Adaptive Optimal Linear Estimators for Enhanced Motion Compensated Prediction

Ken Rose, UCSB

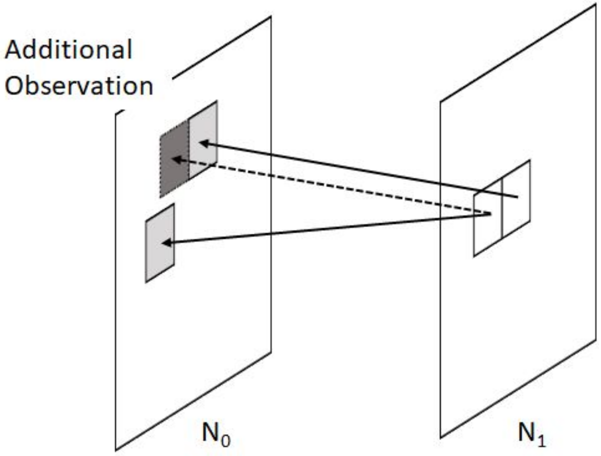
ALLIANCE FOR OPEN MEDIA RESEARCH Symposium 2019

Contributors

- Main UCSB researchers:
  - Bohan Li
  - Wei-Ting Lin
- Interaction and collaboration with Google researchers:
  - Jingning Han
  - Yaowu Xu
  - Debargha Mukherjee
  - Yue Chen
  - Angie Chiang

Background

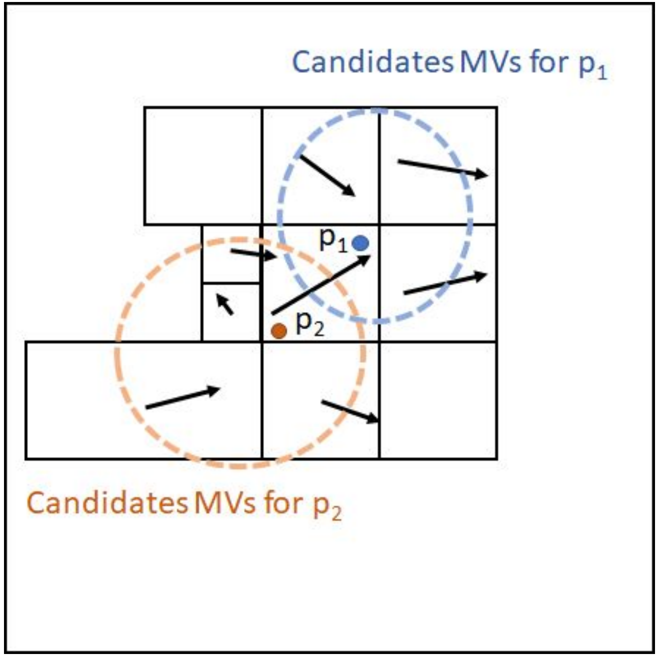
- Conventional motion compensated prediction
  - Block based
  - Only accounts for translational motion
- Motivation: nearby motion vectors point to potentially relevant observations



- Contributions to:
  - Forward prediction
  - Bi-directional prediction

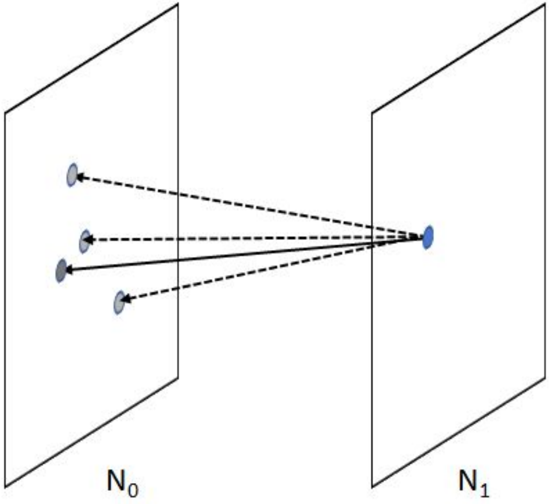
Problem Formulation

- For a given pixel, consider a set of nearby **candidate MVs**



Problem Formulation

- For a given pixel, consider a set of nearby **candidate MVs**
- They point to a set of candidate reference pixels
  - These pixels are treated as **noisy observations** of the current pixel.



ALLIANCE FOR OPEN MEDIA RESEARCH Symposium 2019

Problem Formulation

- For a given pixel, consider a set of nearby **candidate MVs**
- They point to a set of candidate reference pixels
  - These pixels are treated as **noisy observations** of the current pixel.
- Construct **optimal linear estimators** to estimate current pixels from their "noisy observations"

Optimal linear estimator

- The observation vector corresponding to M candidate MVs:

$$\mathbf{x} = (x_1, x_2, \dots, x_M)^T$$

- Linear estimator :

$$\tilde{y} = \mathbf{w}^T \mathbf{x}$$

- The optimal weights satisfy

$$E\{\mathbf{x}\mathbf{x}^T\}\mathbf{w} - E\{\mathbf{x}y\} = 0$$

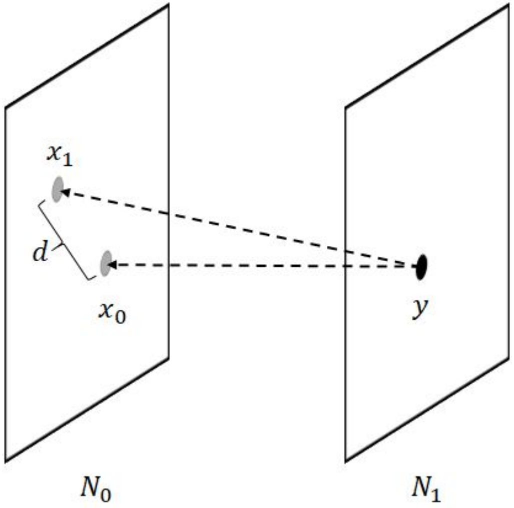
Autocorrelation Matrix

Cross-correlation vector

Auto-Correlation Matrix Estimation

- Each matrix entry specifies the correlation between two pixels in the reference frame
- Modeling pixels in a frame as a first-order Markov process with spatial correlation coefficient ρ_s:

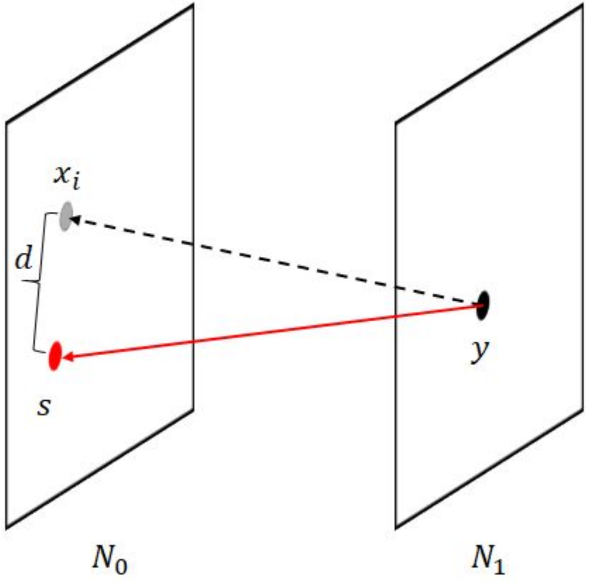
$$E\{x_0 x_1\} = \sigma^2 \rho_s^{||mv_0 - mv_1||}$$



Cross-Correlation Vector Estimation

- Let the true (unknown) motion map pixel y to location s in the reference frame
  - A separable temporal-spatial Markov model yields the cross-correlation:

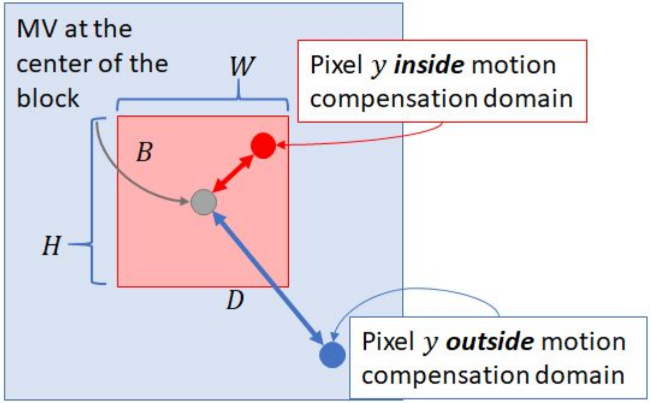
$$E\{x_i y\} = \sigma^2 \rho_t \rho_s^{d(s,x_i)}$$

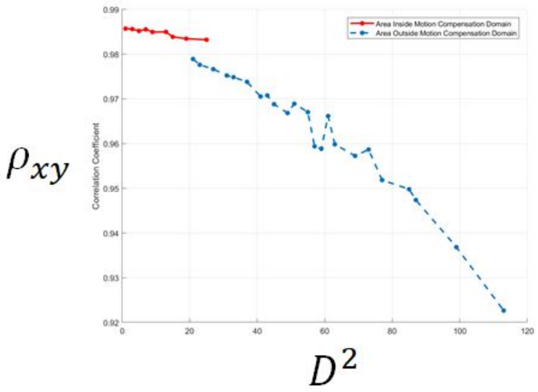


For simplicity we will assume that $\boldsymbol{\rho}_t=1$

Cross-Correlation Vector Estimation

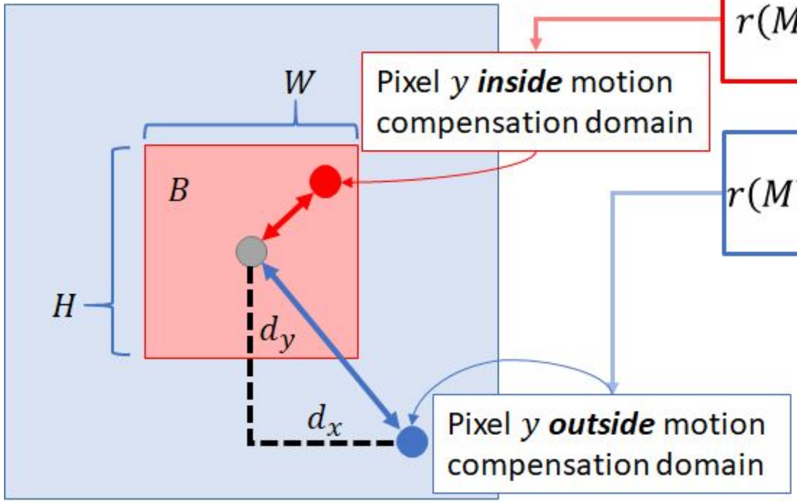
- Naturally, we do not known the true motion and hence *s*
- We need a subterfuge to calculate the cross correlation
- Observation:
  - x_i is derived from motion vector MV_i
  - $E\{x_i y\}$ decays with distance between y and the location of MV_i in the current frame





Cross-Correlation Vector Estimation

- Model the decay of motion vector reliability with distance:



$$r(MV_i, y) = \exp\left\{-\alpha\left[\left(\frac{d_x}{W}\right)^2 + \left(\frac{d_y}{H}\right)^2\right]\right\}$$

$$r(MV_i, y) = \beta\exp\left\{-\alpha\left[\left(\frac{d_x}{W}\right)^2 + \left(\frac{d_y}{H}\right)^2\right]\right\}$$

$0 < \beta < 1$ is the penalty for crossing boundary

Define the probability that MV_i is the true motion vector for pixel y:

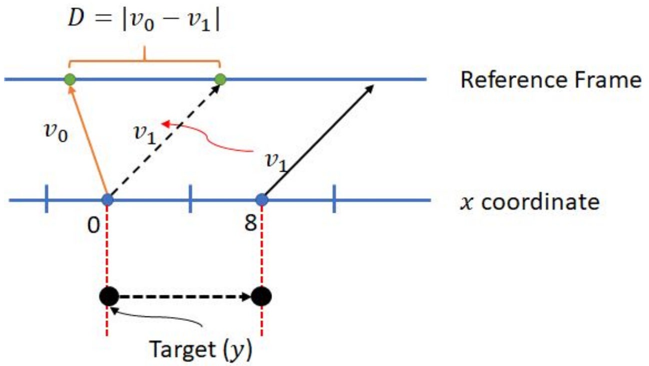
$$p_i(y) = \frac{r(MV_i, y)}{\sum_j r(MV_j, y)}$$

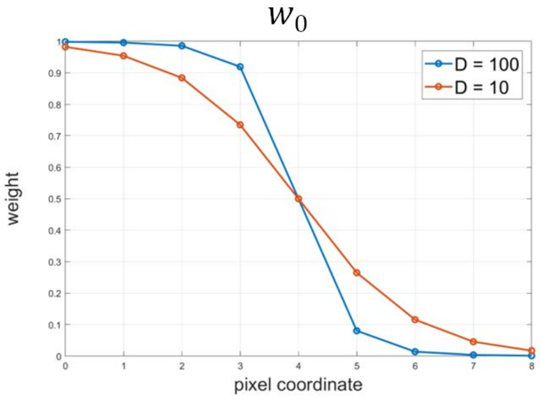
Cross-Correlation Vector Estimation

- The expected distance between observation x_i and s, the true motion compensated location of y on the reference frame:

$$E\{d(s, x_i)\} = \sum_j p_j(y)||mv_i - mv_j||$$

- The cross-correlation: $E\{x_i y\} = \sigma^2 \rho_t \rho_s^{E\{d(s,x_i)\}}$
- 1-D example weight distribution

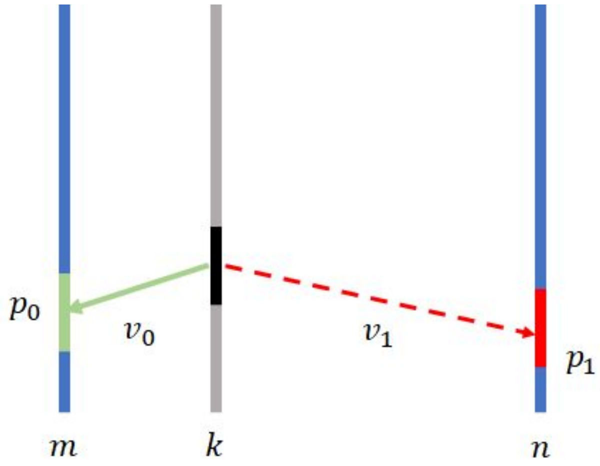




Considering Multiple Reference Frames

- Neighboring motion vectors may point to different reference frames
    - Observations are now not on the same reference frame so the spatial model cannot be directly applied

- Idea: for the autocorrelation matrix:
    - Find a common past frame



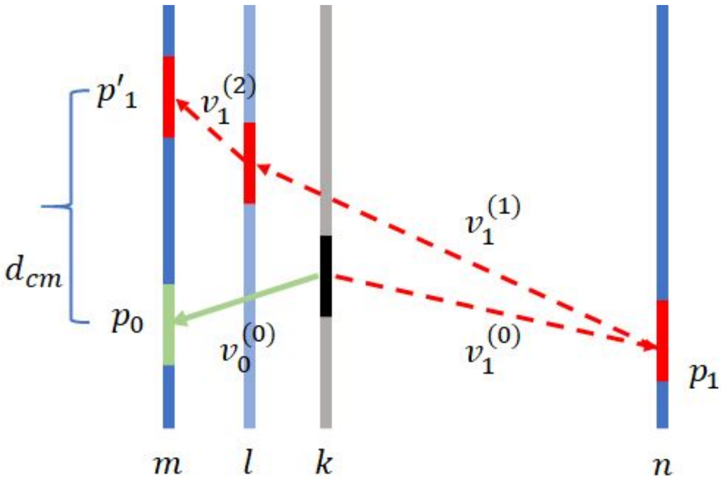
Considering Multiple Reference Frames

Track along the motion vector chain

- Locate the first common reference frame where both observations have precursors
- Since

$$p_1' \cong p_1\,,\;\; \rho_t \cong 1$$

$$E\{p_0 p_1\} \cong E\{p_0 p_1'\} = \sigma^2 \rho_s^{d_{cm}}$$



$$d_{cm}\left(v_0^{(0)}, v_1^{(0)}\right) = \left\| \sum_{i=0}^{m^*} v_0^{(i)} - \sum_{j=0}^{n^*} v_1^{(j)} \right\|$$

where $v_0^{m^*}$ and $v_1^{n^*}$ point to the same reference frame

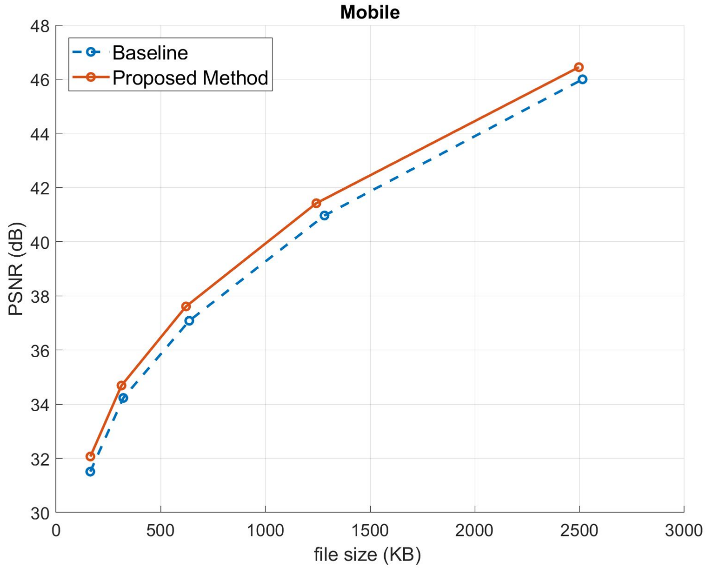
Overall Motion Compensated Prediction

- Derive the optimal per pixel linear estimators
  - Based on the estimated cross-correlation vector and auto-correlation matrix
- Employ the estimators to form the current frame prediction
  - Prediction coefficients account for the distance between and difference in nearby MVs
    - Automatically adapts to local variations

Experimental Results

- Single reference frame per block (no compound mode)
  - Significant performance improvement

coastguard	-5.08
foreman	-5.24
flower	-8.73
mobile	-10.90
bus	-6.14
stefan	-6.01
BlowingBubbles	-7.32
BQSquare	-12.57
Average	-7.75%



Compound Mode Enabled -Work in Progress..

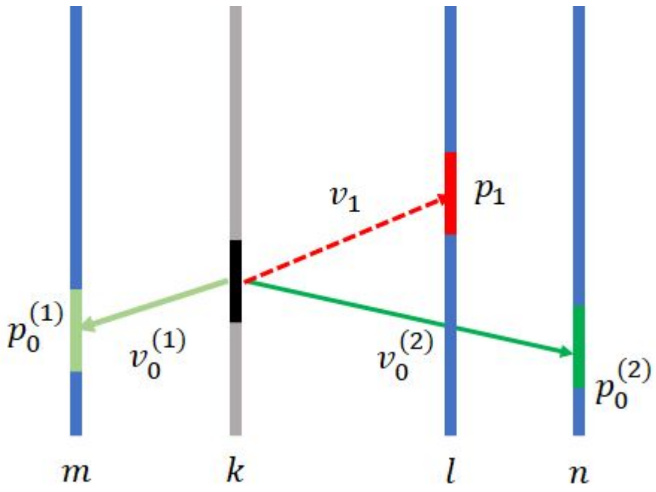
- Compound Prediction is defined as

$$p_0 = w^{(1)}p_0^{(1)} + w^{(2)}p_0^{(2)}$$

- We define distance between p_0 *and* p_1

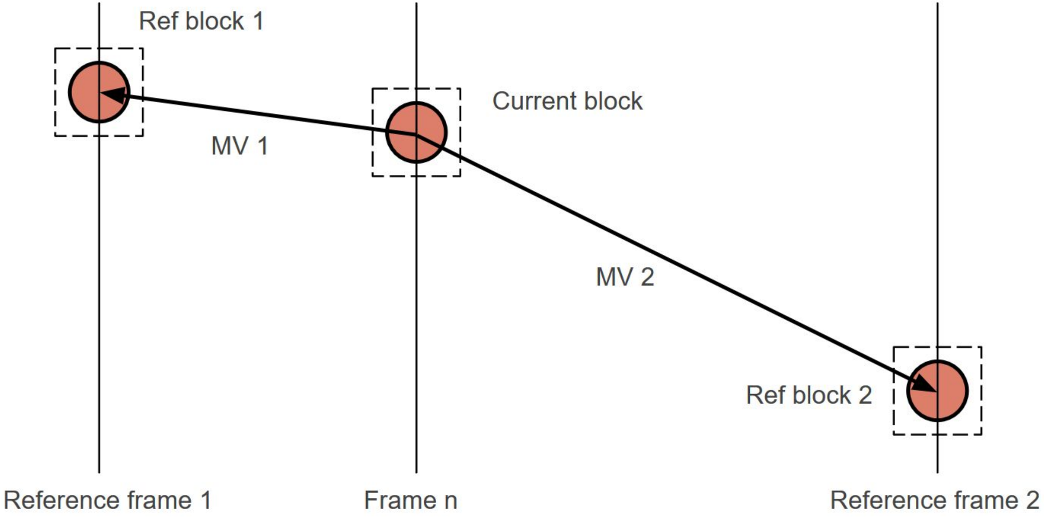
$$d_{cm}(p_0, p_1) = \mathrm{w}^{(1)}\, d_{cm}\left(p_0^{(1)}, p_1\right) + \mathrm{w}^{(2)}\, d_{cm}\left(p_0^{(2)}, p_1\right)$$

- Use "as is" the parameter set from the single reference frame setting
- Preliminary result
  - Average BD-Rate reduction ~ 1.9%
  - Performance expected to improve significantly once actually optimized/trained for compound mode



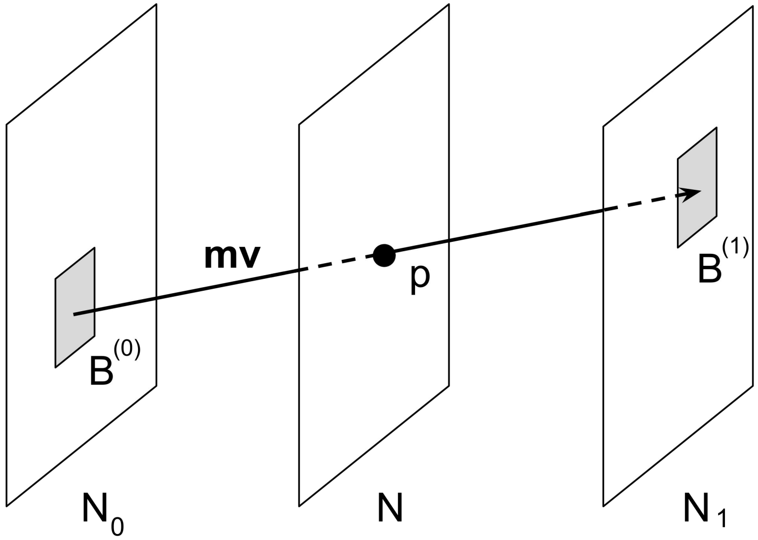
Bi-directional Prediction Background

- Conventional bi-directional motion compensated prediction
  - Block based
  - Only accounts for translational motion
- Important observation: redundancy in motion vectors



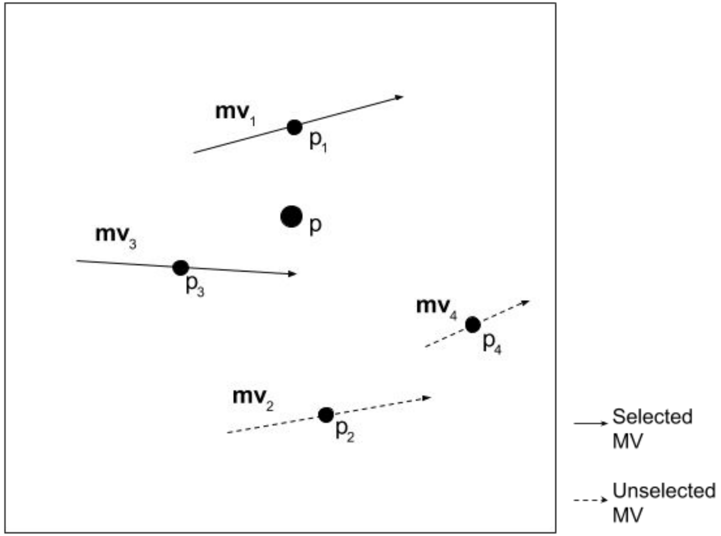
Bi-directional Prediction Background

- "Free" motion information is already available to the decoder
  - Previously decoded MVs
  - Viewed as intersecting the current frame



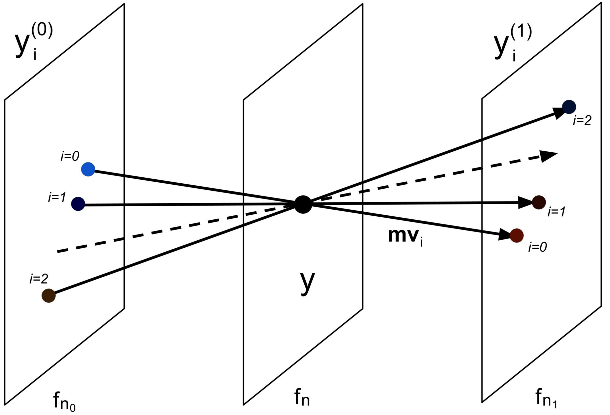
Problem Formulation

- For the current pixel, identify projected MVs that intersect the frame near the current pixel
  - These are the **candidate MVs**



Problem Formulation

- For the current pixel, identify projected MVs that intersect the frame near the current pixel
  - These are the **candidate MVs**
- Use the candidate MVs to obtain pairs of reference pixels
  - By applying the MVs to the current pixel
  - These reference pixels are viewed as **noisy observations**



Problem Formulation

- For the current pixel, identify projected MVs that intersect the frame near the current pixel
  - These are the **candidate MVs**
- Use the candidate MVs to obtain pairs of reference pixels
  - By applying the MVs to the current pixel
  - These reference pixels are viewed as **noisy observations**
- Construct **optimal linear estimators** to estimate current pixels from their "noisy observations"

Optimal linear estimator (déjà vu)

- The observation vector corresponding to M candidate MVs:

  $$\mathbf{x} = (x_1, x_2, ..., x_M)^T$$

- Linear estimator :

  $$\tilde{y} = \mathbf{w}^T \mathbf{x}$$

- The optimal weights satisfy

$$\underbrace{E\{\mathbf{x}\mathbf{x}^T\}}_{\text{Autocorrelation Matrix}}\mathbf{w} - \underbrace{E\{\mathbf{x}y\}}_{\text{Cross-correlation vector}} = 0$$

Cross-Correlation Vector Estimation

- Consider the **true motion trajectory**
- Form a Markov chain

  $s^{(0)}$ - y - $s^{(1)}$

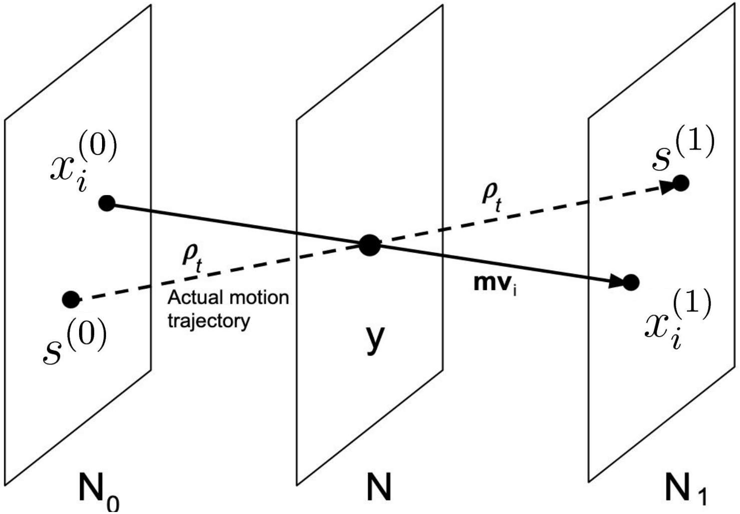
  - With temporal correlation ρ_t

$$y = \rho_t s^{(0)} + n$$

$$s^{(1)} = \rho_t y + n^{(1)}$$

- Between the references, we get:

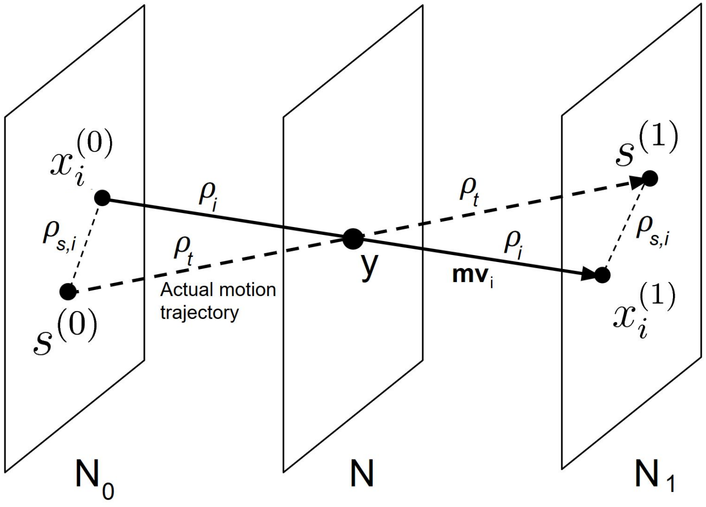
$$\rho_{01} = \frac{E\{s^{(0)} s^{(1)}\}}{\sigma_y^2} = \rho_t^2$$



ALLIANCE FOR OPEN MEDIA RESEARCH Symposium 2019

Cross-Correlation Vector Estimation

- Consider next the **candidate motion vectors**
  - The temporal-spatial separable Markov model
    $\rho_i = \rho_t \rho_{s,i}$
- Similarly, also form a Markov chain
  - When spatial correlation decays exponentially with distance
    $\rho_{01,i} = \rho_i^2$
    $x_i = 0.5x_i^{(0)} + 0.5x_i^{(1)}$
    $E\{x_i y\} = \rho_i \sigma^2 = \sqrt{\rho_{01,i}}\sigma^2$
- Need $\rho_{01,i}$ for the cross correlation
  - By collecting data in the neighboring area of $x_i^{(0)}$ and $x_i^{(1)}$



Estimation of Auto-Correlation Matrix

- $E\{x_i y\} = \rho_i \sigma^2$ write observation as:

  $x_i = \rho_i y + \boxed{z_i}$ “Innovation” part that is uncorrelated with y

- Autocorrelation:

  $$E\{x_{i_1} x_{i_2}\} = \rho_{i_1}\rho_{i_2}\sigma^2 + E\{z_{i_1} z_{i_2}\}$$

- Need $E\{z_{i_1} z_{i_2}\}$
  - The “exponential decay” model
    - Correlation decays with distance

      $$E\{z_{i_1} z_{i_2}\} = exp(-\alpha||\Delta\mathbf{mv}_{i_1,i_2}||)\sigma_{z_1}\sigma_{z_2}$$

      where

      $$\sigma_{z_i}^2 = (1-\rho_i^2)\sigma^2$$

Co-Located Reference Frame

- Obtained the optimal linear estimator
  - Based on the estimated cross-correlation and auto-correlation
- Note: estimate assumes linear motion for MV intersection with current frame
  - Motion offsets degrade the prediction quality
- Solution: use the optimal linear estimate as a “reference frame”
  - Largely co-located with the current frame
  - Offset is eliminated by standard motion compensation
  - Co-located frame also proposed in prior work, albeit at high complexity
    - Generated by extensive optical flow estimation from reconstructed frames

Experimental Results

- Significant performance improvement
  - Complexity much lower - circumvent extensive motion estimation

Sequence	BD-rate change (%)
coastguard	-1.53
mobile	-4.37
foreman	-7.17
flower	-3.19
bus	-2.96
BasketballPass	-4.81
container	-3.95
tempete	-1.30
Average	**-3.75**