GOSUS: Grassmannian Online Subspace Updates with Structured-sparsity

Jia Xu[†], Vamsi K. Ithapu[†], Lopamudra Mukherjee[§], James M. Rehg[‡], Vikas Singh[†]

[†]University of Wisconsin-Madison, [§]University of Wisconsin-Whitewater, [‡]Georgia Institute of Technology

Problem Statement

Online subspace learning in the context of sequential observations involving **structured** perturbations

Motivation:

- Segment video as "foreground"/"background" by online learning of the background subspace
- The deviation from the background subspace is the "foreground"
- But "foreground" does not come in "random" and is almost always "structured"

Our Ideas

Model data as two layers: subspace + perturbation:

- Subspace is modeled on a Grassmannian with online updating along the geodesic
- Spatially contiguous and structured perturbations (people, objects, landmarks in videos) are modeled via group sparsity

Structured Sparsity

- Group operator: A $n \times n$ diagonal matrix D^i

$$D^i_{jj} = \begin{cases} 1 & \text{if element } j \text{ is in group } i; \\ 0 & \text{otherwise.} \end{cases}$$

- Support overlapping and non-overlapping groups
- Groups for background subtraction: coarse-to-fine superpixels
- Groups for face tracking: a structure between landmark regions

Problem Formulation

Denoting $\mathbf{v}$ as observation, U as subspace matrix, $\mathbf{w}$ as coefficient vector, $\mathbf{x}$ as perturbation, we formulate

$$\min_{U^TU=I_d,\mathbf{w},\mathbf{x}} \sum_{i=1}^{l} \mu_i \|D^i\mathbf{x}\|_2 + \frac{\lambda}{2}\|U\mathbf{w} + \mathbf{x} - \mathbf{v}\|_2^2 \qquad (1)$$

- non-convex feasible set: $U^TU = I_d$;
- non-smooth regularizer: mixed norm.

Solve for tuple $(\mathbf{w}, \mathbf{x})$ at fixed U^*

Introducing a slack variable $\mathbf{z}$ to decouple the non-smooth term,

$$\min_{\mathbf{w},\mathbf{x}} \sum_{i=1}^{l} \mu_i \|\mathbf{z}^i\|_2 + \frac{\lambda}{2}\|U^*\mathbf{w} + \mathbf{x} - \mathbf{v}\|_2^2 \qquad (2)$$
$$\text{s.t.} \quad \mathbf{z}^i = D^i\mathbf{x}, \quad i = 1, \cdots, l.$$

The augmented Lagrangian is given by

$$\mathcal{L}(\mathbf{w}, \mathbf{x}, \{\mathbf{z}^i\}, \{\mathbf{y}^i\}) = \sum_{i=1}^{l} \mu_i \|\mathbf{z}^i\|_2 + \frac{\lambda}{2}\|U^*\mathbf{w} + \mathbf{x} - \mathbf{v}\|_2^2 + \sum_{i=1}^{l} {\mathbf{y}^i}^T (D^i\mathbf{x} - \mathbf{z}^i) + \sum_{i=1}^{l} \frac{\rho_i}{2}\|D^i\mathbf{x} - \mathbf{z}^i\|_2^2 \qquad (3)$$

Algorithm 1 ADMM for solving $(\mathbf{w}^*, \mathbf{x}^*)$

In: Subspace: U^*, observation: $\mathbf{v}$, initial: $\mathbf{x}_0$, $\mathbf{z}^i_0$, $\mathbf{y}^i_0$, group operator: D^i, hyper-parameters: λ, μ, ρ
Out: coefficient vector: $\mathbf{w}^*$, structured outliers: $\mathbf{x}^*$
Procedure:

1: **for** $k = 0 \to K$ **do**
2: $A \leftarrow \begin{bmatrix} \lambda I_d & \lambda {U^*}^T \\ \lambda U^* & \lambda I_n + \sum_{i=1}^{l} \rho_i D^i \end{bmatrix}$, /* $A \succ 0$, sparse */
3: $\mathbf{b} \leftarrow \begin{bmatrix} \lambda {U^*}^T \mathbf{v} \\ \lambda \mathbf{v} - \sum_{i=1}^{l} D^i \mathbf{y}^i_k + \sum_{i=1}^{l} \rho_i D^i \mathbf{z}^i_k \end{bmatrix}$
4: $(\mathbf{w}_{k+1}, \mathbf{x}_{k+1}) \leftarrow \min_{\mathbf{w},\mathbf{x}} \|(A[\mathbf{w} \quad \mathbf{x}]^T - \mathbf{b})\|^2$ on GPU
5: $\mathbf{r}^i_k \leftarrow \mathbf{D}^i\mathbf{x}_{k+1} + \frac{\mathbf{y}^i_k}{\rho_i}$
6: $\mathbf{z}^i_{k+1} \leftarrow \max\{\|\mathbf{r}^i_k\|_2 - \frac{\mu_i}{\rho_i}, 0\}\frac{\mathbf{r}^i_k}{\|\mathbf{r}^i_k\|_2}$
7: $\mathbf{y}^i_{k+1} \leftarrow \mathbf{y}^i_k + \rho_i(D^i\mathbf{x}_{k+1} - \mathbf{z}^i_{k+1})$
8: Stop if tolerance conditions satisfied.
9: **end**

Convergence Result

Theorem 1. For $\lambda, \mu_i, \rho_i > 0, \forall i \in \{1, \cdots, l\}$, the sequence $\{(\mathbf{w}_k, \mathbf{x}_k, \{\mathbf{z}^i_k\}, \{\mathbf{y}^i\})\}$ generated by Alg. 1 from any initial point $(\mathbf{w}_0, \mathbf{x}_0, \{\mathbf{z}^i_0\}, \{\mathbf{y}^i_0\})$ converges to $(\mathbf{w}^*, \mathbf{x}^*, \{\mathbf{z}^{i*}\}, \{\mathbf{y}^{i*}\})$, which minimizes (3) at fixed U^*.

Update of U with estimated $(\mathbf{w}^*, \mathbf{x}^*)$

1. Derivative of $\mathcal{L}(.)$ in (3) w.r.t. U,

$$\frac{\partial \mathcal{L}}{\partial U} = \lambda(U\mathbf{w}^* + \mathbf{x}^* - \mathbf{v}){\mathbf{w}^*}^T = \mathbf{s}{\mathbf{w}^*}^T \qquad (4)$$

$\mathbf{s} = \lambda(U\mathbf{w}^* + \mathbf{x}^* - \mathbf{v})$: residual vector.

2. Gradient on the Grassmannian

$$\nabla \mathcal{L} = (I - UU^T)\frac{\partial \mathcal{L}}{\partial U} = (I - UU^T)\mathbf{s}{\mathbf{w}^*}^T = \mathbf{s}{\mathbf{w}^*}^T \qquad (5)$$

3. Compact SVD of $\nabla \mathcal{L} = \mathbf{p}\sigma\mathbf{q}$

$$\mathbf{p} = \frac{\mathbf{s}}{\|\mathbf{s}\|}, \quad \sigma = \|\mathbf{s}\|\|\mathbf{w}^*\|, \quad \mathbf{q} = \frac{\mathbf{w}^*}{\|\mathbf{w}^*\|}$$

4. Update U with a gradient stepsize η along the geodesic direction $-\nabla \mathcal{L}$

$$U(\eta) = U + (\cos(\sigma\eta) - 1)U\mathbf{q}\mathbf{q}^T - \sin(\sigma\eta)\mathbf{p}\mathbf{q}^T \qquad (6)$$

Lemma 1. The subspace updating procedure (6) preserves the column-wise orthogonality of U.

Full Pipeline

Algorithm 2 Main Procedure of GOSUS

In: Observation: V, subspace initialization: U_0, hyperparameters: λ, μ, ρ
Out: Background: B, structured foreground: X
Procedure:

1: **for** $t = 1 \to T$ **do**
2: Solve $(\mathbf{w}^*, \mathbf{x}^*, \{\mathbf{z}^{i*}\}, \{\mathbf{y}^{i*}\})$ by Algorithm 1;
3: (Optional) Update stepsize η_t ;
4: Update U_t by (6);
5: **end**

Remark: relation to stochastic gradient algorithms

- Examples come in a sequential manner, instead of random sampling;
- Gradient of U for each example is computed from the manifold.

Online Background Subtraction

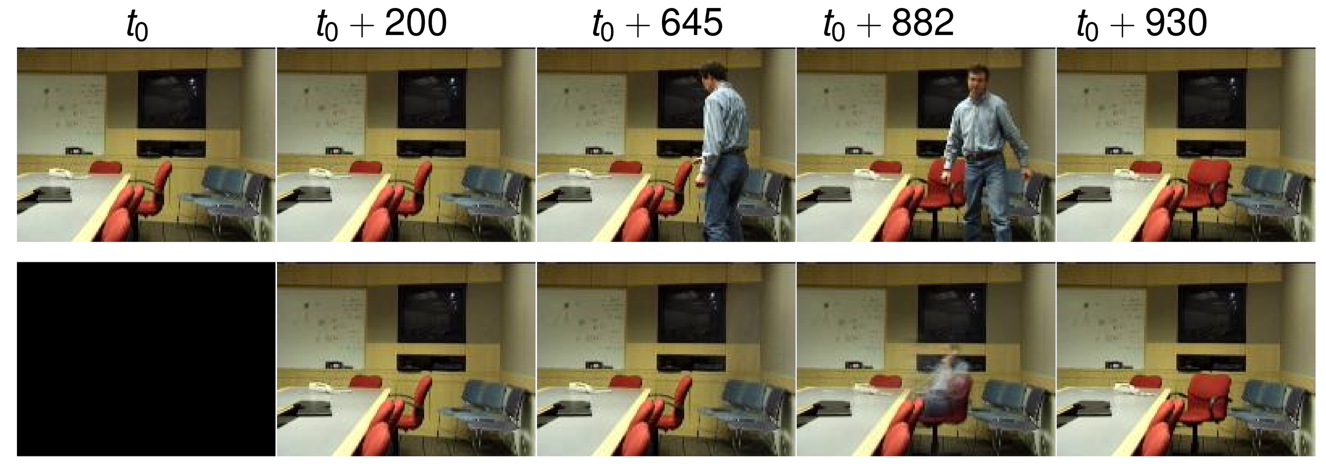

Fig. 1 Effective adapting to intermittent object motion in the background.

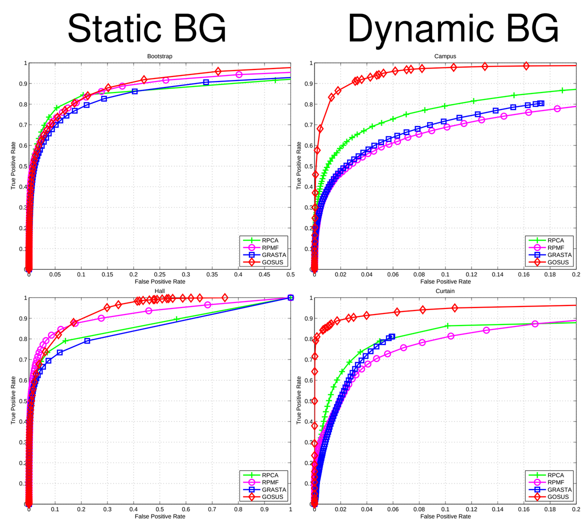

Fig. 2 ROC curves.

Video Datasets	Models			
	RPCA[24]	RPMF[36]	GRASTA[15]	GOSUS
Fountain	0.94	0.94	0.69	**0.99**
Escalator	0.91	0.90	0.90	**0.96**
WavingTrees	0.74	0.84	0.87	**0.98**
Campus	0.90	0.86	0.77	**0.98**
Bootstrap	0.87	0.91	0.87	**0.93**
WaterSurface	0.73	0.84	0.87	**0.97**
Hall	0.82	0.90	0.76	**0.93**
Time of Day	0.80	0.85	0.84	**0.89**
LightSwitch	0.87	**0.92**	0.62	0.88
Curtain	0.87	0.90	0.88	**0.96**
Lobby	0.89	0.94	0.70	**0.95**
ShoppingMall	0.92	0.93	0.90	**0.94**

Tab. 1 Area under ROC.

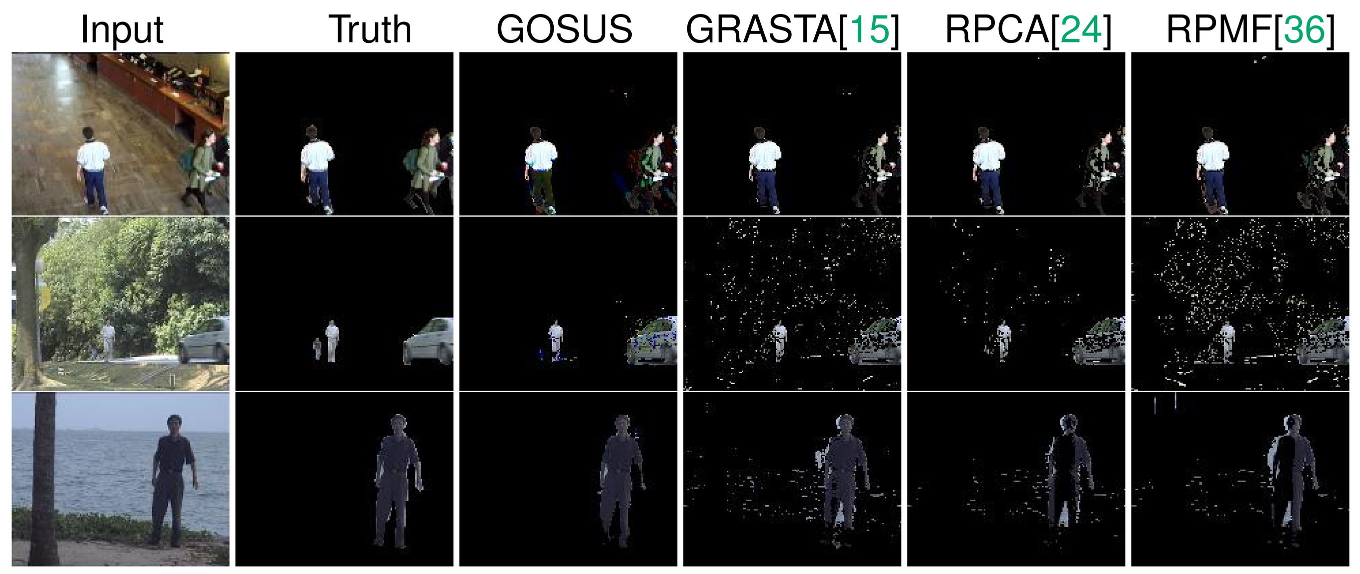

Fig. 3 Qualitative results.

Online Multiple Face Tracking

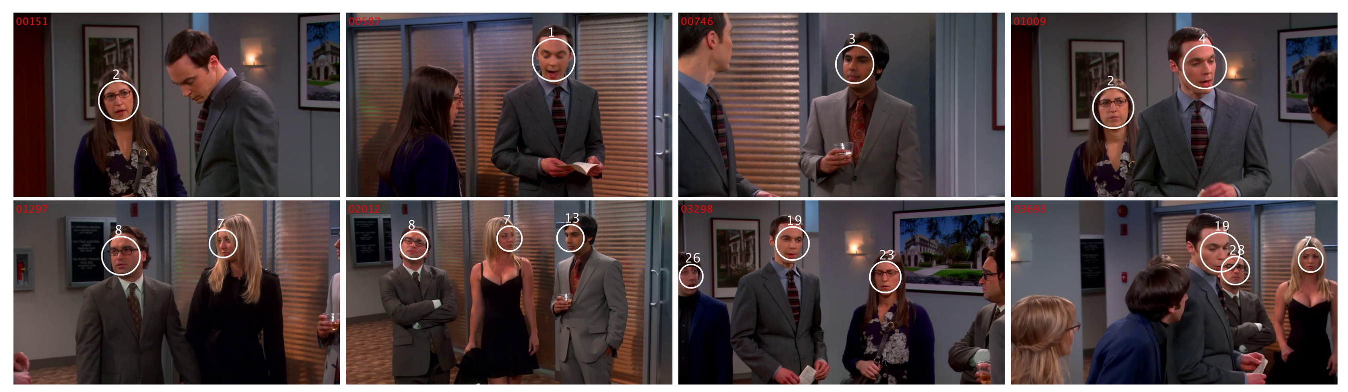

Fig. 4 Examples of multiple face tracking in the Big Bang Theory.

Key Takeaways

1. Online subspace updating schemes on the Grassmannian while allowing structured norm regularization.
2. Imposing prior structures on perturbations improves subspace estimation and is robust to noisy observations.
3. Code available: http://pages.cs.wisc.edu/~jiaxu/projects/gosus/

Acknowledge to NSF RI 1116584

jiaxu@cs.wisc.edu