Retina Image Quality

A Literature Review

2008-10-20

Jörg Meier

Chair of Pattern Recognition (CS 5)
Friedrich-Alexander-University Erlangen-Nuremberg
International Max Planck Research School
Optics and Imaging

MAX-PLANCK-FORSCHUNGSGRUPPE
Institut für Optik, Information und Photonik
Universität Erlangen-Nürnberg

Outline

- Motivation for retina image quality considerations
- Existing methods in literature
  - Automated image quality assessment in general
  - For retinal images
    - CFDM
    - ISC

Motivation for automated quality scores

- Objective measure for image processing algorithms
  - Compression
  - Transmission
- Measure quality of acquired image
  - If bad, acquire again
  - Provide only good images to next process (human/algorithm)
  - Improve acquisition protocol/technique
  - Objective score which images to include in a study
- Evaluation of classification algorithms
  - Correlate image quality to classification accuracy

Literature

- [1] H. R. Sheikh et al.:
  No-Reference Quality Assessment Using Natural Scene Statistics JPEG2000
  IEEE Transactions on Image Processing, 14/,12, December 2005, p. 1918-1927
- [2] G. Ginesu et al.:
  A multi-factors approach for image quality assessment based on a human visual system model
  J Signal Processing: Image Comm., 21/4, April 2006, p. 316-333
- [3] R. Janssen:
  Computational Image Quality
  Book, SPIE Society of Photo-Optical Instrumentation Engineers, July 2001
- [4] A. D. Flemming et al.:
  Automated Assessment of Diabetic Retinal Image Quality Based on Clarity and Field Definition
  J Invest. Ophthalmol. Vis. Sci., 2006, 47/3, p. 1120f
- [5] M. Niemeijer et al.:
  Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening
  J Medical Image Analysis, 2006, 10, p. 888f

Methods to measure image quality [1,2]

- Comparison between reference and processed image (Full-reference, FR)
- Only processed image (No-reference, NR)

- Objective
  - Full-reference (FR)
  - No-reference (NR, "blind measure")
- Subjective
  - Double stimulus Continuous Qualtiy Scale (DS-CQS)
  - Single Stimuls Continuous Qualtiy Scale (SS-CQS)

General methods [3]

- Mean squared error (MSE)
- Peak signal-to-noise ratio (PSNR)
- Visual Differences Predictor (VDM)
- Square-root integral (SQRI)
- Impairment measures
- Color measures

Quality assessment using NSS [1]

- Natural scene statistics (NSS) describe subset in image space showing natural scenes
- NR method, measures distortions in NSS
- Used to measure quality of JPEG2000 compressed images
- Uses NSS in wavelet domain
  - Model of E. P. Simoncelli 1997 and R.W. Buccigrossi 1999
  - Statistics of wavelet coefficients of natural images in given subband and their correlations with other coefficients across scales and orientations

Quality assessment using NSS [1] – method 1/2

- NSS model for wavelet coefficient magnitude C:

$$C = MP + N$$

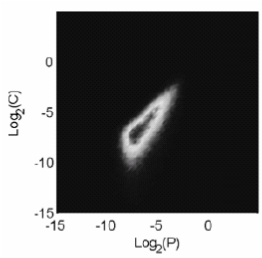
M, N: empirically found distribution M and Gaussian N

$$P = \sum_{i=1}^{n} l_i C_i$$

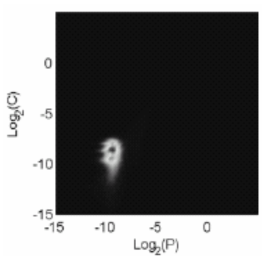
C_i: coefficients from coefficient neighborhood of C in space, scale, and orientation

l_i: linear prediction coefficients

- Joint histogram of P and C changes with quality



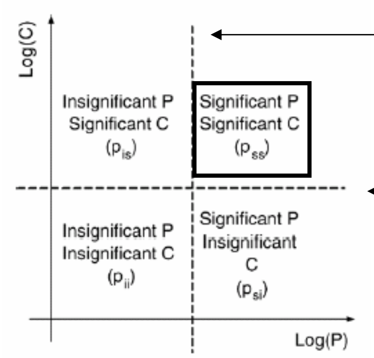
uncompressed image, diagonal subband at finest scale



compressed image (worse quality), certain subband and scale

Quality assessment using NSS [1] – method 2/2

- Simplified values for feature computation



thresholds for each subband is computed (learned in training phase; treshold is lower for smooth images and higher for textured images)

- p_{ss} most important feature
  - Computed in six subbands
  - Non-linear transformation “normalizes” the six p_{ss} to six q_{ss}
  - Final feature vector consists of 4 values (horiz., vertical subbands averaged)

Quality assessment using NSS [1] – evaluation

- LIVE database: http://live.ece.utexas.edu/research/quality
  29 images and their distorted versions (JPEG2000 compressed at different levels); 198 images total
  - Randomly split in test/training set (15/14 images)
  - Algorithm run on luminance component of images only
- 25 Observers assessed perceived quality in 5 categories: bad, poor, fair, good, excellent
  - Scores scaled to values from 1-100
  - Mean opinion score (MOS) computed for each image
    - Root mean-squared error (RMSE) was 7.04 (on scale 1-100)
    - Correlation coefficient 0.92
- Results: RMSE=8.54; correlation 0.91
- Compared to PSNR: RMSE=7.63; correlation 0.93

Using Human visual system (HVS) models [2]

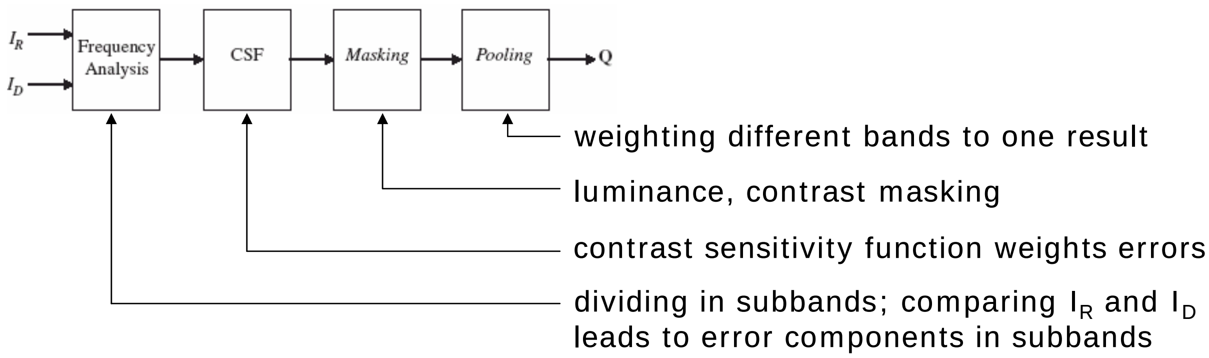
- HVS models: mathematical models try to approximate the human vision of the physical world
- Luminance, contrast masking and contrast sensitivity play important role in HVS models
  - Minimal perceived difference between a starting stimulus and a new one is proportional to the initial strengh of the stimulus

$$\Delta L = k \cdot L$$

  - Strong signal variation (contrast) masks/hides other image details
  - HVS is more sensible for middle frequencies (spatial frequ. of contrast change) → contrast sensitivity function (CSF), approximated by a heuristic function

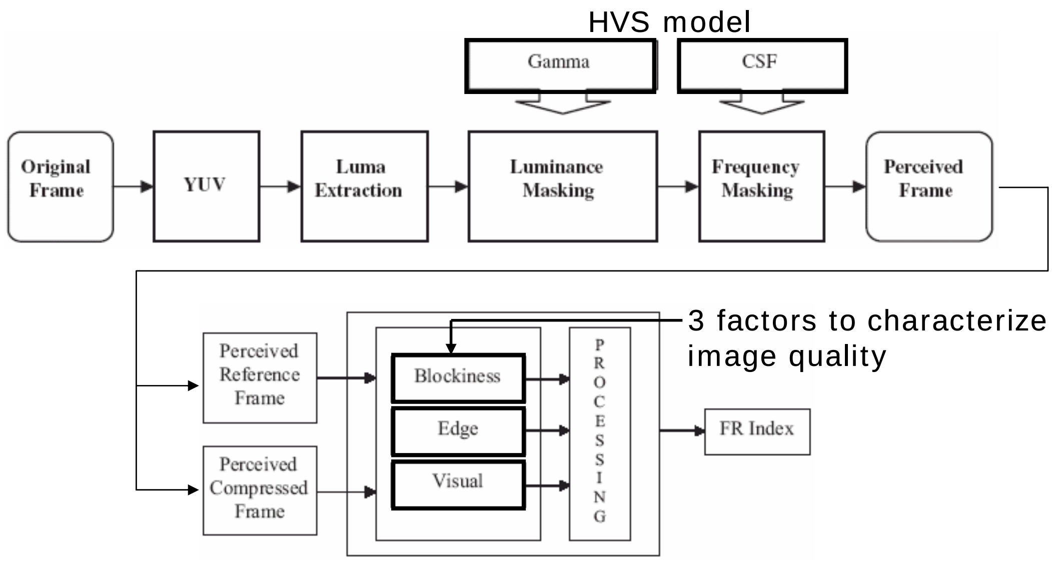
Using Human visual system (HVS) models [2]

- Generic scheme for FR metrics with HVS models



Multi-factor HVS model [2] – method overview

- In [2] the following method is proposed:



Multi-factor HVS model [2] - method

- F1: Blockiness
- F2: Edge
- F3: Visual impairment
- Combination of the 3 factors to final quality score Q
  - Linear combination, scaled to 1-5 (weights found by linear regression in training step):

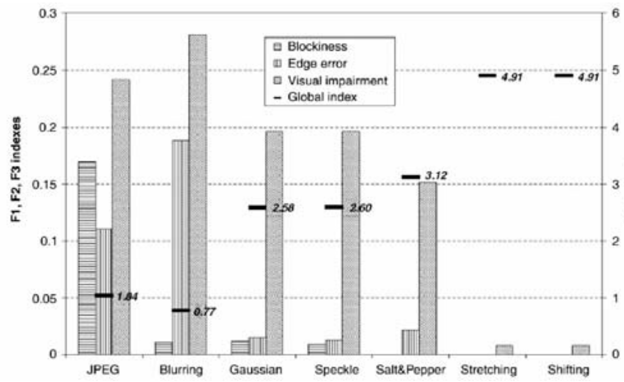
$$Q = 5 \cdot \left[1 - k \cdot \frac{w_1 F_1 + w_2 F_2 + w_3 F_3}{w_1 + w_2 + w_3}\right]$$

Multi-factor HVS model [2] - evaluation

- LIVE database used (JPEG and JPEG2000 images)
  - 29 input images, distorted by JPEG compression; blurring; gaussian, speckle, and salt&pepper noise; streching; shifting
  - 168 images for training, 176 for test
- Method compared with
  - PSNR
  - SSIM (Structural SIMilarity [Wang, Bovik 2004])
  - VQM (ITU-T J.144 standard, 1998, video quality for digital cable transmission)
- Mean opinion score (MOS from 1-100) and Difference-MOS (DMOS) used from the manual assessments
- Criteria and used metric
  - Prediction accuracy: RMSE as difference between DMOS and $DMOS_{predicted}$
  - Monotonicity: Pearson's and Spearman's linear correlation index R_P and R_S
  - Consistency: outlier ratio (OR): ration between outlier and normal points

Multi-factor HVS model [2] - results

- Distortion factors and final metric for an example image



- Result after non-linear correction of score to subjective quality assessment

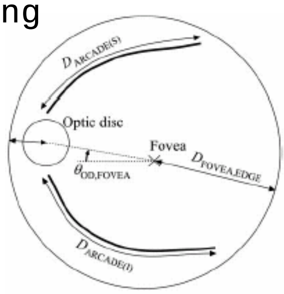
Models	Complete verification data set			
	Pearson	Spearman	RMSE*	OR
FRI	0.9331	0.9268	0.2340	0.0398
PSNR	0.8417	0.8395	0.3513	0.1364
SSIM	0.9371	0.9289	0.2270	0.0227
VQM	0.8369	0.8342	0.3565	0.1379

Retina Image Quality

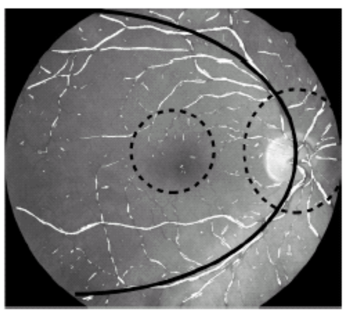
- Not very much publications in this special field
  - Two interesting papers are summarzied in the following
    - [4] A. D. Flemming et al.: CFDM method, IOVS, March 2006
    - [5] M. Niemeijer et al.: ISC method, MIA, Sept. 2006
- Additionally referenced works
  - S. Lee et al.: SPIE conference paper, 1999
    - Global image intensity histogram analysis; one mean histogram of some high quality images as reference
  - M. Lalonde et al.: conference paper, 2001
    - Global edge histogram combined with local intensity histogr.; one mean histogram of some high quality images as reference
  - J. Lowell et al.: conference paper, 2005
    - Similar idea as in [4]; segmenation of vessel tree

Clarity and field definition metric (CFDM) [4]

- Clarity
  - Image shows sufficient detail for automated grading
  - Four categories: excellent, good, fair, inadequate with definitions what each category means
- Field
  - Image shows the desired field of view
  - Three categories: excellent, good, inadequate

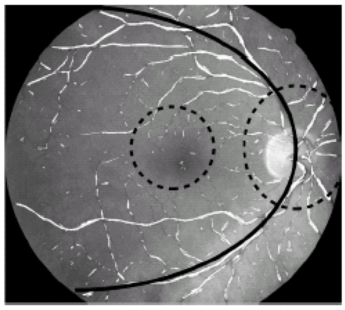


- Segmentation-based technique
  - Analysis of the vessels around the macula
  - Presence of small vessels there is indicator of high image quality



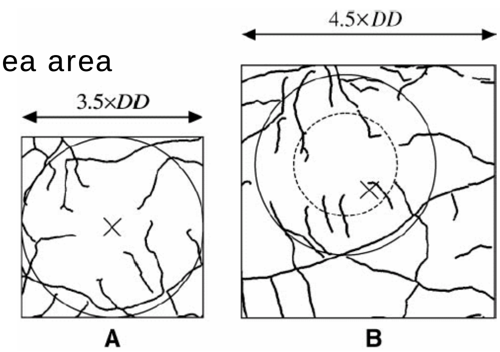
CFDM [4] – segmentation 1/2

- Find temporal arcades
  - Detect large-scale vessels with width between 10-30 pixels (details described in a conference paper 2004)
  - Generalized Hough transform for semielliptical shapes (on subsampled image, factor 32)
- Find optic disc
  - Disc diameter (DD) empirically set to 246 pixels
  - Search space restricted: box (2.4DDx2.0DD) around rightmost point of arcade
  - Detect circular outline of disc by Hough transform (search from 0.7DD to 1.25DD)
- Find fovea
  - Maximize correlation coefficient between image and fovea model in circular region (1.6DD diameter, 2.4 DD from disc)

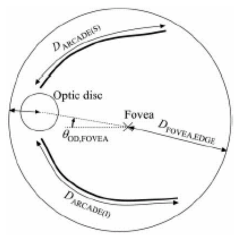


CFDM [4] – segmentation 2/2

- Clarity
  - Measure total length of micro-vessels in fovea area
  - Two alternative measures
    - A: Fovea well detected (high correlation to model)
    - B: Fovea not well detected refinement of circular search region
  - Threshold for total vessel length for *clarity=ok* derived in training



- Field
  - Certain thresholds for measured distances set
    - Empirically
    - and by looking at the ROC (optimizing sensit./select.)



CFDM [4] – evaluation, results

- 395 training, 1039 test images
  - Canon CR5-45NM and CR6-45NM devices with a D30 camera
  - 1600x1440 image size, disc-macula imaging protocol
  - Green channel of images used
- Results of finding the bad images

	Data Set	Sensitivity (CI)	Specificity (CI)
Image clarity	Training	100% = 50/50 (92.9%–100%)	89.9% = 310/345 (86.2%–92.6%)
	Test	100% = 57/57 (93.7%–100%)	90.9% = 893/982 (89.0%–92.6%)
Field definition	Training	91.7% = 33/36 (78.2%–97.1%)	97.5% = 350/359 (95.3%–98.7%)
	Test	95.3% = 82/86 (88.6%–98.2%)	96.4% = 919/953 (95.1%–97.4%)
Overall quality	Training	98.7% = 73/74 (92.8%–99.8%)	90.0% = 288/320 (86.2%–92.8%)
	Test	99.1% = 116/117 (95.3%–99.8%)	89.4% = 824/922 (87.2%–91.2%)

ISC [5] - overview

- Image structure clustering
  - Not segmentation-based, but uses
  - Image filter for looking at different scales and edge directions
  - Unsupervised clustering of feature vector
- Color information is also used
  - ISC performed only on green channel, but
  - Histograms of the color channels finally included in feature vector

ISC [5] - method

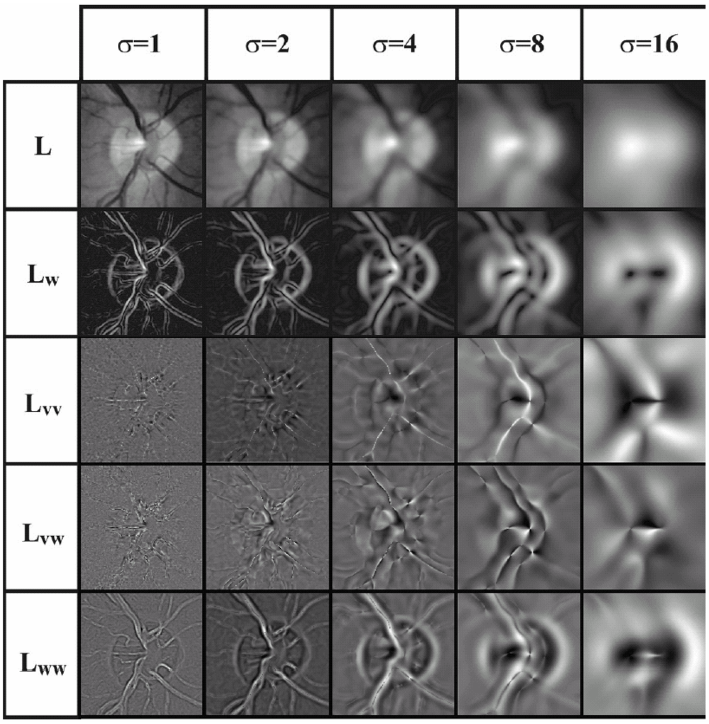
- Filter bank applied to image
  - Gauge coordinates: local coordinate system $(\vec{v}, \vec{w})$ in each point of an image *L:*

$$\vec{w} = (\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}) \qquad \vec{v} \perp \vec{w}$$

  - Filter set: Gaussian derivative applied at five different scales σ on L, L_w, L_{vv}, L_{vw}, L_{ww} (subscripts indicate certain direction)
- Feature vectors selected from filter responses (25 dim per pixel)
  - Scaled to zero mean, unit variance
- Clustered by *k-means* clustering
  - Clustering should find similar image structures
  - Each pixel is assigned to one of the *k* clusters
- Histogram of cluster image taken as feature for quality classification

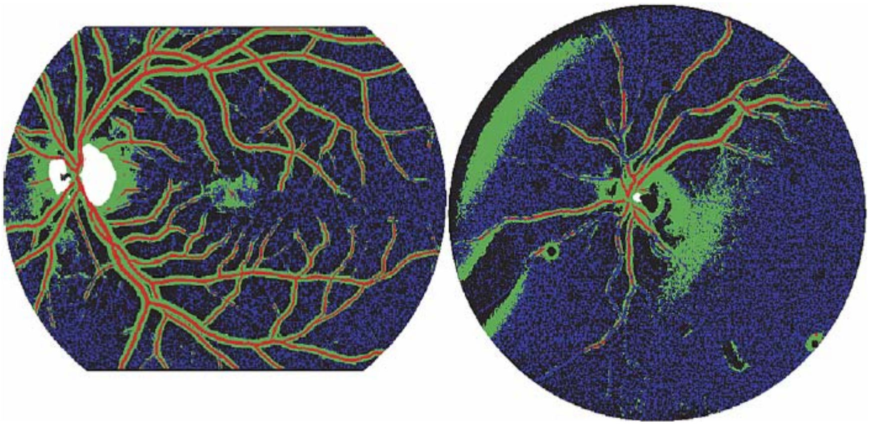
ISC [5] – filter response

- Example



ISC [5] – clustered images

- Cluster number of k=5 experimentally chosen



black: background

blue: background, bright to dark transitions

green: borders of high contrast

red: vessels

white: optic disc

ISC [5] – features

- Features finally used (20 dim.)
  - 5 bins of image structured clusters
  - 5 bins of each color channel histogram
    - Red
    - Green
    - Blue
- Feature selection was applied
  - On a subset of the training set (balanced split)
  - AROC used as criterion (called A_z)
- Different classifiers evaluated, SVM was best
  - Non-linear SVM (parameters optimized with cross-validation on training set by grid search)
  - For SVM no feature selection applied!

ISC [5] – evaluation

- 2000 images from DR screening (from 20 centers)
  - Image sizes from 768x576 (35° FOV) to 2048x1536 (45° FOV)
  - JPEG compressed
  - Cameras: Topcon NW 100, NW 200, Canon CR5-45NM
  - Images resampled for "equal FOV", 530 pixel diameter
  - 3 readers (ophthalmologists)
  - "low quality": reader unable to judge absence/presence of DR
    - otherwise "normal" quality
  - Training and test set (1000/1000), 10% contained pathologies
  - Additional grading of test set by one doctor: four categories

ISC [5] – results

- Classification results

Classifier	A_z	95% CI	Acc.
ISC + histogram			
SVM $c = 16384$, $\sigma = 9.77 \times 10^{-4}$	**0.9968(0.0013)**	**(0.9934, 0.9985)**	**0.974**
QDC*	0.9944(0.0014)	(0.9909, 0.9967)	0.963
LDC	0.9901(0.0021)	(0.9851, 0.9936)	0.951
kNN* $k = 15$	0.9932(0.0019)	(0.9885, 0.9961)	0.958

- System outperformed second observer
  - But almost equal to each other (AROC values)

Lessons learned

- Automated retina image quality assessment for screening applications is a non-reference classification task
- Important part in processing pipeline
- Non-segmentation-based techniques are promising

ISC [4] – filter responses (A4 kowa images)