Introduction to Image Processing

Prof. Alexandre Zaghetto
alexandre.zaghetto@mcgill.ca

McGill University
Department of Electrical and Computer Engineering

Topic 09
Video Coding

1. Introduction

- A change of scene
  - 2000
    - ✓ Most viewers receive analogue television via terrestrial, cable or satellite transmission.
    - ✓ VHS video tapes are the principal medium for recording and playing TV programs, movies, etc.
    - ✓ Cell phones are cell phones, i.e. a mobile handset can only be used to make calls or send SMS messages.
    - ✓ Internet connections are slow, primarily over telephone modems for home users.

1. Introduction

- A change of scene
  - 2000
    - ✓ Internet connections are slow, primarily over telephone modems for home users.
    - ✓ Web pages are web pages, with static text, graphics and photos and not much else.
    - ✓ Video calling requires dedicated video-conferencing terminals and expensive leased lines.
    - ✓ Video calling over the internet is possible but slow, unreliable and difficult to set up.

1. Introduction

- A change of scene
  - 2000
    - ✓ Video calling over the internet is possible but slow, unreliable and difficult to set up.
    - ✓ Consumer video cameras, camcorders, use tape media, principally analogue tape.
    - ✓ Home-made videos generally stay within the home.

1. Introduction

- A change of scene
  - 2010
    - ✓ Most viewers receive digital television via terrestrial, cable, satellite or internet
    - ✓ Greater choice of channels, electronic programme guides and high definition services.
    - ✓ Analogue TV has been switched off in many countries.
    - ✓ Many TV programmes can be watched via the internet.

1. Introduction

- A change of scene
  - 2010
    - ✓ DVDs are the principal medium for playing pre-recorded movies and TV programs.
    - ✓ Movie downloading, hard-disk recording and playback.
    - ✓ Variety of digital media formats.
    - ✓ High definition DVDs, Blu-Ray Disks, are increasing in popularity.

1. Introduction

- A change of scene
    - 2010
        - ✓ Cell phones function also as cameras.
        - ✓ Internet access speeds continue to get faster, enabling widespread use of video-based web applications.
        - ✓ Among other things web pages are movie players with content that changes dynamically.
        - ✓ Video calling over the internet is commonplace.
        - ✓ Consumer video cameras use hard disk or flash memory card media.

1. Introduction

- A change of scene
  - 2010
    - ✓ Editing, uploading and internet sharing of home videos is widespread.
    - ✓ A whole range of illegal activities has been born – DVD piracy, movie sharing via the internet etc.
    - ✓ Video footage of breaking news is more likely to come from a cell phone than a TV camera.

1. Introduction

- A change of scene

  - We will focus on one technical aspect that is key to the widespread adoption of digital video technology – **video compression**.

  - Digital video data tends to take up a **large amount of storage** or transmission capacity.

  - Video compression or video coding is the process of **reducing the amount of data** required to represent a digital video signal.

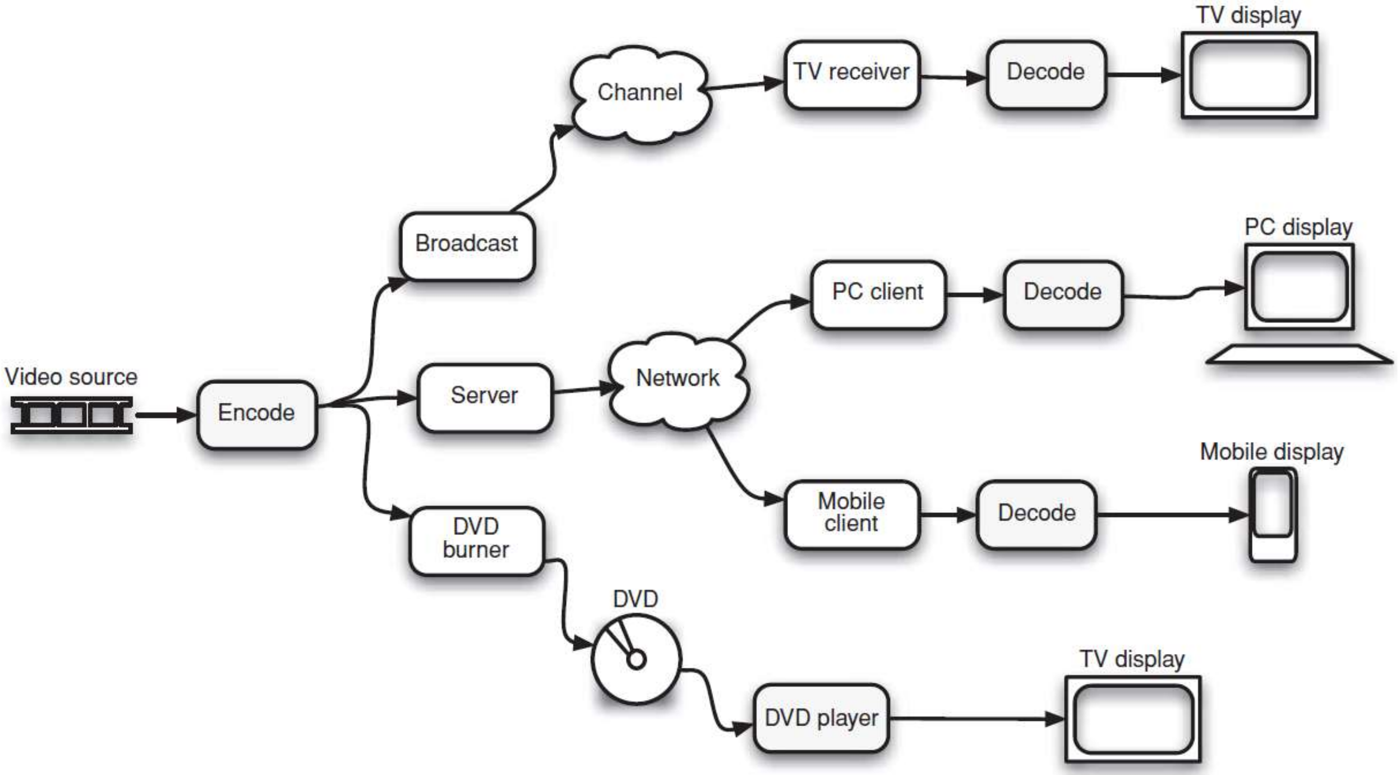
  - It is essential for any application in which storage capacity or transmission bandwidth is constrained.

1. Introduction

- A change of scene
  - Almost all consumer applications for digital video fall into this category:
    - One-way scenario
      - Digital television broadcasting
      - Internet video streaming
      - Mobile video streaming
      - DVD video
    - Two-way scenario
      - Video calling

1. Introduction

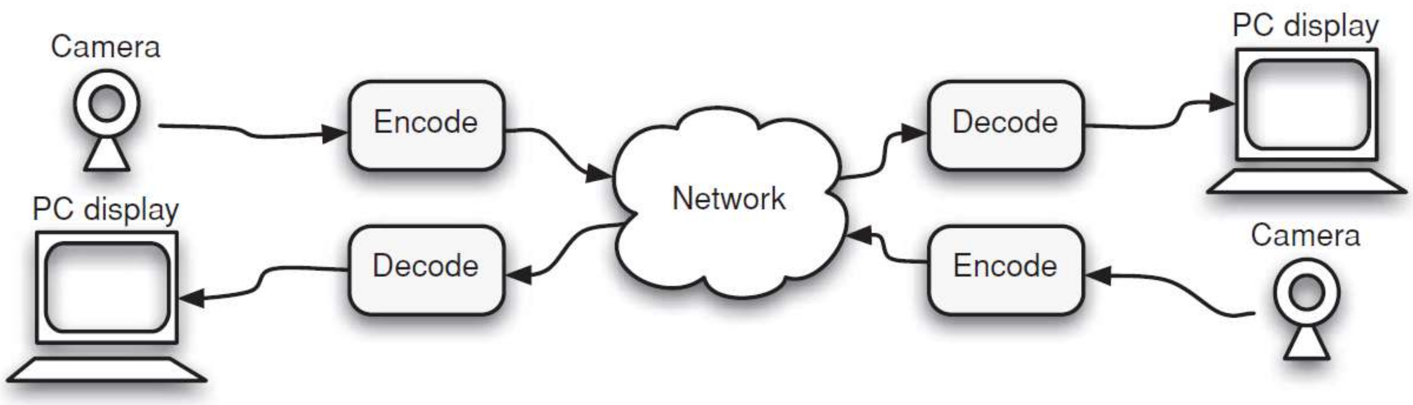
- A change of scene
  - Video coding scenarios, one-way



1. Introduction

- A change of scene

  - Video coding scenarios, one-way



1. Introduction

- Driving the change

  - The consumer applications discussed above represent very **large markets**.

  - A TV company that can pack a **larger number of high-quality** channels into the available bandwidth has a market edge over its competitors.

  - **Better video codec** results in **a better product** and therefore a more **competitive product**.

  - This drive to improve video compression technology has led to **significant investment** in video coding **research** and **development** over the last 15-20 years.

2. Video Formats and Quality

- Natural video scenes

  ➢ Characteristics of a typical natural video scene that are relevant for video processing and compression include:

    ✓ Spatial characteristics such as;
      - texture variation within scene;
      - number and shape of objects; and
      - colour.

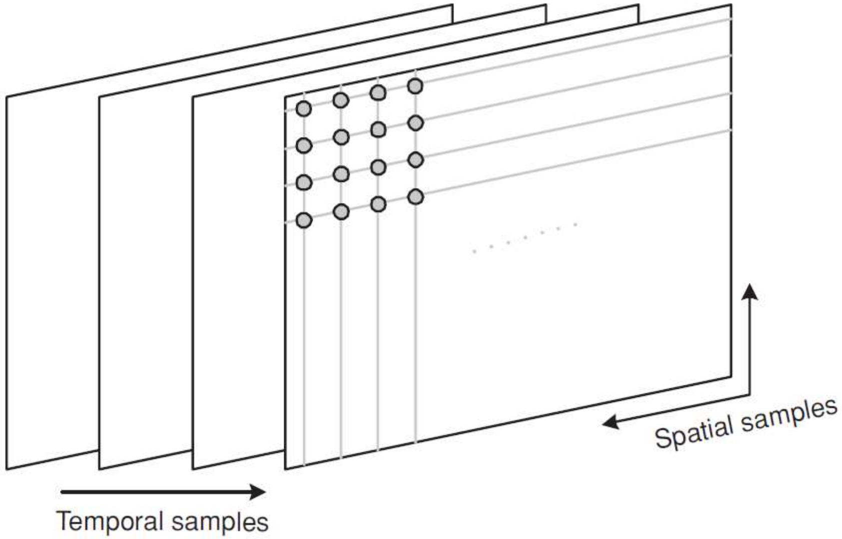
    ✓ Temporal characteristics such as
      - object motion;
      - changes in illumination; and
      - movement of the camera or viewpoint.

2. Video Formats and Quality

- Capture

  - A natural visual scene is **spatially** and **temporally** continuous.

  - Representing a visual scene in digital form involves **sampling** the real scene spatially and temporally.



2. Video Formats and Quality

- Capture

  - To obtain a 2-D sampled image, a camera focuses a 2-D projection of the video scene onto a sensor, such as an array of Charge Coupled Devices (CCDs).

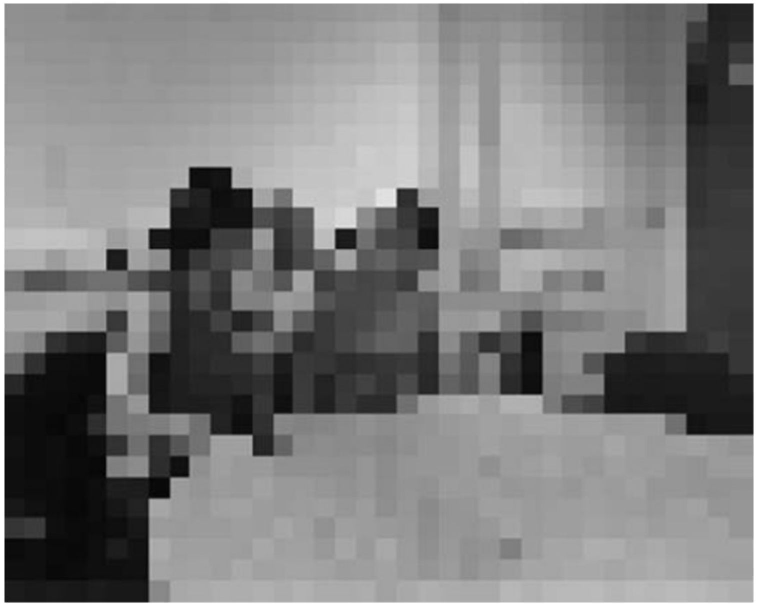
  - In the case of **colour image** capture, each colour component is separately **filtered** and projected onto a CCD array

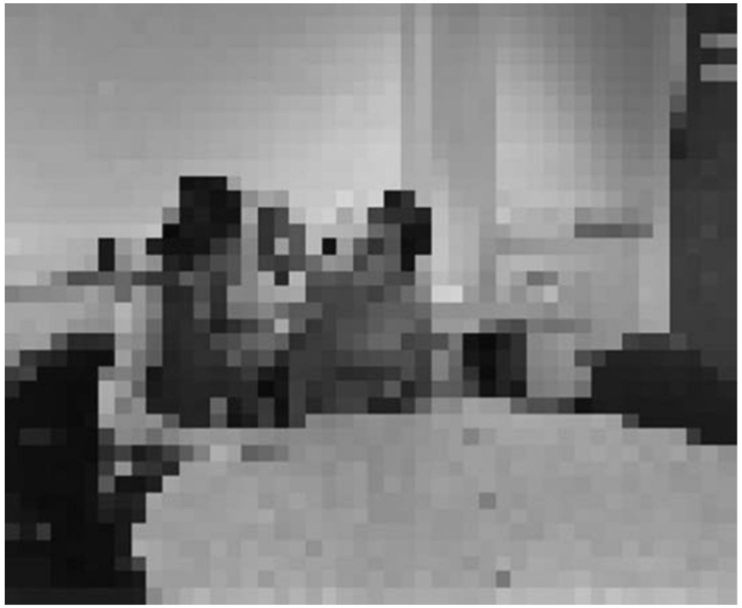
2. Video Formats and Quality

- Spatial sampling
  - The output of a CCD array is an analogue video signal, a varying electrical signal that represents a video image.
  - Sampling the image at a specific time produces a sampled **frame** that has defined values at a set of discrete positions.
  - The number of sampled positions influences the visual quality of the image.

2. Video Formats and Quality

- Spatial sampling





2. Video Formats and Quality

- Temporal sampling

  - A moving video image is formed by taking a rectangular **snapshot** of the signal at periodic time intervals.

  - Playing back the series of snapshots or frames produces the appearance of **motion**.

  - **Higher temporal sampling rate** or frame rate gives apparently **smoother motion** in the video scene

    - ✓ 10 frames/s ➔ very low bit-rate video
    - ✓ 10–20 frames/s ➔ low bit-rate video
    - ✓ 25 or 30 frames/s ➔ Standard Definition TV

2. Video Formats and Quality

- Temporal sampling

  - A moving video image is formed by taking a rectangular **snapshot** of the signal at periodic time intervals.

  - Playing back the series of snapshots or frames produces the appearance of **motion**.

  - **Higher temporal sampling rate** or frame rate gives apparently **smoother motion** in the video scene

    - ✓ 10 frames/s → very low bit-rate video
    - ✓ 10–20 frames/s → low bit-rate video
    - ✓ 25 or 30 frames/s → Standard Definition TV
    - ✓ 50 or 60 frames/s → very smooth motion

2. Video Formats and Quality

- Frames and fields

  - A video signal may be sampled as a series of complete frames (**progressive** sampling) or as a sequence of interlaced fields (**interlaced** sampling).

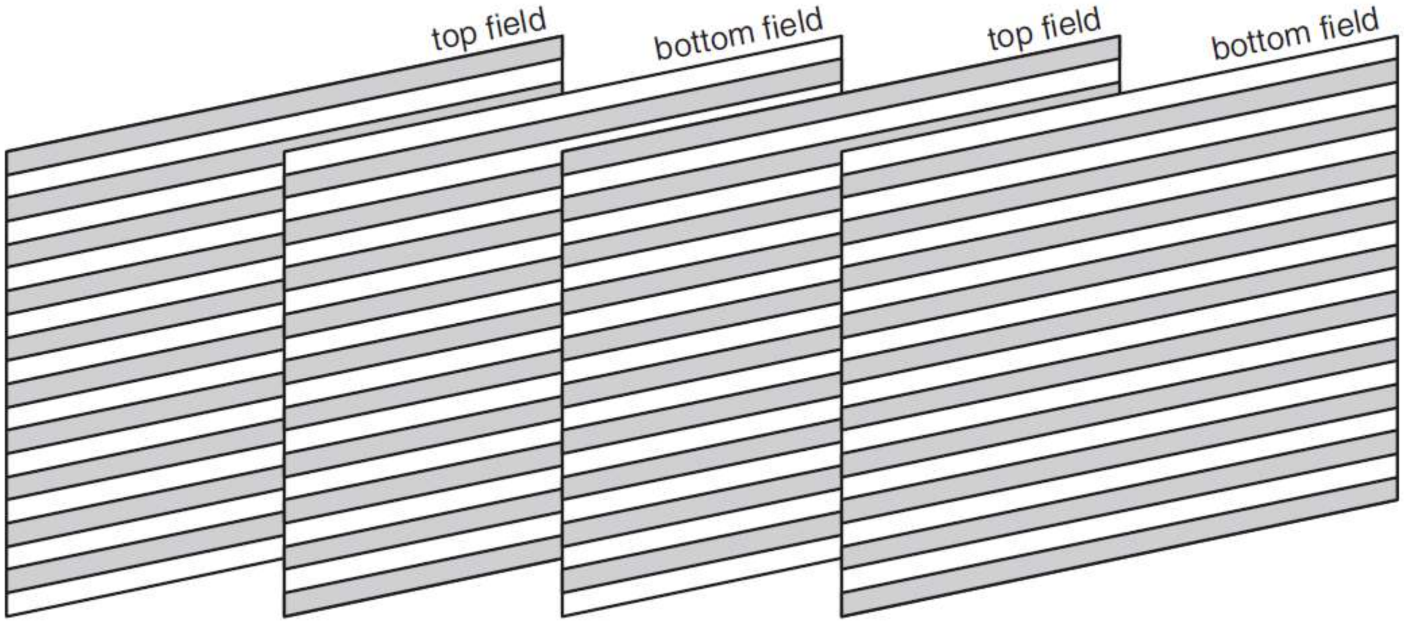
  - In an **interlaced** video sequence, **half** of the data in a frame (**one field**) is typically sampled **at each** temporal sampling **interval**.

  - A field may consist of either the odd-numbered or even-numbered lines within a complete video frame.

  - An interlaced video sequence typically contains a series of fields, each representing half of the information in a complete video frame

2. Video Formats and Quality

- Frames and fields
  - An interlaced video sequence typically contains a series of fields, each representing half of the information in a complete video frame



2. Video Formats and Quality

- Frames and fields
  - An interlaced video sequence typically contains a series of fields, each representing half of the information in a complete video frame

2. Video Formats and Quality

- Frames and fields

  - The advantage of this sampling method is that it is possible to send twice as many fields per second as the number of frames in an equivalent progressive sequence.

  - For example, a PAL video sequence consists of 50 fields/s.

    - When played back, motion **appears smoother** than in an equivalent **progressive** video sequence containing **25 frames** per second.

2. Video Formats and Quality

- Color spaces
  - RGB
    - ✓ A pixel is represented with three numbers that indicate the amount of Red, Green and Blue which define its color.



2. Video Formats and Quality

- Color spaces

  - YCrCb

    - ✓ The human visual system (HVS) is less sensitive to colour than to luminance.

    - ✓ In the RGB colour space the three colours are equally important and so are usually all stored at the same resolution.

    - ✓ It is possible to represent a colour image more efficiently by separating the luminance from the colour information and representing **luma** with a higher resolution than the chrominance or **chroma**.

2. Video Formats and Quality

- Color spaces
  - YCrCb

$$
\begin{aligned}
Y &= 0.299R + 0.587G + 0.114B \\
Cb &= 0.564(B - Y) \\
Cr &= 0.713(R - Y)
\end{aligned}
$$

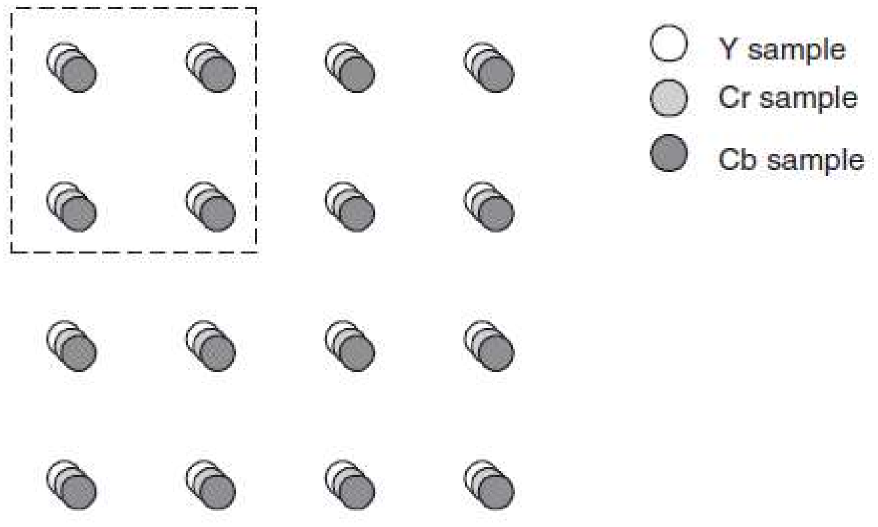
$$
\begin{aligned}
R &= Y + 1.402Cr \\
G &= Y - 0.344Cb - 0.714Cr \\
B &= Y + 1.772Cb
\end{aligned}
$$

2. Video Formats and Quality

- Color spaces

  - Examples of YCrCb sampling formats: 4:4:4

    - ✓ The three components have the same resolution.



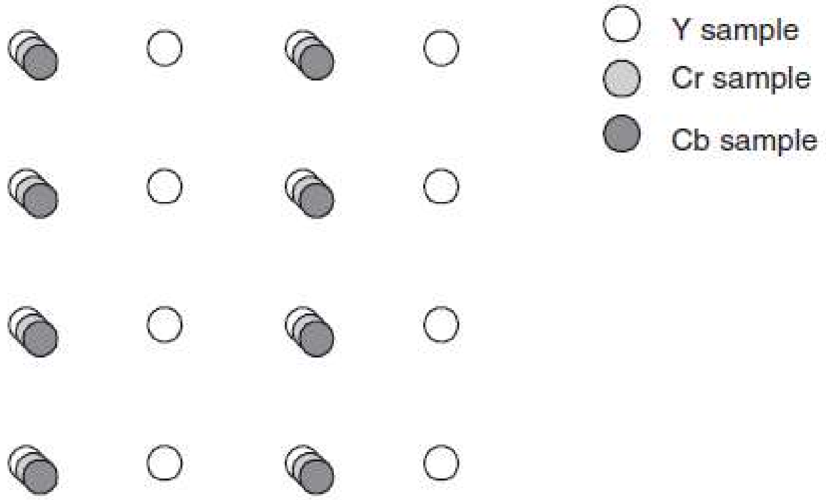
4:4:4 sampling

2. Video Formats and Quality

- Color spaces

  - ➢ Examples of YCrCb sampling formats: 4:2:2

    - ✓ The chrominance components have the same vertical resolution as the luma but half the horizontal resolution.



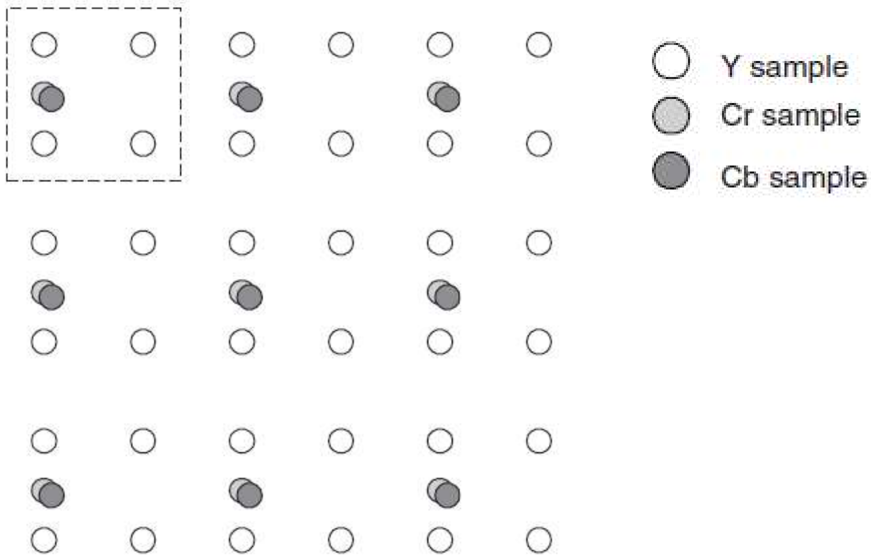
4:2:2 sampling

2. Video Formats and Quality

- Color spaces

  - Examples of YCrCb sampling formats: 4:2:0

    - C_b and C_r each have half the horizontal and vertical resolution.

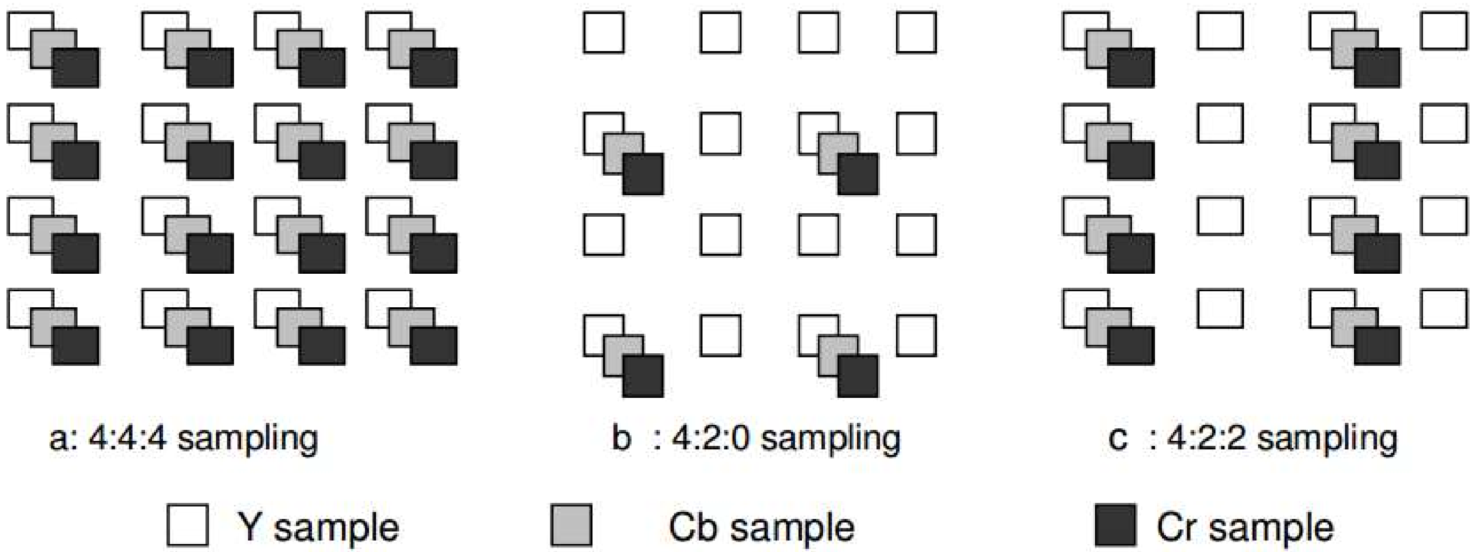


4:2:0 sampling

2. Video Formats and Quality

- Color spaces

  - Summary: 4:4:4 | 4:2:2 |4:2:0

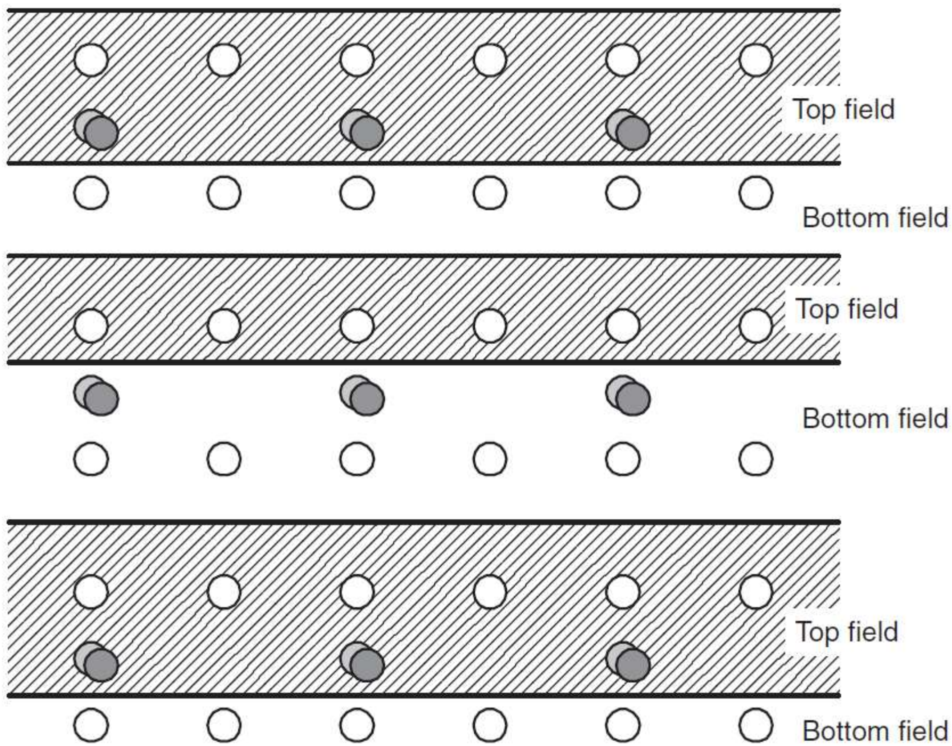


http://www.diva-portal.org/smash/get/diva2:831349/FULLTEXT01.pdf

2. Video Formats and Quality

- Formats

  - Allocation of 4:2:0 samples to top and bottom fields.



2. Video Formats and Quality

- Formats
  - Intermediate formats
    - The Common Intermediate Format (CIF) is the basis for a popular set of formats:

Format	Luminance resolution (horiz. × vert.)	Bits per frame (4:2:0, 8 bits per sample)
Sub-QCIF	128 × 96	147456
Quarter CIF (QCIF)	176 × 144	304128
CIF	352 × 288	1216512
4CIF	704 × 576	4866048

2. Video Formats and Quality

- Formats
  - Intermediate formats

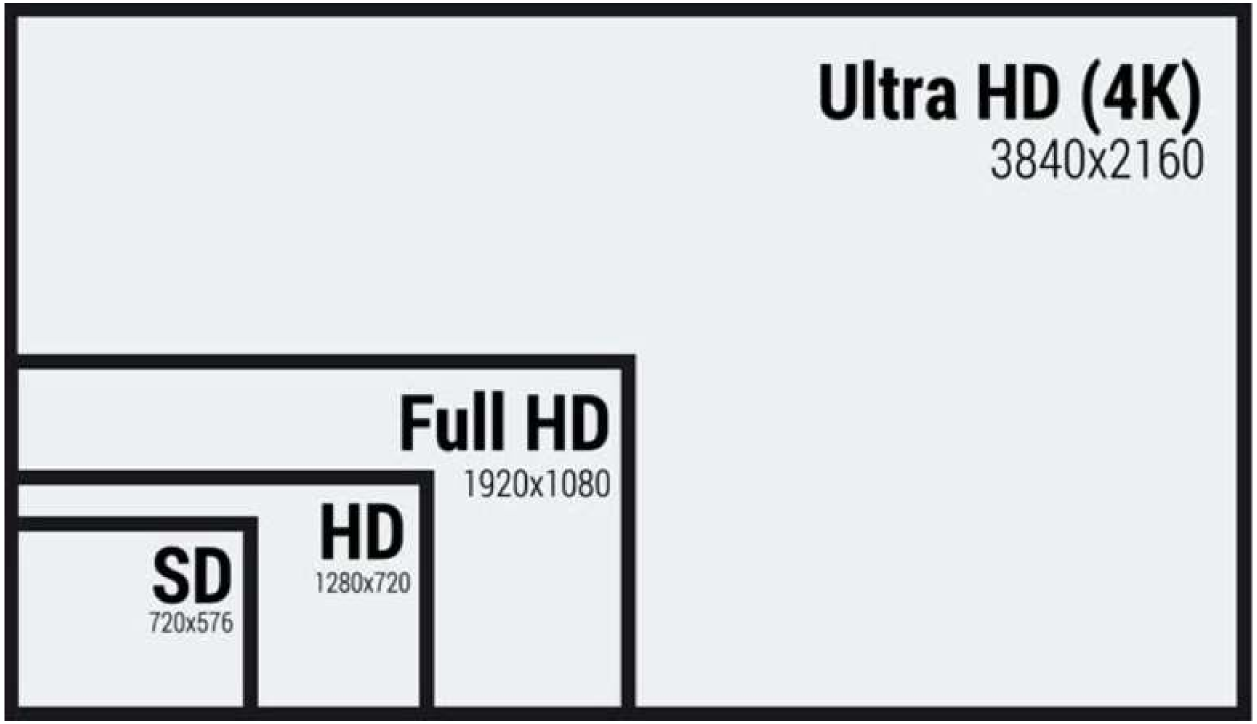
2. Video Formats and Quality

- Formats
  - Standard and High Definition

Format	Progressive or Interlaced	Horizontal pixels	Vertical pixels	Frames or fields per second
720p	Progressive	1280	720	25 frames
1080i	Interlaced	1920	1080	50 fields
1080p	Progressive	1920	1080	25 frames

2. Video Formats and Quality

- Formats
  - Standard and High Definition



2. Video Formats and Quality

- Quality
  - Subjective quality measurement
    - ✓ The perception of visual quality is influenced by:
      - **spatial fidelity,** i.e. how clearly parts of the scene can be seen, whether there is any obvious distortion; and
      - **temporal fidelity**, i.e. whether motion appears natural and ‘smooth’.

2. Video Formats and Quality

- Quality
  - Subjective quality measurement
    - ✓ However, a viewer's opinion of 'quality' is also affected by other factors such as:
      - the viewing environment;
      - the observer's state of mind; and
      - the extent to which the observer interacts with the visual scene.

2. Video Formats and Quality

- Quality
  - Subjective quality measurement
    - ✓ A widely-used quality test procedure is the Double Stimulus Continuous Quality Scale (DSCQS) method.
    - ✓ An assessor is presented with a pair of images or short video sequences A and B, one after the other, and is asked to give A and B a ‘quality score’ by marking on a continuous line with five intervals ranging from ‘Excellent’ to ‘Bad’.
    - ✓ Within each pair of sequences, one is an unimpaired ‘reference’ sequence and the other is the same sequence, modified by a system or process under test.

2. Video Formats and Quality

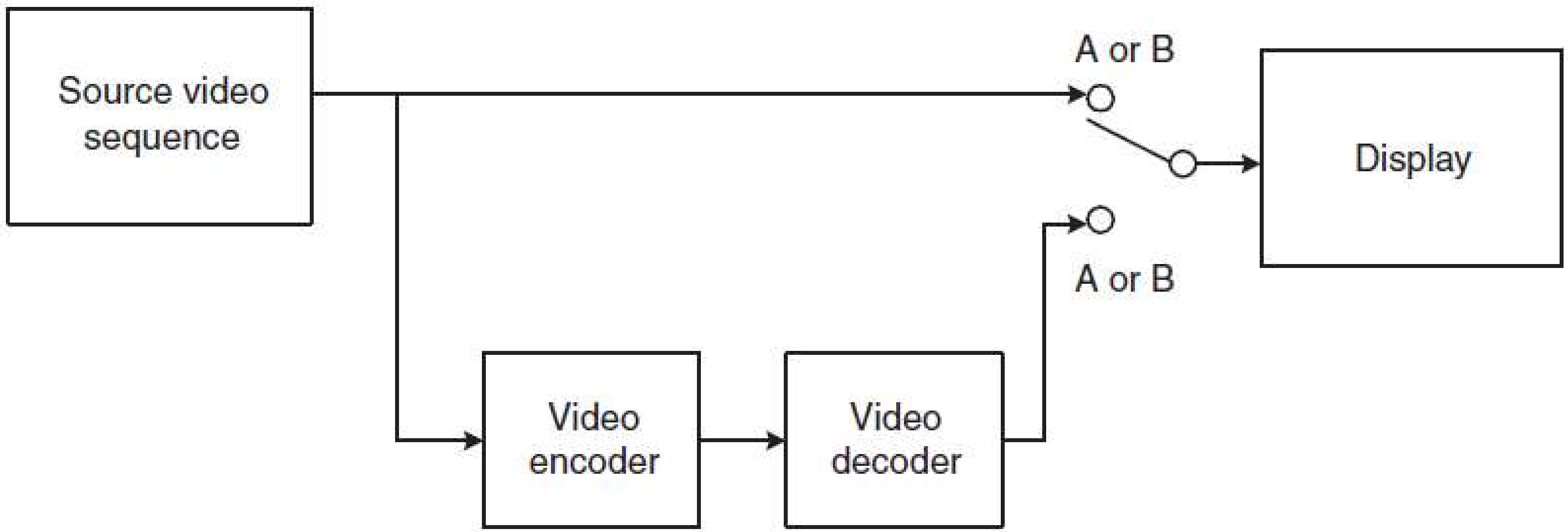
- Quality
  - Subjective quality measurement
    - The order of the two sequences, original and ‘impaired’, is randomized during the test session so that the assessor does not know which is the original and which is the impaired sequence.
    - At the end of the session, the scores are converted to a normalized range and the end result is a score, sometimes described as a ‘mean opinion score’ (MOS) that indicates the relative quality of the impaired and reference sequences.
    - Tests such as DSCQS are accepted as realistic measures of subjective visual quality.

2. Video Formats and Quality

- Quality
  - Subjective quality measurement
    - An 'expert' assessor who is familiar with the nature of video compression distortions or 'artefacts' may give a biased score and it is recommended to use 'non-expert' assessors.
    - This means that a large pool of assessors is required because a non-expert assessor will quickly learn to recognize characteristic artefacts in the video sequences and so will become 'expert'.
    - These factors make it expensive and time consuming to carry out the DSCQS tests thoroughly.

2. Video Formats and Quality

- Quality
  - Subjective quality measurement



2. Video Formats and Quality

- Quality
  - Subjective quality measurement

2. Video Formats and Quality

- Quality

  - Objective quality measurement

    - ✓ **PSNR** is a very popular quality measure, widely used to compare the 'quality' of compressed and decompressed video images.

$$PSNR_{dB} = 10 \log_{10} \frac{(2^n - 1)^2}{MSE}$$

2. Video Formats and Quality

- Quality
  - Objective quality measurement

Figure 2.16 PSNR examples: (a) Original; (b) 30.6dB; (c) 28.3dB

3. Video coding concepts

- Introduction

  - Compression involves a complementary pair of systems, a compressor (**encoder**) and a decompressor (**decoder**).

  - The **encoder** converts the **source** data into a compressed form occupying a **reduced** number of **bits**.

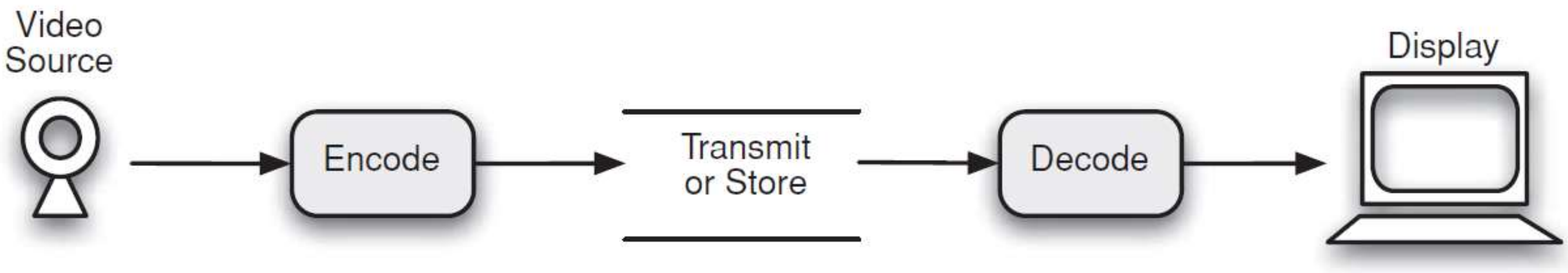
  - The **decoder** converts the **compressed form** back into a representation of the **original** video data.

  - The encoder/decoder pair is often described as a CODEC (enCOder/DECoder).

3. Video coding concepts

- Introduction

  - Compression involves a complementary pair of systems, a compressor (**encoder**) and a decompressor (**decoder**).



3. Video coding concepts

- Introduction

  - Data **compression** is achieved by **removing redundancy:** components that are not necessary for faithful reproduction of the data.

  - Many types of data contain **statistical redundancy** and can be effectively compressed using **lossless** compression.

  - The best that can be achieved with **lossless** image **compression** standards is a compression ratio of around **3–4 times**.

3. Video coding concepts

- Introduction

  - **Lossy** compression is necessary to achieve higher compression.

  - Lossy video compression systems are based on the principle of removing **subjective** redundancy.

  - Most video coding methods exploit both **temporal** and **spatial** redundancy to achieve compression.

  - In the **temporal** domain, there is usually a **high correlation** or similarity between frames of video that were **captured at around the same time**.

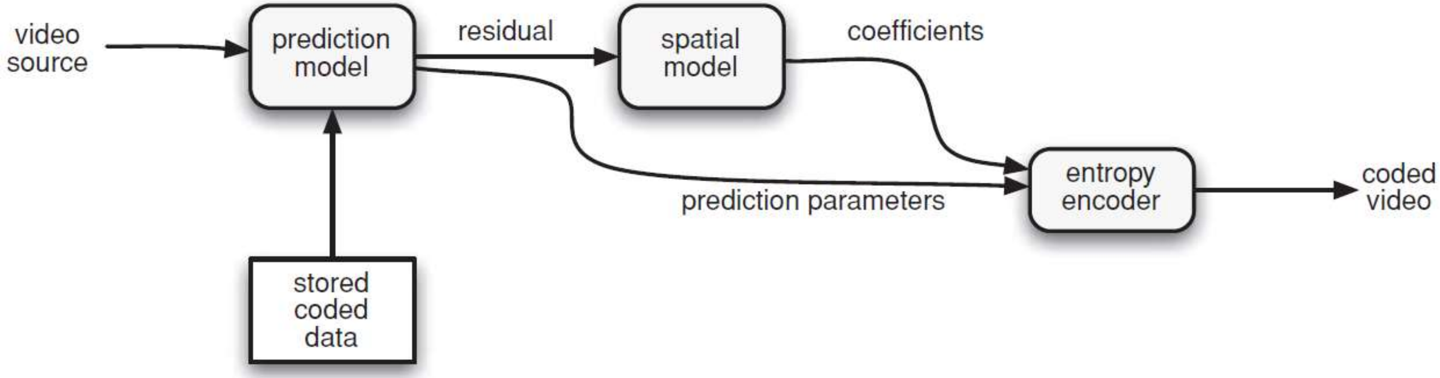
  - In the **spatial** domain, there is usually a **high correlation** between that are **close to each other**.

3. Video coding concepts

- Introduction

3. Video coding concepts

- Video CODEC
  - A video encoder consists of three main functional units:
    - ✓ a prediction model
    - ✓ a spatial model; and
    - ✓ an entropy encoder.



3. Video coding concepts

- Video CODEC
  - Prediction model
    - ✓ Attempts to **reduce redundancy** by exploiting the similarities between neighbouring video frames and/or neighbouring image samples.
    - ✓ The prediction is formed from data in **current, previous** and/or **future** frames.
    - ✓ Created by **spatial extrapolation** from neighbouring image samples, **intra prediction**.
    - ✓ Or by compensating for differences between the frames, **inter**/motion compensated **prediction**.

3. Video coding concepts

- Video CODEC
  - Prediction model
    - The output of the prediction model is a **residual frame**, created by
      - **subtracting** the **prediction** from the **actual** current frame; and
      - a set of **model parameters** indicating
        - the **intra prediction** type; or
        - Describing how the **motion** was **compensated**

3. Video coding concepts

- Video CODEC
  - Prediction model
    - The output of the prediction model is a **residual frame**, created by
      - **subtracting** the **prediction** from the **actual** current frame; and
      - a set of **model parameters** indicating
        - the **intra prediction** type; or
        - Describing how the **motion** was **compensated**

3. Video coding concepts

- Video CODEC
  - Spatial model
    - ✓ The **residual frame** forms the **input** to the **spatial model** which makes use of similarities between local samples in the residual frame to reduce **spatial redundancy**.
    - ✓ Carried out by applying a **transform to the residual samples** and **quantizing** the results.
    - ✓ The output of the spatial model is a set of **quantized transform coefficients**.

3. Video coding concepts

- Video CODEC
  - Entropy encoder
    - ✓ The parameters of the prediction model, i.e. **intra** and **inter** prediction mode(s), **motion vectors**, and the **coefficients**, are compressed by the **entropy encoder**.
    - ✓ It removes **statistical redundancy** in the data.
    - ✓ Produces a **compressed bit stream** or a file that may be transmitted and/or stored.
    - ✓ A **compressed sequence** consists of coded prediction parameters, coded residual coefficients and header information

3. Video coding concepts

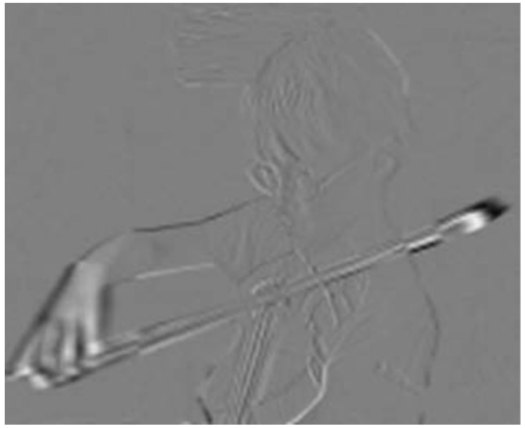
- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ The predicted frame is created from one or more past or future frames known as **reference frames**.
    - ✓ The simplest method of temporal prediction is to use the **previous frame** as the **predictor** for the **current frame**.

3. Video coding concepts

- Prediction Model

  - ➢ Temporal prediction (inter prediction)

    - ✓ Residual formed by subtracting the predictor (frame 1) from the current frame (frame 2).

    - ✓ A lot of energy remains in the residual frame

3. Video coding concepts

- Prediction Model

  - Temporal prediction (inter prediction)

    - ✓ Causes of **changes** between video frames include **motion**, **uncovered regions** and **lighting** changes.

    - ✓ It is possible to estimate the **trajectory** of each pixel between successive video frames, producing a field of trajectories (**optical flow**).

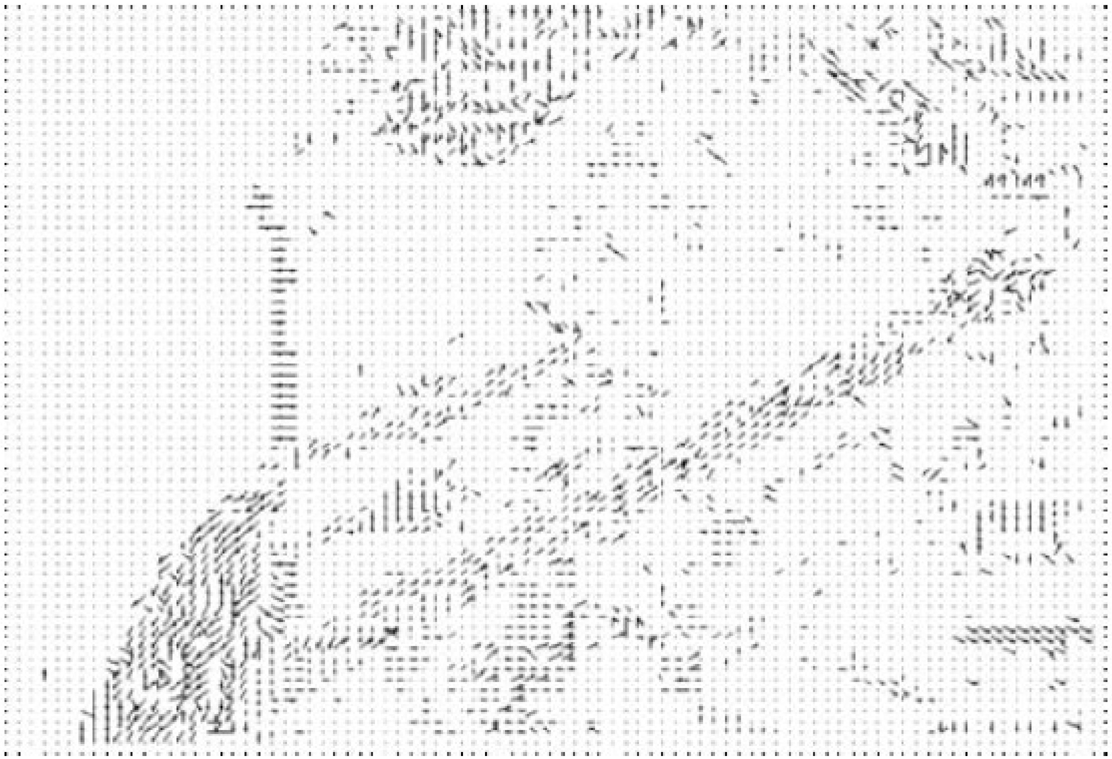
    - ✓ this is not a **practical** method of motion **compensation** for several reasons.

      - An accurate calculation of optical flow is very computationally intensive

3. Video coding concepts

- Prediction Model

  - Temporal prediction (inter prediction)



3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ A practical and widely used method of motion compensation is to compensate for **movement** of '**blocks**' of the current frame.
    - ✓ The following procedure is carried out for each block of MxN samples in the current frame.
      1. Search an area in the reference frame to find a similar MxN-sample region. This process of finding the best match is known as **motion estimation**

3. Video coding concepts

- Prediction Model

  - Temporal prediction (inter prediction)

2. Search an area in the reference frame to find a similar MxN-sample region. This process is known as **motion estimation.**

3. The chosen candidate region becomes the predictor for the current MxN block and is subtracted from the current block to form a **residual (motion compensation)**.

4. The **residual block** and the offset between the current block and the position of the candidate region (**motion vector**) are also encoded.

3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ The decoder uses the received motion vector to re-create the predictor region.
    - ✓ It decodes the residual block, adds it to the predictor and reconstructs a version of the original block.
    - ✓ **Block-based** motion compensation **fits well** with rectangular video frames and with block-based image transforms such as the **Discrete Cosine Transform**.

3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ The **macroblock**, corresponding to a **16 × 16-pixel region** of a frame, is the basic unit for motion compensated prediction in a number of important visual coding standards including **MPEG-2** and **H.264/AVC**.

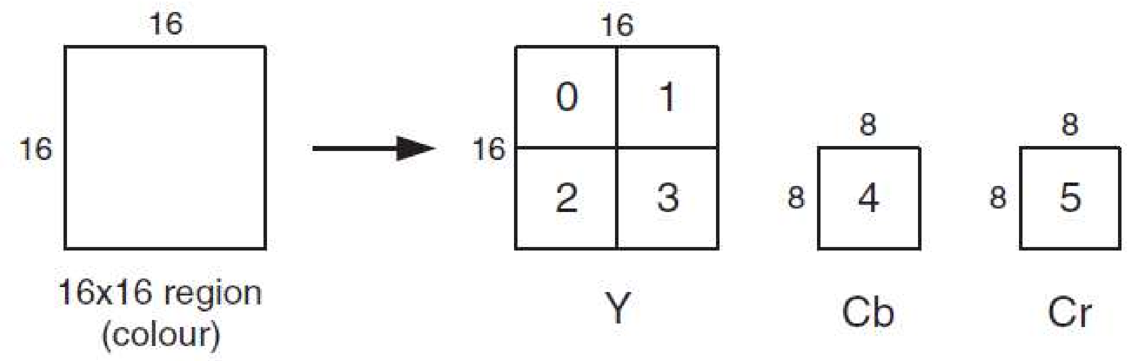
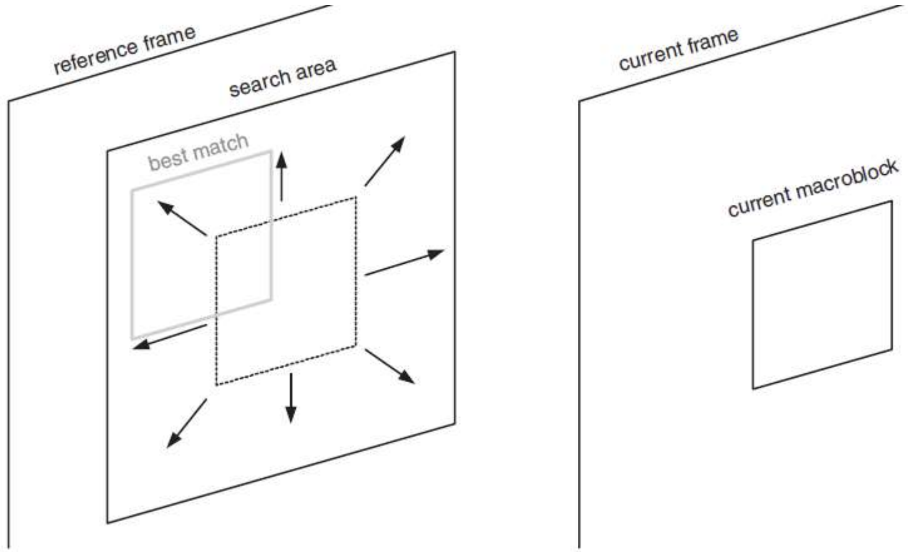


Figure 3.8 Macroblock (4:2:0)

3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ Motion estimation of a macroblock involves a 16 region in a **reference frame** that closely matches the **current macroblock**.



3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ Motion compensation (and block size):

3. Video coding concepts

- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ Motion compensation: no compensation

3. Video coding concepts

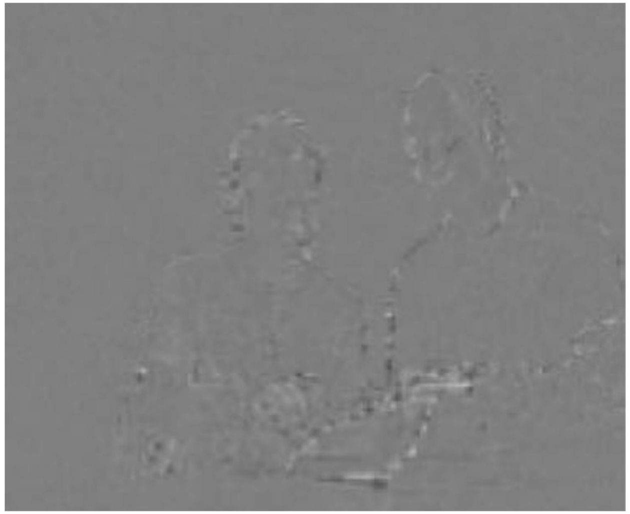
- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ Motion compensation: 16 x 16 pixels

3. Video coding concepts

- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ Motion compensation: 8 x 8 pixels

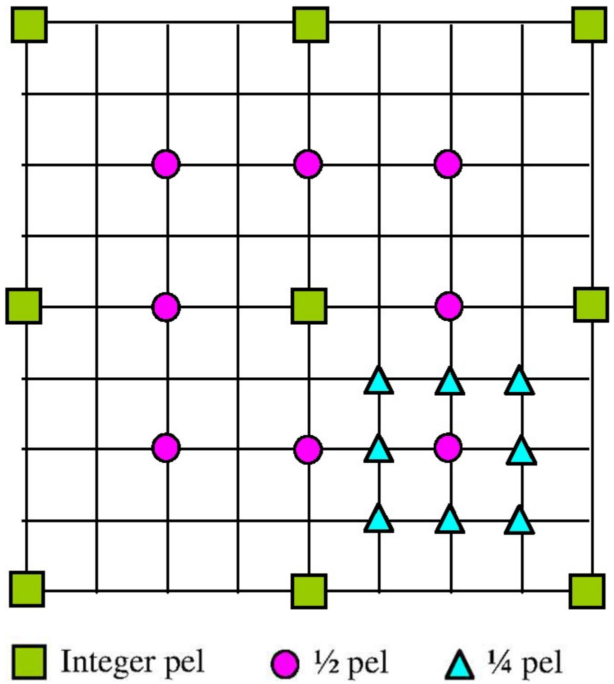
3. Video coding concepts

- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ Motion compensation: 4 x 4 pixels



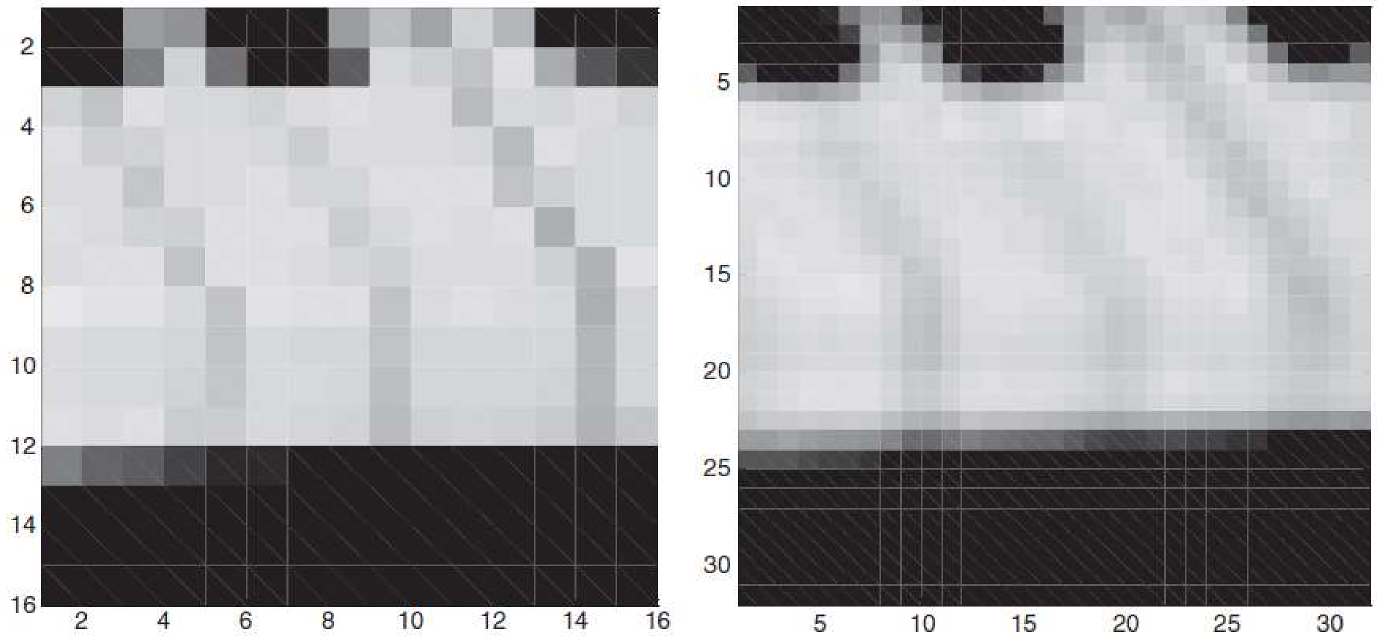
3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ Sub-pixel motion compensation



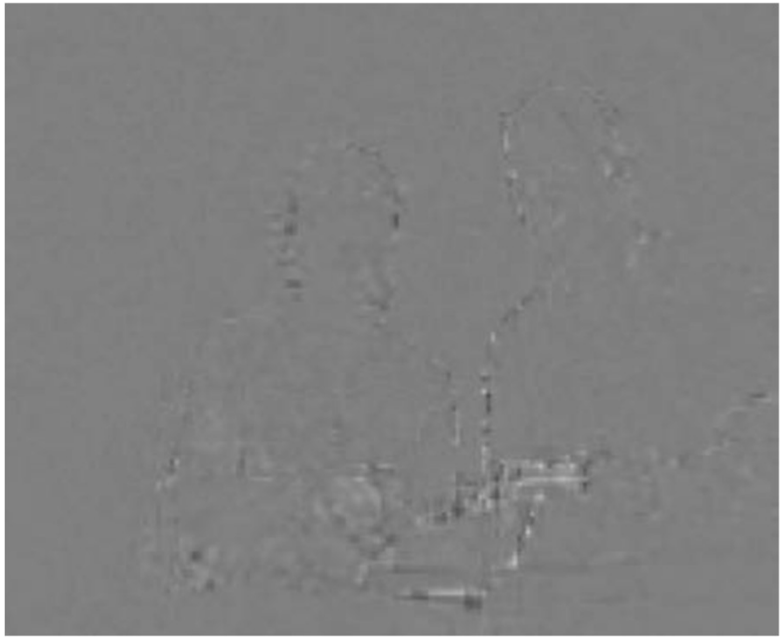
3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ Sub-pixel motion compensation: ½-pixel



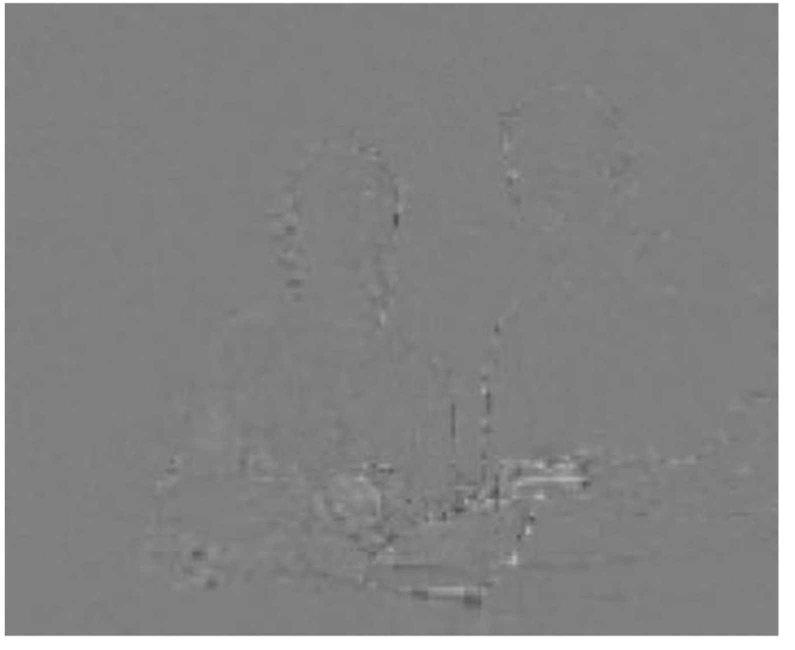
3. Video coding concepts

- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ Sub-pixel motion compensation: 4 x 4, ½-pixel



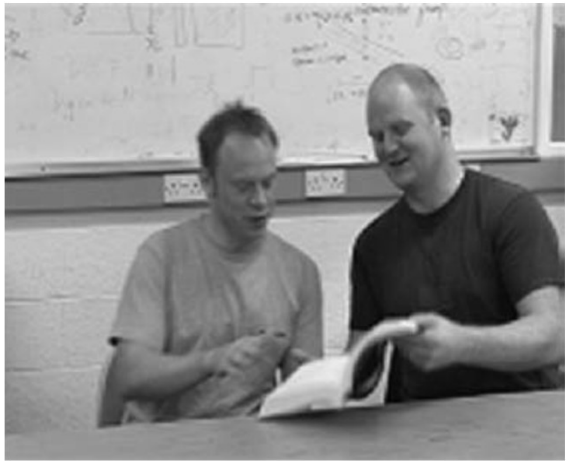
3. Video coding concepts

- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ Sub-pixel motion compensation: 4 x 4, ¼-pixel



3. Video coding concepts

- Prediction Model
  - ➢ Temporal prediction (inter prediction)
    - ✓ In addition to the extra complexity, sub-pixel motion compensation implies coding penalty.



3. Video coding concepts

- Prediction Model
  - Temporal prediction (inter prediction)
    - ✓ In addition to the extra complexity, sub-pixel motion compensation implies coding penalty.

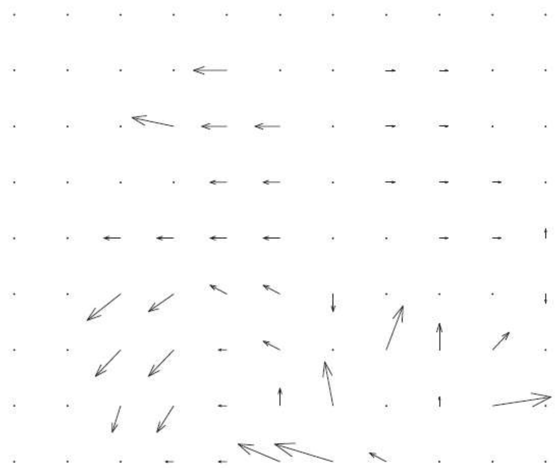


Figure 3.21 Motion vector map : 16 × 16 blocks, integer vectors

3. Video coding concepts

- Prediction Model

  - Temporal prediction (inter prediction)

    - ✓ In addition to the extra complexity, sub-pixel motion compensation implies coding penalty.

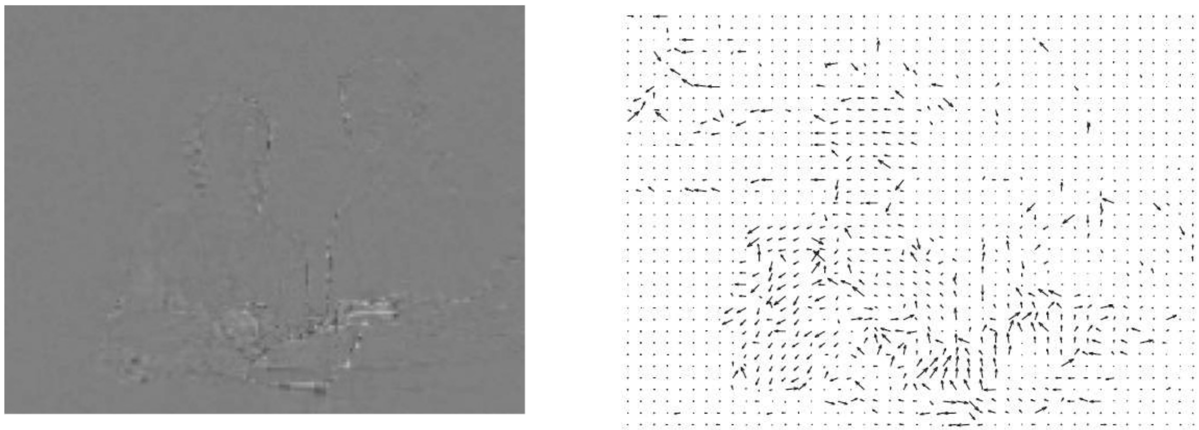
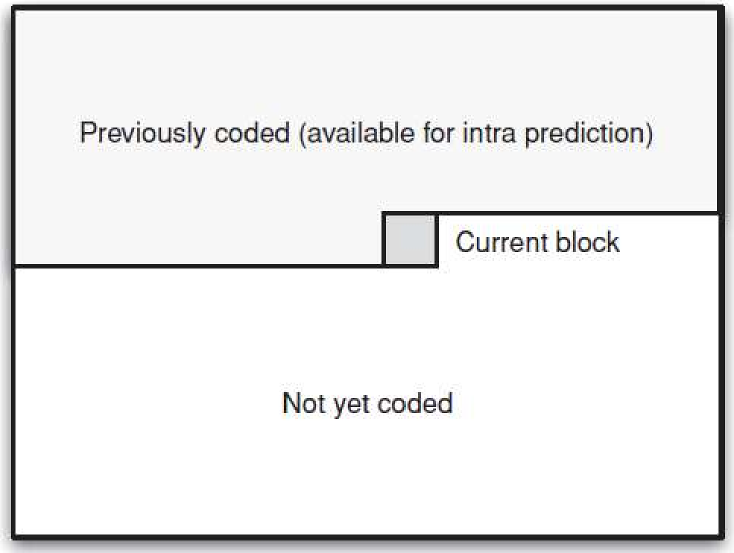


Figure 3.22 Motion vector map : 4 × 4 blocks, 1/4-pixel vectors

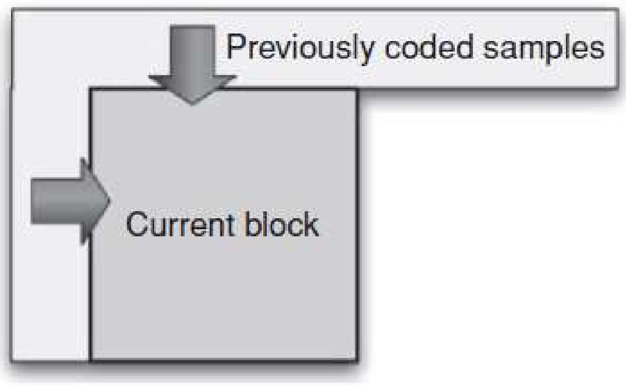
3. Video coding concepts

- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ The prediction for the current block of image samples is created from previously-coded samples in the same frame.



3. Video coding concepts

- Prediction Model
  - Spatial prediction (intra prediction)
    - ✓ Many different approaches to intra prediction have been proposed. H.264/AVC uses spatial extrapolation to create an intra prediction for a block or macroblock.

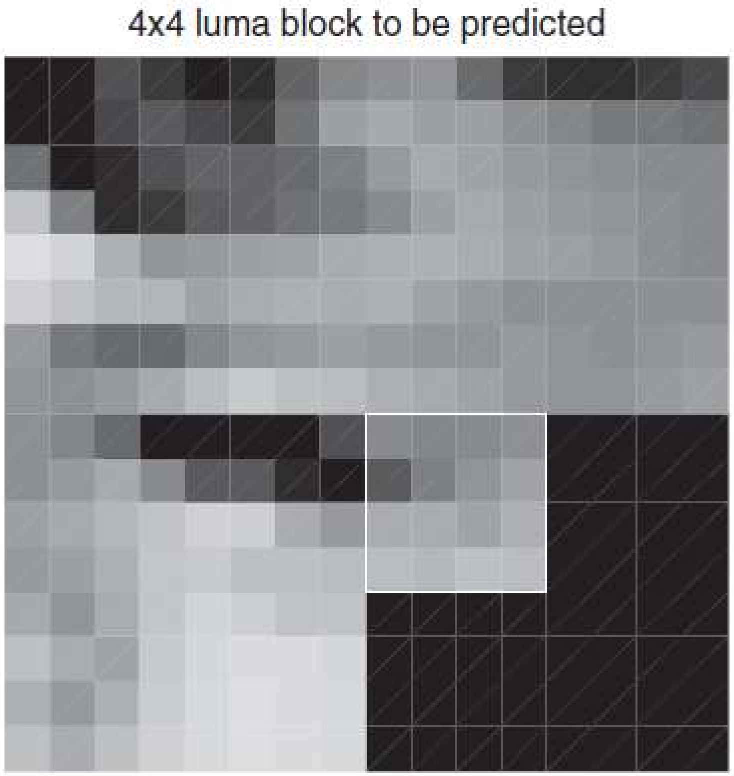


3. Video coding concepts

- Prediction Model
  - Spatial prediction (intra prediction)
    - ✓ One or more prediction(s) are formed by **extrapolating** samples from the top and/or left sides of the current block.
    - ✓ In general, the **nearest samples** are most likely to be **highly correlated** with the samples in the current block and so only the pixels along the top and/or left edges are used to create the prediction block.
    - ✓ Once the **prediction** has been generated, it is **subtracted** from the **current block** to form a residual in a similar way to inter prediction.

3. Video coding concepts

- Prediction Model
  - Spatial prediction (intra prediction)



3. Video coding concepts

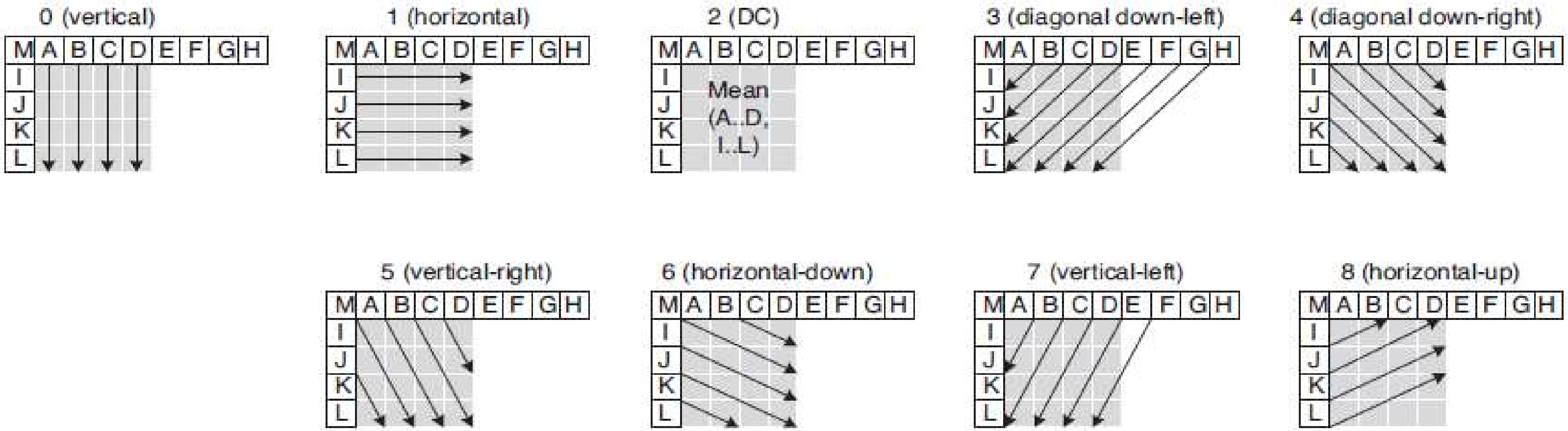
- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ Example: H.264 4x4 luma prediction modes

M	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				

Mode 0 (Vertical)	The upper samples A,B,C,D are extrapolated vertically.
Mode 1 (Horizontal)	The left samples I,J,K,L are extrapolated horizontally.
Mode 2 (DC)	All samples in P are predicted by the mean of samples A..D and I..L.
Mode 3 (Diagonal Down-Left)	The samples are interpolated at a 45° angle between lower-left and upper-right.
Mode 4 (Diagonal Down-Right)	The samples are extrapolated at a 45° angle down and to the right.
Mode 5 (Vertical-Left)	Extrapolation at an angle of approximately 26.6° to the left of vertical, i.e. width/height = $^1/_2$.
Mode 6 (Horizontal-Down)	Extrapolation at an angle of approximately 26.6° below horizontal.
Mode 7 (Vertical-Right)	Extrapolation or interpolation at an angle of approximately 26.6° to the right of vertical.
Mode 8 (Horizontal-Up)	Interpolation at an angle of approximately 26.6° above horizontal.

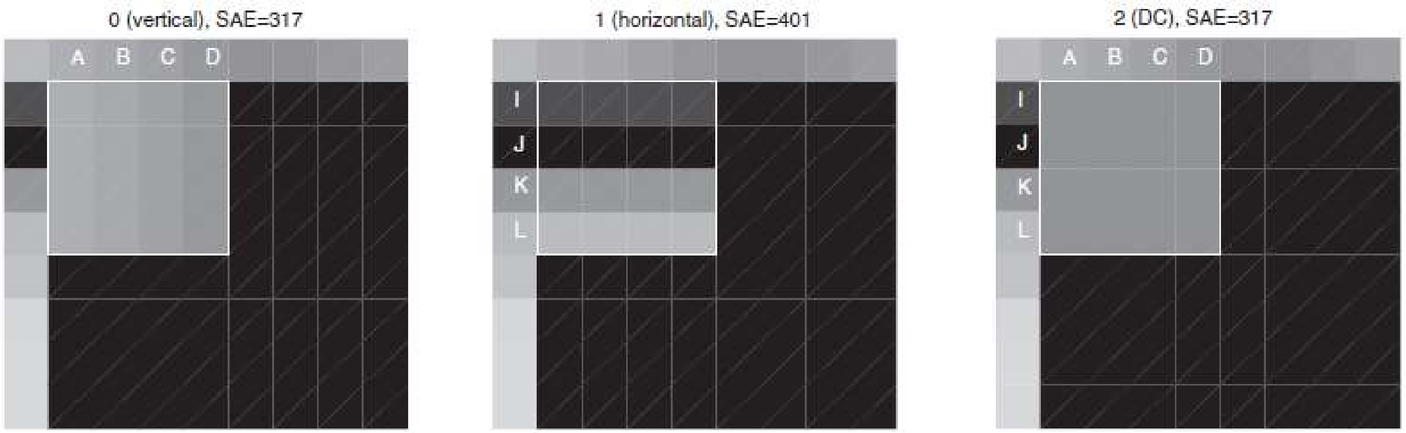
3. Video coding concepts

- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ Example: H.264 4x4 luma prediction modes



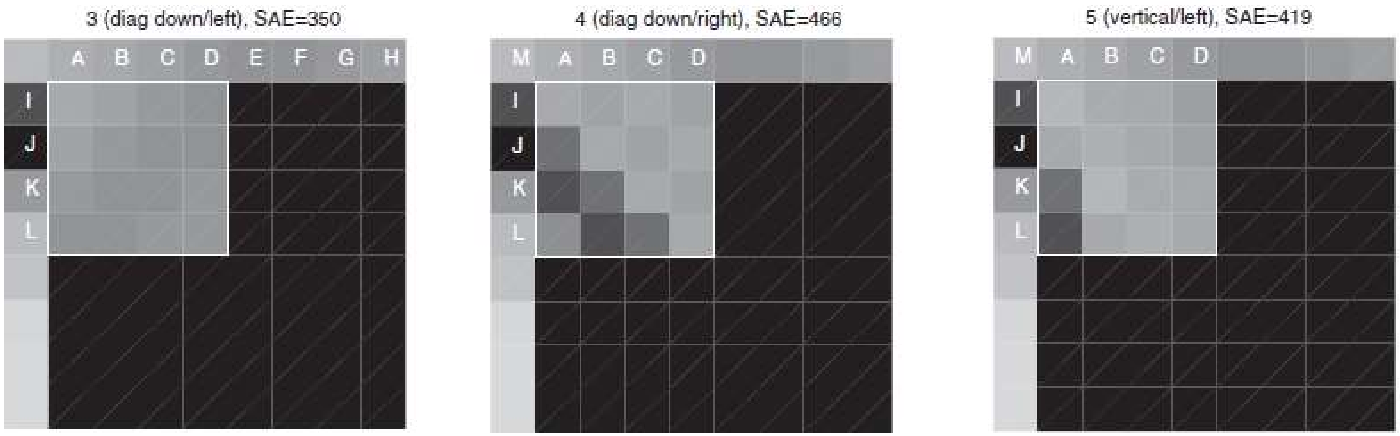
3. Video coding concepts

- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ Example: H.264 4x4 luma prediction modes



3. Video coding concepts

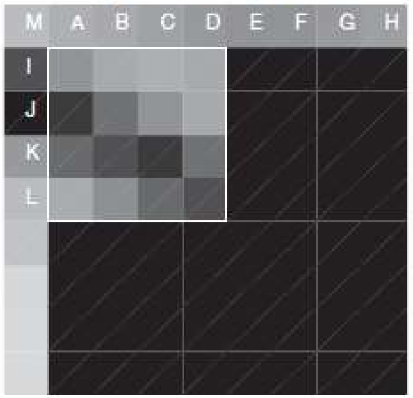
- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ Example: H.264 4x4 luma prediction modes



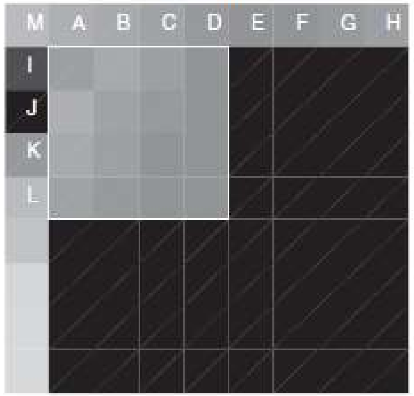
3. Video coding concepts

- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ Example: H.264 4x4 luma prediction modes

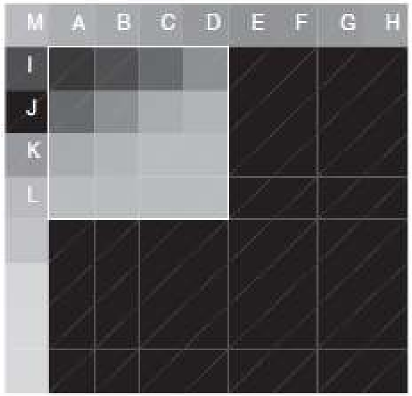
6 (horizontal/down, SAE=530)



7 (vertical/right), SAE=351

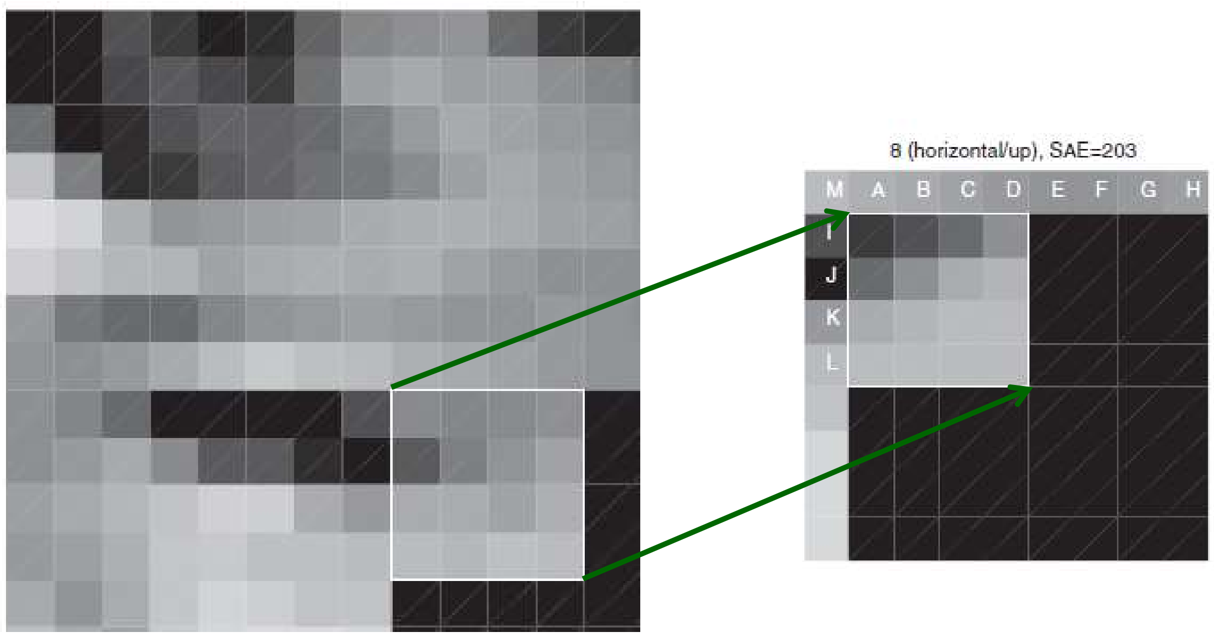


8 (horizontal/up), SAE=203



3. Video coding concepts

- Prediction Model
  - ➢ Spatial prediction (intra prediction)
    - ✓ Example: H.264 4x4 luma prediction modes



3. Video coding concepts

- Spatial model

  - The function of the spatial model is to **further decorrelate** image or residual data and to convert it into a form that can be efficiently compressed using an **entropy coder**.

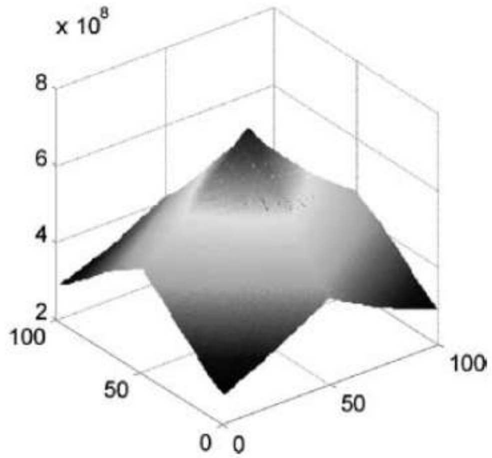


Figure 3.25 2D autocorrelation function of image

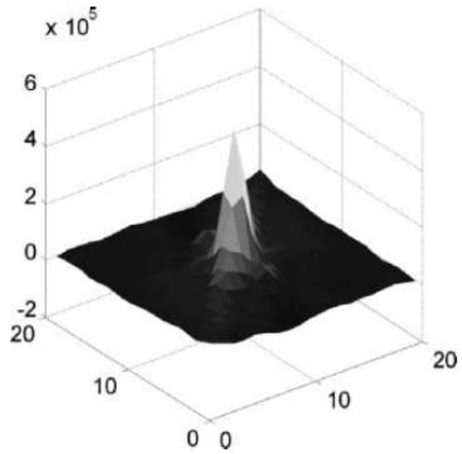


Figure 3.26 2D autocorrelation function of residual

3. Video coding concepts

- Spatial model

    - If the **prediction** is **successful**, the energy in the residual is **lower** than in the original frame and the residual can be represented with **fewer** bits.

    - The function of the **spatial model** is to:

        - ✓ **Further decorrelate** image or residual data; and

        - ✓ Convert it into a form that can be efficiently **compressed** using an **entropy coder.**

3. Video coding concepts

- Spatial model
  - Practical image models typically have three main components:
    - ✓ **Transformation** to **decorrelate** and compact the data;
    - ✓ **Quantization** to **reduce** the **precision** of the transformed data; and
    - ✓ **Reordering** to arrange the data to **group** together **significant values**.

3. Video coding concepts

- Entropy coder

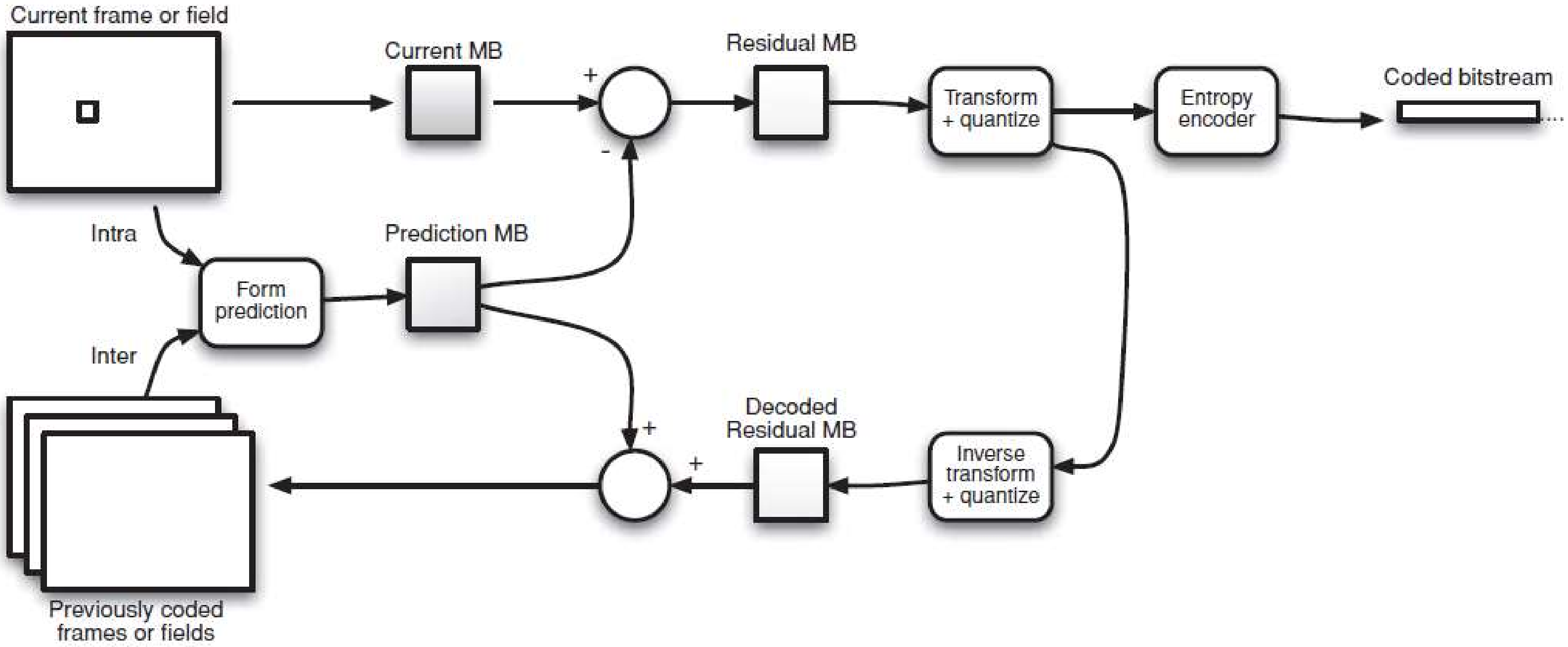
  - The **entropy encod**er converts a series of symbols representing elements of the video sequence into a **compressed bitstream** suitable for transmission or storage.

3. Video coding concepts

- Hybrid DPCM/DCT video CODEC model

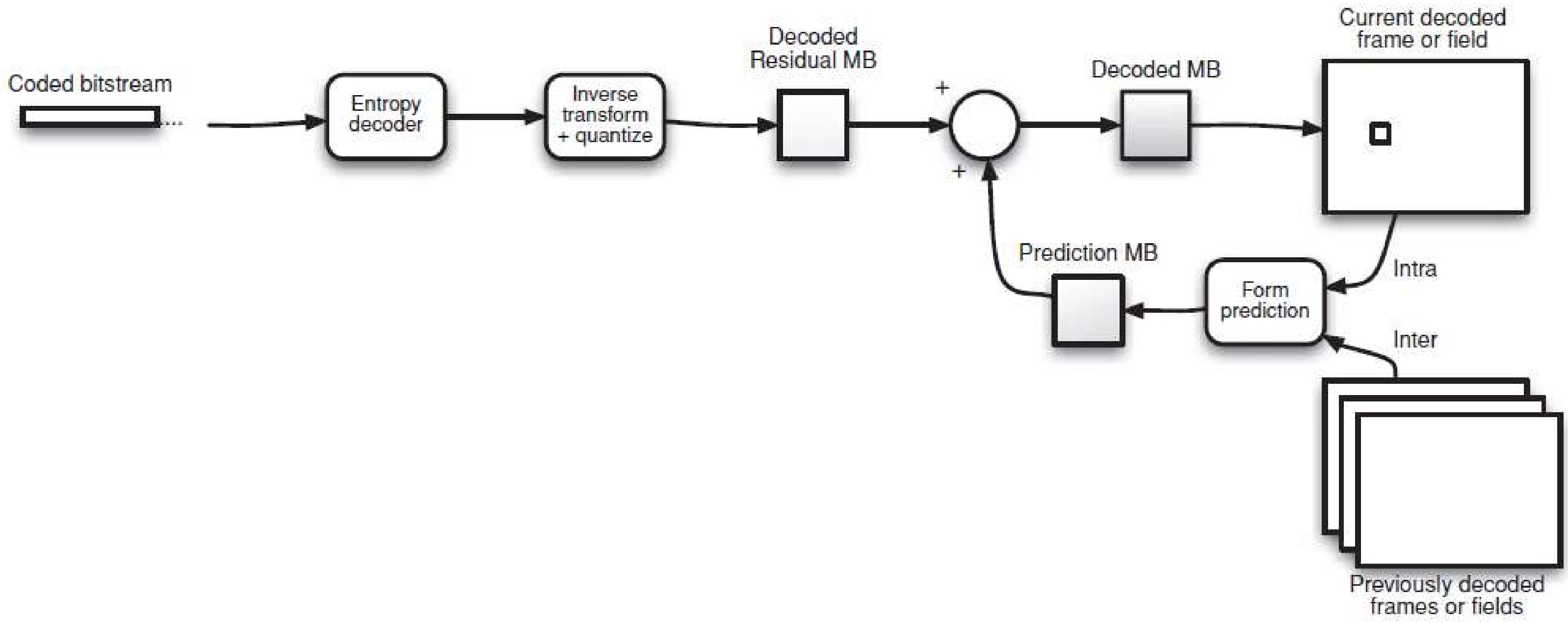
  ➢ Example

    ✓ H.264/AVC encoder



3. Video coding concepts

- Hybrid DPCM/DCT video CODEC model
  - Example
    - ✓ H.264/AVC decoder

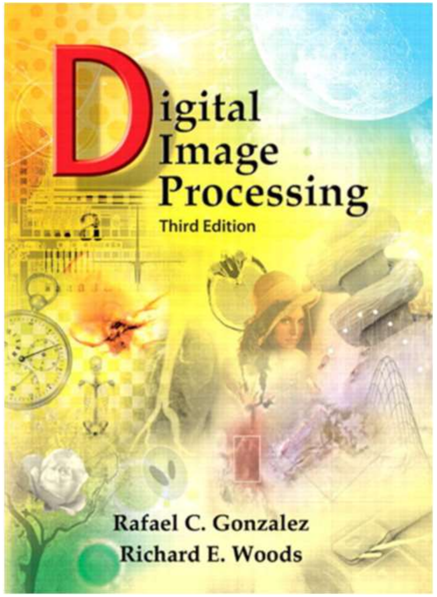


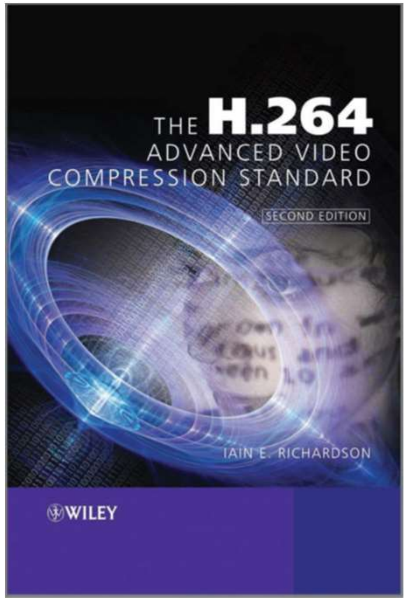
3. Video coding concepts

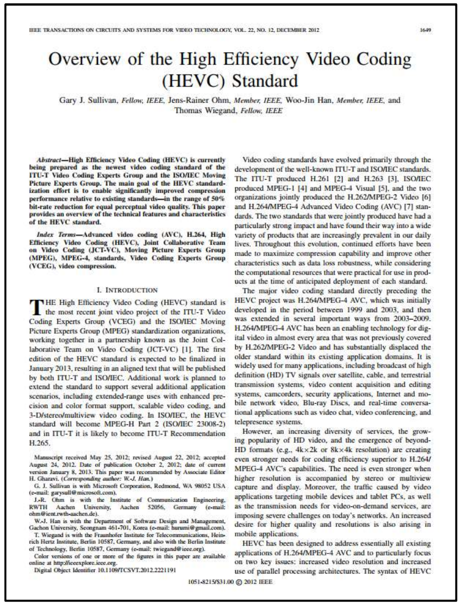
- The video coding tools described, namely:
  - motion compensated prediction;
  - intra-frame prediction;
  - transform coding;
  - quantization; and
  - entropy coding

form the basis of the reliable and effective coding model that has dominated the field of video compression for over 20 years.

4. Further Reading







IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 22, NO. 12, DECEMBER 2012 1649

Overview of the High Efficiency Video Coding (HEVC) Standard

Gary J. Sullivan, *Fellow, IEEE*, Jens-Rainer Ohm, *Member, IEEE*, Woo-Jin Han, *Member, IEEE*, and Thomas Wiegand, *Fellow, IEEE*

***Abstract*—High Efficiency Video Coding (HEVC) is currently being prepared as the newest video coding standard of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group. The main goal of the HEVC standardization effort is to enable significantly improved compression performance relative to existing standards—in the range of 50% bit-rate reduction for equal perceptual video quality. This paper provides an overview of the technical features and characteristics of the HEVC standard.**

***Index Terms*—Advanced video coding (AVC), H.264, High Efficiency Video Coding (HEVC), Joint Collaborative Team on Video Coding (JCT-VC), Moving Picture Experts Group (MPEG), MPEG-4, standards, Video Coding Experts Group (VCEG), video compression.**

I. Introduction

THE High Efficiency Video Coding (HEVC) standard is the most recent joint video project of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) standardization organizations, working together in a partnership known as the Joint Collaborative Team on Video Coding (JCT-VC) [1]. The first edition of the HEVC standard is expected to be finalized in January 2013, resulting in an aligned text that will be published by both ITU-T and ISO/IEC. Additional work is planned to extend the standard to support several additional application scenarios, including extended-range uses with enhanced precision and color format support, scalable video coding, and 3-D/stereo/multiview video coding. In ISO/IEC, the HEVC standard will become MPEG-H Part 2 (ISO/IEC 23008-2) and in ITU-T it is likely to become ITU-T Recommendation H.265.

Manuscript received May 25, 2012; revised August 22, 2012; accepted August 24, 2012. Date of publication October 2, 2012; date of current version January 8, 2013. This paper was recommended by Associate Editor H. Gharavi. (*Corresponding author: W.-J. Han.*)

G. J. Sullivan is with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: garysull@microsoft.com).

J.-R. Ohm is with the Institute of Communication Engineering, RWTH Aachen University, Aachen 52056, Germany (e-mail: ohm@ient.rwth-aachen.de).

W.-J. Han is with the Department of Software Design and Management, Gachon University, Seongnam 461-701, Korea (e-mail: hurumi@gmail.com).

T. Wiegand is with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin 10587, Germany, and also with the Berlin Institute of Technology, Berlin 10587, Germany (e-mail: twiegand@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2012.2221191

Video coding standards have evolved primarily through the development of the well-known ITU-T and ISO/IEC standards. The ITU-T produced H.261 [2] and H.263 [3], ISO/IEC produced MPEG-1 [4] and MPEG-4 Visual [5], and the two organizations jointly produced the H.262/MPEG-2 Video [6] and H.264/MPEG-4 Advanced Video Coding (AVC) [7] standards. The two standards that were jointly produced have had a particularly strong impact and have found their way into a wide variety of products that are increasingly prevalent in our daily lives. Throughout this evolution, continued efforts have been made to maximize compression capability and improve other characteristics such as data loss robustness, while considering the computational resources that were practical for use in products at the time of anticipated deployment of each standard.

The major video coding standard directly preceding the HEVC project was H.264/MPEG-4 AVC, which was initially developed in the period between 1999 and 2003, and then was extended in several important ways from 2003–2009. H.264/MPEG-4 AVC has been an enabling technology for digital video in almost every area that was not previously covered by H.262/MPEG-2 Video and has substantially displaced the older standard within its existing application domains. It is widely used for many applications, including broadcast of high definition (HD) TV signals over satellite, cable, and terrestrial transmission systems, video content acquisition and editing systems, camcorders, security applications, Internet and mobile network video, Blu-ray Discs, and real-time conversational applications such as video chat, video conferencing, and telepresence systems.

However, an increasing diversity of services, the growing popularity of HD video, and the emergence of beyond-HD formats (e.g., 4k×2k or 8k×4k resolution) are creating even stronger needs for coding efficiency superior to H.264/MPEG-4 AVC's capabilities. The need is even stronger when higher resolution is accompanied by stereo or multiview capture and display. Moreover, the traffic caused by video applications targeting mobile devices and tablet PCs, as well as the transmission needs for video-on-demand services, are imposing severe challenges on today's networks. An increased desire for higher quality and resolutions is also arising in mobile applications.

HEVC has been designed to address essentially all existing applications of H.264/MPEG-4 AVC and to particularly focus on two key issues: increased video resolution and increased use of parallel processing architectures. The syntax of HEVC

1051-8215/$31.00 © 2012 IEEE