FIA - WS 2014/15

1 Introduction

1.1 Internet

- loosely hierarchical: tier 1 to tier 3
- communications infrastructure enables distributed applications
- hyper... $\Rightarrow$ concentration of content (Amazon, Google, ...)

1.2 Protocol Layering

- ISO/OSI
- TCP, UDP, ICMP, UDP Like, SCTP, ...
- DHCP, NAT, ...

1.3 Analogue digital conversion

- Fourier transformation
- sampling theorem + Nyquist
- PCM-transmission
- small amplitudes are being encoded more detailed than large amplitudes

1.4 Color Coding

- monochrome, RGB, YCbCr
- RGB:
  - red, green, blue values between [0, {7,31,...,65535}]
  - nonlinear encoding of intensities
  - computer graphics
- YCbCr
  - TV and digital video
  - Y more important $\Rightarrow$ encode with more details
  - more efficient than RGB
  - can handle downsampling better
  - YUV is based color model used in analog color TV
    * YCbCr is scaled and offset version
    * YPbPr is the analog version of YCbCr

2 Digital coding of audio and video

2.1 Rate-Distortion Theory

- lossless compression algorithms
  - allow perfect reconstruction
  - low compression ratios
  - frequently encountered data is encoded more efficiently
- lossy compression algorithms
  - result is only a close approximation of original data
  - trade-off: distortion vs. required rate
  - much higher compression rate than lossless compression

  //TODO: bild einfuegen
- rate and distortion as measures for efficiency of compression and difference between reconstructed and original data
  - goal is to minimize distortion and rate
  - basic problem:
    * minimum expected distortion at a given rate?
    * minimum rate to achieve given distortion?
- distortion measures
  - mean square error $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$
  - signal to noise ratio $SNR = 10log_{10}\frac{\sigma_x^2}{\sigma_d^2}$
  - peak signal to noise ratio $PSNR = 10log_{10}\frac{x_{peak}^2}{x_d^2}$
- in order to maximize efficient communication maximize mutual information between x and y
- rate distortion function

2.2 digital image

- conversion between RGB and YUV
- downsampling J:a:b
  - 4:4:4 $\widehat{=}$ no downsampling
  - 4:2:2 $\widehat{=}$
  - 4:2:0 $\widehat{=}$
- statistical image modeling
  - all natural images occupy a tiny and unknown sloped space of all images
  - pixel intensities are dependent and correlated to the corresponding image $\Rightarrow$ pixel within images are highly correlated
  - correlation of pixels has high impact on image compression
- image transformations
  - negative transformations

  - log transformations
  - power-law transformations

- intensity:
  - change in brightness ⇒ shift of histogram
  - change in contrast ⇒ stretch/? of histogram

- filters based on convolution of neighbor pixels ⇒ enhancement, smoothing, edge, detection, ...

- the hoar transform
  - simplest useful energy compression
  - lossless retransformation is possible
  - transform original image into 4 parts (in 20)
    1. top-left: a+b+c+d, 4 point average, very important
    2. top-right: a-b+c-d, average horizontal gradient, less important
    3. bottom-left: a+b-c-d, average vertical gradient, less important
    4. bottom-right: a-b-c+d, diagonal curvature, less important
  - 1 is more expensive while 2-4 is cheaper to encode
  - reordering required to provide "typical" representation

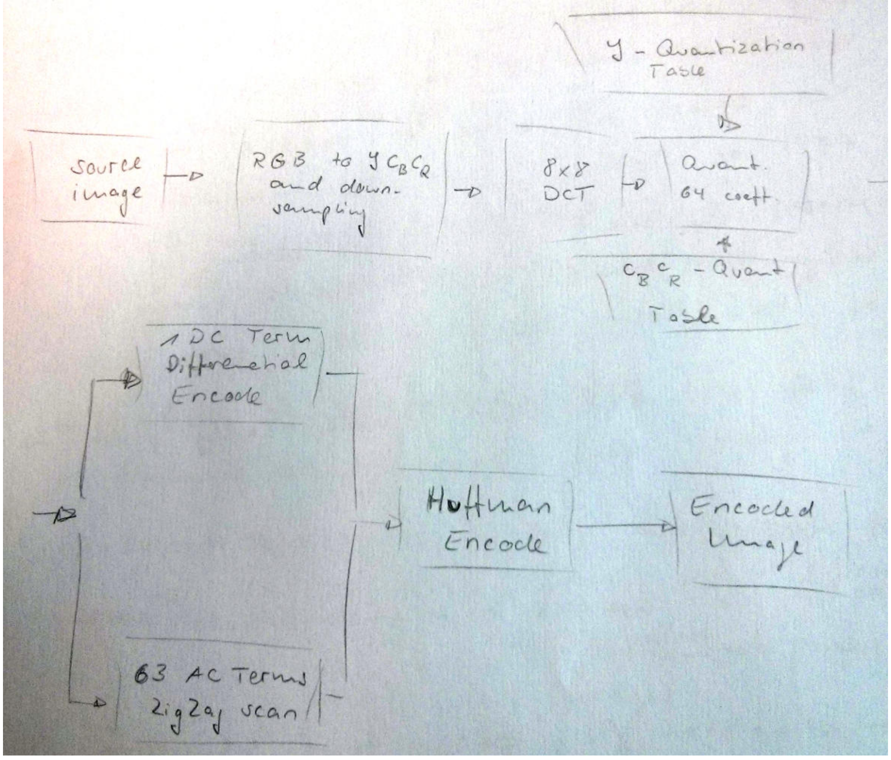
- entropy – quantifying the required bitrate
  - entropy H_x represents the mean number of bits per pixel that are required to encode the image
  - $H_x = \sum_{i=0}^{M-1} p_i - log_2\left(\frac{1}{p_i}\right)$ where $p_i = \frac{\text{pixel in bin i}}{N}$ where N is the number of pixels
  - estimated number of bits needed is $H_x \cdot N$

- the Karkunen-Loeve transform (KLT)
  - is optimal for minimizing bitrate
  - not or seldom used in practice ⇒ slow and complex

- the discrete cosine transformation (DCT)
  - seems to put most energy on a small number of values
  - linear transformation
  - each real world image can be represented by a combination of the DCT bases
  - DCT bases fit better to typical images than FFT or DST
  - first coefficient of DCT is mean of values being transformed and represents average tone of the block. subsequent blocks add further detail
  - JPEG: 8×8 DCT (empirically found to be good)

- need for wavelets
  - signals are available in time-domain but processing frequency information is much more easy
  - transformations (e.g. FFT) translate time-domain signals into frequency domains
    * useful for stationary signals where all frequencies are present at all times
    * not useful for instationary signals: both information required
    * solution: short term transformation—transformation of small fixed timewindows (similar to blocks of images)
  - wavelet analysis uses different sized windows
    * high frequency parts ⇒ small windows ⇒ good time resolution
    * low frequency parts ⇒ big windows ⇒ good frequency resolution

2.3 JPEG compression

- joint photographic experts groups
- 

2.4 digital video

- high corellation between successive frames
- interframe differential pulse code modulation (⇒ vgl. skript skizze)
- frame replenishment ⇒ pixels are only transmitted if a specified threshold is exceeded, otherwise nothing is transmitted
- change detection
  - pixel wise change detector
  - block wise change detector
  - comparison between current and previous frame
    * substract previous frame from current one
    * convert to absolute value
    * generate 3×3 averages
    * check for threshold
- motion compensated coding
  - changes between frames due to moving objects ⇒ estimation of motion vectors
  - prediction and original frame may differ significantly
    * solution: compute and transmit prediction error additionally
    * higher coding efficiency but also higher computational complexity
  - three-stage-coding:

    Motion Analysis ⇒ Prediction and differentiation ⇒ Encoding

- block matching

  - partitioning of frame into nonoverlapping equally spaced and fixed size rectangular blocks
  - smaller block size $\Rightarrow$ better approximation but higher computational complexity
  - 16×16 blocks used in MPEG-1 and -2
  - calculate best displacement vector for each block
  - search strategies:
    * full search
    * 2D logarithmic search
    * diamond search

- grouping of elements
  - sequence $\rightarrow$ frame $\rightarrow$ slice $\rightarrow$ macroblock $\rightarrow$ block $\rightarrow$ pixel

- different frame types
  - I-frame: intra-coded-frame: independent of other frames
  - P-frame: predictively coded frame: depends on previous frame
  - B-frame: bidirectionally predicted frame: depends on previous and subsequent frames

- scalable video encoding – H.264/SVC
  - different frame rates
  - different resolutions
  - variable image quality
  - layered video codec (LVC)
    * base layer is encoded in H.264/SVC
    * enhancement layers allow refinement of base quality
    * different combinations of temporal, spatial and quality refinement is possible
      TODO: Bild einfuegen
    * refinement is represented by 3D cube of cubes
  - temporal scalability: adding B-frames to the base layer
  - spatial scalability: 3 mechanisms
    * interlayer prediction
    * residual prediction
    * motion prediction
  - quality scalability
    * interlayer prediction (like in spatial scaling)
    * up sampling
    * different quantization parameters
  - idea is to adept the video quality to the available bandwidth by removing or adding enhancement layers

- network abstraction layer (NAL)
  - frames are packaged in NAL units
    * payload of 1 package is one layer of one frame
    * layers are separated into different streams
    * MANE (media aware network element)
      · decides package forwarding
      · separate layers on separate flows (???) allow adapted forwarding

- use of LVC and NAL
  - adaption in case of device or network limitations

- reduce storage size by removing enhancement layers
- video broadcasting: base layer free + payed enhancement

- multi description coding
  - n frames/s using 1 description
  - 2n frames/s using 2 description
  - 2 alternatives
    1. independent description: robust + inefficient
    2. → less robust + more efficient
  - robust against package loss

- comparison of different video codings
  - layered > MDC2 > MDC1
  - non variable coding is most efficient but not adaptive to changes in network limit or device limitations

2.5 Estimating quality of media applications

- what is QoE?
  - the overall acceptability of an application or service as perceived subjectively by the end user
  - definition is still evolving
  - closely related to QoS
  - QoE is usually a subjective value
  - assessed by asking or observing users
  - related values:
    * quality of presentation, quality of delivery
    * hourglass model matches QoX with layers of the ISO/OSI model
    * QoS → QoD → QoP → QoE
  - often characterized with MOS (mean opinion score)

- QoE assessment methods and tools
  - subjective QoE
    * ask or observe the user (use questionaires, rating scales, annotations)
  - mean opinion score
    * widely used (politics, marketing, ...)
    * the likely level of satisfaction of a service or product as appreciated by an average user
    * average might be resulting by two extremes
    * use discrete scales and an odd number of steps

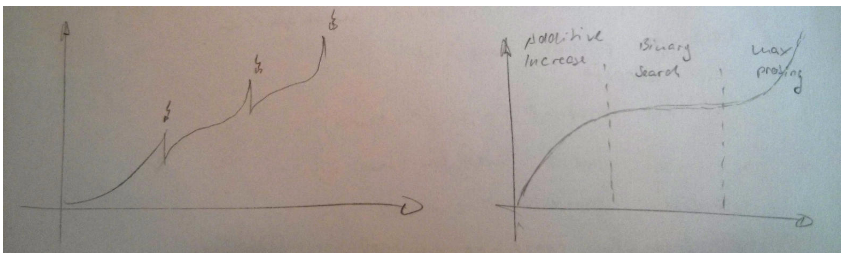
- subjective tests are often expensive and time consuming
  - objective algorithms are desired (MOS predictors)
    * MSE (bad) ⇒ doesn't care about reordering
    * PSNR ⇒ correlation of c.a. 80 % to subjective studies (but sometimes much worse)
    * maximum pixel deviation
    * multiscale structural similarity (MSSIM) ⇒ very good

- region of interest ⇒ salient vs. non-salient regions (e.g. football)

3 data communication and transport protocol

- connection-oriented:
  - setting up a call and release the call afterwards (TCP)
- connection-less:
  - packet switching without call establishment

3.1 single path transport protocol

- UDP and UDP-lite
  - SRC part || DST part || length || checksum || data
    - suiteable for real-time applications (nondelayed transmission)
    - works well for broadcasting (e.g. service discovery, video broadcast)
  - UDP-like replaces length with checksum coverage
  - UDP-like headers must be protected by the checksum
- TCP basics
  - SYN and ACK flags
  - sequence numbers: number of packet ascending
  - acknowledgement number: if ACK is set, this is the next sequence the reciever expects
  - connection establishment (3-way-handshake)
    - client sends SYN with random sequence number A
    - server responds ACK with acknowledgement number $A+1$ and random sequence number B
    - client sends ACK with acknowledgement number $B+1$ and sequence number $A+1$
    - result: full duplex connection
  - key features of TCP
    - orders data transfer
    - retransmission
    - error free data transfer
    - flow control
    - congestion control
  - TCP measures round-trip-time (RTT) and calculates expected RTT for future packets
- compound TCP
  - combines delay-based and loss-based congestion control
  - higher throughput than normal TCP at high RTT
- cubic TCP
  - 

- fair for low RTT, very aggressive for high RTT

- RED – random early detection
  - schedulers for congestion avoidance
  - idea: monitor average queue size and drop packets based on statistical probabilities
  - problems: has strong impact on TCP (retransmission)
- ECN – explicit congestion notification
  - allows end-to-end notifications of network congestion without dropping packets
  - ECN-routers may set a mark in the TCP headers to signal impending congestion to the reciever. the reciever echos a signal to the sender which then can reduce transmission rates
  - can reduce the number of dropped packets
  - may reduce performance on highly congested networks
- DCCP – congestion control without reliability
  - TCP: long living flows to guarantee fairness
  - UDP: short living flows and multicast applications
  - problem: UDP is used for long-living flows with live constraints: streaming, VoIP, online-gaming ⇒ no congestion controll using UDP may result in congestion collapse of the internet
  - requirements:
    * choice of congestion controll mechanisms
    * low per-packet overhead
    * ECN support
    * middlebox traversal (NAT/Firewall)
  - congestion controll mechanisms
    * TCP-like: congestion controll
    * TCP-friendly: congestion controll
  - SCTP – stream controll transmission protocol
    * properties:
      · security: 4-way-handshake
      · resilience: multi-homing, crc32–checksum
      · practical features:
      · IPv4 and IPv6 simultaneously
      · multi-streaming
      · message-oriental
      · support extensions
    * multi-homing
      · SCTP associates two nodes (connection)
      · if primary path between nodes breaks connection switches to another path
    * multi-streaming (vgl. script 3.49)

3.2 multi-path transport protocols

- multi-path TCP: split data transmission via multiple paths
  * enhances reliability and flexibility
  * allows balancing of data transmission via links with different limitations
  * common problem: out-of-order arrival