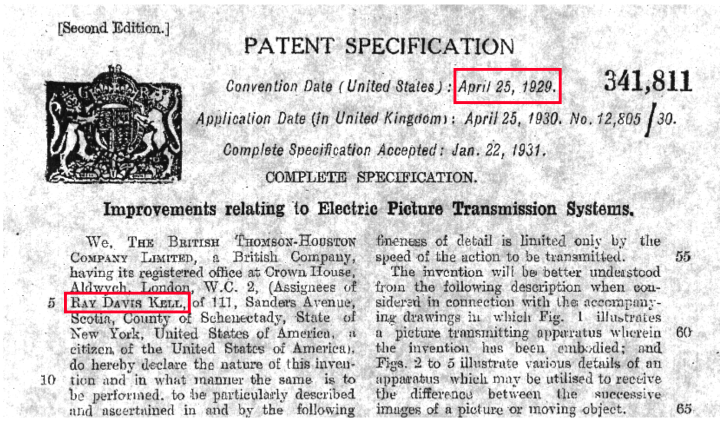
Hybrid Video Coding

- Principle of Hybrid Video Coding
- Motion-Compensated Prediction
- Bit Allocation
- Motion Models
- Motion Estimation
- Efficiency of Hybrid Video Coding



[Second Edition.]

PATENT SPECIFICATION

Convention Date (United States): April 25, 1929. 341,811

Application Date (in United Kingdom): April 25, 1930. No. 12,805 / 30.

Complete Specification Accepted: Jan. 22, 1931.

COMPLETE SPECIFICATION.

Improvements relating to Electric Picture Transmission Systems.

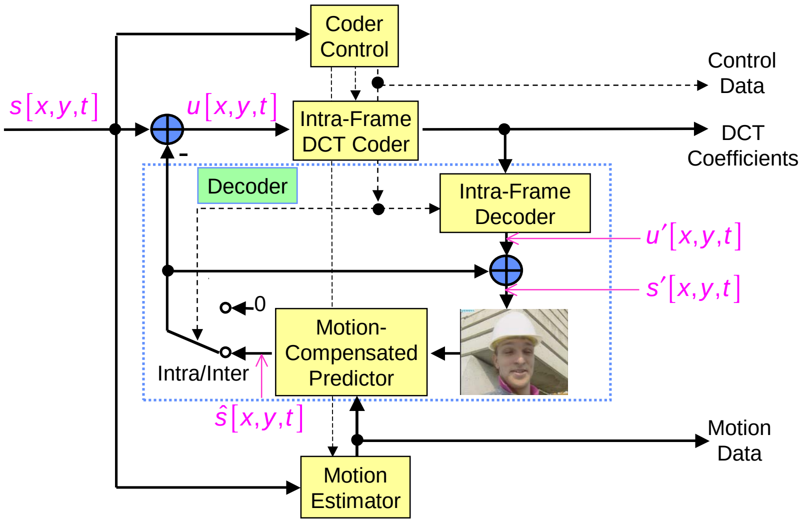
We, THE BRITISH THOMSON-HOUSTON COMPANY LIMITED, a British Company, having its registered office at Crown House, Aldwych, London, W.C. 2, (Assignees of RAY DAVIS KELL, of 111, Sanders Avenue, Scotia, County of Schenectady, State of New York, United States of America, a citizen of the United States of America), do hereby declare the nature of this invention and in what manner the same is to be performed, to be particularly described and ascertained in and by the following

fineness of detail is limited only by the speed of the action to be transmitted.

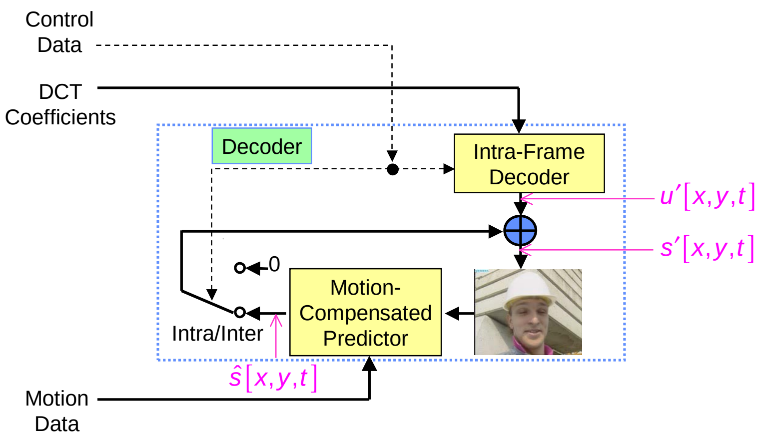
The invention will be better understood from the following description when considered in connection with the accompanying drawings in which Fig. 1 illustrates a picture transmitting apparatus wherein the invention has been embodied; and Figs. 2 to 5 illustrate various details of an apparatus which may be utilised to receive the difference between the successive images of a picture or moving object.

"It has been customary in the past to transmit successive complete images of the transmitted picture."
[...]
"In accordance with this invention, this difficulty is avoided by transmitting only the difference between successive images of the object."

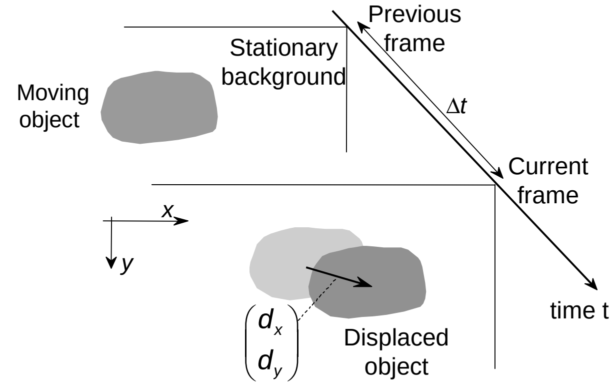
Hybrid Video Encoder



Hybrid Video Decoder



Motion-Compensated Prediction



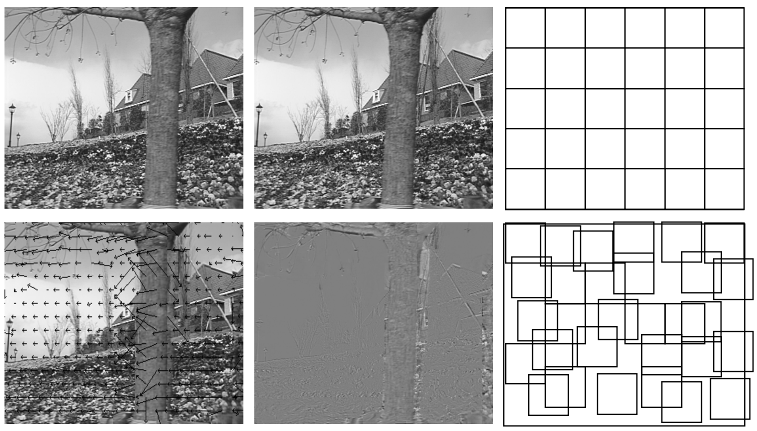
Prediction for the luminance signal $s[x,y,t]$ within the moving object:

$$\boxed{\hat{s}\left[x,y,t\right] = s'(x-d_x, y-d_y, t-\Delta t)}$$

from: Girod

Motion-Compensated Prediction: Example

Frame 1 $s[x,y,t-1]$ (previous) Frame 2 $s[x,y,t]$ (current) Partition of frame 2 into blocks (schematic)

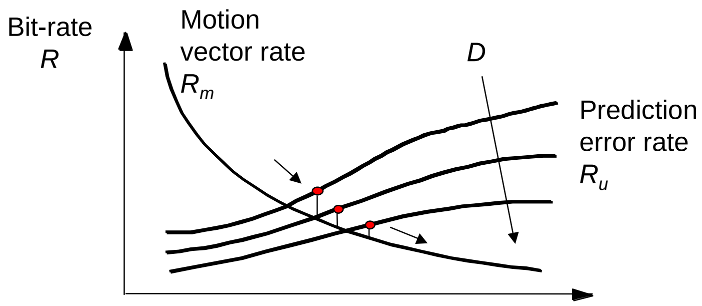


Frame 2 with displacement vectors

Difference between motion-compensated prediction and current frame $u[x,y,t]$

Referenced blocks in frame 1

Rate-constrained Motion Estimation



- Optimum trade-off: $\boxed{\frac{dD}{dR_m} = \frac{dD}{dR_u}}$
- Displacement error variance can be influenced via
  - Block-size, quantization of motion parameters
  - Choice of motion model

from: Girod

Rate-Constrained Coder Control

- Efficiency increase via adding coding modes

What part of the image should be coded using what method ?

- Problem can be posed as: minimize distortion *D* subject to a rate constraint R_c

$$\boxed{\min \{ D \} \text{ s.t. } R<R_c}$$

- Coder Control using Lagrangian optimization:

Solve an unconstrained minimization problem

$$\boxed{\min \{ D + \lambda R \}}$$

[Everett III 1963], [Shoham, Gersho 1988], [Chou, Lookabaugh, Gray 1989]

Lagrangian Optimization in Video Coding

- A number of interactions are often neglected
  - Temporal dependency due DPCM loop
  - Spatial dependency of coding decisions
  - Conditional entropy coding
- *Rate-Constrained Motion Estimation* [Sullivan, Baker 1991]:

$$\min \{D_m + \lambda\, R_m\}$$

Distortion after motion compensation | Lagrange parameter | Number of bits for motion vector

- *Rate-Constrained Mode Decision* [Wiegand, et al. 1996]:

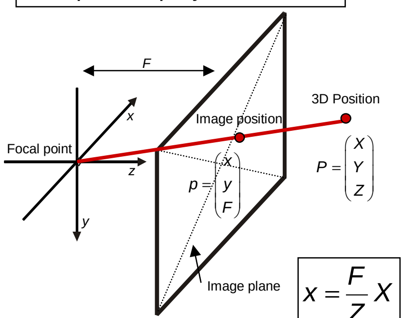
$$\min \{D + \lambda\, R\}$$

Distortion after reconstruction | Lagrange parameter | Number of bits for coding mode

Camera Models

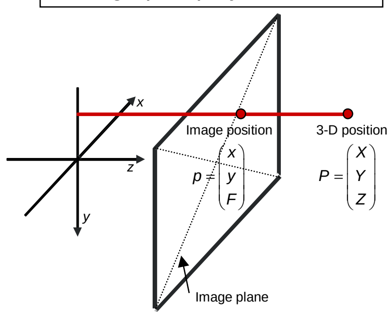
- Natural camera-view scenes are projected from the 3-D world into the 2-D image plane
- A camera model can be used to describe projection, lens and sampling

Perspective projection model



$$x = \frac{F}{Z} X$$
$$y = \frac{F}{Z} Y$$

Orthographic projection model



Can be used when F«Z for all considered points

$$x = X$$
$$y = Y$$

Motion Models

- Motion in 3-D space corresponds to displacements in the image plane
- Motion compensation in the image plane is conducted to provide a prediction signal for efficient video compression
- Efficient motion-compensated prediction often uses side information to transmit the displacements
- Displacements must be efficiently represented for video compression
- Motion models relate 3-D motion to displacements assuming reasonable restrictions of the motion and objects in the 3-D world

Motion Model

$d_x = x' - x = f_x(\mathbf{a}, x, y), \quad d_y = y' - y = f_y(\mathbf{b}, x, y)$

x, y : location in previous image

x', y' : location in current image

$\mathbf{a}, \mathbf{b}$: vector of motion coefficients

d_x, d_y : displacements

Perspective Motion Model

- Mathematical model:

$$d_x = \frac{a_1 + a_2 x + a_3 y}{1 + c_1 x + c_2 y}, \quad d_y = \frac{b_1 + b_2 x + b_3 y}{1 + c_1 x + c_2 y}$$

- Restrictions:
  - Rotation and scaling of a rigid body in 3-D space, but no translation
  - Translation, rotation, and scaling of a planar patch in 3-D
- Advantage: corresponds to perspective projection model
- Disadvantage: hyperbolic motion function

Orthographic Motion Models

- Translational motion model $d_x = a_1, \quad d_y = b_1$
- 4-Parameter motion model: translation, zoom (isotropic Scaling), rotation in image plane

$$d_x = a_1 + a_2 x + a_3 y$$
$$d_y = b_1 - a_3 x + a_2 y$$

- Affine motion model:

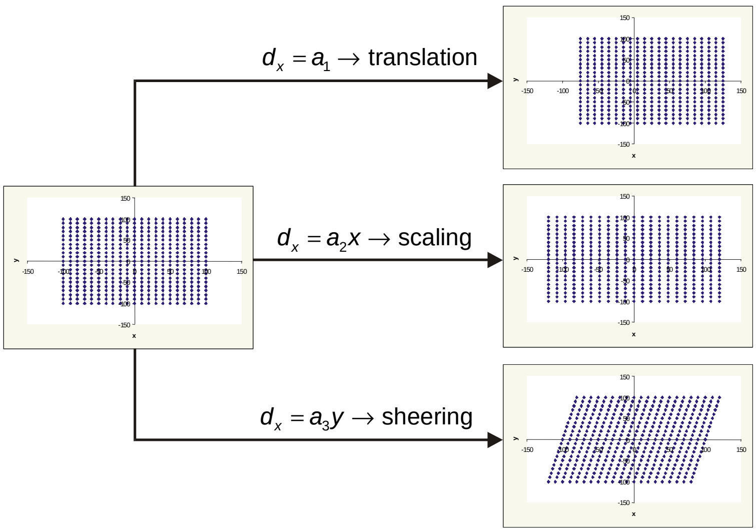
$$d_x = a_1 + a_2 x + a_3 y$$
$$d_y = b_1 + b_2 x + b_3 y$$

- Parabolic motion model

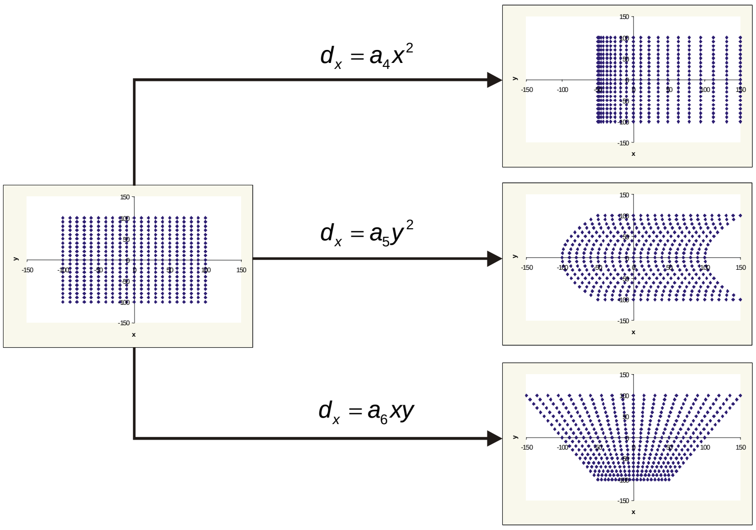
$$d_x = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 y^2 + a_6 xy$$
$$d_y = b_1 + b_2 x + b_3 y + b_4 x^2 + b_5 y^2 + b_6 xy$$

- Can also be viewed as Taylor expansions of perspective motion model

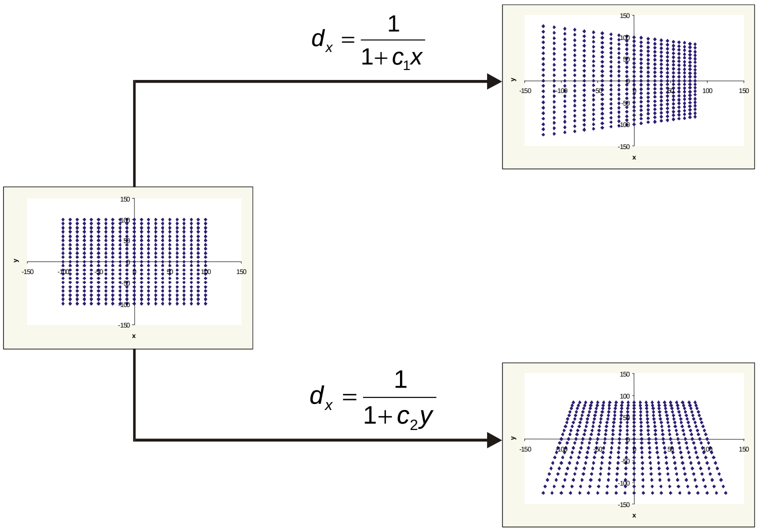
Impact of the Affine Parameters



Impact of the Parabolic Parameters



Impact of the Perspective Parameters



Differential Motion Estimation

- Assume small displacements d_x, d_y:

$$u(x,y,t) = s(x,y,t) - \hat{s}(x,y,t,d_x,d_y)$$
$$\approx s(x,y,t) - s'(x,y,t-\Delta t) - \frac{\partial s'(x,y,t-\Delta t)}{\partial x} \cdot d_x - \frac{\partial s'(x,y,t-\Delta t)}{\partial y} \cdot d_y$$

Displace frame difference

Horizontal and vertical gradient of image signal S

- Aperture problem: several observations required
- Inaccurate for displacements > 0.5 pel
  ➔ multigrid methods, iteration
- Minimize

$$\boxed{\min \sum_{y=1}^{By} \sum_{x=1}^{Bx} u^2(x,y,t)}$$

Gradient-Based Affine Refinement

- Displacement vector field is represented as

$$x' = a_1 + a_2 x + a_3 y$$

- Combination

$$y' = b_1 + b_2 x + b_3 y$$

$$u(x,y,t) \approx s(x,y,t) - s'(x,y,t-\Delta t) + \frac{\partial s'}{\partial x}(a_1 + a_2 x + a_3 y) + \frac{\partial s'}{\partial y}(b_1 + b_2 x + b_3 y)$$

yields a system of linear equations:

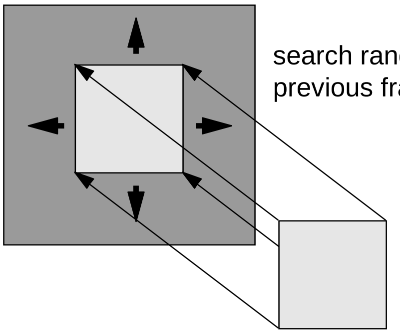
$$u = s - s' - \left(\frac{\partial s'}{\partial x}, \frac{\partial s'}{\partial x} x, \frac{\partial s'}{\partial x} y, \frac{\partial s'}{\partial y}, \frac{\partial s'}{\partial y} x, \frac{\partial s'}{\partial y} y \right) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

- System can be solved using, e.g., pseudo-inverse, minimizing

$$\underset{a_1,a_2,a_3,b_1,b_2,b_3}{\operatorname{argmin}} \sum_{y=1}^{By} \sum_{x=1}^{Bx} u^2(x,y,t)$$

Principle of Blockmatching

- Subdivide every image into square blocks
- Find one displacement vector for each block
- Within a search range, find a best „match" that minimizes an error measure.



Blockmatching: Error Measure

- Mean squared error (sum of squared errors)

$$SSE(d_x,d_y)=\sum_{y=1}^{By}\sum_{x=1}^{Bx}[s(x,y,t)-s'(x-d_x,y-d_y,t-\Delta t)]^2$$

- Sum of absolute differences

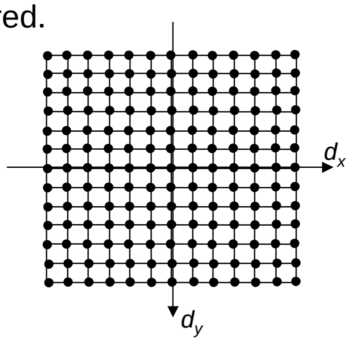
$$SAD(d_x,d_y)=\sum_{y=1}^{By}\sum_{x=1}^{Bx}\left|s(x,y,t)-s'(x-d_x,y-d_y,t-\Delta t)\right|$$

- Approximately same performance
  ➔ SAD less complex for some architectures

Blockmatching: Search Strategies I

Full search

- All possible displacements within the search range are compared.

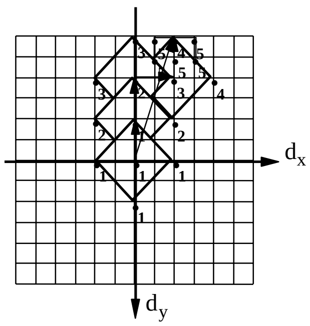


- Computationally expensive
- Highly regular, parallelizable

Block matching: Search Strategies II

2D logarithmic search (Jain + Jain, 1981)

- Iterative comparison of error measure values at 5 neighboring points
- Logarithmic refinement of the search pattern if
  - best match is in the center of the 5-point pattern
  - center of search pattern touches the border of the search range



from: Girod

Block matching: Search Strategies III

Computational Complexity

	Block comparisons a	b
2D logarithmic	18	21
full search	169	169

Example: max. horizontal, vertical displacement = 6
integer-pel accuracy

a - for special vector (2,6)
b - worst case

from: Girod

Block Comparison Speed-Ups

- Triangle inequalities for SAD and SSE

$$\sum_{block} |S_k - S_{k-1}| \geq \left| \sum_{block} (S_k - S_{k-1}) \right| = \left| \sum_{block} S_k - \sum_{block} S_{k-1} \right|$$

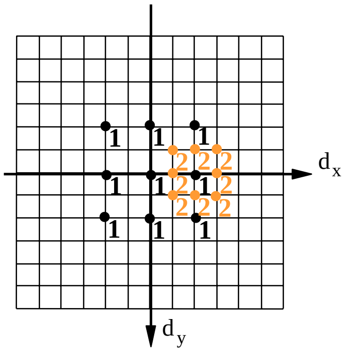
$$\sum_{block} |s_k - s_{k-1}|^2 \geq \frac{1}{N} \left| \sum_{block} (S_k - S_{k-1}) \right|^2 = \frac{1}{N} \left| \sum_{block} S_k - \sum_{block} S_{k-1} \right|^2$$

number of terms in sum

- Strategy:
  1. Compute partial sums for blocks in current and previous frame
  2. Compare blocks based on partial sums
  3. Omit full block comparison, if partial sums indicate worse error measure than previous best result
- Performance: > 20x speed up of full search block matching reported by employing
  1. Sum over 16x16 block
  2. Row wise block projection
  3. Column wise block projection (Lin + Tai, IEEE Tr. Commun., May 97)

Sub-pel Translational Motion

- Motion vectors are often not restricted to only point into the integer-pel grid of the reference frame
- Typical sub-pel accuracies: half-pel and quarter-pel
- Sub-pel positions are often estimated by „refinement“

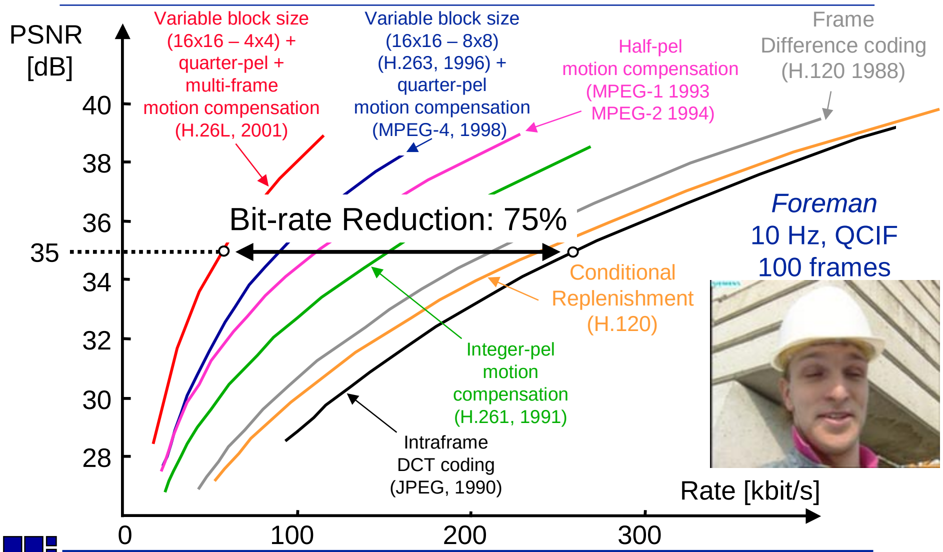


History of Motion Compensation

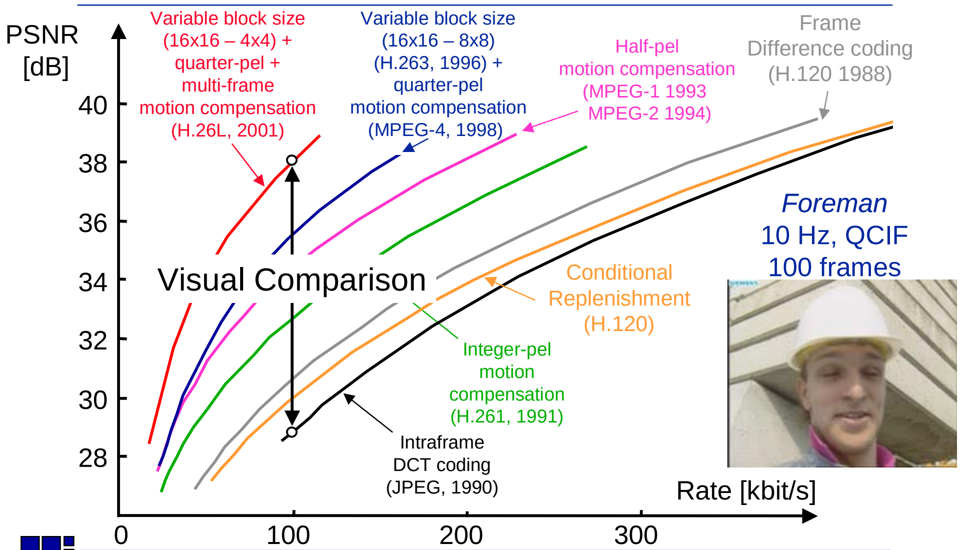
Complexity increase

- *Intraframe coding:* only spatial correlation exploited
  ➔ DCT [Ahmed, Natarajan, Rao 1974], JPEG [1992]
- *Conditional replenishment*
  ➔ H.120 [1984] (*DPCM, scalar quantization*)
- *Frame difference coding*
  ➔ H.120 Version 2 [1988]
- *Motion compensation: integer-pel accurate displacements*
  ➔ H.261 [1991]
- *Half-pel accurate motion compensation*
  ➔ MPEG-1 [1993], MPEG-2/H.262 [1994]
- *Variable block-size (16x16 & 8x8) motion compensation*
  ➔ H.263 [1996], MPEG-4 [1999]
- *Variable block-size (16x16 – 4x4) and multi-frame motion compensation*
  ➔ H.26L [2001]

Milestones in Video Coding



Milestones in Video Coding



Foreman @ 100 kbit/s, 10 Hz

Intraframe DCT coding
No motion compensation
(JPEG, 1990)

Variable block size (16x16 – 4x4) +
quarter-pel +
multi-frame motion compensation
(H.26L, 2001)

Summary

- Video coding as a hybrid of motion compensation and prediction error coding
- Lagrangian bit-allocation rules specify constant slope allocation to motion coefficients and prediction error
- Motion models can represent various kinds of motions
- In practice: affine or 8-parameter model for camera motion, translational model for small blocks
- Differential methods calculate displacement from spatial and temporal differences in the image signal.
- Block matching computes error measure for candidate displacements and finds best match.
- Speed up block matching by search methods and by clever application of triangle inequality
- Hybrid video coding has been drastically improved by enhanced motion compensation capabilities