

Analysis of the Effect of Packet Loss and Delay Jitter on MPEG-4 Video Transmission

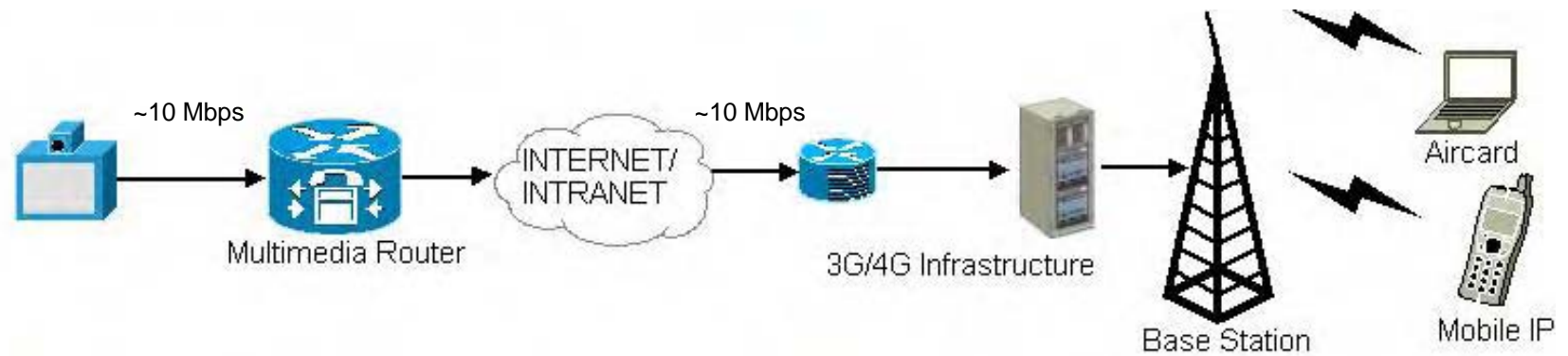
Muhammad Saleem Koul
EE Dept., UT Arlington



Synopsis

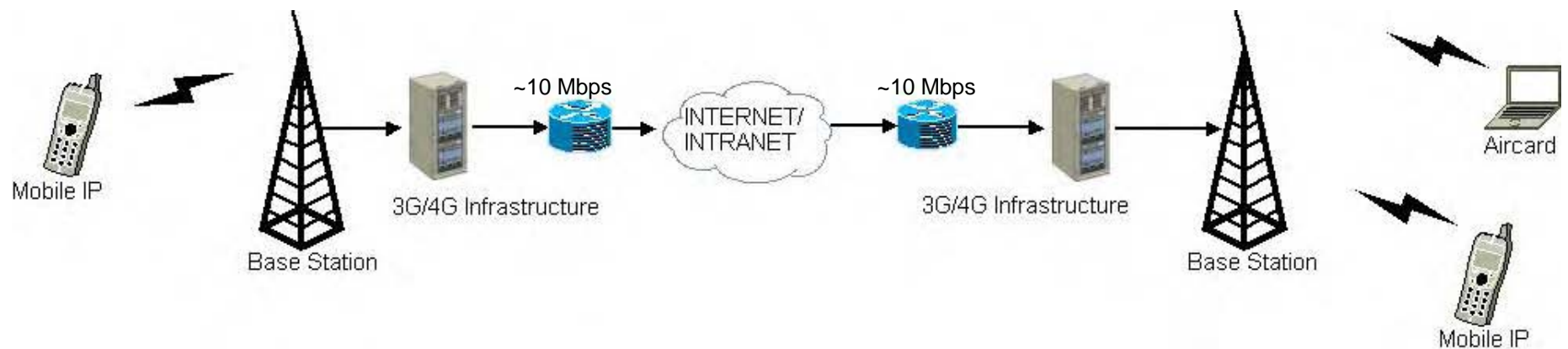
- The advent of Broadband 3G/4G Cellular Technologies bring the boon of Multimedia over Wireless.
- Inherent nature of the transmission medium makes the problems of Packet Loss and Delay variance (jitter) are more severe in Wireless/Cellular networks.
- A Video Quality Assessment (VQA) methodology is introduced, that helps analyze and quantify these effects on the Quality of the received Video.
- Several distinct applications are introduced, that can be accomplished using this methodology.

Typical 3G/4G Network Applications

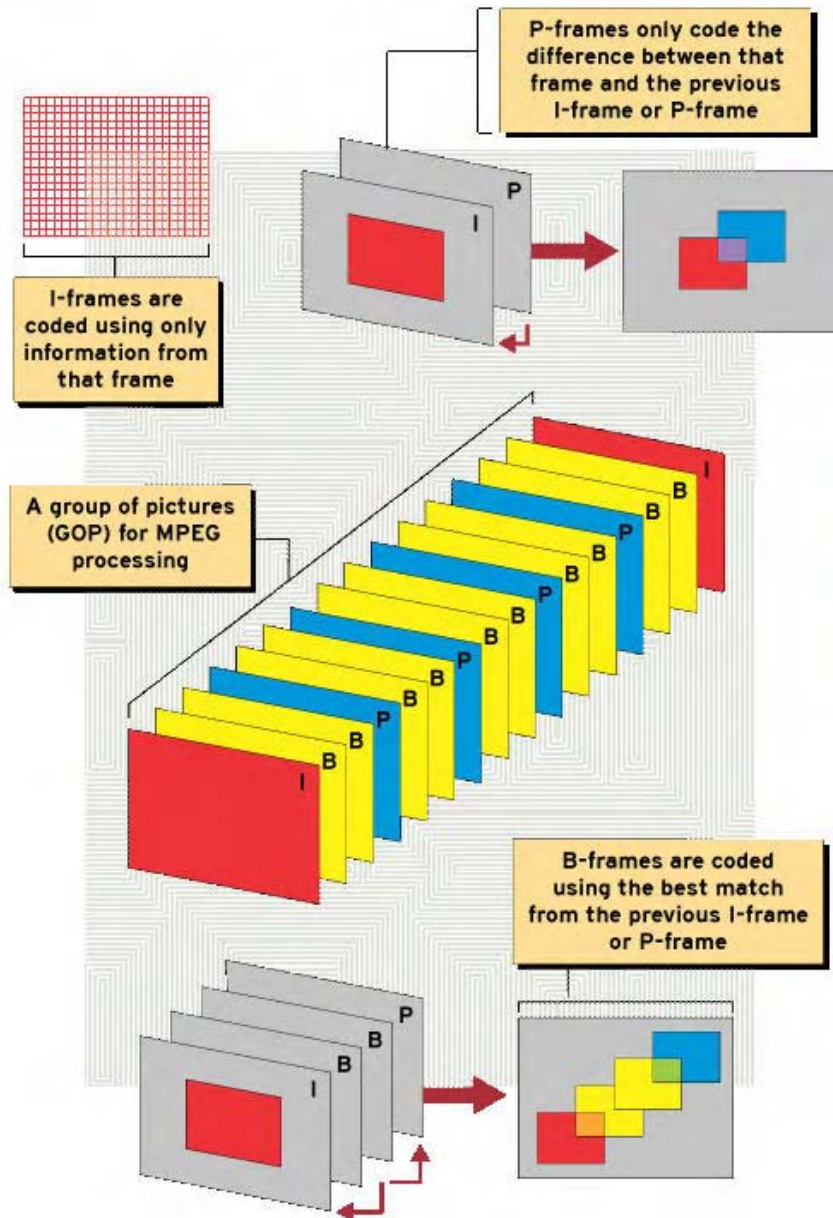


CDMA 2000 1X: 144 kbps
 EVDO Rev A: 500-700kbps
 WiMAX: >3.1Mbps

CDMA 2000 1X: 144 kbps
 EVDO Rev A: 500-700kbps
 WiMAX: >3.1Mbps



Typical MPEG-4 Stream Setup courtesy: streamingmedia.org



Determining Packet Loss and Delay Jitter

courtesy: [Klaue et al.](#)

$$\text{packet loss } PL_T = 100 \frac{n_{T_{recv}}}{n_{T_{sent}}}, \quad \text{where:}$$

T : Type of data in packet (one of all, header, I, P, B, S)

$n_{T_{sent}}$: number of type T packets sent

$n_{T_{recv}}$: number of type T packets received

$$\text{frame loss } FL_T = 100 \frac{n_{T_{recv}}}{n_{T_{sent}}}, \quad \text{where:}$$

T : Type of frame (one of all, header, I, P, B, S)

$n_{T_{sent}}$: number of type T frames sent

$n_{T_{recv}}$: number of type T frames received

inter-packet time $it_{P_0} = 0$

$$it_{P_n} = t_{P_n} - t_{P_{n-1}}$$

where: t_{P_n} : time-stamp of packet number n

inter-frame time $it_{F_0} = 0$

$$it_{F_m} = t_{F_m} - t_{F_{m-1}}$$

where: t_{F_m} : time-stamp of last segment of frame number m

$$\text{packet jitter } j_P = \frac{1}{N} \sum_{i=1}^N (it_i - it_N)^2$$

N : number of packets

it_N : average of inter-packet times

$$\text{frame jitter } j_F = \frac{1}{M} \sum_{i=1}^M (it_i - it_M)^2$$

where:

M : number of frames

it_M : average of inter-frame times



Quality Analysis of Video Transmission over wireless networks

Two main approaches to VQ analysis in
High speed networks [3]

- Video Bit stream based Full Reference VQ Analysis
- Video Traffic Trace based Full Reference VQ Analysis



Video Trace Generation

- Commonly used Video format is YUV. The sequences are several hundred frames long.
- Y=Luminance, and Chrominance (U=Hue, V=Saturation).
- In all our tests we used the luminance 'Y' only, as the HVS is most sensitive to this component.
- A typical [Trace file](#) contains the Frame number, Cumulative Display time, Frame Type [I, P, B], Frame size. The Second (optional) section of the trace contains the PSNR of the Y, U and V components.
- Several softwares are available to generate trace files from encoded MPEG-4 files. A simple trace can be a tcpdump or Ethernet packet port trace at the sender and/or receiver side.



(Objective) Video Quality Analysis of the Received Sequences

- In this section we detail several Video Quality Analysis schemes applied to the transmitted and received video sequences and analyze their results in light of the VQEG final report 'No one objective model outperforms the other in all cases'.
- We compare results from RMSE, PSNR, DVQ and SSIM.
- The VQA methods need to be mapped to a Human Perceivable Score (MOS). The importance of MOS mapping is discussed

Root-Mean-Square Error (RMSE)

- It calculates the “difference” between two images. It can be applied to digital video by averaging the results for each frame.
- For an $M \times N$ image, RMSE can be calculated as:

$$E_{RMS} = \sqrt{\frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [f'(m, n) - f(m, n)]^2}$$



PSNR (Peak Signal to Noise Ratio)

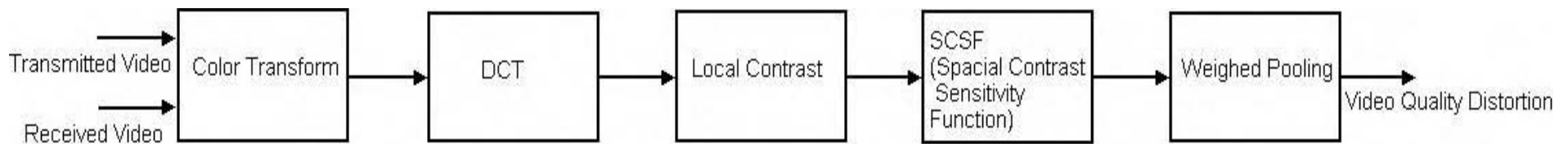
- The most commonly used objective quality metric is the Peak Signal to Noise Ratio (PSNR). For a video sequence of frames. The PSNR (dB) of each frame having N*M pixels can be calculated as:

$$PSNR = 20 \times \log_{10} \left(\frac{255}{RMSE} \right)$$

Where 255 is the Maximum pixel value in the N*M pixel image

DCT based VQ Evaluation

- The conventional video metrics (RMSE and PSNR) do not take into account the spatial and temporal property of human visual perception.
- This DCT based VQ Metric proposed Xiao's [8] is based on Watson's [7].
- The 8x8 based block based distortion is the 'atom' of all current compression based video processing. This kind of block based distortions are very eminent hallmarks in all decoded video sequences. Hence a metric that does the Evaluation in the DCT domain on 8x8 blocks yields significant results:





Structural Similarity Approach

- This approach emphasizes that the Human Visual System (HVS) is highly adapted to extract structural information from visual scenes. Therefore, a measurement of structural similarity (or difference) should provide a good approximation to perceptual image quality.
- The SSIM index is defined as a product of luminance, contrast and structural comparison functions. [8, 9]
- Where μ is the mean intensity, and σ is the standard deviation as a round estimate of the signal contrast. C_1 and C_2 are constants. M is the numbers of samples in the quality map.

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma$$

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

$$\text{MSSIM}(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{j=1}^M \text{SSIM}(\mathbf{x}_j, \mathbf{y}_j)$$



MOS Mapping

■ In Multimedia, the **Mean Opinion Score (MOS)** provides a numerical indication of the perceived quality of received media after compression and/or transmission. The MOS is expressed as a single number in the range 1 to 5, where 1 is lowest perceived quality, and 5 is the highest perceived quality.

Mean Opinion Score (MOS)

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying



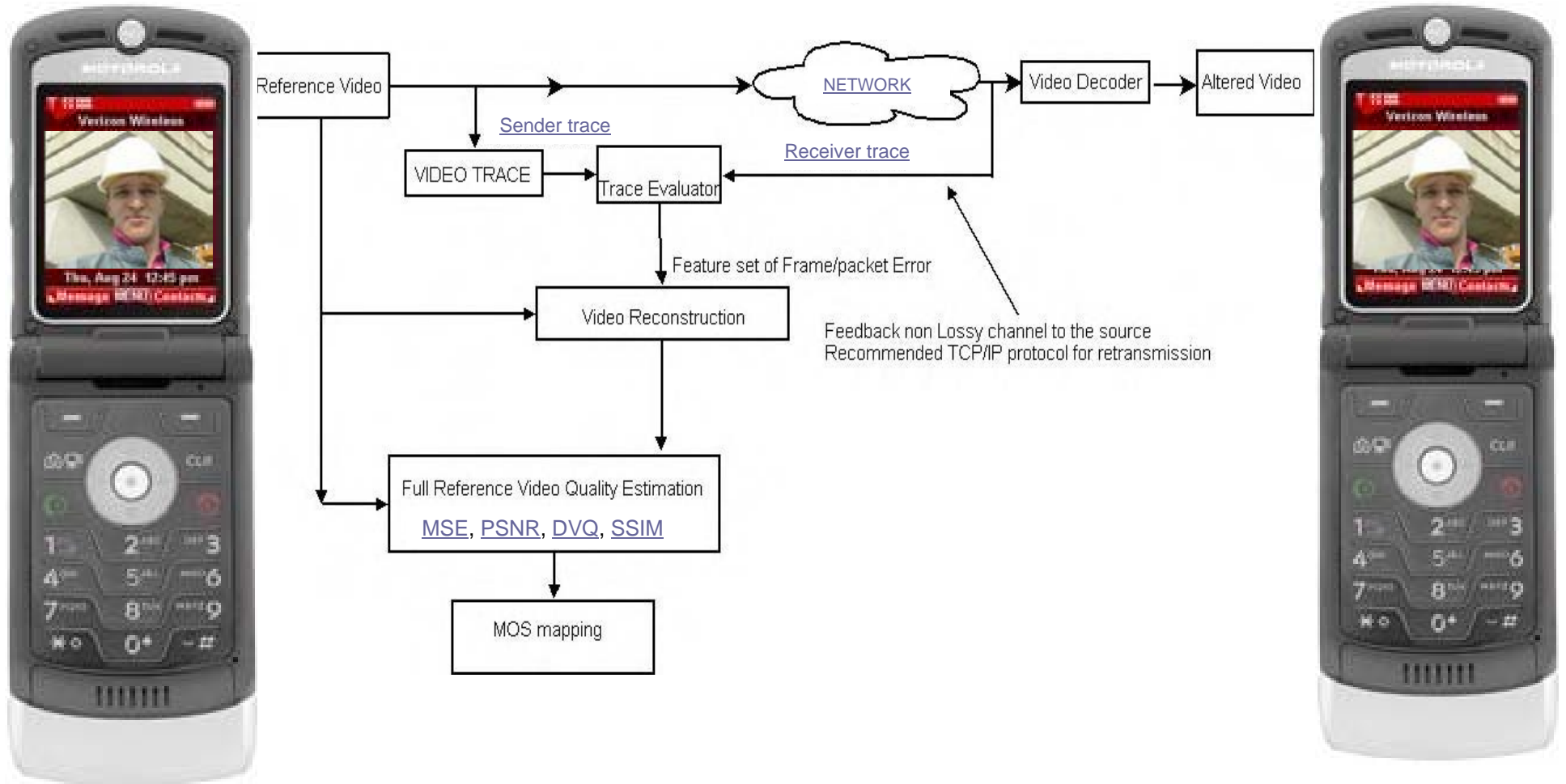
MOS Mapping contd.

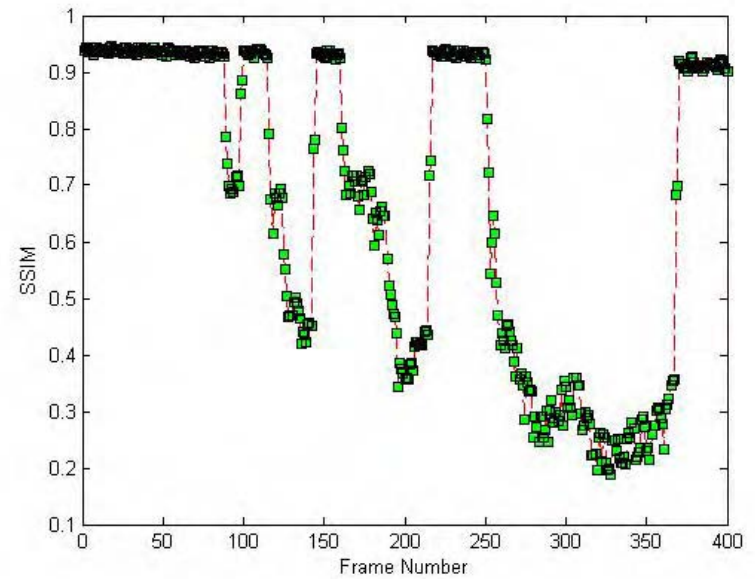
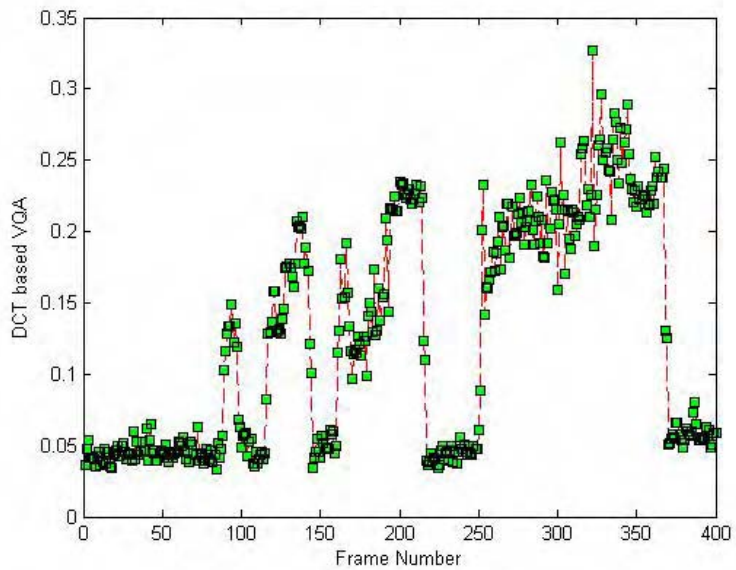
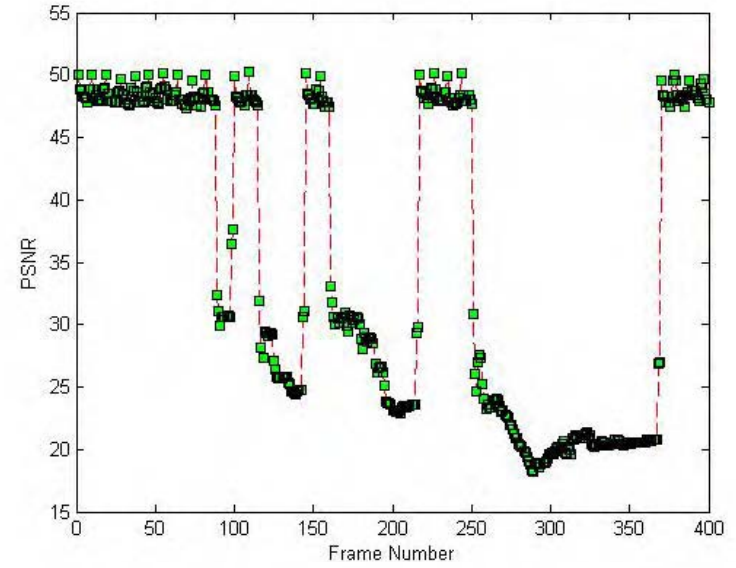
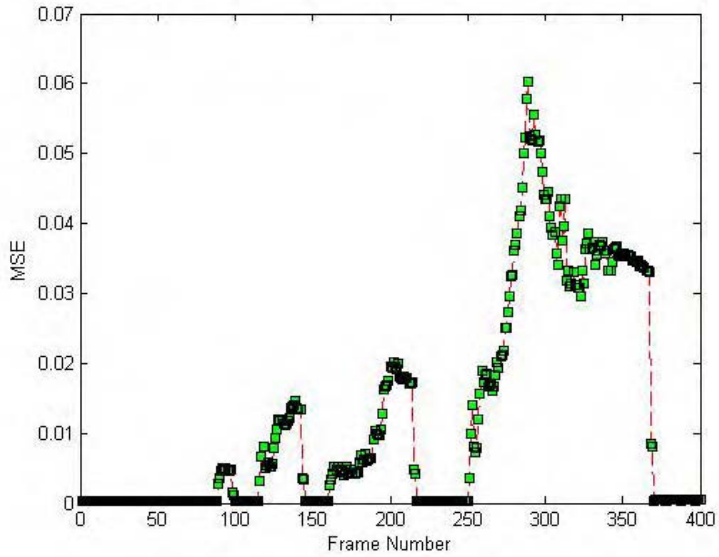
- Current MOS 'Mean Opinion Score' mapping used by most researches provides good correlation models between a VQA method and MOS for still frames.
- Recent research has revealed that certain temporal aspects have to be considered for a complete Quality assessment of Videos.
- A heuristic model is presented by Hauske et al. [11]. It combines spatial quality aspects in terms of PSNR and a temporal parameter in terms of frame rate FR. This model has been cited and used by several recent papers for VQA e.g. [12].

$$Q_m = -0.45(PSNR_{mean}) + 17.9 - (FR - 5) / 10.$$

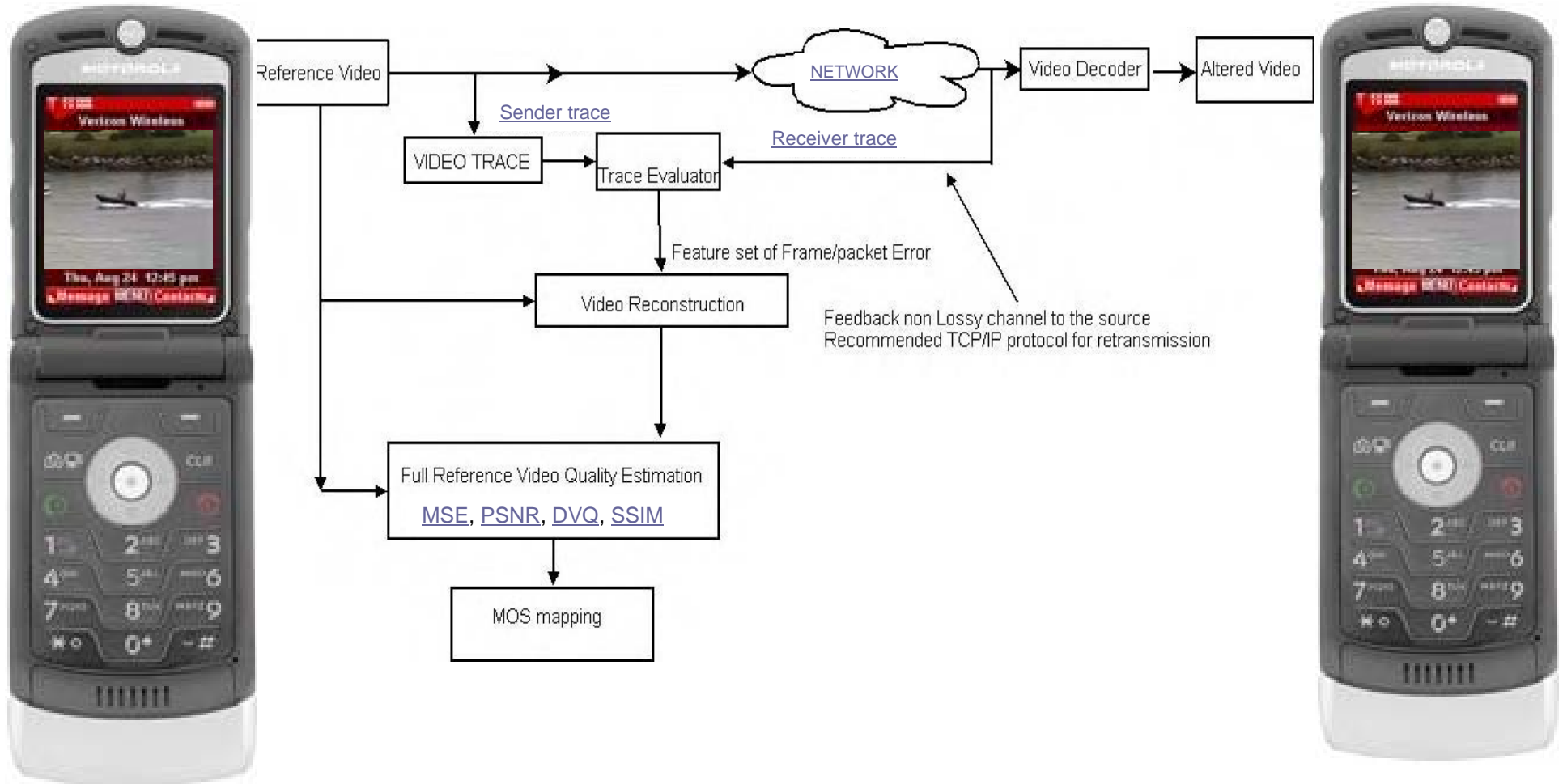
- A new Spatio-temporal Quality assessment model is proposed that accounts for Psychophysical aspects of Human Visual system*.

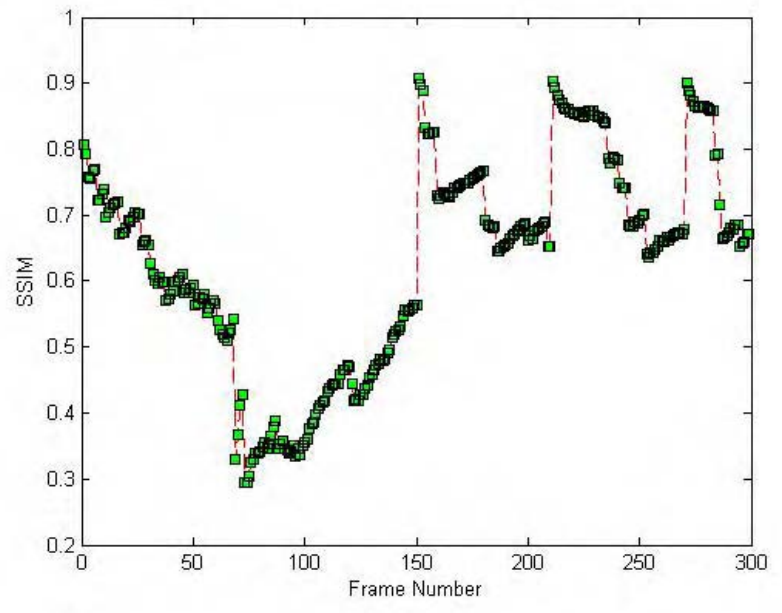
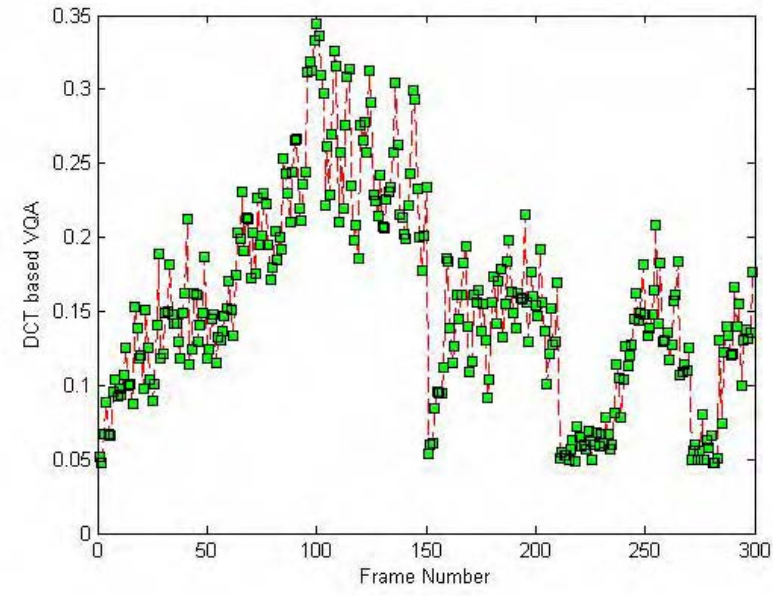
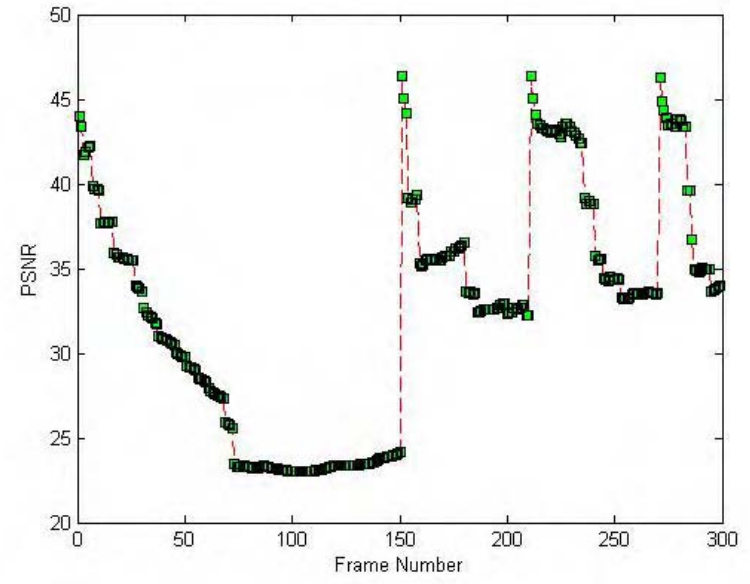
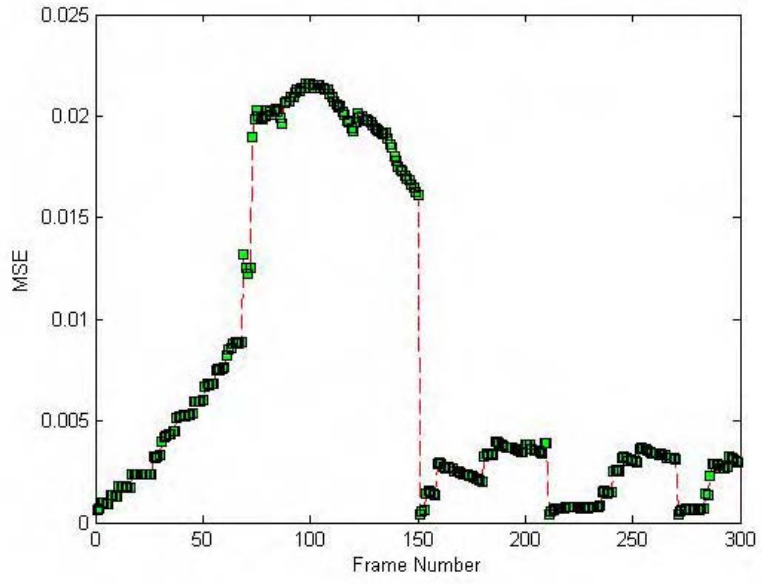
Mobile to Mobile QCIF Clips



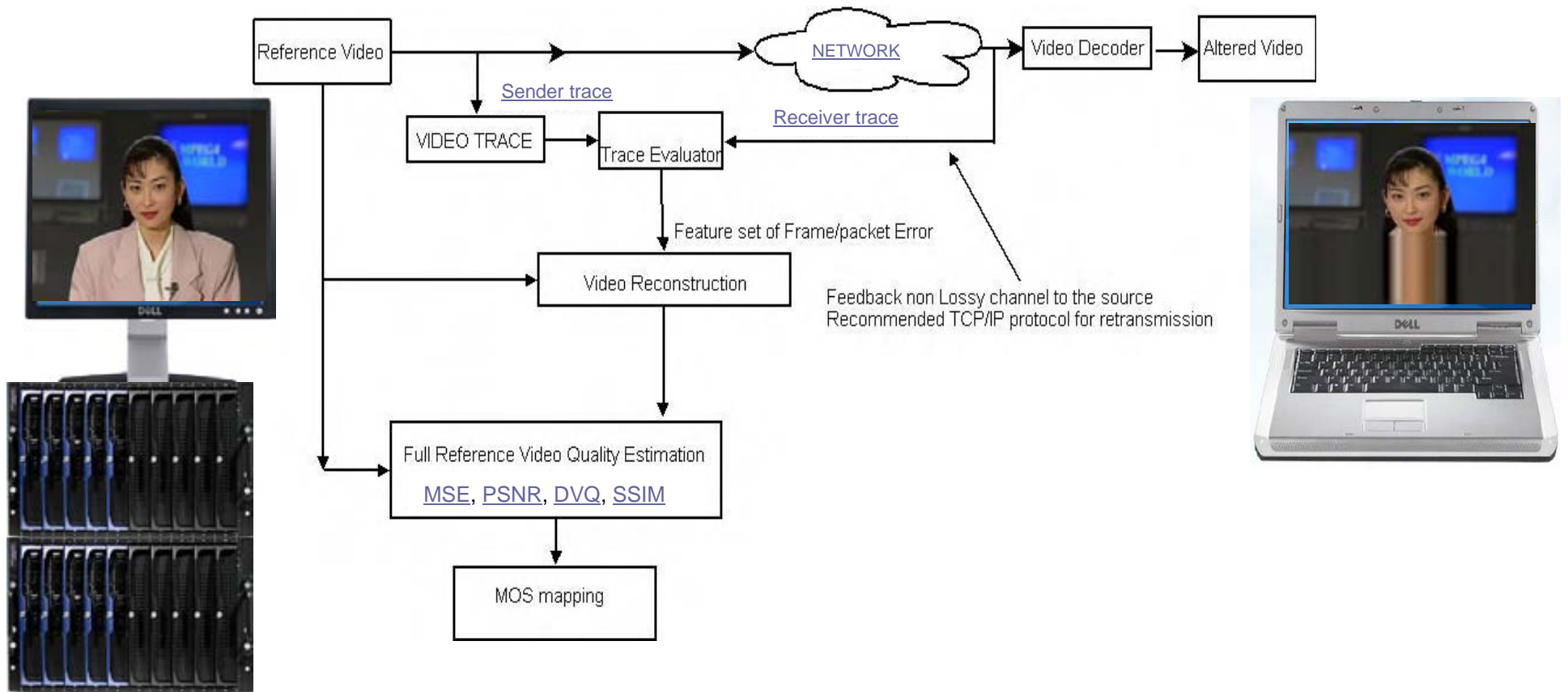


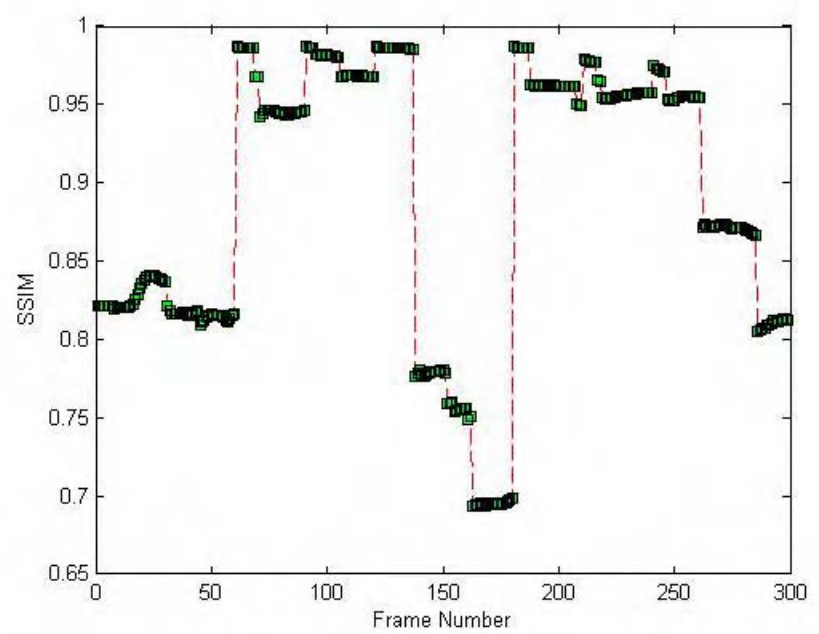
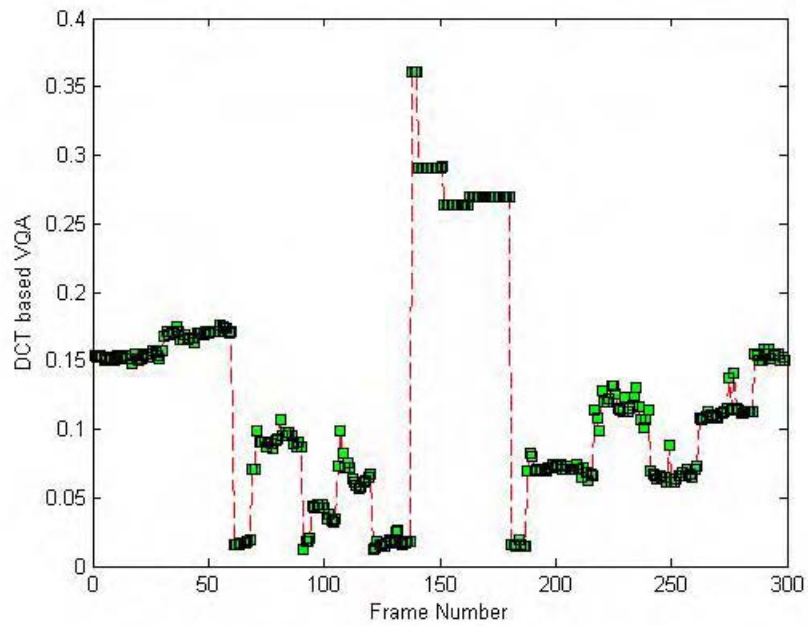
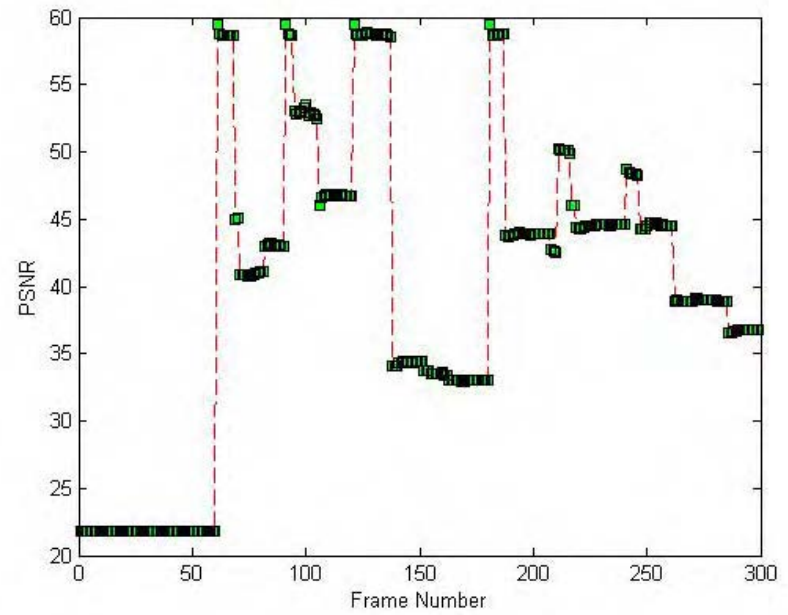
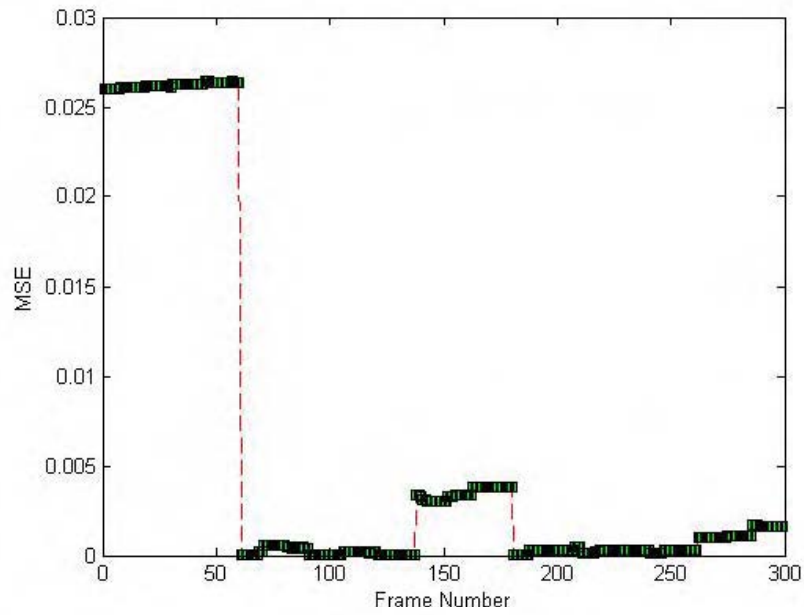
Mobile to Mobile QCIF Clips





PDSN to Mobile Aircard CIF Clips







Conclusions from the VQA

- The tests indicate that that MSE is not a suitable quality measure for Transmission based video quality assessment.
- The tests show distinct similarities between PSNR, DCT based algorithm and SSIM approach.
- The DCT based algorithm falls out of lines from other methods in cases when the images have more texture information than edges.



Applications and Future Work

- A **framework** for Evaluating the Quality of standard Video Transmissions over a wireless infrastructure (system/subsystem).
- A new **methodology** for Wireless service providers, to compare real-time traces of transmitted videos with that of received videos over a feedback loop. This will provide the service providers with real-time video quality analysis over their network.
- A **test bed** that can be used to test novel VQA approaches. It can also be used for new Encoding schemes, compression algorithms, Motion Estimation methods etc.
- A **test bed** to generate training data for Iterative schemes or ANNs, to be used by the Vendor to Quantify the Video Quality Degradation over their infrastructure based on key information regarding the Video Transmission. E.g. RF fading, total packet loss, flow control, channel bandwidth used, duration of the transmission, Frame rate, Video resolution (CIF(Common Intermediate Format), QCIF (Quarter Common Intermediate Format) etc.).



References

- 1) G. H. A. Lo and I. Niemegeers, "Evaluation of mpeg-4 video streaming over umts/wcdma dedicated channels," *Proc. IEEE Int. Conf. Wireless Internet*, vol. 1, pp. 8–10, Jan. 2005.
- 2) B. R. J. Klaue and A. Wolisz, "Evalvid - a framework for video transmission and quality evaluation," *Proc. 13th Intl Conf on Modeling, Techniques and Tools for Computer Performance Evaluation, Urbana, IL*, 2003.
- 3) P. Seeling, *Video traces for network performance evaluation: a comprehensive overview and guide on video traces and their utilization in networking research*. Springer, 2007.
- 4) M. R. P. Seeling and B. Kulapala, "Network performance evaluation using frame size and quality traces of single-layer and two-layer video: A tutorial," *IEEE Communications Surveys and Tutorials*, vol. 24, no. 4, 2004.
- 5) video trace research group at ASU, "Yuv video sequences," <http://trace.eas.asu.edu/yuv/index.html>.
- 6) VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar. 2000. <http://www.vqeg.org/>.
- 7) F. Xiao, "Dct-based video quality evaluation," Final Project for EE392J Stanford Univ. 2000.
- 8) A.B. Watson, "Toward a perceptual video quality metric", *Human Vision, Visual Processing, and Digital Display VIII*, 3299, pp 139-147, 1998.
- 9) Z. Wang and A. C. Bovik, *Modern Image Quality Assessment. Synthesis Lectures on Image, Video and Multimedia Processing*. Morgan and Claypool, 2006.
- 10) Z. Wang, "The SSIM index for image quality assessment," <http://www.cns.nyu.edu/zwang/files/research/ssim/>.
- 11) G. Hauske, T. Stockhammer, and R. Hofmaier, "Quality of LowRate and LowResolution Video Sequences," MoMuC2003, 2003.
- 12) Frank H.P. Fitzek, B. Can, R. Prasad, and M. Katz. "Traffic Analysis and Video Quality Evaluation of Multiple Description Coded Video Services for Fourth Generation Wireless IP Networks", Special Issue of the International Journal on Wireless Personal Communications, 2005.