Application of Directionlets in Video Coding

Vladan Velisavljevic[*], **Marta Mrak**[†], **Ahmet Kondoz**[†]

[*]Deutsche Telekom Laboratories, Germany; email: vladan.velisavljevic@telekom.de
†I-Lab, Centre for Communication Systems Research, University of Surrey, UK; email: {m.mrak, a.kondoz}@surrey.ac.uk

Keywords: Directional transforms, wavelets, Dirac video codec.

Abstract

The goal of the presented work is to apply a directionally-adaptive two-dimensional separable wavelet transform, called *directionlets*, in common wavelet-based video coders. So far, directionlets have been used only in still image coding, where they have provided a sparser representation of images in the transform domain and a better compression performance. We apply directionlets to frames of video sequences in the video coding scheme based on the *Dirac* video codec. The novel method outperforms the traditional non-adaptive method based on the standard wavelet transform in terms of both the rate-distortion performance and the visual quality of reconstructed sequences.

1 Introduction

All modern video communication systems require some form of video compression. Early video coding standards, like MPEG-2, do not provide sufficient compression for growing demands to fit a large number of TV channels in the existing broadcasting systems. New solutions based on the H.264 / AVC standard enable a better utilisation of the available bandwidth using higher compression rates [1]. With a goal to develop innovative compression solutions that meet their specific needs, BBC has developed an advanced video coding system called *Dirac* [2]. The Dirac coder is based on the technology that has been used in the state-of-the-art image and video compression systems. In contrast to other popular video codecs, Dirac is based on an open-source technology and, thus, can become influential in a wide research community.

Similar to the widespread MPEG and H.26x video coding technologies, Dirac uses motion compensation to enhance the compression efficiency. The produced frames (*intra-coded* and *motion-compensated*) are spatially transformed using the wavelet transform (WT), which is also adopted in the still image coding standards, like JPEG-2000 [12].

One of the challenges in application of the WT in video coding is how to exploit efficiently the high spatio-temporal coherence of frames. Recently, an adaptation of the spatial transform to the motion information has been proposed in [7] and [11] with a goal to improve the sparsity of the signal representation and, thus, to achieve a better compression performance and decoding quality. This, so-called, *Motion-Driven Adaptive Transform* (MDAT) is beneficial when intra blocks exist in the motion-compensated frames.

A similar motivation to provide a sparser signal representation in still image coding has led to a construction of *directionlets* [13], which have been built on top of the standard WT as an adaptive directional asymmetric transform. Directionlets are capable of characterizing efficiently oriented elongated features in images, like edges or contours, by adaptation of transform directions to locally dominant directions across the image domain. At the same time, they preserve separability and the conceptual and computational simplicity of the standard WT. Moreover, directionlets have been shown to outperform the standard WT in still image compression algorithms [14].

This achievement motivates us to propose a novel video coding method obtained as a combination of the Dirac coder and directionlets. The standard WT is replaced by directionlets and such a transform is applied to both types of video frames in the Dirac codec. The transform directions are adapted within frames and this adaptation is encoded as side information together with the other coding information. However, in spite of this overhead data, the novel video coding method outperforms the standard Dirac coder in terms of both the rate-distortion performance and the visual quality of decoded video sequences.

The outline of the paper is as follows. Section 2 reviews the basic concepts of the wavelet-based video coding, describes the architecture of the Dirac video encoder and addresses the limitations of the standard WT. The basic principle of the construction of directionlets is described in Section 3. Section 4 introduces the concept of the novel coding algorithm obtained as a combination of directionlets and the Dirac video codec, whereas, in Section 5, the results of encoding using the new method are compared to the results obtained by the standard Dirac coder. Finally, Section 6 concludes this paper.

2 Wavelet-based video coding

The architecture of the Dirac codec is similar to the architecture of other wavelet-based video codecs [3], [10] and also of the standard MPEG and H.26x video coding systems. The Dirac codec consists of the three basic blocks: motion compensation, spatial WT with quantisation and entropy coding, as shown in Figure 1.

The main goal of motion compensation is to exploit temporal redundancy between frames, whereas the task of the spatial WT is to provide a compact representation of the frame content, which commonly have most of the energy located in lower frequency subbands.

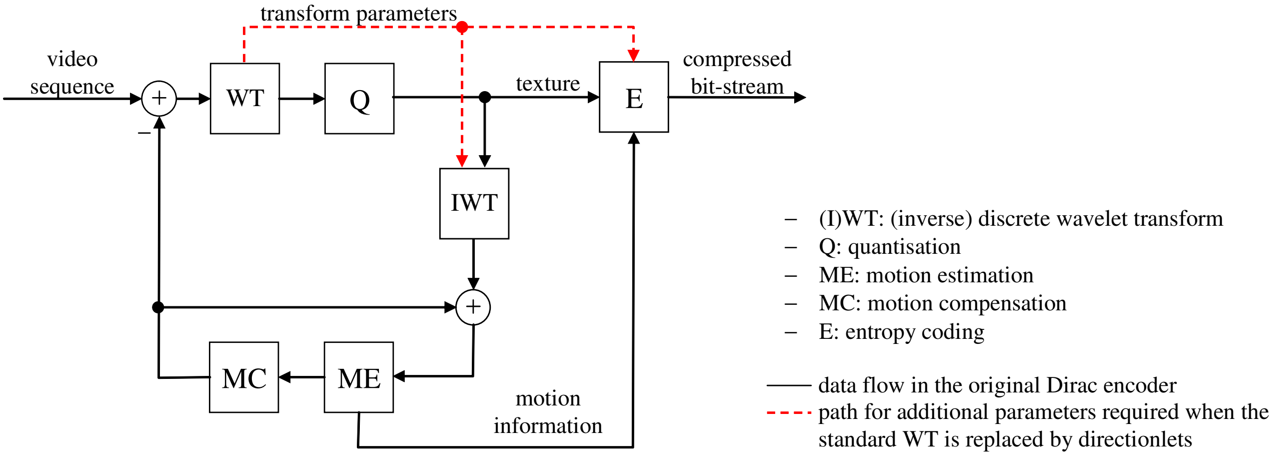


Figure 1 Architecture of the wavelet-based Dirac video codec.

Hence, in this way, the WT introduces a coding gain. The resulting transform coefficients obtained from the spatially transformed frames are efficiently encoded using quantisation and entropy coding. Quantisation is a lossy encoding process, where the choice of quantisation step size determines the resulting bit-rate. However, in turn, the entropy coder performs a lossless compression, where the final data is compressed using an arithmetic coding engine that exploits inter-dependency of the remaining coding symbols.

Since the overall performance of video coding depends on the types of video frames and the implemented transform, we explain next these two characteristics in more details.

2.1 Motion compensation and frame types

A great part of efficiency in video compression relies on an efficient exploitation of temporal redundancies across frames. Since neighbouring frames in a video sequence are highly correlated, the content of a frame can be efficiently estimated using the data from previously encoded neighbouring frames. In this process, which is called *Motion Compensation* (MC), a significant part of the temporal redundancy can be removed.

In general, we distinguish 2 types of frames in video coding: intra-coded frames and motion-compensated frames. The intra-coded frames are independently encoded, similarly to images in still image coding. By contrast, the motion-compensated frames (or error frames) contain prediction error that is related to texture, edges and intra-coded areas and these frames commonly have a noise-like behaviour. The content of these frames mostly depends on the applied motion compensation method. Notice that, in the systems that use motion blocks as basic motion units, the compensated frames can suffer from a noticeable block structure-effect.

2.2 Spatial wavelet transform and its limitations

Two categories of transforms are commonly used in video and still image coding – discrete cosine transform (DCT) or DCT-like transforms (mostly in video coding standards, like MPEG and H.26x) and the WT (in still image coding standards, like JPEG-2000, but also in some video codecs). The transforms in video coding are applied on both the intra-coded and the motion-compensated frames, which are first divided into fixed size blocks.

The WT is especially popular in still image coding because of a good compression performance, but also because it facilitates a broad range of other functionalities, including scalability and multi-resolution image representation. For that reason, the WT has been adopted in video coding applications and it is also implemented as a spatial transform in the Dirac video codec.

One of the main advantages of the WT is its computational simplicity. The filtering operations are separable, where purely 1-D filters are applied along rows and columns. However, this feature of the WT implies a limitation in the efficiency of representation of the objects in images (or frames) characterized by geometric regularity and directions different than horizontal or vertical. For that reason, there have been proposed many other more sophisticated transforms beyond the standard WT that can improve characterization of these directional features, but at a cost of higher computational complexity (e.g. curvelets [6] or bandelets [9]). By contrast, another directionally adaptive transform, called directionlets, is capable of improving the sparsity of representation of images and frames and, at the same time, it retains separability and simplicity of the standard WT, as explained in the next section.

3 Still image coding using directionlets

The construction of directionlets has been explained in detail in [13] and [14]. Here, we give only a brief review of the basic ideas.

Directionlets are constructed as separable 2D basis functions of the so-called *skewed anisotropic wavelet transforms*. These transforms make use of the two concepts: anisotropy and directionality. Anisotropy is obtained by an asymmetric iteration of transform steps along two transform directions, that is, the transform is applied more along one than along the

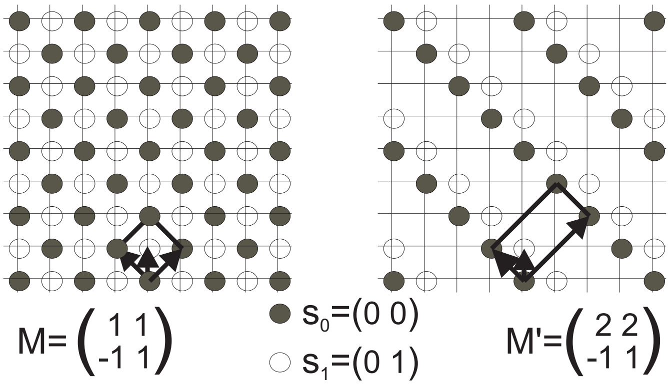


Figure 2 An example of the construction of directionlets based on integer lattices for pair of directions 45° and -45°.

other direction. Directionality is a result of the construction along two skewed transform directions (not necessarily horizontal or vertical) built using integer lattices. An example of the construction of directionlets is shown in Figure 2 for directions 45° and -45°, whereas two examples of the basis functions are illustrated in Figure 3 (b) and (c) using the frequency decomposition as shown in (a).

Filtering using the corresponding high-pass (HP) wavelet filters along any pair of directions imposes *directional vanishing moments* (DVM). This property of directionlets allows for efficient capturing of oriented features and a sparser representation of natural images than the representation provided by the standard WT.

However, images have geometrical oriented features that vary over space and, thus, directionality in image is a local characteristic defined in a small neighbourhood. Since directionlets are limited to have up to 2 DVM across different directions, the chosen orientations are adapted locally in each neighbourhood across the image domain. Thus, this implies a need for *spatial segmentation* as a way of partitioning an image into smaller segments with one or few dominant directions per segment. In [14], a quad-tree spatial segmentation has been used for image coding and we adopt the same segmentation method in our video coding algorithm because of the simplicity, as explained next.

4 Directionlets in video coding

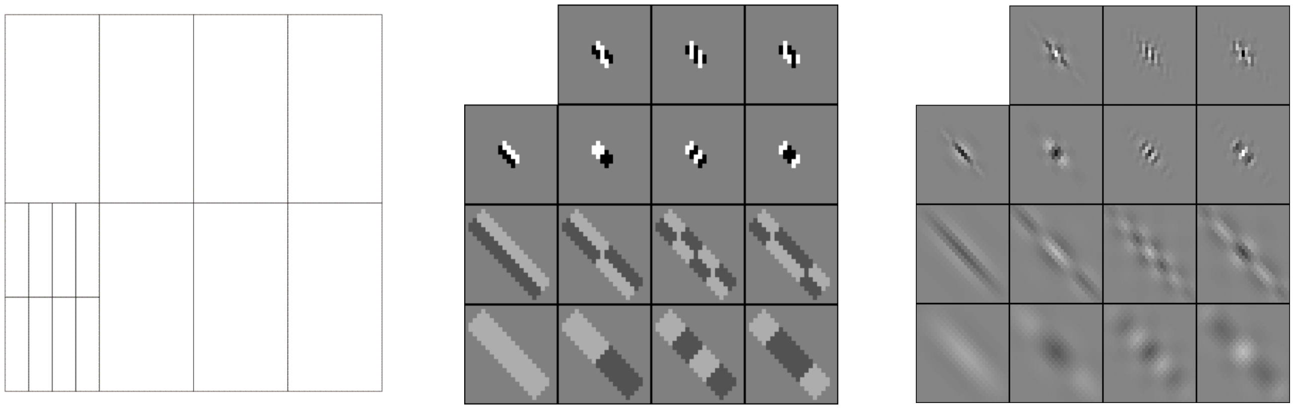
In this section we describe a newly proposed method that is based on integration of directionlets in the Dirac video codec.

Recall that, in the Dirac codec, both intra-coded and motion-compensated frames are transformed using the WT, as shown in Figure 1. The transform coefficients are quantised and encoded using an entropy coder. At the encoder, the reference frames are synthesised from the quantised coefficients to allow the decoder to use the same reference for compensation.

In our novel video coding method, we modify the Dirac codec so that the standard WT is replaced by directionlets, that is, we address the following blocks: the spatial transform (the block marked as WT in Figure 1) and the inverse spatial transform (IWT). These modifications are shown by dashed lines in the data flow chart in Figure 1.

In the transform of each frame, two nested optimization processes are performed to minimize the energy in the HP subbands: (a) spatial quad-tree segmentation and, for each spatial segment, (b) search for the optimal transform directions. First, the frames are segmented into full spatial quad-trees with a predetermined maximal segmentation depth m. Then, the optimal pair of transform directions d_n^* chosen from a predetermined set D is assigned to each segment n in the spatial trees. This optimization is performed by an exhaustive search so that the resulting HP energy is minimised. That is, the optimal pair of transform directions is determined as

$$d_n^* = \arg\min_{d\in D} P_W(d,n)\,, \tag{1}$$



(a) The decomposition in frequency for two iterations

(b) The basis functions for the Haar scaling and wavelet functions

(c) The basis functions for the 9/7 biorthogonal scaling and wavelet functions

Figure 3 Directionlets allow for an asymmetric iteration of the filtering and subsampling operations applied along two different directions.

where $P_W(d,n) = \sum_i |W_{i,n}^{(d)}|^2$ and $W_{i,n}^{(d)}$ are HP wavelet coefficients produced by the pair of directions d. Finally, the dynamic programming method [5] is used to check all the segments bottom-top and to prune the corresponding sub-trees, if necessary, to reduce the total energy of the HP coefficients. That is, the sub-tree rooted at segment n is pruned if

$$P_W(d_n^*, n) \le \sum_{k=1}^{4} P_W(d_{n_k}^*, n_k), \tag{2}$$

where the segments n_k are children of the segment n.

The resulting optimal segmentation tree and the optimal transform directions at each terminating segment are encoded as a side information and sent using additional bits. Notice that the amount of these side bits is significantly smaller than the amount of bits required to encode the transform coefficients. Apparently, the exact length of the side bit stream depends on the structure of the optimal tree, but it can be upper bounded by the following proposition.

Proposition 1: The number of side bits per frame required to encode the optimal spatial tree structure of maximal depth m and the choice of transform directions from the set D is upper bounded by $4^m \cdot (1/3 + \log_2|D|)$, where $|D|$ is the cardinal number of the set D.

Proof: To encode the spatial tree, one bit per segment is required for all segments except the ones at the level m. Thus, the number of these bits is upper bounded by

$$N_1 = \sum_{i=0}^{m-1} 4^i = 1/3 \cdot \left(4^m - 1\right).$$

To encode the choice of transform directions, $\log_2|D|$ bits per terminate segment are required in average (assuming that the choices of directions are entropy encoded) and, thus, the number of these bits is upper bounded by $N_2 = 4^m \cdot \log_2|D|$. Therefore, the total number of the side bits is bounded by $N_1 + N_2 < 4^m \cdot (1/3 + \log_2|D|)$.■

In the experiments (as also explained in Section 5), $m = 2$ and the set $D = \{(0°, 90°), (0°, 45°), (0°, -45°), (90°, 45°), (90°, -45°)\}$. Thus, the maximal number of the additional bits per frame is smaller than 43. Notice that there exists a high correlation among the optimal tree structures and transform directions in neighbour frames. This correlation can be exploited to further reduce the length of the side bit stream. However, we do not address that issue here, but leave it for future work.

The computational complexity of the entire coding algorithm is increased because of the optimization process by the constant $m \cdot |D|$. However, even though this increase is not negligible, the order of the complexity remains the same and the real-time implementation is still achievable.

We summarize next the entire novel video coding algorithm.

Algorithm 1: Dirac video codec with directionlets

```
Initialization: Set segm_level ← 0,
---
Recursive sequence:
Step 1: If segm_level < m, then:
    • Apply one level of quad-tree segmentation,
    • For each of the 4 generated segments go recursively
      to Step 1 with segm_level ← segm_level + 1,
Step 2: For each pair of transform directions from the set D,
    apply directionlets to the current segment,
Step 3: Choose the optimal pair of directions $d_n^*$ using (1),
    and record the corresponding $P_W(d_n^*, n)$,
Step 4: If segm_level < m, then:
    • If (2) holds, then prune the sub-tree rooted in n,
    • Else, update $P_W\left(d_n^*, n\right) \leftarrow \sum_{k=1}^{4} P_W\left(d_{n_k}^*, n_k\right)$,
---
Encoding sequence:
Step 5: Apply the spatial transform on the optimal
    segmentation tree using the optimal pairs of transform
    directions for each segment,
Step 6: Encode the tree structure and the transform directions
    as side information.
```

Notice that the subbands produced by directionlets still have the same structure in terms of the number of low and high-pass subbands and coefficients therein as in case of the standard WT. Therefore, we use the same quantisation and entropy coding process as in the original Dirac codec.

At the decoder side, first the spatial segmentation tree is synthesised and then the corresponding inverse transforms with the optimal transform directions are applied on the quantised coefficients.

5 Experiments

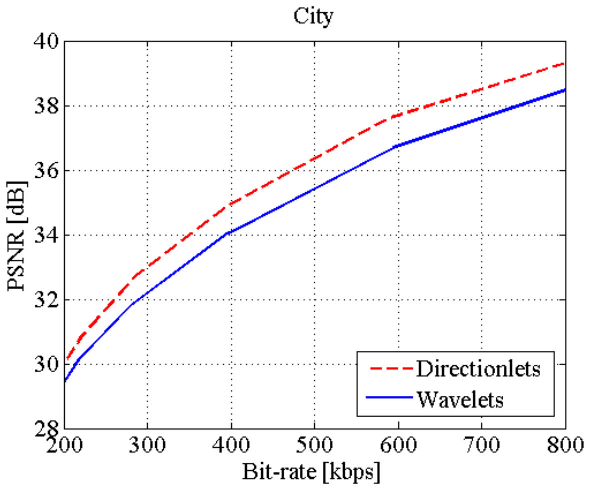
The tests have been performed using Dirac video codec with an integrated support for directionlets. Directionlets are used as the actual spatial transform and the coding results are compared to the results obtained using the standard WT.

Both the standard WT and directionlets use the biorthogonal 9/7 filter-bank [4], which has been widely accepted in image coding and has also been successfully deployed in video coding. The methods are applied on the test sequences "City", "Crew" and "Foreman" that have the CIF resolution (352×288) and are sampled at 15 fps. The sequences are encoded at a wide range of bit-rates. Directionlets are applied on the Y component of the sequences and the obtained PSNR results are averaged over the entire sequence. The results are presented for the entire bit-rate (encoding the bits for all components - YUV and side information). Notice that, despite the additional bits used to encode the side information, directionlets outperform the standard WT in terms of the coding performance across the entire bit-rate interval. Notice also that the compression artefacts that appear in case of the standard WT at low bit-rates are significantly reduced by implementation of directionlets leading to a better visual quality of the decoded frames, as shown in Figure 5.

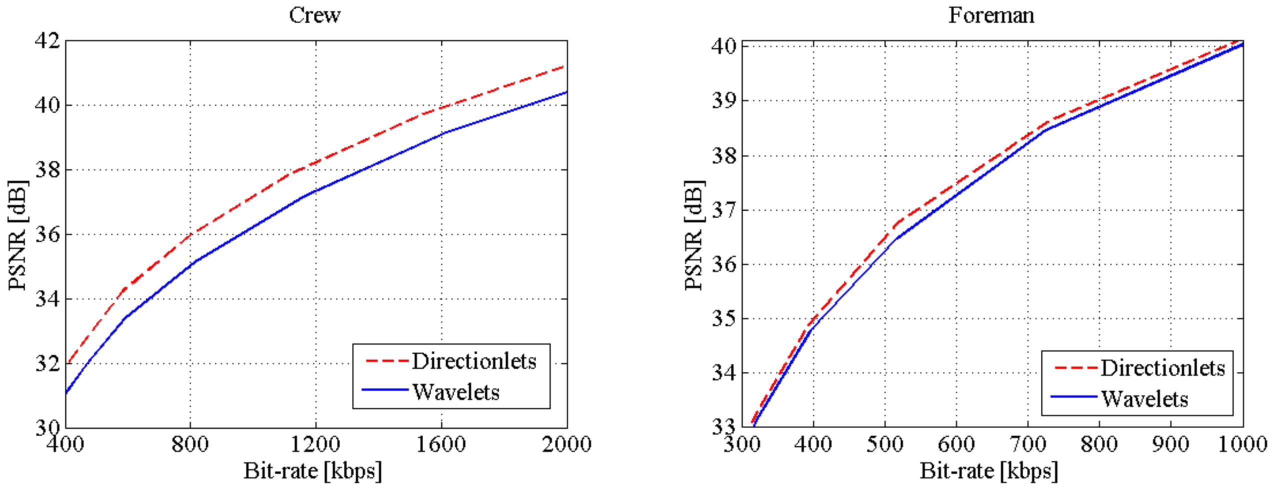
Spatial segmentation applied in the novel coding method might result in a blocking effect, especially at very low bit rates, since the neighbor regions within frames are transformed and quantised using different transform directions. This effect can lead to visually annoying distortions in decoded sequences. To reduce this blocking effect, several deblocking methods can be used, which have also been used in the old JPEG still image coding method. However, this enhancement is left for future work.

6 Conclusions

We have proposed a novel video coding method based on a directionally adaptive spatial transform, called directionlets. The adaptive transform replaces the standard WT in the Dirac codec and allows for a sparser representation of video frames in the transform domain, which leads to an enhanced compression performance. Moreover, the visual quality of the decoded video sequences is significantly improved. At the same time, the order of computational complexity of the coding algorithm is retained the same.



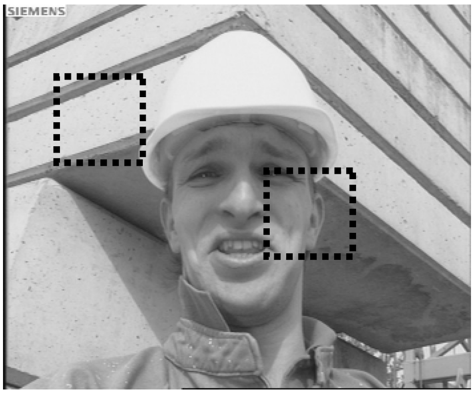
a) Test sequence "City"



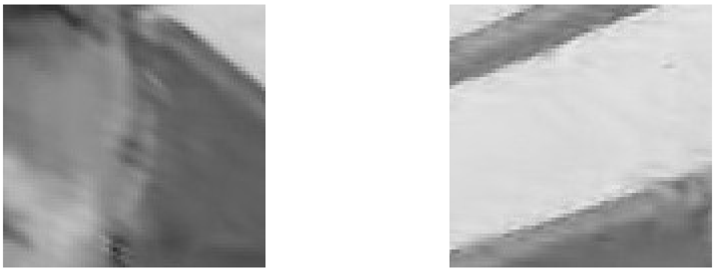
b) Test sequence "Crew"

c) Test sequence "Foreman"

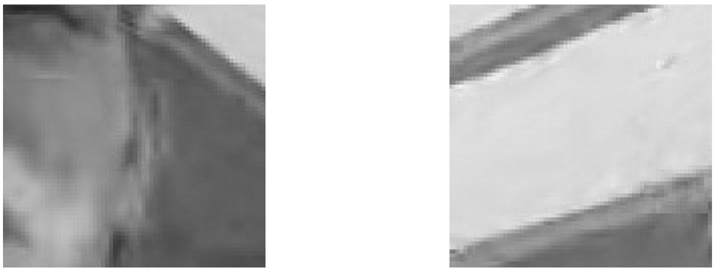
Figure 4 Comparison of the compression performance of the Dirac video codec with the standard WT and directionlets for 3 standard test video sequences.



a) original frame from the "Foreman" sequence with selected details



b) wavelets; PSNR = 30.34 dB; 225 kbps



d) directionlets; PSNR = 30.91 dB; 227 kbps

Figure 5 Decoding of the video sequence Foreman at the same quality level using the two methods. Notice that the compression artefacts are significantly reduced in case of the Dirac codec based on directionlets, as compared to the method based on the standard WT.

Acknowledgment

Part of the presented work was developed within VISNET II, a European Network of Excellence, funded under the European Commission IST FP6 programme.

References

[1] *Advanced Video Coding for Generic. Audiovisual Services*, ITU-T Rec. H.264 & ISO/IEC 14496-10, May 2003.

[2] *Dirac video codecs*, URL: www.bbc.co.uk/rd/projects/dirac/

[3] N. Adami, A. Signoroni, R. Leonardi, "State-of-the-Art and Trends in Scalable Video Compression with Wavelet-based Approaches," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 17, Iss. 9, pp. 1238 - 1255, September 2007.

[4] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, “Image Coding using Wavelet Transform”, *IEEE Trans. on Image Processing,* vol. 1, pp. 205-220, Apr. 1992.

[5] L. Breiman, J. Friedman, R. A. Olshen, C. J. Stone, “Classification and Regression Trees”, Wadsworth, Belmont, CA, 1984.

[6] E. J. Candes, D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges", in *Curve and Surface Fitting*, A. Cohen, C. Rabut, L. L. Schumaker, Eds. Saint-Malo: Vanderbilt University Press, 1999.

[7] M. Mrak, E. Izquierdo, "Spatially Adaptive Wavelet Transform for Video Coding with Multi-Scale Motion Compensation," Proc. IEEE International Conference on Image Processing (ICIP 2007), September 2007.

[8] M. Mrak, N. Sprljan, E. Izquierdo, "Motion estimation in temporal subbands for quality scalable motion coding," *Electronics Letters*, Vol. 41, No. 19, September 2005.

[9] E. L. Pennec, S. Mallat, "Sparse geometric image representations with bandelets", *IEEE Trans. Image Processing*, vol. 14, pp. 423-438, Apr. 2005.

[10] N. Sprljan, M. Mrak, G.C.K. Abhayaratne, E. Izquierdo, "A Scalable Coding Framework For Efficient Video Adaptation", *Proc. 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005)*, April 2005.

[11] N. Sprljan, M. Mrak, E. Izquierdo, "Motion Driven Adaptive Transform Based on Wavelet Transform for Enhanced Video Coding", *Proc. 2nd International Mobile Multimedia Communications Conference (Mobimedia)*, September 2006.

[12] D. Taubman, M. W. Marcellin, *JPEG2000 Image Compression: Standards and Practice*, Kluwer Academic Publishers, 2002.

[13] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, P. L. Dragotti, "Directionlets: Anisotropic Multidirectional Representation with Separable Filtering", *IEEE Trans. on Image Proc.*, Vol. 15, pp. 1916-1933, July 2006.

[14] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, “Space-Frequency Quantization for Image Compression with Directionlets”, *IEEE Trans. on Image Proc.*, Vol. 16, pp. 1761-1773, July 2007.