Prenc – Predict Number of Video Encoding Passes with Machine Learning

Steve Göring, Rakesh Rao Ramachandra Rao, Alexander Raake

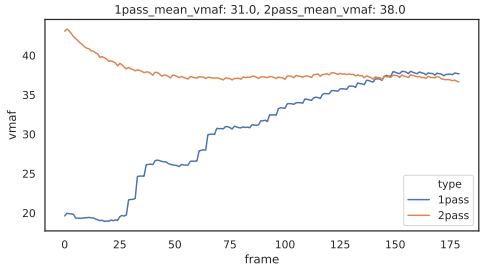
Audiovisual Technology Group,
Technische Universität Ilmenau, Germany;
Email: [steve.goering, rakesh-rao.ramachandra-rao,
alexander.raake]@tu-ilmenau.de

May 7, 2020

TECHNISCHE UNIVERSITÄT
ILMENAU

Motivation

TECHNISCHE UNIVERSITÄT
ILMENAU



- video providers: several advanced encoding strategies
- quality difference between **two and one pass** encoded videos
- important for quality models
- general idea: reverse engineer encoding settings based on bitstream data

$\rightarrow$ prenc = **pre**diction of number of **enc**oding passes

Related work

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Related work

TECHNISCHE UNIVERSITÄT
ILMENAU

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Related work

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Related work

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Related work

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Related work

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

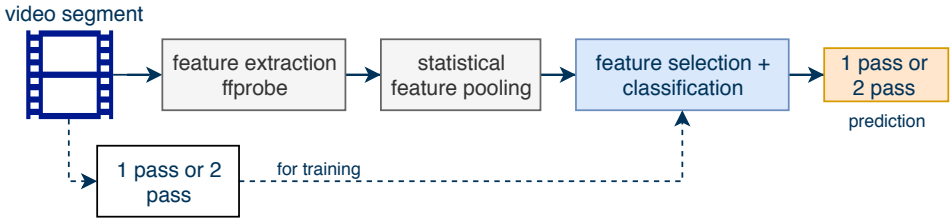
Related work

- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Related work

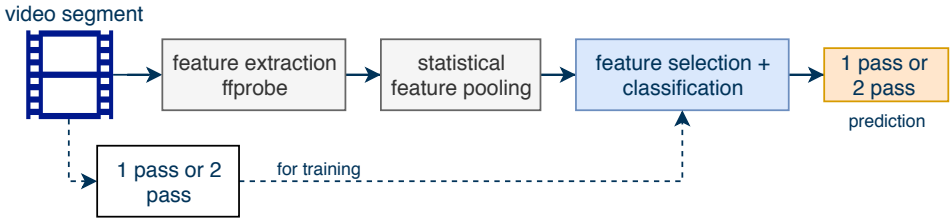
- bitstream based quality models:
  - ITU-T's P.1203: *Raake et al., Robitza et al., ITU-T* [4, 6, 1]
  - Mode 3 model of P.1204: *ITU-T* [2]
  - SVR based: *Shahid, Rossholm, and Lövström* [7]
- from pixels to bitstream:
  - QP prediction h.264: *Tagliasacchi and Tubaro* [8]
  - GOP period estimation: *Ramakrishna, Mazumdar, and Bora* [5]
  - source video resolution: *Katsavounidis, Aaron, and Ronca* [3]

Our Approach– *prenc*



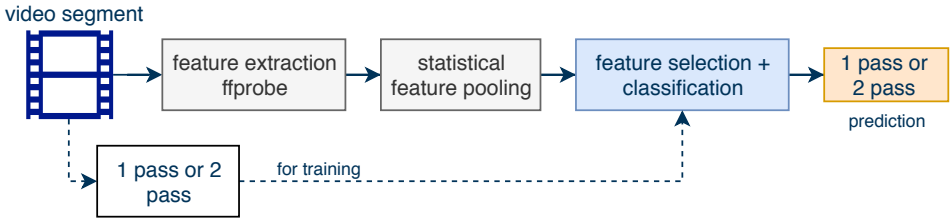
- 56 features based on ffprobe, *codec*
- framesizes: $mean_{all}, std_{all}, q^i_{all}$, $mean_{I,P,B}, std_{I,P,B}, q^i_{I,P,B}$; $\forall i \in [0, 10]$
- frametypes: r_I, r_P, r_B
- several ML algorithms applicable, e.g. RF, SVM

Our Approach– *prenc*



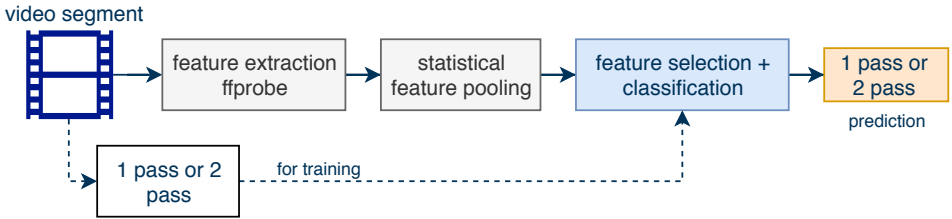
- 56 features based on ffprobe, *codec*
- framesizes: $mean_{all}, std_{all}, q^i_{all}$, $mean_{I,P,B}, std_{I,P,B}, q^i_{I,P,B}$; $\forall i \in [0, 10]$
- frametypes: r_I, r_P, r_B
- several ML algorithms applicable, e.g. RF, SVM

Our Approach– *prenc*



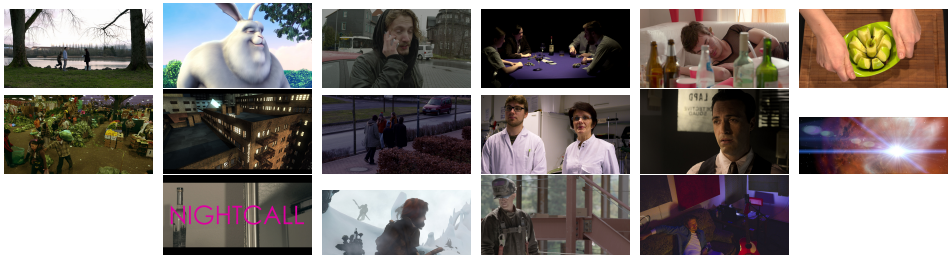
- 56 features based on ffprobe, *codec*
- framesizes: $mean_{all}, std_{all}, q^i_{all}$, $mean_{I,P,B}, std_{I,P,B}, q^i_{I,P,B}$; $\forall i \in [0, 10]$
- frametypes: r_I, r_P, r_B
- several ML algorithms applicable, e.g. RF, SVM

Our Approach– *prenc*



- 56 features based on ffprobe, *codec*
- framesizes: $mean_{all}, std_{all}, q^i_{all}$, $mean_{I,P,B}, std_{I,P,B}, q^i_{I,P,B}$; $\forall i \in [0, 10]$
- frametypes: r_I, r_P, r_B
- several ML algorithms applicable, e.g. RF, SVM

Evaluation – Dataset



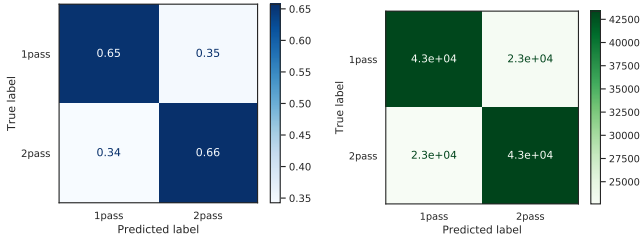
- 10 / 16 src videos own; all $\geq$ 3 minutes video duration
- represent several short video genres
- uncompressed, 4:2:2 chroma sub-sampling, most 10 bit

Evaluation –Encoding

Resolution	Bitrates [Mbit/s]
360p	[0.25, 0.5, 1.0]
480p	[0.3, 0.6, 1.2]
720p	[0.5, 1.0, 2.0]
1080p	[2.0, 4.0, 8.0]
1440p	[3.0, 6.0, 12.0]
2160p	[4.0, 8.0, 16.0]

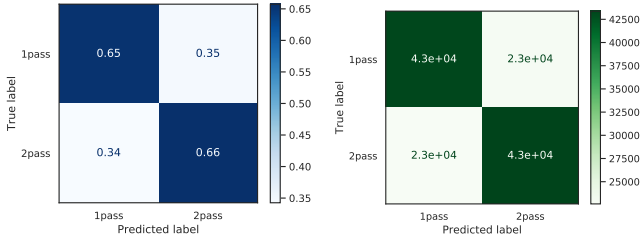
- 1-pass and 2-pass fixed bitrate encoding (50% each)
- several bitrate and resolutions; h.264 and h.265
- 72 different encoding settings for a given video
- encoding performed using FFmpeg 4.1.3
- DASH segmentation after encoding (4 s segment length) $\rightarrow$ 131.976 segments

Evaluation – Prediction



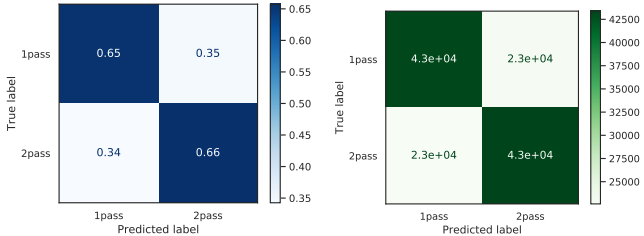
- 10-fold evaluation (10 repetitions each)
- with several algorithms: RF, SVC, GBC, KNN
- feature selection: $FS(0)$, $FS(0.5)$, $FS(1.0)$
- **best:** RF model with $FS(0)$ and 150 trees

Evaluation – Prediction



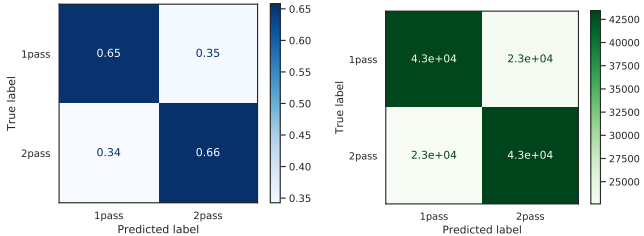
- 10-fold evaluation (10 repetitions each)
- with several algorithms: RF, SVC, GBC, KNN
- feature selection: $FS(0)$, $FS(0.5)$, $FS(1.0)$
- **best:** RF model with $FS(0)$ and 150 trees

Evaluation – Prediction



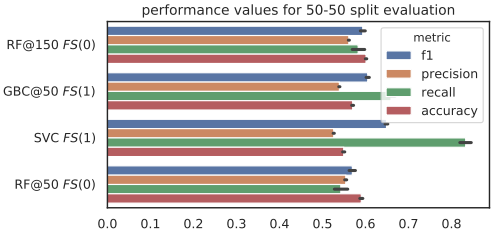
- 10-fold evaluation (10 repetitions each)
- with several algorithms: RF, SVC, GBC, KNN
- feature selection: $FS(0)$, $FS(0.5)$, $FS(1.0)$
- **best:** RF model with $FS(0)$ and 150 trees

Evaluation – Prediction



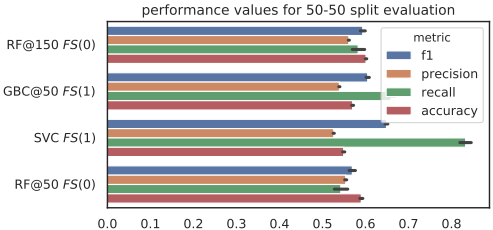
- 10-fold evaluation (10 repetitions each)
- with several algorithms: RF, SVC, GBC, KNN
- feature selection: $FS(0)$, $FS(0.5)$, $FS(1.0)$
- **best:** RF model with $FS(0)$ and 150 trees

Evaluation – Prediction



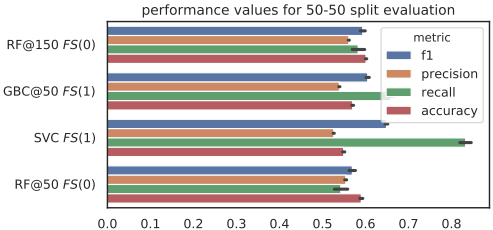
- 50% - 50% split with source video overlapp (10 repetitions each)
- RF: $FS(0)$, 50/150 trees; SVC: $FS(1)$; GBC: $FS(1)$, 50 trees
- **best:** SVC

Evaluation – Prediction



- 50% - 50% split with source video overlapp (10 repetitions each)
- RF: $FS(0)$, 50/150 trees; SVC: $FS(1)$; GBC: $FS(1)$, 50 trees
- **best:** SVC

Evaluation – Prediction



- 50% - 50% split with source video overlapp (10 repetitions each)
- RF: $FS(0)$, 50/150 trees; SVC: $FS(1)$; GBC: $FS(1)$, 50 trees
- **best:** SVC

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT ILMENAU

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

TECHNISCHE UNIVERSITÄT
ILMENAU

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

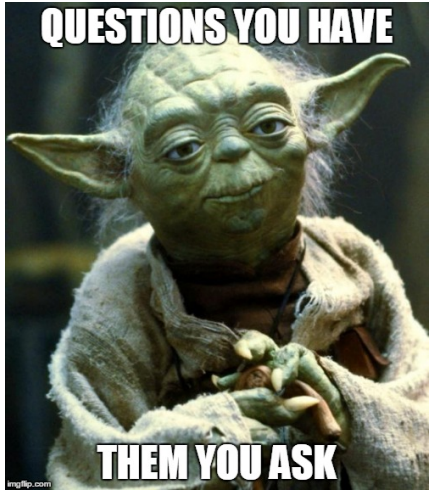
- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Conclusion, Summary and Future Work

- overview of **prenc**
  - prediction of the number of encoding passes
  - large scale dataset and evaluation
- mode 1 features
  - seem to be feasible, results can be improved
  - RF and SVC based models best
- open and next steps:
  - evaluate to predict other encoding settings
  - include higher features (mode 3)

Thank you for your attention

TECHNISCHE UNIVERSITÄT
ILMENAU



. . . . . are there any questions?

References I

[1] ITU-T. *Recommendation P.1203 - Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport*. Tech. rep. Int. Telecommunication Union, 2016.

[2] ITU-T. *Recommendation P.1204.3 - Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full bitstream information*. Tech. rep. Int. Telecommunication Union, 2019.

[3] Ioannis Katsavounidis, Anne Aaron, and David Ronca. "Native resolution detection of video sequences". In: *Annual Technical Conference and Exhibition, SMPTE 2015*. SMPTE. 2015, pp. 1–20.

References II

[4] Alexander Raake et al. “A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1”. In: *QoMEX*. Erfurt: IEEE, May 2017.

[5] Perla Ramakrishna, Aniruddha Mazumdar, and Prabin K Bora. “Blind forensics method for GOP period detection in motion compensated video”. In: *Twenty Second National Conference on Communication*. IEEE. 2016, pp. 1–6.

[6] Werner Robitza et al. “HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 – Open Databases and Software”. In: *9th ACM Multimedia Systems Conference*. Amsterdam, 2018.

References III

[7] Muhammad Shahid, Andreas Rossholm, and Benny Lövström. "A no-reference machine learning based video quality predictor". In: *QoMEX*. IEEE. 2013, pp. 176–181.

[8] Marco Tagliasacchi and Stefano Tubaro. "Blind estimation of the QP parameter in H. 264/AVC decoded video". In: *WIAMIS*. IEEE. 2010, pp. 1–4.