PERFORMANCE EVALUATION OF MOBILE VIDEO QUALITY ESTIMATORS

Michal Ries

Institute of Communications and Radio-Frequency Engineering,

Vienna University of Technology, Austria.

2nd Seminar on Service Quality Evaluation in Wireless Networks, 12th of June 2007

Overview

- Subjective quality tests
- Content based video quality estimation
- Motion based video quality estimation
- Ensemble based video quality estimation
- Performance evaluation of proposed metrics

Scope of the Work

- GOAL: Design low complexity video quality metric for mobile video streaming.

- Define motion and color features suitable for content classification and quality estimation.

- Quality and content estimation within one cut.
  - Temporal video segmentation

- Introduce new statistical method for content classification.

Subjective Quality Tests

- Test methods: (rec. ITU-T P.910)
  - Absolute Category Rating (ACR)
  - Scaling: 5 grade MOS scale. 1: Bad
    5: Excellent

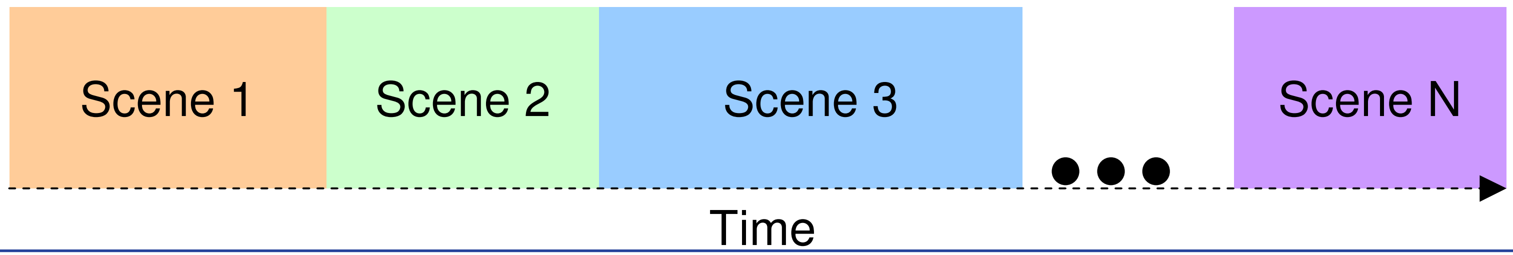
- Test conditions:
  - Terminal: VPA IV UMTS/WLAN
  - Codec: H.264 baseline profile
  - Resolution: SIF (240 x 320)
  - Two rounds with identical conditions
  - 26 persons in training set
  - 10 persons in evaluation set – different sequences

Content Classes

- We have defined five most frequent content classes:

Design of Video Quality Estimator

- Video quality estimation is based on character of the sequence.
- The video quality is estimated within two cuts :
  - Scene change detector allows for temporal segmentation of video stream.
- Scene change detection is based on variable threshold algorithm.

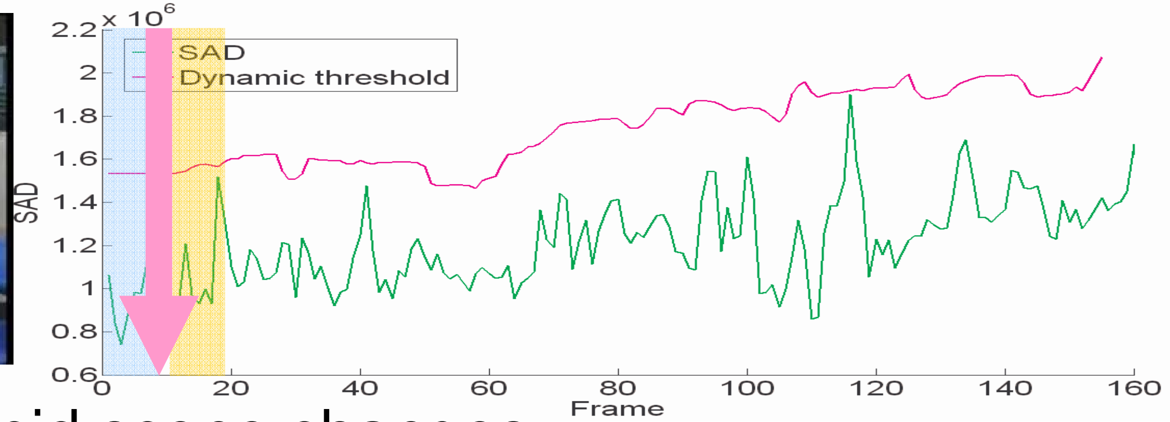


Scene Change Detector

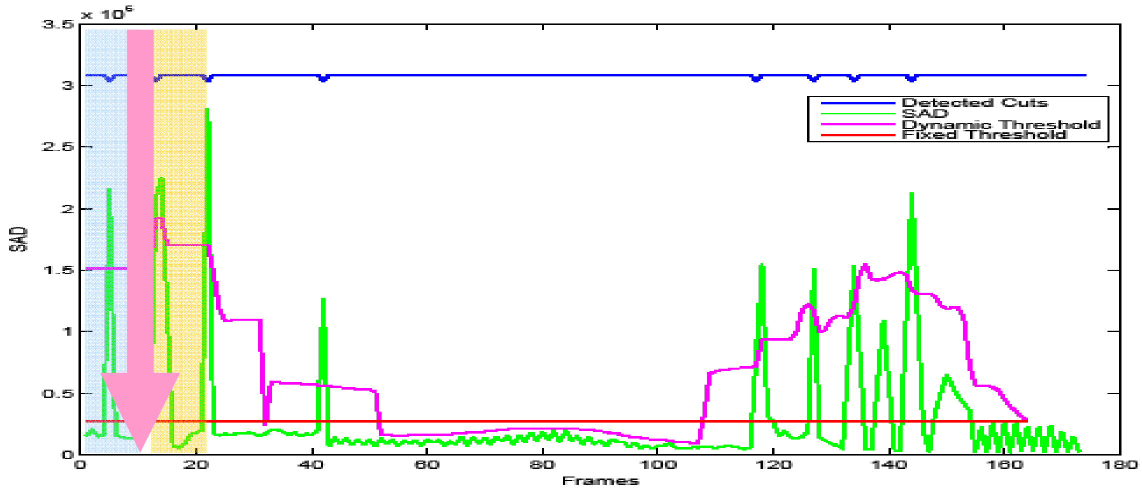
- Statistical features are computed over sliding window [n – 10, n + 10], 10 previous and 10 following frames.
- Scene change is detected using a local sequence statistical features of sum of absolute differences (SAD):
  - Empirical mean $$m_n = \frac{1}{N+1} \sum_{i=n-N-1}^{n-1} X_i$$
  - Standard deviation: $$\sigma_n = \sqrt{\frac{1}{N} \sum_{i=n-N-1}^{n-1} (X_i - m_n)^2}$$
- The thresholding function: $\mathrm{T}_n = a \cdot m_n + b \cdot \sigma_n$

Scene Change Detector

- Dynamic video without cuts



- Video with rapid scene changes

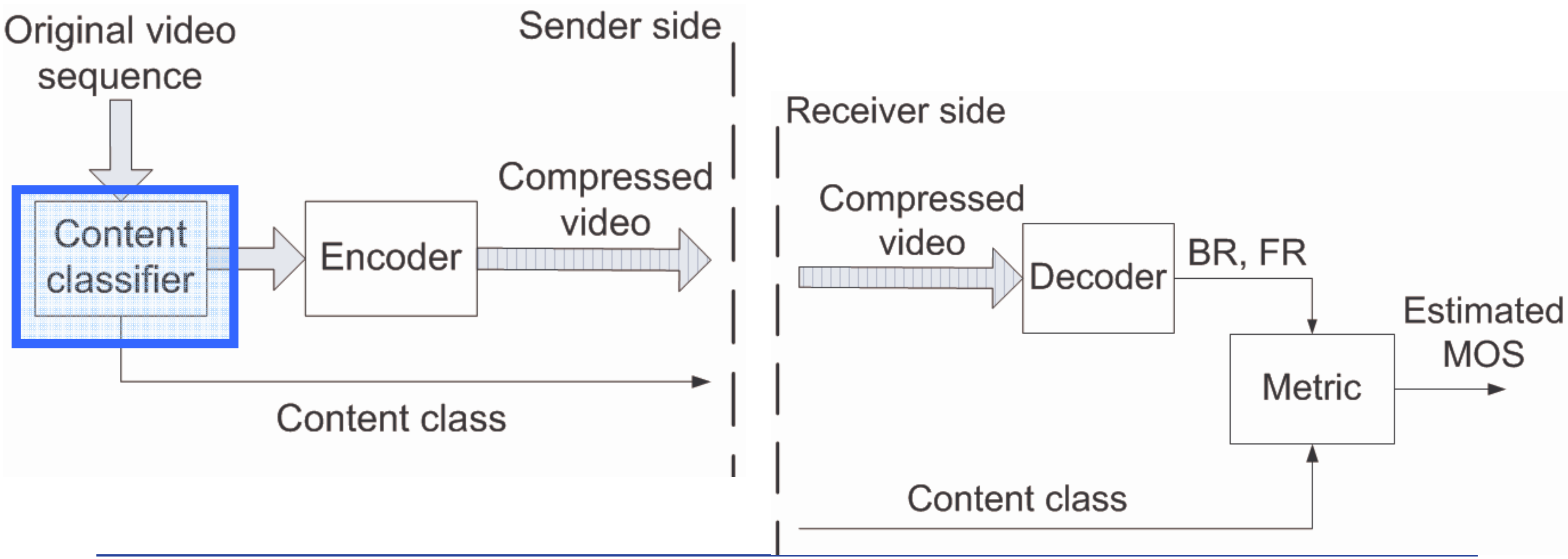


Content Based Video Quality Estimation- Highlights

- Content classification
  - Motion and color parameters
  - Hypothesis testing
- Design of video quality estimator
  - Metric design
  - Prediction performance
- Conclusions

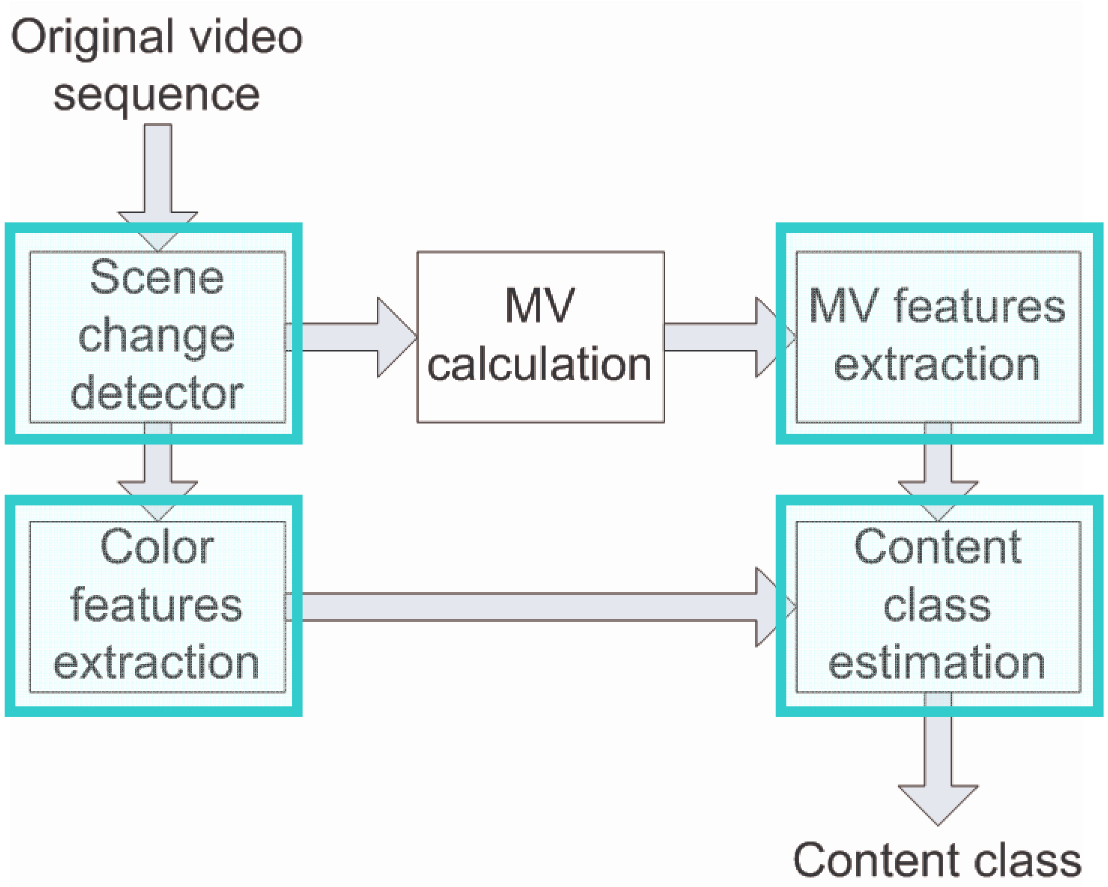
Design of Video Quality Estimator

- Two step approach:
  - content class estimation between two cuts on sender side,
  - subjective video quality estimation on receiver side.



Content Classification

- Content classification character of the movement and color features.



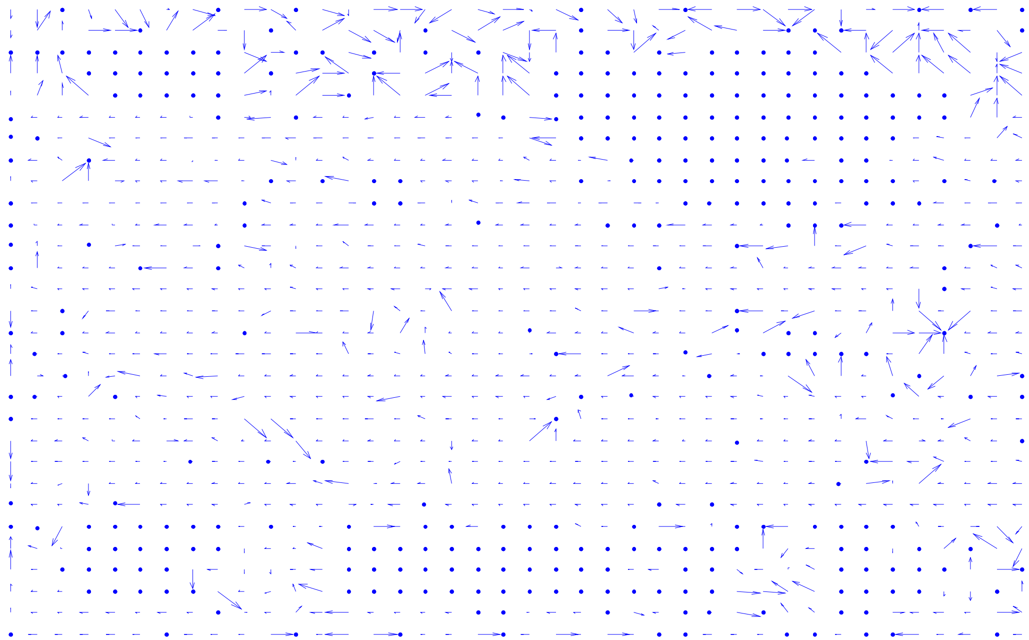
Motion and Color Parameters

- Color and motion parameters are extracted within one cut and calculated for each frame.
- Content classification is based on following features:

- Color features:
  - Percentage of green pixels in the frame,
  - Green color is determined by 5 bins from 64 bin color space.

Motion and Color Parameters

- Type and character of movement is defined:
  - Percentage of zero motion vectors (MV) in one frame.
  - Mean size of non zero MVs in one frame normalized by v



  - Per … direction (the … ame.
  - Dor … Vs. The dire … or (170; 190

INSTITUT FÜR
NACHRICHTENTECHNIK UND
HOCHFREQUENZTECHNIK

Michal Ries

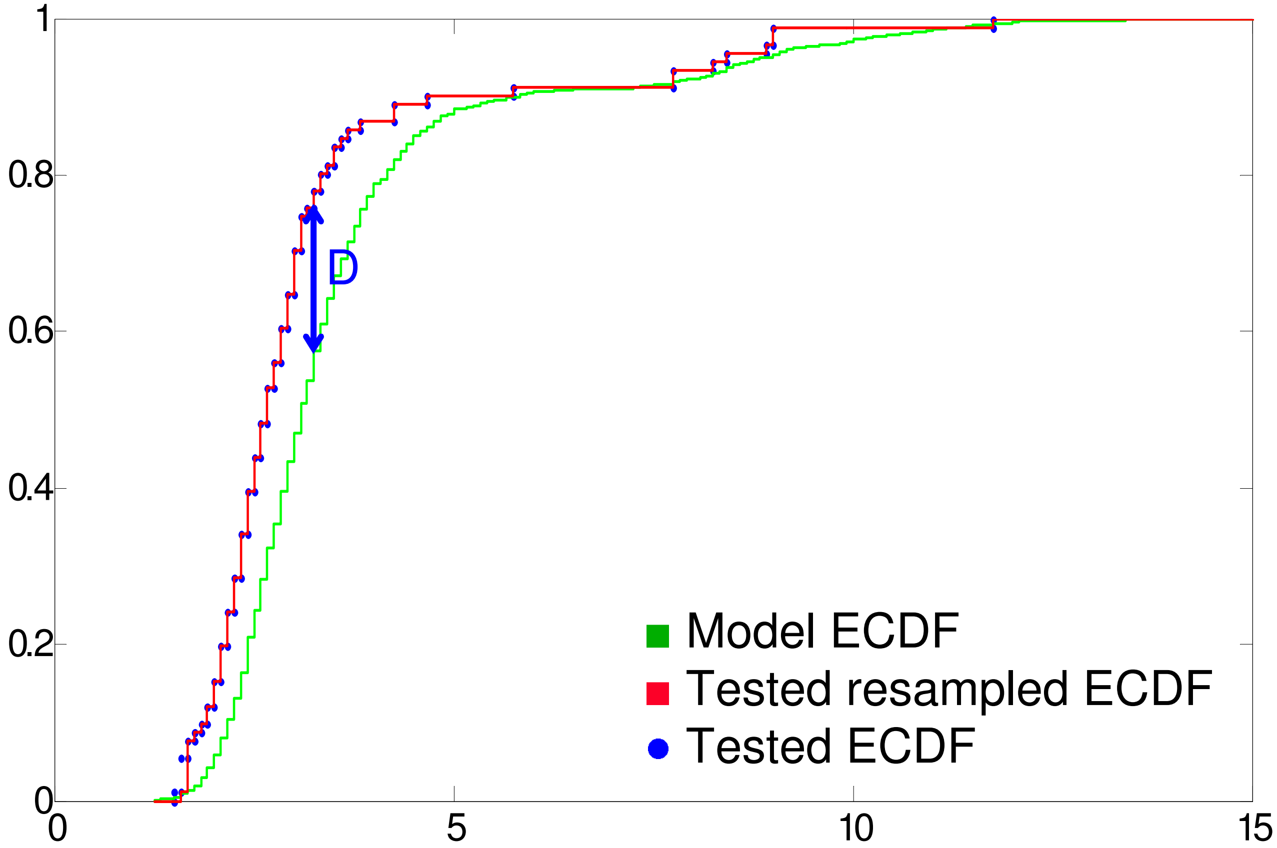
Hypothesis Testing

- We set up hypotheses for all content classes based on color and motion features.
- Find out the most suitable test for our empirical data sets:
  - Seek for difference between two datasets
  - Non parametric and distribution free
- Kolmogorov-Smirnov (KS) test is the most suitable
  - Max difference between two CDFs or ECDFs

$$D_n = \max\left[abs\left(F_n(x) - F_m(x)\right)\right]$$

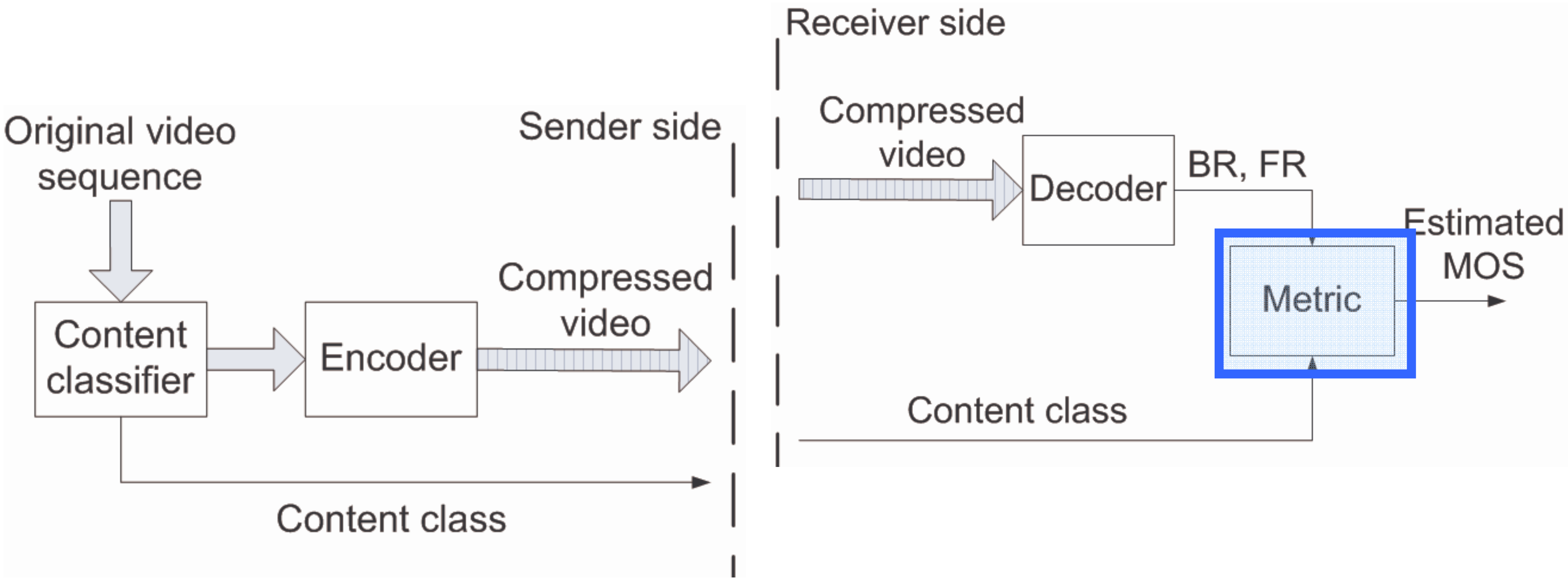
Hypothesis Testing - Data Processing

- ECDF of percentage of MVs pointing in the dominant direction.



Metric Design

- Low complexity reference free video quality metric.
- The content is signaled parallel with video stream.
- Metric is based on a zero complexity streaming/codec parameters.



Metric Design

- Reference-free video quality metric does not require knowledge about the original (non-compressed) sequence only content class is signaled.

- One video quality metric model with different coefficients for each content class:

$$MOS_V = K + A \cdot BR + \frac{B}{BR} + C \cdot FR + \frac{D}{FR}$$

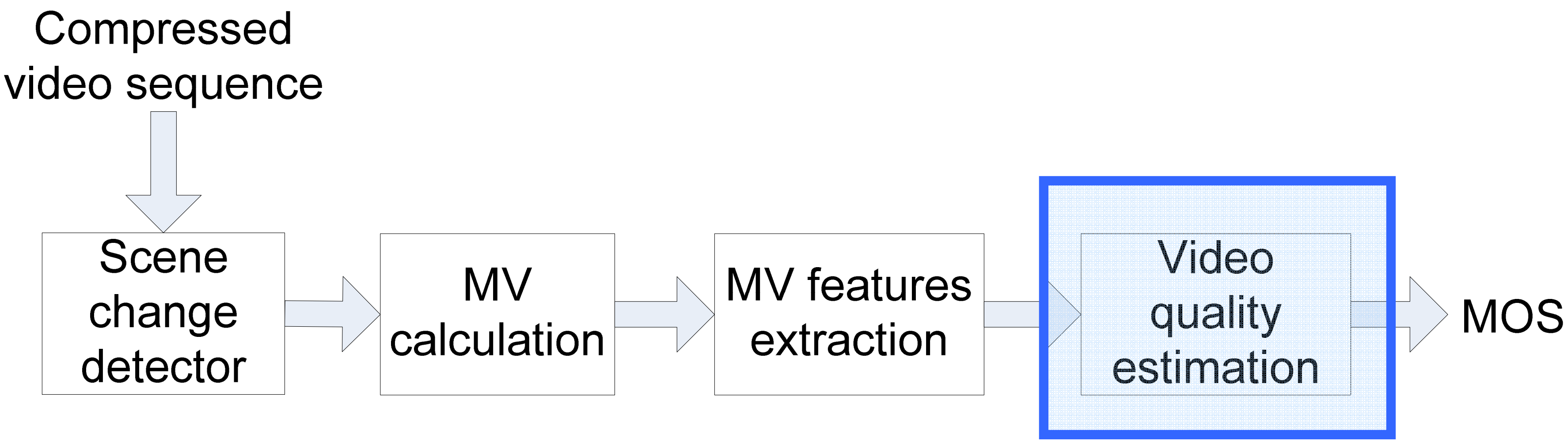
- Model coefficients vary substantially for the content classes – each class has two zero coefficients.

Full Reference Free Video Quality Estimation-Highlights

- Design motion based video quality metric for most frequently streamed content types
- Define the most relevant motion features based on multivariate statistical analysis.
- Single model estimation approach
- Ensemble based estimation approach

Design of Video Quality Estimator

- One step approach:
  - subjective video quality estimation on receiver side.
  - Quality is estimated with single or with ensemble of models



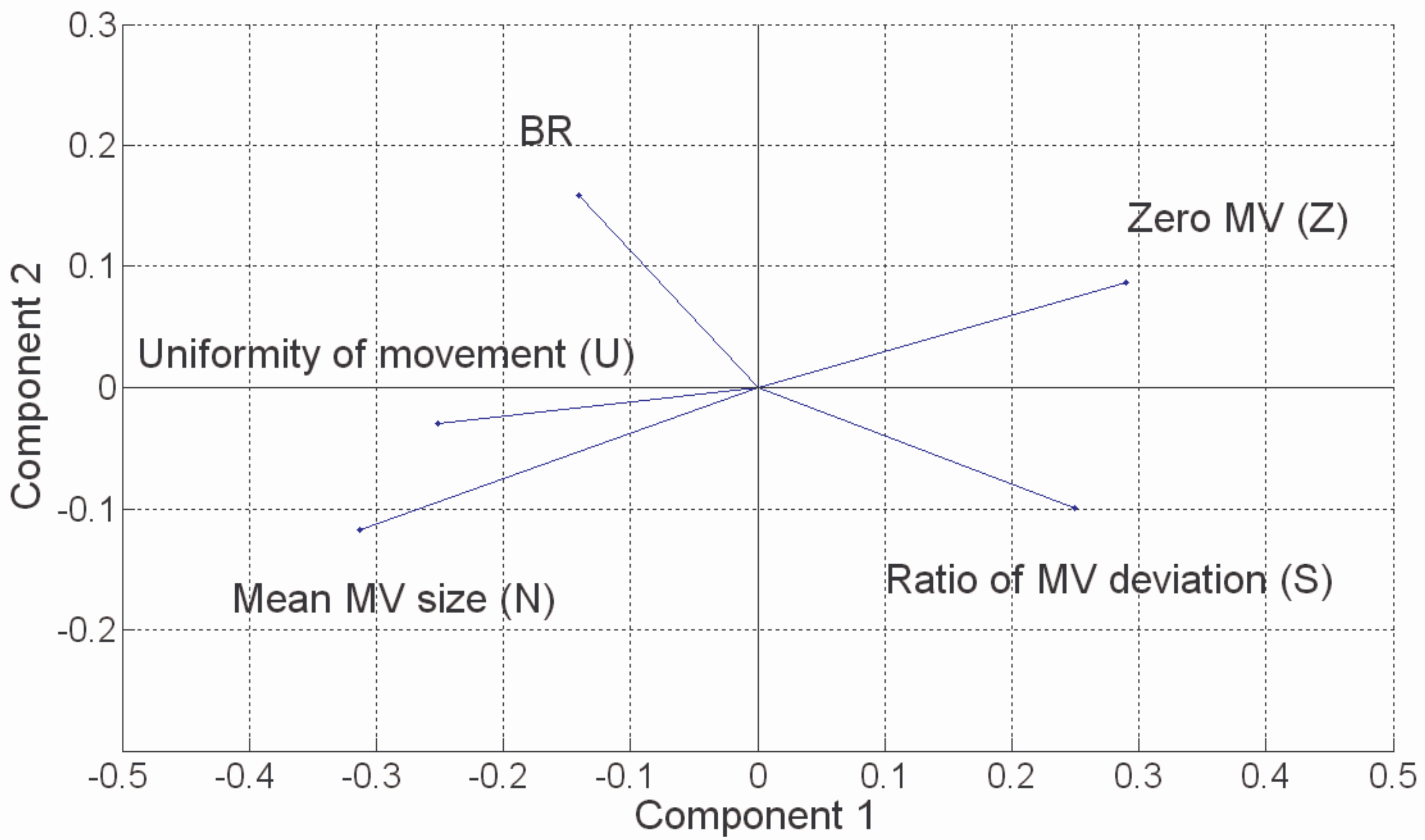
Motion Parameters

- MV features, BR and FR were investigated:
  - mean size of all MV
  - standard deviation of MV sizes
  - histograms of MV directions
  - variance of MV directions
  - proportion of horizontal movement
  - proportion of dominant MV direction

- Principal Component Analysis (PCA) was carried out to verify further applicability of the investigated parameters.

Motion Parameters

- Visualization of PCA results for all content classes.



Motion Parameters

- Type and character of movement is defined:

  Z - Percentage of zero motion vectors (MV) in one **cut**.

  N - Mean size of non zero MV in one **cut** normalized over width of sequence resolution (normalized to max length of MV).

  U - Percentage of MVs pointing in the dominant direction (the most frequent direction of MV) in one **cut**.

  S - Ratio of MV deviation within one shot refers to proportion of standard MV deviation within one shot.

Design of Video Quality Estimator

- Reference-Free video quality metric does not require knowledge about the original.

- Video quality metric model for all content classes:

$$MOS_V = k + a \cdot BR + c \cdot Z + d \cdot S^e + f \cdot N^2 + g \cdot \ln(U) + h \cdot S \cdot N$$

I'd like to Ask the Audience, …

- Ensemble based systems adopted following process for decision making:
  - Seeking additional opinions before making a decision.
  - Goal in doing so, is to improve our confidence that we are making the right decision, by weighing various opinions.

- Suitable application for ensemble based system is mapping objective data on subjective MOS results.
- Design of ensemble based perceptual quality metric for mobile video streaming services.

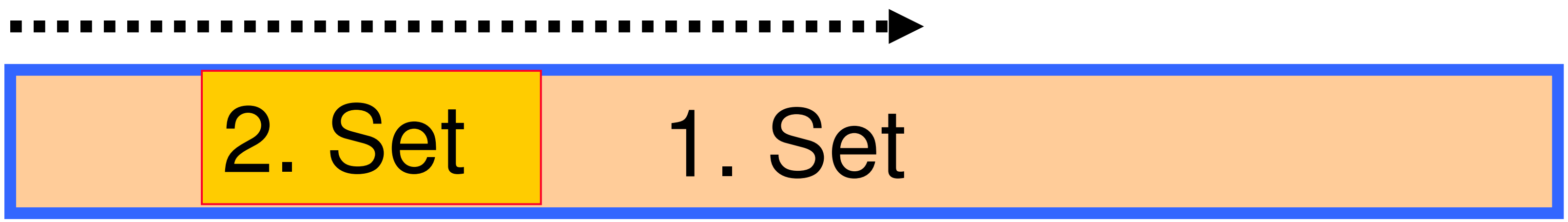
Ensemble of Models

- The ensemble generalization error is always smaller than or equal to the expected error of the individual models.

- An ensemble should consist of well trained but diverse models in order to increase the ensemble ambiguity:
  - estimators with significantly different decision boundaries from the rest of the ensemble set.

Ensemble of Models

- Cross validation improves generalization property of our ensemble set.
- Cross validation scheme:
  - Our data set is divided in two subsets and the models are trained on the first set.
  - The models are evaluated on the second set, the model with the best performance becomes ensemble member.
  - The data set is divided with light overlapping with previous subsets into two new subsets and the models are trained on the first set.



Ensemble of Models

- We use six estimation models in our ensemble.

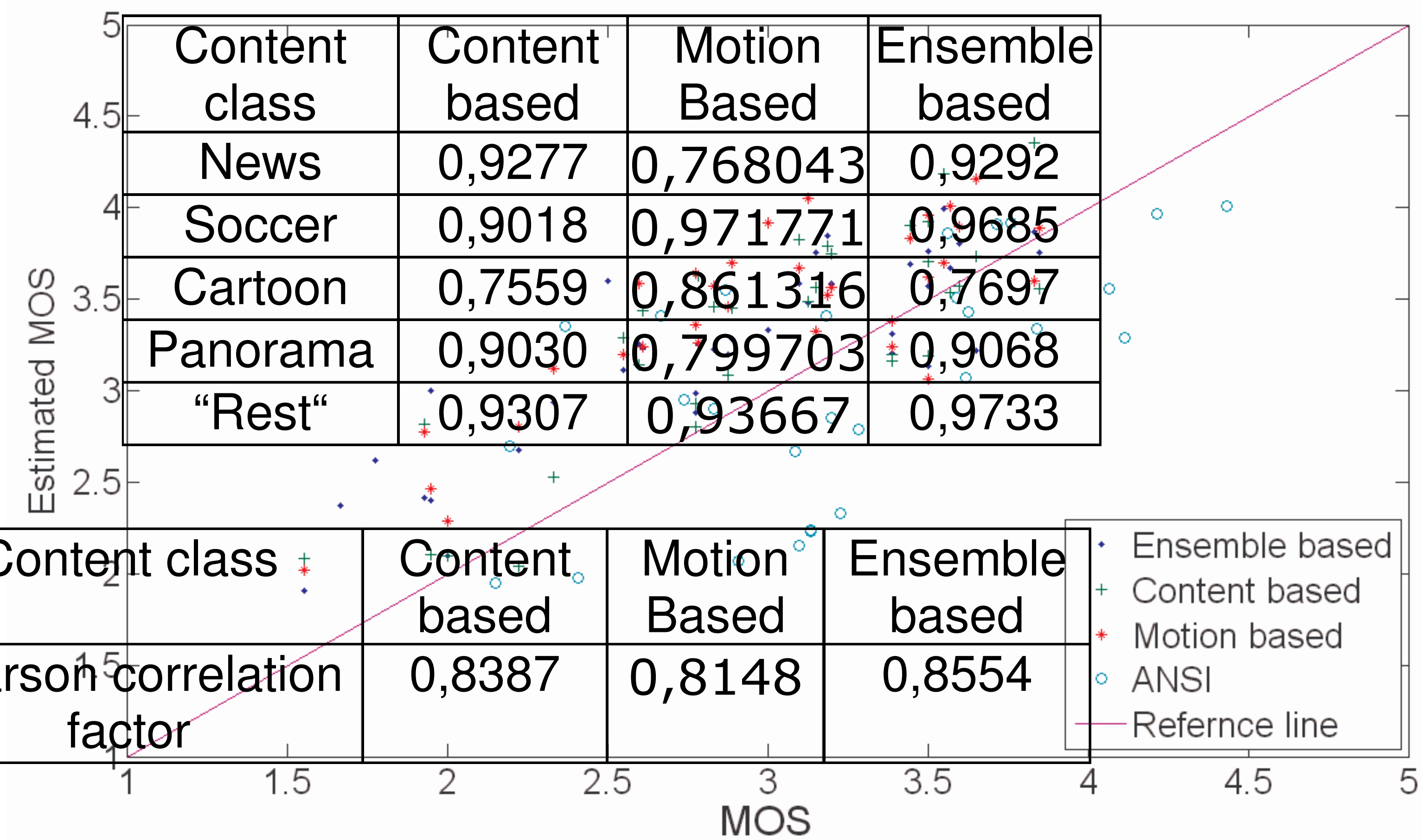
- Estimation models:
  - ***k*-nearest neighbor rule** (kNN) decision rule assigns to an unclassified sample point the classification of the nearest sample point of a set of previous classified points,
  - **artificial neural network** (ANN).
    - 90 neurons in hidden layer
    - learning method is improved resilient propagation with back propagation

Estimator‘s Performance

- Prediction performance (with evaluation set) of content based video quality metric:

Content class	Content based	Motion Based	Ensemble based
News	0,9277	0,768043	0,9292
Soccer	0,9018	0,971771	0,9685
Cartoon	0,7559	0,861316	0,7697
Panorama	0,9030	0,799703	0,9068
“Rest“	0,9307	0,93667	0,9733

Content class	Content based	Motion Based	Ensemble based
Pearson correlation factor	0,8387	0,8148	0,8554



Conclusions

- Content estimation based on KS-test is robust and an easily extendable method to new content classes and parameters.
- Content estimation on sender side allows for low complexity metrics on user/receiver side.
- Motion based video quality estimation allows for full reference free estimation.
- All proposed methods are less complex than the ANSI metric.

The End

Thank you for your attention

Tests setup

Questionnaire:

Nummer **Bewertung**

1 ***1*** ***2*** ***3*** ***4*** ***5***

Nicht gunügend **Genügend** **Befriedigend** **Gut** **Sehr gut**

• Nehmen Sie an, sie hätten für diesen Clip bezahlt. Waren Sie mit der Qualität des Clips zufrieden?

. **Ja** **Nein**

.

.

x ***1*** ***2*** ***3*** ***4*** ***5***

Nicht gunügend **Genügend** **Befriedigend** **Gut** **Sehr gut**

• Nehmen Sie an, sie hätten für diesen Clip bezahlt. Waren Sie mit der Qualität des Clips zufrieden?

Ja **Nein**

Future Work

- Audiovisual quality estimation
  - Extension of actual work on video quality

- Video and audiovisual quality of **CS** streaming

- Extension of content classification
  - New content classes
  - Audio content classification

Hypothesis Testing - Data Processing

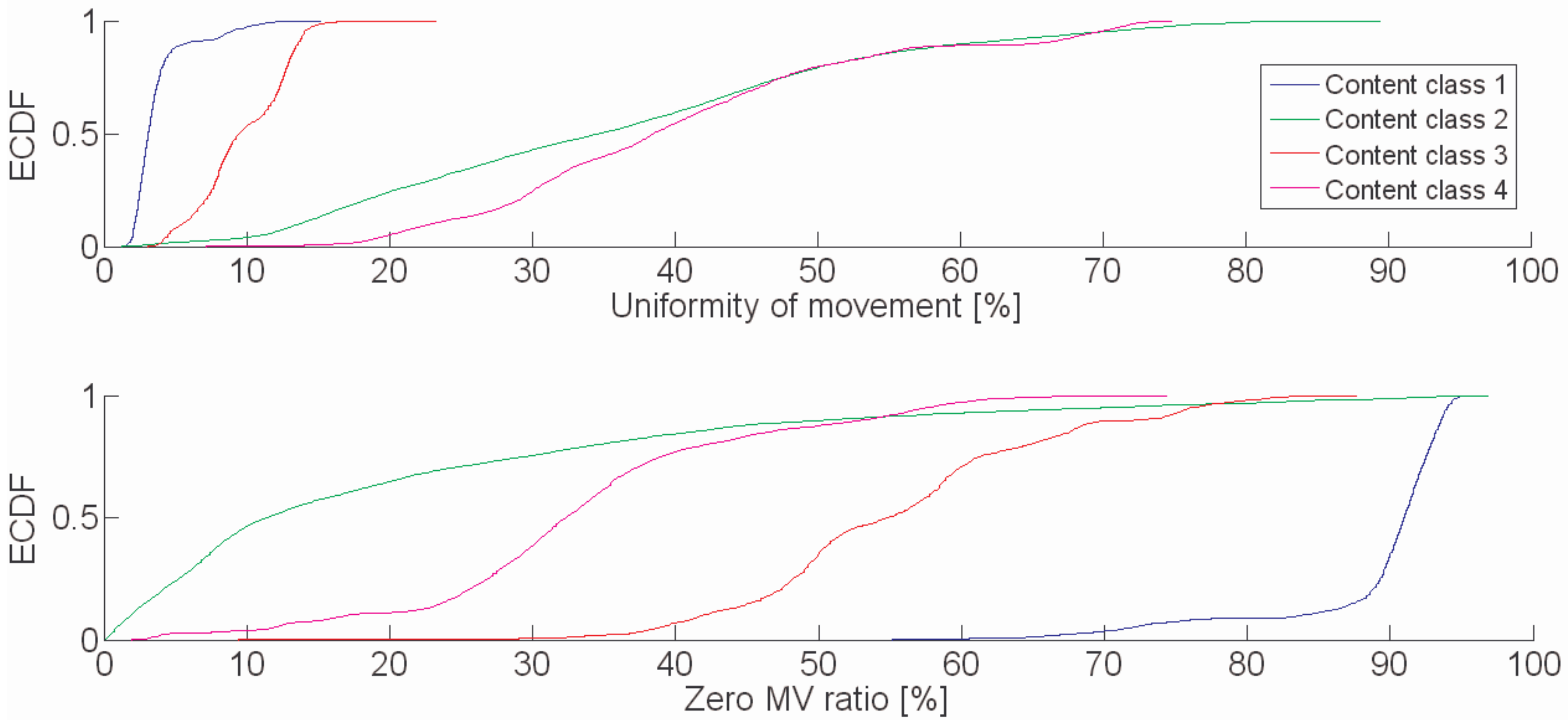
- Define critical D values
  - Maximal deviation from defined content classes.

- Calculate ECDFs for evaluated sequence within one cut.

- ECDF alignment:
  - execute KS-test for all defined parameters,
  - content class estimation.

Hypothesis Testing - Data Processing

- Estimate ECDF for defined content classis:
  - from sequence set of certain content class,
  - for all defined parameters.



Artificial Neural Network - Design

- Training:
  - Method – Automated regularization (Bayesian regularization with Levenberg-Marquardt training),
  - More efficient by preprocessing and postprocessing steps,
  - 54 vectors with three inputs: FR, BR, f_{SI13} and one target: MOS.

- Generalization:
  - Underfitting: too few hiden units,
  - Overfitting: too many hidden units,
  - Improvements: large vs. small enough network or much more points in training data than network parameters.

Used formulas

- Correlation factor

$$\mathrm{r} = \frac{(\mathbf{x} - \bar{\mathrm{x}})^T (\mathbf{y} - \bar{y})}{\sqrt{\left((\mathbf{x} - \bar{\mathrm{x}})^T (\mathbf{x} - \bar{\mathrm{x}})\right)\left((\mathbf{y} - \bar{y})^T (\mathbf{y} - \bar{y})\right)}}$$

- Mean squared error (MSE)

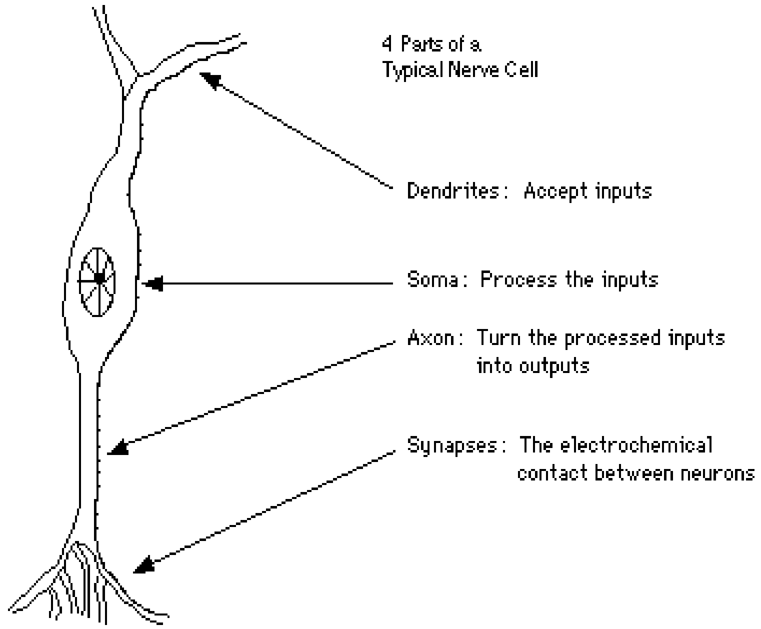
$$MSE = \frac{1}{N} (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})$$

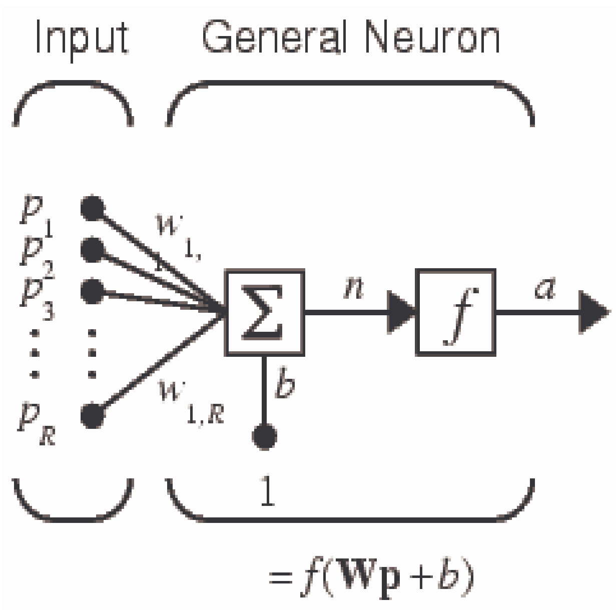
- Linear regression

$$\mathbf{y} = m\mathbf{x} + b$$

Artificial Neural Network - Neuron

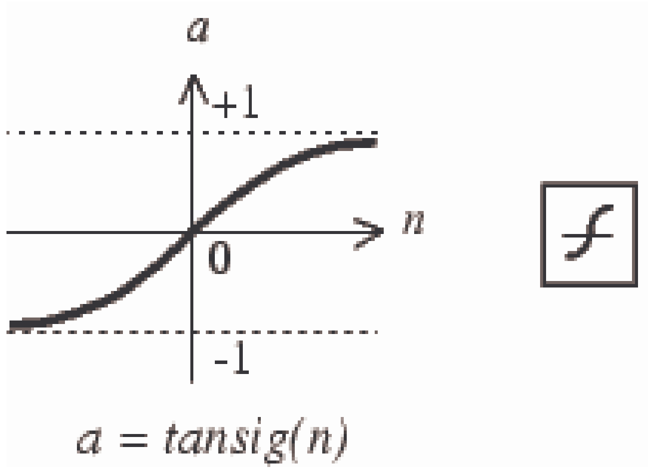
- Inspired by biological nervous systems
- Composed of large number of highly interconnected neurons
- After the training can behave like a “real“ human evaluating the streams

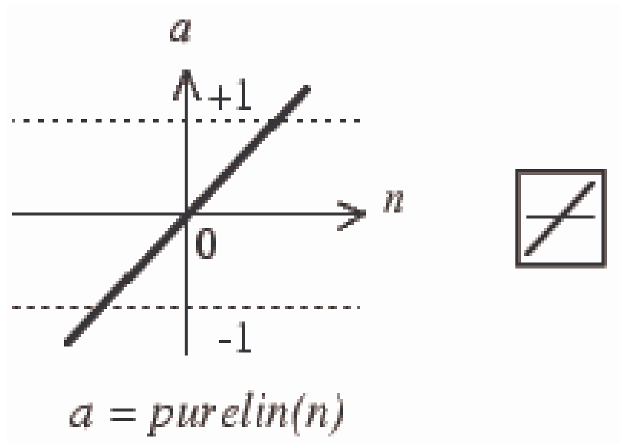




Transfer functions

- Tan-Sigmoid
  - Generates outputs between -1 and 1 as the neuron's net input goes from negative to positive infinity,
  - Nonlinear transfer function allows the network to learn nonlinear and linear relationships between input and output vectors.
- Linear
  - The linear output layer lets the network produce values outside the range -1 to +1.





Regularization

- Performance function, used for training feedforward neural networks, is the mean sum of squares of the network,

- Modify the performance function by adding a term that consists of the mean of the sum of squares of the network weights and biases,

$$\mathrm{msereg} = \gamma\,\mathrm{mse} + (1-\gamma)\,\mathrm{msw} \qquad \mathrm{msw} = \frac{1}{n}\sum_{j=1}^{n} w_j^2$$

- Causes that network have smaller weights and biases, and this will force the network response to be smoother and less likely to overfit,
- The problem - difficult to determine the optimum value for the performance ratio parameter γ.

Training algorithm

- Backpropagation learning:
  - Updates the network weights and biases in the direction in which the performance function decreases most rapidly - the negative of the gradient.

$$x_{k+1} = x_k - \alpha_k g_k$$

- Levenberg-Marquardt training:

$$x_{k+1} = x_k - \left[J^T J + \mu I\right]^{-1} J^T e$$

- Newton's method:

$$x_{k+1} = x_k - A_k^{-1} g_k$$