HIERARCHICAL LEARNING OF SPARSE IMAGE REPRESENTATIONS USING STEERED MIXTURE-OF-EXPERTS

Rolf Jongebloed[*]*, Ruben Verhack*[*†]*, Lieven Lange*[*]*, and Thomas Sikora*[*]

[*]Technische Universität Berlin - Communication Systems Lab, Germany
[†]Ghent University - imec, IDLab, Department of Electronics and Information Systems (ELIS), Belgium

ABSTRACT

Previous research showed highly efficient compression results for low bit-rates using Steered Mixture-of-Experts (SMoE), higher rates still pose a challenge due to the non-convex optimization problem that becomes more difficult when increasing the number of components. Therefore, a novel estimation method based on Hidden Markov Random Fields is introduced taking spatial dependencies of neighboring pixels into account combined with a tree-structured splitting strategy. Experimental evaluations for images show that our approach outperforms state-of-the-art techniques using only one robust parameter set. For video and light field modeling even more gain can be expected.

Index Terms— Steered Mixture-of-Experts, Sparse Representation, Hidden Markov Random Field, Denoising, Image Signal Processing, Inference

1. INTRODUCTION

The *Steered Mixture-of-Experts* (SMoE) framework which has been introduced in [1] for coding of still images yields a compact sparse representation. This allows for very efficient compression as only the parameters of a Mixture Model need to be stored. Furthermore, easy access to MPEG-7-like low- and mid-level image features on bit-level are provided [2] which can be used for image processing tasks, e.g. classification and comparison. The unifying vision incorporated higher dimensional image modalities, in [3] and [4] the framework has been extended to video and light field coding, respectively. In all of these works [1, 3, 4] it is shown that state-of-the-art coding results for low bit-rates can be achieved. Unfortunately, the higher the bit-rate, the more parameters need to be estimated and thus rendering the optimization problem more non-convex. Consequently, the estimation of these parameters becomes a crucial part within the entire framework pipeline. This work proposes a novel estimation approach for the case of still images. Note that the proposed method is extendable to video and light field modeling for which more potential gain can be achieved as the sparse continuous representation does not suffer from the *curse of dimensionality*, in contrast to traditional discrete dense representation techniques.

Mixture-of-Experts (MoE) approaches follow the divide-and-conquer principle [5]. Each expert acts as a regression function weighted by a gating function. This achieves soft partitioning of the input space to determine in which regions the experts are trustworthy. In the SMoE framework the alternative model is used [5] in which the parameters of the MoE coincides with the parameters of a *Gaussian Mixture Model* (GMM) that models the joint probability density of input and output variables. The GMM parameters are found by maximizing the joint likelihood of the input and output space using the *Expectation Maximization* (EM) algorithm [6]. As a result, each component steers along the direction of highest correlation. However, the more kernels are used for modeling, the more important is the initialization of their parameters. The resulting GMM trained by the EM algorithm is highly dependent on the initialization. EM easily becomes trapped in one of the local maxima of the likelihood function [7]. Therefore, a lot of effort has been made before to overcome these limitations, e.g. in [8]. The authors proposed a Split-and-Merge approach in which Split-and-Merge candidates are searched under certain criteria after the usual EM algorithm has converged. Each candidate tuple has to be trained until convergence and will be accepted in case the likelihood has increased. In [7] a greedy algorithm is proposed which produces a sequence of GMMs with increasing number of kernels. For adding a new component the algorithm generates several candidates and chooses the one with the highest likelihood. In general, such techniques result in well-trained GMMs but optimization is very time-consuming for large numbers of components such as in our case.

In [1], the initialization problem is mitigated by dividing the image in blocks of equal size. Each block is modeled by the EM algorithm independently using a different amount of components determined by the so-called *Spatial Activity Analysis* based on a 2D-DCT. Additionally, between 5 and 10 split-and-merge iterations as in [8] have been performed in each block. Nevertheless, the distribution of the kernels remains inadequate and the block-division creates block artifacts. As a consequence, the maximum potential of the model remains unexploited.

In this paper we propose a novel approach to infer these pa-

978-1-5386-1737-3/18/$31.00

rameters to guarantee that regions with high spatial activities get more attention by having a higher density of experts. The main idea of this technique is to start the modeling with few components and train them until convergence. The resulting GMM is considered as a root node within a tree-structured splitting strategy. After convergence, the weighted sum squared error (WSSE) is computed for each component to determine which component needs to be split. Further, the training data is split into several segments which boundaries are defined by the highest influence of each expert. Thus, several training subsets emerge that are used to train new sub-GMMs independently and considered as child nodes of the root node. This procedure will be repeated for all nodes until no component fulfills the split criterion. This ensures that regions with high spatial activity are split in more distinct areas represented by more components.
Additionally, in each node a *Hidden Markov Random Field* (HMRF) is assumed to take the neighborhood of the pixels into account. HMRF are frequently used for several computer vision tasks such as image segmentation [9] and depth inference [10]. After the building of the tree is converged, the resulting GMM is extracted as initialization of the standard EM algorithm [6].

2. STEERED MIXTURE-OF-EXPERTS

2.1. Introduction

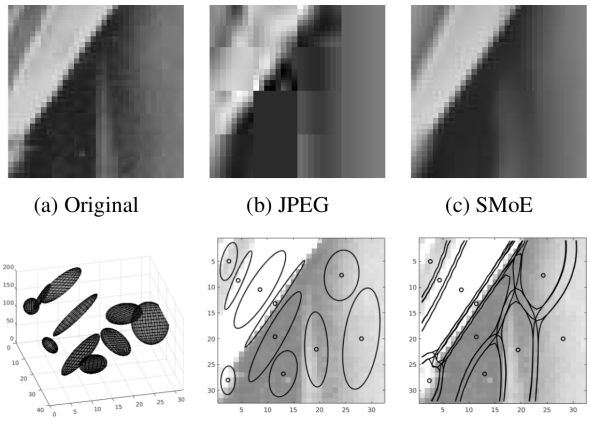
The goal in the *Steered Mixture-of-Experts* (SMoE) approach is to predict the expected amplitude of a pixel given the location of the pixel. This underlying stochastic process is modeled as a multimodal and multivariate Mixture Model with K modes. In such stochastic processes *Gaussian Mixture Models* (GMM) are frequently used for modeling.
The joint pdf $p_{XY}(x, y)$, x being the location and y being the amplitude, is:

$$p_{XY}(x, y) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x, y \,|\, \mu_k, \Sigma_k) \tag{1}$$

where the parameters are π_k, μ_k and Σ_k, respectively being the mixing proportions (or priors), means and covariances with
$\sum_{k=1}^{K} \pi_k = 1, \mu_k = \begin{bmatrix} \mu_{X,k} \\ \mu_{Y,k} \end{bmatrix}, \Sigma_k = \begin{bmatrix} \Sigma_{XX,k} & \Sigma_{XY,k} \\ \Sigma_{YX,k} & \Sigma_{YY,k} \end{bmatrix}$.

The estimation of these parameters can be done by the *Expectation Maximization* (EM) algorithm assuming the image training data $D = \{x_i, y_i\}_{i=1}^{N}$.
Through the conditional pdf $Y|X$ [1]

$$p_{Y|X}(y|x) = \sum_{k=1}^{K} w_k(x) \mathcal{N}(x \,|\, m_k(x), \sigma_k^2) \tag{2}$$



(a) Original (b) JPEG (c) SMoE

(d) Mixture model (e) Top view (f) Softmaxed

Fig. 1: An example of the modeling with 10 components and reconstruction of a 32x32 pixel crop from *Lena*. (Source: [1])

with $m_k(x)$ being the expected value of the conditional, $w_k(x)$ being the conditional mixing proportions and σ_k^2 being the conditional variance defined as follows:

$$m_k(x) = \mu_{Y,k} + \Sigma_{YX,k} \Sigma_{XX,k}^{-1} (x - \mu_{X,k}) \tag{3}$$

$$w_k(x) = \frac{\pi_k \mathcal{N}(x \,|\, \mu_{X,k}, \Sigma_{XX,k})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x \,|\, \mu_{X,j}, \Sigma_{XX,j})} \tag{4}$$

$$\sigma_k^2 = \Sigma_{YY,k} - \Sigma_{YX,k} \Sigma_{XX,j}^{-1} \Sigma_{XY,k} \tag{5}$$

The regression function defined as the expected amplitude y given the location x can be derived:

$$E[Y|X = x] = m(x) = \sum_{k=1}^{K} w_k(x) m_k(x). \tag{6}$$

The reconstructed value at location x is a weighted sum over all K experts. As such, all experts collaborate towards the definition of the regression function. The conditional mixing weights in Eq. 4 serve as the gating function as in [5] and ensure global support of experts which are defined by Eq. 3.

The compression capability of SMoE is shown in Fig. 1 with a 32x32 crop of *Lena* compared to JPEG, each at $0.35\,\mathrm{bit/sample}$. In SMoE the edges are reconstructed with excellent quality. Fig. 1(d) depicts the Mixture Model which contains the steered 3-D ellipsoid Gaussian components. They define the experts m_k as 2-D steering planes for regression. The top view in Fig. 1(e) which shows the components projected onto the 2-D pixel domain illustrates how the kernels steers along edges. Through the softmaxed gating weights the steering planes are windowed which can be seen in Fig. 1(f). The emerging windows are of arbitrary shape.

2.2. Hidden Markov Random Field Model

Given the image training data $D = \{x_i, y_i\}_{i=1}^{N}$ the inference of a configuration of classes $\boldsymbol{Z} = \{z_i\}_{i=1}^{N}$ with $z_i \in$

$\{1, \ldots, K\}$ is done by accomplishing the Maximum a Posteriori (MAP) criterion [11]:

$$\boldsymbol{Z}^{\star} = \arg\max_{\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}} \{P(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z}, \Theta) P(\boldsymbol{Z})\} \tag{7}$$

where $\boldsymbol{Z}$ is arranged as a field of same size as the image and the prior probability $P(\boldsymbol{Z})$ is a Gibbs distribution. The joint likelihood probability in Eq. 7 is

$$\begin{aligned} P(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z}, \Theta) &= \prod_{i=1}^{N} P(x_i, y_i | \boldsymbol{Z}, \Theta) \\ &= \prod_{i=1}^{N} P(x_i, y_i | z_i, \theta_{z_i}) \end{aligned} \tag{8}$$

where $P(x_i, y_i | z_i, \theta_{z_i})$ are multivariate Gaussian distributions with parameters $\Theta = [\theta_1, \theta_2, \ldots, \theta_K]$ with

$$\theta_j = (\mu_j, \Sigma_j), \tag{9}$$

respectively being the means and covariances of the Gaussian distributions. As the prior probability can be written as:

$$P(\boldsymbol{Z}) = \frac{1}{Z} \exp\left(-\beta U(\boldsymbol{Z})\right) \tag{10}$$

and, respectively, the Gaussian probability as:

$$P(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z}, \Theta) = \frac{1}{Z'} \exp\left(-U(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z})\right), \tag{11}$$

where Z and Z' are a normalization factors and $U(\boldsymbol{Z})$ and $U(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z})$ are energy functions, the MAP criterion can be reformulate to a minimum problem:

$$\boldsymbol{Z}^{\star} = \arg\min_{\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}} \{U(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z}, \Theta) + \beta \cdot U(\boldsymbol{Z})\} \tag{12}$$

where the likelihood energy with given $\boldsymbol{Z}$ and Θ is:

$$\begin{aligned} U(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{Z}, \Theta) &= \sum_{i=1}^{N} U(x_i, y_i | z_i, \Theta) \\ &= \sum_{i=1}^{N} \Bigg[\frac{([x_i, y_i]^T - \mu_{z_i})^T \Sigma_{z_i}^{-1} ([x_i, y_i]^T - \mu_{z_i})}{2} \\ &\qquad + \ln(\sqrt{|\Sigma_{z_i}|}) \Bigg]. \end{aligned} \tag{13}$$

The prior energy function $U(\boldsymbol{Z})$ can be formulated as follows:

$$U(\boldsymbol{Z}) = \sum_{c \in C} V_c(\boldsymbol{Z}), \tag{14}$$

where $V_c(\boldsymbol{Z})$ is the clique potential and C is the set of all possible cliques. The parameter β weights the neighborhood of the pixels.
The clique potential is defined as follows:

$$V_c(z_i, z_j) = \frac{1}{2}(1 - \delta_{z_i, z_j}) \tag{15}$$

where δ_{z_i, z_j} is the Kronecker Delta function. The inference of the parameters of the model is done by the HMRF-EM as in [11].

2.3. The Tree-Structured Approach and Split Criterion

The main idea of the proposed method is to start the modeling with few components and split that component which produces the highest amount of prediction error. After convergence of HMRF-EM algorithm, the weighted sum squared error (WSSE) is used to determine which component has to be split. The WSSE is calculated as follows:

$$\text{WSSE}_k = \sum_{i=1}^{N} w_k(x_i) \cdot \left(m_k(x_i) - y_i\right)^2. \tag{16}$$

The component having the heighest WSSE is split:

$$k_{\text{Split}} = \arg\max_{k} \left\{\text{WSSE}_k\right\}. \tag{17}$$

The resulting new components will only be trained with a subset D_l defined through Eq. 12. Only samples that are labeled to the k_{Split}-th component belong to the subset:

$$D_l = \{x_i, y_i | z_i = k_{\text{Split}}\}_{i=1}^{N}. \tag{18}$$

As a result, the modeling of the new components can be considered as a sub-model also trained by the HMRF-EM.
After the sub-model is converged the WSSE for each component will be calculated again to determine the next split candidate. Note that the calculation of the WSSE is done over the entire image training data D and all existing components. The procedure results in a tree-structured approach which can be seen in Fig. 2 which illustrates the split algorithm on a 32x32 crop of *Cameraman*. The amount of new components within the sub-model l depends on how many unconnected regions within the subset D_l exists (as shown in Fig. 2). The component marked in green within the root node is split in 5 new components modeled in its child node as its subset D_l consists of 5 unconnected regions. The minimum amount of new components is two.
This procedure is repeated as long as the current split candidate has a $\text{WSSE}_{k_{\text{Split}}}$ higher than a predefined threshold θ_{th}. The aforementioned steps are summarized in Alg. 1.

Algorithm 1: Our Proposed Method

1 Create root node with few components
2 Define subset $D_l = D$
3 Run HMRF-EM to model the new components using subset D_l until convergence
4 Determine split candidate k_{Split} (Eq. 17)
5 **if** $WSSE_{k_{Split}} > \theta_{th}$ **then**
6 Define subset D_l (Eq. 18)
7 Add new components
8 Go to Step 3
9 **else** Run EM using D to model all components

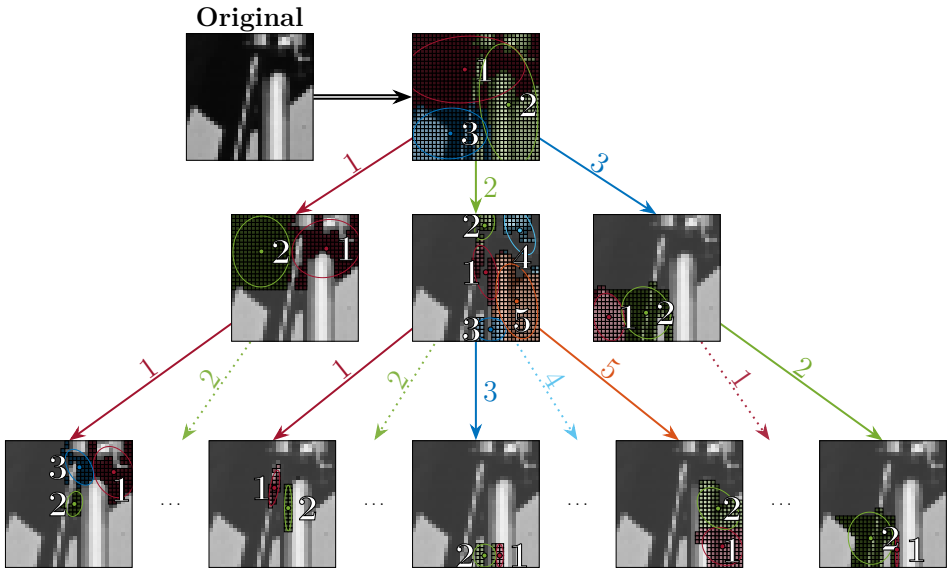


Fig. 2: Visualization of the split algorithm on a 32x32 example. The modeling starts with three kernels trained with all samples. The coloring of samples represents the membership to the kernels defined by Eq. 12. New kernels are only trained with that membership assigned to its parent kernel and thus only build sub-memberships within a sub-model. Samples which are not used in the corresponding sub-model are smoothly faded out. Dotted arrows point to non-plotted sub-models due to space limitations. Note that the numbering of components is done in each node independently.

3. EXPERIMENTS AND RESULTS

3.1. Image Approximation

Exhaustive experimental evaluations have been made to determine the neighboring weighting β as mentioned in Subsection 2.2 which led to $\beta = 2$ giving the best results. These experiments have shown that low-texturized regions are sparsely populated by kernels while in regions with high spatial activities of arbitrary shape are densely represented by more components. Fig. 3 illustrates one of these results on a 256x256 crop of *Lena* (top right) as a kernel density heat map. In this example the modeling was started with 5 components and run until $K = 400$ components were reached. Obviously, kernels are densely located at the feathers. In contrast to the background (top-left corner), which is very sparsely represented as its content is very low. For comparison a model with the same amount of kernels also illustrated as a kernel density heat map is shown in Fig. 3 (top left). This model is initialized by the K-means++ algorithm [12], which is a common technique to initialize the centers of GMMs [13]. One can see that the kernels in that model are approximately uniform distributed which leads to an overrepresented background and to a underrepresented foreground such as the feathers and the eye in Fig. 3.
Consequently, this is also reflected within the corresponding reconstruction images (Fig. 3, bottom row). The proposed method is able to reconstruct the feathers with much more details. This visual gain can be also seen in the eye. In contrary, the K-means++ initialized reconstruction suffers from some visual artifacts. An exception is the structures of the hat. As the kernel density is still sparse at this location in our case, it is not able to reconstruct these structures appropriately. Note that these structures will be reconstructed if the split algorithm runs further towards more kernels.
The quality gain is also reflected in objective metrics as our method is able to achieve $27.34\,\mathrm{dB}$ while the K-means++ initialized approach achieves $25.72\,\mathrm{dB}$ at the same amount of components. Note the more kernels are used, the more is the visual as well as the objective gain.

3.2. Denoising Application

In image processing tasks and coding, it is commonly a requirement to be robust against noise corruption. Therefor, it is beneficial to analyze the robustness of our modeling.
We can show that our approach is on one hand robust against noise corruption and on the other hand even provides the ability of denoising which is shown in Fig. 4. A 256x256 crop of *Peppers* has been corrupted by additive white Gaussian noise with standard deviation $\sigma = 20$ (see Fig. 4, left). Our approach modeled with $K = 1100$ is still able to reconstruct the original image quite well (see Fig. 4, middle). Compared to the noisy images it yields gains in PSNR as well as in SSIM values up to $8\,\mathrm{dB}$ and 0.434, respectively. The higher the standard deviation of the noise is, the higher is the

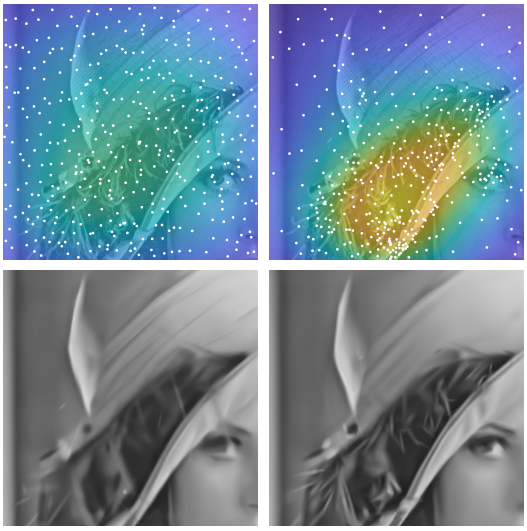


Fig. 3: Comparison of models with $K = 400$ (upper row) depicted as kernel density heat map and reconstruction images (lower row) with a 256x256 crop of *Lena*: K-means++ initialized (left), proposed (right)

22.18 dB / 0.357 30.16 dB / 0.791 31.99 dB / 0.833

Fig. 4: Denoising of *Peppers* corrupted by noise with $\sigma = 20$ (left) by proposed with $K = 1100$ (middle) and by BM3D [14] (right). The two numbers under each image indicates the corresponding PSNR and SSIM values.

gain we can achieve. Fig. 4 shows a result also for a state-of-the-art denoising algorithm for comparison (right). The *Block-matching and 3D filtering* (BM3D) algorithm [14] outperforms our algorithm. However, BM3D needs the information about the standard deviation of the additive noise for yielding these results while our method acts "blind". It is possible to incorporate noise structure information into our SMoE approach to achieve further gains and to reduce remaining speckle artifacts. A possible path may be to use the irregular sampling structure for processing the kernels on a graph representation. This will be subject to future work.

3.3. Modeling Experiments

As mentioned before, the main goal of this work is to develop a modeling technique which is able to estimate model parameters exploiting best possible potential with one robust parameter set. For modeling experiments the image is divided into

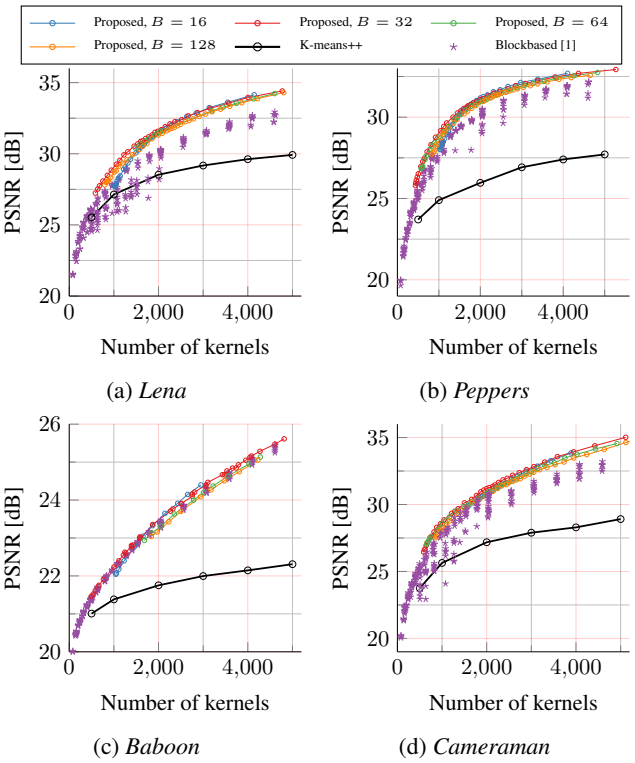


Fig. 5: Above plots illustrate the consistent PSNR gains of our proposed method compared to the state-of-the-art and standard K-means++ initialization.

blocks of equal size, similar to [1]. On each block the proposed modeling approach is done independently. In contrary to [1], the amount of kernels per block is determined through the predefined threshold θ_{th}. This enables that high-texturized regions are represented by more components to underrun the split criterion. Furthermore, using the tree-structured splitting strategy as proposed the kernels are already well initialized in contrast to initializing with K-means++ per block [1]. After each block was modeled independently, the EM algorithm has been used globally over all blocks until convergence.

Due to computational complexity blocksizes of $B = \{16, 32, 64, 128\}$ and the neighborhood weighting $\beta = 2$ were used. The predefined threshold θ_{th} which implicitly determines the total amount of kernels varied in the range of $[0.15, 8]$. The minimum amount per block is two. Note that the WSSE is calculated with normalized amplitude values.

We compare our results with models which were trained by the block-based technique as in [1] and additionally with models which were trained globally by the EM algorithm initialized by K-means++. Fig. 5 depicts curves of reconstruction qualities in PSNR over the total numbers of kernels for test images *Lena*, *Peppers*, *Baboon* and *Cameraman*, each in 512x512 pels. While the naïve modeling with the global EM algorithm initialized by K-means++ is not competitive at

all with increasing number of kernels, one can see that our approach outperforms the modeling as in [1] in most cases. In general, a considerable gain is achieved using blocksize $B = 32$ which provides the overall best results. For *Lena* a gain up to approximately $1.46\,\mathrm{dB}$ is yielded. An exception is *Baboon*. Due to its predominately high frequency content the potential for gain might be reasonable for $K > 10000$. Nevertheless, our results for *Baboon* are slightly better than the best results from [1] with $B = 32$. The experiments in [1] were done on a range of blocksizes and spatial activity sensitivity values to determine the optimal parameter set for each total amount of kernels independently. Note that we need to fix ours to only one set to outperform these results.

4. CONCLUSIONS AND FUTURE WORK

We presented a novel estimation scheme of SMoE models for still images. By introducing combining HMRFs and a tree-splitting approach, we have shown a significant gain in reconstruction quality. Since the optimization becomes increasingly non-convex as the number of parameters rises. As such, the estimation within SMoE framework becomes an essential part as it is responsible for the reachable quality at high bitrate cases. By using the splitting strategy it is ensured that high-textured regions are more densily populated by kernels while the sparsity in flat regions is maintained with merely one parameter set. This results in improved reconstruction qualities with the same amount of components compared to other techniques. Moreover, using the proposed estimation scheme SMoE provides the capability of denoising noise-corrupted images in a "blind" manner yielding considerable results. The denoising competence could still be further improved in the future by incorporating the standard deviation of the additive noise.

The estimation scheme is extendable towards video and light field modeling. In this case, the correlation of neighboring pixels in time and disparity, respectively, is sharply higher than in spatial dimensions, the weighting neigboorhood parameter needs to be configured for each dimension individually.

5. REFERENCES

[1] R. Verhack, T. Sikora, L. Lange, G. Van Wallendael, and P. Lambert, "A universal image coding approach using sparse steered Mixture-of-Experts regression," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 2142–2146.

[2] T. Sikora, "The MPEG-7 Visual Standard for Content Description - An Overview.," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 11, no. 6, pp. 696–702, 2001.

[3] 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)*, Dec 2016, pp. 1–5.

[4] R. Verhack, T. Sikora, L. Lange, R. Jongebloed, G. Van Wallendael, and P. Lambert, "Steered mixture-of-experts for light field coding, depth estimation, and processing," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, July 2017, pp. 1183–1188.

[5] S. E. Yuksel, J. N. Wilson, and P. D. Gader, "Twenty Years of Mixture of Experts," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 8, pp. 1177–1193, Aug 2012.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[7] J. J. Verbeek, N. Vlassis, and B. Krse, "Efficient Greedy Learning of Gaussian Mixture Models," *Neural Computation*, vol. 15, no. 2, pp. 469–485, 2003.

[8] Z. Zhang, C. Chen, J. Sun, and K. L. Chan, "EM algorithms for Gaussian mixtures with split-and-merge operation.," *Pattern Recognition*, vol. 36, no. 9, pp. 1973–1983, 2003.

[9] L. Zhang and Q. Ji, "Image Segmentation with a Unified Graphical Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1406–1425, Aug 2010.

[10] A. Saxena, S. H. Chung, and A. Y. Ng, "3-D Depth Reconstruction from a Single Still Image," *Int. J. Comput. Vision*, vol. 76, no. 1, pp. 53–69, Jan. 2008.

[11] Q. Wang, "HMRF-EM-image: Implementation of the Hidden Markov Random Field Model and its Expectation-Maximization Algorithm," *CoRR*, vol. abs/1207.3510, 2012.

[12] D. Arthur and S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, USA, 2007, SODA '07, pp. 1027–1035, Society for Industrial and Applied Mathematics.

[13] J. Blömer and K. Bujna, "Simple Methods for Initializing the EM Algorithm for Gaussian Mixture Models," *CoRR*, vol. abs/1312.5946, 2013.

[14] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *Trans. Img. Proc.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.