CLIM 2021 –Rennes

Gating Networks

-

Edge-Aware Sparse Representations for Image Processing and Compression

Thomas Sikora

Director, Communication Systems Lab (TUB)

Ruben Verhack
(TUB, Uni Gent, iMinds Lab)

Rolf Jongebloed (TUB)
Eric Bochinski (TUB)
Lieven Lange (TUB)
Jonas Geistert (TUB)

Gating Networks

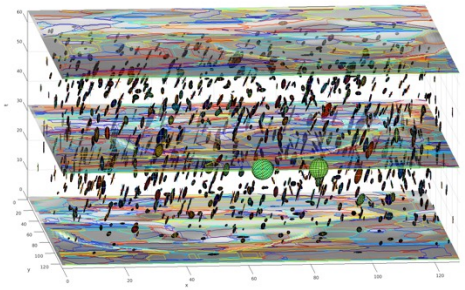
Outline

I. 30 years of Video Coding (Hybrid DCT/DPCM)

I. Disruptive Strategies for Image and Video Coding?

II. Universal Coding - What are Gating Networks

III. Some Results for Image and Video Coding



Happy Anniversary

30 years of MPEG/ITU-type compression for video

30 years of MC-DPCM/Transform

Motion-compensated DPCM Transform Coding

Evolution of Video Coding Standards

- JPEG (1990)
- H.261 (1990)
- MPEG-1 (1991)
- MPEG-2 (1994)

- MPEG-4 (2000)
- H.263/MPEG-4 (2005)
- H.264 (2010)
- MPEG-HEVC (2014)

30 years of DCT-DPCM technology

Today's Video Coding Technology is

Hybrid DCT/DPCM

Google VP8/VP9

Improvements are mainly due to larger size content

Where is this going?

- I quote (influential members from MPEG):
- „... We find it incredibly difficult to squeeze out more performance from HEVC/VCEG …
- … we need to work on something completely new (a new and disruptive format)”

Press the „Reset Button“ and nothing is going to happen.
Virtually NO ideas where to start with „disruptive“ coding approaches.

Well: ALMOST!

Learning for Video Compression with Recurrent Auto-Encoder and Recurrent Probability Model

Ren Yang, *Student Member, IEEE,* Fabian Menzter, *Student Member, IEEE,*
Luc Van Gool, *Member, IEEE,* and Radu Timofte, *Member, IEEE*

2020

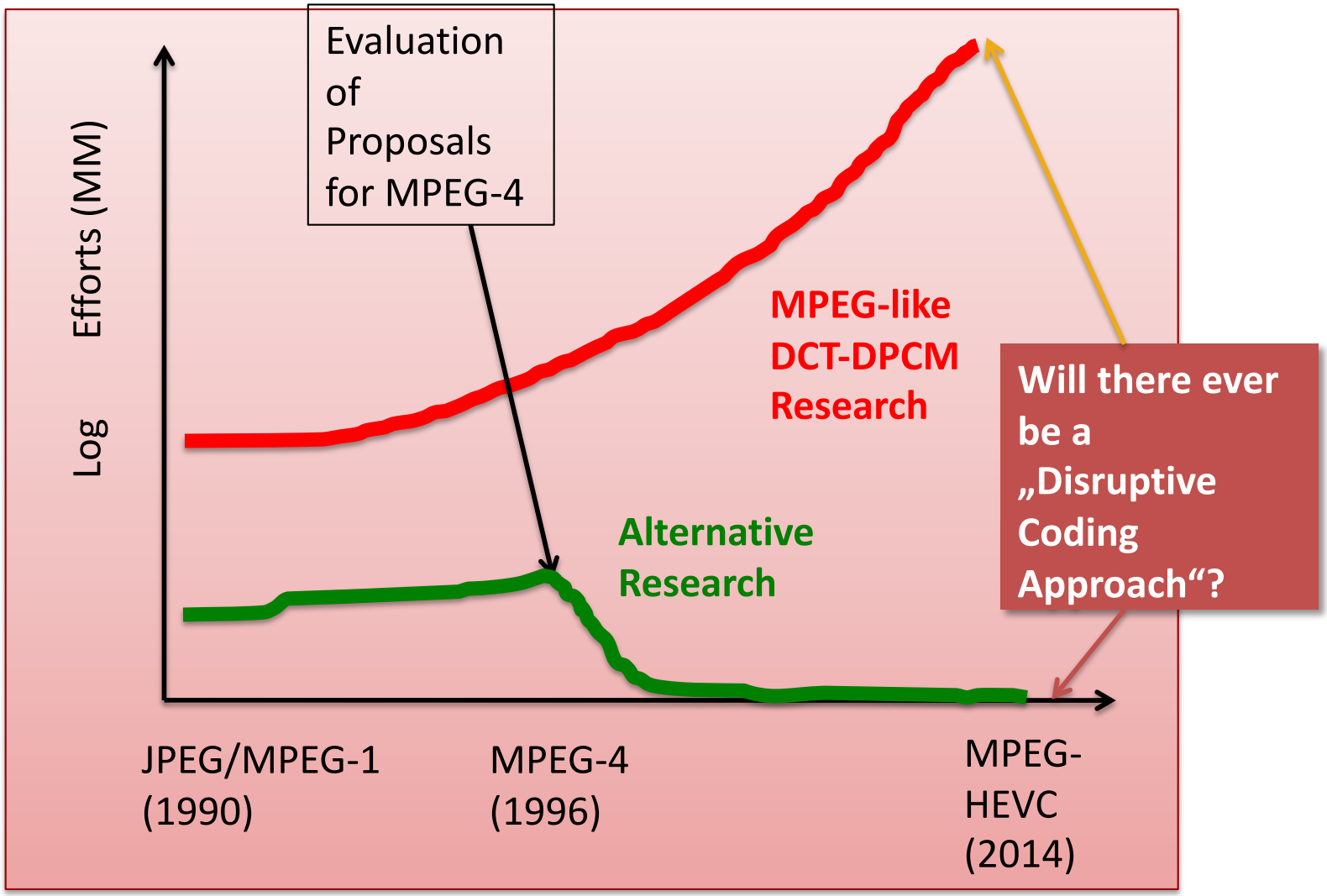
DVC: An End-to-end Deep Video Compression Framework

Guo Lu[1], Wanli Ouyang[2], Dong Xu[3], Xiaoyun Zhang[1], Chunlei Cai[1], and Zhiyong Gao *[1]

[1]Shanghai Jiao Tong University, {luguo2014, xiaoyun.zhang, caichunlei, zhiyong.gao}@sjtu.edu.cn
[2]The University of Sydney, SenseTime Computer Vision Research Group, Australia
[3]The University of Sydney, {wanli.ouyang, dong.xu}@sydney.edu.au

2019

30 Years of JPEG/MPEG-like Coding



Alternative Research around 1996

(Examples)

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 7, NO. 1, FEBRUARY 1997

Segmentation-Based Video Coding System Allowing the Manipulation of Objects

Philippe Salembier, Ferran Marqués, Montse Pardàs, Josep Ramon Morros, Isabelle Corset, Sylvie Jeannin, Lionel Bouchard, Fernand Meyer, and Beatriz Marcotegui

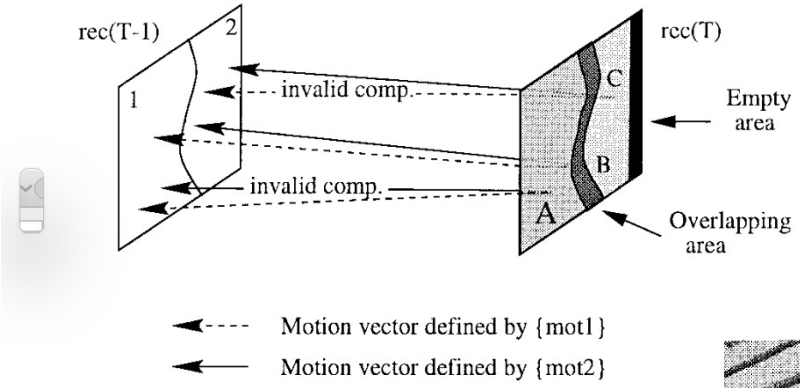
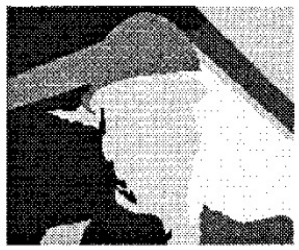


Fig. 7. Backward motion compensation of partition.

Original frame



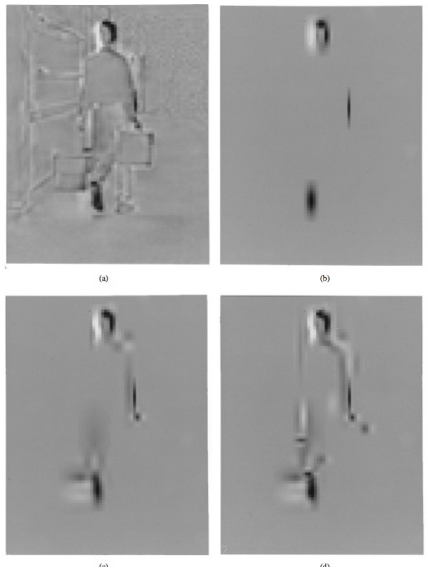
Final partition

Coded frame

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 7, NO. 1, FEBRUARY 1997

Very Low Bit-Rate Video Coding Based on Matching Pursuits

Ralph Neff and Avideh Zakhor



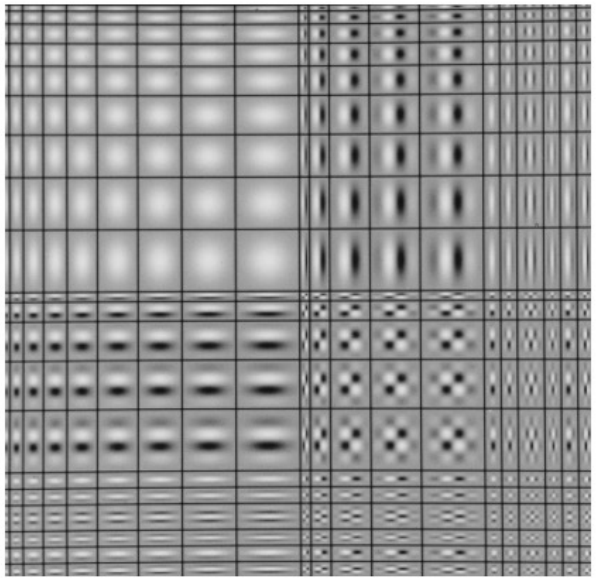


Fig. 2. The 2-D separable Gabor dictionary.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 15, NO. 6, JUNE 1993 545

3-D Motion Estimation in Model-Based Facial Image Coding

Haibo Li, *Student Member, IEEE,* Pertti Roivainen, and Robert Forchheimer

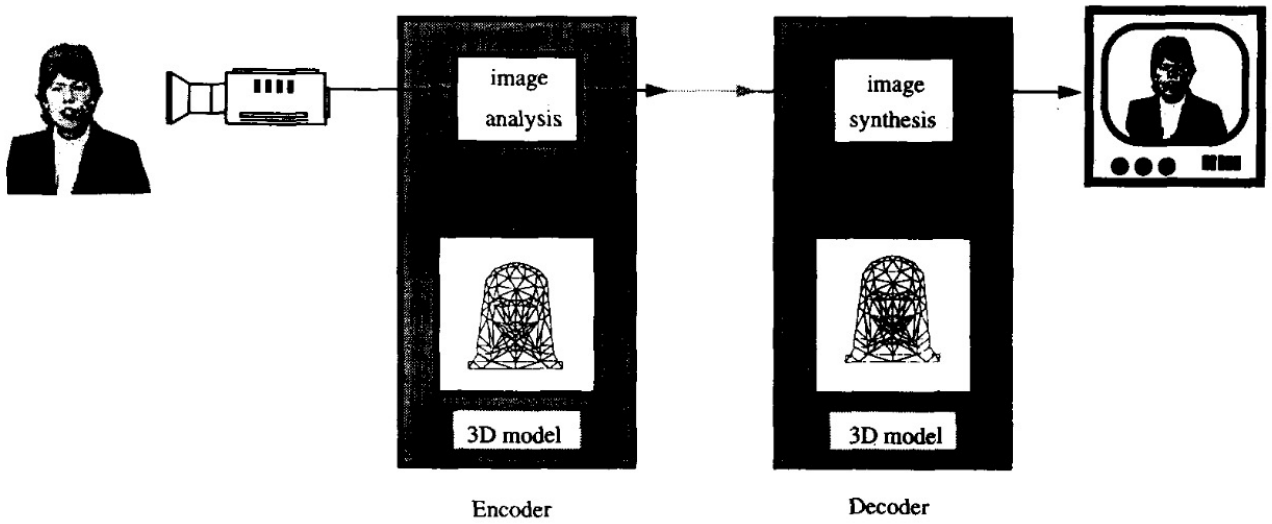


Fig. 1. Model-based image coding—A 3-D object oriented system.

J.-R. Ohm

"Three-dimensional subband coding with motion compensation,"

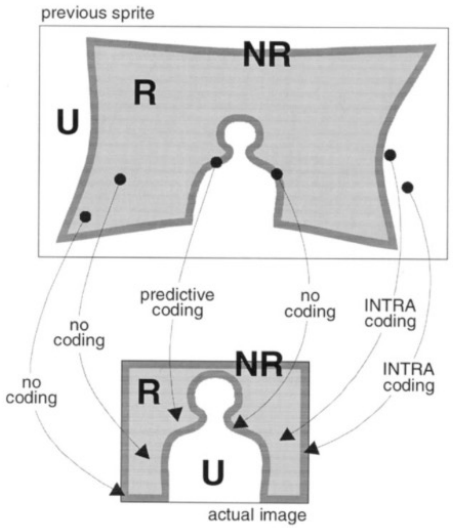
in *IEEE Transactions on Image Processing, vol. 3, no. 5, pp. 559-571, Sept. 1994*

We still see work on 3D wavelets for video compression around – mainly for „scalable coding“.

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 9, NO. 8, DECEMBER 1999

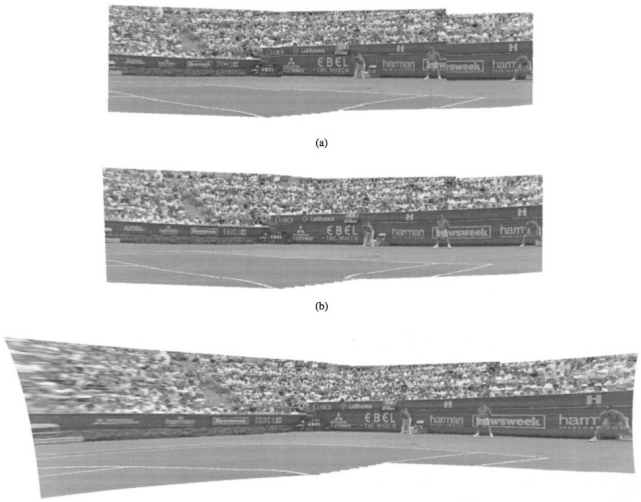
Long-Term Global Motion Estimation and Its Application for Sprite Coding, Content Description, and Segmentation

Aljoscha Smolić, Thomas Sikora, *Senior Member, IEEE*, and Jens-Rainer Ohm, *Member, IEEE*



(a) (b) (c)

Fig. 5. Second frame of the *Stefan* sequence: (a) original image and (b) coding mask. White: outside the object; black: not coded; light gray: predictive coding; dark gray: INTRA coding. (c) Resulting coding VOP.



(a) (b) (c)

Fig. 6. Sprites for sequence *Stefan* generated by (a) accumulation of short-term parameters using an affine motion model, (b) direct estimation of long-term parameters using an affine motion model, and (c) direct estimation of long-term parameters using a parabolic motion model.

We still see work around – but improvements only in the background.

Coding Performance of Sprite Coding

No temporal DPCM coding in the background

Towards a „Disruptive Coding Approach“

What is it we are looking for?

- **Break-Through in Performance**
- **Sparse Representation**
  - *that allows exploration of long-range spatio-temporal statistical dependencies between pixels*
  - *Edge preserving*
- **End-to-End Optimization – Machine Learning**
  - *Differentiable set of non-linear equations*
- **Universal Coding Approach for 1D/2D/3D/N-D Pixel formats**
- **Drastic departure from DCT-DPCM – block-based approaches**

Towards a „Disruptive Coding Approach“

Question why we should need certain techniques that have been around for such a long time!

Revisit basic assumptions:

Do we need

- Block-Processing?
- Transform Coding (DCT-like)?
- DPCM (Prediction Error Coding)?
- Motion Vectors?

Images are not organized around „blocks“ – but segments!

Is this the only way we can represent textures?

DPCM means „quantization of quantization errors“. Is that a good idea?

Motion vectros seem to be the „Holy Grale“ of video. Can we do without them?

The Re-Emergence of Gating Networks

To the rescue of the „complexity barrier“?!

To the rescue of the „compression challenge“?!

Happy Anniversary

30 years of Mixtures-of-Experts

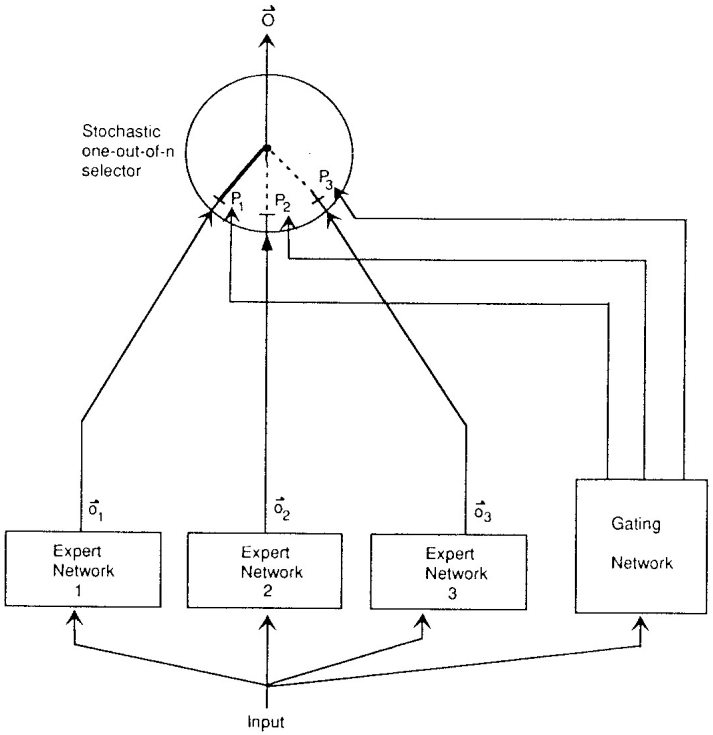
Gating Networks

[PS] Adaptive **mixtures** of local experts.

RA **Jacobs**, MI **Jordan**, SJ Nowlan, GE Hinton - Neural …, 1991 - csri.utoronto.ca

We present a new supervised learning procedure for systems composed of many separate networks, each of which learns to handle a subset of the complete set of training cases. The new procedure can be viewed either as a modular version of a multilayer supervised …

☆ Zitiert von: 4103 Ähnliche Artikel Alle 24 Versionen



Mixtures-of-Experts

Gating Networks

Gated Experts Networks

Mixture-of-Experts Networks

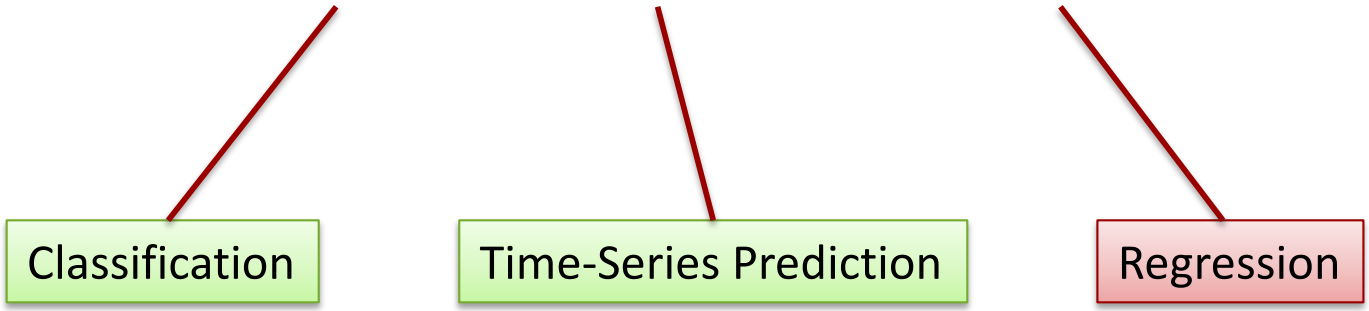
Many names for the same thing!

Interesting overview paper on Gating Networks (Mixtures-of-Experts)

IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 23, NO. 8, AUGUST 2012 1177

Twenty Years of Mixture of Experts

Seniha Esen Yuksel, *Member, IEEE*, Joseph N. Wilson, *Member, IEEE*, and Paul D. Gader, *Fellow, IEEE*



„And when he saw her looking so lovely in her sleep, he could not turn away his eyes; and presently he stopped and kissed her. ....
And she awaked, and opened her eyes, and looked very kindly on him“.

Rosamond tale (Sleeping Beauty)

Work appeared – but not with much impact.

The Re-Emergence of Gating Networks

OUTRAGEOUSLY LARGE NEURAL NETWORKS: THE SPARSELY-GATED MIXTURE-OF-EXPERTS LAYER

Noam Shazeer[1], Azalia Mirhoseini*†[1], Krzysztof Maziarz*[2], Andy Davis[1], Quoc Le[1], Geoffrey Hinton[1] and Jeff Dean[1]

[1]Google Brain, {noam,azalia,andydavis,qvl,geoffhinton,jeff}@google.com
[2]Jagiellonian University, Cracow, krzysztof.maziarz@student.uj.edu.pl

2017

Goal: Conditional Computation

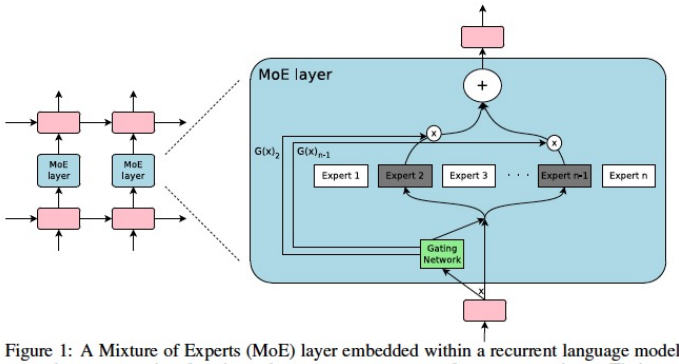


Figure 1: A Mixture of Experts (MoE) layer embedded within a recurrent language model. In this case, the sparse gating function selects two experts to perform computations. Their outputs are modulated by the outputs of the gating network.

Kiss the „Gating Networks“ awake. My guess is that we will see much renewed interest in Gating Netrworks.

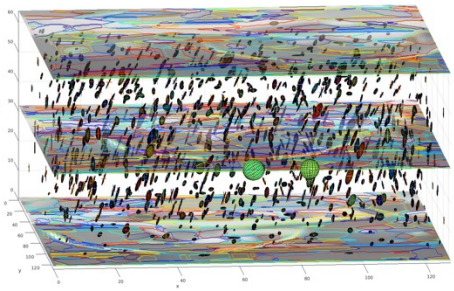
Our Work on Gating Networks for Pixel Data

Exploration of long-range statistical dependencies

-in high-dimension pixel data

Sparse & edge-aware representation

- *Compression*
- *Denoising*
- *N-D Segmentation*
- *Graph-Representation*
- *Feature Extraction*



We use Gating Networks differently – for „regression“

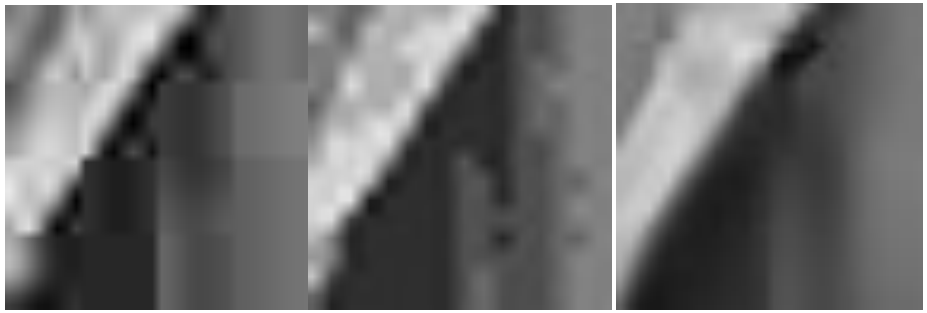
Apparently interesting for many applications beyond „compression“

Gated-Experts Compression

Let us start with a result!

Original

0.81 SSIM **26 dB**



JPEG

JPEG 2000

BPG (HEVC-Intra)

0.93 SSIM

31.7 dB

GE Network Compression

coded at same bpp (without header bits)

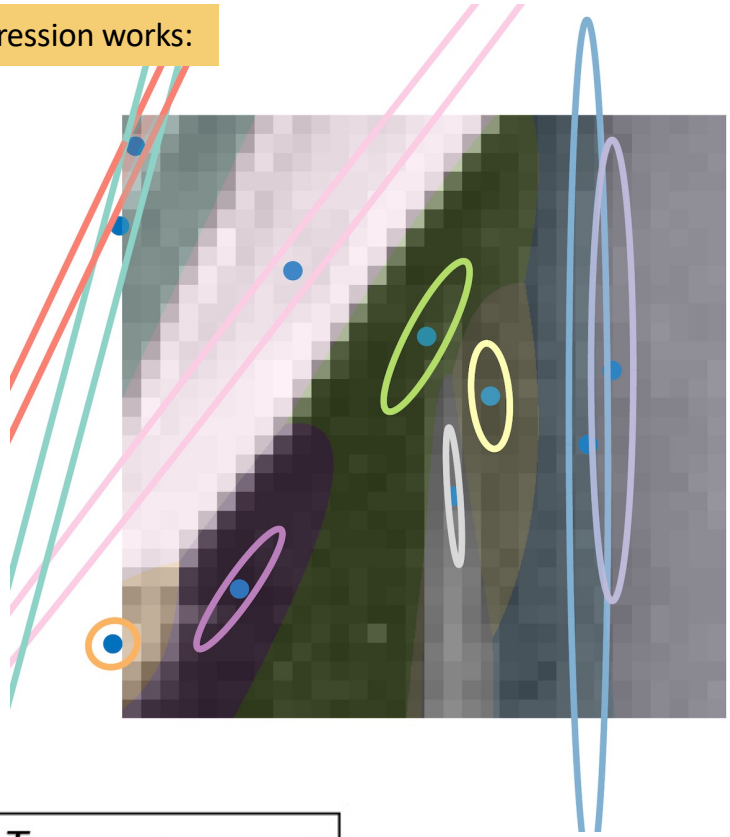
Excellent edge reconstruction – no common compression artifacts

In-built noise reduction capabilities?

Coding Swarms of „Atoms“

This is how Gated Experts Network compression works:

Original



Decoded

7 Parameters/Atom

$$e^{-\left(\underline{x}-\underline{\mu}_k\right)^T \cdot \underline{R}_k \cdot \left(\underline{x}-\underline{\mu}_k\right)}$$

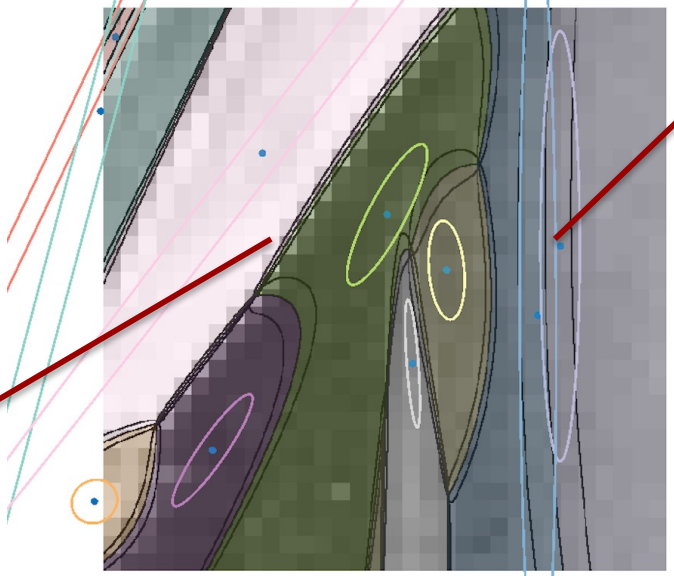
Atom/Kernel

Compression is based on edge-aware (steering) „atoms“. In this example we use simple steered Gaussian functions.
A great wealth of atom functions can be used. The parameters are tuned in an optimization framework,
quantized and coded. The decoder reconstructs the image using a regression formula – using edge-aware gating functions.

Soft-Gated Regions (Soft-Max Windows)

Overlaping gates for smooth transitions

Original



Decoded

Sharp transitions around edges

A 2D-Soft-Gating Function provides for sharp transitions around edges and can overlap for smooth transitions.

$$y(\underline{x}) = \sum_{k=1}^{K=10} m_k \cdot \frac{e^{-\left(\underline{x}-\underline{\mu}_k\right)^T \underline{R}_k \left(\underline{x}-\underline{\mu}_k\right)}}{\sum_{j=1}^{10} e^{-\left(\underline{x}-\underline{\mu}_j\right)^T \underline{R}_j \left(\underline{x}-\underline{\mu}_j\right)}}$$

This is the regression function: an expert function (here a scalar With a grey value) is gated with a 2D-Soft-Gatinting window function

Edge-aware & Sparse Representation

Experts

2D Soft-Gating Function (Window)

Mixture of Steered Gaussians

The Most Simple Network

It is actually not required that atom functions steer to reproduce sharp edges and smooth transitions.
It is also possible to use a network with simple radial Gaussian functions. Two radial Gaussian functions are sufficient to reproduce an edge in an image.

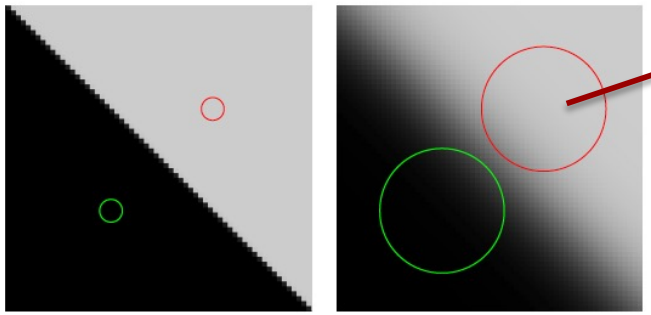
- Constant Expert – Radial Gating Function

K=2

$$y_p(\boldsymbol{x}) = \sum_{i=1}^{K} m_i \cdot \frac{\exp\left(-S\left\|\boldsymbol{x}-\boldsymbol{\mu}_i\right\|^2\right)}{\sum_{j=1}^{K}\exp\left(-S\left\|\boldsymbol{x}-\boldsymbol{\mu}_j\right\|^2\right)}.$$

In this example the locations of the atoms are fixed, but the bandwidth S changes, to control sharpness of the edge.

$$\frac{\exp\left(-S\left\|\boldsymbol{x}-\boldsymbol{\mu}_i\right\|^2\right)}{}$$



(a) Small bandwidth (b) Large bandwidth

Fig. 2: Edge reconstruction with simple radial kernels

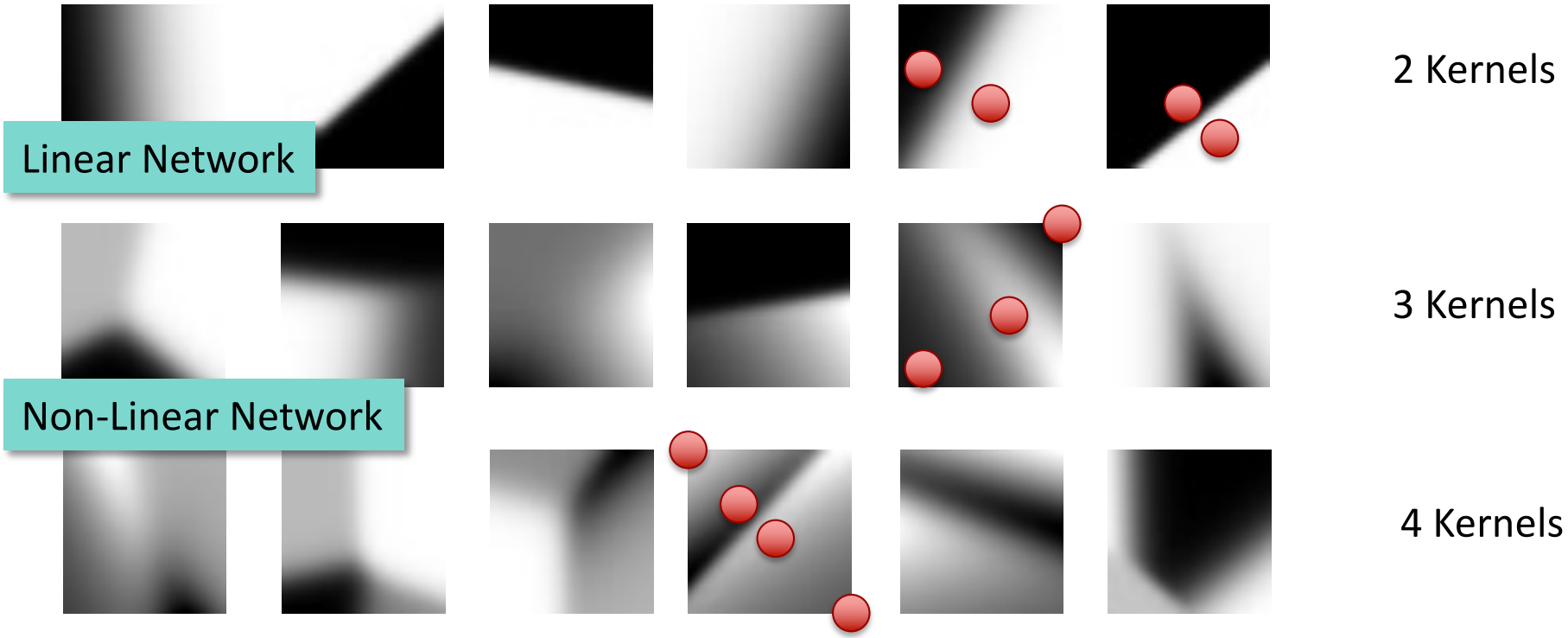
Edge-aware representation

2 Kernels reproduce lines with sharp or smooth transitions

Examples of Block Patterns

All Kernels with same bandwidth

It is also possible to fix bandwidth S, and change the location of the atoms, to control sharpness of the edge.

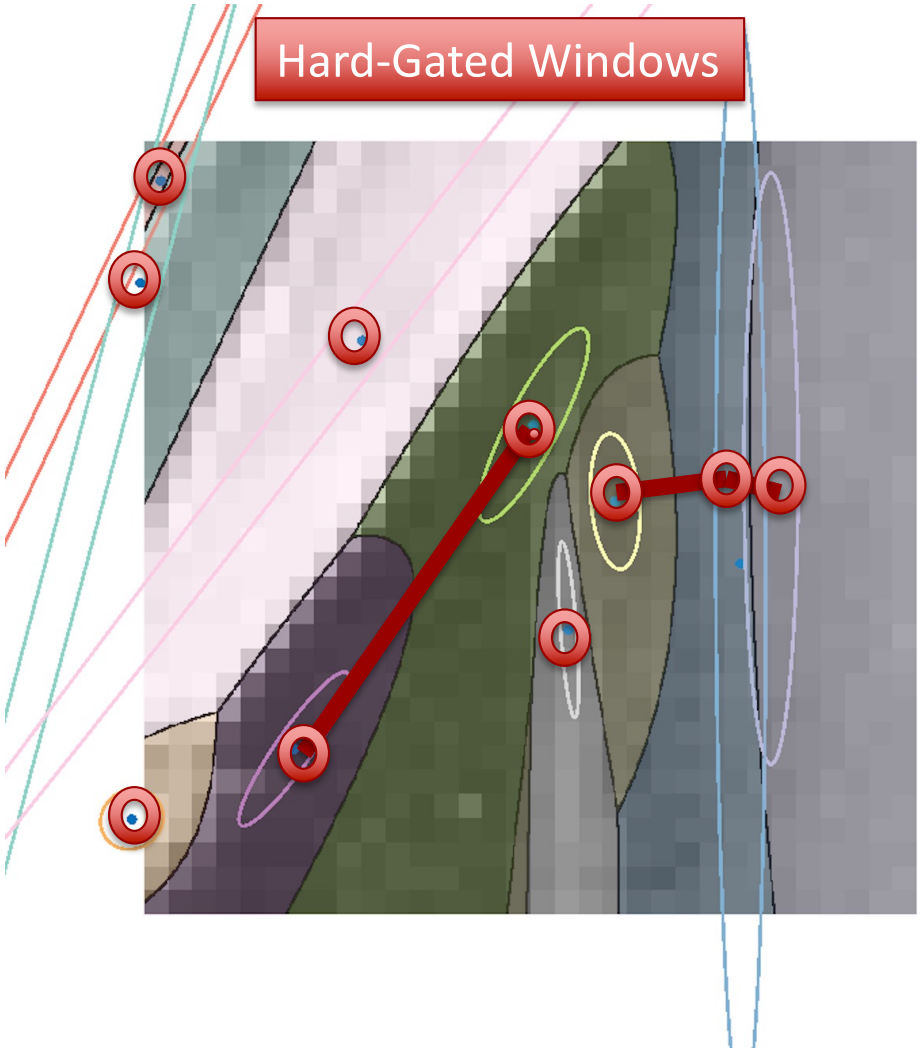


Tok, M., Jongebloed, R., Lange, L., Bochinski, E., & Sikora, T. (2018). An MSE Approach For Training And Coding Steered Mixtures Of Experts. 2018 Picture Coding Symposium (PCS), 273-277.

Segmentation & Graph Representation

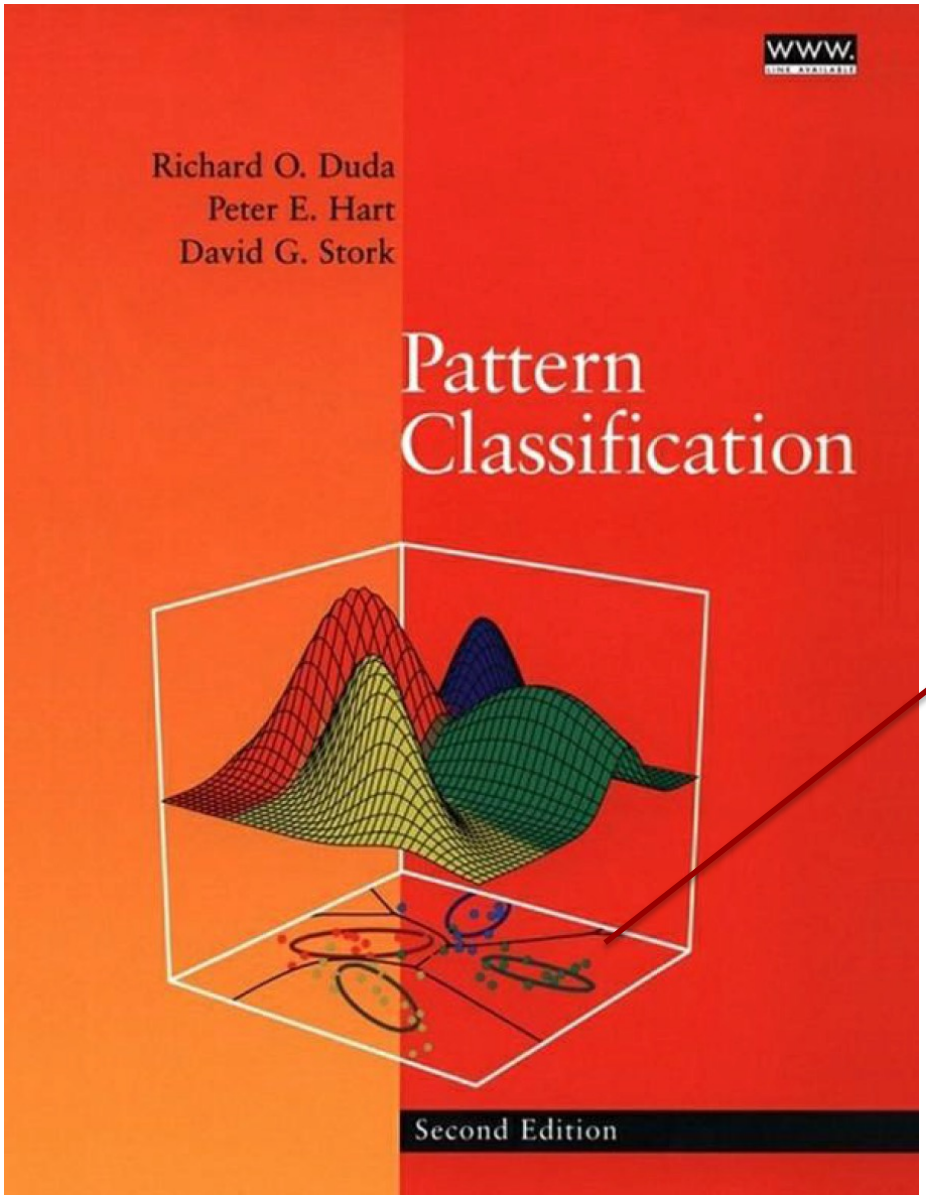
Original

Hard-Gated Windows



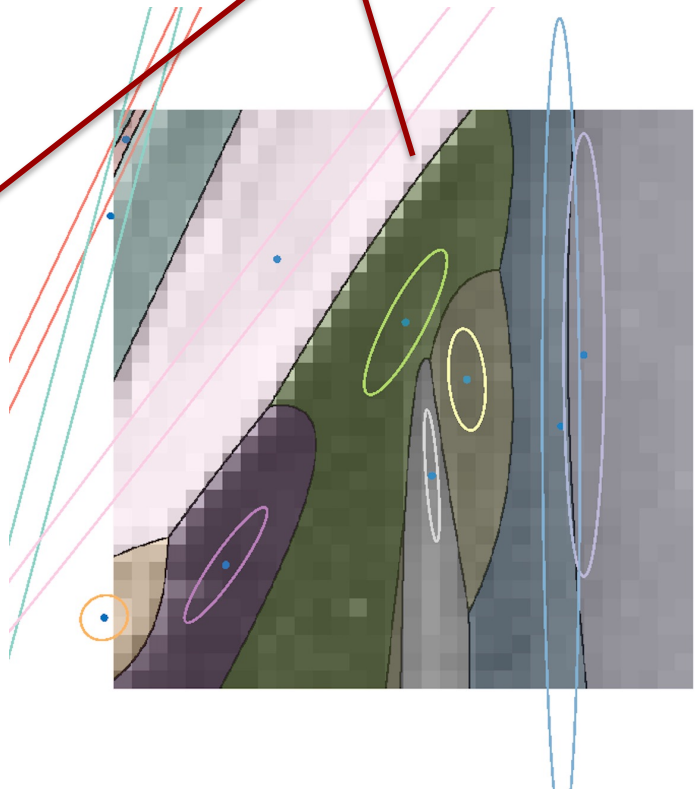
connected

un-connected



Strong similarity with Bayesian Classification.
Decision boundaries are derived through Hard-Gating of a Soft-Gating region.
Except we use the strategy for regression.

Decision Boundaries



Compression Results for Images

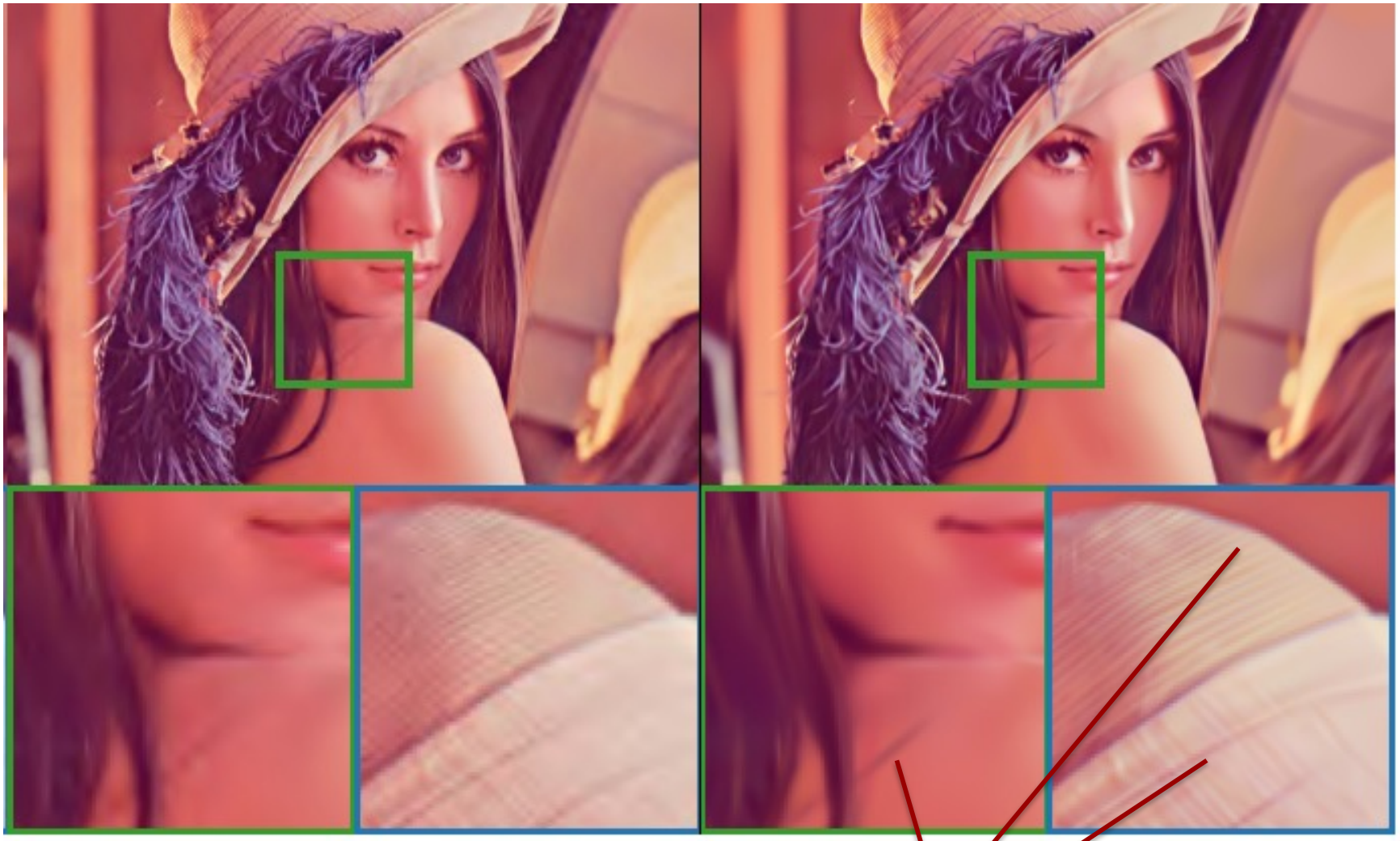
Does this work for compression of entire Images?

Indeed very efficient for compression.

same rate

R. Jongebloed, E. Bochinski, L. Lange and T. Sikora, "Quantized and Regularized Optimization for Coding Images Using Steered Mixtures-of-Experts," *2019 Data Compression Conference (DCC), Snowbird, UT, USA, 2019, pp. 359-368.*

Bochinski, Erik et al. “Regularized Gradient Descent Training of Steered Mixture of Experts for Sparse Image Representation.” 2018 25th IEEE International Conference on Image Processing (ICIP) (2018): 3873-3877.



Excellent reconstruction of edges and textured regions!!!

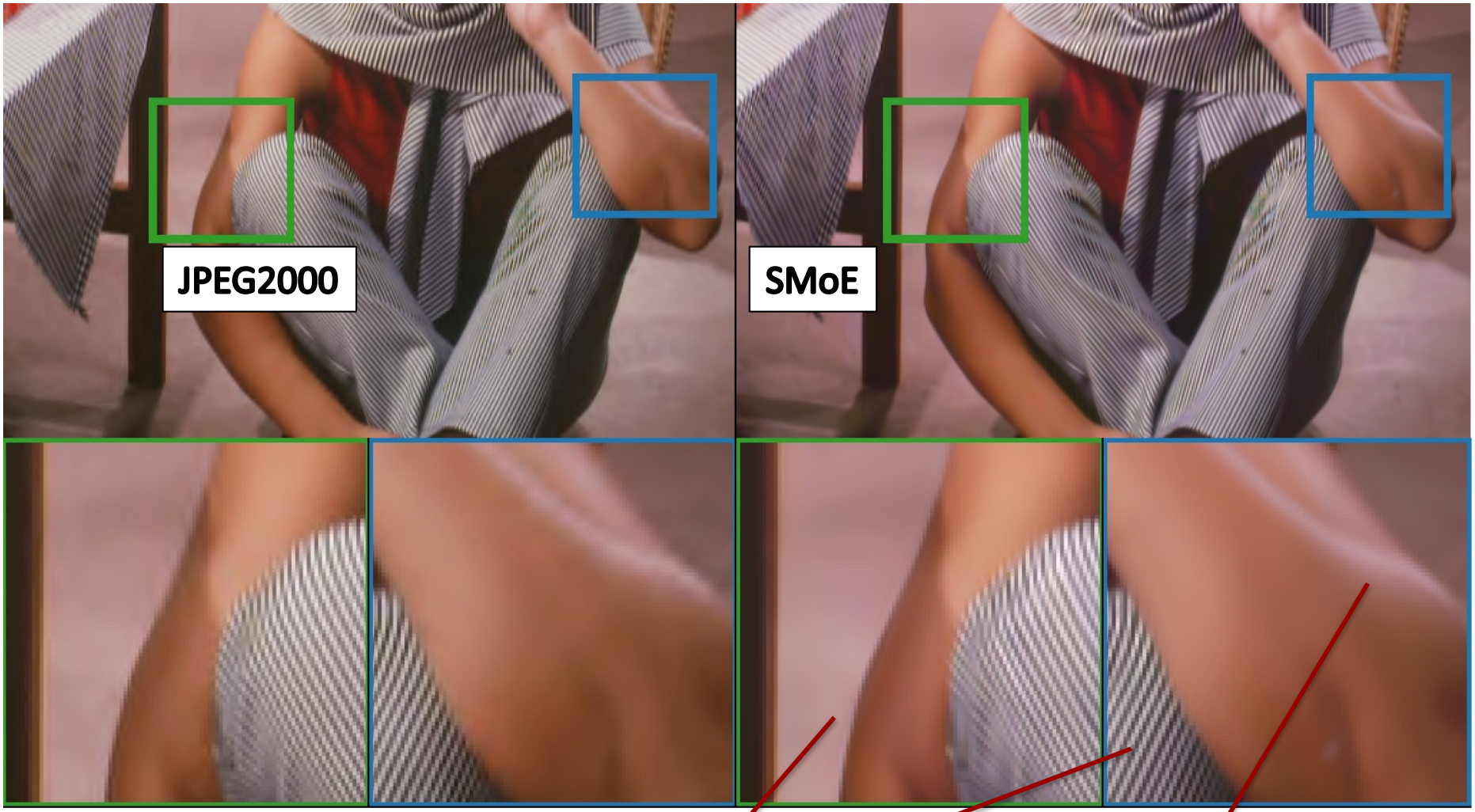
R. Jongebloed, E. Bochinski, L. Lange and T. Sikora, "Quantized and Regularized Optimization for Coding Images Using Steered Mixtures-of-Experts," *2019 Data Compression Conference (DCC), Snowbird, UT, USA, 2019, pp. 359-368.*

Bochinski, Erik et al. “Regularized Gradient Descent Training of Steered Mixture of Experts for Sparse Image Representation.” 2018 25th IEEE International Conference on Image Processing (ICIP) (2018): 3873-3877.

Results for Images

SMoE Gated Expert Compression

same rate



Again excellent reconstruction of edges and textured regions – as well as of smooth transitions!!!

Gating

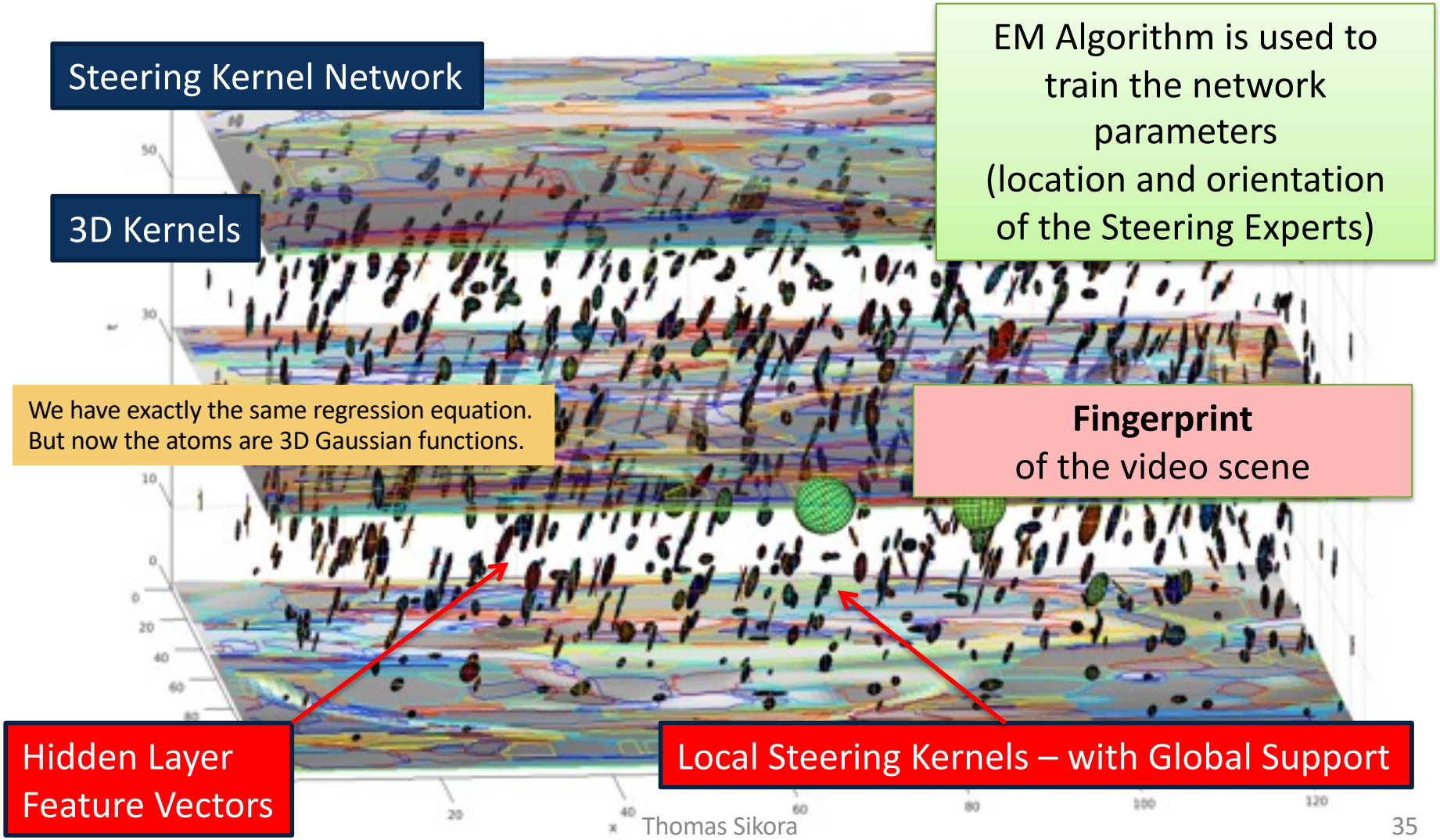
The 2D-Soft-Gatied regions are here displayed as Hard-Gated regions.

Sparsity

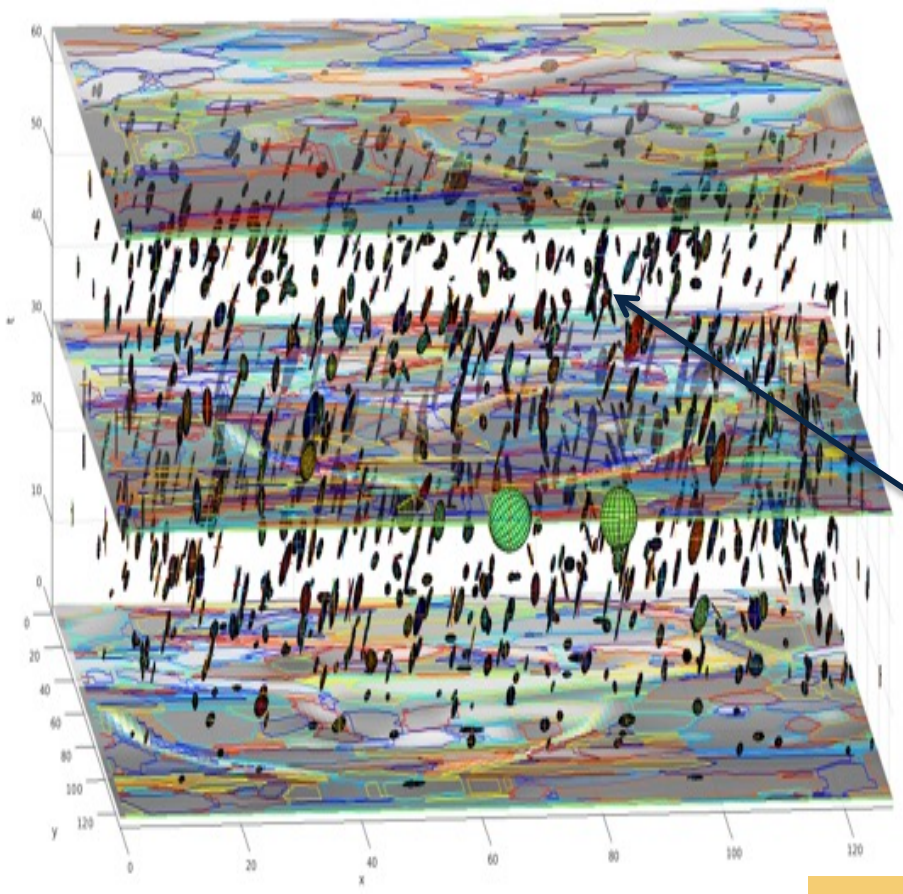
1 kernel responsible for extremely large number of pixels

And Video Compression?

(4-D GMM Model - crop with 128x128 pels x 64 frames)



Implicit Coding of Motion



265 x 265 x 64 pixels of video sequence.

In temporal direction the Kernels steer along motion of pixels.

Spatial and temporal „correlation“ is one and the same concept!!!!

No explicit motion vectors – rather spatio- temporal correlation flow!

Excellent reconstruction of edges (no blocks, no DPCM, no Transforms, no motion vectors used in this compression framework)!!!

Reconstruction of Mobile & Calendar

128 x 128 x 64 pixels

6D Gaussian Steering Kernels
- 3D Location
- 3D Color Space

Results 2016

Original H.264 Gating Network

L. Lange, R. Verhack, and T. Sikora, “Video representation and coding using a sparse steered mixture-of-experts network,” in 2016 PICTURE CODING SYMPOSIUM (PCS) , Nuremberg, Germany, 2016, pp. 1–5.

Gradient Descent Algorithm in TensorFLow

Gradient Descent training of the atom parameters greatly improves quality of reconstruction.

(a) Original crop

(b) GMM-EM, 100 components (25.59 dB)

(c) GMM-MSE, 100 components (31.83 dB)

@Erik Bochinski

6 dB Gain!!!!

9 parameters/expert

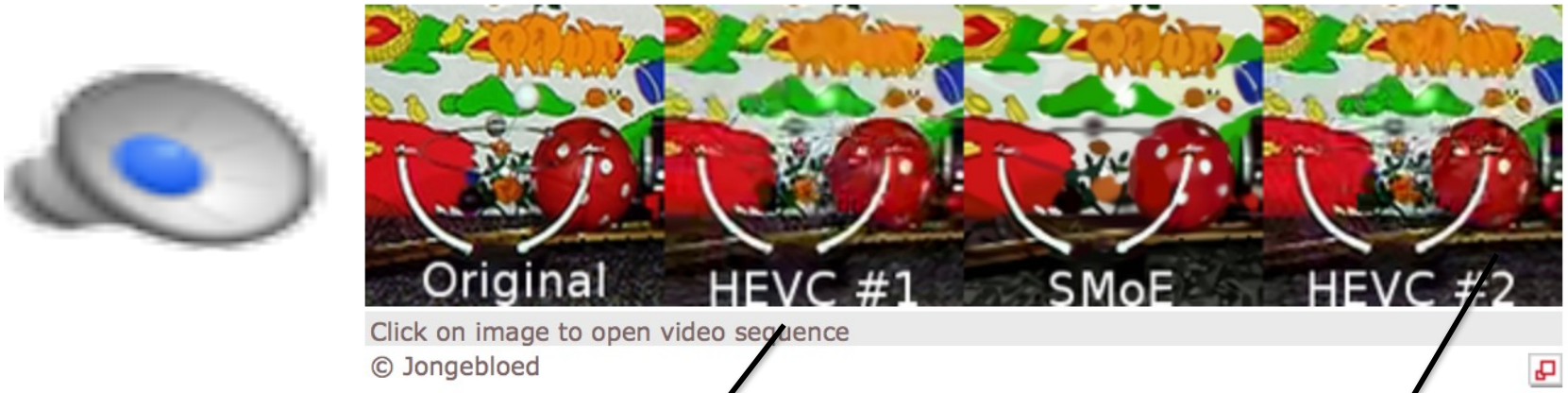
Bochinski, Erik et al. “Regularized Gradient Descent Training of Steered Mixture of Experts for Sparse Image Representation.” 2018 25th IEEE International Conference on Image Processing (ICIP) (2018): 3873-3877.

Coding of Mobile & Calendar

@Rolf Jongebloed

128 x 128 x 64 pixels

SMoE Gating Network



10% less

20% more bits than SMoE

Results for Gated Experts compression greatly improves quality of decoded video compared to HEVC !!!

Coding of Mobile & Calendar

@Rolf Jongebloed

128 x 128 x 64 pixels

SMoE Gating Network

See life video demo on our web pages:

https://www.nue.tu-berlin.de/fileadmin/fg97/04_Forschung/Coding/SMoE/mobile_hevc_vvc_smoe_same_bitrate.mp4

Coding of Mobile & Calendar

128 x 128 x 64 pixels

SMoE Gating Network

Click on image to open video sequence

$$y(\underline{x}) = \sum_{k=1}^{K=10} m_k \cdot \frac{e^{-\left(\underline{x}-\underline{\mu}_k\right)^T \underline{R}_k \cdot \left(\underline{x}-\underline{\mu}_k\right)}}{\sum_{k=1}^{10} e^{-\left(\underline{x}-\underline{\mu}_j\right)^T \underline{R}_j \cdot \left(\underline{x}-\underline{\mu}_j\right)}}$$

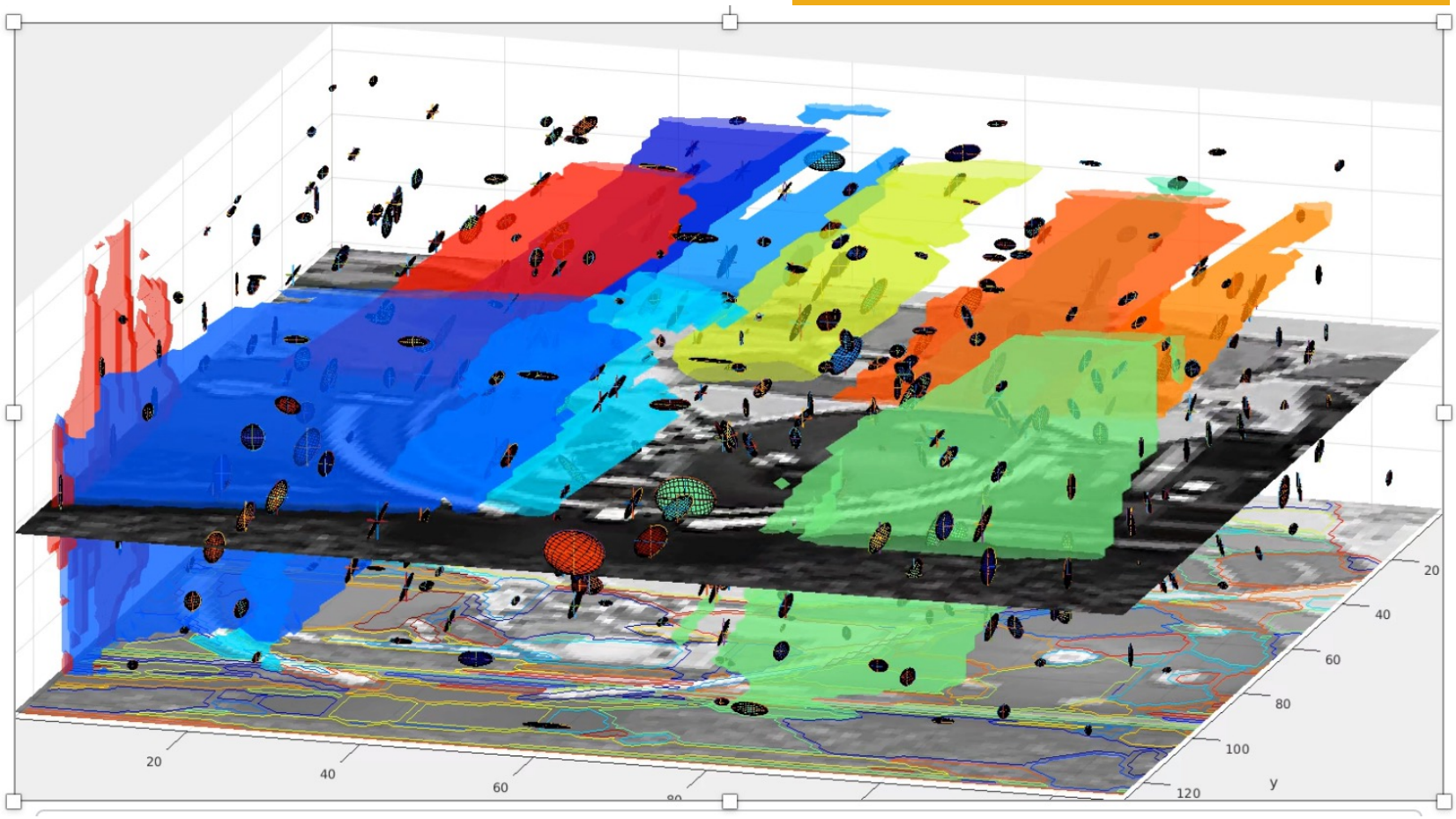
Experts

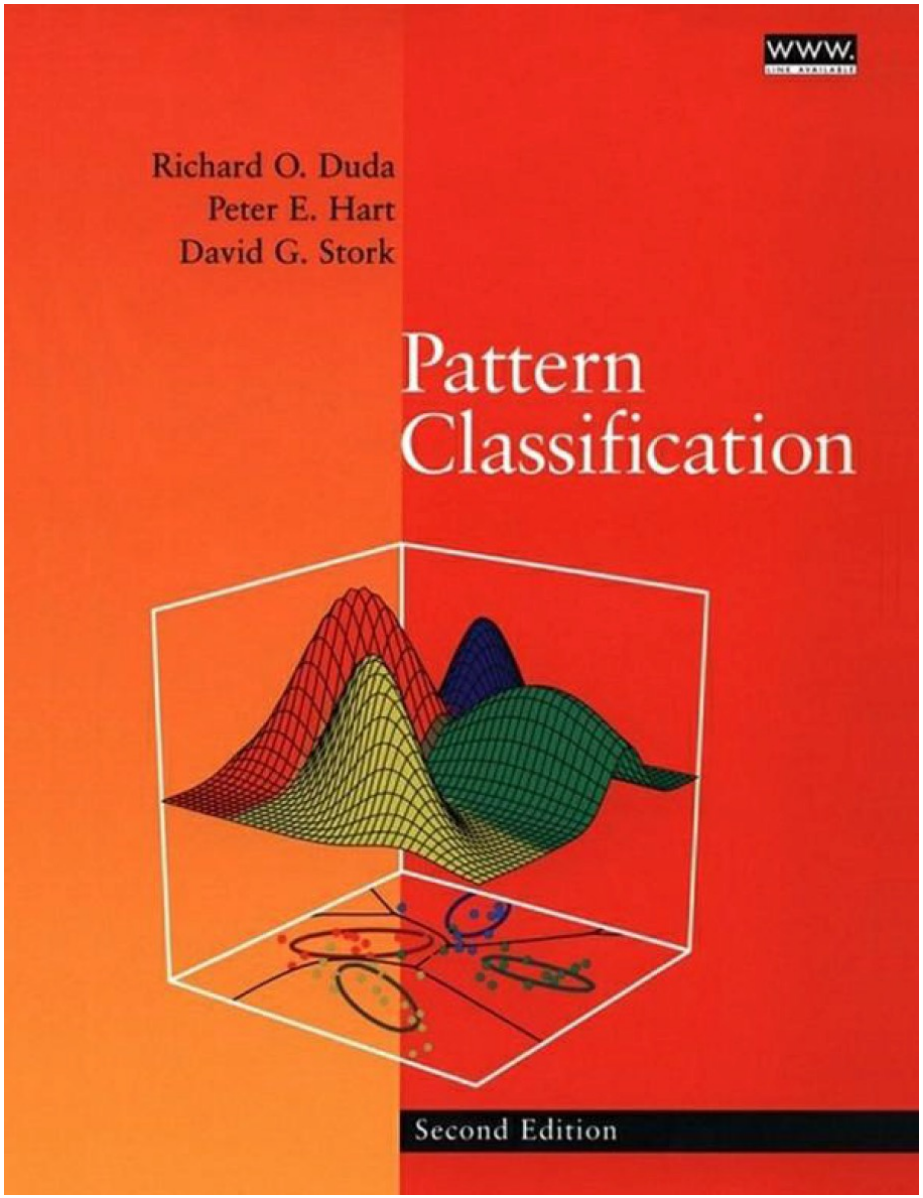
Soft-Gates

Now same simple and differentiable equation for video – but Gaussian atoms are now 3D!

Global Spatio-Temporal Support of Kernels

3D Kernels produce 3D Soft-Gates





The following illustrations are taken from chapter 2 of the book.

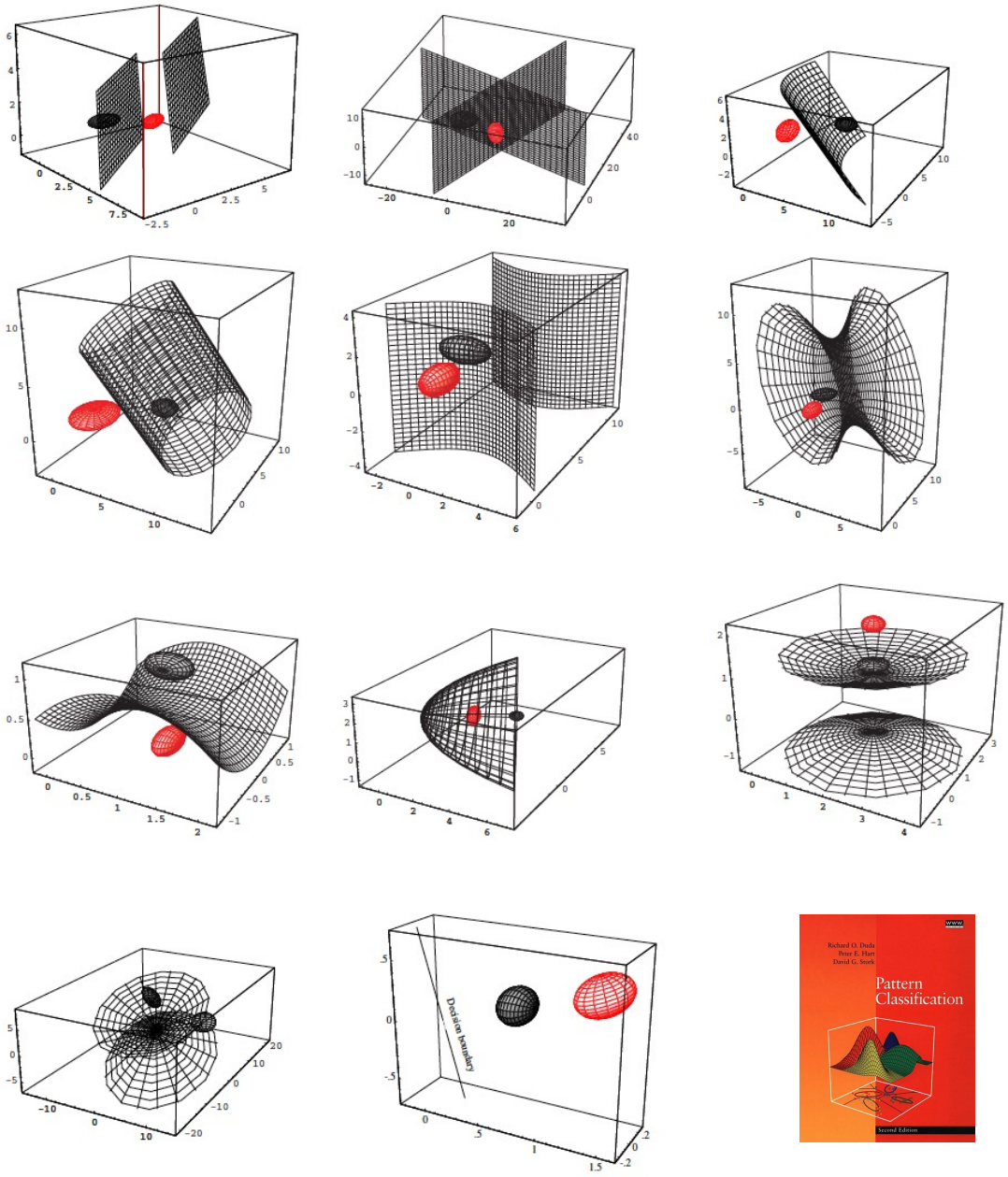
Gated-Experts Networks can be given a Bayesian interpretation if the atoms are Probability Density Functions – which is interesting to consider.

In this case decisions boundaries outlined in chapter 2 of the book, even though described for Bayesian classification, can be used to derive segmentation capabilities in images with Gated-Experts Networks.

Arbitrary three-dimensional Gaussian kernels yield two-dimensional hyper-quadratics as boundaries (3D-edges for video).

There are even degenerate classes in which" the "edge" is a straight line."

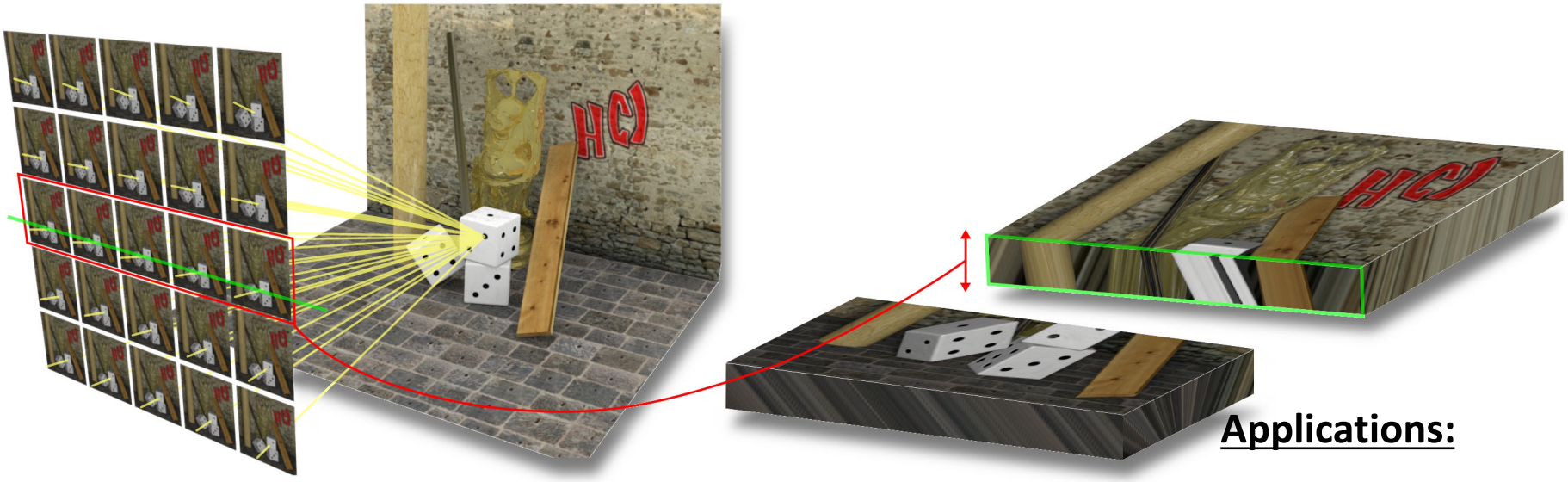
It is clear that more than 2 atoms can generate tremendously complex edge patterns.



Figures from Duda book

Lightfield Coding

And how about sparse representation of even higher dimensional pixel data? It works just the same. Same simple set of gated non-linear equations. The Gaussian atom functions are now in general N-dimensional to fit the N-dimensional pixel raster !!!



Source: Uni Thübingen

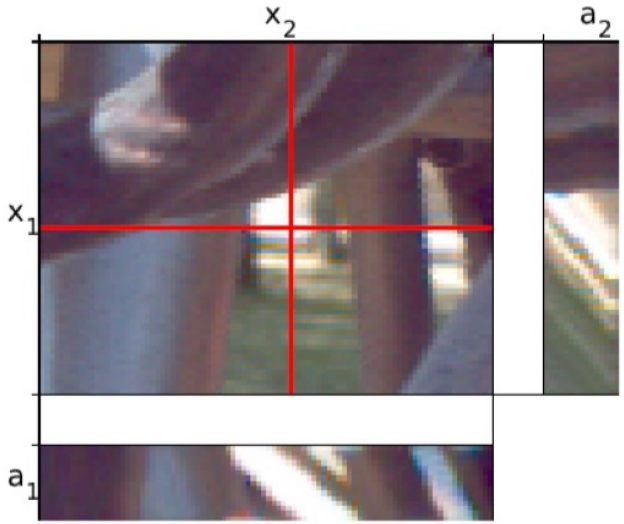
Applications:

- 3D TV
- 3D Displays
- Interactive Vision
- 3D Reconstruction

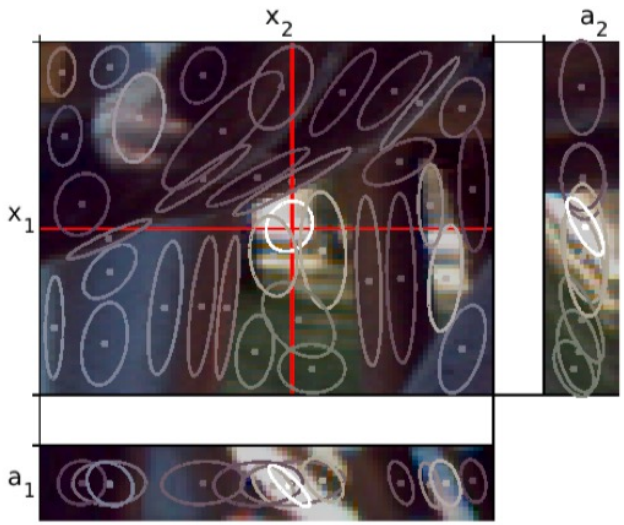
R. Verhack, T. Sikora, G. Van Wallendael and P. Lambert, **"Steered Mixture-of-Experts for Light Field Images and Video: Representation and Coding,"** in *IEEE Transactions on Multimedia*.

Best Paper Award 2021, IEEE Trans. Multimedia

4-D Pixel Grid

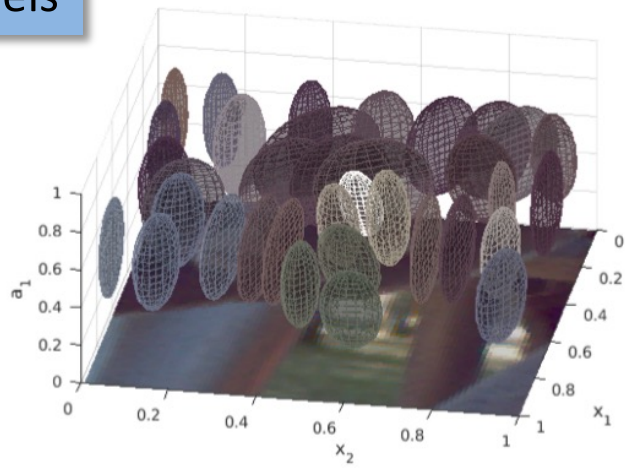


(a) Original

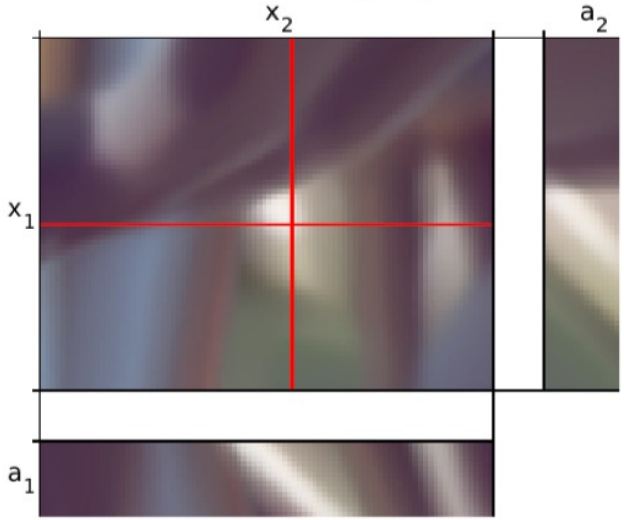


(b) GMM (4D projection)

8-D Gaussian Kernels



(c) GMM (3D projection)



(d) SMoE reconstruction

More Details in ....

REGULARIZED GRADIENT DESCENT TRAINING OF STEERED MIXTURE OF EXPERTS FOR SPARSE IMAGE REPRESENTATION

Erik Bochinski, Rolf Jongebloed, Michael Tok, and Thomas Sikora

Technische Universität Berlin
Communication Systems Group

ICIP 2018

Jongebloed, Bochinski, Sikora,"Regularized Optimization with Gradually Increasing Quantization for Coding Videos Using Steered Mixtures-of-Experts“, ICASSP 2020.

Optimization Framework in TensorFlow publicly avalaible

Get the software from our website!!!

Where from here?

The field of Gated Experts Networks for sparse representation of images is completely new – a wide and open field to explore. Both theoretically as well a from an application point of view. We believe that the approach can be adopted and of value for many fields that take advantage of sparse representations, such as:

- *Compression*
- *Denoising*
- *Segmentation*
- *Graph Signal Processing*
- *Super-Resolution (after all we have a closed-form mathematical model for the pixel data)*
- *etc*

Part of the work done with support of the „Google Faculty Research Award 2016 -in Machine Perception".