Joint Source and Channel Coding techniques for 3D Video

Valentina Pullano
XXV cycle
Supervisor: Giovanni E. Corazza
January 25th 2012
Overview

• State of the art
  • 3D videos
  • Technologies for 3D video acquisition & rendering
  • Video Coding Standards
  • VQE Metrics

• Ongoing activities
  • LA-FEC & Unequal Time Interleaving
  • PSNR & Sequence alignment technique

• Future Developments
  • Multidimensional Layer Aware FEC for 2D and 3D video

• Publications and Short courses
STATE OF THE ART
3D video:

• 3D video is a hot-topic
• Add the third dimension to traditional video: depth
  • Visual perception results more natural and incisive
• 3D technologies are the next revolution in the history of TV and multimedia content fruition
• Three dimensions → Three major factors of success
  • Technology: Ability to capture, show and process 3D
  • Content: Availability of interesting 3D contents
  • Quality: Attractive to consumer
• Stereoscopy: phenomenon at the base of the 3D perception
  • The phenomenon was first discovered by Euclid in 208 b.C.
  • The term Stereoscopy addresses any technique able to create the illusion of depth in an image
Technologies for 3D video acquisition & rendering:

- **Content acquisition:**
  - Stereoscopic cameras
  - Video color 2D + depth map
  - 2D to 3D conversion
  - Computer generated imagery (CGI)

- **3D scene representation**
  - Stereoscopic (SbyS, O&U, Interlaced)
  - Image plus depth
  - Multiview

- **3D display**
  - Head Mounted Display
  - Stereoscopic Display → Needs of special glasses
  - Autostereoscopic Display → No glasses
  - Holographic and volumetric display
Video coding standard:

• High efficiency compression and protection techniques are required:
  • Double storage space in comparison to monoscopic video
  • Double bandwidth resources and higher computational complexity
  • Heterogeneity of receiver devices and network conditions

• Standard H.264/AVC (MPEG4 Part 10)
  • State of the art standard in video compression
  • Joint Video Team: ITU-T VCEG & ISO/IEC MPEG
  • Layered organization: VCL & NAL
    • The compression meets the network
  • H.264/AVC video coding extensions:
    • Scalable Video Coding (SVC)
    • Multiview video Coding (MVC)
    • Multiple Description Coding

• Efficient Forward Error Correction techniques
  • Upper Layer Coding
  • Raptor and RaptorQ codes
Video Quality Evaluation Metrics:

- **Subjective quality assessment**
  - Most accurate & reference for multimedia quality
  - Standardized methodology (ITU-R Recommendation BT.500-10 "Methodology for subjective assessment of the quality of television pictures")
  - Cons: Time consuming and high human resources are needed

- **Objective Video Quality Evaluation**
  - Mathematical models
  - Applicable to compressed and/or uncompressed domain
  - PSNR (Peak Signal to Noise Ratio) is the widely used objective metric
  - The Human Visual Perception is difficult to be taken in account
ONGOING ACTIVITIES

LA-FEC & UNEQUAL TIME INTERLEAVER FOR SVC VIDEO
**Motivation:**

- **Scalable video coding (SVC) adds new features to broadcast services**
  - Support of heterogeneous devices, Graceful degradation,…
- **Upper layer Forward Error Correction (FEC) defined to cope with errors the physical layer cannot cope with**
  - The transmission of multi-layer video such as Scalable Video Coding (SVC) and MVC (Multiview Video Coding) is subject to a particular FEC optimization
- **Unequal Time Interleaver for increasing the time diversity of the signal and the robustness against burst errors**
- **Fast Zapping Service Provisioning:**
  - lower quality and faster decoding for the base layer
  - higher quality and longer time interleaver for the enhancement layer
Scalable Video Coding:

- A scalable video stream is a compressed representation of a video s.t.:
  - The representation is made up of \textit{layers}.
  - Layers provide \textit{incremental refinement} of the video sequence.
  - \textit{Efficient} representation in terms of reconstructed image quality for a given rate.

- Different kind of scalability:
  - SNR scalability \(\rightarrow\) quality, bit-rate scalability
  - Space scalability \(\rightarrow\) resolution scalability
  - Time scalability \(\rightarrow\) frame-rate scalability
**SVC: Dependencies within Media Stream + FEC (1/2)**

- Inter-layer prediction causes dependencies within a layered media stream
  - Dependencies lead to different importance of quality layers
    → If parts of the base layer is lost, typically the enhancement layer becomes useless

- **Solution**: Unequal Error Protection (UEP) is applied, such that the base layer is more heavily protected
SVC: Dependencies within Media Stream + FEC (2/2)

- Two independent source blocks over a certain amount of data.
- FEC algorithm (Raptor or RaptorQ, ...) generates parity data
  - UEP increases protection of more important layer
- In cases, where the base layer cannot be corrected, the enhancement layer cannot be used due to missing references
  → Enhancement layer parity data is useless

**Solution:** Generate protection of enhancement layer across dependent layers (Layer-Aware FEC)
SVC: LA-FEC approach

- Layer-Aware FEC generates parity bits of dependent layer across predicted layer
  - The total number of redundancy symbols remains constant and redundancy symbols of different layers in the same dependency path can be jointly used for decoding
- LA-FEC parity bits protect both, base and enhancement layer
  - LA-FEC increases protection of more important layer without any increase in bit rate
  - Due to FEC generation follows existing dependencies within the media stream, LA-FEC never performs worse in terms of video quality

Raptor and RaptorQ codes are used as FEC
LA-FEC & Unequal Time Interleaving

- **Time Interleaving** increases the time diversity of a message by spreading the symbols of a media layer over a longer time period.

\[ \text{d}_{\text{inter}} : \text{required minimum time for decoding all the symbols of the SB} \]

- **LA-FEC and UI**
  - **Shorter time interleaving** is provided for the base layer with weaker robustness but shorter delay for enabling fast zapping services.
  - **Long time interleaving** is provided by the SVC enhancement layer with stronger robustness against burst losses but longer delay.
  - Thanks to the Layer-Aware FEC scheme adopted, the base layer also benefits from the improved time diversity of the enhancement layer.
Simulation Results

PSNR[dB] vs TB

Code Rate 0.5 and L_B = 1.4

Error Rate

TB

AVC-NoInterleaver
SVC-NoInterleaver
LAFEC-NoInterleaver
AVC-Interleaver
SVC-Interleaver
LAFEC-Interleaver

AVC-Interleaver
SVC-Interleaver
LAFEC-Interleaver
Simulation Results

FEC decoding probability vs Channel Error Rate for Base Layer with $L_B = 1.4$

![Graph showing FEC decoding probability vs Channel Error Rate for different interleaver types.](image)
Simulation Results

FER vs Channel Error Rate for Base Layer with $L_B = 1.4$
ONGOING ACTIVITIES

PSNR EVALUATION AND SEQUENCE ALIGNMENT BY MEANS OF A SLIDING WINDOWS MECHANISM
Motivation

- Video streaming applications are susceptible to packet losses and errors
  - Visible destructive distortions
  - Loss of synchronization between audio and video
  - Misalignment between the transmitted and the reconstructed video flow

- Goals:
  - Defining an objective metric which can be evaluated rapidly and automatically and which possesses a clear significance in terms of the subjective experience
  - Establishing methodologies for the design and analysis of digital video transmission systems
**PSNR (Peak Signal To noise Ratio)**

- **PSNR** is the most used objective quality metric
- It is computationally lightweight, applicable to any content type, source coding independent and easily interpreted
- **PSNR** is defined as the ratio of the squared useful signal peak over the mean squared error in decibel

$$\text{PSNR}(i, j) = 10 \log_{10} \frac{(2^P - 1)^2}{\text{MSE}(i, j)}$$

- The **MSE** is computed as the average quadratic pixel by pixel difference between the original video frame and the decoded video frame

$$\text{MSE}(i, j) = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [f_i(x, y) - g_j(x, y)]^2$$

- The **Y-PSNR** is evaluated performing the comparison considering the luminance frame only
The proposed method

1) Sliding window mechanism & Sequence alignment recovery

\[ L_W = L_{or} - L_{rec} + 1 - \text{losses}_{detected}(k - 1) \]

\[ \text{PSNR}(h, k) = 10 \log \frac{(2^P - 1)^2}{\text{PSNR}(h, k)} \]

\[ h = \hat{h}_{(k-1)} + 1, \ldots, \hat{h}_{(k-1)} + L_W \text{ with } \hat{h}_0 = 0 \]

\[ \hat{h}_k = \arg \max_h \text{PSNR}(h, k) \]

\[ l(k) = \hat{h}_k - (\hat{h}_{(k-1)} + 1) \]

\[ \text{losses}_{detected}(k) = l(k - 1) + l(k) \text{ where } l(0) = 0 \]

2) Full sequence reconstruction

In case of losses the missing frames are recovered by means of forward frame duplication technique

3) Total and Partial PSNR evaluation

The Total PSNR is evaluated between the origina sequence and the realigned one

The partial PSNR is evaluated over the subset of reconstructed frames

\[ \text{PSNR}_{\text{total}}(k) = \text{PSNR}_{\text{total}}(k - 1) + \text{PSNR}(\hat{h}_k, k) + \sum_{n=1}^{l} \text{PSNR}(\hat{h}_k - n, k) \]

\[ \text{PSNR}_{\text{partial}}(k) = \text{PSNR}_{\text{partial}}(k - 1) + \text{PSNR}(\hat{h}_k, k) \]
FUTURE DEVELOPMENTS

MULTIDIMENSIONAL LAYER-AWARE FORWARD ERROR CORRECTION
Multidimensional LA-FEC SVC/MVC

- LA-FEC approach can be extended to dependency structures in multiple dimension of SVC
- At the moment there are three dependency dimensions $D_1$, $D_2$, $D_3$ according to the temporal, spatial and fidelity dimensions in SVC
  - In each dimension several layers $l_{D_i}$ each layers depends on all lower layers of the same dimension and partially on the layers of other dimensions
  - All redundancy symbols FEC $l_{D1}$, $l_{D2}$, $l_{D3}$ are generated over all depending layers
  - The redundancy symbols within a particular dependency path can be jointly used for correcting all source symbols of that path
  - The base layer is included in all FEC symbols → errors in the base layer can be corrected using redundancy symbols belonging to multiple path
Multidimensional LA-FEC SVC/MVC
Publications

• **Journal Papers:**
  "Mobile TV Long Time Interleaving and Time Zapping"
  Authors: C. Hellge, M. Hensel, V.Pullano, T.Schierl, G.E.Corazza
  Paper to be submitted to: Special Issue on IEEE Transaction on Multimedia Cloud-Based Mobile Media: Infrastructure, Services and Applications

• **Conference Papers**
  "PSNR evaluation and recovery of video frame alignment in case of frame losses in real-time streaming applications"
  Authors: V.Pullano, A.Vanelli-Coralli, G.E.Corazza
  Paper to be submitted to: ASMS/SPSC 2012 5-7 Sept 2012 Baiona, Spain
Short Courses

• **International School of Scientific Computation and MATLAB**
  High computational & Grid Computing and Matlab Parallel Programming
  July 2010 - Palermo, Italy

• **Ph.D. School on Information Engineering**
  February 2010 – Naples, Italy

• **Ph. D. Course on Multimedia Databases**
  University of Bologna, Italy
  Prof. Ilaria Bertolucci

• **English Courses Upper Intermediate Level**
FUTURE DEVELOPMENTS

UNIVERSAL MODELING OF 2D/3D VIDEO TRANSMISSION OVER MEMORYLESS AND CORRELATED CHANNELS
Universal Modellind of 2D/3D video transmission

• Following from all the researches in the field the ultimate objective of my PhD. is the definition of an universal transmission chain

• No specific standards and methods are used
  • The new video coding standard HEVC (High Efficiency Video Coding) currently under development by the Joint Collaborative Team on Video Coding (ISO/IEC- MPEG and ITU-T VCEG) will overcome these issues
  • Further analysis will be done in this direction with the objective of extending the LA-FEC solution to the 3D case, enabling a full scalable encoder which allows to encode in a single bit stream any possible video flow in order to address the heterogeneous devices and network landscape with a scalable 3D video service
The proposed method

• The windowed PSNR method enable the evaluation of video quality in the presence of unknown frame losses in the uncompressed domain.
• The method is composed by different steps:
  1) Sliding window mechanism & Sequence alignment recovery
     The sliding window is shifted over the reconstructed sequence and the Y-PSNR is evaluated between the frame of the received video and all the frames contained in the window.
     The maximum PSNR value is find out inside the window and compared with the treshold for defining if and which frame have been lost
  2) Full sequence reconstruction
     In case of losses the missing frames are recovered by means of forward frame duplication technique
  3) Total and Partial PSNR evaluation
     The Total PSNR is evaluated between the origina sequence and the realigned one
     The partial PSNR is evaluated over the subset of reconstructed frames