GOSUS: Grassmannian Online Subspace Updates with Structured-sparsity

Supplement Material

Anonymous ICCV submission

Paper ID 342

1. Detailed analysis of technical results in the main paper

This section includes additional details of a few observations and theorems summarized in the main paper.

1.1. The matrix in the linear system has a special structure

Observation 1. *For $\lambda > 0, {U^*}^T U^* = I_d, \rho_i > 0, \forall i \in \{1, \cdots, l\}$, we have $A \succ 0$.*

Proof. Observe that A is given by,

$$A \leftarrow \begin{bmatrix} \lambda I_d & \lambda {U^*}^T \\ \lambda U^* & \lambda I_n + \sum_{i=1}^{l} \rho_i D^i \end{bmatrix} \tag{1}$$

and denoting $Q = \sum_{i=1}^{l} \rho_i D^i$, we have (for any $(\mathbf{w}, \mathbf{x})$),

$$\begin{aligned} \begin{bmatrix} \mathbf{w} \\ \mathbf{x} \end{bmatrix}^T A \begin{bmatrix} \mathbf{w} \\ \mathbf{x} \end{bmatrix} &= \begin{bmatrix} \lambda(\mathbf{w}^T + \mathbf{x}^T U^*) & \lambda(\mathbf{w}^T {U^*}^T + \mathbf{x}^T) + \mathbf{x}^T Q \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{x} \end{bmatrix} \\ &= \lambda(\mathbf{w}^T \mathbf{w} + \mathbf{x}^T U^* \mathbf{w} + \mathbf{w}^T {U^*}^T \mathbf{x} + \mathbf{x}^T \mathbf{x}) + \mathbf{x}^T Q \mathbf{x} \\ &= \lambda \|\mathbf{x} + U^* \mathbf{w}\|_2^2 + \mathbf{x}^T Q \mathbf{x} \end{aligned} \tag{2}$$

Let us check both terms in (2). Observe that $\forall \mathbf{x}$, $\mathbf{x}^T Q \mathbf{x} \geq 0$. Next, as $\lambda \|\mathbf{x} + U^* \mathbf{w}\|_2^2 \geq 0$, for the LHS of the identity in (2), we have

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{x} \end{bmatrix}^T A \begin{bmatrix} \mathbf{w} \\ \mathbf{x} \end{bmatrix} \geq 0 \tag{3}$$

For the equality to hold in (3), we need to have both the terms, $\|\mathbf{x} + U^* \mathbf{w}\|_2^2$ and $\mathbf{x}^T Q \mathbf{x}$ equal to 0. But $Q \succ 0$ because $\mathbf{x}^T Q \mathbf{x} = 0$ only when $\mathbf{x} = 0$. Further, $\mathbf{w} = 0$ whenever $\mathbf{x} = 0$, for $\|\mathbf{x} + U^* \mathbf{w}\|_2^2$ to be zero. Hence equality in (3) holds only when $\mathbf{w}$ and $\mathbf{x}$ are zero.

□

1.2. Convergence properties

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

Proof. Our proof emulates the convergence proof in [2]. We first show that model with a fixed U agrees with the standard ADMM formulation in [2].

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{x}, \mathbf{z}^i} \quad & \sum_{i=1}^{l} \mu_i \|\mathbf{z}^i\|_2 + \frac{\lambda}{2} \|U\mathbf{w} + \mathbf{x} - \mathbf{v}\|_2^2 \\ \text{s.t.} \quad & \mathbf{z}^i = D^i \mathbf{x}, \end{aligned} \tag{4}$$

If we denote $f(\mathbf{w}, \mathbf{x}) = \frac{\lambda}{2} \|U\mathbf{w} + \mathbf{x} - \mathbf{v}\|_2^2$, $g(\{\mathbf{z}^i\}) = \sum_{i=1}^{l} \mu_i \|\mathbf{z}^i\|_2$, our problem (4) is really a special case of (3.1) in [2].

We next need to show that (4) satisfies the two main assumptions made in the convergence proof given in [2].

Assumption (i) $f(\mathbf{w}, \mathbf{x})$ and $g(\{\mathbf{z}^i\})$ are both convex, proper and closed.
Assumption (ii) $\mathcal{L}(\mathbf{w}^*, \mathbf{x}^*, \{\mathbf{z}^{i*}\}, \{\mathbf{y}^{i*}\})$ has a saddle point.

For notational simplicity, denote $\mathbf{c} = \begin{bmatrix}\mathbf{w} & \mathbf{x}\end{bmatrix}^T$, $\hat{U} = \begin{bmatrix}U & I_n\end{bmatrix}$ and $\hat{D}^i = \begin{bmatrix}0_{n\times d} & D^i\end{bmatrix}$ ($0_{n\times d}$ is a $n \times d$ zero matrix). Using this notation, $f(\mathbf{w}, \mathbf{x}) = f(\mathbf{c}) = \frac{\lambda}{2}\|\hat{U}\mathbf{c} - \mathbf{v}\|_2^2$. Here, $f(\mathbf{c})$ is convex. By non-negativity of the norm squared function, $f(\mathbf{c}) \geq 0 > -\infty$, and taking $\mathbf{c} = 0$, we have $f(\mathbf{c}) = \|v\|_2^2 < \infty$. Hence, $f(\mathbf{c})$ is proper. Further, the domain of $\mathbf{c}$ is $\mathcal{R}^{n+d}$ and $f(\mathbf{c})$ is continuous on that domain. Following the closure property of proper convex functions [3], we see that $f(\mathbf{c})$ is closed.

Following similar arguments as above, consider $g_1(\mathbf{z}^i) = \|\mathbf{z}^i\|_2$. Since $\|.\|_2$ is convex and increasing and as $\mathbf{z}^i = D^i x$, using the composition rule, $g_1(\mathbf{z}^i)$ is convex. Using the non-negativity of the norm and by taking $\mathbf{x} = 0$ which gives $g_1(\mathbf{z}^i) < \infty$, we have $g_1(\mathbf{z}^i)$ is proper. Finally, observe that $g_1(\mathbf{z}^i)$ is a continuous function of $\mathbf{x}$, and the domain on $\mathbf{x}$ ($\mathcal{R}^n$) is closed. Hence, $g_1(\mathbf{z}^i)$ a closed proper convex function. The non-negative sum of closed proper convex functions is also closed proper convex when the domain of summation remains unchanged. This concludes the proof of the first assumption.

For the second part, using the new notation, the augmented Lagrangian is,

$$\mathcal{L}_0(\mathbf{w}, \mathbf{x}, \{\mathbf{z}^i\}, \{\mathbf{y}^i\}) \sim \mathcal{L}_0(\mathbf{c}, \{\mathbf{z}^i\}, \{\mathbf{y}^i\}) = \sum_{i=1}^{l} \mu_i \|\mathbf{z}^i\|_2 + \frac{\lambda}{2}\|\hat{U}\mathbf{c} - \mathbf{v}\|_2^2 + \sum_{i=1}^{l} {\mathbf{y}^i}^T(\mathbf{z}^i - \hat{D}^i\mathbf{c}) \tag{5}$$

First, observe that the domain of $\mathbf{c}$, $\mathbf{z}^i$, and $\mathbf{y}^i$ is $\mathcal{R}^{n+d}$, $\mathcal{R}^n$, and $\mathcal{R}_+^n$ respectively, which are compact and convex sets. Fixing $\mathbf{y}^i$s for $i = 1, \ldots, l$, $\mathcal{L}_0$ is a convex function of $\mathbf{c}$ and $\mathbf{z}^i$. This follows from the fact that the first two terms in (5) are convex and the last term is affine in the primal parameters (when ${\mathbf{y}^i}$'s are fixed). So, there exists a triple, $(\mathbf{c}^*, \{{\mathbf{z}^i}^*\}, \{{\mathbf{y}^i}^*\})$ such that

$$\mathcal{L}_0(\mathbf{c}^*, \{{\mathbf{z}^i}^*\}, \{{\mathbf{y}^i}^*\}) \leq \mathcal{L}_0(\mathbf{c}, \{\mathbf{z}^i\}, \{{\mathbf{y}^i}^*\}).$$

Further, for a fixed ($\mathbf{c}$,$\mathbf{z}^i$), $\mathcal{L}_0$ is an linear combination of affine functions in $\mathbf{y}^i$s. Hence it is concave. So, there exists a triple, $(\mathbf{c}^*, \{{\mathbf{z}^i}^*\}, \{{\mathbf{y}^i}^*\})$

$$\mathcal{L}_0(\mathbf{c}^*, \{{\mathbf{z}^i}^*\}, \{\mathbf{y}^i\}) \leq \mathcal{L}_0(\mathbf{c}^*, \{{\mathbf{z}^i}^*\}, \{{\mathbf{y}^i}^*\}).$$

Thus, $\mathcal{L}_0$ has a saddle point $(\mathbf{c}^*, \{{\mathbf{z}^i}^*\}, \{{\mathbf{y}^i}^*\})$ in the primal–dual domain.

□

1.3. The updating scheme for U: Part one

$$U(\eta) = U + (\cos(\sigma\eta) - 1)U\mathbf{q}\mathbf{q}^T - \sin(\sigma\eta)\mathbf{p}\mathbf{q}^T \tag{13}$$

Recall the compact SVD of $\nabla\mathcal{L}$ by (12) in the main paper ($\mathcal{L}$ is the Lagrangian),

$$\nabla\mathcal{L} = \frac{\mathbf{s}}{\|\mathbf{s}\|} \times \|\mathbf{s}\|\|\mathbf{w}^*\| \times \left(\frac{\mathbf{w}^*}{\|\mathbf{w}^*\|}\right)^T = \mathbf{p}\sigma\mathbf{q}^T$$

Here, we approximate the full SVD with

$$\nabla\mathcal{L} = S\Sigma V^T = \begin{bmatrix}\mathbf{p} & \mathbf{p}_2 & \cdots & \mathbf{p}_d\end{bmatrix} \times \mathrm{diag}(\sigma, 0, \cdots, 0) \times \begin{bmatrix}\mathbf{q} & \mathbf{q}_2 & \cdots & \mathbf{q}_d\end{bmatrix}^T$$

where $S = \begin{bmatrix}\mathbf{p} & \mathbf{p}_2 & \cdots & \mathbf{p}_d\end{bmatrix}$, $\Sigma = \mathrm{diag}(\sigma, 0, \cdots, 0)$, $V = \begin{bmatrix}\mathbf{q} & \mathbf{q}_2 & \cdots & \mathbf{q}_d\end{bmatrix}$, $\mathbf{p}_2, \cdots, \mathbf{p}_d$ and $\mathbf{q}_2, \cdots, \mathbf{q}_d$ are slack orthonormal basis, which will be omitted by the zero singular values.

By Thm. 2.65 in [1], we can update our subspace U with stepsize η by

$$U(\eta) = \begin{bmatrix}UV & -S\end{bmatrix} \begin{bmatrix}\cos(\Sigma\eta) \\ \sin(\Sigma\eta)\end{bmatrix} V^T$$

Observe that the above update involves full matrix operations. It can be simplified as,

$$\begin{aligned}
U(\eta) =& UV\cos(\Sigma\eta)V^T - S\sin(\Sigma\eta)V^T \\
=& UV\cos(\mathrm{diag}(\sigma\eta, 0, \cdots, 0))V^T - S\sin(\mathrm{diag}(\sigma\eta, 0, \cdots, 0))V^T \\
=& UV\mathrm{diag}(\cos(\sigma\eta), 1, \cdots, 1)V^T - S\mathrm{diag}(\sin(\sigma\eta), 0, \cdots, 0)V^T \\
=& UV\mathrm{diag}(1, 1, 1, \cdots, 1)V^T + UV\mathrm{diag}(\cos(\sigma\eta) - 1, 0, \cdots, 0)V^T - S\mathrm{diag}(\sin(\sigma\eta), 0, \cdots, 0)V^T \\
=& UVI_dV^T + UV\mathrm{diag}(\cos(\sigma\eta) - 1, 0, \cdots, 0)V^T - S\mathrm{diag}(\sin(\sigma\eta), 0, \cdots, 0)V^T \\
=& U + (\cos(\sigma\eta) - 1)U\mathbf{q}\mathbf{q}^T - \sin(\sigma\eta)\mathbf{p}\mathbf{q}^T
\end{aligned}$$

The last identity is what appears in the main paper.

1.4. The updating scheme for U: Part two

Lemma 1. *The subspace updating procedure (13) preserves the column-wise orthogonality of U.*

Proof. The residual vector $\mathbf{s}$ is orthogonal to all the columns of U, thus we have

$$U^T\mathbf{p} = U^T\frac{\mathbf{s}}{\|\mathbf{s}\|} = 0$$

Also, $\mathbf{p}, \mathbf{q}$ are unary vectors, hence $\mathbf{q}^T\mathbf{q} = 1, \mathbf{p}^T\mathbf{p} = 1$. Now we show $U(\eta)^TU(\eta) = I_d$:

$$\begin{aligned}
U(\eta)^TU(\eta) &= \left(U + (\cos(\sigma\eta) - 1)U\mathbf{q}\mathbf{q}^T - \sin(\sigma\eta)\mathbf{p}\mathbf{q}^T\right)^T \left(U + (\cos(\sigma\eta) - 1)U\mathbf{q}\mathbf{q}^T - \sin(\sigma\eta)\mathbf{p}\mathbf{q}^T\right) \\
&= U^TU + (\cos(\sigma\eta) - 1)U^TU\mathbf{q}\mathbf{q}^T - \sin(\sigma\eta)U^T\mathbf{p}\mathbf{q}^T + (\cos(\sigma\eta) - 1)\mathbf{q}\mathbf{q}^TU^TU \\
&\quad + (\cos(\sigma\eta) - 1)^2\mathbf{q}\mathbf{q}^TU^TU\mathbf{q}\mathbf{q}^T - (\cos(\sigma\eta) - 1)\sin(\sigma\eta)\mathbf{q}\mathbf{q}^TU^T\mathbf{p}\mathbf{q}^T \\
&\quad - \sin(\sigma\eta)\mathbf{q}\mathbf{p}^TU - (\cos(\sigma\eta) - 1)\sin(\sigma\eta)\mathbf{q}\mathbf{p}^TU\mathbf{q}\mathbf{q}^T + \sin^2(\sigma\eta)\mathbf{q}\mathbf{p}^T\mathbf{p}\mathbf{q}^T \\
&= I_d + (\cos(\sigma\eta) - 1)\mathbf{q}\mathbf{q}^T - 0 + (\cos(\sigma\eta) - 1)\mathbf{q}\mathbf{q}^T + (\cos(\sigma\eta) - 1)^2\mathbf{q}\mathbf{q}^T \\
&\quad - 0 - 0 - 0 + \sin^2(\sigma\eta)\mathbf{q}\mathbf{q}^T \\
&= I_d + (2\cos(\sigma\eta) - 2 + \cos^2(\sigma\eta) - 2\cos(\sigma\eta) + 1 + \sin^2(\sigma\eta))\mathbf{q}\mathbf{q}^T \\
&\qquad (\ 2\cos(\sigma\eta) \text{ cancels out and } \cos^2(\sigma\eta) + \sin^2(\sigma\eta) = 1\ ) \\
&= I_d
\end{aligned}$$

Thus, the subspace updating procedure preserves the column-wise orthogonality of U.

□

2. More Details on Experiments

In the main paper, we presented representative ROC curves for 6 video datasets. Figs. 1 and Fig. 2 give the ROC curves for all 12 videos. Observe that the ROCs in Fig. 2 are zoomed in versions of those in Fig. 1.

References

[1] T. Arias, A. Edelman, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20:303–353, 1998. 2

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 1

[3] R. T. Rockafellar. *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970. 2

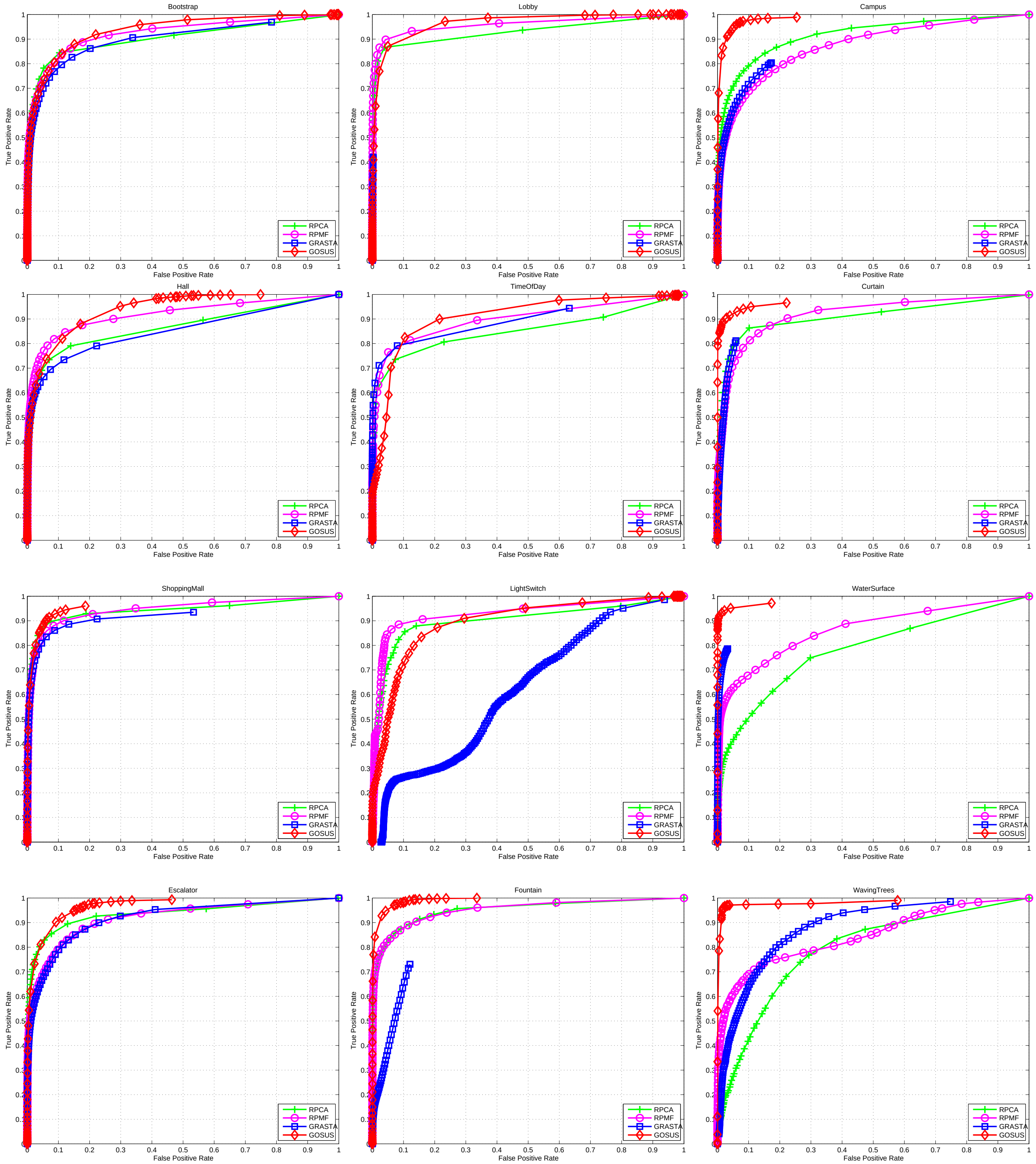

Figure 1: ROC curves of 12 datasets for three different dataset categories showing the performance of RPCA, RPMF, GRASTA and GOSUS.

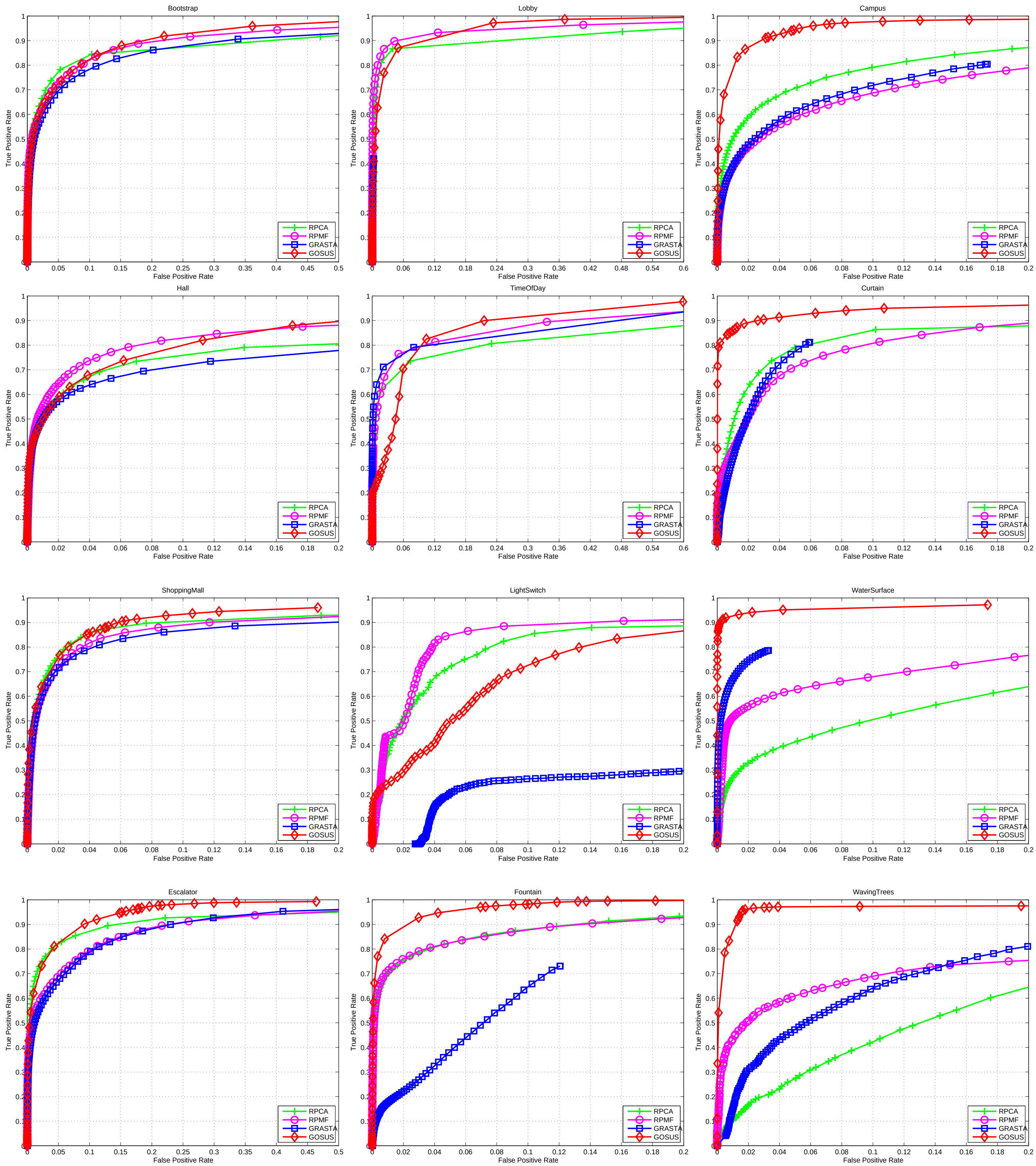

Figure 2: Zoomed in (adjusted false positive rate range) ROC curves of 12 datasets for three different dataset categories showing the performance of RPCA, RPMF, GRASTA and GOSUS.