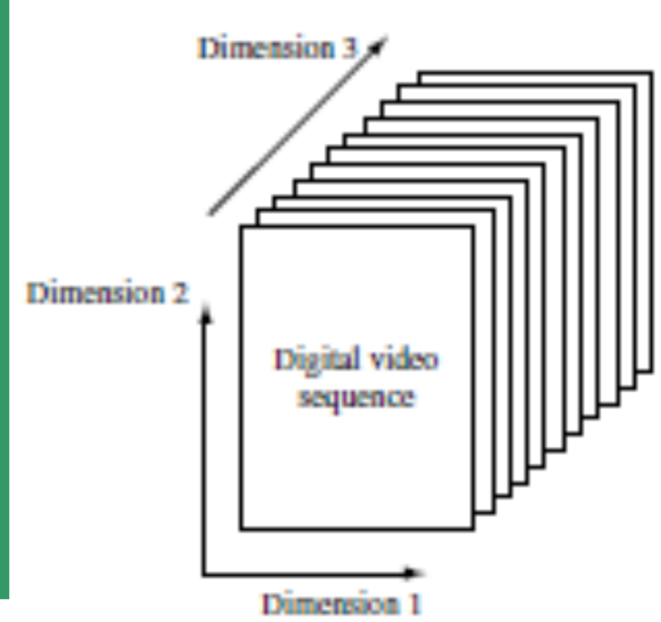
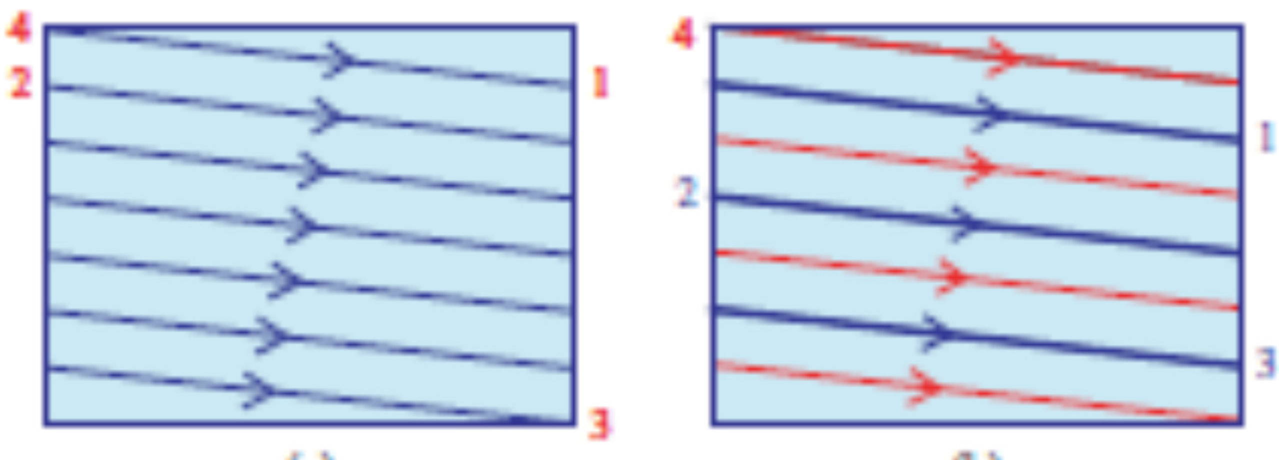
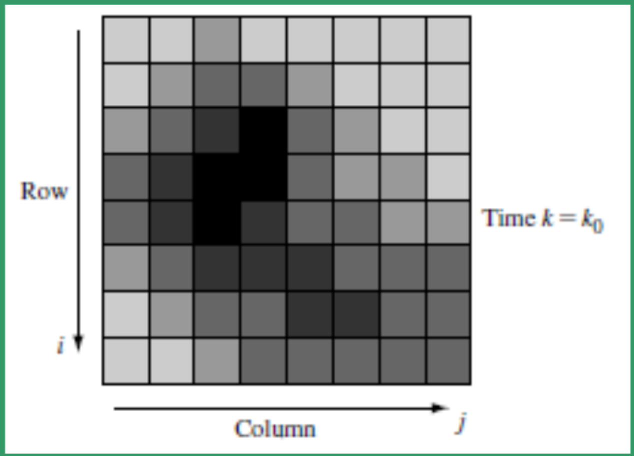
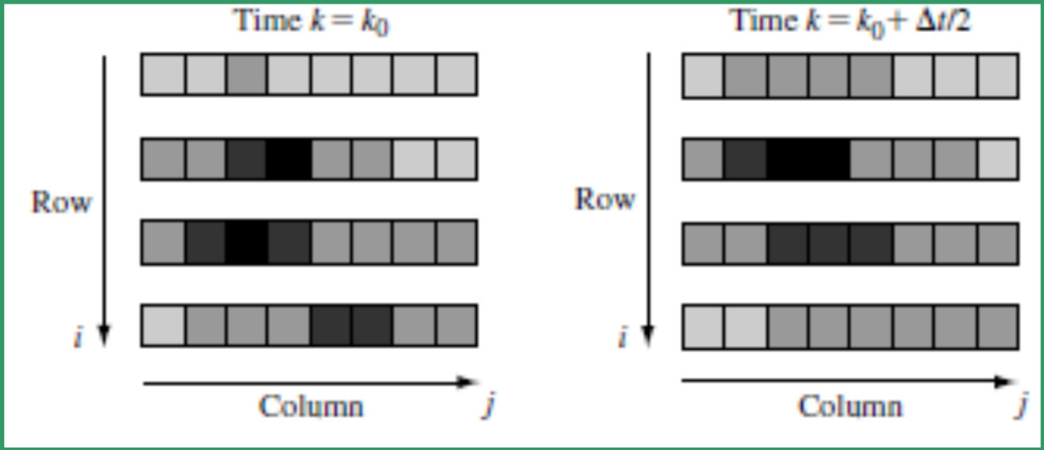

Dimension 3
Dimension 2
Digital video sequence
Dimension 1


4
2
1
3
4
1
2
3

(a) (b)


Row
i
Time $k = k_0$
Column
j


Time $k = k_0$
Row
i
Column
j
Time $k = k_0 + \Delta t/2$
Row
i
Column
j

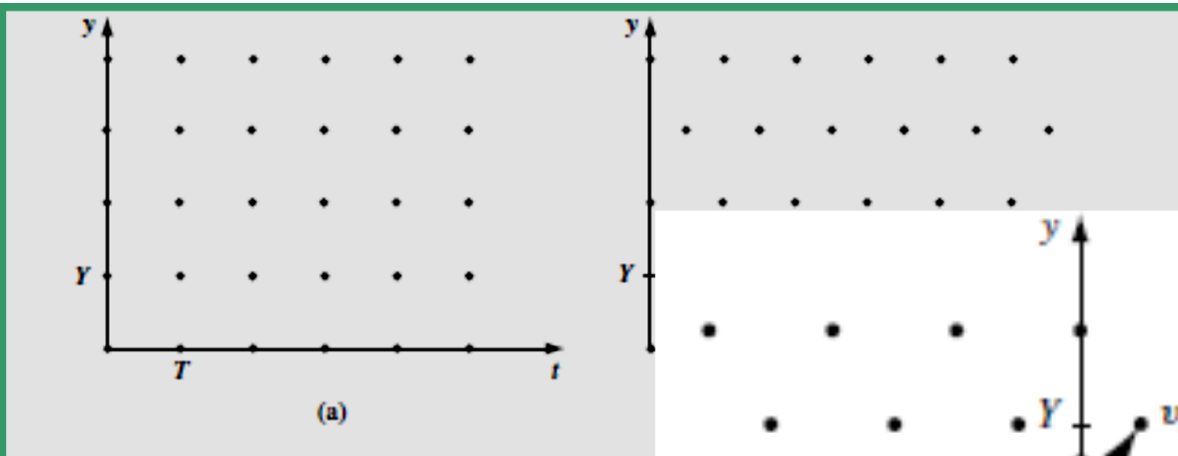
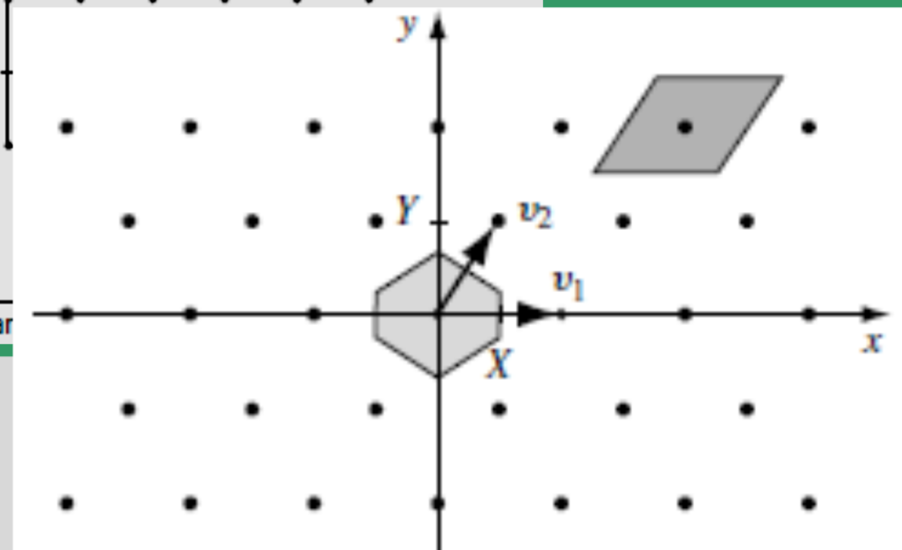


FIGURE 2.2

Two-dimensional vertical-temporal lattices. (a) Rectar



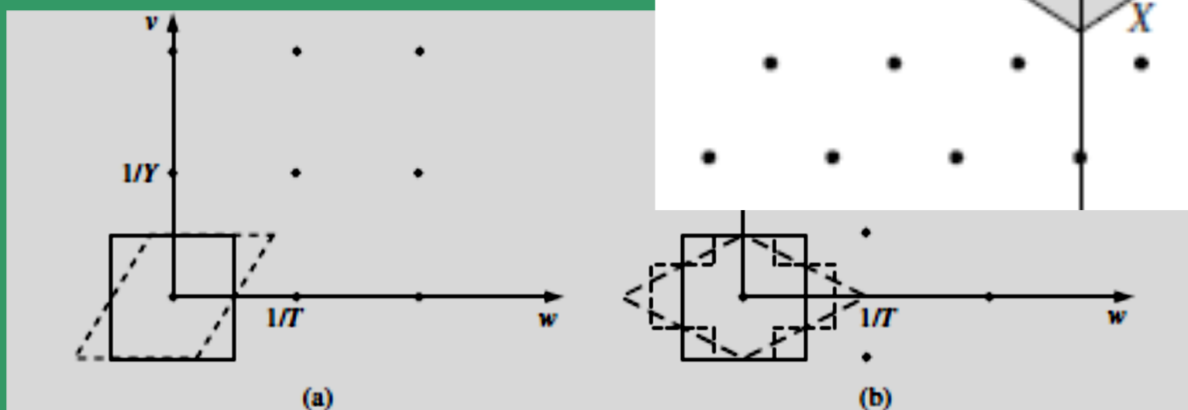
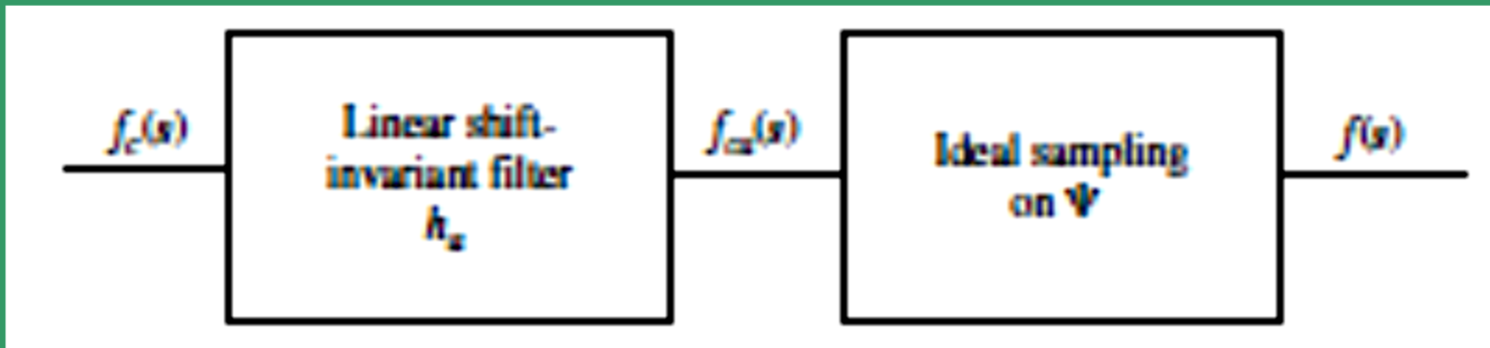


FIGURE 2.3

Reciprocal lattices of the two-dimensional vertical-temporal lattices of Fig. 2.2 with several possible unit cells. (a) Rectangular lattice Λ_R^*. (b) Hexagonal lattice Λ_H^*.

$$V_\Lambda = \begin{bmatrix} 2X & X \\ 0 & Y \end{bmatrix}$$

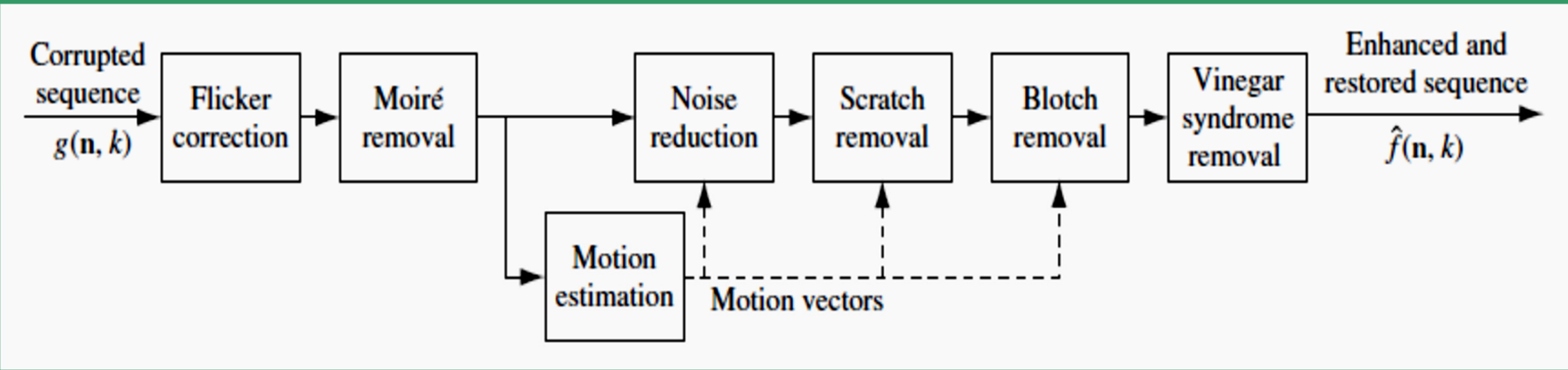


System for sampling a time-varying image.

$$\begin{bmatrix} X & 0 & 0 \\ 0 & Y & 0 \\ 0 & 0 & T \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} X & 0 & 0 \\ 0 & Y & 0 \\ 0 & T/2 & T \end{bmatrix}$$

$$f_o(\mathbf{x}, nT_2) = \sum_m f(\mathbf{c}(mT_1; \mathbf{x}, nT_2), mT_1) h(nT_2 - mT_1).$$

- Sampling lattice
- Framerate conversion;
- Resampling
- Deinterlacing
- restructuring

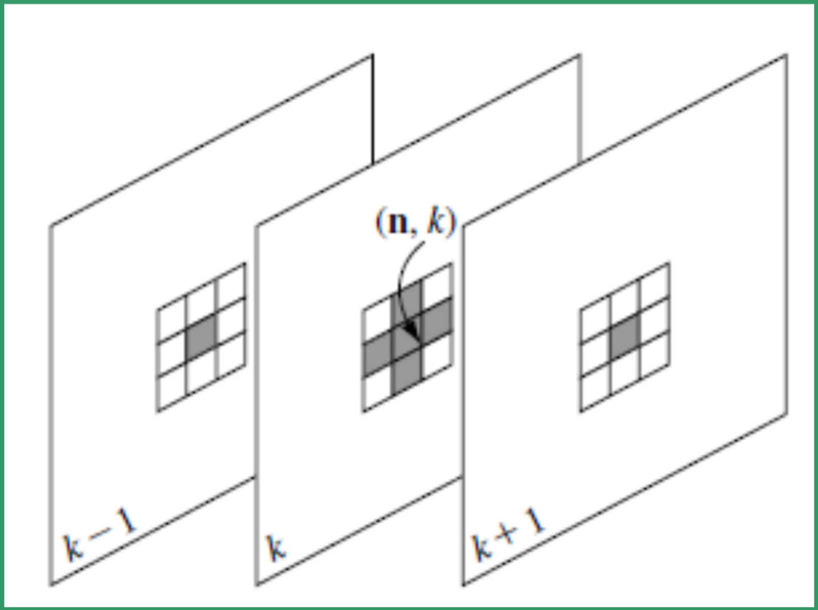


$$g(\mathbf{n},k) = f(\mathbf{n},k) + w(\mathbf{n},k)$$

SPATIOTEMPORAL NOISE FILTERING:
- **Linear filters – 3-D Wiener, recursive filtering;**
- **Order-statistical filters – multi-stage median,**
- **Multi-resolution filters**

$$\hat{f}(\mathbf{n},k) = \sum_{l=-K}^{K} h(l)g(\mathbf{n},k-l)$$

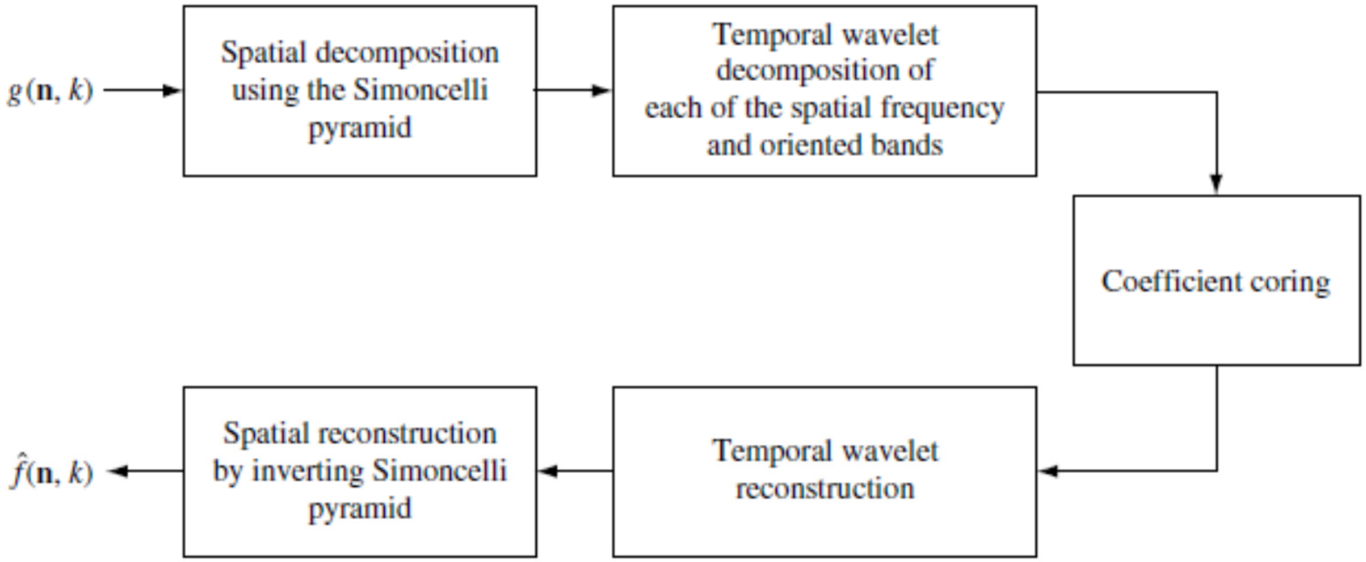
$$\hat{f}(\mathbf{n},k) = \sum_{(\mathbf{m},l)\in S} h(\mathbf{m},l)g(\mathbf{n}-\mathbf{m},k-l)$$

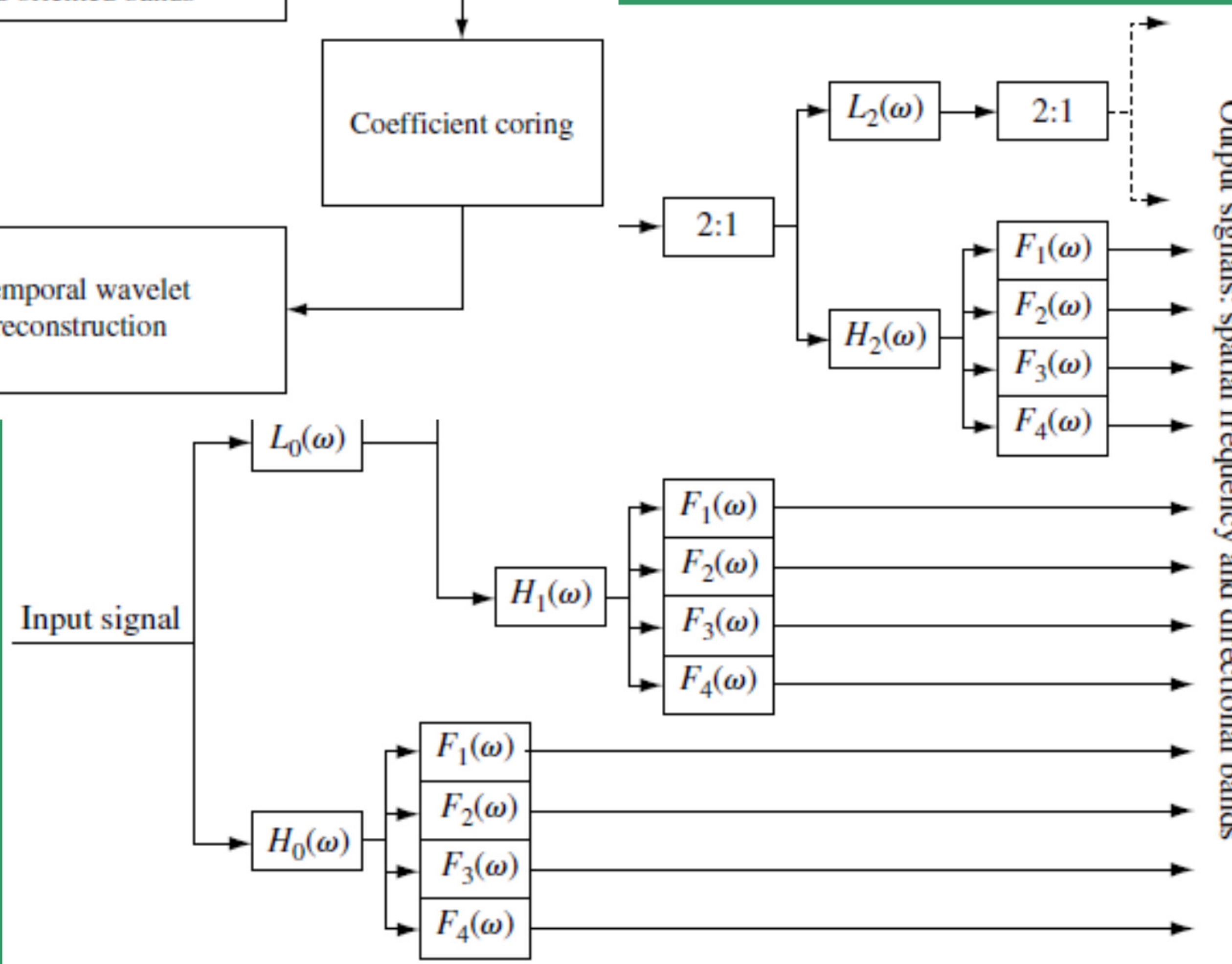


$$\hat{f}(\mathbf{n},k) = \hat{f}_b(\mathbf{n},k) + \alpha(\mathbf{n},k)\left[g(\mathbf{n},k) - \hat{f}_b(\mathbf{n},k)\right]$$

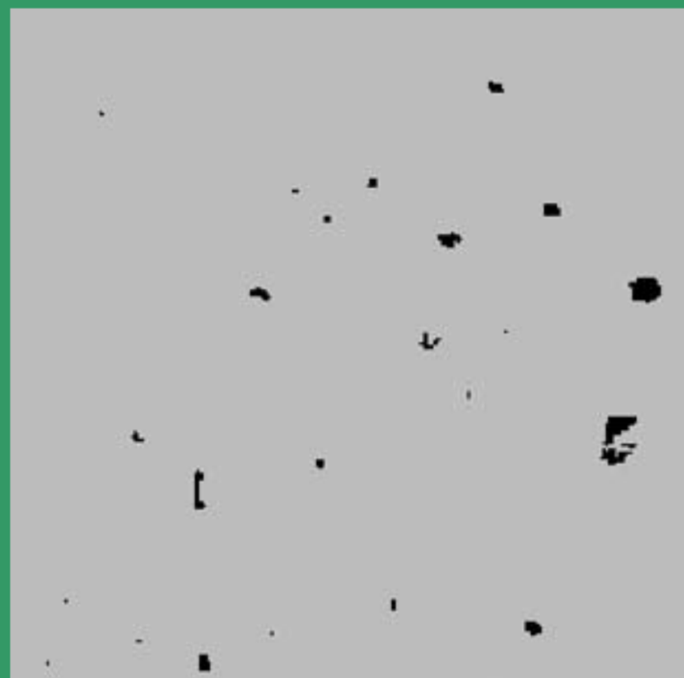
where $\hat{f}_b(\mathbf{n},k)$ is the prediction of the original *k*th frame on the basis of previously filtered frames and $\alpha(\mathbf{n},k)$ is the filter gain for updating this prediction with the observed *k*th frame.

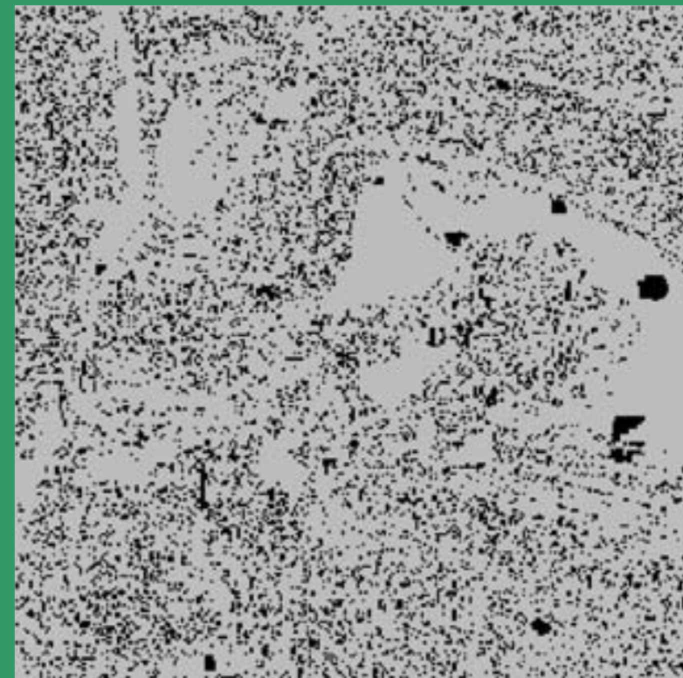
$$\alpha(\mathbf{n},k) = \max\left(1 - \frac{\sigma_w^2}{\sigma_g^2(n,k)}, 0\right)$$


$g(\mathbf{n}, k)$
Spatial decomposition using the Simoncelli pyramid
Temporal wavelet decomposition of each of the spatial frequency and oriented bands
Coefficient coring
Temporal wavelet reconstruction
Spatial reconstruction by inverting Simoncelli pyramid
$\hat{f}(\mathbf{n}, k)$


Input signal
$L_0(\omega)$
$H_0(\omega)$
$H_1(\omega)$
2:1
$L_2(\omega)$
2:1
$H_2(\omega)$
$F_1(\omega)$
$F_2(\omega)$
$F_3(\omega)$
$F_4(\omega)$
Output signals: spatial frequency and directional bands

Blotch Detection





A pixel-based blotch detector is known as the spike-detector index (SDI).

$$\mathrm{SDI}(\mathbf{n},k) = \min\left(\left(g(\mathbf{n},k) - g\left(\mathbf{n} - \mathbf{d}(\mathbf{n};k,k-1),k-1\right)\right)^2, \left(g(\mathbf{n},k) - g\left(\mathbf{n} - \mathbf{d}(\mathbf{n};k,k+1),k+1\right)\right)^2\right)$$

$$b(\mathbf{n},k) = \begin{cases} 1 & \text{if} \quad \mathrm{SDI}(\mathbf{n},k) > T \\ 0 & \quad \text{otherwise} \end{cases}$$

Rank order difference (ROD) detector

$$\mathrm{ROD}_i(\mathbf{n},k) = \begin{cases} r_i - g(\mathbf{n},k) & \text{if } g(\mathbf{n},k) \leq \mathrm{medium}(\mathbf{r}) \\ g(\mathbf{n},k) - r_{|S|-i} & \text{if } g(\mathbf{n},k) > \mathrm{medium}(\mathbf{r}) \end{cases} \quad \text{for } i = 1,2,\ldots,\frac{|S|}{2}$$

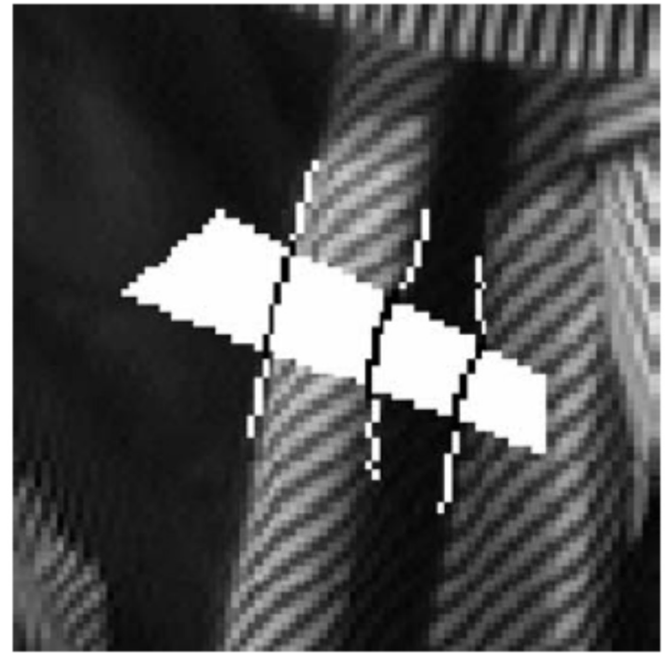
$$b(\mathbf{n},k) = \begin{cases} 1 & \text{if } \exists\, i \text{ such that } \mathrm{ROD}_i(\mathbf{n},k) > T_i \\ 0 & \quad \text{otherwise} \end{cases}$$

Simplified Rank order difference (sROD) detector

$$\mathrm{sROD}(\mathbf{n},k)=\begin{cases}\min(\mathbf{r})-g(\mathbf{n},k) & \text{if } g(\mathbf{n},k)<\min(\mathbf{r})\\ g(\mathbf{n},k)-\max(\mathbf{r}) & \text{if } g(\mathbf{n},k)>\max(\mathbf{r})\\ 0 & \text{elsewhere}\end{cases}$$

$$b(\mathbf{n},k)=\begin{cases}1 & \text{if } \mathrm{sROD}(\mathbf{n},k)>T\\ 0 & \text{otherwise}\end{cases}.$$

Video Inpainting



Original

Restoration Result

VINEGAR SYNDROME REMOVAL

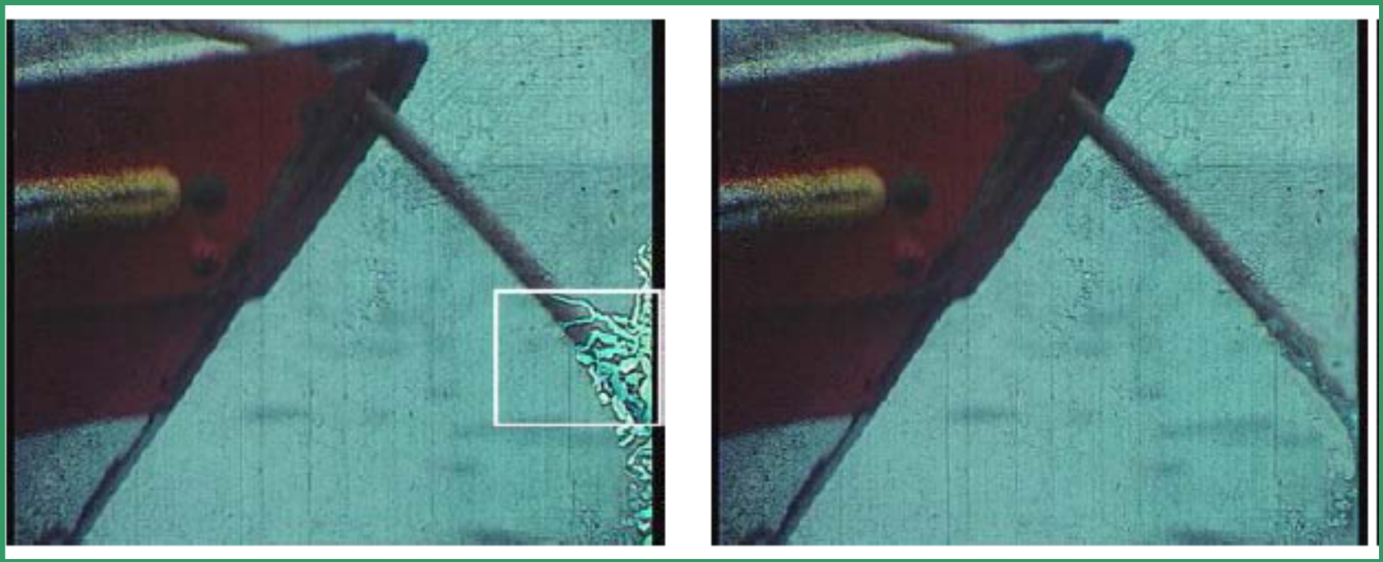
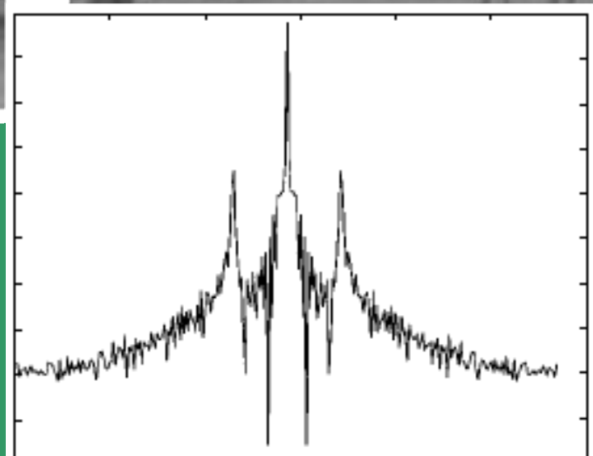
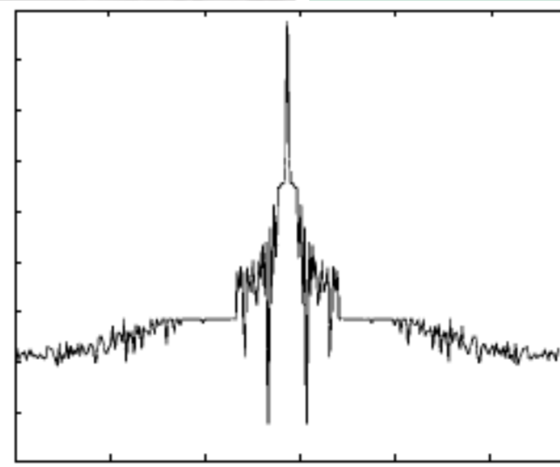


Image affected by moiré





Scratch Removal

Kalman filter, Wavelets, damped sinusoids, auto-regression models

Video Stabilization and Mosaicing

$$\begin{bmatrix} \hat{x}_{1,1} & \cdots & \hat{x}_{M,1} \\ \vdots & \ddots & \vdots \\ \hat{x}_{1,N} & \cdots & \hat{x}_{M,N} \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_N \end{bmatrix} \begin{bmatrix} \hat{X}_1 & \cdots & \hat{X}_M \end{bmatrix}$$

The above a SFM eqn. derived, for solution of the RHS.

x^ are image coordinates, forming the measurement matrix;

P_js are projection matrix, for (N) different camera positions/poses;
$P_j = K \cdot Rj$ denotes the 2*3 projection matrix of the camera.

X^ are 3-D point locations; M points

The measurement matrix is decomposed into a 2N*3 projection matrix and 3*M structure matrix (Tomasi and Kanade – 1992 – IJCV).

$$R(I_M, I_t) = D(I_M, I_t) + G(I_M, I_t),$$

$$D(I_M, I_t) = \sum_{r \in R} \left[I_M(r) - I_t(p(r;m)) \right]^2, \quad \text{and}$$

$$G(I_M, I_t) = \sum_{r \in R} \left[\nabla I_M(r) - \nabla I_t(p(r;m)) \right]^2.$$

Basic Equations used

$$\begin{pmatrix} u_{i,j} \\ v_{i,j} \end{pmatrix} = K \cdot \begin{bmatrix} R_j & T_j \end{bmatrix} \underline{X}_i$$

$$\hat{x}_{i,j} = P_j \cdot \hat{X}_i$$

$$x_{i,j} = K \cdot (R_j X_j + T_j)$$

$$P = K \cdot R \cdot \begin{bmatrix} I & -T \end{bmatrix}$$

$$x_1 = H_1 x_p,$$

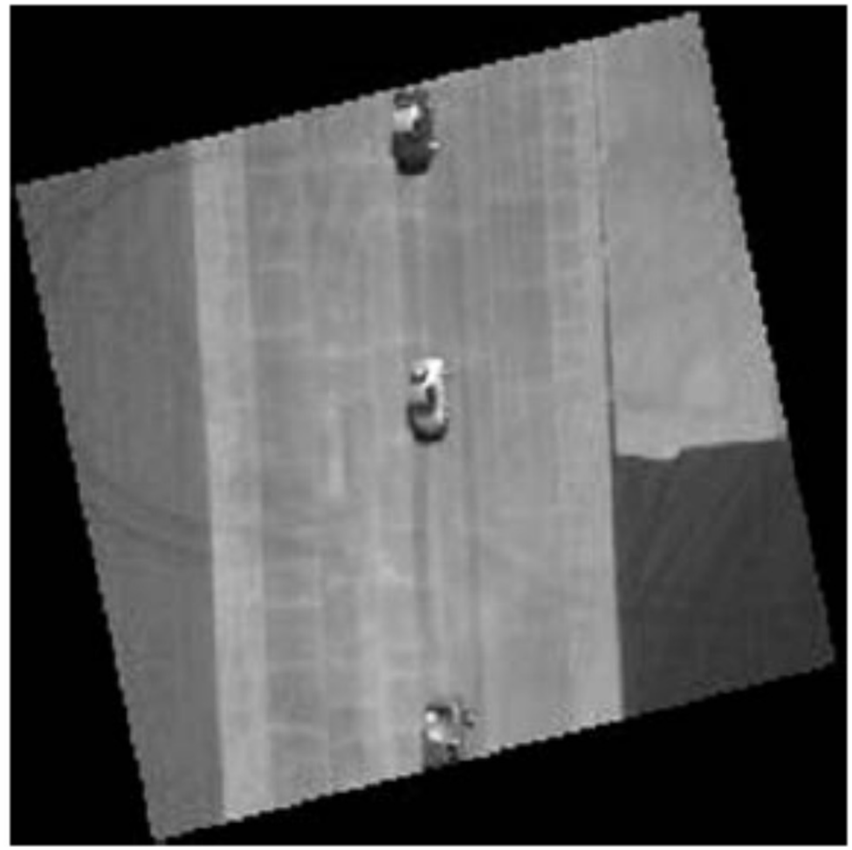
$$x_p = H_2 x_2, \quad \text{and}$$

$$x_1 = H_1 H_2 x_2 = H x_2$$

The homography between the image $\boldsymbol{I_i}$ and the mosaic is represented as $P_M = H_i \cdot P_i$, where $H_i = T_i F_i$.

F_i is the homography for ortho-correction; Or fronto-parallel view.

(a) Original view



(b) Fronto-parallel view

Methods of Video Segmentation

- Scene Change
- Spatio-temporal integration
- Dominant and multiple motion segmentation
- Motion field and MAP
- Semi-automatic

- **Spatio-temporal integration**

$$\mathrm{FD}_{k,r}(\mathbf{x}) = s(\mathbf{x},k) - s(\mathbf{x},r)$$

$$z_{k,r}(\mathbf{x}) = \begin{cases} 1 & \text{if } |\mathrm{FD}_{k,r}(\mathbf{x})| > T \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{FDN}_{k,r}(\mathbf{x}) = \frac{\sum_{\mathbf{x}\in\mathcal{N}} |s(\mathbf{x},k) - s(\mathbf{x},r)|\,|\nabla s(\mathbf{x},r)|}{\sum_{\mathbf{x}\in\mathcal{N}} |\nabla s(\mathbf{x},r)|^2 + c}$$

$$\mathrm{FDM}_k(\mathbf{x}) = s(\mathbf{x},k) - \bar{s}(\mathbf{x},k)$$

$$\bar{s}(\mathbf{x},k) = (1-\alpha)s(\mathbf{x},k) + \alpha\bar{s}(\mathbf{x},k-1), \quad k = 1,\ldots$$

$$\bar{s}(\mathbf{x},0) = s(\mathbf{x},0).$$

- **Dominant and multiple motion segmentation**

$$\begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} -\sum I_x I_t \\ -\sum I_y I_t \end{bmatrix}$$

$$R(\mathbf{x}, k) = \frac{\lambda_{\min}}{\lambda_{\max}}$$

$$\bar{s}(\mathbf{x}, k) = (1 - \alpha) s(\mathbf{x}, k) + \alpha \, warp(\bar{s}(\mathbf{x}, k - 1), s(\mathbf{x}, k)), \quad k = 1, \ldots,$$

$$\bar{s}(\mathbf{x}, 0) = s(\mathbf{x}, 0)$$

Clustering in motion parameter space

$$\bar{\eta}^2 = \sum_{\mathbf{x} \in \mathcal{B}} ||\mathbf{v}(\mathbf{x}) - \tilde{\mathbf{v}}(\mathbf{x})||^2$$

The motion parameters for blocks with acceptably small residuals are selected as the seed models.
Then, the seed model parameter vectors are clustered to find the *K* representative affine Motion models. The clustering procedure can be described as follows: given *N* seed affine parameter vectors A1,A2, . . . ,A*N*, where

$$\mathbf{A}_n = \begin{bmatrix} a_{n,1} \\ a_{n,2} \\ a_{n,3} \\ a_{n,4} \\ a_{n,5} \\ a_{n,6} \end{bmatrix}, \quad n = 1,\ldots,N, \tag{6.10}$$

find K cluster centers $\bar{\mathbf{A}}_1, \bar{\mathbf{A}}_2, \ldots, \bar{\mathbf{A}}_K$, where $K \ll N$, and the label k, $k = 1, \ldots, K$, assigned to each affine parameter vector $\mathbf{A}_n$ which minimizes

$$\sum_{n=1}^{N} \mathcal{D}(\mathbf{A}_n, \bar{\mathbf{A}}_k).$$

$$\mathcal{D}(\mathbf{A}_n, \mathbf{A}_k) = \mathbf{A}_n^T \mathbf{M} \mathbf{A}_k$$

Other methods:

- Hough Transform
- MAP based (Bayes)
- MLE (log-likelihood)
- Region based
- semi-automatic

Other Issues not covered:

- Video indexing, Retrieval browsing etc.

- Video Summarization

- Video Content classification;

- Database organization

- Active learning of semantic content from video

- Relevance feedback

- Video Quality assessment

- Current Compression standards codecs (changes faster than Android ver., or models in mobile)

- Video Communication/broadcast n LAN, wireless

- Multimodal video analysis (speech, audio, text etc. wth visual)

- Video Security (piracy etc.)