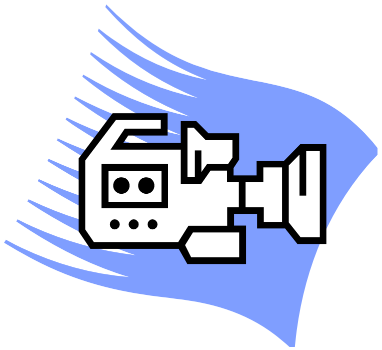
Multimedia Computing

Video: Basic Concepts



Topics

- Color models in video
- Fundamental concepts in video
- Basic motion computing methods

Video color transforms

- Methods of dealing with color in digital video largely derive from older methods of coding color for analog TV.
- Luminance is separated from color information.
  - YIQ is used to transmit TV signals in North America and Japan.
  - In Europe, video tape uses the PAL codings, which are based on TV that uses a matrix transform called YUV.
  - Finally, digital video mostly uses a matrix transform called YCbCr that is closely related to YUV.

YUV color model

- Luminance is calculated as

$$Y=0.299R+0.587G+0.114B$$

- Chrominance refers to color differences U, V

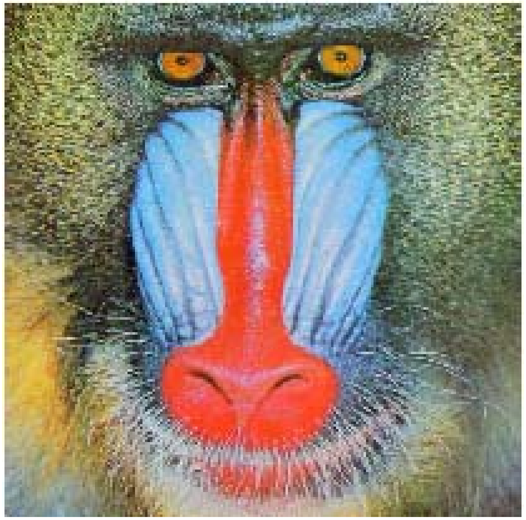
$$U=B-Y;\ V=R-Y$$

- Then

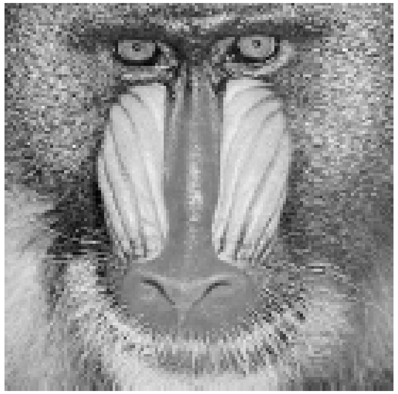
$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- For a gray image, the chrominance (U, V) is zero. Color TV can be displayed on a black/white TV by just using the Y signal.

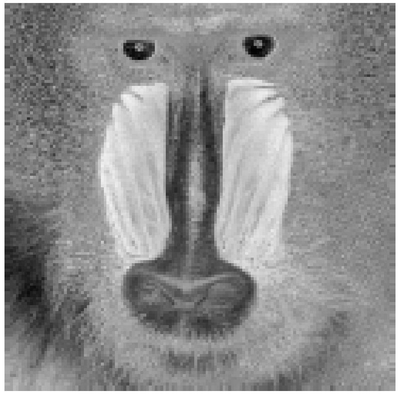
Example



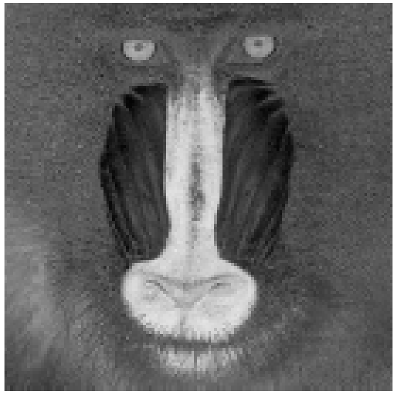
Original color image



Y



U



V

YIQ color model

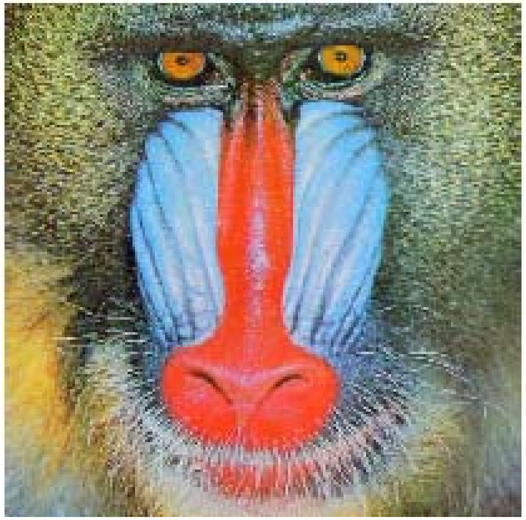
- YIQ is used in NTSC color TV broadcasting. Again, gray pixels generate zero (I,Q) chrominance signal.
- I and Q are a rotated version of U and V .
- Y in YIQ is the same as in YUV; U and V are rotated by 33^0:

  $I = 0.877283(R - Y)\cos(33^0) - 0.492111(B - Y)\sin(33^0)$

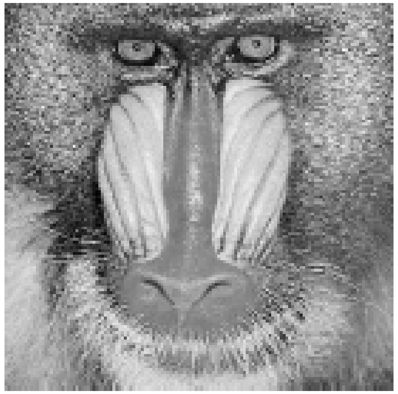
  $Q = 0.877283(R - Y)\sin(33^0) + 0.492111(B - Y)\cos(33^0)$
- This leads to the following matrix transform:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595879 & -0.274133 & -0.321746 \\ 0.211205 & -0.523083 & 0.311878 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

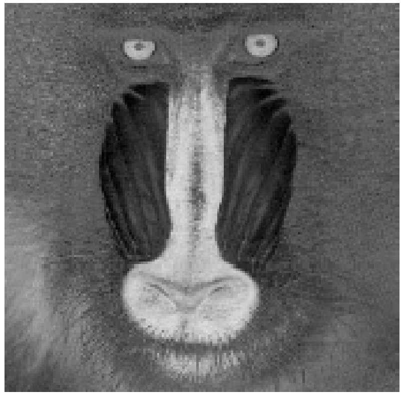
Example



Original color image



Y



I

Q

YCbCr color model

- The Recommendation ITU-R BT.601-4 standard for digital video uses another color space, YCbCr, closely related to the YUV transform.
- Normalize R, G, B in [0, 1], there is

$$Cb = ((B-Y)/1.772)+0.5$$

$$Cr = ((R-Y)/1.402)+0.5$$

- In practice, Rec. 601 species 8-bit coding, with a maximum Y value of only 219, and a minimum of 16. Cb and Cr have a range of ±112 and offset of 128.

YCbCr color model

- Normalize R, G, B in [0, 1], we obtain YCbCr in [0, 255] via the transform:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

- The YCbCr transform is used in JPEG image compression and MPEG video compression.

Topics

- Color models in video
- Fundamental concepts in video
- Basic motion computing methods

Types of video signals

- **Component video:** Higher-end video systems make use of three separate video signals for the red, green, and blue image planes. Each color channel is sent as a separate video signal.
  - Most computer systems use Component Video, with separate signals for R, G, and B signals.
  - Component Video gives the best color reproduction since there is no “crosstalk” between the three channels.
  - Component video, however, requires more bandwidth and good synchronization of the three components.

Types of video signals

- **Composite video**: color (“chrominance") and intensity (“luminance") signals are mixed into a single carrier wave. Chrominance is a composition of two color components (I and Q, or U and V).
  - The chrominance and luminance components can be separated at the receiver end and then the two color components can be further recovered.
- Since color and intensity are wrapped into the same signal, some interference between the luminance and chrominance signals is inevitable.

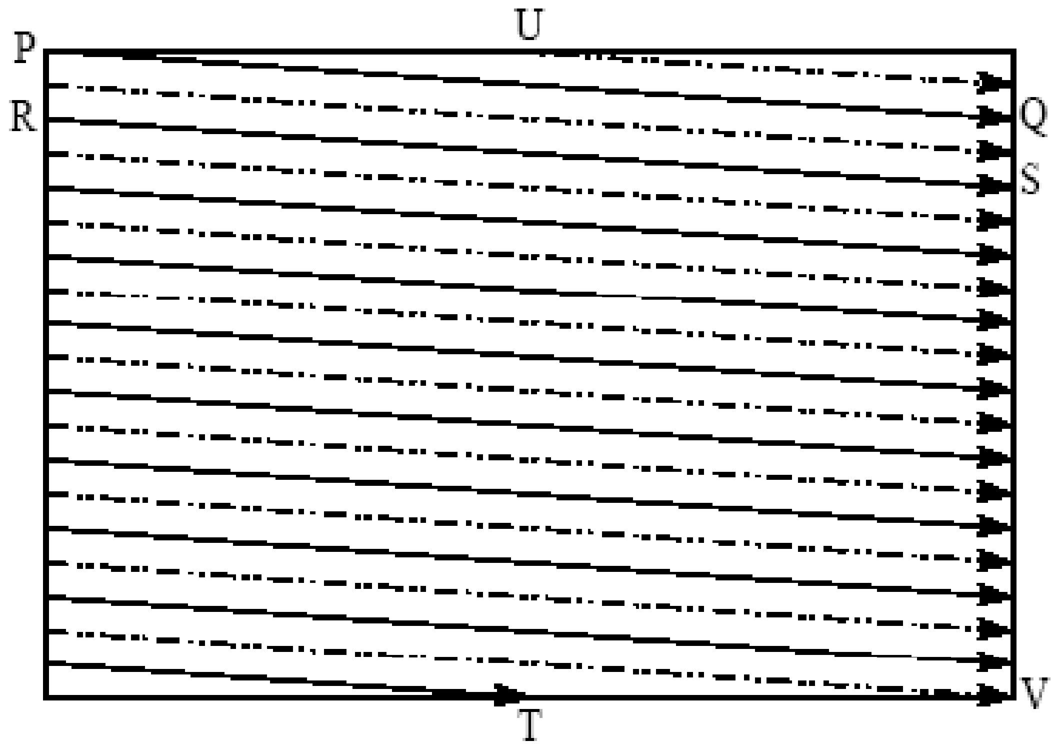
Types of video signals

- As a compromise, **S-Video** (separated video) uses two wires, one for luminance and another for a composite chrominance signal.
- So there is less crosstalk between the color information and the crucial gray-scale information.
- The reason for placing luminance into its own part of the signal is that black/white information is most crucial for visual perception.
- We can send less accurate color information than intensity information.

Analog video

- In analog video, an analog signal f(t) samples a time-varying image.
- "Progressive" scanning traces through a complete picture (a frame) row-wise for each time interval.
- In TV and some monitors and multimedia standards, another system, called "interlaced" scanning is used:
  - The odd-numbered lines are traced first, and then the even-numbered lines are traced.

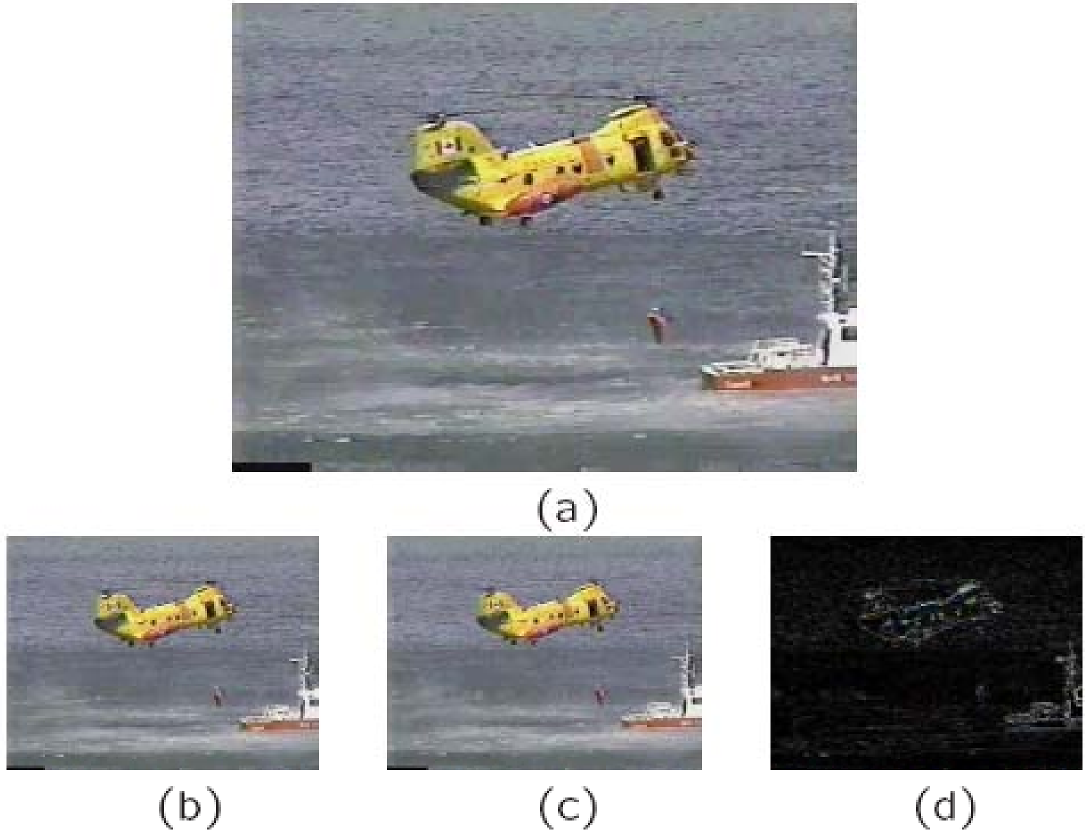
Analog video



- First the solid (odd) lines are traced, P to Q, then R to S, etc., ending at T; then the even field starts at U and ends at V.

Example

- Because of interlacing, the odd and even lines are displaced in time from each other. This is generally not noticeable except when very fast action is taking place on screen, when blurring may occur.
  - For example, the moving helicopter is blurred more than is the still background.



NTSC video

- **NTSC** (National Television System Committee) TV standard is mostly used in North America and Japan. It uses the familiar 4:3 aspect ratio and uses 525 scan lines per frame at 30 frames per second (fps).
- NTSC uses the YIQ color model, and the technique of quadrature modulation is employed to combine I (in-phase) and Q (quadrature) signals into a single chroma signal C.

PAL video

- PAL (Phase Alternating Line) is a TV standard widely used in Western Europe, China, India, and many other parts of the world.
- PAL uses 625 scan lines per frame, at 25 frames/second, with a 4:3 aspect ratio and interlaced fields.
- PAL uses the YUV color model. It uses an 8 MHz channel and allocates a bandwidth of 5.5 MHz to Y, and 1.8 MHz each to U and V.

SECAM video

- SECAM stands for *Systeme Electronique Couleur Avec Memoire*, the third major broadcast TV standard.
- SECAM also uses 625 scan lines per frame, at 25 frames per second, with a 4:3 aspect ratio and interlaced fields.
- SECAM and PAL are very similar. They differ slightly in their color coding scheme:
  - In SECAM, U and V signals are modulated using separate color subcarriers.
  - They are sent in alternate lines, i.e., only one of the U or V signals will be sent on each scan line.

Comparison

TV System	Frame Rate (fps)	# of Scan Lines	Total Channel Width (MHz)	Bandwidth Allocation (MHz)		
				Y	I or U	Q or V
NTSC	29.97	525	6.0	4.2	1.6	0.6
PAL	25	625	8.0	5.5	1.8	1.8
SECAM	25	625	8.0	6.0	2.0	2.0

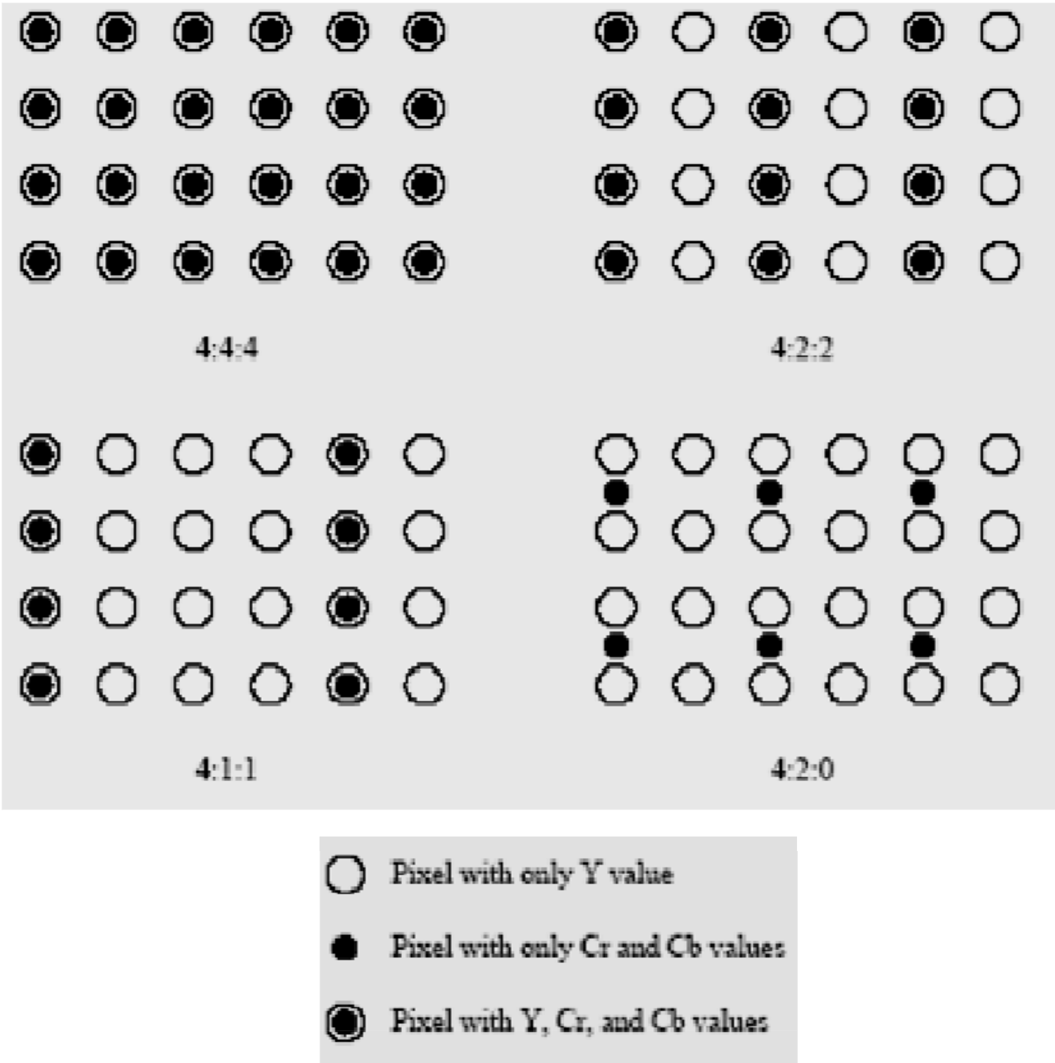
Digital video

- The advantages of digital representation for video are many.
  - Video can be stored on digital devices or in memory, ready to be processed (noise removal, cut and paste, etc.), and integrated to various applications;
  - Direct access is possible, which makes nonlinear video editing achievable as a simple task;
  - Repeated recording does not degrade image quality;
  - Ease of encryption and better tolerance to channel noise.

Chroma subsampling

- Since humans see color with much less spatial resolution than they see black and white, it makes sense to "decimate" the chrominance signal.
  - The chroma subsampling scheme "4:4:4" indicates that no chroma subsampling is used: each pixel's Y, Cb and Cr values are transmitted, 4 for each of Y, Cb, Cr.
  - We also have "4:2:2", "4:1:1" and "4:2:0" schemes.

Chroma subsampling



CCIR standards for digital video

- **CCIR** is the Consultative Committee for International Radio.
- One of the most important standards it has produced is CCIR-601, for component digital video.
  - This standard has since become an international standard for professional video applications adopted by certain digital video formats including the popular DV video.

CIF and QCIF standards

- CIF stands for Common Intermediate Format specified by the CCITT (International Telegraph and Telephone Consultative Committee).
  - The idea of CIF is to specify a format for lower bitrate.
  - CIF is about the same as VHS quality. It uses a progressive (non-interlaced) scan.
- QCIF stands for "Quarter-CIF".

Digital video specifications

	CCIR 601 525/60 NTSC	CCIR 601 625/50 PAL/SECAM	CIF	QCIF
Luminance resolution	720 × 480	720 × 576	352 × 288	176 × 144
Chrominance resolution	360 × 480	360 × 576	176 × 144	88 × 72
Color Subsampling	4:2:2	4:2:2	4:2:0	4:2:0
Aspect Ratio	4:3	4:3	4:3	4:3
Fields/sec	60	50	30	30
Interlaced	Yes	Yes	No	No

Topics

- Color models in video
- Fundamental concepts in video
- Basic motion computing methods

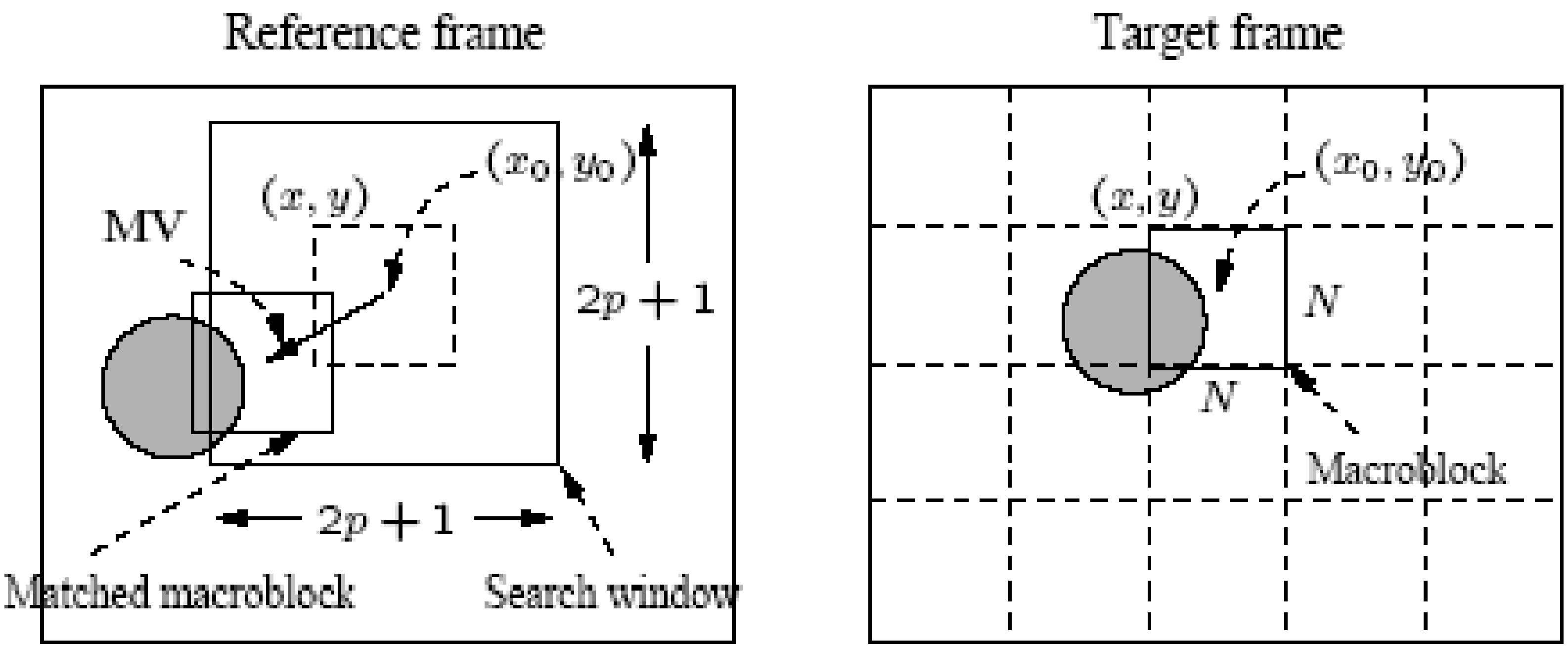
Motion in image sequence

- A video consists of a time-ordered sequence of frames, i.e. images.
- Consecutive frames in a video are similar, i.e. temporal redundancy exists.
- Temporal redundancy can be exploited for applications such as video compression, super-resolution, etc.
- Motion estimation is a key issue in exploiting the video temporal redundancy.

Block-based motion estimation

- Each image is divided into macroblocks. Motion estimation is performed at the macroblock level.
- The current image frame is referred to as Target Frame.
- A match is sought between the macroblock in the Target Frame and the most similar macroblock in a previous and/or future frame, referred to as Reference Frame.
- The displacement of the reference macroblock to the target macroblock is called a motion vector (MV).

Search window



- MV search is usually limited to a search window of size (2p+1)x(2p+1).

Mean Absolute Difference (MAD)

- The difference between two macroblocks can be measured by their Mean Absolute Difference (MAD):

$$MAD(i,j) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)|$$

N – size of the macroblock,

k and l – indices for pixels in the macroblock,

i and j – horizontal and vertical displacements,

$C(x+k, y+l)$ – pixels in macroblock in Target frame,

$R(x+i+k, y+j+l)$ – pixels in macroblock in Reference frame.

Block matching criterion

- The goal of the search is to find a vector (i, j) as the motion vector MV = (u,v), such that MAD(i, j) is minimum:

$$(u, v) = [\ (i, j)\ |\ MAD(i, j) \ \text{ is } \ \text{minimum}$$

$$i \in [-p, p],\ \ j \in [-p, p]\ ]$$

Full search

- Full search: sequentially search the whole (2p+1)x (2p+1) window in the reference frame.
- A macroblock centered at each of the positions within the window is compared to the macroblock in the target frame pixel by pixel, i.e. by using the MAD of the two blocks.
- The vector (i,j) that offers the least MAD is set as the MV (u, v) for the macroblock in the target frame.

Full search algorithm

```
begin
   min_MAD = LARGE_NUMBER;          /* Initialization */
   for i = −p to p
     for j = −p to p
       {
         cur_MAD = MAD(i, j);
         if cur_MAD < min_MAD
           {
             min_MAD = cur_MAD;
             u = i;       /* Get the coordinates for MV. */
             v = j;
           }
       }
end
```

Complexity of full search

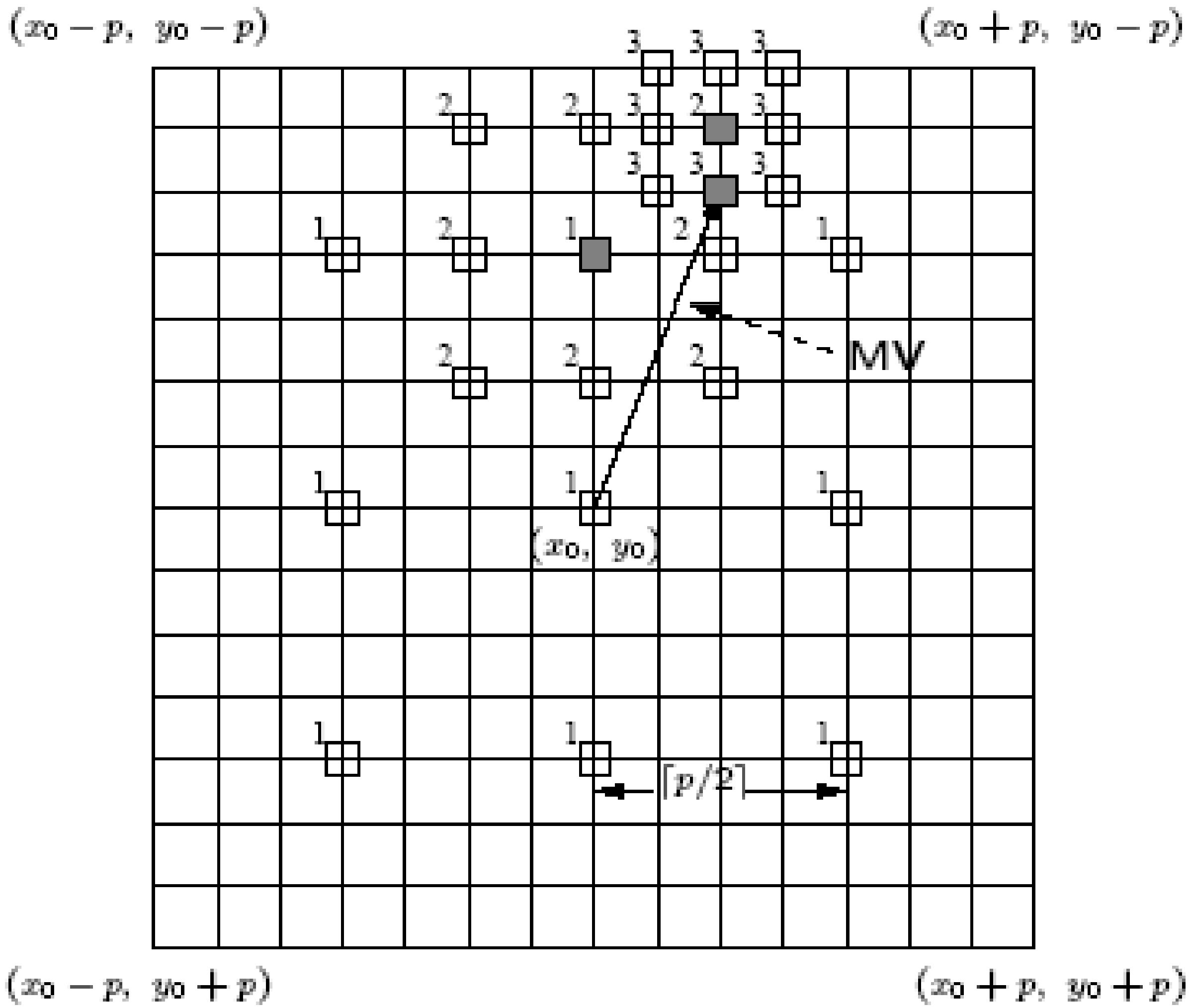
- Full search method is very costly.
- Assuming each pixel comparison requires three operations: subtraction, absolute value, addition.
- Then the cost for obtaining a motion vector for a single macroblock is

$$(2p+1)\times(2p+1)\times N^2\times 3 \Rightarrow O(p^2N^2)$$

2D logarithmic search

- **Logarithmic search**: a cheaper version, that is suboptimal but still usually effective.
- 2D Logarithmic search of motion vectors takes several iterations and is similar to a binary search.
- Initially only nine locations in the search window are used as seeds for a MAD-based search; they are marked as “1”.
- After the one that yields the minimum MAD is located, the center of the new search region is moved to it and the step-size is reduced to half.
- In the next iteration, the nine new locations are marked as “2”, and so on.

2D logarithmic search



The complexity of 2D logarithmic search is O(log*p*N²).

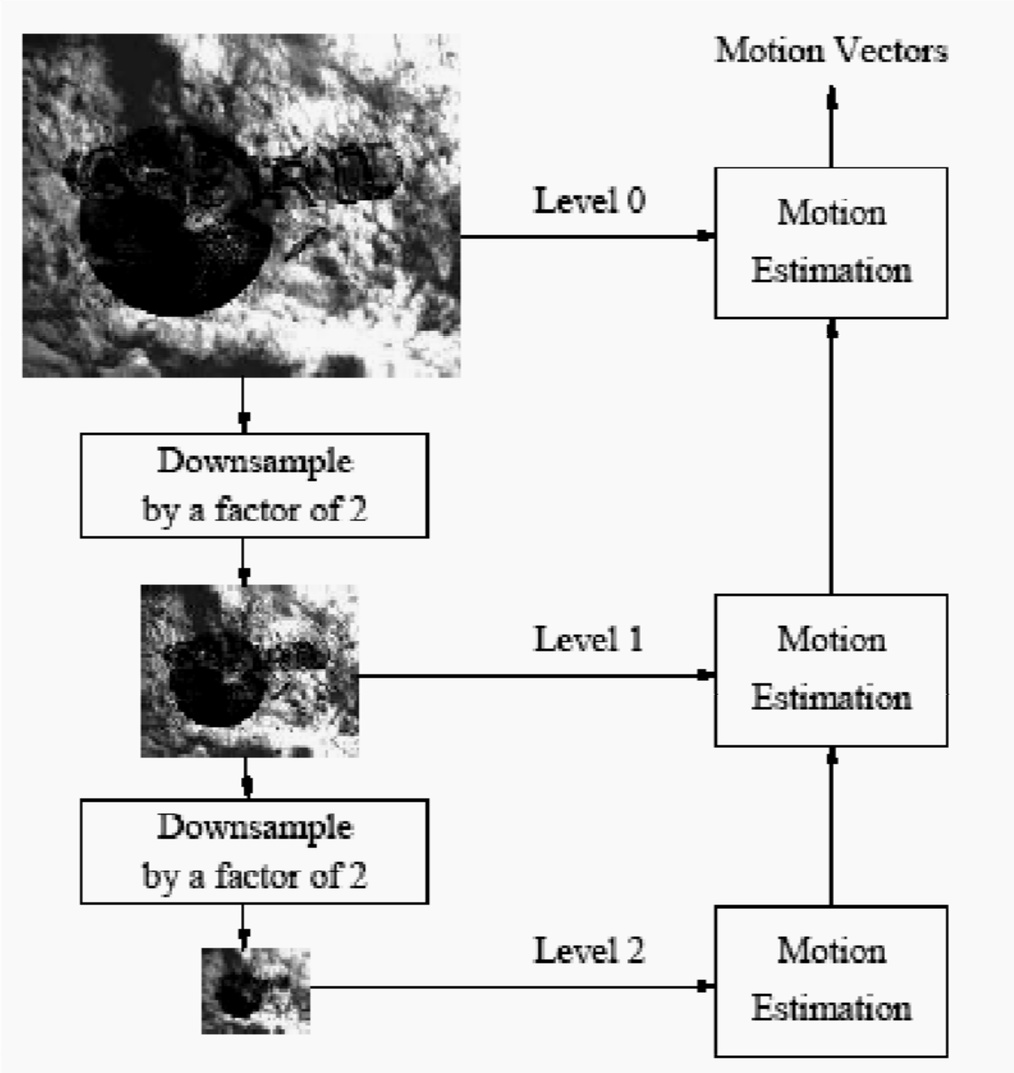
2D logarithmic search algorithm

```
begin
  offset = ⌈p/2⌉;
  Specify nine macroblocks within the search window in the Reference frame,
  they are centered at (x0, y0) and separated by offset horizontally and/or
  vertically;
  while last ≠ TRUE
    {
      Find one of the nine specified macroblocks that yields minimum MAD;
      if offset = 1 then last = TRUE;
      offset = ⌈offset/2⌉;
      Form a search region with the new offset and new center found;
    }
end
```

Hierarchical search

- The search can benefit from a hierarchical approach in which initial estimation of the motion vector can be obtained from images with a significantly reduced resolution. Example:
  - A three-level hierarchical search: the original image is at Level 0, images at Levels 1 and 2 are obtained by down-sampling from the previous levels by a factor of 2, and the initial search is conducted at Level 2.
- Since the size of the macroblock is smaller and p can also be proportionally reduced, the number of operations required is greatly reduced.

Hierarchical search: example



Hierarchical search rule

- Given the estimated motion vector (u^k,v^k) at Level k, a 3x3 neighborhood centered at $(2u^k,2v^k)$ at Level k − 1 is searched for the refined motion vector.
- In other words, the refinement is such that at Level k − 1 the motion vector (u^{k-1},v^{k-1}) satisfies:

$$(2u^k-1 \le u^{k-1} \le 2u^k+1,\ 2v^k-1 \le v^{k-1} \le 2v^k+1)$$

Hierarchical search algorithm

```
begin
  // Get macroblock center position at the lowest resolution Level k
  x_0^k = x_0^0/2^k;    y_0^k = y_0^0/2^k;
  Use Sequential (or 2D Logarithmic) search method to get initial estimated
  MV(u^k, v^k) at Level k;
  while last ≠ TRUE
   {
     Find one of the nine macroblocks that yields minimum MAD
     at Level k − 1 centered at
     (2(x_0^k + u^k) − 1 ≤ x ≤ 2(x_0^k + u^k) + 1,  2(y_0^k + v^k) − 1 ≤ y ≤ 2(y_0^k + v^k) + 1);
     if k = 1   then last = TRUE;
     k = k − 1;
     Assign (x_0^k, y_0^k) and (u^k, v^k) with the new center location and MV;
   }
end
```

Computation cost comparison

Search Method	OPS_per_second for 720 × 480 at 30 fps	
	$p = 15$	$p = 7$
Sequential search	29.89×10^9	7.00×10^9
2D Logarithmic search	1.25×10^9	0.78×10^9
3-level Hierarchical search	0.51×10^9	0.40×10^9

Comparison of Computational Cost of Motion Vector Search using the three methods

An example application of motion estimation

- Temporal color demosaicking

Limitation of spatial demosaicking

- Spatial demosaicking exploits only the correlation within a frame.
- Spatial demosaicking can not faithfully reconstruct the features when the input image has very high frequency components.

Full color image

Simulated CFA mosaic image

Demosaicking Results of Spatial Methods

Original image

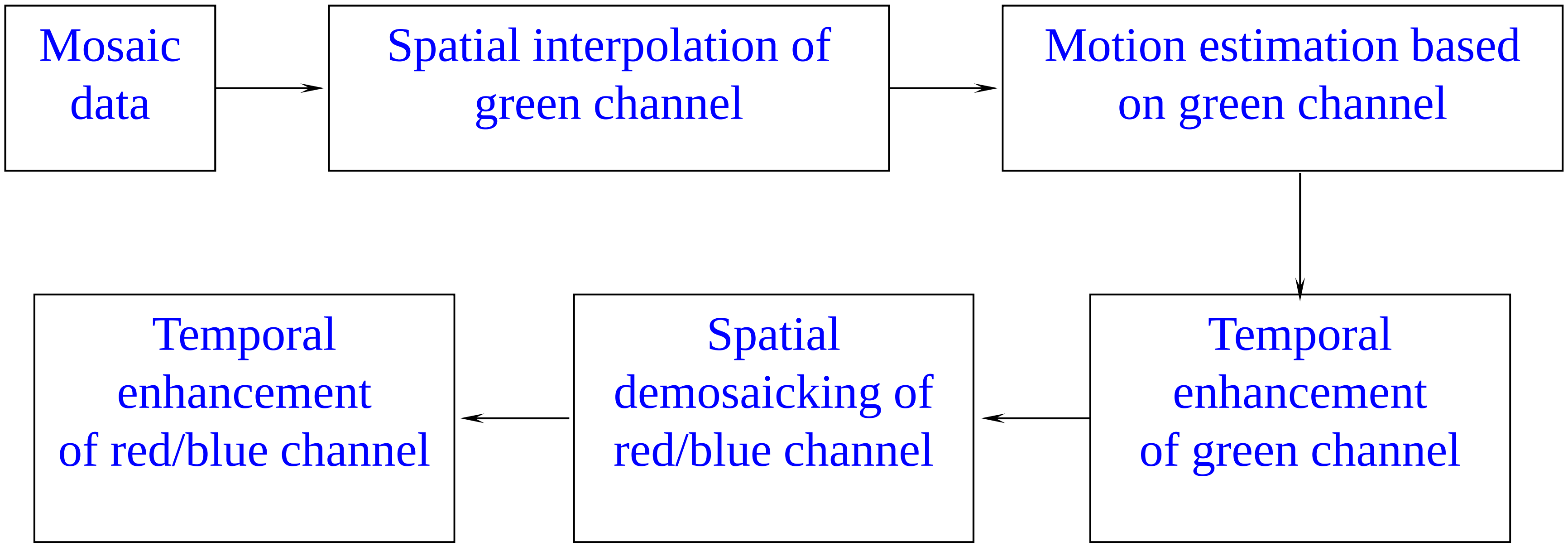
Spatial method 1

Spatial method 2

Temporal demosaicking

- Additional information is needed to overcome the difficulties in spatial demosaicking.
- The temporal dimension of a sequence of mosaic images can reveal new information that would be otherwise not available within a frame.
- The temporal correlation between adjacent frames can be exploited through motion estimation and data fusion to improve the accuracy of color demosaicking.

Temporal demosaicking



Flow Chart of Temporal Demosaicking Scheme

X. Wu and Lei Zhang, "Improvement of color video demosaicking in temporal domain," *IEEE Trans. Image Processing,* vol. 15, no. 10, pp. 3138-3151, Oct. 2006.

Experiment

Full color sequence

Simulated mosaic data

- The spatial demosaicking algorithms used for comparison:
  - *S1: The second order Laplacian correction by Hamilton and Adams.*
  - *S2: The variable number of gradients by Chang et al.*
  - *S3: The LMMSE based directional filtering and fusion by Zhang and Wu.*

Full color image

S1

S2

S3

Proposed temporal demosaicking

S1

S2

S3

Proposed temporal demosaicking

References

- Ze-Nian Li and Marks S. Drew, *Fundamentals of Multimedia,* Pearson Education, Inc.