Perceptual Video Quality Measurement for Streaming Video over Mobile Networks

Senthil Shanmugham

Master's Thesis Defense

27th June 2006

Committee:

Dr. John Gauch (Chair)

Dr. Arvin Agah

Dr. Joe Evans

Acknowledgements

- Dr. John Gauch
- Dr. Arvin Agah and Dr. Joe Evans
- Dr. Jim Black and Dr. Claudio Lima (Sprint ATL)

Presentation Outline

- Introduction
- Background
- Digital Video Quality
- KUIM Video Pipeline
- Implementation Details
- Performance Evaluation
- Conclusion
- Future Work

Introduction

- The Internet will be an important source of video distribution
- Best-effort video delivery without any Quality of Service (QoS) guarantees
- Network bandwidth, packet losses and frame jitter are the main factors effecting video quality

Introduction

- Subjective quality
  - User perceived quality of the video
  - Time consuming and expensive
- Objective quality
  - Produce results comparable to subjective methods
  - Easy, real-time and done without user intervention
- Pixel-based metrics
  - Mean-Squared Error (MSE)
  - Peak Signal-to-Noise Ratio (PSNR)

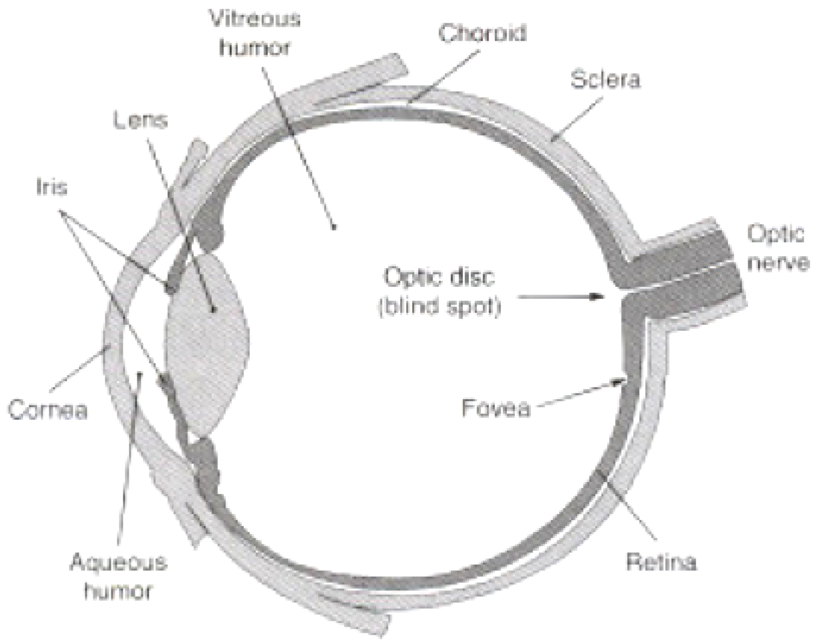
Thesis Goals

- Develop an effective method for measuring perceptual visual quality of mobile streaming video
- Generate consistent quality scores for video sequences comparable to subjective measurements
- Models should be based on the properties of Human Visual System (HVS)
- Data will be generated using SprintPCS EVDO-Rev 0 mobile network
- Results will be compared with the Mean Opinion Score (MOS) generated by NetQual setup at ATL

Background

- Visual perception is the most important of all senses - 80 to 90% of all the neurons in the brain
- HVS can be divided two main parts: eyes and visual pathways
- Retina – information is preprocessed before sent to the brain
- Complexity
  - Considerable optical differences between individuals
  - Component of the eye undergoes constant changes throughout life

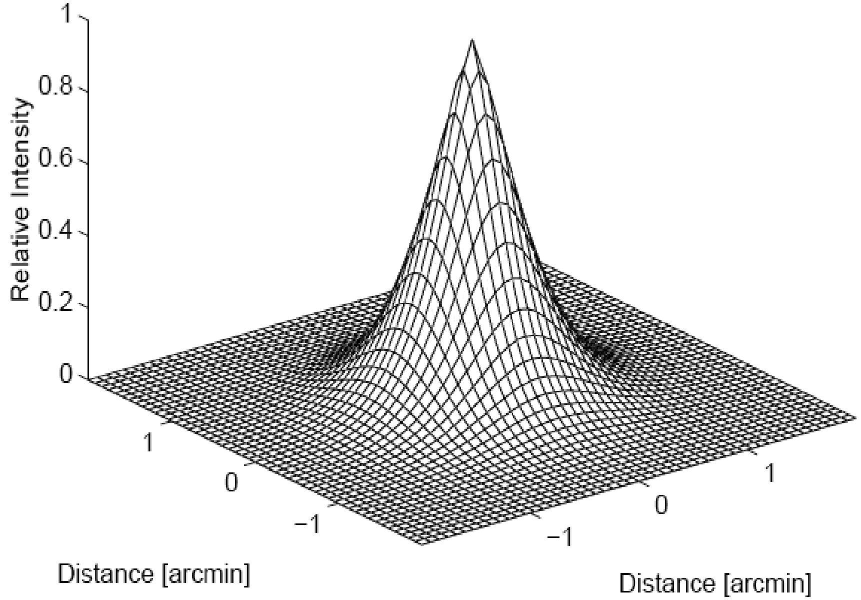
Background



The human eye (transverse section of the left eye) (Winkler, 2004)

Background

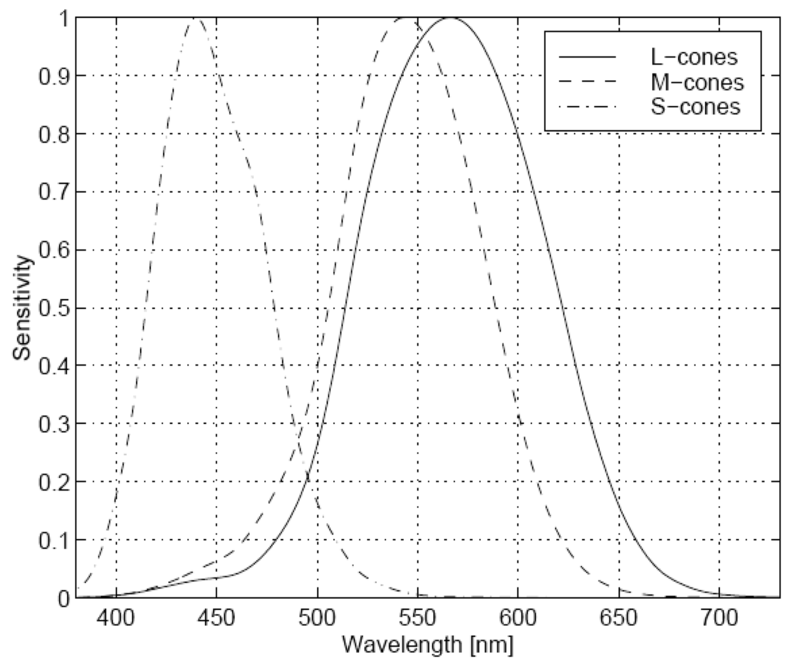
- Quality of the eye – reflection of the visual stimulus on the retina
- The image is the distorted version of the input and the most important one is blurring
- Point spread or Line spread function - to identify the amount of blurring



Point spread function of the human eye as a function of visual angle (Westheimer, 1986)

Background

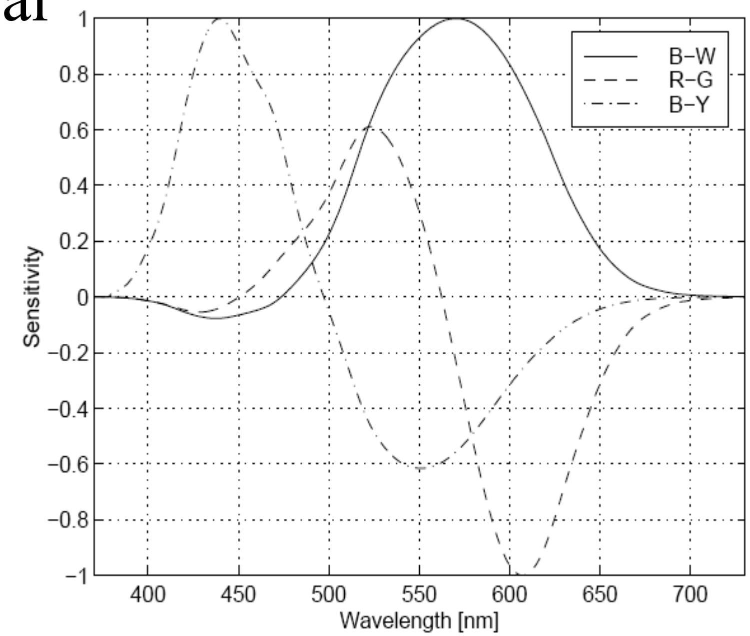
- Photoreceptor mosaic at the back of the retina
- Responsible for sampling the image and converting into information
- Two types of photoreceptors – rods and cones
- L- , M- and S-cones



Normalized absorption spectra of three cones (Stockman and Sharp, 2000)

Background

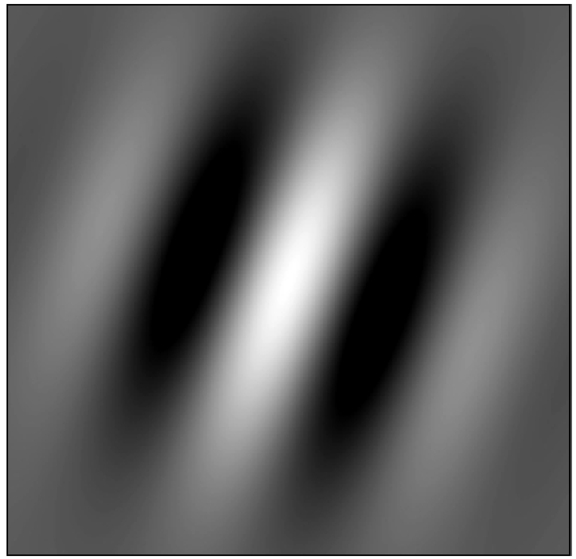
- Light is defined by spectral power distribution
- Trichromacy of human color vision
- Reddish yellow is perceived as orange where as we cannot perceive reddish green
- Opponent color theory
- The principle components are (White-Black) W-B, (Red-Green) R-G and (Blue-Yellow) B-Y



Normalized spectral densities of three opponent colors (Poirson and wandell, 1993)

Background

- Receptive fields of primary visual cortex
- Light and dark shades denotes excitatory and inhibitory regions, respectively
- These characteristics of human visual system are used in the design quality models and metrics



Idealized receptive field of primary visual cortex (Winkler, 2004)

Digital Video Quality

- The main goal is reduce bandwidth and storage requirements without compromising quality
- Compression and Transmission of digital video results visual artifacts
- Compression artifacts are blocking, ringing, blurring and mosquito noise
- MPEG, H.263, RealMedia and Windows Media

Digital Video Quality

- Compressed video is transferred over packet-switched network
- Wire or wireless channel at physical layer and TCP/UDP at transport layer
- Header contains sequencing, timing and signaling information
- Streaming video needs additional protocols like RTP and RTSP

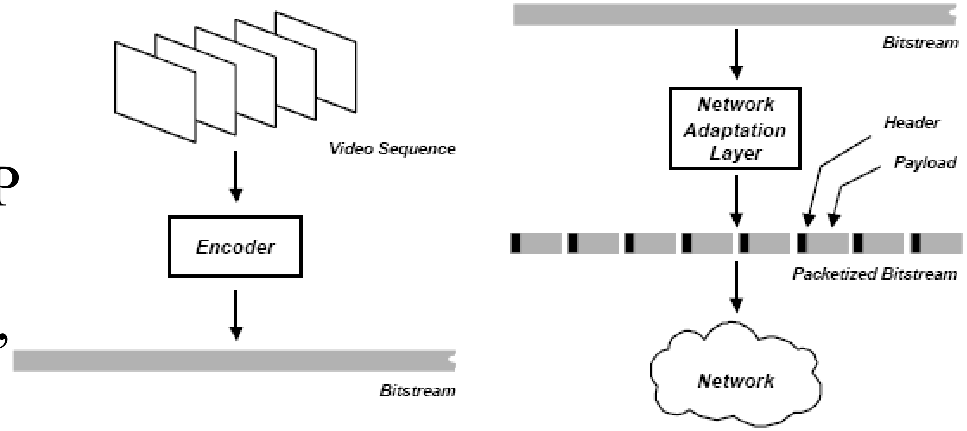
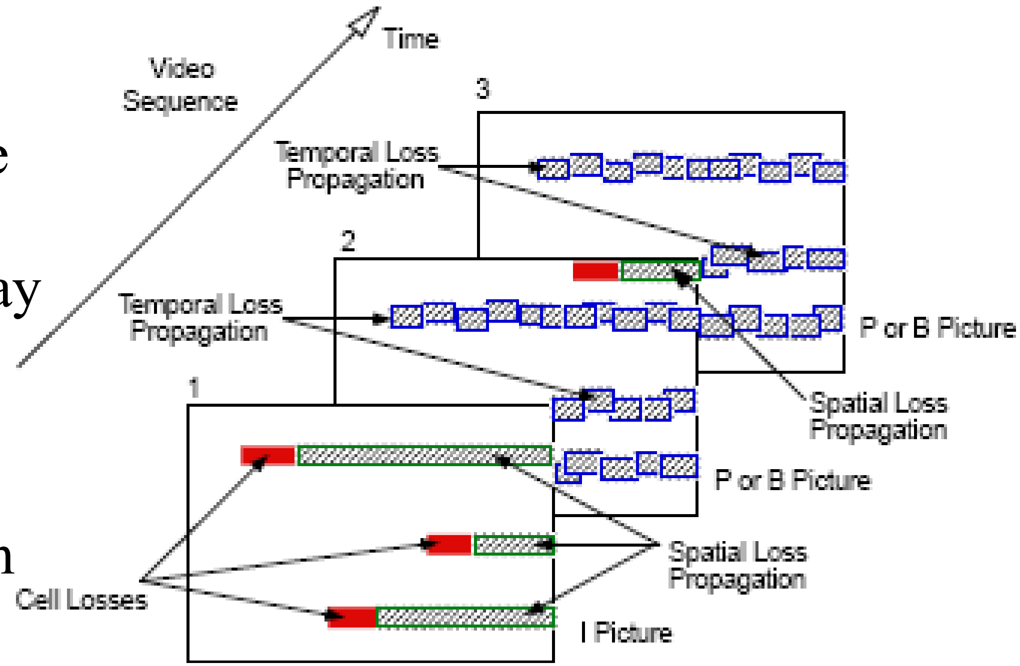


Illustration of video transmission system (Winkler, 2001)

Digital Video Quality

- Packets may be delayed or lost during transmission
- Quality of the video may be depend upon the lost frame
- MPEG macroblock loss may result in temporal loss propagation until synchronized
- Visual effects depends upon the ability of the decoder to identify and conceal errors



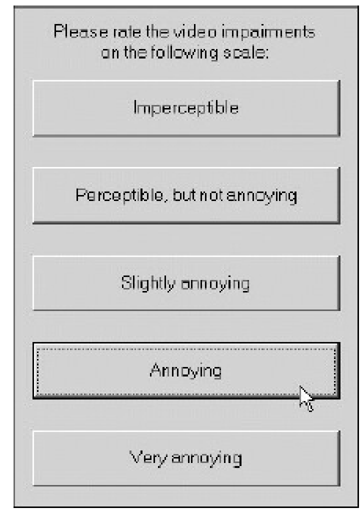
Spatial and temporal loss propagation in a MPEG-compressed video (Winkler, 2001)

Subjective Video Quality

- Recommendation ITU-R BT.500-10 “Methodology for the subjective assessment of the quality of television pictures” has been used for many years
- Subjects are asked to rate the test sequence based upon the reference sequence on a continuous quality scale.
- Double Stimulus Continuous Quality-Scale Method (DSCQS) and Single Stimulus Continuous Quality Evaluation (SSCQE)

Subjective Video Quality

Typical subjective video quality assessment laboratory



Subjective quality assessment metrics corresponding to quality score from 1 to 5

Objective Video Quality

- More reproducible and portable but should have good correlation with subjective scores
- Full Reference Method – The reference and distorted videos are compared to arrive at a quality score
- Reduced Reference Method – Features from the reference and distorted video are compared to arrive at a quality score
- No Reference Method – No reference frame is needed and the quality score is based on the distorted video only

Objective Video Quality

(a) Original (b) PSNR = 32 dB (c) PSNR = 32 dB

$$PSNR = 10\log\left(\frac{255^2}{MSE \text{ between frames}}\right)$$

The same amount of noise after inserting to original image (a) at two different parts of the image. (Winkler, 2004)

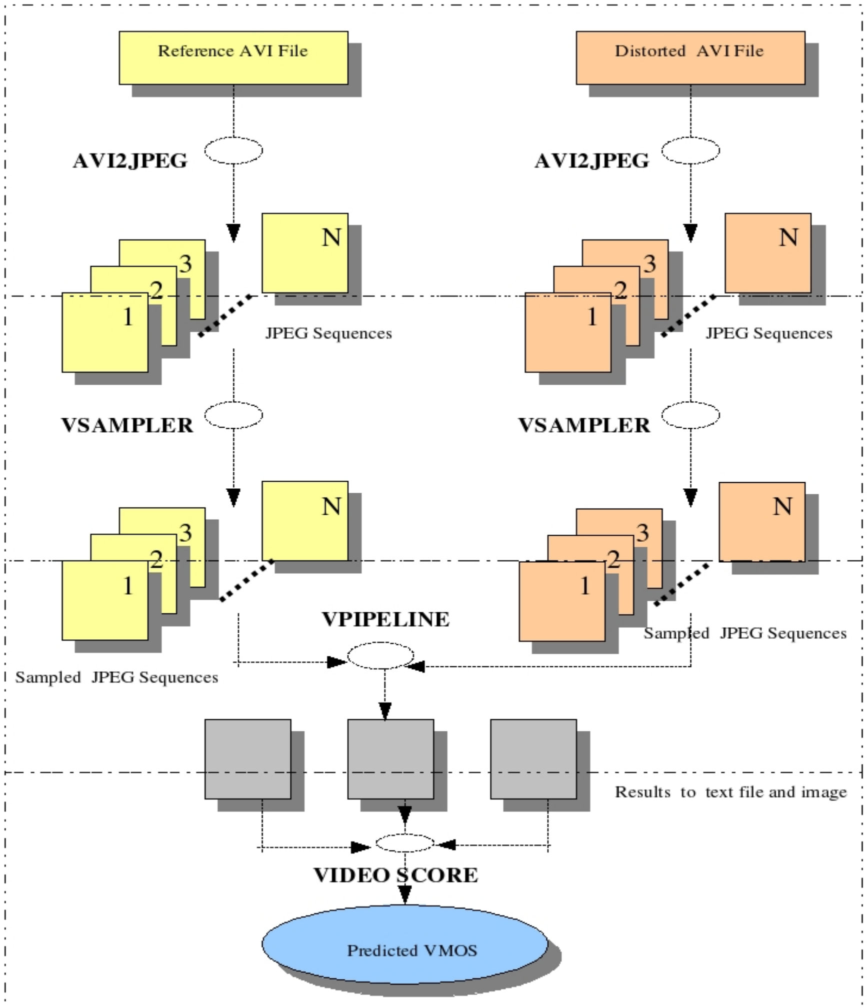
KUIM Video Quality Pipeline

- We have implemented an objective video quality system by extending existing KUIM tools
- Simulates the visual pathways of the HVS
- Color perception, spatio-temporal contrast sensitivity and multi-channel representation of the HVS
- Full reference method – requires both reference and distorted videos

KUIM Video Quality Pipeline

- AVI2JPEG – conversion of the original AVI video into a sequence of JPEG frames
- Vsampler – Temporal sampling is done to remove duplicate frames and to recover from frame loss
- Vpipeline – takes the two videos as input and calculates the distortion measure
- Vscore – Based on the distortion measure comes up with the Predicted Video Mean Opinion Score (VMOS)

KUIM Video Quality Pipeline

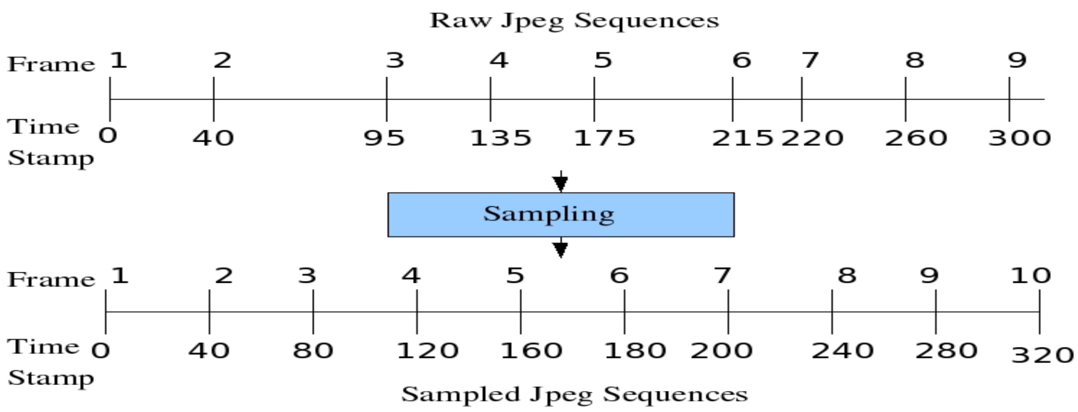


Implementation Details - AVI2Jpeg

- Conversion of the original AVI video into a sequence of JPEG frames
- It skips the initial block of header and extracts the uncompressed video frames
- Initial blue frames are the synchronization frames
- The blue frames were discarded and the comparison was done only for the video content
- Extracted video frames are then converted to Jpeg images using KUIM JPEG Library

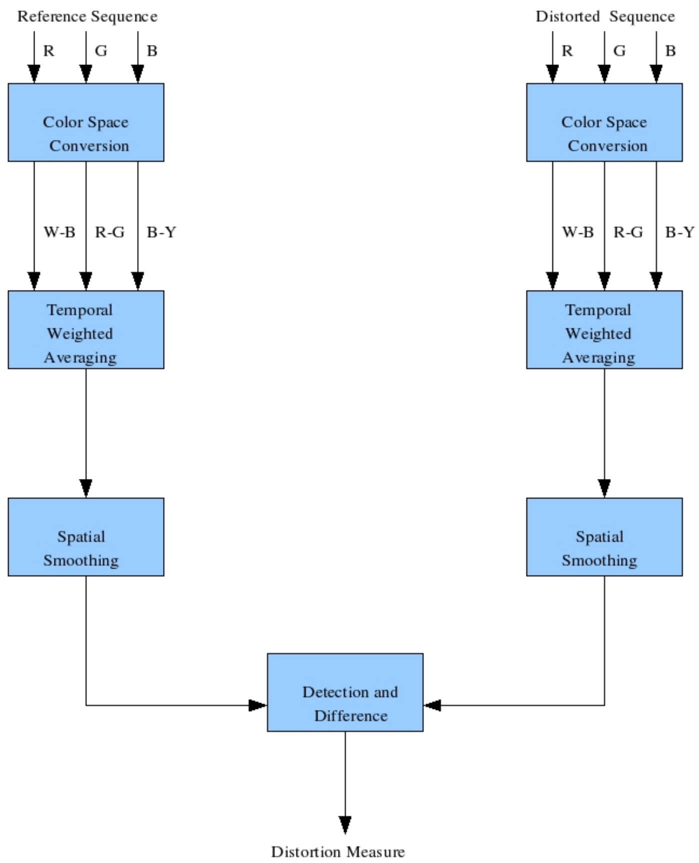
Implementation Details - Vsampler

- Temporal sampling is done to remove duplicate frames and insert new frames as needed
- Important step in a full-reference method where we do frame-by-frame comparison
- The frame was sampled using nearest neighbourhood at 40us for a video transmitted at 25fps for 6 seconds



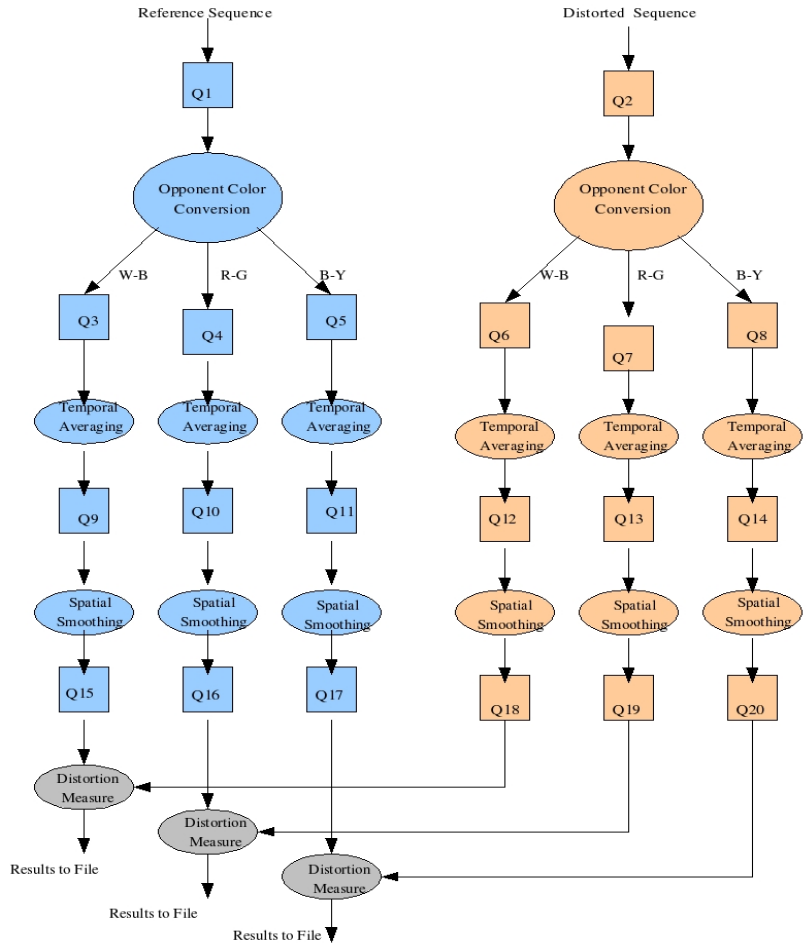
Implementation Details - Vpipeline

- Convert the images in RGB color space to opponent color space W-B, R-G and B-Y
- Perform temporal weighted averaging (window size = 5)
- Perform binomial spatial smoothing
- Calculate and compute the distortion measure



Implementation Details - Vpipeline

- KUIM_QUEUE
- KUIM_PIPELINE
- KUIM_COLOR
- Temporal Averaging
- Opponent Color Conversion
- Distortion Differences
- Queue Status
- Display and Store the results



Implementation Details - Vscore

- Information from the various channels within the primary visual cortex is integrated in the subsequent brain areas
- Same process was done for our models by gathering data from all the channels and coming up with the distortion measure
- The quality score was calculated after analyzing the distortion measure
- This quality score was compared against the SwissQual's VMOS for performance evaluation

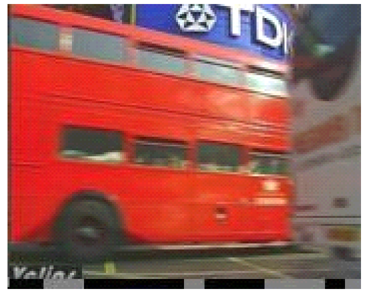
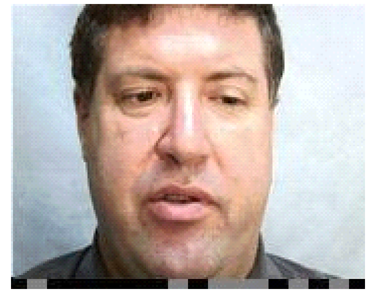
Implementation Details - Vscore

$$Q = ((a / Average) + (b / Max) + (c / Top10)) / 3$$

where 'Average' is average value of the all the pixel differences
'Max' is the maximum value
'Top10' is the average of the top ten largest pixel differences
a, b and c are KUIM quality constants

VIDEO SEQUENCES	**MOTION CONTENT**	**a**	**b**	**c**
Woman(CW)	Low	19.26	64.88	29.65
Traffic (PC)	High	54.02	162.8	78.81
Man (CA)	Low	17.73	61.60	22.70

Performance Evaluation

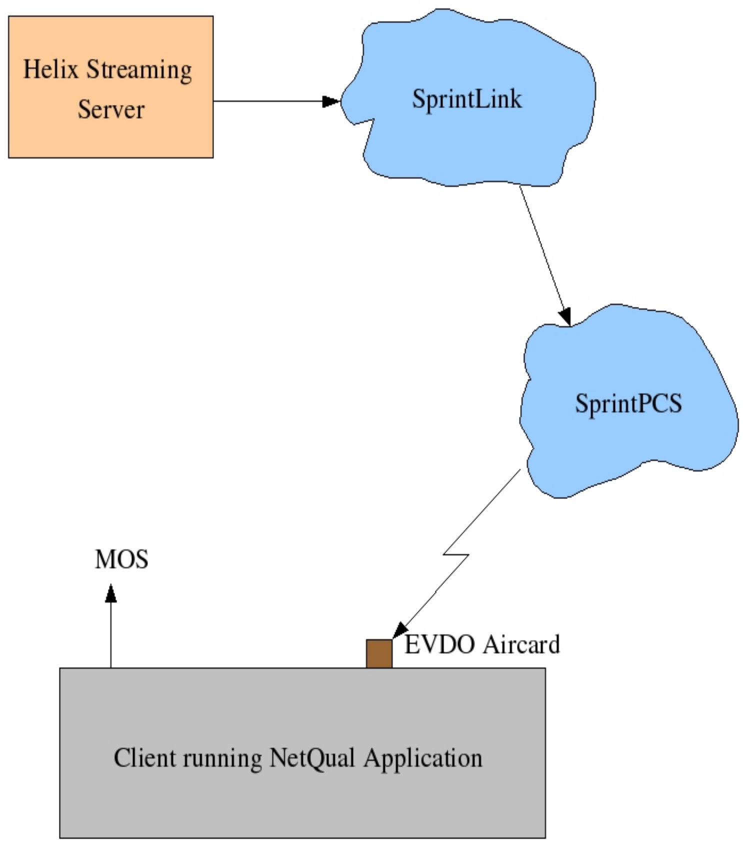




- 3 types of videos in QCIF format at 25 fps
- Two low motion content – woman drinking water outside a cafe and a man talking
- One high motion content – auto traffic outside Piccadilly Circus

MOS	USER EXPERIENCE
5	Imperceptible / Excellent
4	Perceptible / Good
3	Slightly annoying / Fair
2	Annoying / Poor
1	Very annoying / Bad

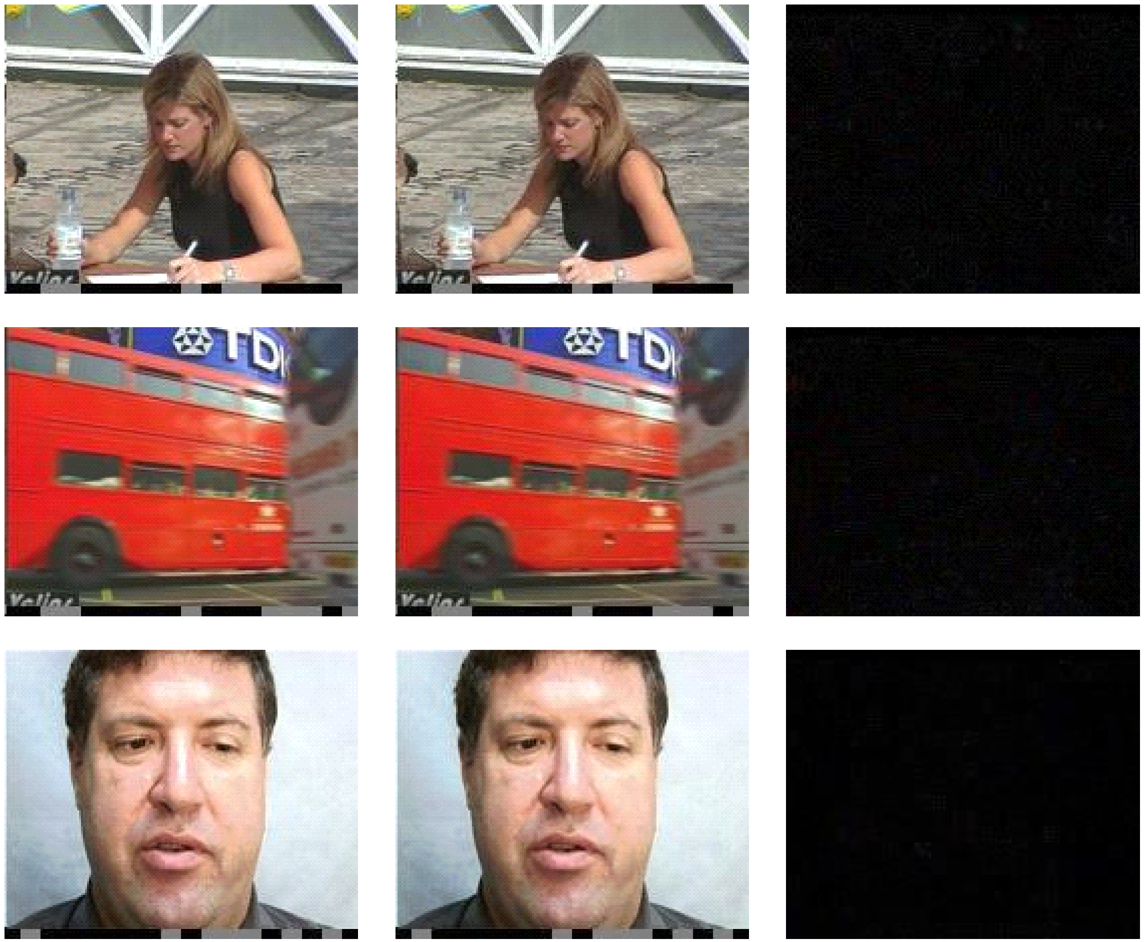
Performance Evaluation

- SwissQual's NetQual setup at Sprint ATL
- Helix Multi-media server, client running NetQual application test set and EVDO Samsung A600 PCS Vision phone
- MPEG-4, H.263 and MPEG-2 transport streams



Network Set-Up for Data Generation for Test Sequences

Reference, Distorted and Pixel Differences for Woman, Car and Man test sequences in RGB Color Space



Woman

Car

Man

Reference | Distorted | RGB Differences

W-B, R-G and B-Y components of the test sequences after opponent color conversion for Woman, Car and Man test sequences, respectively

Woman

Car

Man

W-B R-G B-Y

W-B, R-G and B-Y components of the test sequences after temporal weighted averaging for Woman, Car and Man test sequences, respectively

W-B, R-G and B-Y components of the test sequences after binomial spatial smoothing for Woman, Car and Man test sequences, respectively

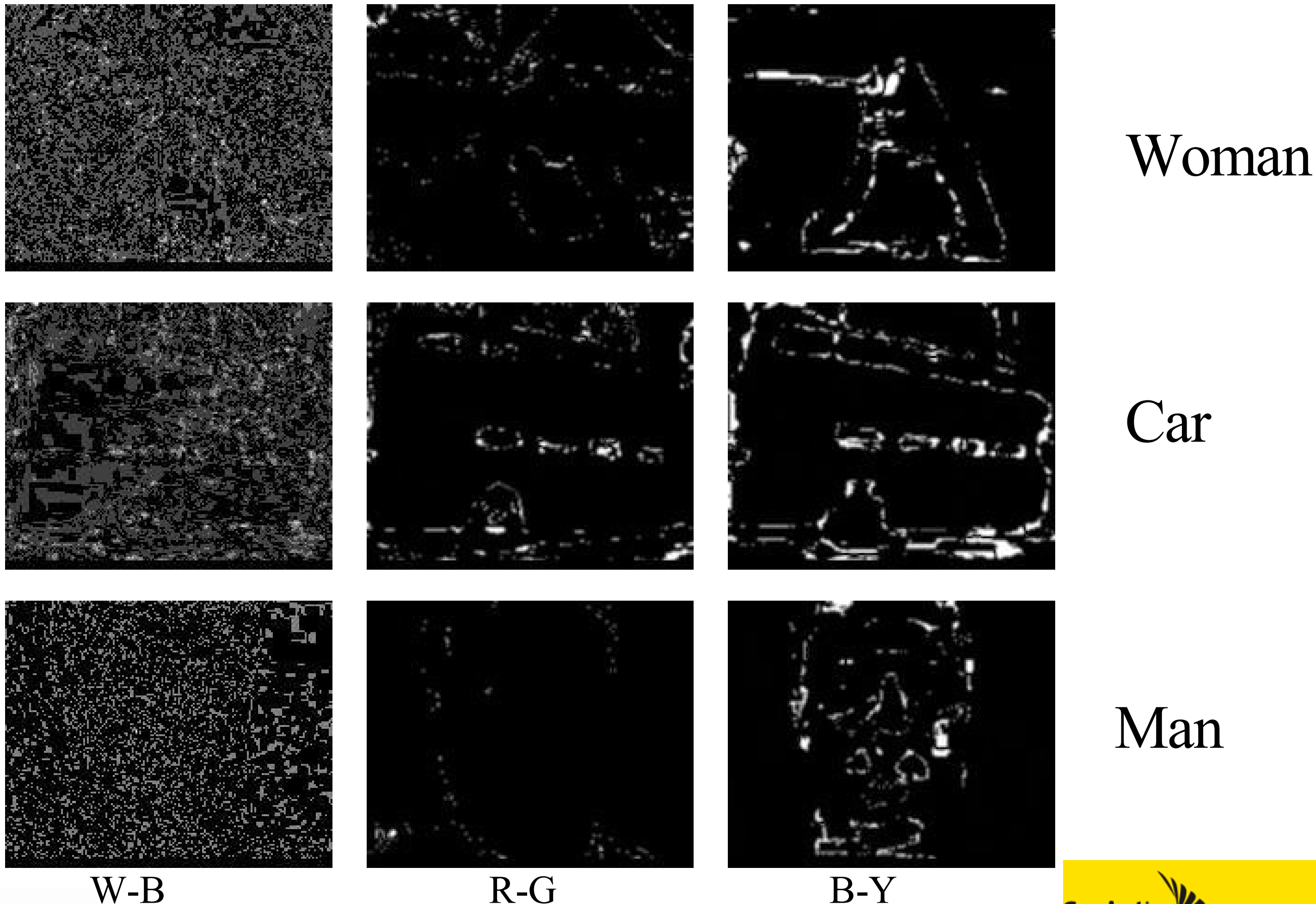
Woman

Car

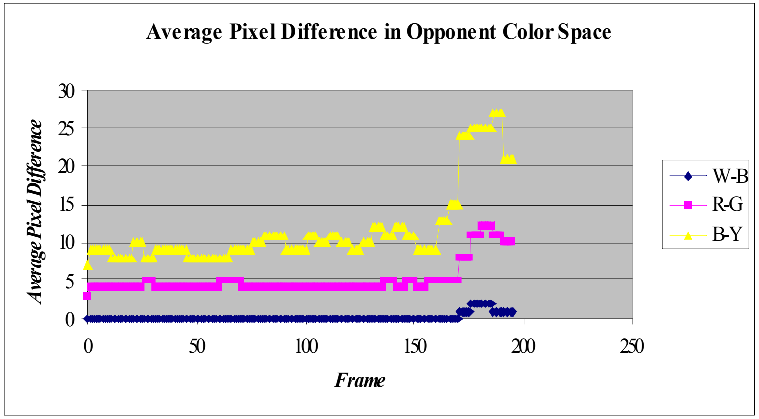
Man

W-B R-G B-Y

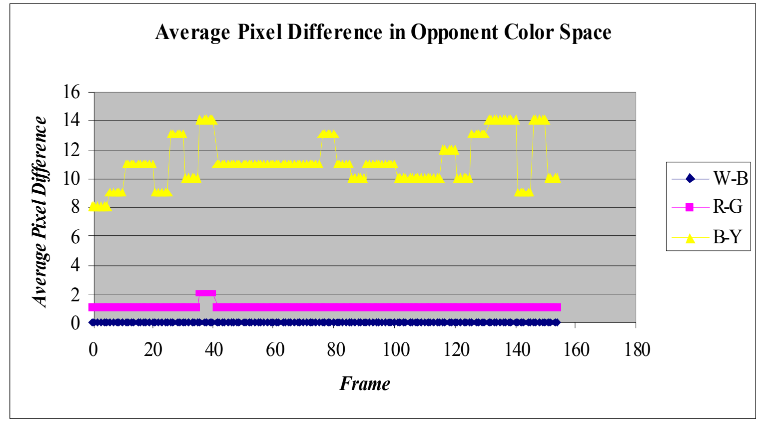
Frame difference between the reference and distorted sequences after processing through KUIM perceptual software pipeline



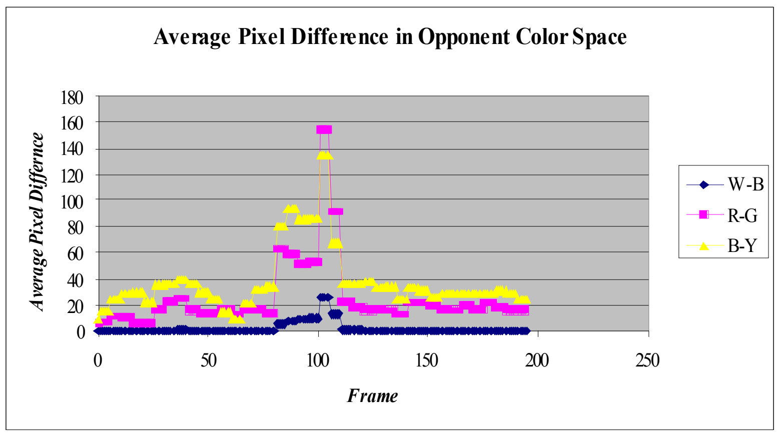
Average pixel difference between the reference and distorted sequence



Woman

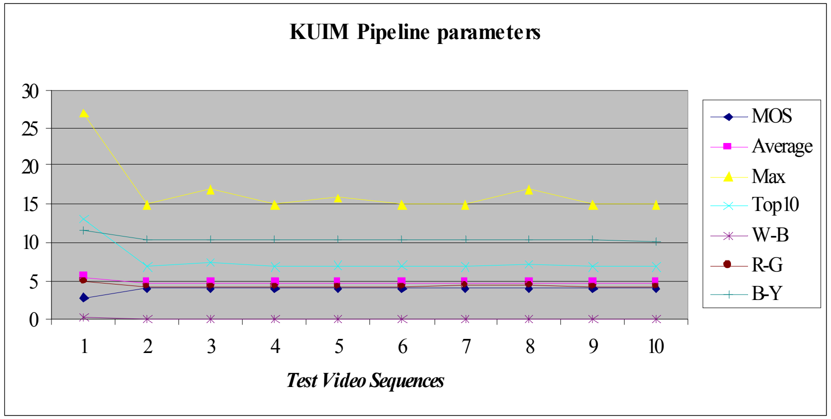


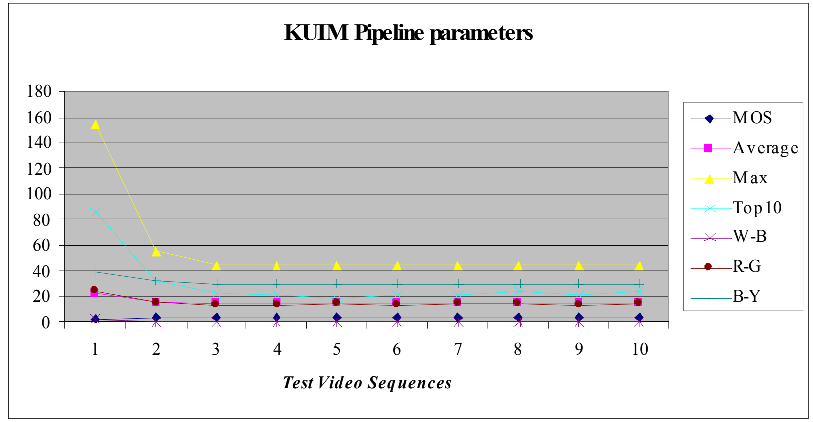
Car



Man

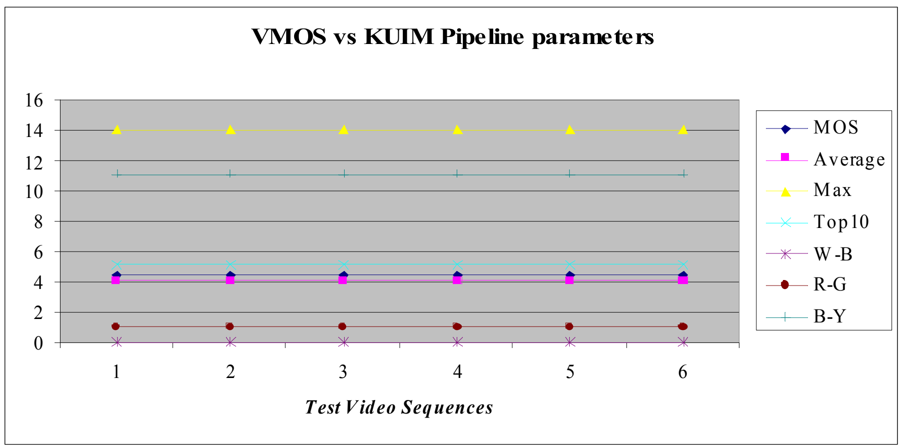
KUIM Pipeline parameters





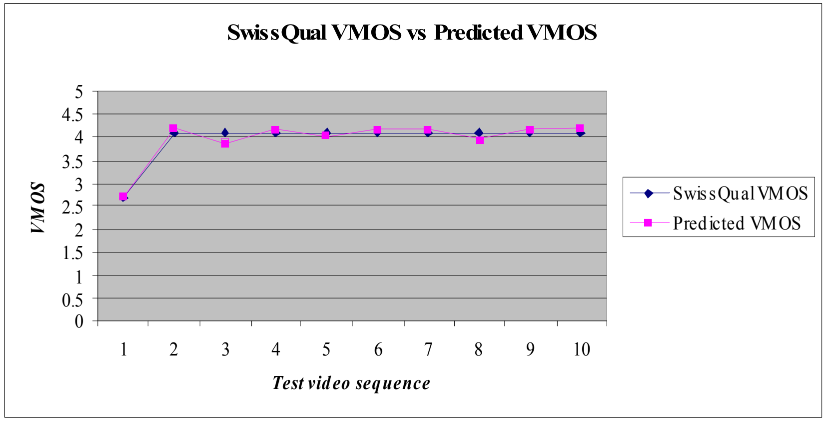
Woman

Man

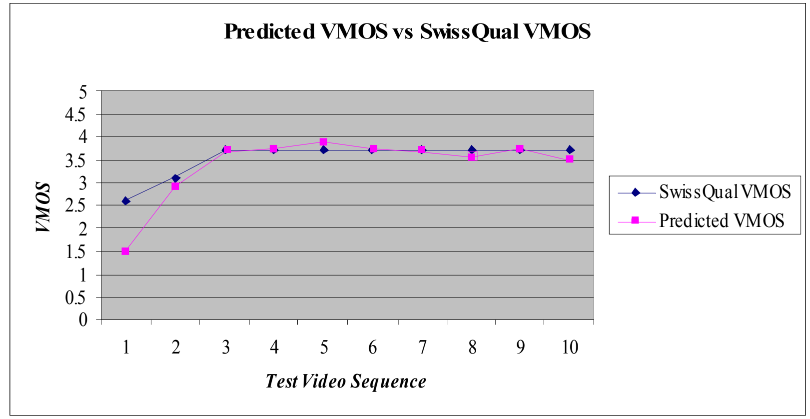


Car

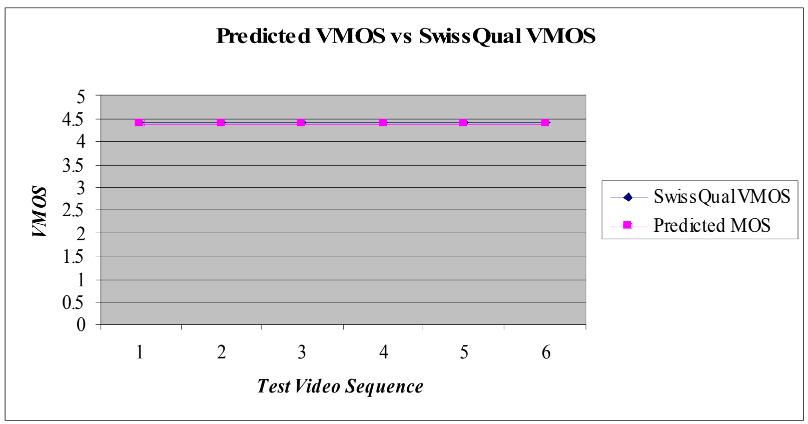
Predicted VMOS vs SwissQual VMOS



Woman



Car



Man

Performance Evaluation

VIDEO SEQUENCES	MOTION CONTENT	SEQUENCE NAME	SWISSQUAL VMOS	PREDICTE D VMOS
Woman	Low	CW_2.7_45_005005	2.7	2.70
Woman	Low	CW_4.1_45_001005	4.1	4.20
Woman	Low	CW_4.1_45_002005	4.1	3.92
Woman	Low	CW_4.1_45_003005	4.1	4.17
Woman	Low	CW_4.1_45_004005	4.1	4.02
Woman	Low	CW_4.1_45_006005	4.1	4.16
Woman	Low	CW_4.1_45_007005	4.1	4.17
Woman	Low	CW_4.1_45_008005	4.1	3.91
Woman	Low	CW_4.1_45_009005	4.1	4.17
Woman	Low	CW_4.1_45_010005	4.1	4.21

Woman - SwissQual VMOS vs Predicted VMOS

Performance Evaluation

VIDEO SEQUENCES	MOTION CONTENT	SEQUENCE NAME	SWISSQUAL VMOS	PREDICTED VMOS
Car	High	PC_2.6_45_009008	2.6	1.48
Car	High	PC_3.1_45_004008	3.1	2.91
Car	High	PC_3.7_45_001008	3.7	3.66
Car	High	PC_3.7_45_002008	3.7	3.72
Car	High	PC_3.7_45_003008	3.7	3.90
Car	High	PC_3.7_45_005008	3.7	3.72
Car	High	PC_3.7_45_006008	3.7	3.67
Car	High	PC_3.7_45_007008	3.7	3.55
Car	High	PC_3.7_45_008008	3.7	3.72
Car	High	PC_3.7_45_010008	3.7	3.49

Car - SwissQual VMOS vs Predicted VMOS

Performance Evaluation

VIDEO SEQUENCES	MOTION CONTENT	SEQUENCE NAME	SWISSQUAL VMOS	PREDICTE D VMOS
Man	Low	CA__4.4_45_001009	4.4	4.39
Man	Low	CA__4.4_45_002009	4.4	4.39
Man	Low	CA__4.4_45_003009	4.4	4.39
Man	Low	CA__4.4_45_004009	4.4	4.39
Man	Low	CA__4.4_45_005009	4.4	4.39
Man	Low	CA__4.4_45_006009	4.4	4.39

Man - SwissQual VMOS vs Predicted VMOS

Overall Accomplishments

- KUIM Video quality pipeline
- AVI2JPEG
- Temporal sampling of the distorted video
- Predicted VMOS in good correlation with the SwissQual VMOS
- Extensive collection of papers on Digital Video Quality – Video library
- Data generated using full reference as well as no-reference frames at ATL with SwissQual VMOS

Conclusion

- Video quality assessment and optimizing user experience based on errors in video capture, storage, transmission and display
- Models based on HVS
- Based on constraints like the quality of the displayed video and user's viewing conditions
- Methods to measure perceptual video quality that predict human perception of video quality

Future Work

- Visual quality assessment without any reference frames
- Estimate the video quality in real-time and without any user intervention
- Reduced reference model of estimating video quality
- Quality metrics for both audio and video
- Automatic selection of a, b and c weights based on video content (high/low motion)

Thank You!