Integrating a Spatial Just Noticeable Distortion Model in the Under Development HEVC Codec

Matteo Naccari and Fernando Pereira

ICASSP, Prague, May 24th, 2011

INSTITUIÇÕES ASSOCIADAS:

INSTITUTO SUPERIOR TÉCNICO

instituto de telecomunicações

Outline

1. Context and objective
2. Technical contributions:
   - HEVC extension towards integer DCT and higher block sizes
   - HEVC extension towards the Mode Dependent Directional Transform (MDDT)
3. Experimental results
4. Conclusions and future work

1. Context and objective

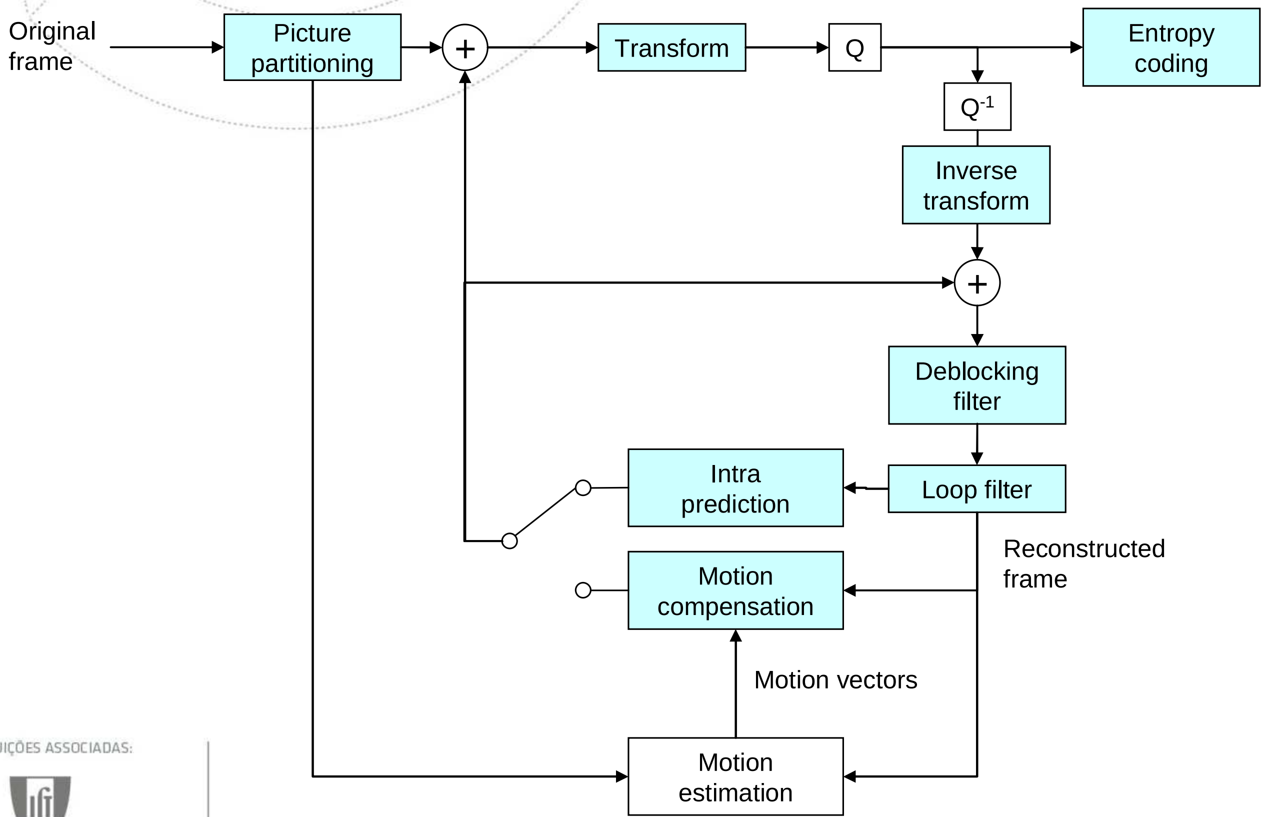
High definition videos everywhere

- Recent advantages in video capturing and display will allow to have videos with definitions higher than High Definition Television (HDTV)

- Current network infrastructure will hardly manage these new higher definitions
- Thus, **compression efficiency must be improved** beyond H.264/AVC performance
  - **January 2010**: Joint (MPEG+ITU) Call for Proposals (CfP) on Video Compression Technology (MPEG Doc. N11113)

MPEG+ITU CfP outcome and the emerging HEVC standard

- Twenty seven proposals were submitted for evaluation
- Some of them provided same H.264/AVC subjective quality at half the bitrate
- From this promising result, some actions followed:
  - Launching the **High Efficiency Video Coding (HEVC)** standardization project, managed by the Joint Collaborative Team on Video Coding (JCT-VC)
  - Investigation of the most promising coding tools in the **Test Model under Consideration** (TMuC, currently HM)



Main novelties:

- Block sizes larger than 16×16
- New transforms beside DCT (MDDT, ROT)
- New intra prediction modes
- Wiener in-loop filter

Human visual system properties

- The Human Visual System (HVS) is not equally sensitive to distortion

Lower distortion areas

Higher distortion areas

- Therefore, one can perform:
  - **Coarser quantization** → HVS **less sensitive** areas
  - **Finer quantization** → HVS **more sensitive** areas

Just noticeable distortion and objective of this work

- The HVS distortion sensitivity is quantified by the Just Noticeable Distortion (JND):

⇓

JND: the minimum visibility threshold below which no change can be perceived by the HVS

- The JND threshold is computed by a JND model and (usually) for each coded coefficient
- With the JND threshold, it is possible to **perceptually tune** the quantization step Q:

$$Q_{JND} = Q \times T_{JND}$$

⇓

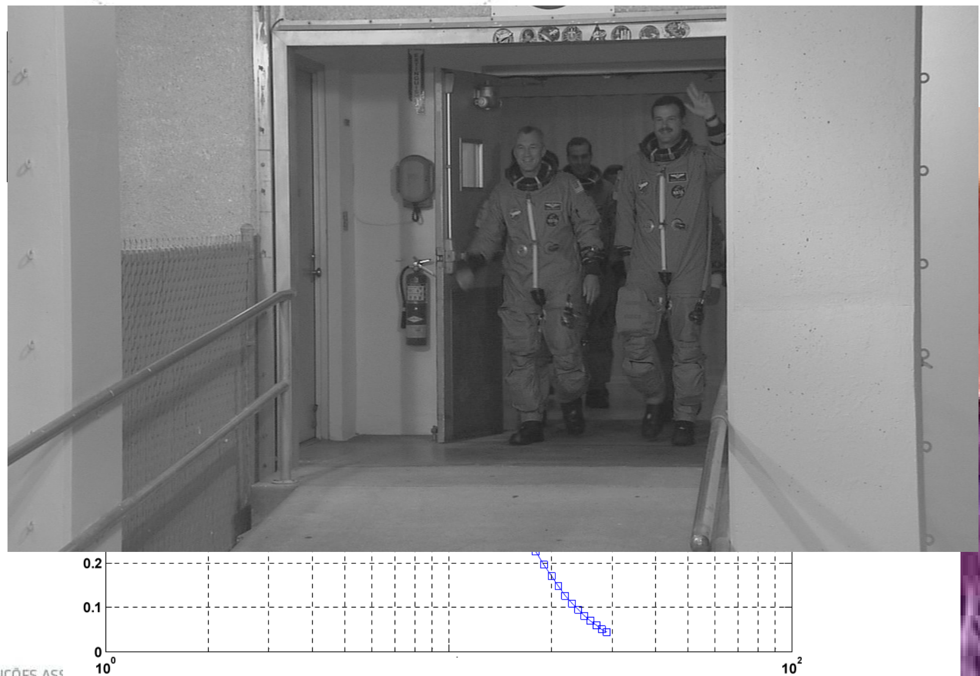
Objective: Integrate a JND model in the HEVC codec and design the extensions required to accommodate the novel adopted coding tools

2. Technical contributions

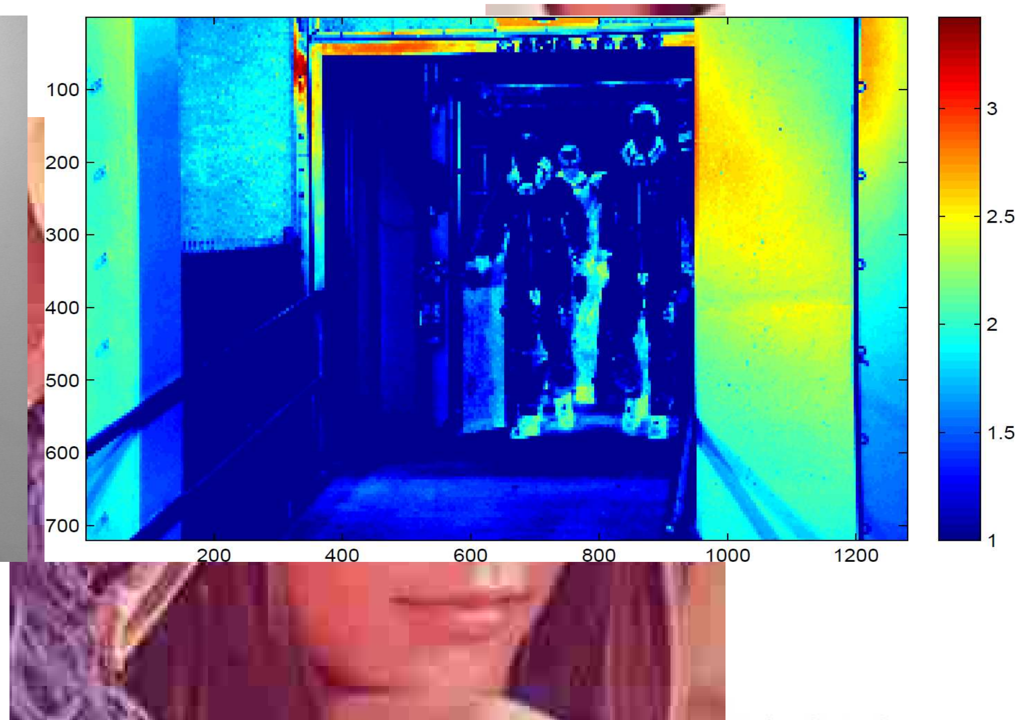
Adopted JND model

- The adopted JND model takes into account the HVS spatial masking mechanisms contributing to the final JND thresholds
- The JND thresholds are computed for **each frequency band** (*i,j*) in each *k* coded image block

$$JND(i,j,k) = JND_{band}(i,j) \cdot JND_{lum}(k) \cdot JND_{pat}(i,j,k)$$



Spatial frequency (ω)



JND model extension for HEVC

- **State-of-the-art JND models assume the following conditions**:
  - 8×8 block size for frequency decomposition
  - Floating point DCT for frequency decomposition
- **The HEVC introduced the following novelties**:
  - Several block sizes for frequency decomposition (4×4, 8×8, 16×16, 32×32)
  - Integer DCT and MDDT for 4×4 and 8×8 intra predicted blocks
- **To integrate the JND model in the HEVC codec the following extensions are proposed**:
  - Extensions for integer DCT and larger block sizes
  - Extensions for the MDDT

$$JND(i,j,k) = \boxed{JND_{band}(i,j)} \cdot JND_{lum}(k) \cdot JND_{pat}(i,j,k)$$

INSTITUIÇÕES ASSOCIADAS:

INSTITUTO SUPERIOR TÉCNICO

instituto de telecomunicações

Extension towards integer DCT and higher block sizes

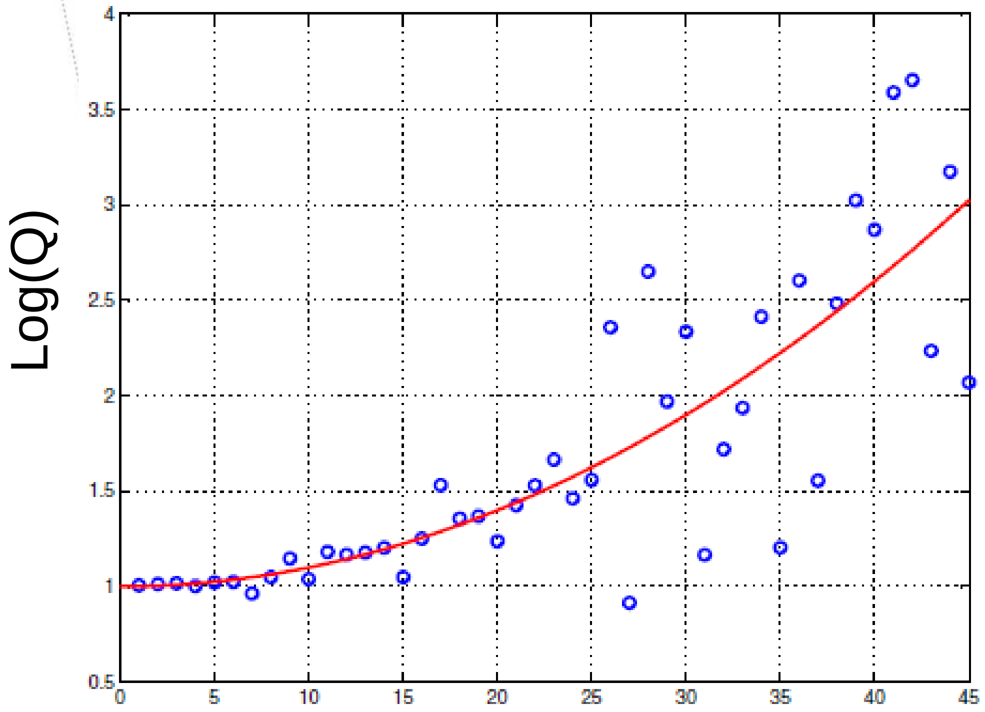
- For 4×4 and 8×8 block size, the DCT is the same as the H.264/AVC standard
- Therefore, a good JND_{band} model corresponds to the frequency weighting matrices provided with the H.264/AVC reference software
- For block sizes higher than 8×8, the following perceptual error metric M can be considered:

$$M = spatial\ pooling\left(\frac{\left| C(i,j,k) - \hat{C}(i,j,k) \right|}{JND(i,j,k)} \right)$$

Metric M quantifies, in JND units, the quantization error perceptual impact

Extended JND_{band} model by Andrew Watson's findings

- For any metric M value m and integer DCT, Andrew Watson found a quadratic relationship between the radial frequency f_r and the logarithm of the quantization step Q corresponding m



$$f_r = \sqrt{i^2 + j^2}$$

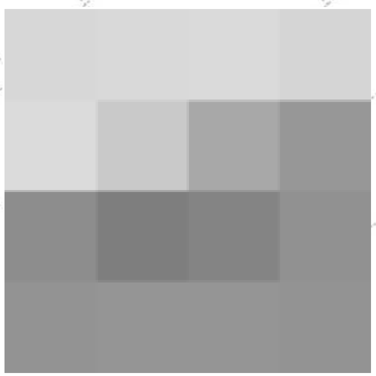
- By spatial pooling definition, Q corresponds to the JND threshold:

$$JND_{band}(i, j) = JND_{\min} \cdot e^{\ln(2)\cdot\left(\frac{f_r(i,j)}{w}\right)^2}$$

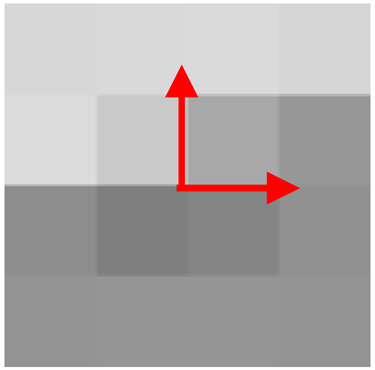
Extension towards the MDDT

- A frequency transform can be also seen as a reference system change
- MDDT rotates input data along the intra prediction direction

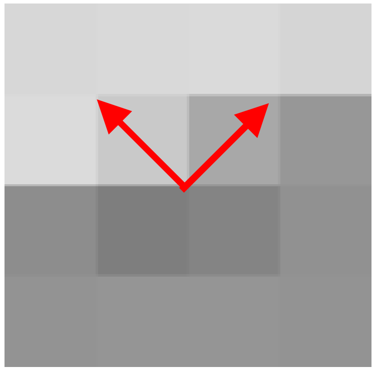
Input prediction residuals



DCT reference system



MDDT reference system



- It is assumed that also the HVS frequency band sensitivity rotates along this direction

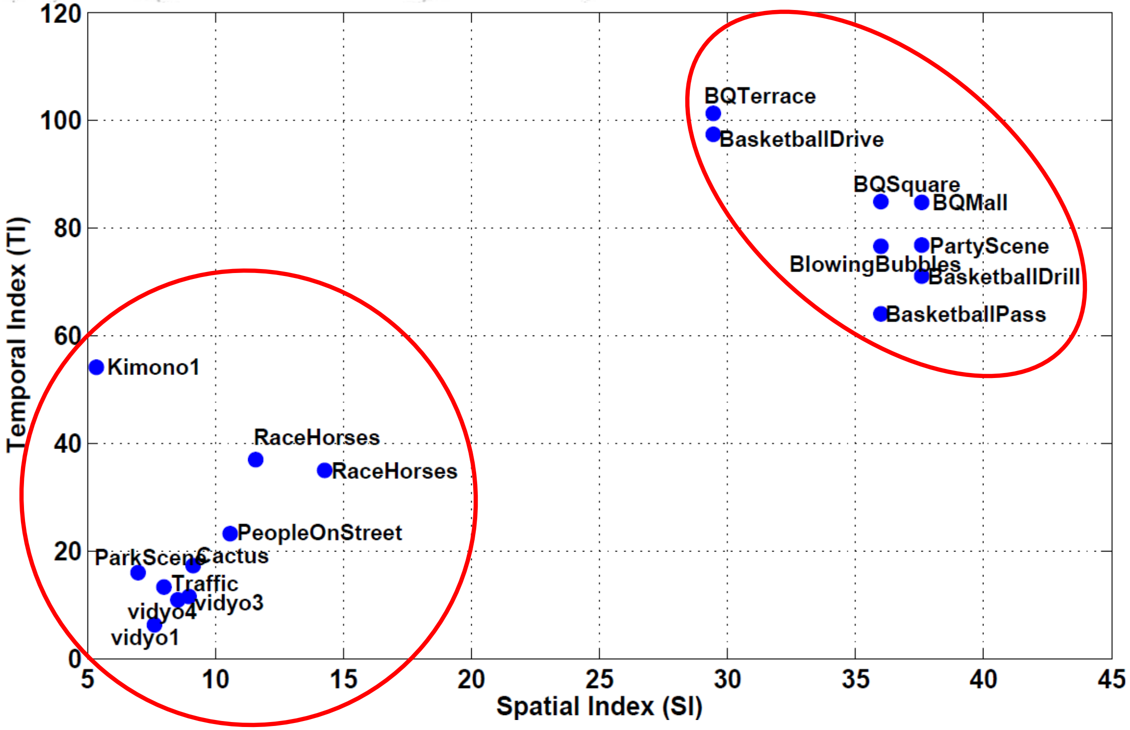
$$JND_{band}(i,j) = \Phi\left(JND_{band}^{DCT}, \alpha\right)$$

α: prediction direction angle

3. Experimental results

Test conditions

- JCT-VC high definition test video sequences (class A and B)
- JCT-VC “*random access – high efficiency*” coding parameters setting
- One representative sequence for each spatio-temporal resolution class



- ITU spatial and temporal indexes (SI-TI)

Assessment and benchmarking

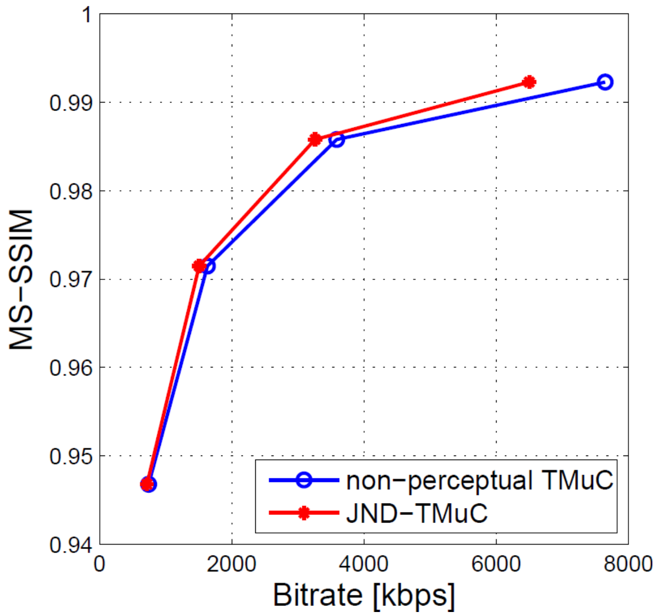
- **Assessment methodology**:
  - ***Distortion measure***: the Multi Scale Structural SIMilarity (MS-SSIM) metric is computed over the luminance component and averaged along all video frames
  - ***Objective quality score***: the metric resolving power is computed for the MS-SSIM to obtain a common subjective quality level
- **Benchmarking**:
  - ***JND-TMuC***: HEVC TMuC codec with the proposed JND model integration
  - ***TMuC***: HEVC TMuC codec without the proposed JND model integration

Results

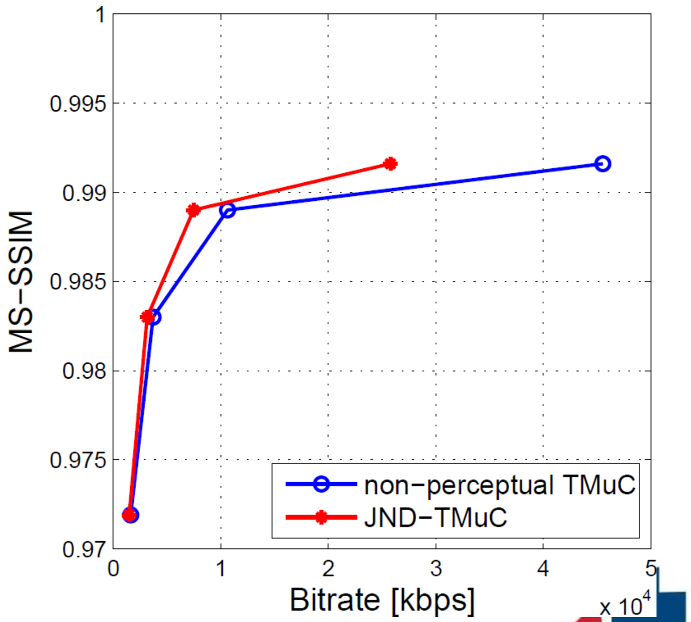
Sequence name	QPI	TMuC MS-SSIM	Bitrate [kbps]		Δ rate [%]
			TMuC	JND-TMuC	
ParkScene	22	0.9923	7646.9	6499.4	-15.00
	27	0.9858	3586.8	3260.5	-9.10
	32	0.9715	1634.8	1508.3	-7.73
	37	0.9468	744.9	720.4	-3.30
BQTerrace	22	0.9916	45517.2	25758.7	-43.40
	27	0.9890	10620.6	7471.5	-29.65
	32	0.9830	3700.5	3151.8	-14.83
	37	0.9719	1488.4	1631.9	-8.79
Average	-	-	-	-	-16.47

- **JND-TMuC provides the same perceptual quality and good bitrate reductions regarding the TMuC codec**
- **Highest bitrate reductions for the BQTerrace sequence: very bright areas**

Parkscene sequence



BQTerrace sequence



4. Conclusions and future work

Conclusions and future work

- A spatial JND model has been extended towards the HEVC coding tools and integrated in the TMuC (HM) codec
- **Advantages**:
  - **Good bitrate reductions** can be achieved regarding the TMuC codec at the same perceptual quality
  - The **available rate can be perceptually allocated** for each coded coefficient (JND model estimated at the decoder side, Naccari and Pereira, ICIP 2010)
- **Disadvantages**:
  - Increased computational complexity at both side (decoder side JND model estimation)
  - Bitstream lacks compliance with the HEVC format (not so problematic at this stage)
- **Future work**:
  - Inclusion in the JND model of temporal masking mechanisms
  - Subjective assessment for the proposed JND model extensions
  - Adaptive Loop Filter (ALF) perceptual optimization

Questions?

Thanks for your attention!