Super-Resolution With Fuzzy Motion Estimation

Matan Protter & **Michael Elad**
Computer-Science Department
The Technion - Israel

Peyman Milanfar & **Hiro Takeda**
Electrical Engineering Department
UC Santa-Cruz - USA

SIAM Conference on Imaging Science
Session on Locally Adaptive Patch-Based Image and Video Restoration – Part II
July 9th, 2008 San-Diego

SIAM Conference on
IMAGING SCIENCE
July 7-9, 2008
Town and Country Resort
& Convention Center
San Diego, CA

Motivation

Low-quality video sequences are quite common: webcams, cellular phones, security cameras, …

Super-Resolution could (in principle) reconstruct **better optical resolution** from these sequences, but …

This reconstruction requires highly **accurate motion** estimation

The implication: classical Super-resolution algorithms are limited to handle sequences with **global motion**

Can we bypass this limitation?

Motivation

Yes, we can!

In this talk we present a new **Super-Resolution Reconstruction (SRR)** algorithm that relies on fuzzy (probabilistic) motion estimation, and can work on arbitrary image sequences

Agenda

1. **Super-Resolution (SR) – Introduction**

   The model, the classic approach, and the limitations

2. **The Concept of Fuzzy Motion**

   The idea, who uses it, and why

3. **The Proposed SR Algorithm**

   How does fuzzy motion fit in? the evolved algorithm

4. **Results**

   Several videos, and conclusions

Agenda

1. **Super-Resolution (SR) – Introduction**

   The model, the classic approach, and the limitations

2. **The Concept of Fuzzy Motion**

   The idea, who uses it, and why

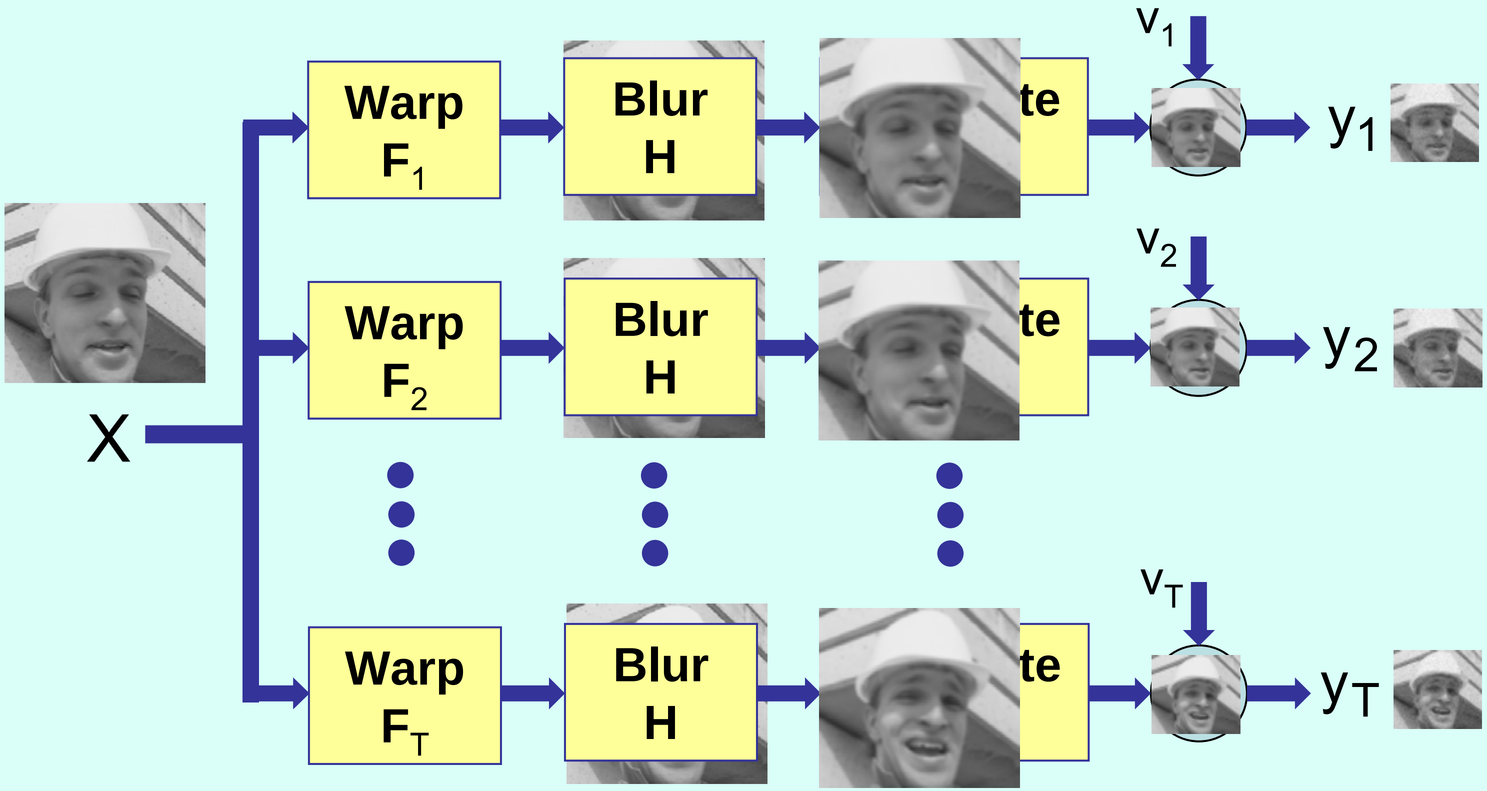
3. **The Proposed SR Algorithm**

   How does fuzzy motion fit in? the evolved algorithm

4. **Results**

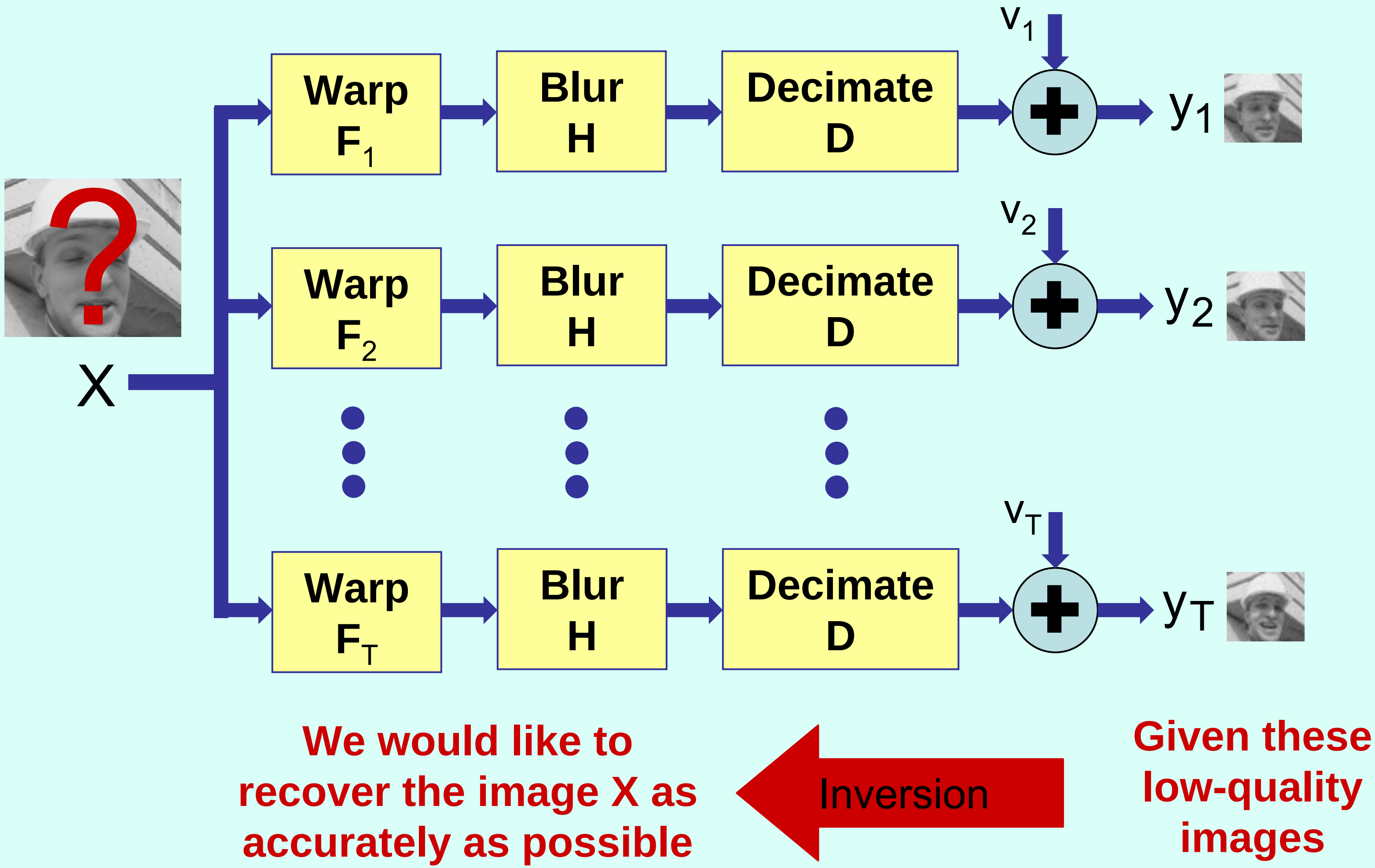
   Several videos, and conclusions

The Imaging Model



$$\left\{ y_t = \mathbf{DHF}_t X + v_t \right\}_{t=1}^{T}$$

Super-Resolution Reconstruction (SRR)



Super-Resolution Reconstruction (SRR)

- The model we have is: $\left\{ y_t = \mathbf{DHF}_t X + v_t \right\}_{t=1}^{T}$

- Define the desired image as the minimizer of the following function:

$$\hat{X} = \min_{X} \sum_{t=1}^{T} \left\| \mathbf{DHF}_t X - y_t \right\|_2^2$$

- Iterative solvers can be applied for this minimization, and their behavior is typically satisfactory, BUT …

- Solving the above requires the knowledge of:
  - **D** – a common decimation operation,
  - **H** – A common blur operation, and
  - $\mathbf{F}_t$ – the warp operators, relying on exact motion estimation.

Super-Resolution Reconstruction (SRR)

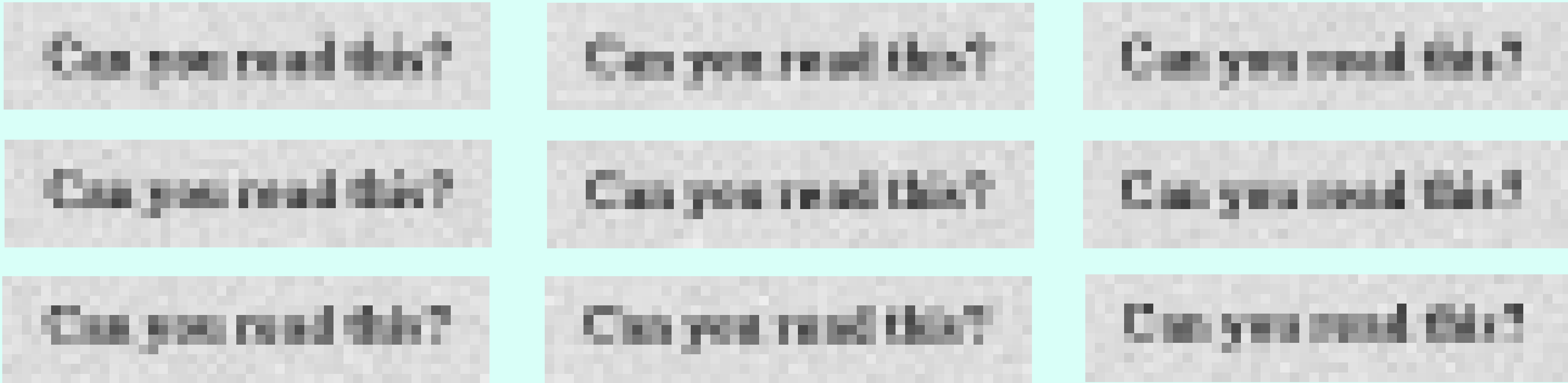
- The model we have is: $\left\{y_t = \mathbf{DHF}_t X + v_t\right\}_{t=1}^{T}$

Since the warp operators, $\mathbf{F}_t$, are hard to obtain in general, SRR algorithms are typically limited to sequences having **global motion** characteristics.

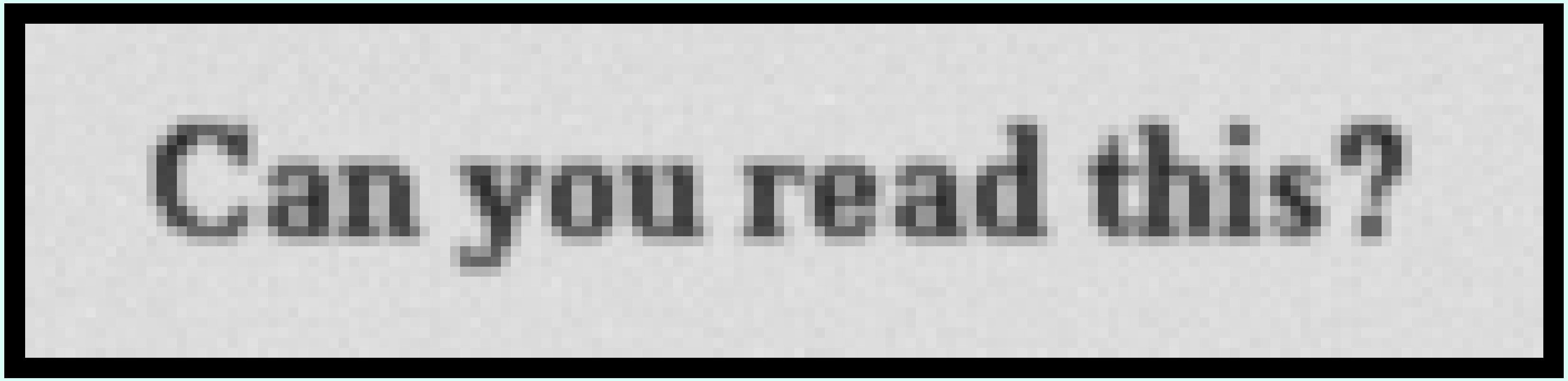
Is there no hope for sequences with general motion?

- **H** – A common blur operation, and
- F_t – the warp operators, relying on exact motion estimation.

SRR – Just a Small Example



3:1 scale-up in each axis using 9 images, with pure global translation between them



Agenda

1. **Super-Resolution (SR) – Introduction**

   The model, the classic approach, and the limitations

2. **The Concept of Fuzzy Motion**

   The idea, who uses it, and why

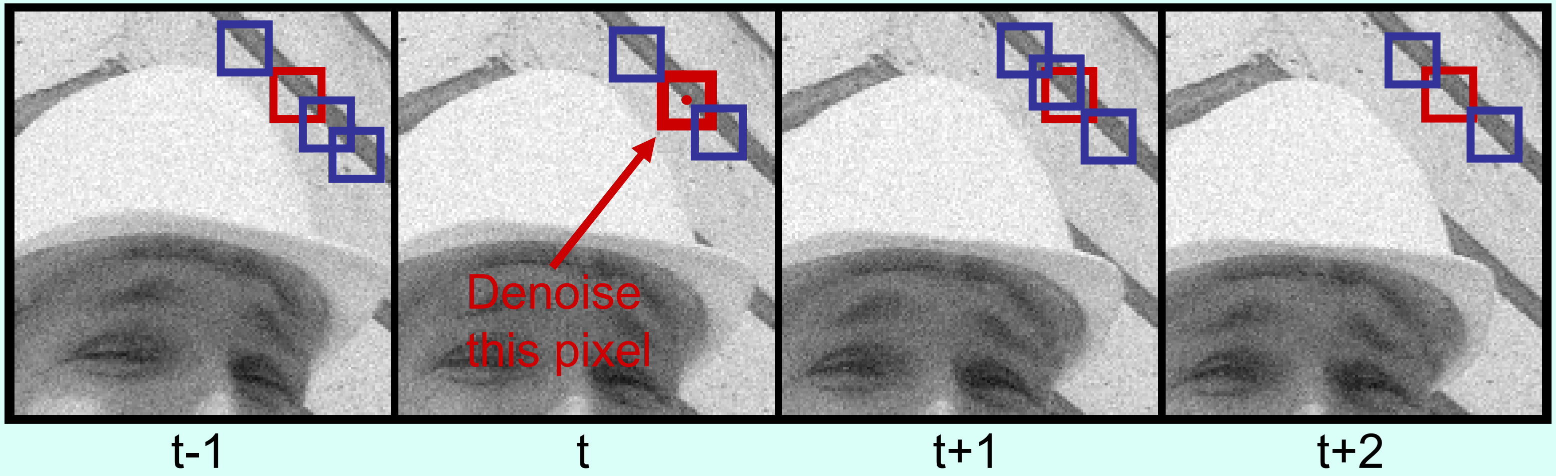
3. **The Proposed SR Algorithm**

   How does fuzzy motion fit in? the evolved algorithm

4. **Results**

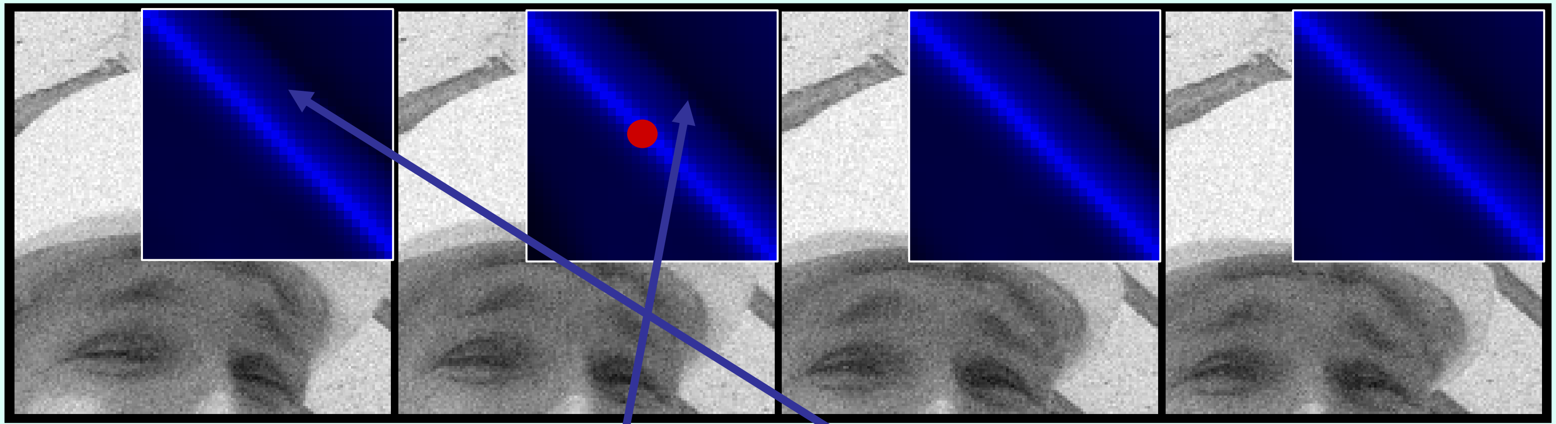
   Several videos, and conclusions

The Core Intuition



- Classic approach: Average the pixels along the motion trajectories.
- Practically: (i) Find the corresponding areas in the other images, and
  (ii) Average the center pixels in these patches.
- Alternative approach: exploit spatial redundancy, i.e., use other relevant patches as well.
- Using more relevant patches implies stronger noise suppression.

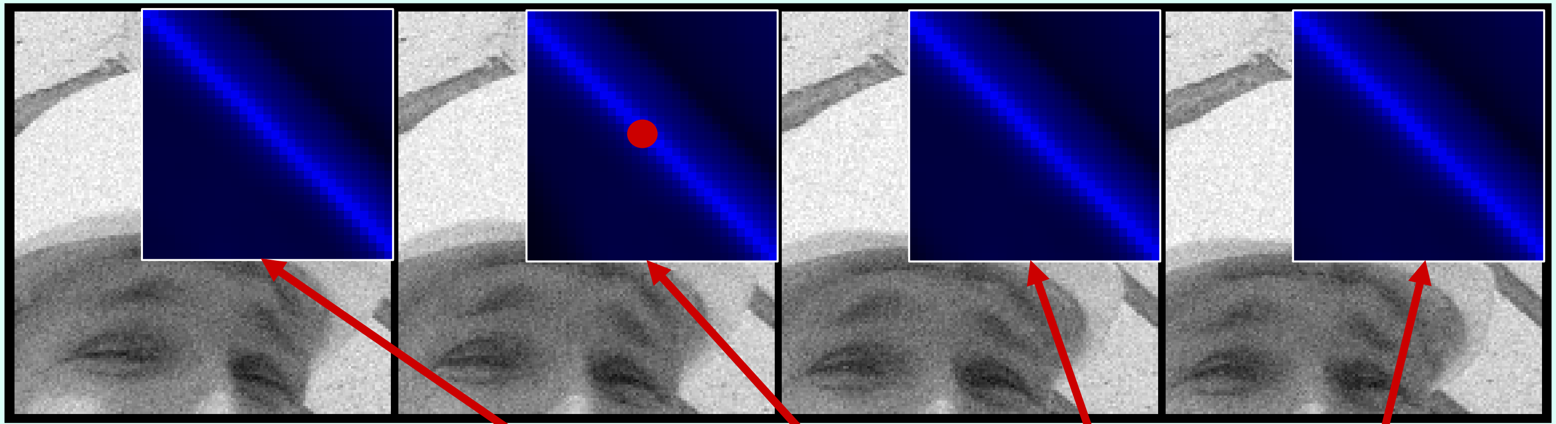
Fuzzy Motion Estimation



This idea could be interpreted as **fuzzy motion**:

- Traditionally: the pixel y[m,n,t] is tied to it's origin y[m-dm,n-dn,t-1].

Fuzzy Motion Estimation



t-1 t t+1 t+2

This idea could be interpreted as **fuzzy motion**:

- Traditionally: the pixel y[m,n,t] is tied to it's origin y[m-dm,n-dn,t-1] .
- Fuzzy approach: y[m,n,t] is tied to ALL pixels in its 3D neighborhood y[m-dx,n-dy,t-dt] for -D≤dx,dy,dt≤D, with a confidence weight (i.e. relative probability) **w[m,n,t,dm,dn,dt]** .

Our Inspiration: Image Sequence Denoising

- Classic video denoising methods estimate motion trajectories and filter along them, i.e. relaying strongly on optical flow estimation.
- A recent group of algorithms presents a new trend of avoiding explicit motion estimation:
  - Non-Local-Means (NLM): Buades, Coll & Morel (2005).
  - Adaptive Window NLM: Boulanger, Kervrann, & Bouthemy (2006).
  - 3D-DCT and Shrinkage: Rusanovskyy, Dabov, Foi, & Egiazarian (2006).
  - Sparse Representations and Learned Dictionary: Protter & Elad (2007).
- All these achieve state-of-the-art results.

Could we leverage on this knowledge and develop novel SRR algorithms that avoid motion estimation ?

Agenda

1. **Super-Resolution (SR) – Introduction**

   The model, the classic approach, and the limitations

2. **The Concept of Fuzzy Motion**

   The idea, who uses it, and why

3. **The Proposed SR Algorithm**

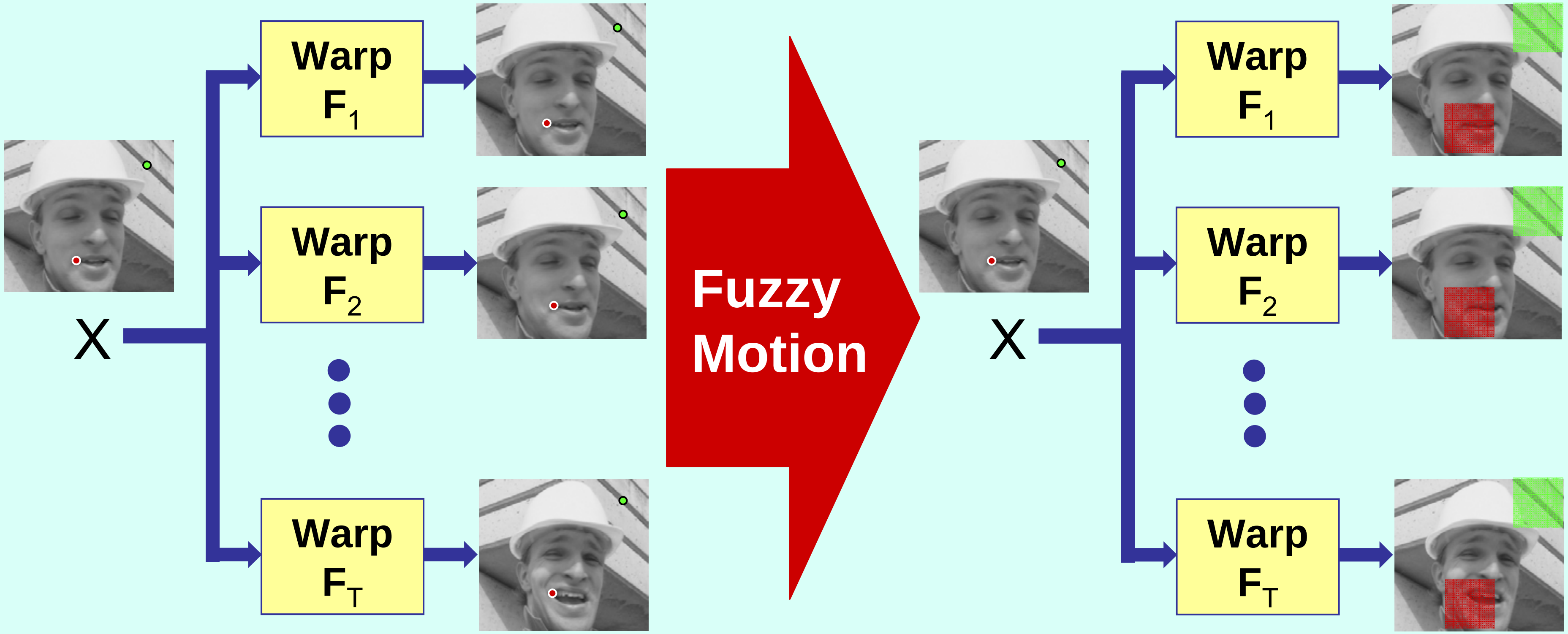
   How does fuzzy motion fit in? the evolved algorithm

4. **Results**

   Several videos, and conclusions

Using Fuzzy Motion – The Core Principle

$$\hat{X} = \min_{X}\left\{\sum_{t=1}^{T} \left\| \mathbf{DHF}_t X - y_t \right\|_2^2 + \Pr(X)\right\}$$



Using Fuzzy Motion – The Formulation

- We use a set of global shift operators that apply all the shifts [dx,dy] in the range [-D,D], i.e. $M=(2D+1)^2$: $\left\{\mathbf{G}_k\right\}_{k=1}^{K}$
- The original formulation is: $\hat{X} = \min_{X}\left\{\sum_{t=1}^{T}\left\|\mathbf{DHF}_t X - y_t\right\|_2^2 + \Pr(X)\right\}$
- Use the new displacement operators, and allow all of them to co-exist: $$\hat{X} = \min_{X}\left\{\sum_{t=1}^{T}\sum_{k=1}^{K}\left\|\mathbf{DHG}_k X - y_t\right\|_2^2 + \Pr(X)\right\}$$
- Some displacements are more likely than others (pixel-wise), and thus weights are needed: $$\hat{X} = \min_{X}\left\{\sum_{t=1}^{T}\sum_{k=1}^{K}\left\|\mathbf{DHG}_k X - y_t\right\|_{\mathbf{W}_{k,t}}^2 + \Pr(X)\right\}$$

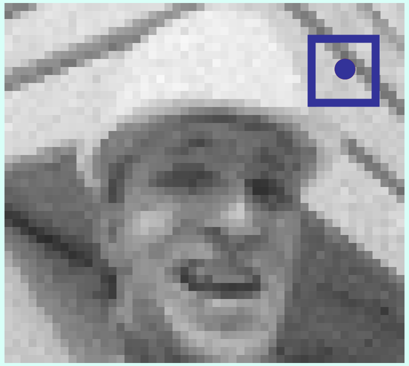
Using Fuzzy Motion – The Weights

- How are the weights computed?
  - $\mathbf{W}_{k,t}$ should reflect the probability that $\mathbf{DHG}_k X = y_t$
  - $\mathbf{W}_{k,t}$ is a diagonal matrix, with varying entries along the main diagonal, reflecting the different movements pixels undergo.
  - $\mathbf{W}_{k,t}[m,n]$ computation:
    - Extract patch around X[m,n].
    - Extract patch around Yt[m+dm,n+dn].
    - Compute the (Euclidean) distance between patches.
    - Compute: $W_{k,t}[m,n] = \exp\left\{-\frac{d^2}{2\sigma^2}\right\}$

Reference image

$X = \text{ScaleUp}\{y_1\}$

Any other image



$Y_t = \text{ScaleUp}\{y_t\}$

Using Fuzzy Motion – Deblurring Aside

$$\hat{X} = \min_{X} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\mathbf{D}\mathbf{H}\mathbf{G}_k X - y_t\right)^{\mathbf{T}} \mathbf{W}_{k,t} \left(\mathbf{D}\mathbf{H}\mathbf{G}_k X - y_t\right) + \mathbf{Pr}(X)$$

H and $\mathbf{G}_k$ are commutative since they are LSI operators

$$\hat{X} = \min_{X} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\mathbf{D}\mathbf{G}_k\mathbf{H} X - y_t\right)^{\mathbf{T}} \mathbf{W}_{k,t} \left(\mathbf{D}\mathbf{G}_k\mathbf{H} X - y_t\right) + \mathbf{Pr}(X)$$

- Let us define Z=**H**X as the blurred-SR image.
- We separate the reconstruction to 2 steps:
  - Recovery of Z (fusion): $\hat{Z} = \min_{X} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\mathbf{D}\mathbf{G}_k Z - y_t\right)^{\mathbf{T}} \mathbf{W}_{k,t} \left(\mathbf{D}\mathbf{G}_k Z - y_t\right)$
  - Recovery of X (deblurring): $\hat{X} = \min_{X} \left\| \mathbf{H}X - \hat{Z} \right\|_2^2 + \lambda \Pr(X)$

Using Fuzzy Motion – The Numerical Scheme

- Little bit of **annoying** algebra leads to the following **pleasant** formula:

$$\hat{Z}[m,n]=\sum_{t=1}^{T}\boxed{\sum_{k\in C}}\mathbf{W}_{k,t}[m,n,dm,dn]\cdot y_t\left[\boxed{\frac{m+dm}{s},\frac{n+dn}{s}}\right]$$

This summation is over all the displacements -D≤dm,dn≤D, such that •••

These indices are both integers (s is the resolution factor, restricted to be an integer)

- Bottom line: Z is computed as a **locally adaptive** weighted averaging of pixels from the low-resolution images in a limited neighborhood.
- The deblurring stage is done using a classical technique (e.g., TV deblurring).

Agenda

1. **Super-Resolution (SR) – Introduction**

   The model, the classic approach, and the limitations

2. **The Concept of Fuzzy Motion**

   The idea, who uses it, and why

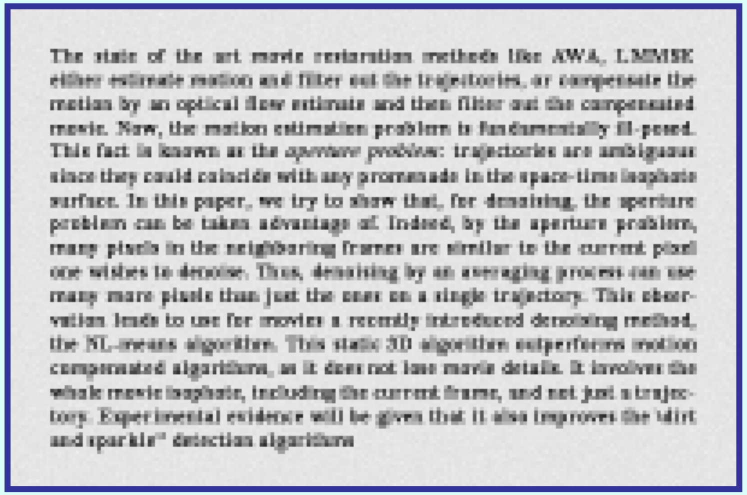
3. **The Proposed SR Algorithm**

   How does fuzzy motion fit in? the evolved algorithm

4. **Results**

   Several videos, and conclusions

Results 1: Naïve Experiment


The state of the art movie restoration methods like AWA, LMMSE either estimate motion and filter out the trajectories, or compensate the motion by an optical flow estimate and then filter out the compensated movie. Now, the motion estimation problem is fundamentally ill-posed. This fact is known as the *aperture problem*: trajectories are ambiguous since they could coincide with any promenade in the space-time isophote surface. In this paper, we try to show that, for denoising, the aperture problem can be taken advantage of. Indeed, by the aperture problem, many pixels in the neighboring frames are similar to the current pixel one wishes to denoise. Thus, denoising by an averaging process can use many more pixels than just the ones on a single trajectory. This observation leads to use for movies a recently introduced denoising method, the NL-means algorithm. This static 3D algorithm outperforms motion compensated algorithms, as it does not lose movie details. It involves the whole movie isophote, including the current frame, and not just a trajectory. Experimental evidence will be given that it also improves the \dirt and sparkle" detection algorithms

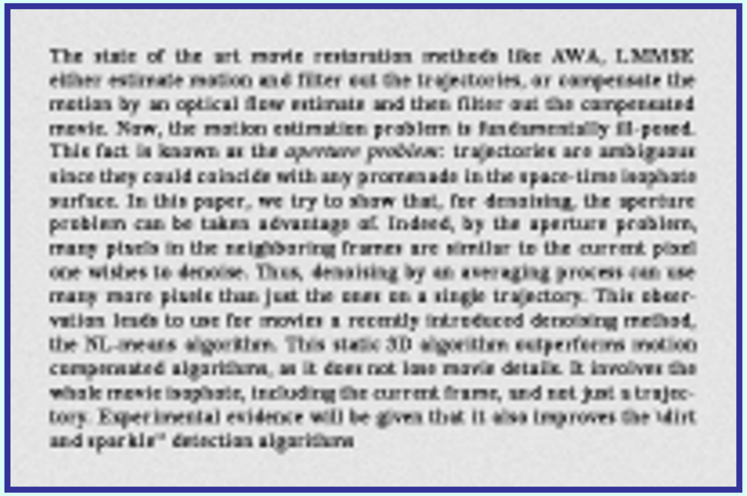
Input Image (1 of 9) created synthetically from a high-res. Image using

(i) 3x3 uniform blur,

(ii) integer global shifts,

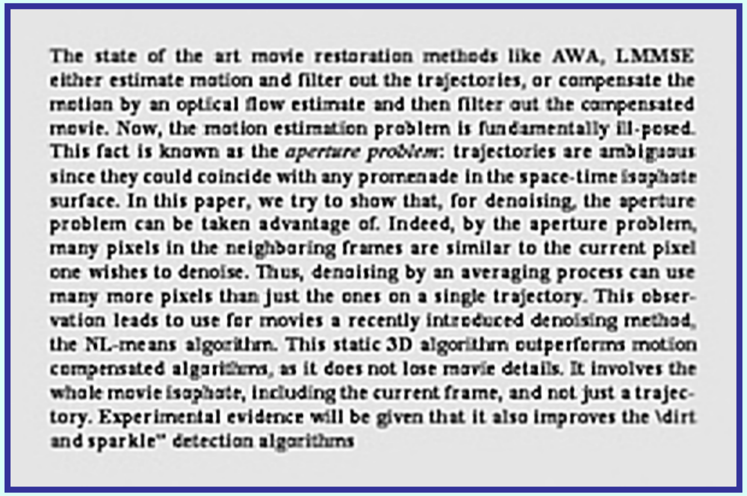
(iii) 3:1 decimation, and

(iv) noise std = 2


The state of the art movie restoration methods like AWA, LMMSE either estimate motion and filter out the trajectories, or compensate the motion by an optical flow estimate and then filter out the compensated movie. Now, the motion estimation problem is fundamentally ill-posed. This fact is known as the *aperture problem*: trajectories are ambiguous since they could coincide with any promenade in the space-time isophote surface. In this paper, we try to show that, for denoising, the aperture problem can be taken advantage of. Indeed, by the aperture problem, many pixels in the neighboring frames are similar to the current pixel one wishes to denoise. Thus, denoising by an averaging process can use many more pixels than just the ones on a single trajectory. This observation leads to use for movies a recently introduced denoising method, the NL-means algorithm. This static 3D algorithm outperforms motion compensated algorithms, as it does not lose movie details. It involves the whole movie isophote, including the current frame, and not just a trajectory. Experimental evidence will be given that it also improves the \dirt and sparkle" detection algorithms

Lanczos Interpolation

Algorithm Result


The state of the art movie restoration methods like AWA, LMMSE either estimate motion and filter out the trajectories, or compensate the motion by an optical flow estimate and then filter out the compensated movie. Now, the motion estimation problem is fundamentally ill-posed. This fact is known as the *aperture problem*: trajectories are ambiguous since they could coincide with any promenade in the space-time isophote surface. In this paper, we try to show that, for denoising, the aperture problem can be taken advantage of. Indeed, by the aperture problem, many pixels in the neighboring frames are similar to the current pixel one wishes to denoise. Thus, denoising by an averaging process can use many more pixels than just the ones on a single trajectory. This observation leads to use for movies a recently introduced denoising method, the NL-means algorithm. This static 3D algorithm outperforms motion compensated algorithms, as it does not lose movie details. It involves the whole movie isophote, including the current frame, and not just a trajectory. Experimental evidence will be given that it also improves the \dirt and sparkle" detection algorithms

Results: Miss America

Input Sequence
(30 Frames)

Created from original high-res. sequence using 3x3 uniform blur, 3:1 decimation, and noise with std = 2

Original Sequence
(Ground Truth)

Lanczos
Interpolation

Algorithm Result

Window Size = 13x13,
Filtering Parameter σ=2.2,
D (search area) = 6,
2 Iterations

Results: Foreman

Input Sequence (30 Frames)

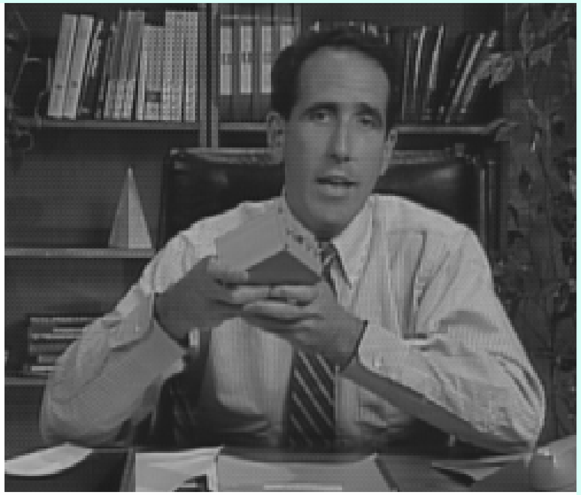
Original Sequence (Ground Truth)

Lanczos Interpolation

Algorithm Result

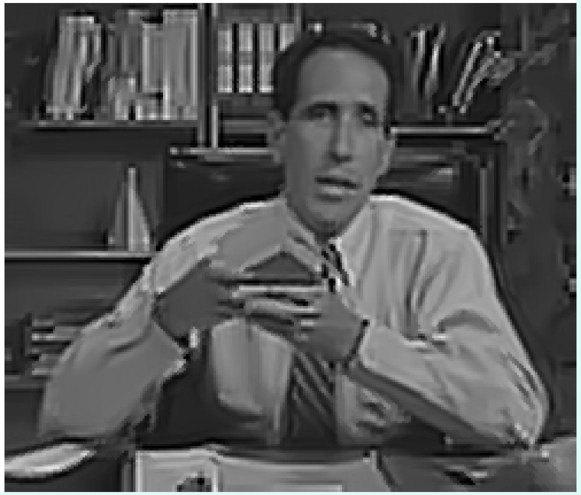
Results: Salesman

Input Sequence (30 Frames)



Original Sequence (Ground Truth)

Lanczos Interpolation



Algorithm Result

Results: Suzie

Input Sequence (30 Frames)

Original Sequence (Ground Truth)

Lanczos Interpolation

Algorithm Result

Summary

- Super-Resolution Reconstruction: improving video resolution.
- Classical SRR approach requires an explicit motion estimation:
  - Must be very accurate.
  - Typically, only **global motion** sequences can be processed reliably.
- Our novel approach uses fuzzy motion estimation:
  - Can process **general content movies**.
  - Gives high quality, almost **artifact-free** results.
  - The eventual algorithm is **very simple**.
  - It is based on **local processing** of image patches - parallelizable.
  - Computational complexity: High! There are ways to improve this.
  - These are just our first steps – better results could be obtained.
- Future work: Many options! …. Stay tuned.