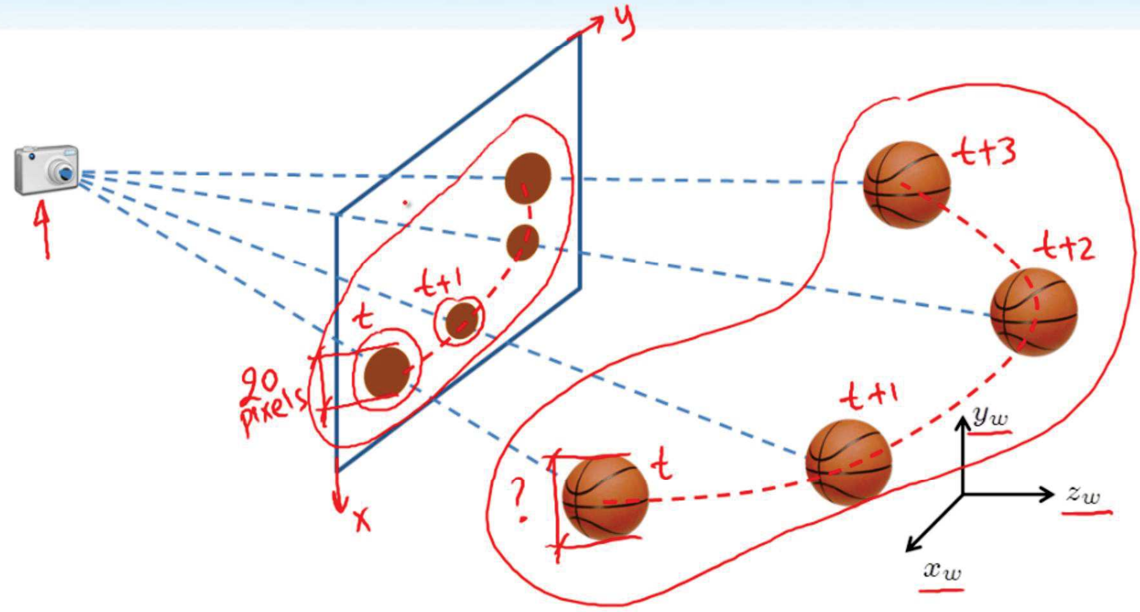
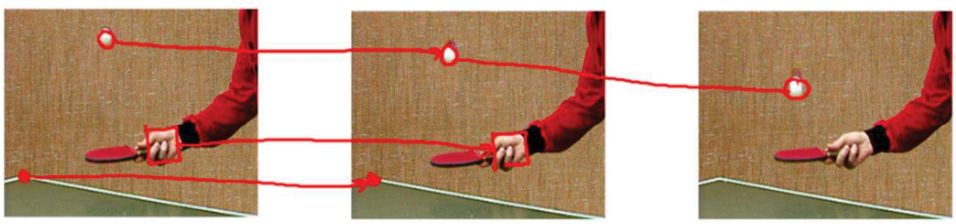
2D vs. 3D Motion



Basic Idea



$x_{k-1}(n_1, n_2)$ $x_k(n_1, n_2)$ $x_{k+1}(n_1, n_2)$

$I(x, y, t-1)$

"True" Motion vs. Optical Flow

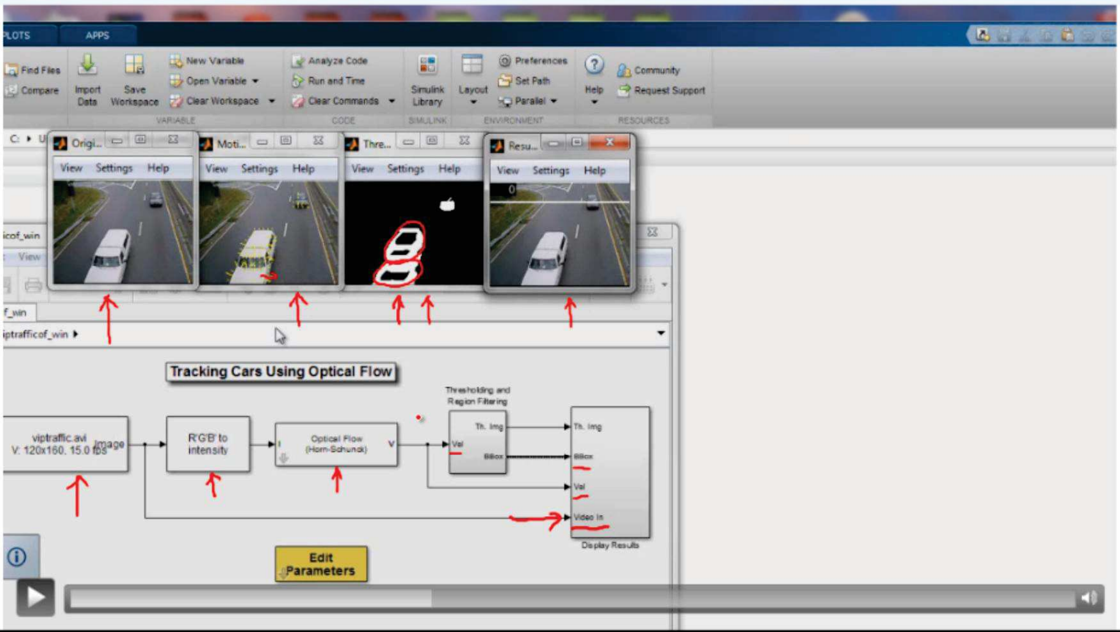
Non-zero true motion
Zero optical flow

Zero true motion
Non-zero optical flow

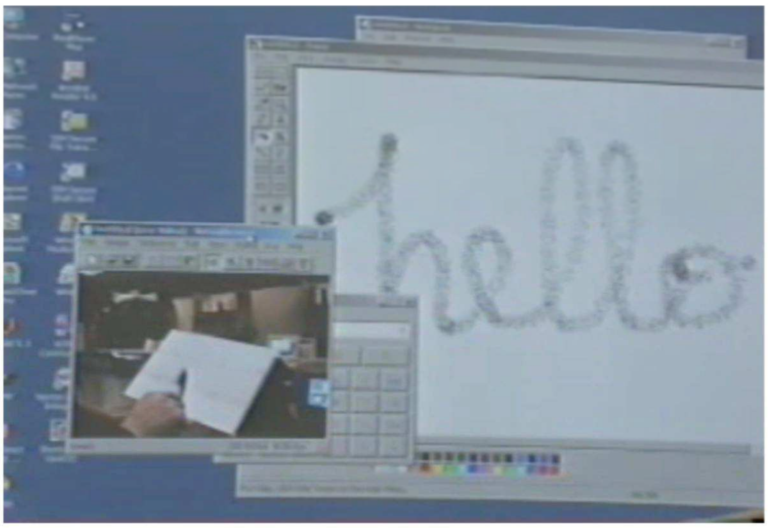
Motion Estimation Applications

- Object Tracking
- Human Computer Interaction (HCI)
- Temporal Interpolation
- Spatio-Temporal Filtering
- Compression

Object Tracking

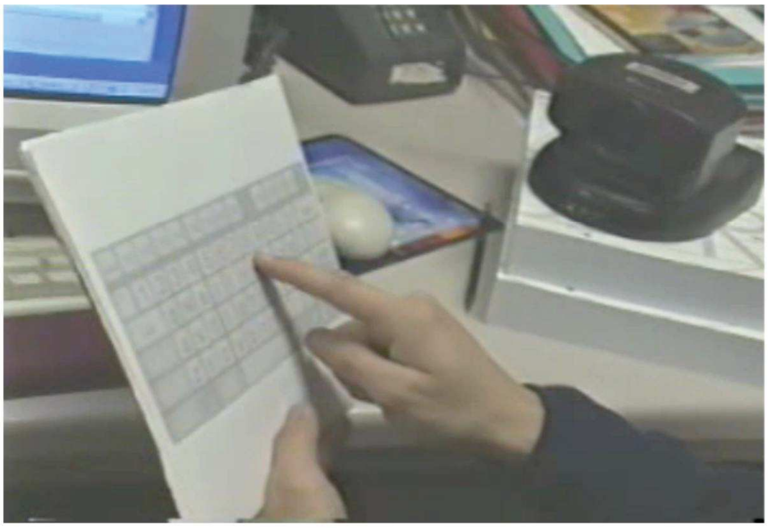


Human Computer Interaction



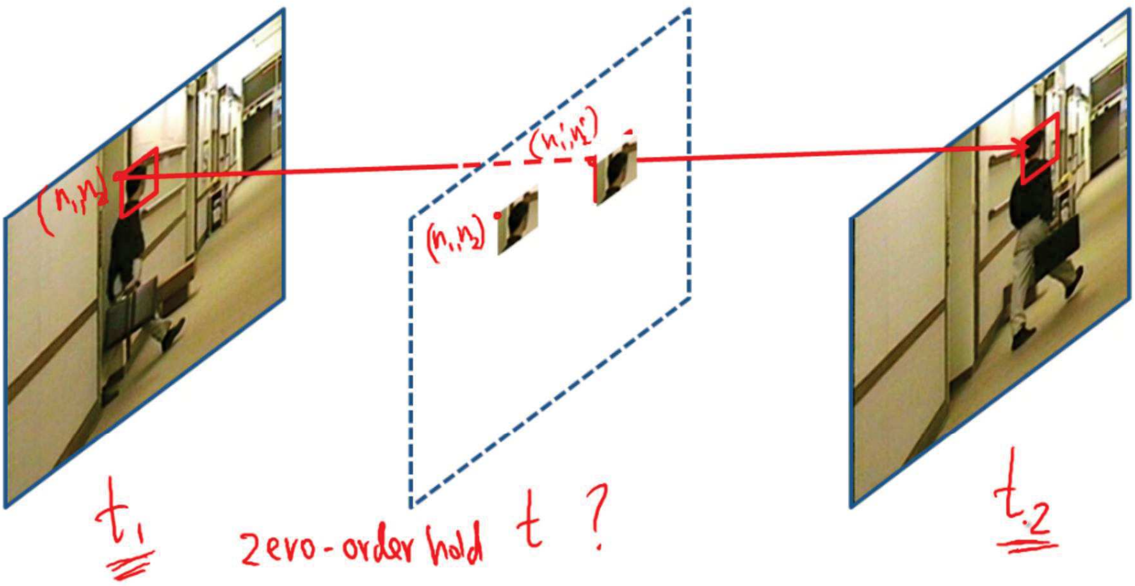
Courtesy of Prof. Ying Wu, Dept of EECS, Northwestern University

Human Computer Interaction

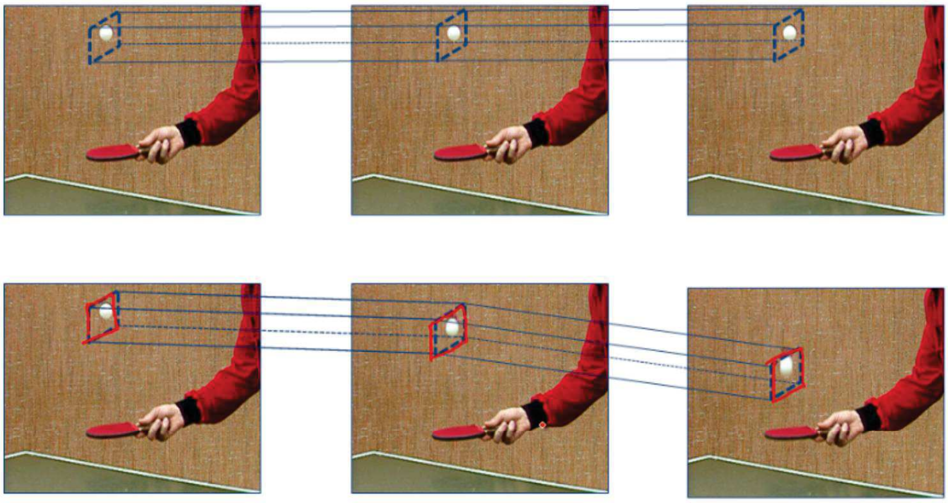


Courtesy of Prof. Ying Wu, Dept of EECS, Northwestern University

MC Temporal Interpolation



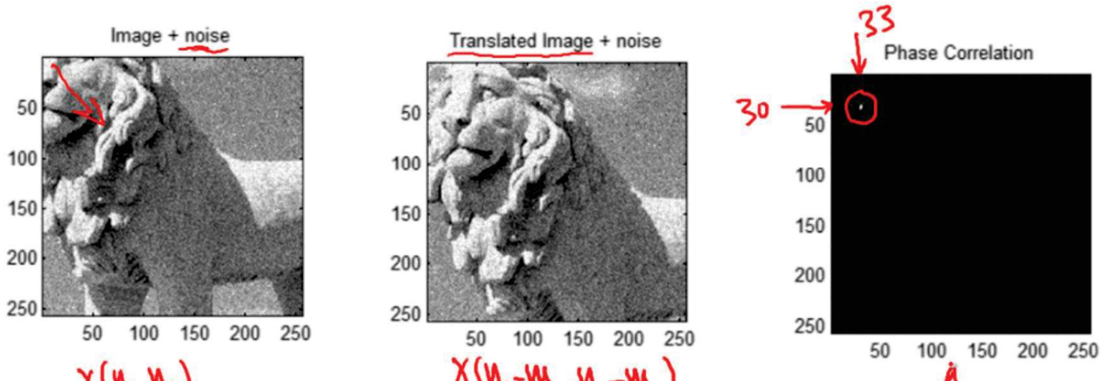
Motion Compensated Temporal Filtering



Classification of ME Methods

- Direct Methods
  - Phase Correlation
  - Block Matching
  - Spatio-temporal gradient
    - Optical Flow
    - Pel-Recursive
- Indirect Methods
  - Feature Matching

Phase Correlation Example



$x(n_1,n_2)$ $\quad$ $x(n_1-m_1, n_2-m_2)$

$X(\omega_1,\omega_2)$ $\quad$ $\boxed{e^{-j\omega_1 m_1} e^{-j\omega_2 m_2}} X(\omega_1,\omega_2)$

Phase Correlation

An image registration method

$$x_{t-1}(n_1,n_2) \leftrightarrow X_{t-1}(k_1,k_2) \qquad x_t(n_1,n_2) \leftrightarrow X_t(k_1,k_2)$$

Assume $\quad x_t(n_1,n_2) = x_{t-1}((n_1-m_1)_{N_1}, (n_2-m_2)_{N_2})$ $\quad N_1 \times N_2$ $\quad 0-(N_1-1)$ $\quad 0-(N_2-1)$

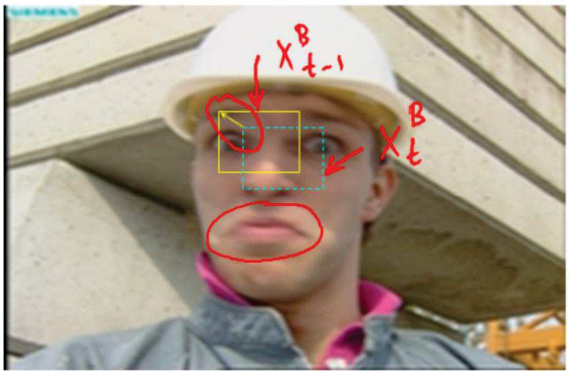
Then $\quad X_t(k_1,k_2) = X_{t-1}(k_1,k_2) e^{-j\frac{2\pi}{N_1}m_1k_1} e^{-j\frac{2\pi}{N_2}m_2k_2}$

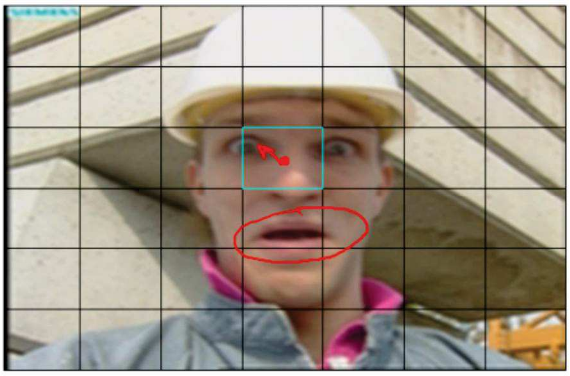
complex conjugate

Form: $$C(k_1,k_2) = \frac{X_t(k_1,k_2)\cdot X^*_{t-1}(k_1,k_2)}{|X_t(k_1,k_2)X^*_{t-1}(k_1,k_2)|} = \frac{|X_{t-1}(k_1,k_2)|^2\, e^{-j\frac{2\pi}{N_1}m_1k_1} e^{-j\frac{2\pi}{N_2}m_2k_2}}{|X_{t-1}(k_1,k_2)|^2}$$

$$c(n_1,n_2) = \delta(n_1-m_1, n_2-m_2)$$

Block Matching





X_t

Basic underlying assumptions:

1. no change in the ambient lighting.
2. objects are rigid
3. objects are translated in the 3D world on a plane parallel to the image place
4. no objects appeared or left the scene

Matching Criteria

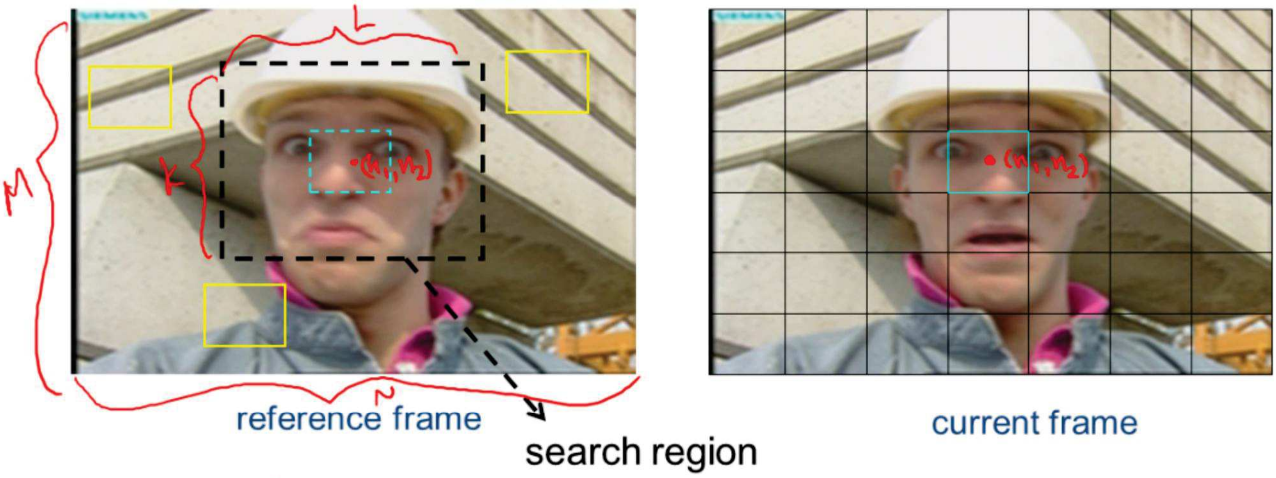
- A similarity or dissimilarity measure between regions (blocks)

$$\epsilon(d_1, d_2) = \sum_{(m_1, m_2) \in \mathcal{N}} \Phi\Big(x_t(n_1 + m_1, n_2 + m_2), x_{t-1}(n_1 + m_2 + d_1, n_2 + m_2 + d_2)\Big)$$

- Examples of $\Phi()$
  - Correlation Function
  - Mean Squared Error (MSE)
  - Mean Absolute Error (MAE) or Mean Absolute Difference (MAD)

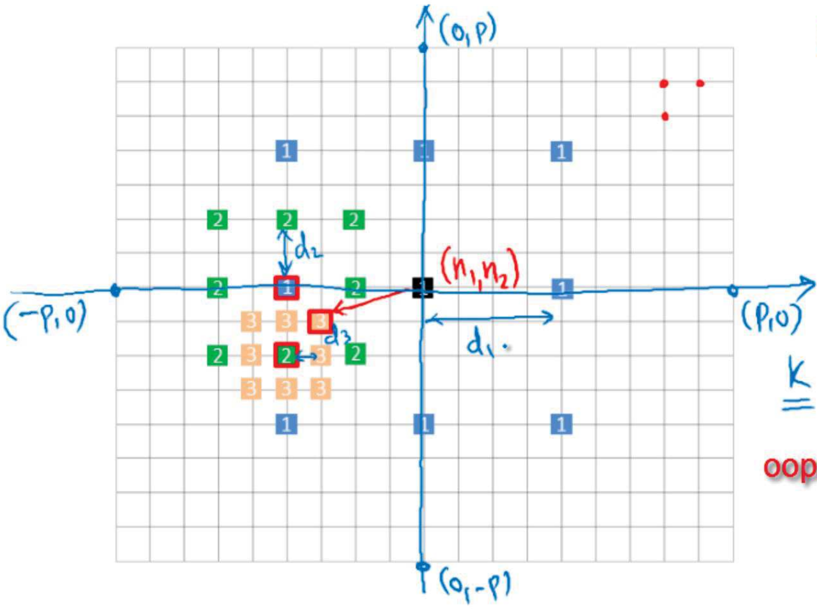
$$\epsilon(d_1, d_2) = \sum_{(m_1, m_2) \in \mathcal{N}} |x_t(n_1 + m_1, n_2 + m_2) - x_{t-1}(n_1 + m_2 + d_1, n_2 + m_2 + d_2)|$$

Search Region



M × N matching comparisons
K × L " "

2D Logarithmic Search



$(2p+1) \times (2p+1)$

$k = \lceil \log_2 P \rceil$

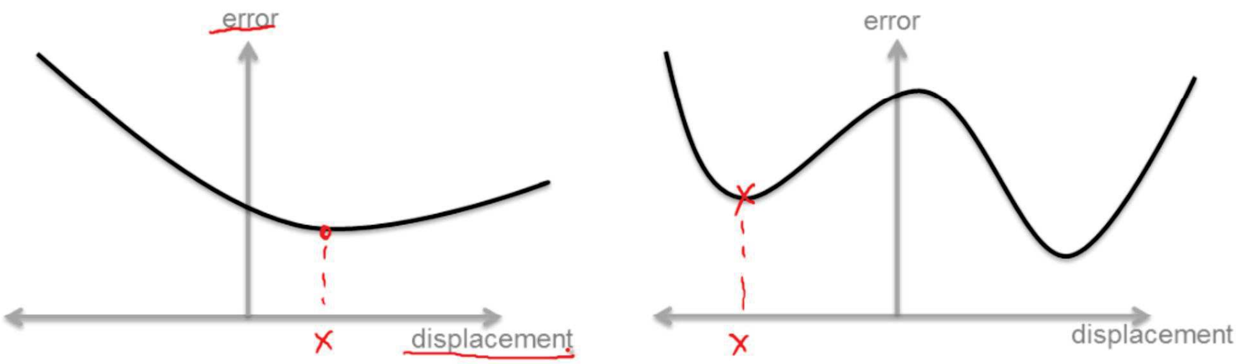
P=7, 15×15

k=3

$k \begin{cases} d_1 = 2^{k-1} \\ d_2 = \frac{d_1}{2} = 2^{k-2} \\ d_3 = \frac{d_2}{2} = 2^{k-1} \end{cases}$

oops! >>> $d_3 = 2^{k-3}$

Global vs. Local Search



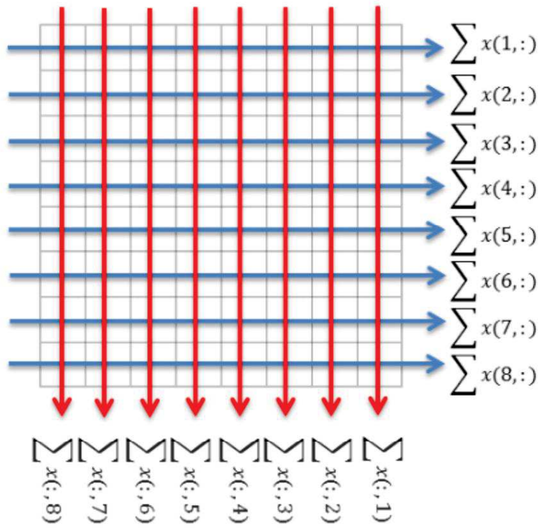
Pixel Sub-Sampling

1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4
1	2	1	2	1	2	1	2
3	4	3	4	3	4	3	4

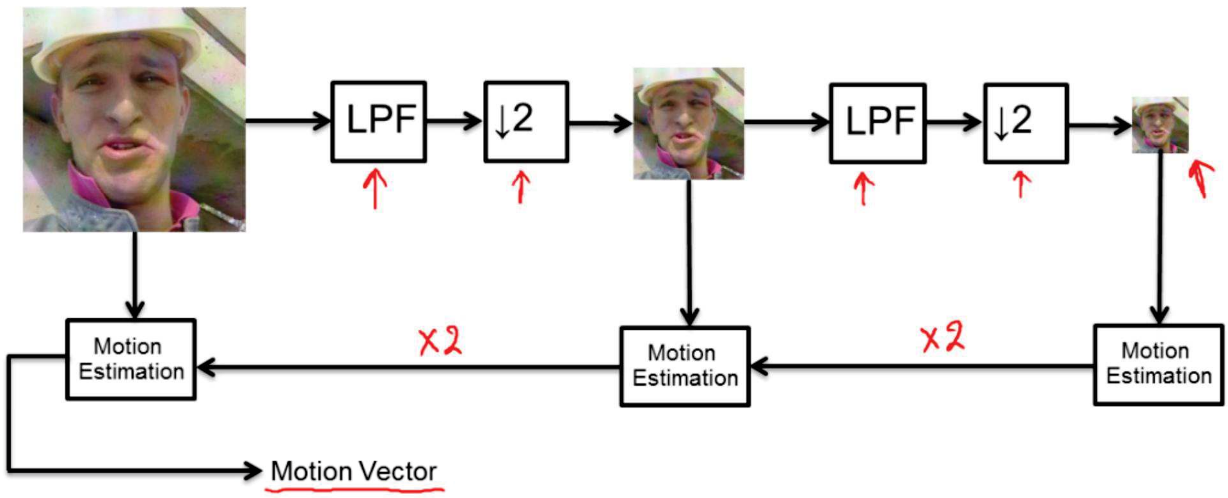
8

8

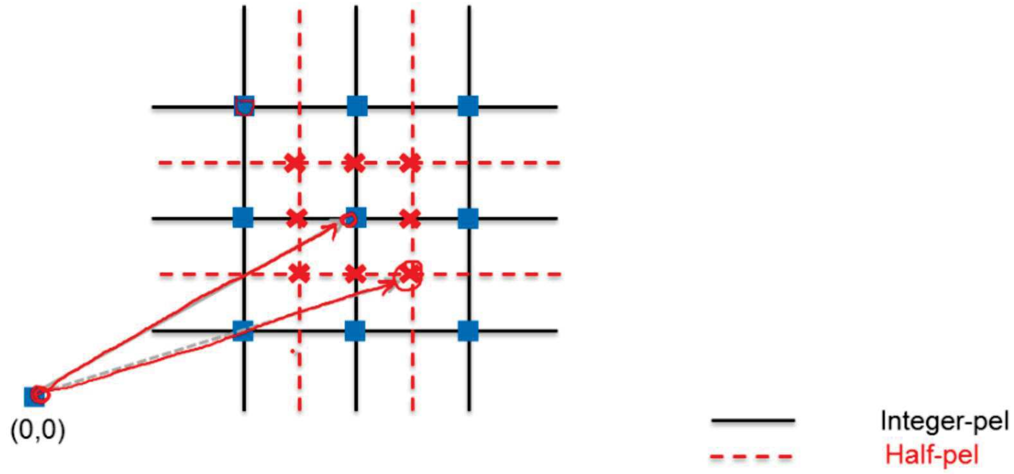
Pixel Projection



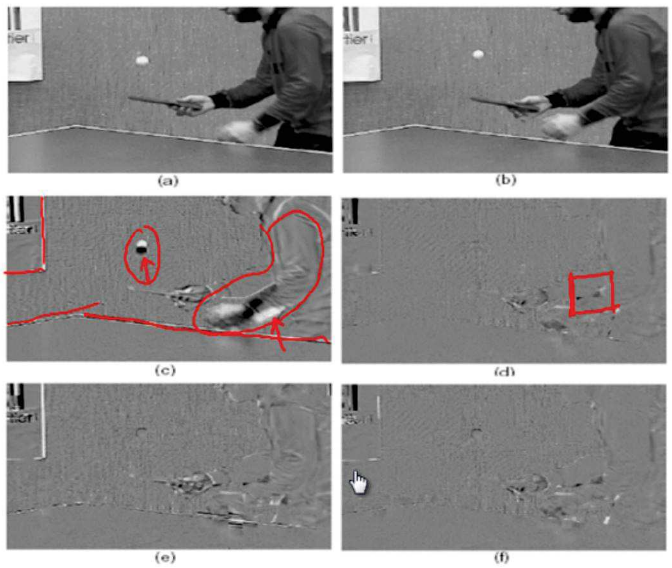
Hierarchical Motion Estimation



Sub-pixel Motion Estimation



Experimental Comparison



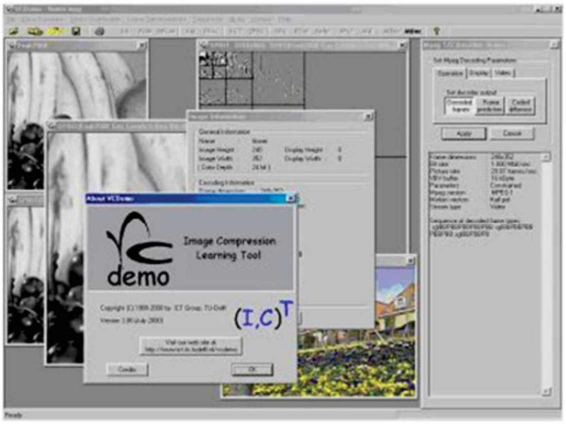
(a) Reference picture; (b) current picture;

(c) Frame difference; (d) full-search;

(e) Logarithmic search; (f) three level hierarchical search

V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, Springer, 1997.

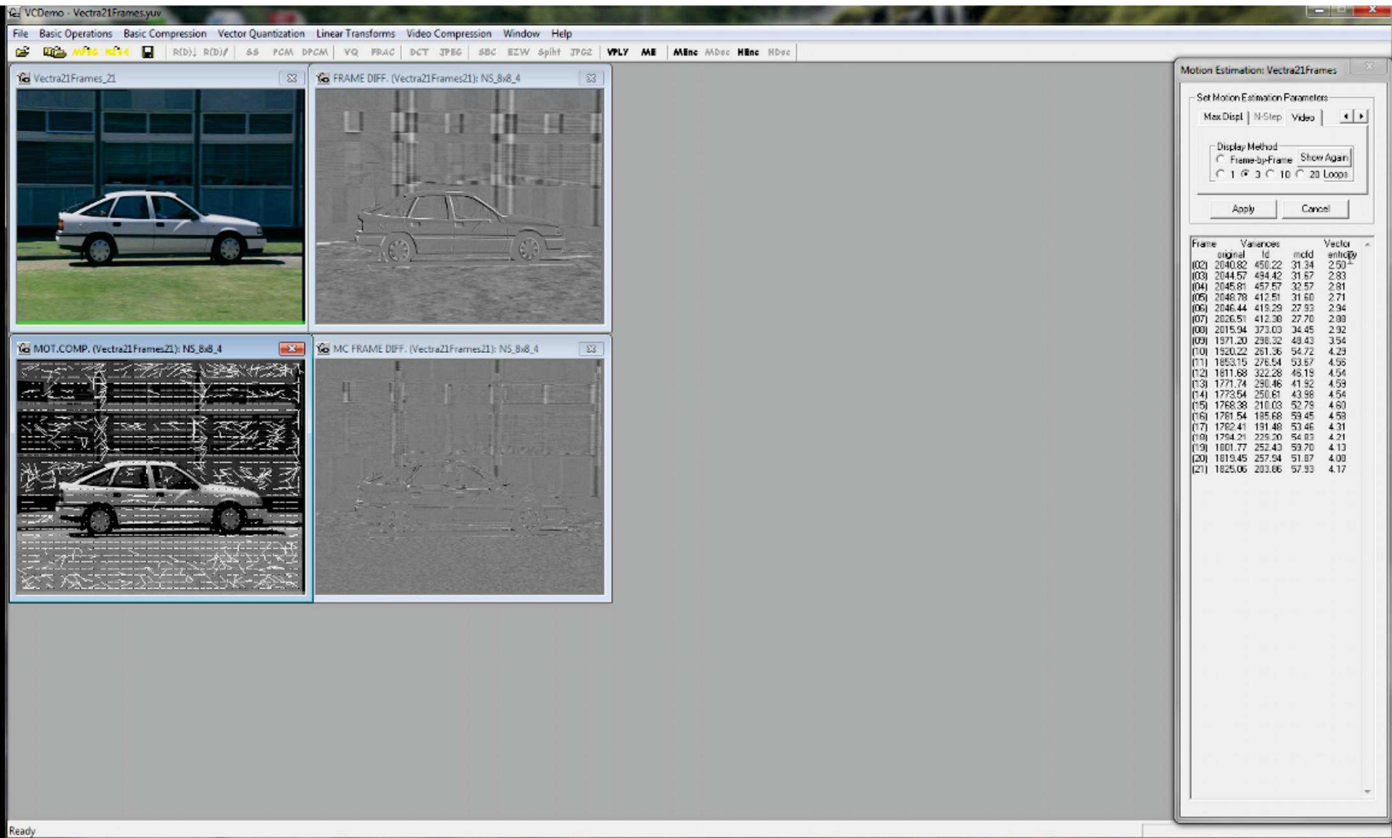
VcDemo



Tool to explore the possibilities of compression theory for images and videos

http://siplab.tudelft.nl/content/image-and-video-compression-learning-tool-vcdemo

Information and Communication Theory Group (ICT), Delft University of Technology, The Netherlands



Optical Flow Approach

Constant brightness constraint

$$I(x,y,0) = I(x+u, y+v, \tau)$$

Taylor series expansion

$$I(x+u,\cancel{y}+v,\tau) = I(x,\cancel{y},0) + \underbrace{\frac{\partial I(x,y,0)}{\partial x}}u + \underbrace{\frac{\partial I(x,y,0)}{\partial y}}v + \underbrace{\frac{\partial I(x,y,0)}{\partial t}}\tau + \mathrm{H.O.T.}$$

$$I_x u + I_y v + I_t \tau = 0$$

Or

$$\boxed{I_x V_x + I_y V_y + I_t = 0}$$

Optical Flow Approach

Consider a neighborhood of the pixel:

$$\begin{matrix} I_x(q_1)V_x + I_y(q_1)V_y = I_t(q_1) \\ I_x(q_2)V_x + I_y(q_2)V_y = I_t(q_2) \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ I_x(q_n)V_x + I_y(q_n)V_y = I_t(q_n) \end{matrix}$$

$$\underbrace{\begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ \vdots & \vdots \\ I_x(q_n) & I_y(q_n) \end{bmatrix}_{n\times 2}}_{A} \underbrace{\begin{bmatrix} V_x \\ V_y \end{bmatrix}_{2\times 1}}_{x} = \underbrace{\begin{bmatrix} I_t(q_1) \\ I_t(q_2) \\ \vdots \\ I_t(q_n) \end{bmatrix}_{n\times 1}}_{b}$$

We obtain $Ax = b$ $\quad x \rightarrow \boxed{A} \rightarrow b$

Min-norm Least-Squares Solution $A^T A x = A^T b \rightarrow x = (A^T A)^{-1} A^T b$

Regularized Solution $(A^T A + \lambda C^T C)x = A^T b \rightarrow x = (A^T A + \lambda C^T C)^{-1} A^T b$

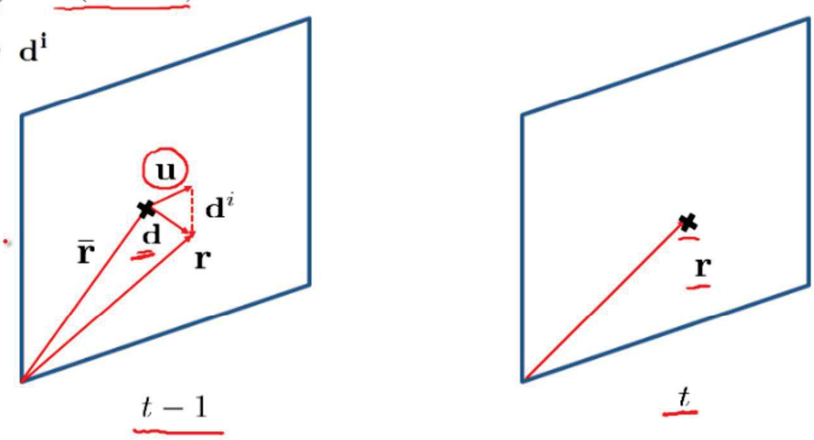
Pel-Recursive Algorithms

Constant brightness constraint

$$I(\mathbf{r},t) = I(\mathbf{r}-\mathbf{d},t-1) = I(\bar{\mathbf{r}},t-1)$$

Assume an initial estimate $\mathbf{d}^{\mathbf{i}}$

and set $\mathbf{u} = \mathbf{d} - \mathbf{d}^{\mathbf{i}}$



Pel-Recursive Algorithms

Displaced Frame Difference

$$\Delta(\mathbf{r},\mathbf{u}) = I(\mathbf{r},t) - I(\mathbf{r}-\mathbf{d}^{\mathbf{i}},t-1) = I(\mathbf{r},t) - I(\bar{\mathbf{r}}+\mathbf{u},t-1)$$

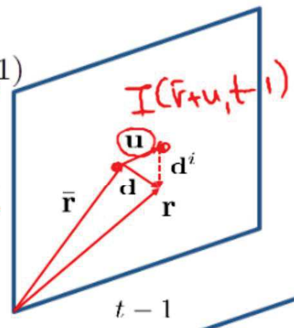
Taylor series expansion

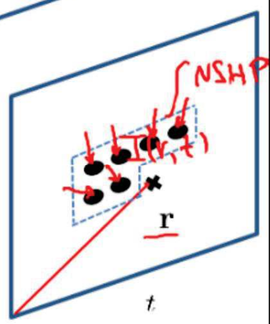
$$I(\bar{\mathbf{r}}+\mathbf{u},t-1) = I(\bar{\mathbf{r}},t-1) + \nabla^T I(\bar{\mathbf{r}}+\mathbf{u},t-1)\mathbf{u} + \epsilon(\mathbf{r},\mathbf{u})$$

$I(r,t)$

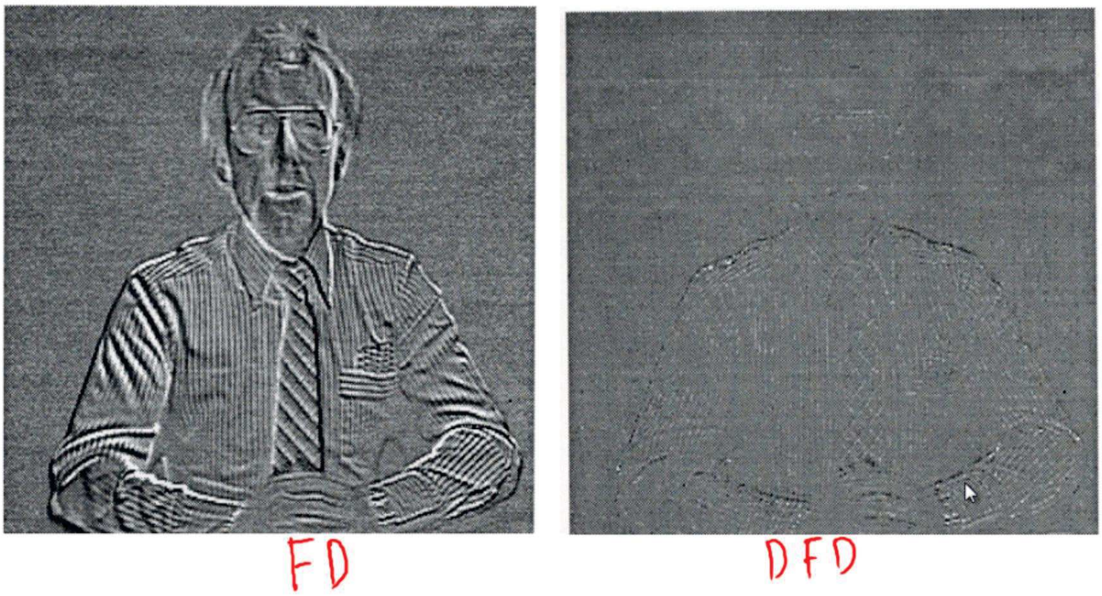
Finally $$\boxed{\Delta(\mathbf{r},\mathbf{u}) = -\nabla^T I(\mathbf{r}-\mathbf{d}^{\mathbf{i}},t-1)\mathbf{u} - \epsilon(\mathbf{r},\mathbf{u})}$$

1. recursive computability
2. $d^i \longrightarrow u$ $\quad u = d - d^i$

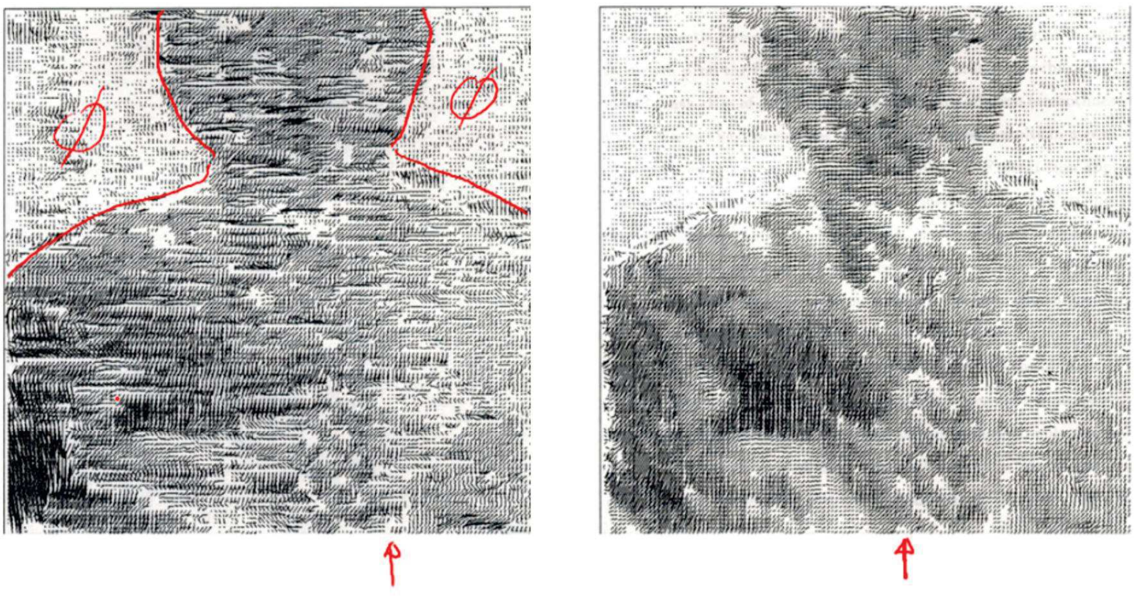




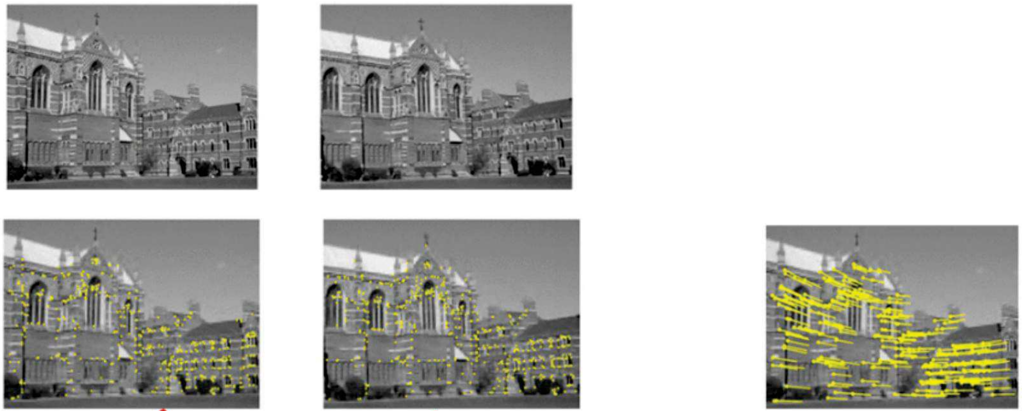
Frame Difference-- DFD



Estimated Motion Vectors



Feature-Based Methods



B. Triggs, A. Zisserman , **Feature Based Methods for Structure and Motion Estimation,** B. Triggs, A. Zisserman, R. Szeliski (Eds.): Vision Algorithms'99, LNCS 1883, pp. 278–294, 2000, Springer-Verlag Berlin Heidelberg 2000.

Harris corners, SIFT, SURF

Importance of Color

- Color is a powerful descriptor that can be used for various tasks, e.g., segmentation, object detection, tracking, and identification
- Humans can distinguish thousands of color shades and intensities, as compared to about only two dozen shades of gray