COMPONENT-BASED IMAGE CODING USING NON-LOCAL MEANS FILTERING AND AN AUTOREGRESSIVE TEXTURE MODEL

Johannes Ballé, Bastian Jurczyk, and Aleksandar Stojanovic

Institut für Nachrichtentechnik
RWTH Aachen University
Aachen, Germany

ABSTRACT

While noise is usually regarded as a problem of the image formation process, we observe that it is also frequently part of natural texture. In this paper, we present a concept for improved compression of noisy texture in natural images. Since noise is problematic to decorrelation-based compression methods, we propose to perform image decomposition by denoising, followed by separate compression of the components. The denoised component is encoded using conventional methods, while the texture component is compressed by encoding parameters of a texture model. It turns out that at similar bit rates, our method can improve visual quality.

Index Terms— image coding, noise modeling, compression by synthesis, multi-layer image model

1. INTRODUCTION

One observation that can be made about current production-class video coding techniques is that they are all designed to optimize pixel fidelity. The quality metrics used by the overwhelming majority of proposed coding schemes is the peak signal-to-noise ratio (PSNR), although it is well-known that this approach has weaknesses. For instance, certain texture in natural images need not be reconstructed exactly in order to be transparent. This is evident from results in work on texture synthesis methods such as [1], suggesting that it could be practical to synthesize visually similar texture instead. It can be expected that exploitation of this fact, if feasible, results in major improvements of coding efficiency for certain types of signals.

Recently, proposals have been made to integrate texture synthesis algorithms into frameworks for image and video coding [2, 3, 4, 5]. These schemes involve texture analysis, segmentation, and classification of the signal into "structure" and "texture" parts. The reconstruction is performed by conventional decoding of the structure segments and synthesis of the texture segments. The employed synthesis methods [6, 7] basically constitute pixel- or patch-based "inpainting" of the segments identified as texture. However, a lot of care has to be taken to perform the segmentation and classification in a *conservative* way so as not to subject the wrong parts of the signal to synthesis, resulting in visually implausible results. Most authors respond to this problem by constraining the classification and synthesis processes. In this context, it is important to consider the "semantic gap" [8]: identifying semantically irrelevant texture is and remains a difficult problem, because it partly depends on the viewer's individual knowledge and interpretation of the presented image or video sequence. Hence, unless the problem of image understanding is solved, it is probably necessary for general-purpose methods to restrict synthesis to textons which appear on a small scale in order to minimize possible semantic interpretation.

In this paper, we propose an alternative concept which specializes on "noisy" texture, which is characterized by very small-scale textons. The scheme is based on recent advances in denoising techniques, specifically, on the non-local type of processes [9]. Considering the difference signal between an image signal and the output of the denoising process, we can make two observations:

Firstly, the algorithm can be adjusted such that the signal contains almost exclusively structureless noise. This has been noted in [9] (the difference signal is called "method noise" by the authors). Secondly, "regular" texture, i.e., recurring texture features, are preserved due to the filtering across similar pixel neighborhoods. Considering that the process filters not only noise resulting from the imagery equipment, but also "noise" inherent in certain types of texture, we can use this kind of denoising process to perform a *decomposition* of the signal into "noisy texture" and structure. Thus, we propose decomposition as a means to distinguish between the two components of the signal (as opposed to segmentation and classification), followed by *component-based* coding.

Theoretically, this approach has three advantages: Firstly, the signal subjected to conventional coding still contains structure such as edges and recurring patterns, but has decreased entropy [10]. We expect that this results in more efficient prediction of repetitive elements, which can be implemented as an extension to conventional coding methods [11]. Secondly, for modeling and synthesis of the noise component at the decoder, autoregressive (AR) processes can be used, which constitute a well-known model for this

type of signal and are analogous to established models for audio coding [12]. Thirdly, the desired level of "exactness" of the reconstruction signal can be controlled by altering the strength of the denoising filter, trading off signal energy between the two components. We expect that the avoidance of hard segmentation decisions might also have a positive effect on the number of visible artifacts. For instance, no special treatment of colour variations over regions is needed, as mentioned in [5].

In the remaining sections, we present a possible coding scheme that implements component-based coding of still images. The denoising algorithm is briefly reviewed in Section 2. The noise model and parameter coding scheme is presented in Sections 3 and 4, respectively. Results are shown in Section 5. Section 6 concludes the paper.

2. DECOMPOSITION

The Non-Local Means (NLM) algorithm [9] estimates pixel values of the original, uncorrupted image based on a weighted average of pixel values with similar neighborhoods. For each pixel position x in the image u, the estimated pixel value is defined as

$$\hat{u}(x) = \frac{1}{C(x)} \sum_{y \in \mathcal{W}} w(x,y) \cdot u(y) \tag{1}$$

with $\mathcal{W}$ representing a search window given as a set of displacement vectors y and a normalizing factor $C(x) = \sum_{y \in \mathcal{W}} w(x,y)$ chosen such that the sum of weights amounts to 1. The weights are defined by

$$w(x,y) = \exp\left(-\frac{\|u_{\mathcal{N}}(x) - u_{\mathcal{N}}(x+y)\|_G}{2h^2} \right) \tag{2}$$

with $u_{\mathcal{N}}(x)$ representing the vectorized pixel values of a neighborhood $\mathcal{N}$ centered at position x. $\|\cdot\|_G$ denotes a weighted Euclidean norm such that the positions in $\mathcal{N}$ are weighted according to a 2D Gaussian mask G. h acts as the filtering parameter controlling filter strength. Throughout the rest of this paper, we shall refer to $\hat{u}$ as the structure component and to $n = u - \hat{u}$ as the noise component.

3. NOISE MODEL

The model used to parameterize the noise component is equivalent to a piecewise autoregressive process with an additional external term (ARX). It can be described using the following recursive equation, using the notation from above:

$$n(x) = \varepsilon_{\sigma(x)}(x) + A^T(x) \cdot n_{\mathcal{N}_A}(x) + X^T(x) \cdot \hat{u}_{\mathcal{N}_X}(x) \tag{3}$$

In this equation, the first is a Gaussian, zero-mean innovation term. The second and third term correspond to linear dependencies between the pixels of the noise component and its neighbors and between the noise component and the structure component, respectively. In order to allow synthesis using a single iteration over the image space, $\mathcal{N}_A$ is restricted to a non-symmetric half plane (NSHP) mask.

Since the model parameters need to be encoded as side information, they need to be quantized. In this proposal, the image space is partitioned into square blocks with model parameters for each block assumed constant. The parameters $A(x)$ and $X(x)$ are subjected to a vector quantization scheme analogous to certain speech coding schemes [12], while $\sigma(x)$ is linearly quantized. Finally, centroid vectors A_k, X_k are linearly quantized, as well. The estimation of all parameters is performed simultaneously with the vector quantization step. A detailed treatment of the algorithm can be found in [13].

4. CODING SCHEME

In the proposed coding scheme (Fig. 1), the structure component is encoded using the same method which is used for the reference codec, the JasPer implementation of the JPEG2000 standard [14]. The noise component is encoded using a separate scheme elaborated upon below. We rely on JasPer's rate control mechanism in order to arrive at the same bit rate for the reference codec and our scheme. The quality of the reconstruction can only be assessed subjectively since evaluation of the PSNR would not yield meaningful results.

For entropy coding, we use an independent implementation of the Context-Adaptive Binary Arithmetic Coding (CABAC) engine [15], libcabac. The model parameters fed to the entropy coding step are: $B(i)$ (VQ indices), A_k, X_k (ARX model coefficients), and $\sigma(i)$ (ARX model mean prediction error). Here, i represents a block position and k the VQ codebook index.

The encoding of $B(i)$ employs a predictive scheme requiring two CABAC contexts, CTX_B_PRED_PLANE and CTX_B_PRED_SEL. Blocks are scanned row by row, and information from the block to the left (A) and above (B) of the current block are used for prediction of the VQ index. If A and B are available and share the same VQ index, a bit is encoded using CTX_B_PRED_PLANE to signify whether the index should be copied to the current block. Otherwise, if any of A and B are available, the bit is encoded using CTX_B_PRED_SEL, and if the prediction is to be used but the predictors from A and B are different, a further bit is encoded using the bypass engine to select which one is to be used. If the predictors are not used or neither A nor B are available, the VQ index is encoded using a fixed-length binarization and the bypass engine.

The linear quantization indices of A_k and X_k are encoded using an Exp-Golomb binarization and the bypass engine. Since the parameters contain sign information, the quantization intervals are centered on zero and on multiples of the quantization step size.

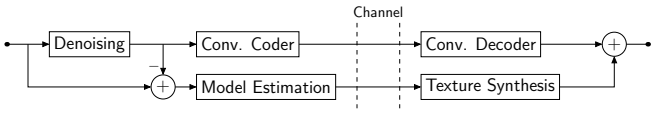

Fig. 1. Overview of Proposed Coding Scheme

$\sigma(i)$ is also binarized using an Exp-Golomb code, although the first four bits are encoded using separate contexts. The contexts are switched depending on the VQ index of the block. Hence, the number of contexts used amounts to four times the codebook size. For $\sigma(i)$, no sign information is needed and the quantization intervals run between zero and multiples of the quantization step size. The reconstruction value of the zeroth quantization index is moved from the center of the interval to zero, effectively leading to a dead-zone quantizer. This is a requirement since it must be possible to reconstruct regions completely without noisy texture.

On the decoder side, first the structure component is decoded using JasPer, then the model parameters are decoded and used to synthesize the noise component. Finally, the components are added to produce the final reconstruction (Fig. 1, right-hand side). Note that, in contrast to schemes using exemplar-based synthesis methods, the added computational complexity mainly appears at the encoder, which might be a practical benefit, depending on the application.

5. RESULTS

As an example, we present visual results for the well-known image "Barbara" in Figure 2. Generally, the amount of reconstructed detail is quite similar for both methods. However, the noisy texture (carpet) in the reconstruction using our method much closer resembles the original texture. Comparison to the texture in the JPEG2000 reconstruction indicates that it is more efficient to represent such texture using our model than using conventional methods. A likely explanation is that a high number of small, random transform coefficients is needed for such texture, which can only be quantized to zero level at but the highest target rates.

In Figure 3, a limitation of the proposed method is demonstrated. Although (very low) camera noise is plausibly synthesized, fine detail is lost. This is due to the simplicity of the decomposition approach. It is obvious that the denoising algorithm needs to be adjusted to the noise energy. In our experiments, we simply use a constant setting of h. Since we do not exclusively target camera noise, but also noisy texture, we would actually need to adapt the denoising process to varying characteristics of texture. This remains a topic for future work.

Further results are presented on the authors' web site at http://www.ient.rwth-aachen.de/~balle/icip09.

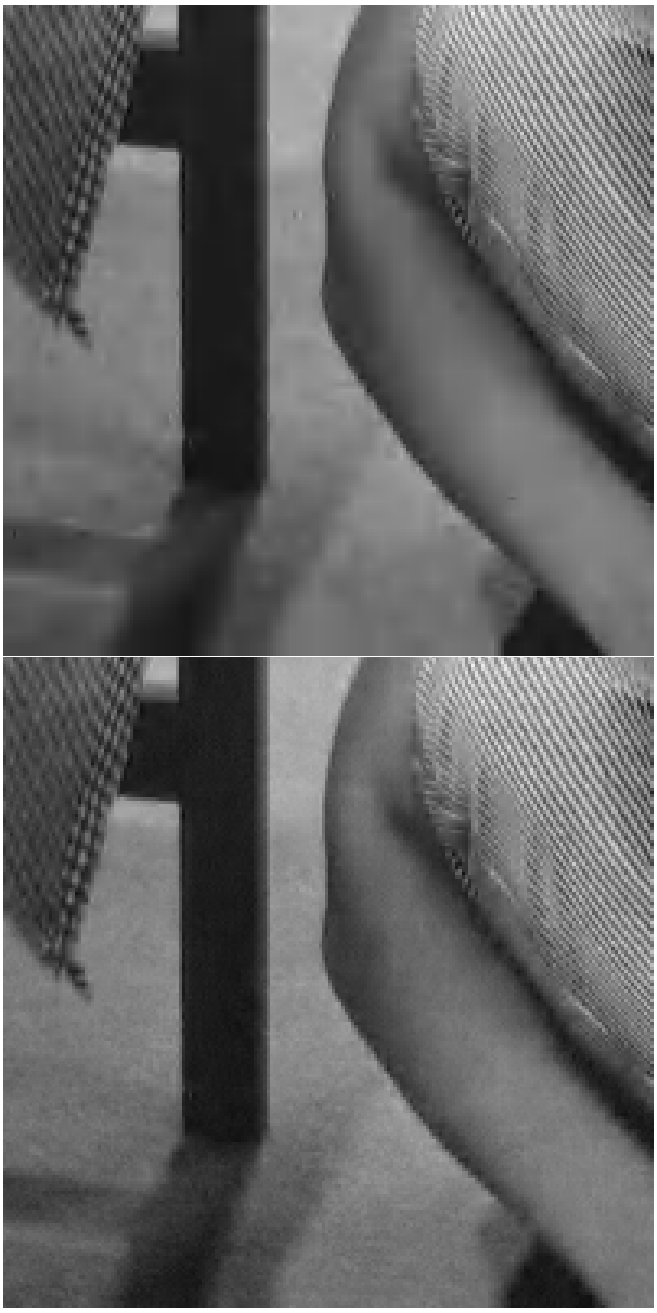

Fig. 2. "Barbara" (200×200 crop). Top: JPEG2000 reconstruction. Bottom: reconstruction using our method at equal bit rate. Parameters: $\mathcal{N}_A : 11 \times 6$ (NSHP mask), $\mathcal{N}_X : 1 \times 1$, block size 8×8, codebook size 4, quantization step size for A_k and X_k 0.00001, quant. step size for $\sigma(i)$ 0.5, target rate 0.8 bpp.

6. CONCLUSION

In this paper, we demonstrate that specialized coding schemes for noisy texture are feasible and pose a viable alternative or complement to other methods based on exemplar-based tex-

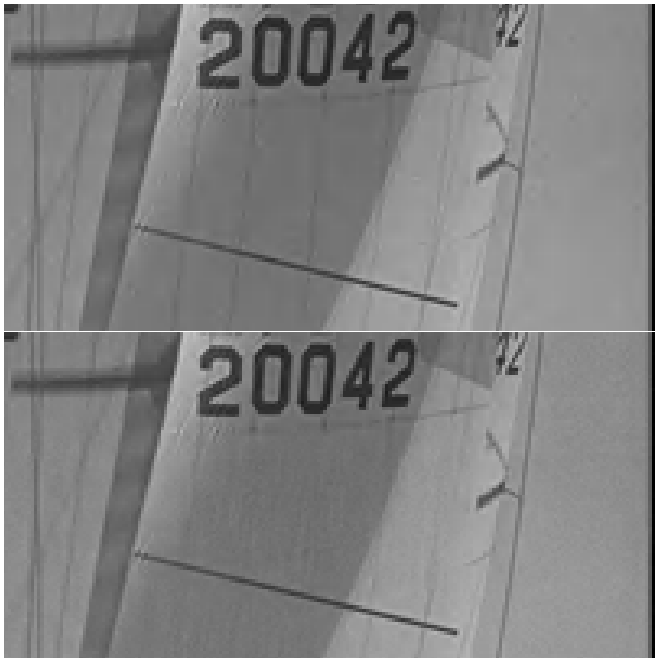

Fig. 3. "Kodim10" (200 $\times$ 100 crop), from [16]. Top: JPEG2000 reconstruction. Bottom: reconstruction using our method at equal bit rate. Same parameters as in Figure 2.

ture synthesis [4, 5]. By compressing a decomposed signal using separate compression methods, visual quality of noisy texture is improved. Further research is needed to improve coding efficiency of the noise model parameters and investigate adaptive adjustment of the decomposition algorithm, as too agressive filtering can lead to visual artifacts. Future work will also include the extension of the concept to color images and video.

7. REFERENCES

[1] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. of IEEE International Conference on Computer Vision ICCV*, pp. 1033–1038, Sept. 1999.

[2] M. Bosch, F. Zhu, and E. J. Delp, "Spatial texture models for video compression," in *Proc. of IEEE International Conference on Image Processing ICIP*, vol. 1, pp. 93–96, 2007.

[3] P. Ndjiki-Nya, T. Hinz, and T. Wiegand, "Generic and robust video coding with texture analysis and synthesis," in *Proc. of IEEE International Conference on Multimedia and Expo ICME*, pp. 1447–1450, July 2007.

[4] D. Liu, X. Sun, F. Wu, S. Li, and Y.-Q. Zhang, "Image compression with edge-based inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1273–1287, Oct. 2007.

[5] J. Byrne, S. Ierodiaconou, D. Bull, D. Redmill, and P. Hill, "Unsupervised image compression-by-synthesis within a JPEG framework," in *Proc. of IEEE International Conference on Image Processing ICIP*, 2008.

[6] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proc. of International Conference on Computer Graphics and Interactive Techniques SIGGRAPH*, pp. 479–488, July 2000.

[7] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: Image and video synthesis using graph cuts," in *Proc. of International Conference on Computer Graphics and Interactive Techniques SIGGRAPH*, pp. 277–286, July 2003.

[8] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1349–1380, Dec. 2000.

[9] A. Buades, B. Coll, and J. M. Morel, "On image denoising methods," tech. rep., Centre de Mathématiques et Leurs Applications, Cachan, France, 2004.

[10] S. P. Awate and R. T. Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 364–376, Mar. 2006.

[11] J. Ballé and M. Wien, "Extended texture prediction for H.264/AVC intra coding," in *Proc. of IEEE International Conference on Image Processing ICIP*, vol. 6, pp. 93–96, 2007.

[12] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer, 1992.

[13] J. Ballé and M. Wien, "A quantization scheme for modeling and coding of noisy texture in natural images," in *Proc. of IASTED Conference on Signal and Image Processing SIP*, 2009. To be published.

[14] *ITU-T Rec. T.800 & ISO/IEC 15444-1: JPEG 2000 Image Coding System: Core Coding System*.

[15] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 620–636, July 2003.

[16] "CIPR still images: Kodak." Downloaded from http://www.cipr.rpi.edu/resource/stills/kodak.html, June 2009.