Universität
Konstanz

VQEG Meeting - Boulder, Colorado, USA

Kongfeng Zhu[1], Marcus Barkowsky[2], Minmin Shen[1],
Patrick Le Callet[2], and Dietmar Saupe[1]

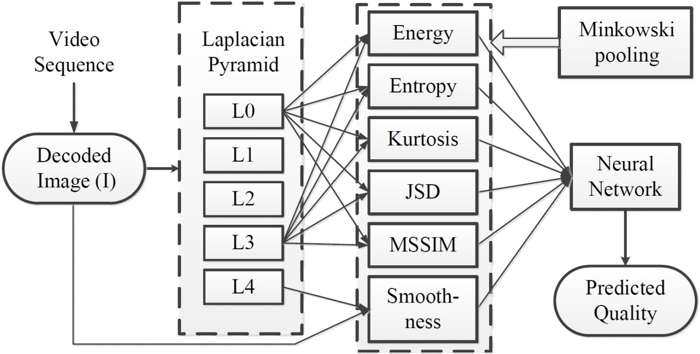
[1]Department of Computer and Information Science,
University of Konstanz, Germany

[2]LUNAM Universite de Nantes, Nantes, France

January 21, 2014

Training-based NR VQA

Universität Konstanz



Step 1
Laplacian Decomposition

Step 2
Features Extraction

Step 3
Quality Prediction

Motivation

Universität Konstanz

Neural network: overfitting

- Large number of parameters
- Small number of videos for training

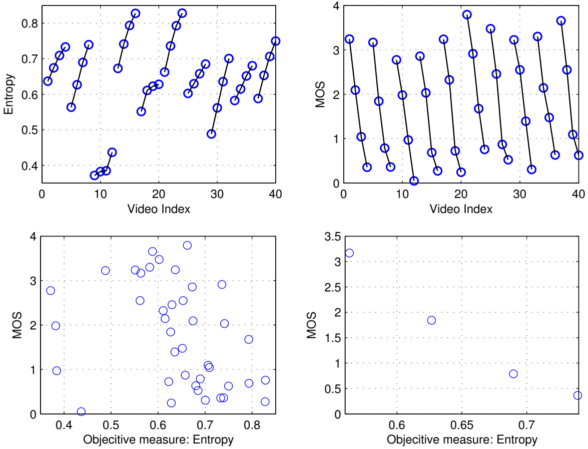
Goal and approach

- New non-linear mapping model
- Small number of parameters
- Analysis of the influence of video content

MOS VS. Entropy

Universität Konstanz

LIVE mobile video database



Notation

Universität Konstanz

Let

- one video: the nth video in the mth video set
- the predicted quality: $\hat{y}(m,n)$
- six extracted video-level features: $f_i(m,n)$, $i = 1, \cdots, 6$
- $m = 1, \cdots, M, n = 1, \cdots, N$
- M: the total number of video sets in the database
- N: the number of videos in one set

Non-linear mapping

University Konstanz

1) Global linear function

$$y'(m,n) = \Sigma_i w_i f_i(m,n), \tag{1}$$

2) Local alignment

$$y''(m,n) = s(m)y(m,n) + b(m), \tag{2}$$

3) Quality calibration

$$\hat{y}(m,n) = g(y''(m,n)), \tag{3}$$

$$g(x) = \frac{\beta_1 - \beta_2}{1 + \exp(-(x - \beta_3)/|\beta_4|)} + \beta_2$$

Prediction: scale s and offset b

Universität Konstanz

- Prediction of the offset b

$$\hat{b}(m) = as(m), \quad m = 1, \cdots, M. \tag{4}$$

- Prediction of the scale s

$$\hat{s}(m) = \alpha_3 f_2^3(m) + \alpha_2 f_2^2(m) + \alpha_1 f_2(m) + \alpha_0, \tag{5}$$

for $m = 1, \cdots, M$, and $f_2(m)$ is the video-level entropy of the reference video in the mth video set.

Video database

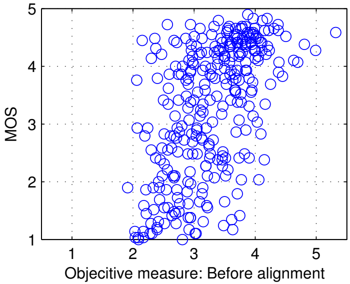
IRCCyN/IVC influence content video VGA database

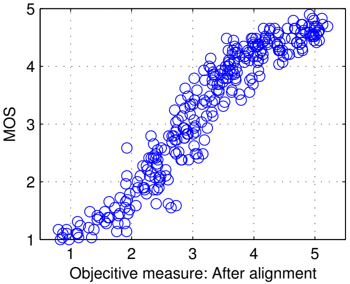
- 60 sets of videos ($M = 60$)
- 5 H.264 videos in each ($N = 5$)
- Resolution: 768×432
- The mean opinion score (MOS): $[0, 5]$
- H.264/SVC coding without transmission errors

Before and after supervised alignment

Universität Konstanz

Scatter plots

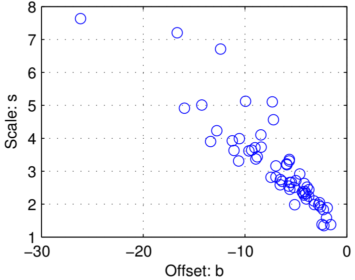


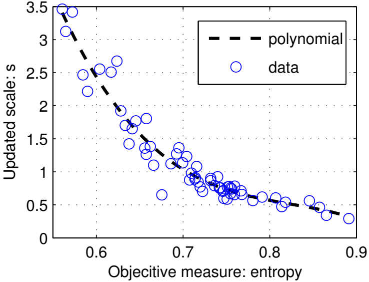


Parameter prediction

Universität Konstanz

Scatter plots

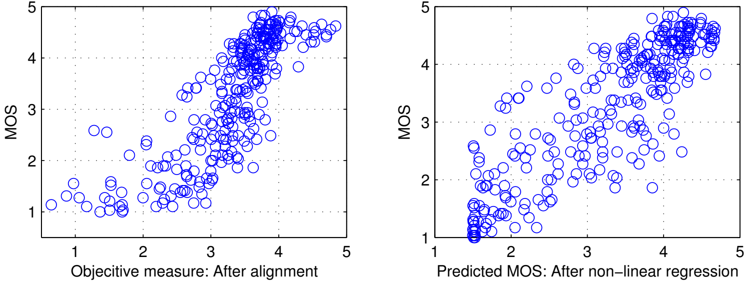




Before and after calibration

Universität Konstanz

Scatter plots



Performance Evaluation

Universität Konstanz

	LCC	SROCC	RMSE	MAE
Before alignment	0.3873	0.3698	1.8813	1.4919
After alignment	0.7996	0.8475	0.6985	0.5906
After regression	0.8554	0.8475	0.5894	0.4630
Supervised	0.9667	0.9581	0.2911	0.2196

- LCC: the Linear (Pearson's) Correlation Coefficient
- SROCC: the Spearman's Rank Ordered Correlation Coefficient
- RMSE: the Root Mean Squared Error
- MAE: the Mean Absolute Error

Summary

Universität Konstanz

Conclusions

- Studied the influence of video content on NR VQA
- Proposed a non-linear mapping strategy for NR VQA
- Improved the performance by designing local alignment
- Required small number of parameters
- Avoided overfitting in training-based methods

Limitation and Future work

- Only test in one video database with one type of distortion
- Develop more methods for local alignment according to video content

For more information:

Universität Konstanz

Read

K. Zhu, K. Hirakawa, V. K. Asari, and D. Saupe,“A no-reference video quality assessment based on Laplacian pyramids”, IEEE International Conference on Image Processing, Melbourne, Australia, Sep. 2013.

Contact

Kongfeng Zhu: kongfeng.zhu@uni-konstanz.de

Visit

http://www.informatik.uni-konstanz.de/saupe/

Appendix

Universität Konstanz

Backup Slides

Intra-subband features

Universität Konstanz

Let us view the subband coefficients in L_n as stationary random processes, where the random variable $X = L_n(i,j)$ have the same probability mass function $p(x)$.

1) Energy

$$E_n = \log_{10}\left(\Sigma_i \Sigma_j L_n^2(i,j)\right)$$

2) Entropy

$$H_n(X) = -\Sigma p(x) \log p(x)$$

3) Kurtosis

$$\kappa_n(x) = E(x-\mu_x)^4/\sigma_x^4$$

where

- x is the intensity, $p(x)$ is the probability mass function
- μ_x is the mean of x, and σ_x is the standard deviation

Inter-subband features

1) JSD (Jensen Shannon divergence)
A measure of the distance between two probability distributions

$$\mathrm{KLD}(p||q) = \sum p(x) \log(p(x)/q(x))$$
$$\mathrm{JSD}(p||q) = \frac{1}{2}(\mathrm{KLD}(p||r) + \mathrm{KLD}(q||r))$$

where

- $r(x) = (p(x) + q(x))/2$
- $p(x)$ and $q(x)$ as two probability mass functions of two images

Inter-subband features:

Universität Konstanz

The SSIM: Structural SiMilarity Index (Z. Wang, 2004)

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}$$

2) MSSIM (the mean structural similarity index)

- $\mathrm{MSSIM}(L_n, L_m)$: the mean of SSIM index map
- quantifies the dependency between L_n and L_m

3) Smoothness

a measure of the relative size of flat area

- S_{SSIM} is the SSIM index map between I and L_{N-1}
- the local region is flat when the SSIM index is $\geq T_0$
- T_0 is set to 0.95 in the experiment