Full-Reference Visual Quality Assessment for Synthetic Images: A Subjective Study

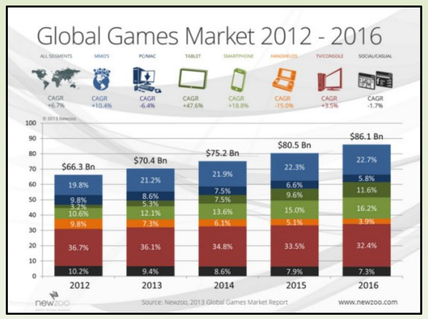
THE UNIVERSITY OF TEXAS AT AUSTIN

Ms. Debarati Kundu and Prof. Brian L. Evans
Embedded Signal Processing Laboratory (ESPL)
debarati@utexas.edu, bevans@ece.utexas.edu

WNCG Wireless Networking & Communications Group

Motivation

- *Problem*: Automate Image Quality Assessment (IQA) for synthetic scenes
  - Give designers of video games and animated films immediate feedback for rendering artifacts
  - Give video game designers immediate feedback on transmission artifacts in cloud gaming
- *Approach*: Evaluate IQA algorithms for high-resolution synthetic scenes
  - Develop public database of pristine and distorted synthetic images
  - Conduct subjective testing for visual quality assessment
  - Correlate IQA algorithms with subjective test results

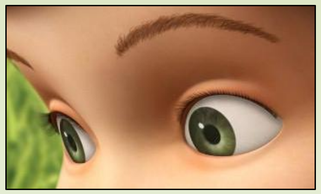


ESPL Synthetic Image Database

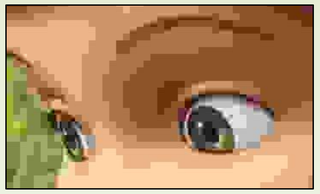
- 25 pristine reference images
- Used less severe distortions than natural images
- 500 distorted images
  - 5 distortion types
  - 4 distortion levels for each image and each distortion type

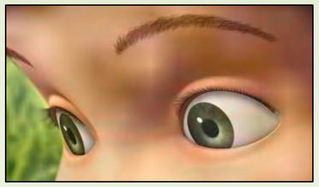

Blur


Interpolation


Original

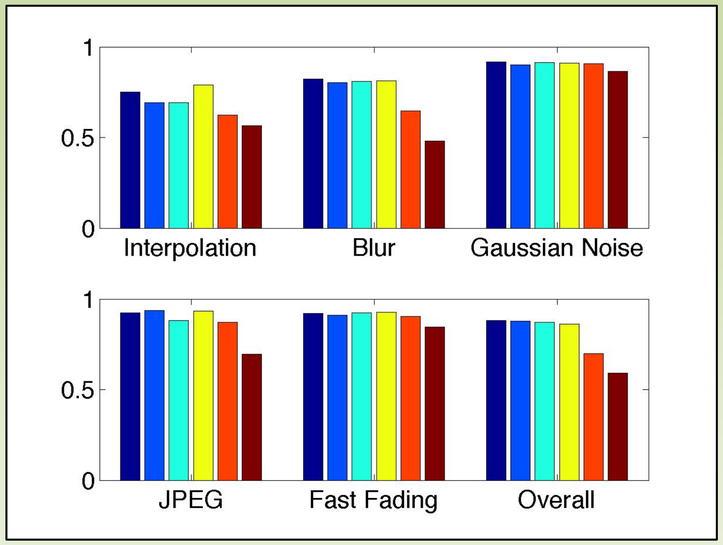

Gaussian Noise


JPEG Compression


Fast Fading

Performance of Full-Reference Algorithms

- Evaluate 23 image quality assessment algorithms



Spearman's Rank Ordered Correlation Coefficient between leading full-reference metrics and subjective opinion scores

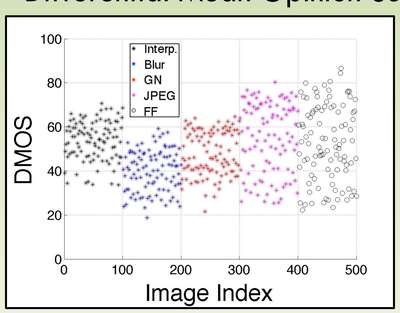
Spectral Residual Based Similarity (SR-SIM)	Feature Similarity Index (Color) (FSIMc)
Visual Saliency Induced Index (VSI)	Most Apparent Distortion (MAD)
Multiscale Structural Similarity Index (MS-SSIM)	Peak Signal-to-Noise Ratio (PSNR)

Synthetic Scenes

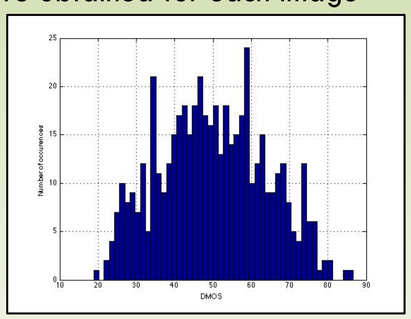
- Sources of graphics data
  - Animation studios
  - Kinect, video games
- Artifacts
  - Interpolation, banding, ringing, noise, blur
  - JPEG and wireless distortions
- Multiple artifacts may occur at same time

Subjective Study Methodology

- Evaluated on Dell U2412M 24-inch displays
- 64 subjects evaluated every image over three sessions
- Single Stimulus Continuous Evaluation on a scale of [0,100]
- Reference and distorted images evaluated in same session
- Differential Mean Opinion Score obtained for each image



Scatter plot of DMOS scores



Histogram of DMOS scores

Conclusion

- Processing Artifacts
  - Interpolation: MAD ✓
  - Blur and Gaussian Noise: SR-SIM ✓
- Transmission Artifacts
  - JPEG Compression : FSIMc ✓
  - Fast Fading: MAD ✓
- Overall: SR-SIM ✓
- Saliency-inspired pooling strategies perform well
- PSNR does reasonably well for additive noise & fast fading
- Interpolation and Blur
  - Less severe distortions in the database
  - Result in near-threshold artifacts

Future Work

- Conduct subjective tests for larger number of graphics artifacts
- Evaluate no-reference image quality measures on database
- Applicability of natural video statistics in animation sequences

http://users.ece.utexas.edu/~bevans/papers/2015/imagequality/