Rate-distortion optimised transform competition for intra coding in HEVC

Adrià Arrufat [#*1], Pierrick Philippe [#2], Oliver Déforges [*3]

[#] *Orange Labs,* [*] *IETR/INSA*

[#] *4, Rue du Clos Courtel, 35512 Cesson-Sévigné, FRANCE,* [*] *UMR CNRS 6164, 35043 Rennes, FRANCE*

[1] adria.arrufat@orange.com
[2] pierrick.philippe@orange.com
[3] olivier.deforges@.insa-rennes.fr

***Abstract*—State of the art video coders are based on prediction and transform coding. The transform decorrelates the signal to achieve high compression levels. In this paper we propose improving the performances of the latest video coding standard, HEVC, by adding a set of rate-distortion optimised transforms (RDOTs). The transform design is based upon a cost function that incorporates a bit rate constraint. These new RDOTs compete against classical HEVC transforms in the rate-distortion optimisation (RDO) loop in the same way as prediction modes and block sizes, providing additional coding possibilities. Reductions in BD-rate of around 2% are demonstrated when making these transforms available in HEVC.**

***Index Terms*—HEVC, DST, transform coding, rate-distortion optimisation, adapted transform design**

I. Introduction

High Efficiency Video Coding (HEVC) is the latest video coding standard, finalised in January 2013 jointly by MPEG and the ITU-T. It provides a bit-rate reduction of up to 50% with regards to the previous standard, H.264/MPEG-4 AVC [1].

HEVC is based on transform coding, a technique that takes errors in prediction, commonly named residuals, and transforms them from spatial to frequency domain to concentrate the signal energy into fewer coefficients. To generate such residuals in intra coding, HEVC tests different prediction modes and block sizes to find the best performing residual in a rate-distortion (RD) space using the transform assigned for their block size.

Transform coding in the HEVC standard has been a very important field of study during its standardisation. One of the most relevant changes with regards to previous coding standards is the replacement of the discrete cosine transform (DCT) in favour of the discrete sine transform (DST) for the 4×4 intra prediction (IP) luma residuals. According to [1], this change provides approximately 1% bit rate reduction in intra-predictive coding.

Currently, HEVC selects the optimal residual in RD by choosing the best combination of transform size and intra-prediction mode. The residual is represented in the transform domain according to its transform unit (TU) size, that is, DST for 4×4 luma component and DCT for all other cases.

978-1-4799-6139-9/14/$31.00 ©2014 IEEE

Although HEVC allows skipping the transform step for 4×4 blocks, which implicitly provides an alternate transform, a reduced number of transform might not allow harnessing the inherent varied properties of the residuals. Some work has already been done in this area with the mode dependent directional transforms (MDDTs) [2], where a transform is designed specifically for each prediction mode based on the Karhunen-Loève Transform (KLT). However, authors in [3] show that one transform per intra prediction mode is not enough, since residuals exhibit different statistical properties, even those coming from the same prediction mode. Hence, a set of trained transforms is proposed to minimise the RD cost.

In order to adapt to different residual statistics, we extend the HEVC through an additional step in the RDO loop. In this method different transforms are tested together with prediction modes, prediction unit (PU) sizes and TU sizes.

A similar approach has been carried out in [3], where the quad-tree partitioning is performed using HEVC core transforms to find the optimal residual and test it against a set of RDOTs. As such, the transform decision is not in the quad-tree loop. The suggested approach allows achieving up to 1.6% BD-rate reductions while the complexity is moderately increased on the encoder side. A more systematic approach is used here, with the transform decision being made inside the RDO loop.

In this work we propose a framework which allows transform competition in video coding.

This paper is organised as follows. Section II introduces the concept of rate-distortion optimised transforms as well as a way of designing one single transform adapted to a set of residual blocks. This is extended towards the design of multiple transforms, each specialised on a subset of the residual blocks. A design example implemented in HEVC is shown in section III, followed by the results on HEVC, discussed in section IV.

II. Rate-distortion optimised transforms

Most desirable properties of transforms used in image processing are the energy compaction, in order to concentrate the prediction residuals energy on fewer coefficients, and the orthogonality, so that they are easily invertible and energy preserving. However, those properties are not sufficient for a transform to perform well in state-of-the-art video encoders: the data in the transform domain should require as few bits as possible to be stored. This is modelled inside HEVC through a bit rate constraint and tested in the RDO loop. Reproducing the actual video coder entropy is too complex when designing a transform. Therefore, [4] and [3] have worked on a simplified solution, involving the ℓ_0 norm, which counts the number of non-zero coefficients to incorporate the bit-rate constraint. This is equivalent to an additional sparsity constraint on the transformed coefficients.

A. Single transform design

The transform design is proposed and proved in detail in [4]. This method is able to find the optimal transform for some input data, given a transform to initialise the algorithm and a constraint on the coefficients sparsity, as explained below.

$$\mathbf{A}_{opt} = \arg\min_{\mathbf{A}} \sum_{\forall i} \min_{\mathbf{c}_i} \left(\|\mathbf{x}_i - \mathbf{A}^T \mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_0 \right) \quad (1)$$

Where $\mathbf{x}_i$ is a block of the training set, $\mathbf{c}_i$ are its quantised transformed coefficients using the transform $\mathbf{A}$. The constraint in the cost function is the ℓ_0 norm of the coefficients, i.e. the number of non-zero coefficients. The Lagrange multiplier λ tunes constraint. It depends on the quantisation applied to the coefficients, as demonstrated in [4].

The suggested design involves an iterative algorithm where the optimal coefficients are found for a given transform. Then, the transform is updated to match the optimal coefficients. Those two steps are performed until convergence is reached, with the value of the metric being stabilised.

An illustration on the iterative design for 4×4 IP luma residuals is provided in figure 1. This figure shows the performances of the DCT and DST according to the metric on equation (1) and the evolution of the RDOT in the iteration loop. One can notice how an improved compaction in the RD plane can be achieved using this learning algorithm. The appropriateness of the DST over the DCT for this kind of residuals is corroborated with this metric, as it is significantly decreased for the DST. The performance of the DST, which is considered to be close to the optimal KLT for the 4×4 IP luma residuals [5], can be reached using the algorithm after some iterations.

B. Multiple transform design

As stated before, designing one transform that captures the properties of all residuals blocks is not possible. This section proves that distortion defined in equation (1) can be further reduced using several transforms.

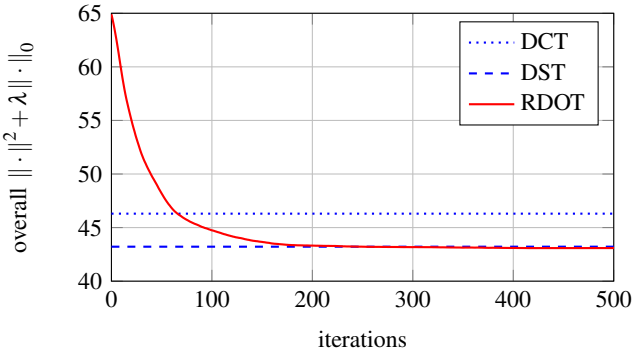

Figure 1. Rate-distortion optimisation of a random transform

As suggested in [4], in order to design a set of transforms, an additional clustering step is carried out. Residuals classification is achieved by computing the distortion from equation (1) and assigning the residual to the transform with the smallest distortion. Once all residuals have been classified, a RDOT is learnt on each class, that is, on all residuals assigned to the same transform. Then, residuals are reclassified with these newly obtained transforms and so forth.

A choice has been made to keep the current HEVC transforms and append additional adapted transforms. Thus, the set of transforms is conservative as it is able to reproduce HEVC choices. For this reason, the chosen transform configuration consists of the HEVC default transform for the current block size (DST for 4×4 blocks, DCT for the others) plus an additional set of N transforms that will be used in case they outperform HEVC in terms of RD. Consequently, each transform configuration will be referred to as $1+N$: DST or DCT plus N complementary transforms.

input : Residuals to classify $\mathbf{x}$
output: RDOTs

Initialise with DST or DCT and N random transforms
while *!convergence* **do**
  foreach *block* $\mathbf{x}$ **do**
    foreach *transform* $\mathbf{A}_n$ **do**
      $\delta_n = \|\mathbf{x} - \mathbf{A}_n^T \mathbf{c}\|^2 + \lambda \|\mathbf{c}\|_0$
    end
    for $n = 0$ **to** $N+1$ **do**
      Class$_n$.append ($\mathbf{x}$ using $\mathbf{A}_n$)
    end
  end
  for $n = 1$ **to** $N+1$ **do**
    Learn a RDOT on Class$_n$
  end
end

Algorithm 1: Clustering using multiple transforms

Figure 2 demonstrates how distortion is decreased with the number of transforms. We notice that a significant improvement can be achieved by adding from 2 ($1+2$ set) to 32 ($1+32$ set) companion transforms.

Multiple transform design using the metric defined in equation (1) has highlighted the fact that better performances can be

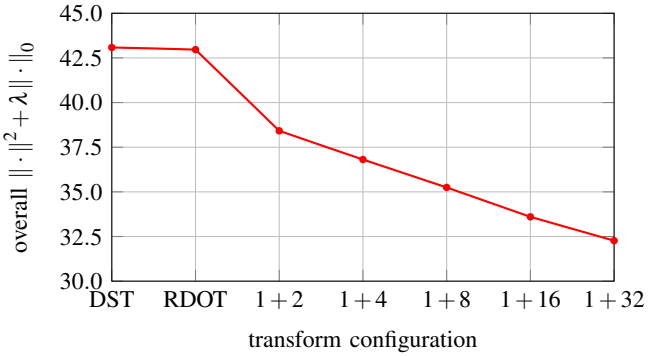

Figure 2. Distortion with multiple transforms for 4×4 residuals

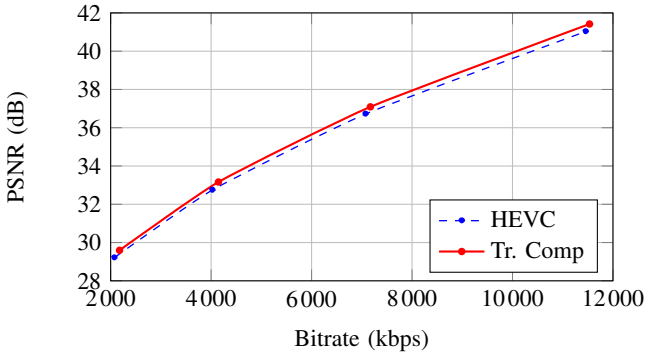

Figure 3. BD-rate savings for Blowing Bubbles: -3.38%

achieved. Although the gains in terms of RD seem important, one needs to confirm that these potential improvements turn into actual gains when applied to a complete coding scheme.

Next step is designing a system to take advantage of using multiple transforms in the state-of-the-art video coder HEVC.

III. Proposed Design

The goal of multiple transform design is to improve HEVC in terms of Bjøntegaard Distortion-rate (BD-rate) [6]. Hence, some considerations have been made when selecting the target block size, the amount of residual blocks and the number of transforms per block size.

One can observe in table I that about half of the surface of the HEVC test sequences is covered by 4×4 and 8×8 blocks. Therefore, those are the selected block sizes to learn the transforms. Working on small block sizes offers a good trade-off in terms of performance and design complexity.

For the sake of transform consistency among different bit-rates, more than 700 000 intra prediction residuals coming from numerous external sequences with quantisation parameters (QPs) and resolutions consistent with the common test conditions [7] have been used to learn the transforms for the 4×4 and 8×8 TUs.

A number of RDOTs has been chosen for each TU size, taking into account the variability of residuals and the relative overhead weight for each size. For the 4×4 blocks, 4 extra RDOTs have been learnt. Despite the potential improvement pointed out by figure 2 when increasing the number of transforms, having to signal the transforms in HEVC introduces an extra overhead that counterbalances the improvements obtained. Thus, 4 additional transforms provide a good trade-off between the signalling cost and the improvements. For analogue reasons, the number of additional transforms chosen for the 8×8 residuals has been set to 16.

Finally, a straightforward signalling mechanism to inform the decoder about the selected transform has been established. The signalling is based on a bit flag which indicates whether the HEVC default transform (DST for 4×4 and DCT for 8×8) has been used. In case it has not, the flag is followed by the transform index represented by a fixed length codeword. In our scheme, the flag is coded using context adaptive binary arithmetic coding (CABAC) with a dedicated context.

IV. Experimental Results

The proposed system, implemented in the HEVC reference software, consists of two RDOT sets: the first one has 4 extra RDOTs for 4×4 residuals whereas the second one has 16 extra RDOTs for 8×8 residuals. All transforms are available at both encoder and decoder sides. While the encoder is carrying out the RDO loop, it chooses the best performing transform and signals it with the corresponding code. The decoder only reads the signalled transform index and performs one single inverse transform to the current residual.

Although training has been done on intra prediction residuals coming from external sequences using the HEVC all intra (AI) common test configuration [7], which details all the coding parameters, the system has been tested in both AI and random access (RA) coding modes.

An example of BD-rate curve is displayed on figure 3. The resulting bit stream has a larger size which can be seen by the shift to the right of the bit rates. However, the quality improvements are consistent across all QPs, with the peak signal-to-noise ratio (PSNR) value being consistently improved for all rates. In spite of the objective quality enhancements, no visual improvements have been observed in the decoded sequences.

Table I contains the average amount of surface coverage by the different block sizes in AI, as well as the performances of the proposed system compared to the HEVC reference software. Improvements on all sequence resolutions are achieved, but most notably on lower resolution sequences. The average gain on AI sequences is around 2% and around 1% for the RA sequences. The significant improvements in RA are explained as the intra coded blocks, which are improved with this technique, serve as better quality predictors.

Obtained gains are correlated to the amount of surface covered by the target blocks (4×4 and 8×8) in the HEVC reference software, as stated in table I. Most improved sequences correspond to those whose surface covered by 4×4 and 8×8 is bigger (up to 84% in BlowingBubbles), whereas modest improvements or even some losses are found on the higher resolution sequences, which present a small coverage of those blocks. Therefore, sequences making less use of our target block sizes are less subject to improvement.

Despite the significant gain in terms of BD-rate, a great part of the improvement is consumed by the signalling mechanism,

	Sequence	Surface by block sizes (HEVC) 4×4	8×8	16×16	32×32	Y BD-rate AI	RA
Class A (2560×1600)	PeopleOnStreet	18%	30%	33%	19%	-1.28%	-0.34%
	Traffic	14%	25%	33%	28%	-1.66%	-1.80%
	NebutaFestival	3%	7%	20%	70%	-0.18%	-0.05%
	SteamLocomotiveTrain	1%	7%	22%	70%	0.00%	0.51%
	Average	**9%**	**17%**	**27%**	**47%**	**-0.78%**	**-0.42%**
Class B (1920×1080)	BasketballDrive	8%	22%	29%	41%	-1.04%	-0.07%
	BQTerrace	14%	40%	22%	24%	-1.61%	-0.94%
	Cactus	16%	23%	31%	30%	-1.85%	-1.38%
	Kimono1	3%	9%	23%	65%	-0.30%	-0.24%
	ParkScene	17%	23%	29%	31%	-1.28%	-1.19%
	Average	**12%**	**23%**	**27%**	**38%**	**-1.22%**	**-0.76%**
Class C (832×480)	BasketballDrill	29%	33%	25%	13%	-2.76%	-2.63%
	BQMall	24%	31%	27%	18%	-1.83%	-1.28%
	PartyScene	47%	35%	16%	3%	-3.06%	-2.14%
	RaceHorses	18%	24%	30%	28%	-1.87%	-0.80%
	Average	**21%**	**27%**	**28%**	**24%**	**-2.38%**	**-1.72%**
Class D (416×240)	BasketballPass	22%	26%	28%	24%	-2.33%	-0.98%
	BQSquare	47%	25%	18%	10%	-2.80%	-1.94%
	BlowingBubbles	50%	34%	14%	2%	-3.38%	-2.14%
	RaceHorses	26%	31%	27%	16%	-2.23%	-0.96%
	Average	**36%**	**29%**	**22%**	**13%**	**-2.69%**	**-1.51%**

Table I
AI HEVC SURFACE COVERAGE AND Y BD-RATE SAVINGS REFERRED TO HEVC

Configuration 4×4	Configuration 8×8	Y BD-rate No Signalling	Y BD-rate Signalling
$1+4$	—	-8.42%	-1.40%
—	$1+16$	-7.01%	-2.46%
$1+4$	$1+16$	-11.27%	-3.38%

Table II
SIGNALLING IMPACT ON Y BD-RATE IMPROVEMENT

as shown in table II. This table reports numbers computed for the BlowingBubbles item. Some configurations have been tested, combining both blocks sizes and various numbers of RDOTs. Table II reveals the potential gains if the decoder was able to guess the chosen transform by the encoder without any signalling. This table also indicates that signalling 4×4 blocks is more expensive than 8×8: the ratio between the number of conveyed pixels and the amount of signalling is more favourable for the 8×8 case. As a result, improving the signalling system by e.g. exploiting some spatial redundancies would certainly improve the bit-rate reduction (up to 10%).

The encoder complexity has been notably increased. With the suggested system, whenever a 4×4 or 8×8 is about to be transformed, it is tested with all the candidate transforms in the RDO loop. Due to the recursive implementation of the quad-tree partitioning, the number of computations per block escalates rapidly: an approximate factor of 8 is currently noticed. It is worth mentioning that the implementation of the transform competition has not been optimised in such a way that all the combinations would not need to be investigated. The decoder, on the other hand, only needs to recover the transform index and apply the designed transform. However, the processing time has been slightly increased due to the non-separable transform design from (1), which will be addressed in future works by using separable transforms.

V. CONCLUSION

Experimental results on this paper have evidenced that extending the RDO loop to enable multiple transforms provides consistent bit-rate savings over the HEVC scheme.

These bit rate improvements come at the expense of increased complexity in the encoder side, up to a factor of 8, since many more coding alternatives are made available and need to be evaluated in the RD sense. No fast decision of the best transform has been investigated so far, but it is felt that the complexity can be vastly reduced: this is part of the forthcoming investigation. Current results serve as a proof of concept. On the other side, the decoder can keep its simplicity as it only has to apply the signalled transform.

Future work includes focusing on bigger TUs to improve performance on larger resolutions, since the amount of surface covered by 16×16 and 32×32 blocks is much bigger than on smaller sequences. Furthermore, a separable approach is being worked on to reduce the complexity.

REFERENCES

[1] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 2116–2119.

[3] F. Zou, O. Au, C. Pang, J. Dai, X. Zhang, and L. Fang, "Rate-distortion optimized transforms based on the Lloyd-type algorithm for intra block coding," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 6, pp. 1072–1083, 2013.

[4] O. Sezer, O. Harmanci, and O. Guleryuz, "Sparse orthonormal transforms for image compression," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 149–152.

[5] J. Han, A. Saxena, and K. Rose, "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 726–729.

[6] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T, Austin, Texas, Tech. Rep. VCEG-M33, April 2001.

[7] F. Bossen, "Common test conditions and software reference configurations," ITU-T, Geneva, Switzerland, Tech. Rep. JCTVC-I1100, May 2012.