|  |  |
| --- | --- |
| Hearing: [20Hz-20KHz]: Regular = 1-5KHz Speech:[200Hz-8KHz]  Sample it, quantize it, compress it  **Nyquist-fs > 2\*fm**  Remove certain frequency to handle sampling, low-pass filters  More quantization, more precise **Bit rate** = Bit/sample \* sample/sec  Nyquist Lime: max data = 2\*Sample Rate\*log2(V), V = 2bits  Post Code Modulation = Uniform Quantization levels  Nonuniform Quantization levels  **DifferentialPCM**: dt = Xt – Xt-1 **Signal to Noise** = Avg Signal Power/ Avg Noise Power  **SNR(db)** = 10log10(SNR) **Noise** = 2A/N **Signal Energy** = A2/2 , Noise Energy = A2/(3N2)  **Color**: Electromagnetic wave(400nm to 700nm) **Hue**: Pure color (Dominant wavelength of the light) **Saturation**: Intensity of a specific color **Brightness(lightness):**Perceived intensity  **Image Size** = Height x Width(in pixels)(Bits/pixel to get it in bits) 320X240, 640x480, 1920x1080 **Digital Image-Quantization** =number of bits per pixel  B\W(2b), Grey(8b), Color(RGB, CMY, YIQ)(24b), Index table  Spatial Domain: Planar region of intensity values at time t  Frequency: Organizing pixels according to their changing intensity  Visual Resolution: Depends on Image size and view distance  **Human Eye: 16fps** for it to recognize as a continuous sequence  Sample Rate (25fps EU, 30fps US)  **Flicker:** Perceived if frame rate or refresh rate of screen too low (<50Hz)  Display Ration  Conventional tv(width/height)(4/3), HDTV(16/9), Cinema(1.85:1, 2.35:1)  Intensity (brightness): luminance, Color: chrominance  **Depth perception**-Visual ability to perceive world in 3D and the distance of an object  **Depth sensation** is corresponding term for animals (it is not known whether they perceive depth in the same subjective way that humans do)  **PAL (NTSC)** video standard- Y is luminance, UV(IQ) are chrominance  **YUV** from RGB: Y = .299R + .587G + .114B, U = 0.492 (B - Y), V = 0.877 (R - Y)  **YIQ** from RGB: Y = .299R + .587G + .114B, I = .74 (R - Y) - .27 (B - Y), Q = 0.48  **NTSC(USA, Jap)**525 lines, 59.94Hz, Interlaced, 480 lines visible, 29.97fps(480i)  **PAL(Europ, China)**625 lines, 50Hz, Interlaced, 576 lines visible, 25(576i)  Cinema 24Hz non-interlaced, SECAM 50Hz intrlaced  Interlaced: Odd line would be refreshed and then even  Progressive: Line refreshed in order  Cinema  **HDTV (Digital) Resolutions:** 1920x1080 (1080p) – Standard HD (HDTV)  2160p, …, 096x2304 (4096p) – 4K High HD  Frame rate: HDTV - 50 or 60 frames per second, HDTV – 120 fps  Interlaced (i) and/or progressive (p) formats  Conventional TVs – use interlaced formats  Computer displays (LCDs) – use progressive scanning  **YCrCB -**Subset of YUV that scales and shifts the chrominance values into range 0..1  Y = 0.299R + 0.587G + .114B, Cr = ((B-Y)/2) + 0.5, Cb = ((R-Y)/1.6) + 0.5  **Color space compression**: YUV444 24 bpp, YUV422 16 bpp , YUV411 12 bpp  DVD video-1997, 704x480 at 29.97 fps, 704x576 at 25 fps, Bitrate: 9.8 Mbps  Blu-ray-2006, 1920i (@59.94 fps) – interlaced, 1920p (@24 fps) – progressive, Bitrate: 40 Mbps**3DTV-**Refresh rate > = 120, synchronized shutter glasses to enable different views for different eyes **Fixed-length coding** - smallest number of bits per symbol needed is *L ≥ log2(N)* bits per symbol **Theoretical minimum average** number of bits per code word is known as Entropy (H) **Entropy** – measure of uncertainty in random variable**: H** = **Compression Ratio** = Avg L(Mess)/Avg L(codeword)  Symmetric compression- Same time to encode/decode, Asymmetric-slow for encode and fast for decode **Run-length Encoding** (RLE): abcccccccccdeffffggg 🡪 abc!9def!4ggg  `**Avg Codeword Length**: L =  JPEG (Joint Photographic Experts Group)-Prep, Processing-Fourier Transforms, Quantization, Lossless **Images** are divided into data units, called blocks –DCT operates on blocks Lossy mode – blocks of 8x8 pixels; lossless mode – data unit 1 pixel  **Non-interleaved**: scan from left to right, top to bottom for each color component  I**nterleaved**: compute one “unit” from each color component, then repeat  full color pixels after each step of decoding  but components may have different resolution  Image Processing, Shift values [0, 2P - 1] to [-2P-1, 2P-1 - 1]  Forward DCT Convert from spatial to frequency domain  **F(0,0)=DC coefficient**, Determines fundamental color of the block  **F(0,1)-end are called AC coefficients**, their frequency is non-zero in one or both directions  Separate DC from AC components: DC components change slowly, thus will be encoded using difference encoding  Quantization: In JPEG – use quantization tables Fq(u,v) = F(u,v)/Quv  **JPEG-2000** goes to 1:20, Low bit rate compression performance  Lossless and lossy compression, Large Images: > 64Kx64K  Single decompression architecture, Transmission in noisy environment  CGI, Compound Document (Picture with text), Progressive transmission by pixel accuracy and resolution **Random code-stream access and processing**  User defines “regions-of-interest” in the image to be randomly accessed and/or decompressed with less distortion than the rest of images **Wavelet-based coder**  This method permits coding of still images with high coding efficiency as well as spatial and SNR (signal-to-noise ratio) scalability at fine granularity  **JPEG-2000**: transformation from RGB to YCBCR or YUV , Irreversible Color Transform or **Reversible Color Transform**   * DWT (Discrete Wavelet Transform) extracts information from the source image at different scales, locations and Orientations * JPEG-2000 uses two techniques in wavelet-based coder | * + 2D wavelets, multi-scale transforms * **Wavelet is defined** as a set of basic functions, derived from the same prototype function * Prototype function is known as “mother wavelet”, Examples: “Mexican Hat” wavelet, Haar wavelet * Wavelet transform coders process high and low frequency parts of image independently, DCT methods have difficulties with high frequency * **Wavelet** method transforms image as a whole (not subdivided into pixel blocks), No blocking artifacts occur, Wavelet coders degrade gracefully   **FWT:LL HL,LH, HH**   * First filtered along the x dimension, down-sampled by factor 2 without loss of information, down sampled along the y dimension resulting in four sub-images   MPEG information: data compression 1.5 Mbps, Lum and 2 Chro(2:1:1) 8 x< 768, y<576, 8b per component, No Quant table  Aspect ratios: 1:1 (CRT), 4:3 (NTSC), 16:9 (HDTV)  Refresh frequencies: 23.975, 24, 25, 29.97, 50, 59.94, 60 Hz  Macroblock(16X16 L, 8X8 C); Process one component at a time  I-Frame(Reference), P(Prev I or P), B(Prev I and Next P)  SSD = SAD =  Target – ½(past + future)  **H.261**-Video Code for Conferencing   * 1988, Symmetric, Rate of pX64Kbps, I and P, BR(40kbps-2Mbps), CIF (352x288 luma, 176x144 chroma), QCIF (176x144 luma, 88x72 croma) using 4:2:0 sampling scheme, MacBl, 16x16 luma,8x8 chro 4:2:0 YCbCr color space, DCT, Scalar quantization, zig-zag, RLE, DCT Coeff. May go to 0, with few DC and AC coeff blocky, deblocking filter, staircase noiseand mosquito noise, low pass filters   **H.263** – video coding for low bit rate communications 1996   * low bit rate transmission PB-frames mode, Temporal, Spatial and SNR scalability, b-picture are created * PB-frame consists of one P-picture which is predicted from last decoded P-picture and B-picture which is predicted from last decoded P-picture and the P-picture currently being decoded   **H.264/MPEG-4 AVC Part 10 2003**   * Blue-ray, standard for YouTube and iTunes, supported by Flash and Silverlight, satellite and cable tv service * YCbCr 4:2:2 and YCbCr 4:4:4, Scalable Video Coding (SVC)( Encoding of high-quality video stream that contains one or more subset of bitstreams), Multi-view Video Coding (MVC) -Construction of bit-streams that represent more than one video of a video scene   + Multi-view High Profile (arbitrary number of views);   + Stereo High Profile (two-view stereoscopic video); * Use previously-encoded pictures as references, Allow up to 16 reference frames, Use variable block size from 16x16 to 4x4, Use multiple motion vectors per macro-block (one or two per partition where partition can be a block of 4x4), new transform design * Quant scaling matrices selected by encoder based on perception optimization * Context-adaptive variable-length coding, Context-adaptive binary arithmetic coding (CABAC) * H.264 – major leap forward towards scalable coding and multi-view capabilities   **H.265/HEVC/MPEG-H Part 2**   * Low bitrate target – target 2:1, Improve resolutions (8K by 4K and 4K by 2K) and frame rates(120>), 1080p50/60 services, encode 10X more complex, decode 2x-3x complex * AVC- macro-blocks 8x8 and 4x4 transform sizes * HEVC- Coding unit size 64x64 to 8x8, 32x32, 16x16, 8x8 and 4x4 transform sizes   Arithmetic Coding- Biggest to smallest  **MPEG-1** was optimized for CD-ROM and apps for 1.5 Mbps  **MPEG-2 adds to MPEG-1: More aspect ratios: 4:2:2, 4:4:4 ,Progressive and interlaced frame coding Four scalable modes:** spatial scalability, data partitioning, SNR scalability, temporal scalability **Lack support for content** manipulation (rm date stamp, off current score)**MPEG-4:**Support object-based features for content**,** Enable dynamic rendering of content, convergence among digital video, synthetic environments, and the Internet, media object, hierarchical representation(scene>person, furniture and 2d background)  **Content-based Interactivity**: Achieves different qualities for different objects with a fine granularity in spatial resolution, temporal resolution and decoding complexity  Needs coding of video objects with arbitrary shapes  **Data rate** = sampling rate \* quantization bits \* channels  **MPEG-1 Audio Encoding: Precision** 16 bits,  **Sampling frequency**: 32KHz, 44.1 KHz, 48 KHz **3 compression layers:** Layer 1, Layer 2, Layer 3 (MP3**)**  **Layer 3:** 32-320 kbps, target 64 kbps**, Layer 2:** 32-384 kbps, target 128 kbps**, Layer 1:** 32-448 kbps, target 192 kbps  **Filter bank** divides input into multiple sub-bands (32 equal frequency sub-bands)  Compresses by **removing acoustically irrelevant parts** of audio signals  Takes advantage of **human auditory systems** inability to hear quantization noise under **auditory masking**  **MPEG4-Audio:** More sample frequencies (8-96 kHz), Arbitrary bit rates and variable frame length, MDCT (modified discrete cosine transform),**P3** Lossy speech coding(code-excited linear prediction), Structured Audio (e.g., MIDI) |