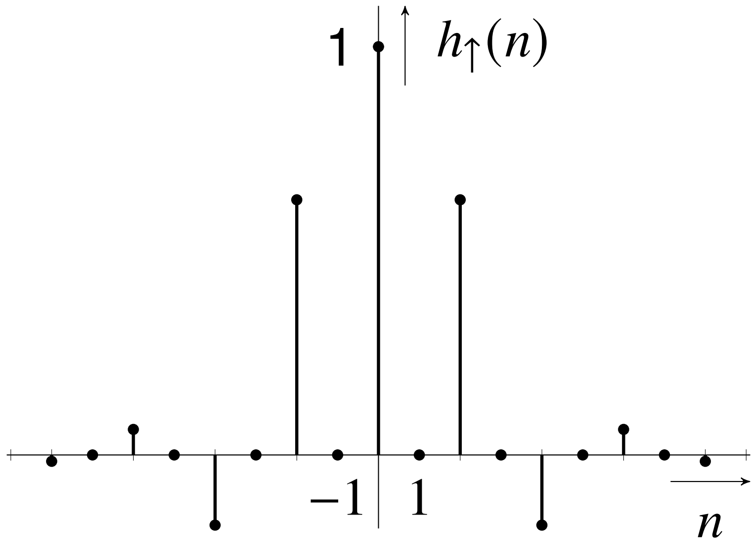
Figure: Frequency response of different downsampling filters

- Bicubic upsampling filter has to be applied several times since we need to “backshift” the phase



(a) bicubic upsampling filter

- The upsampling filter is derived from the half-pel interpolation filters used in HEVC
  - We need to insert a 1 at position zero and 0s at the odd sample positions



(b) upsamling filter derived from HEVC interpolation filters

Figure: different upsampling filters