Extremely Randomised Trees for Computational Complexity Reduction of Omnidirectional Intra Video Coding

J. Filipe[1⋆], J. Carreira[1,2], L. Tavora[2], S. Faria[1,2], A. Navarro[1,3], P. Assuncao[1,2]
[1]Instituto de Telecomunicações, [2]Instituto Politécnico de Leiria/ESTG, PT;
[3]Universidade de Aveiro
⋆jose.filipe@ieee.org

Motivation

- Image and video data represent the majority of all internet traffic
- New applications and emerging services (e.g., 4K/8K video, cloud gaming and smart surveillance systems) are push requirements for more efficient video coding standards
- The next-gen coding standard, Versatile Video Coding (**VVC**) was recently finished, achieving higher coding efficiency gain at cost of increased computational complexity
- **VVC** is **7 to 9 times more complex than** High Efficiency Video Coding (**HEVC**)
- **Fast computational methods are of utmost importance to ease adoption of the this standard and to overcome implementation constraints**

Proposed Method

- Three Extremely Randomised Trees models
  - Quaternary Tree Partition
  - Binary Tree Partition
  - Ternary Tree Partition
- Predicts maximum depths of each partition type at CTU level
- Prediction from models used as **early termination**

Features

Table 1: List of Features

ID	Features
1	Latitude of the centre point of the CTU
2	Secant of the latitude of the centre point of the CTU
3	Spatial Information
4	Std. Dev. of Sobel filtered CTU along x
5	Std. Dev. of Sobel filtered CTU along y
6	Std. Dev. of Sobel filtered bottom left fourth of CTU, along y

- Set of 6 features, selected from 56 features using Recursive Feature Elimination
  - Features 1 and 2 were specifically designed with ERP projection in mind
  - Features 3 to 6 are known to be informative regarding the partition scheme (from literature)

Models

- Models predict maximum partition depth for each partition type at CTU level
- Training conditions
  - 10 ERP sequences, recommended by VVC CTCs
  - 10-Fold Cross-Validation (1 sequence for testing and 9 training)
  - All intra configuration, next profile, $QP = 22$

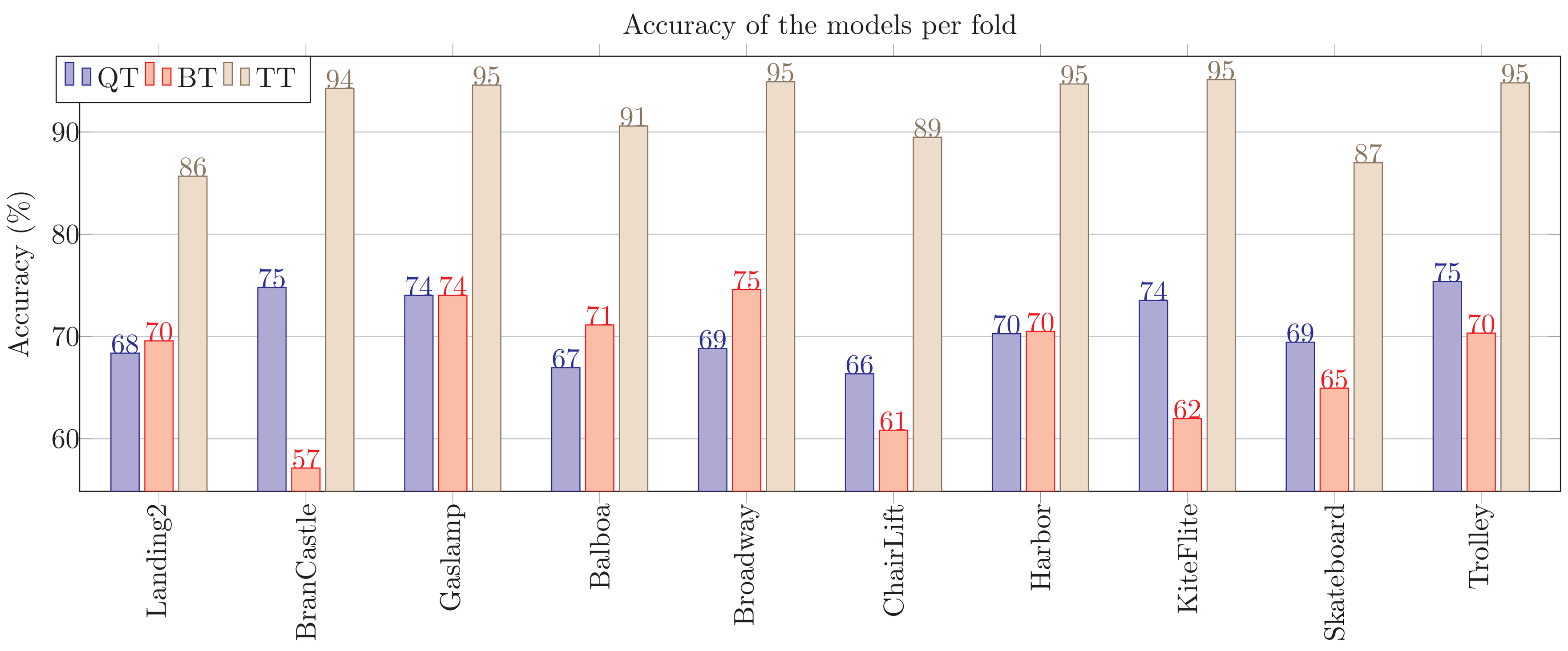


Table 2: Average Accuracy of the models

QT	BT	MT
$71 \pm 6\%$	$67 \pm 12\%$	$92 \pm 7\%$

Results

- Experimental Setup,
  - 10 test sequences with resolution of 4432×2216 pixels
  - VTM 8.0
  - All intra configuration, next profile, set of 4 QPs (22, 27, 32, 37)
  - **Coding efficiency evaluated using Bjontegaard Delta Rate (BD-Rate)**
  - **Coding complexity evaluated using:**

$$\Delta T = \frac{1}{4} \cdot \sum_{q} \frac{T^q_{pm} - T^q_{ref}}{T^q_{ref}}, \; q \in \{22, 27, 32, 37\} \quad (1)$$

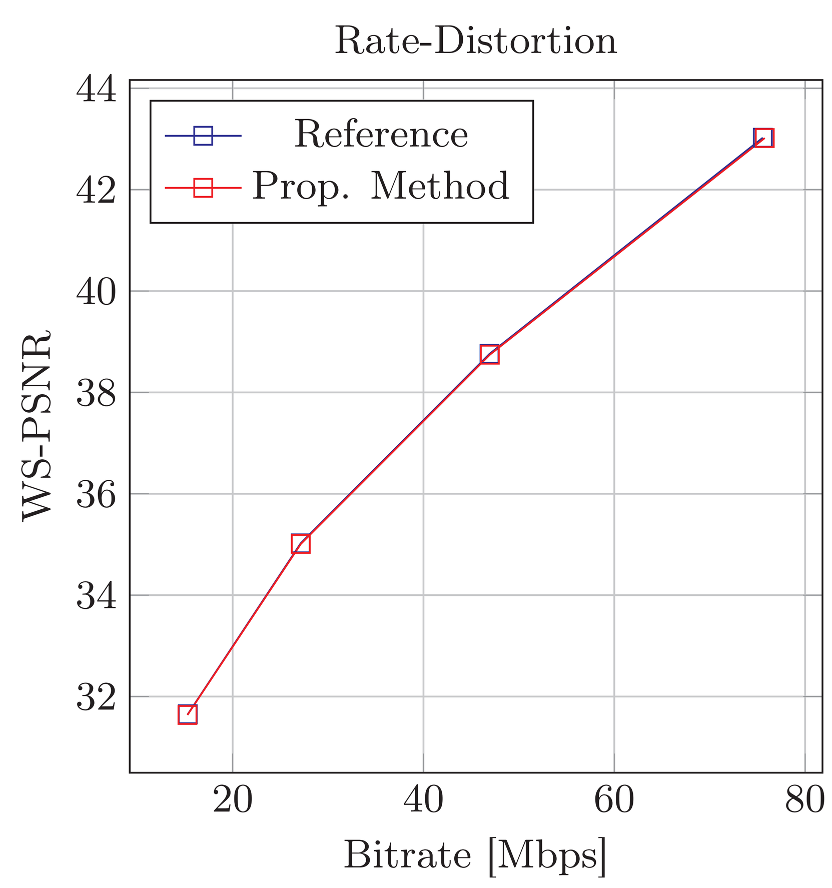
T^q_{ref} - Time spent to encode the sequence using regular VVC with $QP = q$
T^q_{pm} - Time spent to encode the sequence using the proposed method with $QP = q$

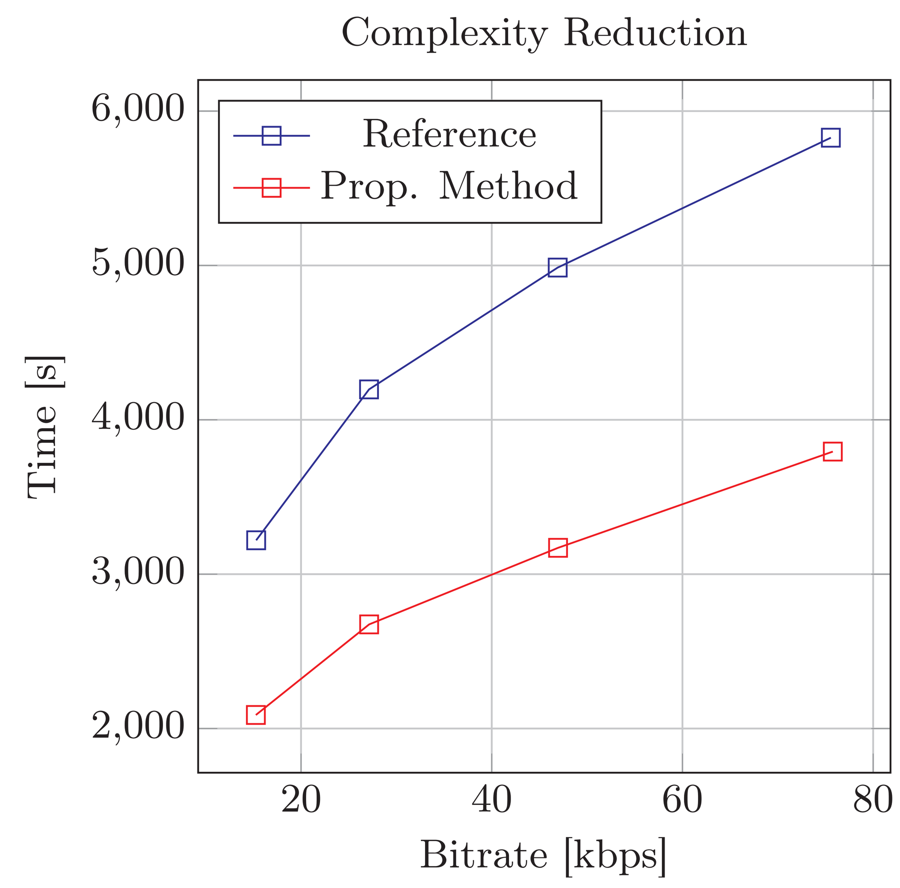
Table 3: Proposed method vs. reference VVC

Sequence	BD-Rate (%)	ΔT (%)
Harbor	0.79	−26.00
KiteFlite	0.42	−28.46
Balboa	0.36	−31.43
BranCastle	0.24	−35.71
Broadway	0.36	−31.23
Landing2	0.30	−35.88
SkateBoardInLot	0.96	−36.45
ChailiftRide	1.06	−35.92
Trolley	0.42	−31.27
Gaslamp	0.72	−25.69
Average	0.42	−31.35

- The Proposed method significantly reduces the complexity, when compared to the VVC reference implementation
- A negligible loss of coding efficiency (below 1%) for 9 out of the 10 sequences
- Average complexity reduction: 31.35%
- Average increase in bitrate for a given visual quality: 0.42%
- Percentage of complexity reduction that a given method can achieve per each 1% of BD-Rate loss: 74.30
- Outperforms other state of the art methods

BranCastle Case Study





Conclusions

- A novel algorithm that leverages 3 ERT models to predict maximum partition depth of each type (QT, BT, TT) in Intra frames of omnidirection sequences in ERP format
- Prediction is used to early terminate RDO process when the maximum depth is reached
- Proposed method achieves complexity reduction of about 31%, with negligible coding efficiency loss 0,42%, outperforming the state of the art low complexity solutions

Supported by Programa Operacional Regional do Centro, project ARoundVision CENTRO-01-0145-FEDER-030652 and FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020, Portugal.