Robust Object Tracking via Adaptive Sparse Representation

Rahman Khorsandi (r.khorsandi@umiami.edu)

Mohamed Abdel-Mottaleb

University of Miami
Department of Electrical and Computer Engineering
Coral Gables, FL 33146

Building the Dictionary

- Assumption: There are k subjects and each subject has n_i training images.
- Each training image is represented by extracted feature vector, $v \in R^m$.

$$A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R^{m \times n_i}$$

$$A = [A_1, A_2, \dots, A_k] \in R^{m \times n}$$

- $y \in R^m$ is the extracted feature vector from a test image.

$$y = Ax$$

Sparse Representation Solution

- The extracted feature from the test image, $y \in R^m$, can be expressed as a linear combination of n_i training images, $\{v_i^1, \ldots, v_i^{n_i}\}$, of the same subject:

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \cdots + \alpha_{i,n_i} v_{i,n_i} \doteq A_i \mathrm{x}_i$$

$$y = A\mathrm{x} \in R^m \qquad \mathrm{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathrm{x}_i \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Wright, Yang, Ganesh, Sastry, and Ma. *Robust Face Recognition via Sparse Representation*, PAMI 2009

Sparse Representation

- We are looking for a sparse solution x:

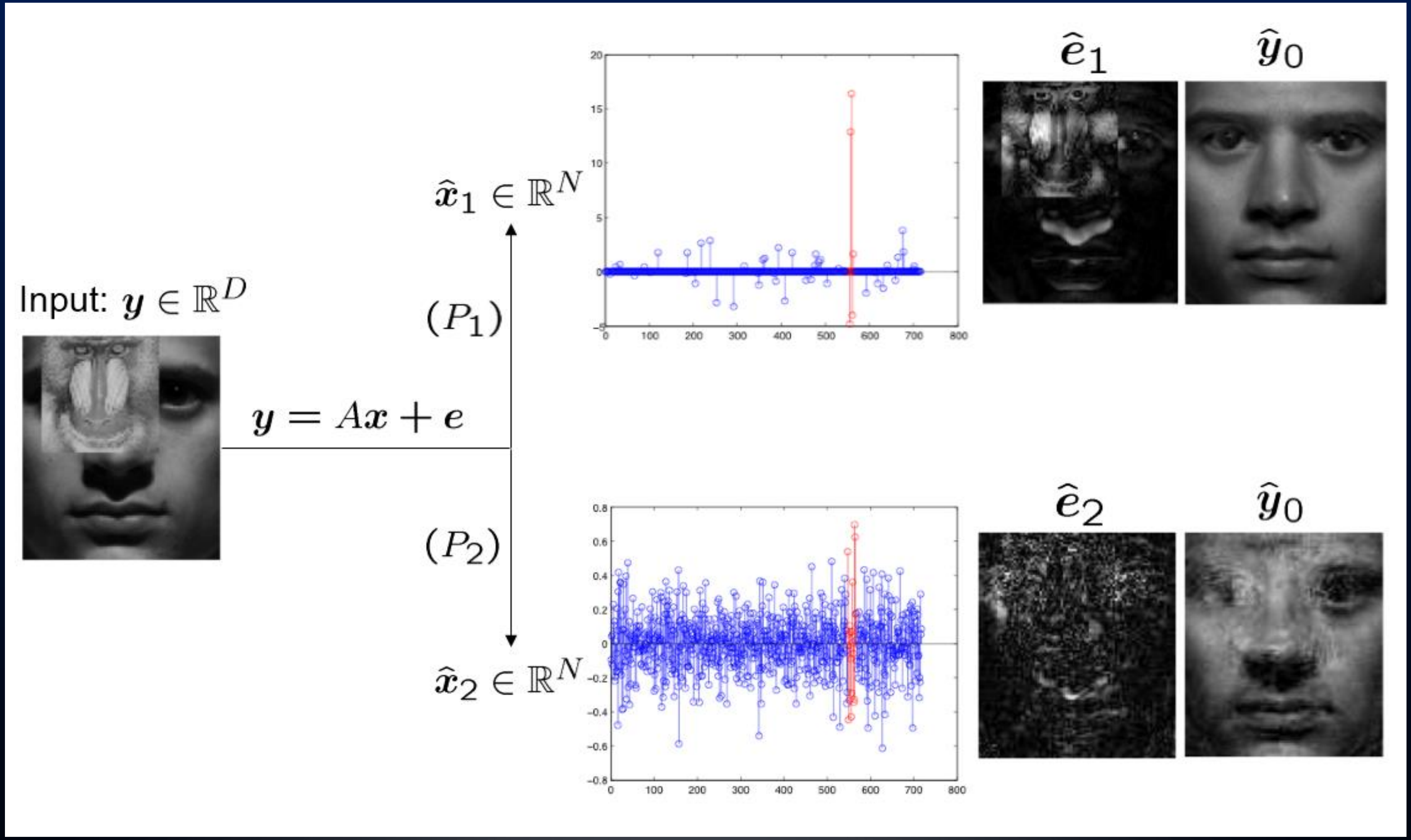
$$(P_0) \qquad \hat{x}_0 = argmin_x \|x\|_0 \; subject\ to\ y = Ax$$

$$\Updownarrow$$

$$(P_1) \qquad \hat{x}_1 = argmin_x \|x\|_1 \; subject\ to\ y = Ax$$

- The problem (P_1) can be solved via Linear programming, and the solution is stable under moderate noise [Candes & Tao'04, Donoho'04]

L1 versus L2 Solution



Wright, Yang, Ganesh, Sastry, and Ma. *Robust Face Recognition via Sparse Representation*, PAMI 2009

Smoothed L0 Norm (SL0)

- L0 Norm
  - discontinuous function
  - Highly sensitive to noise
  - Combinatorial search is needed for minimizing
- The idea of SL0 is based on the approximation of the discontinuous function by a continuous one.

$$f_\sigma(x) \triangleq 1 - e^{\left(\frac{-x^2}{2\sigma^2}\right)}$$

$$f_\sigma(x) \approx \begin{cases} 0, if \ |x| \ll \sigma \\ 1, if \ |x| \gg \sigma \end{cases}$$

Smoothed L0 Norm (SL0)

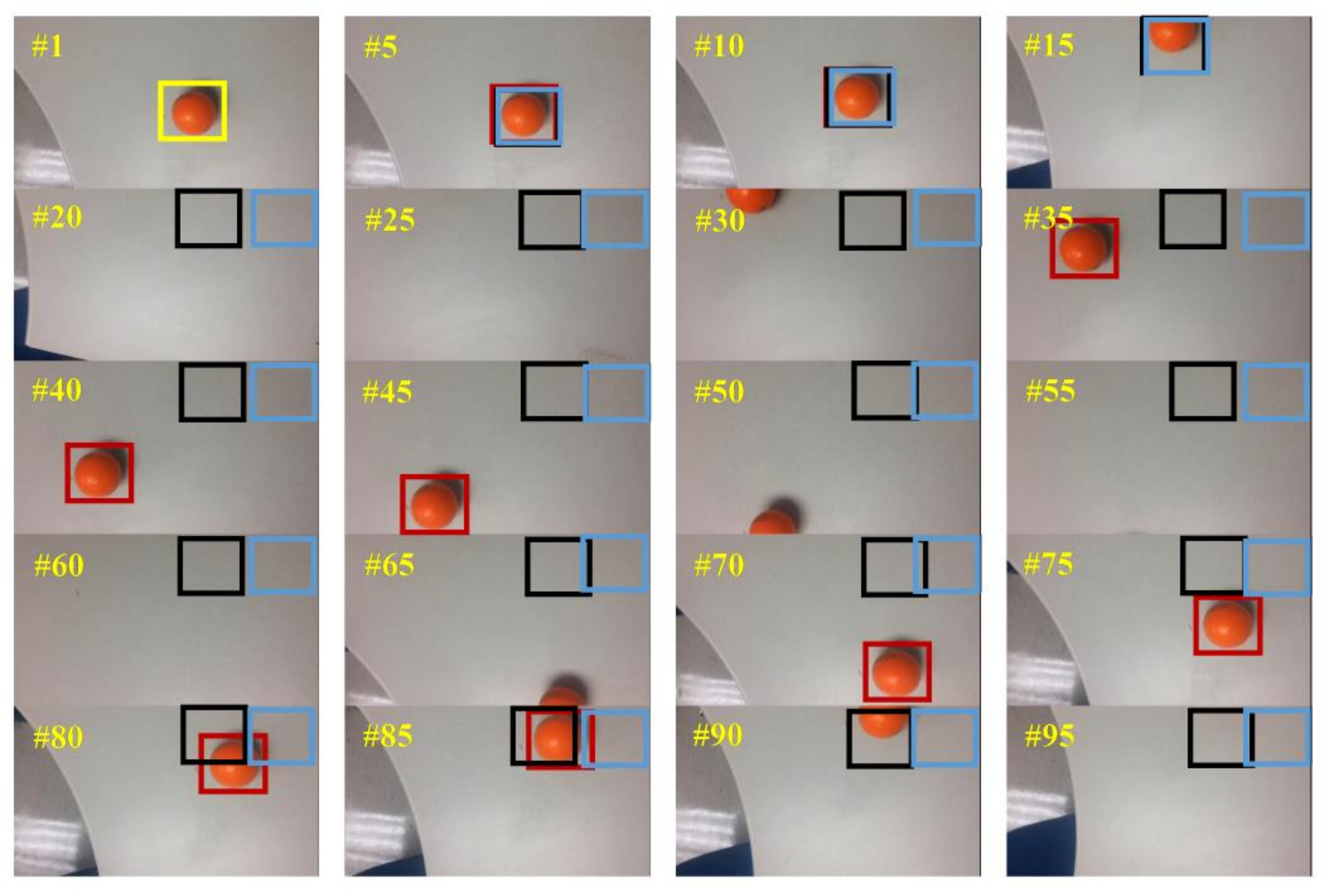
$$F_\sigma(\boldsymbol{x}) = n - \sum_{i=1}^{n} f_\sigma(x_i)$$

- Hence, we can conclude that for small values of σ:

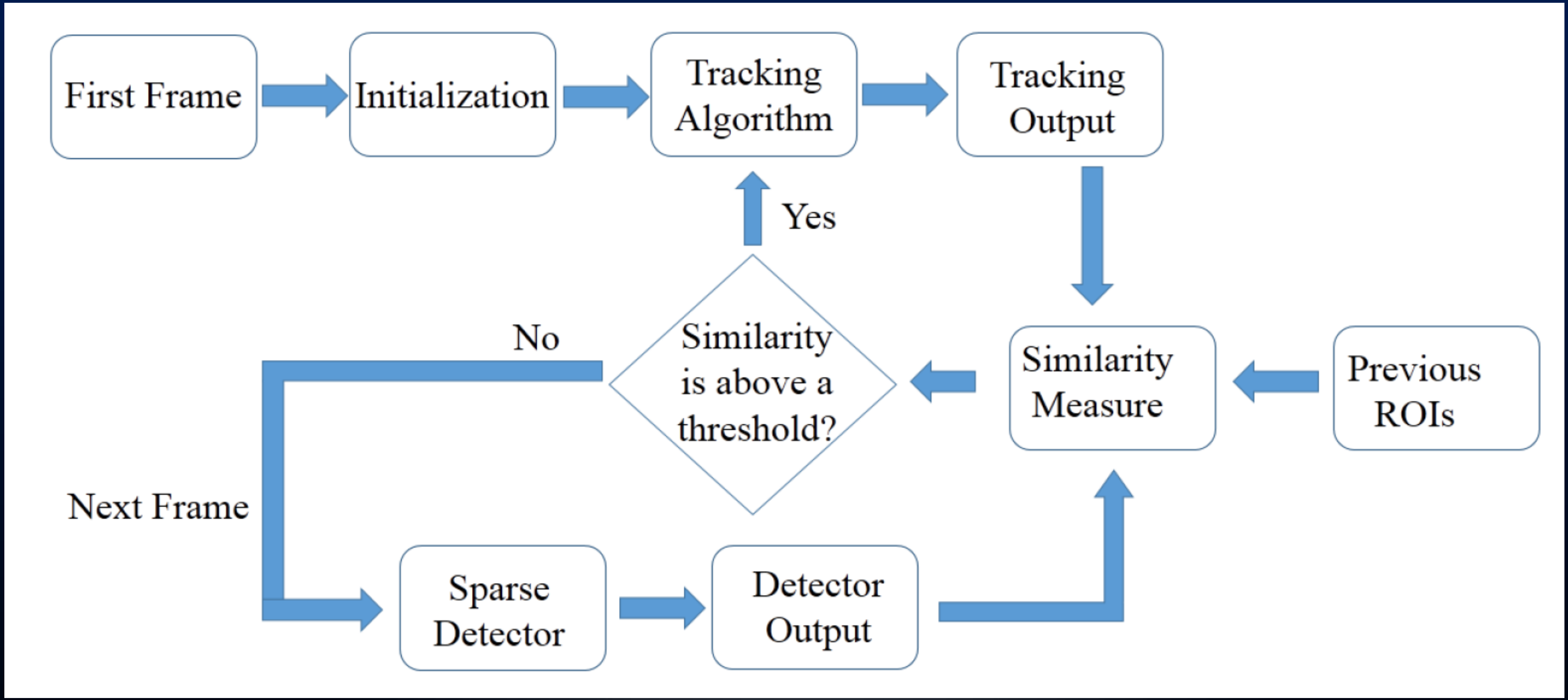
$$\|\boldsymbol{x}\|_0 \approx F_\sigma(\boldsymbol{x})$$

To find the minimum $l0$ norm solution, $F_\sigma(\boldsymbol{x})$ should be minimized.

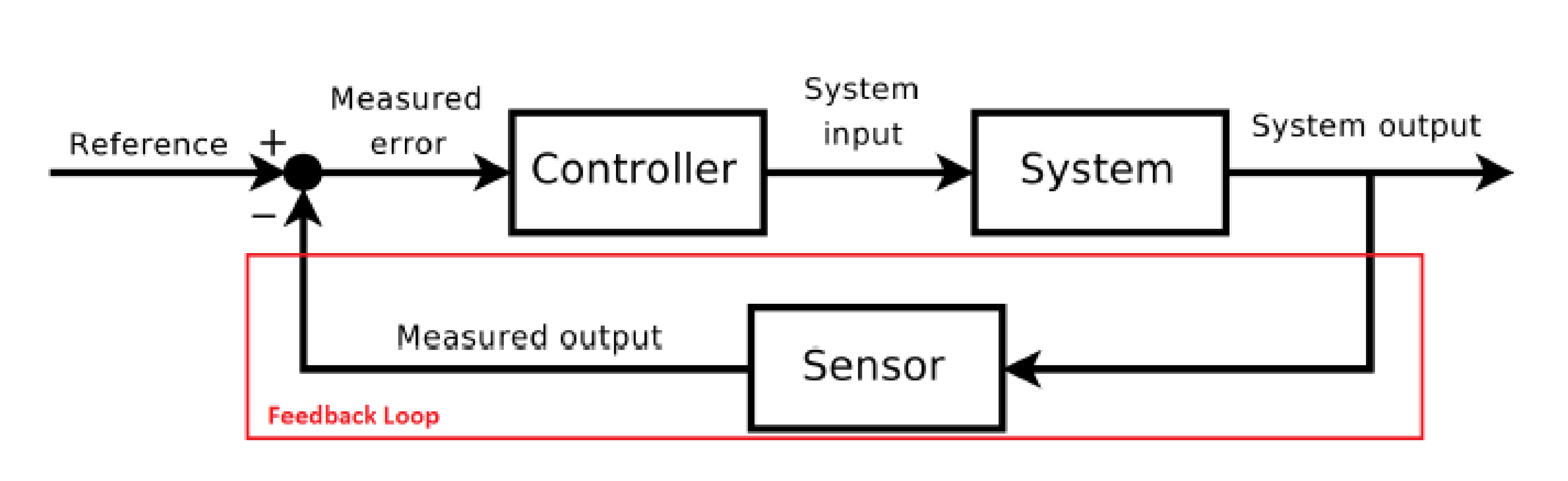
A Simple Tracking Example



Proposed Tracking System



Closed-loop Feedback Control System



Proposed Method

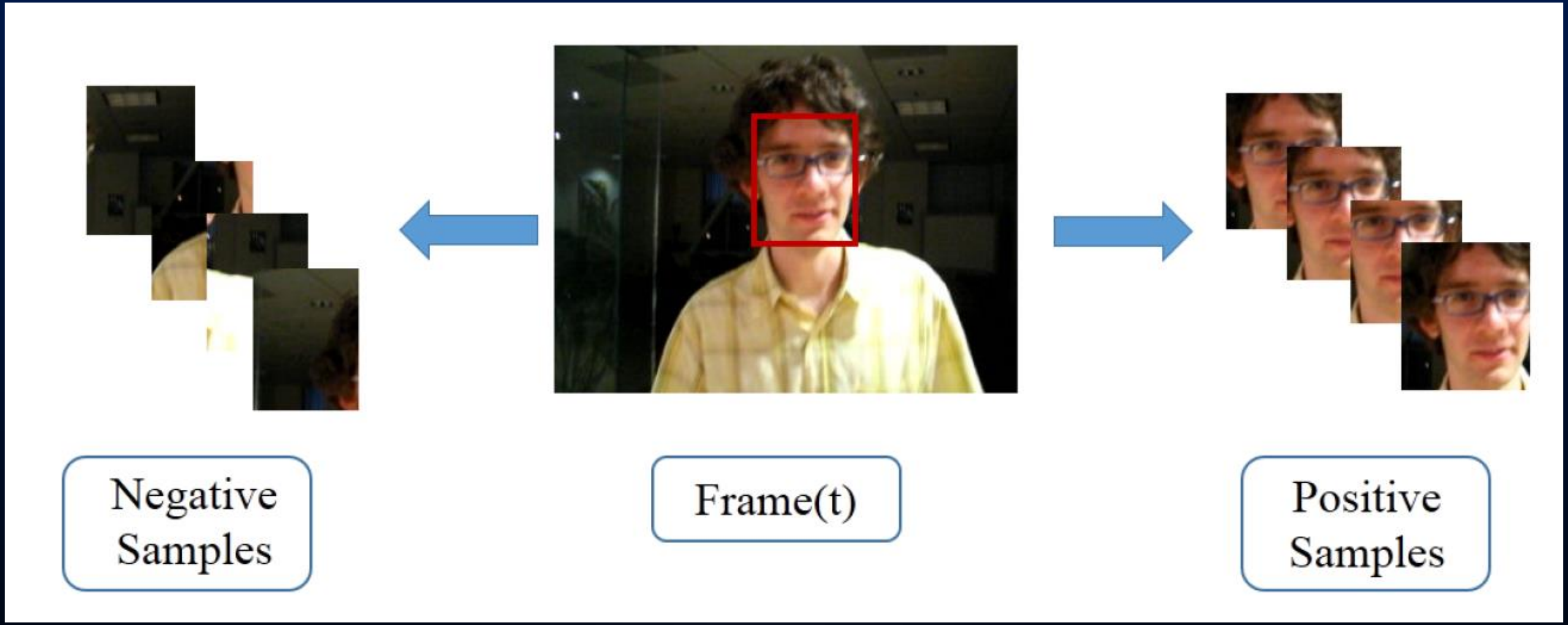
- Time (t)
- Query sample (y)
- Dictionary (A)
- Error (e)
- Sparse coefficients (z)

$$\mathbf{y}(t+1) \approx \mathbf{A}(t)\mathbf{z}(t)$$

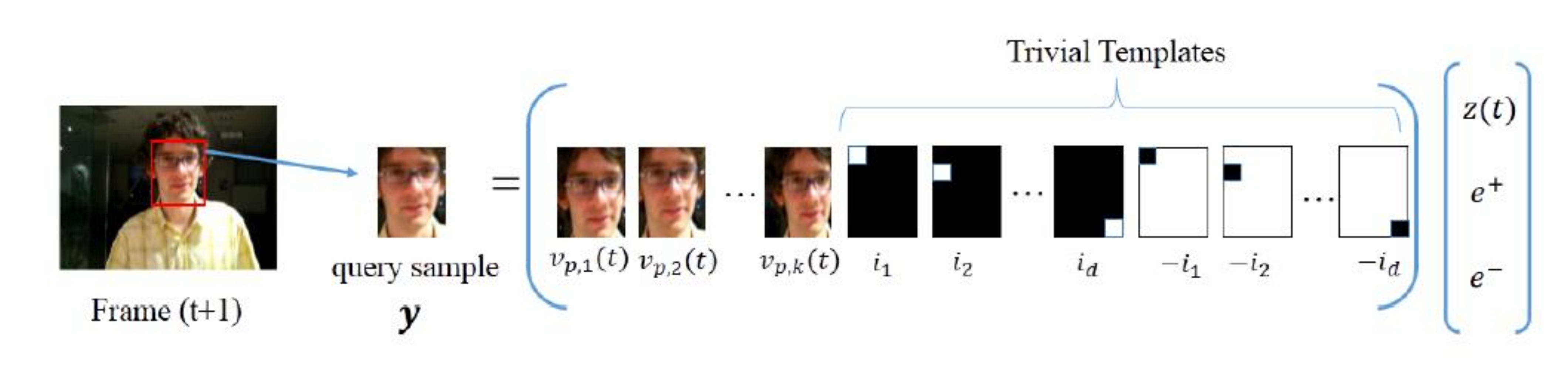
$$\mathbf{y}(t+1) = \mathbf{A}(t)\mathbf{z}(t) + \mathbf{e}$$

$$\mathbf{y}(t+1) = [\mathbf{A}(t), \mathbf{I}] \begin{bmatrix} \mathbf{z}(t) \\ \mathbf{e} \end{bmatrix}$$

Positive and Negative Samples



Representation of a query sample in frame t+1



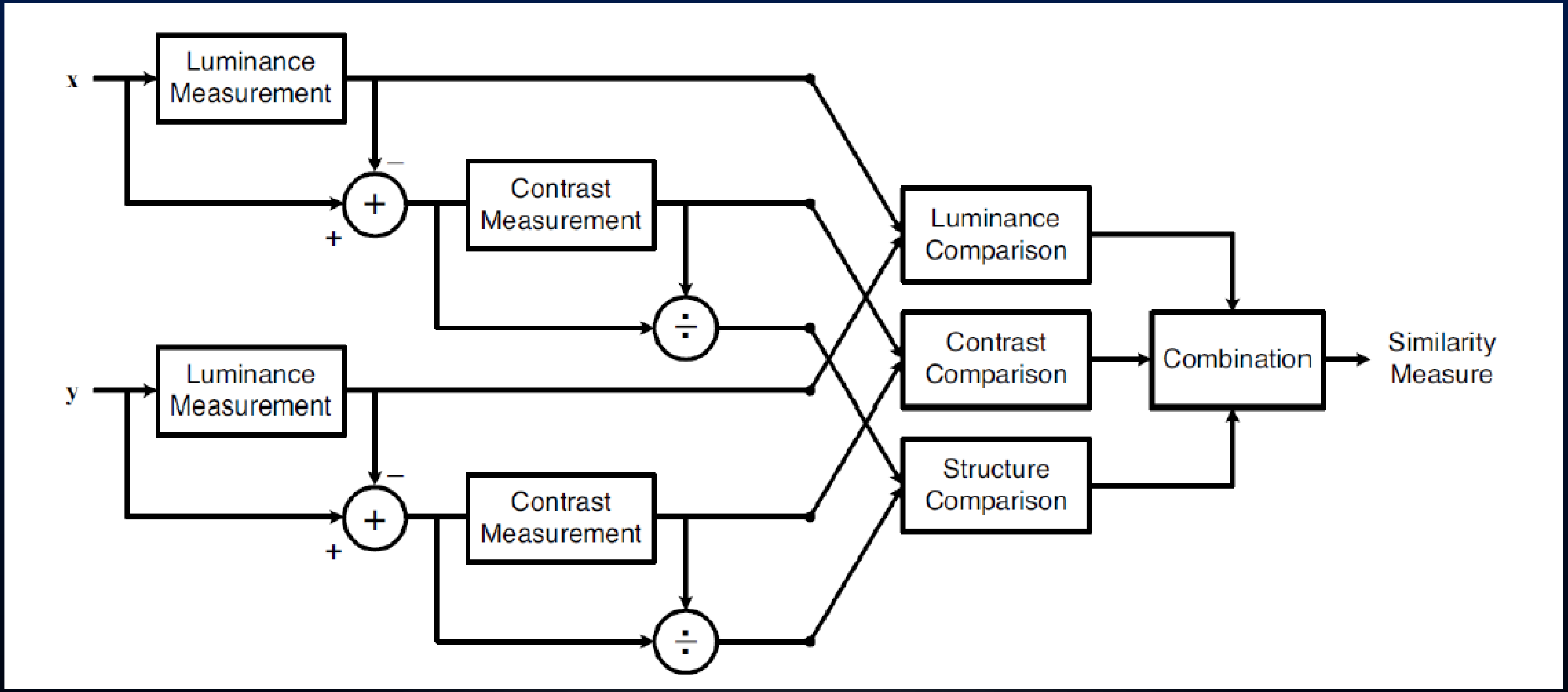
Proposed Method

$$\mathbf{y}(t+1) = [\mathbf{A}(t), \mathbf{I}, -\mathbf{I}] \begin{bmatrix} \mathbf{z}(t) \\ \mathbf{e}_P \\ \mathbf{e}_N \end{bmatrix} \doteq \mathbf{B}(t)\mathbf{x}(t), \quad s.t. \quad \mathbf{x} \succeq 0$$

$$\widehat{\mathbf{x}}_0(t) = argmin \|\mathbf{x}(t)\|_0$$

$$s.t. \quad \mathbf{y}(t+1) = \mathbf{B}(t)\mathbf{x}(t), \quad \mathbf{x}(t) \geq 0$$

Structural Similarity Index Measurement (SSIM)



Wang Zhou, Bovik, Alan C., Sheikh, Hamid R., and Simoncelli, Eero P.*Image Qualifty Assessment: From Error Visibility to Structural Similarity*. IEEE Transactions on Image Processing, Volume 13, Issue 4, pp. 600–612, April 2004

Compressive sensing for feature reduction

- Φ satisfies the restricted isometry property of order k<m and isometry constant $0 \le \delta < 1$ if for all k-sparse signal $\mathbb{u}$:

$$\Phi \in \mathbf{R}^{m \times n} (m \ll n)$$

$$\mathbf{u} \in \mathbf{S}_k = \{\alpha \in \mathbf{R}^n : \| \alpha \|_0 = k\}$$

$$(1-\delta) \| \mathbf{u} \|_2^2 \le \| \Phi \mathbf{u} \|_2^2 \le (1+\delta) \| \mathbf{u} \|_2^2$$

Cont. Compressive sensing

- In the theory of compressive sensing, if the matrix Φ follows the restricted isometry property:

$$\mathbf{z} = \Phi \mathbf{v}$$

Results

- Tracking results for more than 10 various publicly available video sequences
- The Challenging factors:
  - Large pose variation
  - Full and partial occlusion
  - Large scale change
  - Scene blur
  - Significant lighting condition variations
  - Disappearance

Success Rate

Sequence	FCT	CT	CS	Frag	OAB	MIL	MTT	ASLA	Proposed Method
Ball	21	18	15	17	20	16	18	15	87
Biker	35	84	5	3	66	1	9	10	75
Tiger	52	50	62	19	24	34	24	14	68
Chasing	79	47	67	21	71	65	96	63	81
David	98	94	8	8	32	71	41	34	97
Sylvester	77	69	57	34	65	77	67	82	79
Shaking II	88	55	12	34	74	41	93	82	84
Panda	84	90	1	9	83	80	11	71	75
Football	76	74	35	26	31	77	67	7	69
Dark Car	36	53	6	0	14	48	59	57	55
FaceOcc2	99	99	39	54	49	97	88	93	98
Goat	77	26	26	14	46	27	39	37	71
Average SR	68.5	63.2	27.7	19.9	47.9	52.8	51	47.0	78.5

Center Location Error (CLE)

Sequence	FCT	CT	CS	Frag	OAB	MIL	MTT	ASLA	Proposed Method
Ball	45	48	55	52	48	52	50	58	18
Biker	12	6	176	107	10	44	68	109	6
Tiger	23	25	48	39	42	27	61	49	8
Chasing	10	12	9	56	9	13	4	47	8
David	11	14	72	73	57	19	125	57	13
Sylvester	9	14	84	47	12	9	18	9	15
Shaking II	15	46	255	119	18	58	16	27	18
Panda	6	10	157	56	8	7	47	9	10
Football	13	14	43	144	37	13	9	207	24
Dark Car	9	10	89	116	11	9	7	8	8
FaceOcc2	12	16	29	57	36	17	19	20	13
Goat	18	103	137	140	71	109	99	95	25
Average CLE	15.2	26.5	96.1	83.8	29.9	31.4	43.5	57.9	13.1

Thank You

Questions/Comments?