A Nonconvex Projection Method for Robust PCA

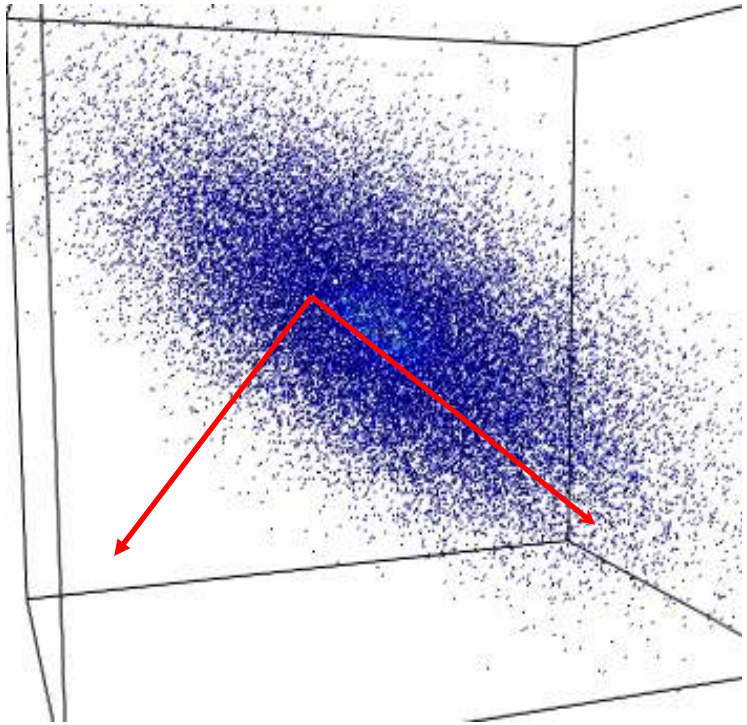
Aritra Dutta[⋆], Filip Hanzely[⋆], Peter Richtárik[⋆,Ω,α]

King Abdullah University of Science and Technology[⋆], University of Edinburgh[Ω], MIPT[α]

Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

The Classical PCA

Set of possibly correlated variables → a set of linearly uncorrelated variables



- Proposed by Karl Pearson in 1901
- Again invented independently by Harold Hotelling in 1930s
- Comes in different names- Karhunen Loève transform, singular value decomposition, EYM theorem, factor analysis…

Given a matrix $A \in R^{m \times n}$ solve:

$$U\Sigma_k V^T = \arg\min_{\substack{X \\ r(X) \leq r}} ||A - X||_F^2$$

Fundamental flaw:

1. Equal weighting on each points!
2. Robust to only Gaussian Noise.

Robust Principal Component Pursuit

Robust principal component pursuit is a matrix decomposition model in which we wish to decompose A into the sum of a low-rank matrix L and an error matrix S : A = L + S.

Data matrix

=

Low-rank matrix

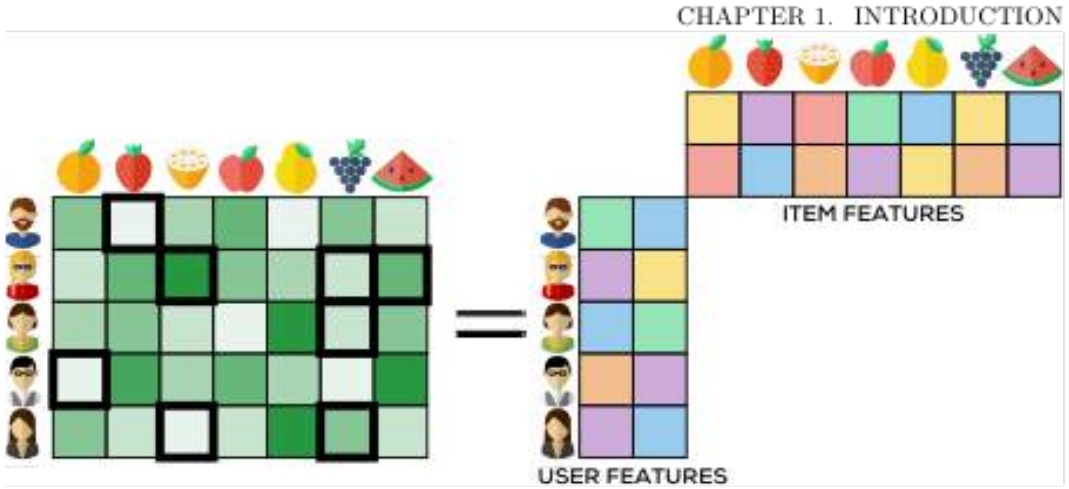
+

Sparse matrix

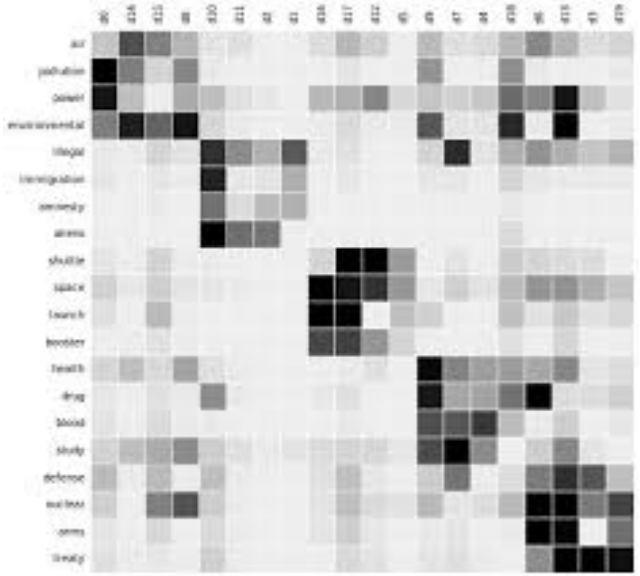
The celebrated principal component pursuit uses the ℓ_0 norm to address the sparsity and one obtain:

$$\min_{L,S} \operatorname{rank}(L) + \lambda \|S\|_{\ell_0} \quad \text{subject to} \quad A = L + S,$$

Applications

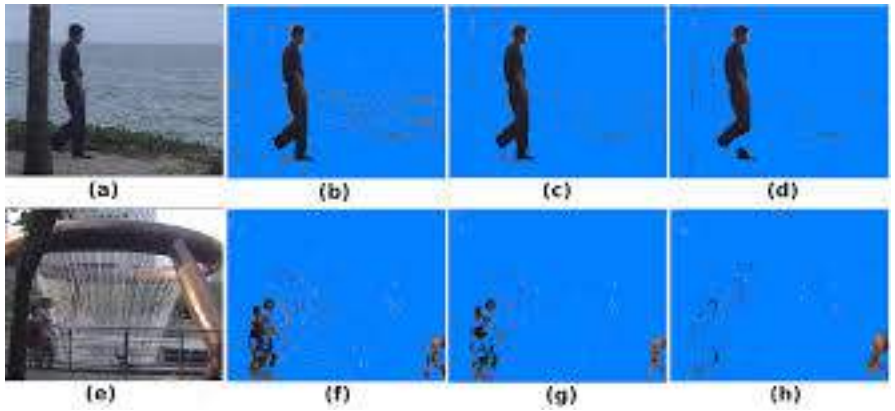


Recommender systems



Common words + key words

Latent Semantic Indexing



Background Estimation

Tracking

1. EEG images
2. fMRI images
3. Sparse Subspace clustering
4. Hyperspectral Image denoising
5. Contrast-filled vessels extraction
6. …

Several Surrogate Formulations Available

By using surrogate constraints and optimization objective functions, Wright et al. 2009; Lin, Chen, and Ma 2010; Cande`s et al. 2011 proposed the celebrated Robust Principal Component Analysis as follows:

RPCA

$$\min_{L,S} \|L\|_* + \lambda \|S\|_{\ell_1} \quad \text{subject to} \quad A = L + S,$$

Moving the constraint to the objective as a penalty, together with adding explicit constraints on the target rank r and target sparsity s leads to the following formulation (Zhou and Tao 2011) as:

Matrix Completion

$$\begin{aligned} &\min_{L,S} \|A - L - S\|_F^2 \\ &\text{subject to} \quad \operatorname{rank}(L) \leq r \ \text{ and } \ \|S\|_0 \leq s. \end{aligned}$$

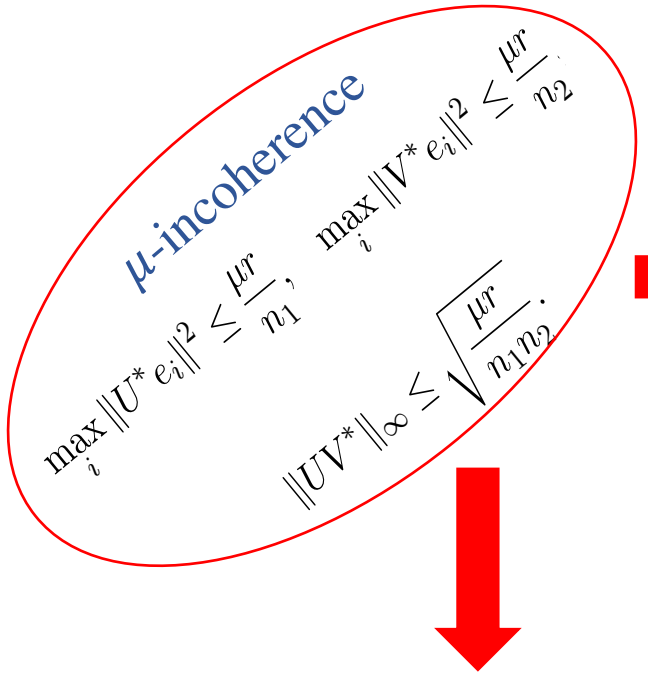
Cherapanamjeri, Gupta, and Jain 2017; Chen et al. 2011; Tao and Yang 2011 proposed the robust matrix completion (RMC) problem as:

Robust Matrix Completion

$$\begin{aligned} &\min_{L,S} \|\mathcal{P}_\Omega(A - L - S)\|_F^2 \\ &\text{subject to} \quad \operatorname{rank}(L) \leq r \ \text{ and } \ \|\mathcal{P}_\Omega(S)\|_0 \leq s', \end{aligned} \qquad \text{where} \qquad (\mathcal{P}_\Omega[X])_{ij} = \begin{cases} X_{ij} & (i,j) \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

Issues

There is a serious identifiability issue…



Theorem 1.1 *Suppose L_0 is $n \times n$, obeys (1.2)–(1.3), and that the support set of S_0 is uniformly distributed among all sets of cardinality m. Then there is a numerical constant c such that with probability at least $1 - cn^{-10}$ (over the choice of support of S_0), Principal Component Pursuit (1.1) with $\lambda = 1/\sqrt{n}$ is* exact, *i.e. $\hat{L} = L_0$ and $\hat{S} = S_0$, provided that*

$$\mathrm{rank}(L_0) \le \rho_r n \, \mu^{-1} (\log n)^{-2} \quad \text{and} \quad m \le \rho_s \, n^2. \tag{1.4}$$

Above, ρ_r and ρ_s are positive numerical constants. In the general rectangular case where L_0 is $n_1 \times n_2$, PCP with $\lambda = 1/\sqrt{n_{(1)}}$ succeeds with probability at least $1 - cn_{(1)}^{-10}$, provided that $\mathrm{rank}(L_0) \le \rho_r n_{(2)} \, \mu^{-1} (\log n_{(1)})^{-2}$ and $m \le \rho_s \, n_1 n_2$.

Theorem 1.2 *Suppose L_0 is $n \times n$, obeys the conditions (1.2)–(1.3), and that Ω_{obs} is uniformly distributed among all sets of cardinality m obeying $m = 0.1n^2$. Suppose for simplicity, that each observed entry is corrupted with probability τ independently of the others. Then there is a numerical constant c such that with probability at least $1 - cn^{-10}$, Principal Component Pursuit (1.5) with $\lambda = 1/\sqrt{0.1n}$ is exact, i.e. $\hat{L} = L_0$, provided that*

$$\mathrm{rank}(L_0) \le \rho_r \, n \mu^{-1} (\log n)^{-2}, \quad \text{and} \quad \tau \le \tau_s. \tag{1.6}$$

Above, ρ_r and τ_s are positive numerical constants. For general $n_1 \times n_2$ rectangular matrices, PCP with $\lambda = 1/\sqrt{0.1 n_{(1)}}$ succeeds from $m = 0.1 n_1 n_2$ corrupted entries with probability at least $1 - cn_{(1)}^{-10}$, provided that $\mathrm{rank}(L_0) \le \rho_r \, n_{(2)} \mu^{-1} (\log n_{(1)})^{-2}$.

And here lies the motivation…

• Nonconvex Feasibility and Alternating Projections

Set feasibility problem aims to find a point in the intersection of a collection of closed sets, that is:

$$\text{Find} \;\; x \in \mathcal{X} \qquad \text{where} \qquad \mathcal{X} \overset{\text{def}}{=} \cap_i^m \mathcal{X}_i \neq \emptyset, \tag{8}$$

A natural alternating projection algorithm to solve (8) is Algorithm 1.

Algorithm 1: Alternating projection method for set feasibility

Input : $\Pi_i(\cdot)$ – Projector onto $\mathcal{X}_i$ for each $i \in \{1, \dots m\}$, starting point x_0

for $k = 0, 1, \dots$ **do**

  Choose via some rule i (e.g., cyclically or randomly)

  $x_{k+1} = \Pi_i(x_k)$

end

Output : x_{k+1}

Now we consider the following reformulation of RPCA:

$$\text{Find} \quad M \overset{\text{def}}{=} [L, S] \in \mathcal{X} \overset{\text{def}}{=} \cap_{i=1}^3 \mathcal{X}_i \neq \emptyset, \tag{9}$$

where

$$\begin{aligned} \mathcal{X}_1 &\overset{\text{def}}{=} \{M \mid L + S = A\} \qquad (10) \\ \mathcal{X}_2 &\overset{\text{def}}{=} \{M \mid \text{rank}(L) \leq r\} \\ \mathcal{X}_3 &\overset{\text{def}}{=} \{M \mid \|S_{(i,.)}\|_0 \leq \alpha n \;\text{ and }\; \|S_{(.,j)}\|_0 \leq \alpha m \\ &\qquad \text{for all } i \in [m], j \in [n].\} \end{aligned}$$

Algorithm 2: Alternating projection method for RPCA

Input : $A \in \mathbb{R}^{m \times n}$ (the given matrix), rank r, sparsity level $\alpha \in (0, 1]$

Initialize : L_0, S_0

for $k = 0, 1, \ldots$ **do**

$\tilde{L} = \frac{1}{2}(L_k - S_k + A)$

$\tilde{S} = \frac{1}{2}(S_k - L_k + A)$

$L_{k+1} = H_r(\tilde{L})$

$S_{k+1} = \mathcal{T}_\alpha(\tilde{S})$

end

Output : L_{k+1}, S_{k+1}

Projection of $[L_0, S_0]$ onto set of $[L, S]$ such that $L + S = A$

$[\frac{1}{2}(L_0 - S_0 + A), \frac{1}{2}(S_0 - L_0 + A)]$

Projection of L_0 onto rank r constraint

Operator H_r – rank r SVD of L_0

Projection of S_0 on the sparsity constraint (approximate)

Keep largest α fraction of values in each row and each column

Algorithm 3: Alternating projection method for RMC

Input : $A \in \mathbb{R}^{m \times n}$ (the given matrix), rank r, sparsity level $\alpha \in (0, 1]$

Initialize : L_0, S_0

for $k = 0, 1, \ldots$ **do**

$\tilde{L} = \frac{1}{2}\mathcal{P}_\Omega(L_k - S_k + A)$

$\tilde{S} = \frac{1}{2}\mathcal{P}_\Omega(S_k - L_k + A)$

$L_{k+1} = H_r(\tilde{L} + \mathcal{P}_{\Omega^c}(L_k))$

$S_{k+1} = \mathcal{T}_\alpha(\tilde{S})$

end

Output : L_{k+1}, S_{k+1}

Projection of $[L_0, S_0]$ onto set of $[L, S]$ such that $L_\Omega + S_\Omega = A_\Omega$

$[(L_0)_{\Omega^C} + \frac{1}{2}(L_0 - S_0 + A)_\Omega, (S_0)_{\Omega^C} + \frac{1}{2}(S_0 - L_0 + A)_\Omega]$

Convergence of Algorithm 2

Local linear convergence

Euclidean distance to optimum

$$d_{\mathcal{X}_1 \cap \mathcal{X}_\cap}([L_k \;\; S_k]) < c^k d_{\mathcal{X}_1 \cap \mathcal{X}_\cap}([L_0 \;\; S_0])$$

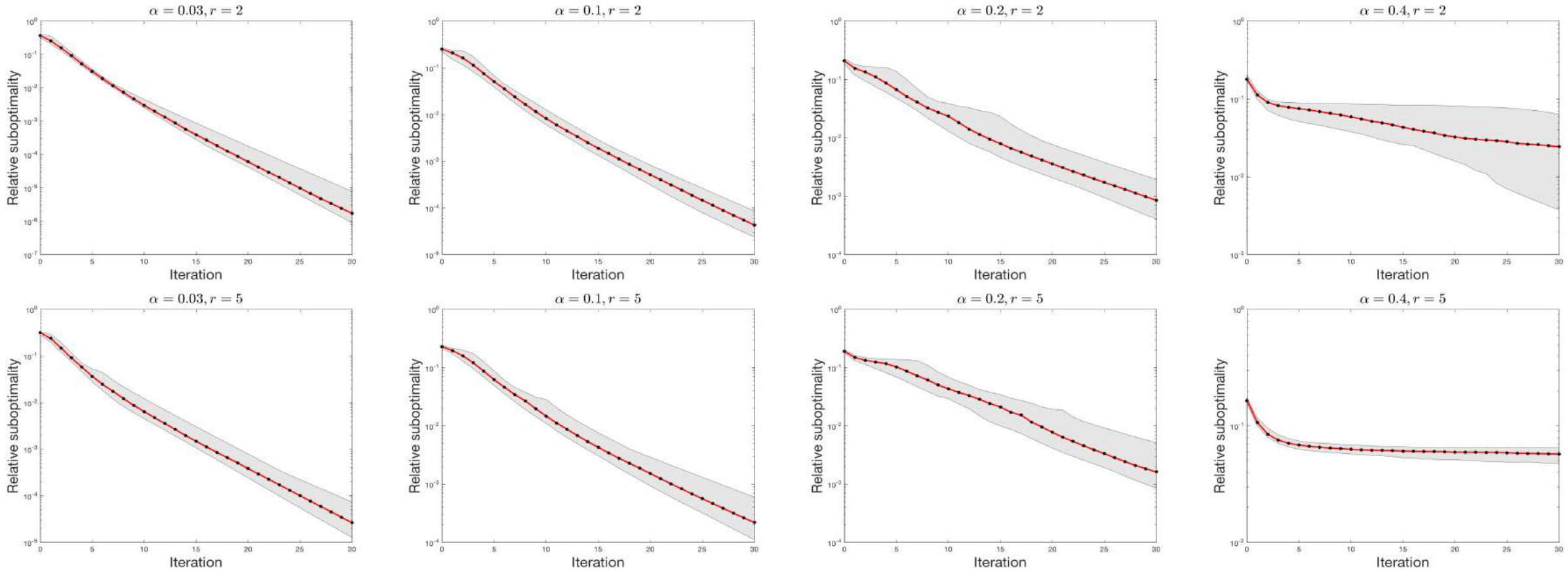
Cosine of angle between tangent spaces of $\mathcal{X}_1$, $\mathcal{X}_\cap$

Define $\mathcal{X}_\cap = \mathcal{X}_2 \cap \mathcal{X}_3$

Convex relaxation - ℓ_1 norm for sparsity and nuclear norm (sum of the singular values) for low-rank

Global linear rate

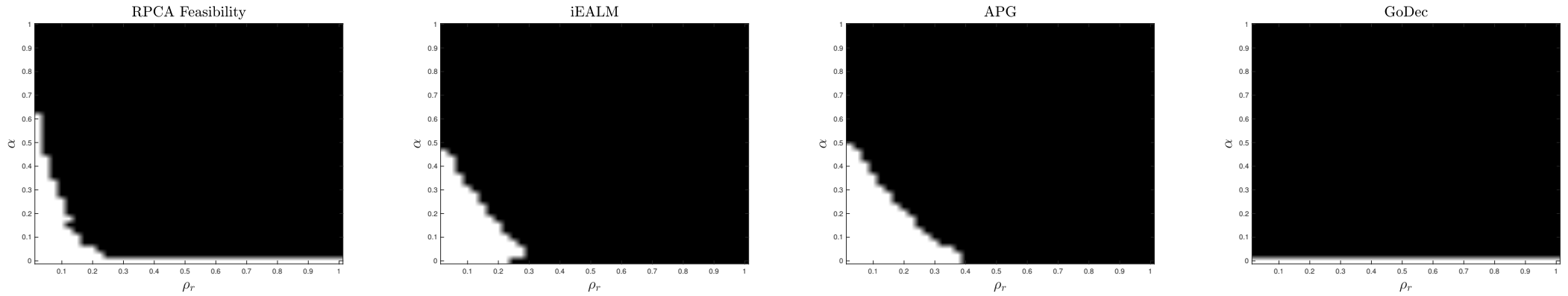
Convergence of Algorithm 2



Sensitivity of Algorithm 2 on the initialization. The best, the worst, and the median case are plotted for each iteration.

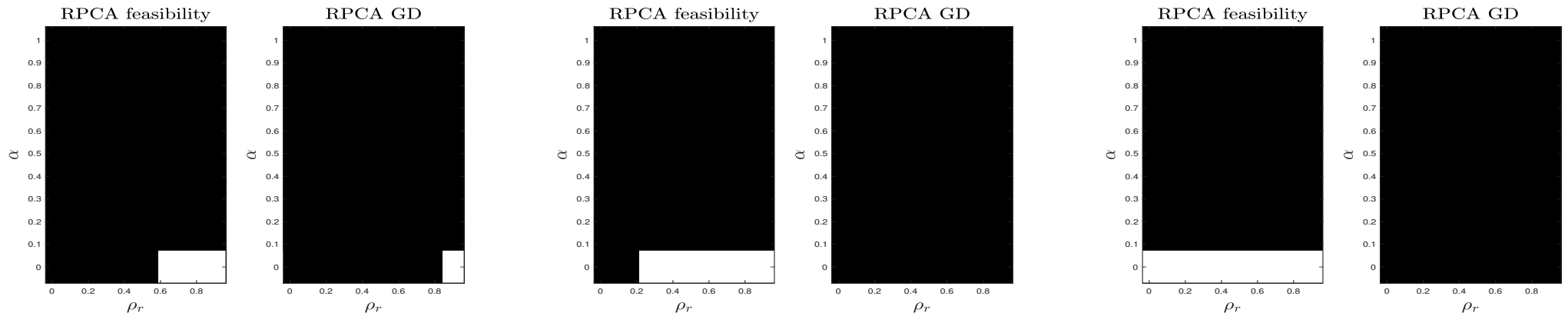
Phase Transition Diagrams - Results on Synthetic Data

White denotes success and black denotes failure



(a) (b) (c) (d)

Figure 1: Phase transition diagram for RPCA F, iEALM, APG, and GoDec with respect to rank and error sparsity. Here, $\rho_r = \text{rank}(L)/m$ and α is the sparsity measure. We have $(\rho_r, \alpha) \in (0.025, 1] \times (0, 1)$ with $r = 5 : 5 : 200$ and $\alpha = \texttt{linspace}(0, 0.99, 40)$. We perform 10 runs of each algorithm.



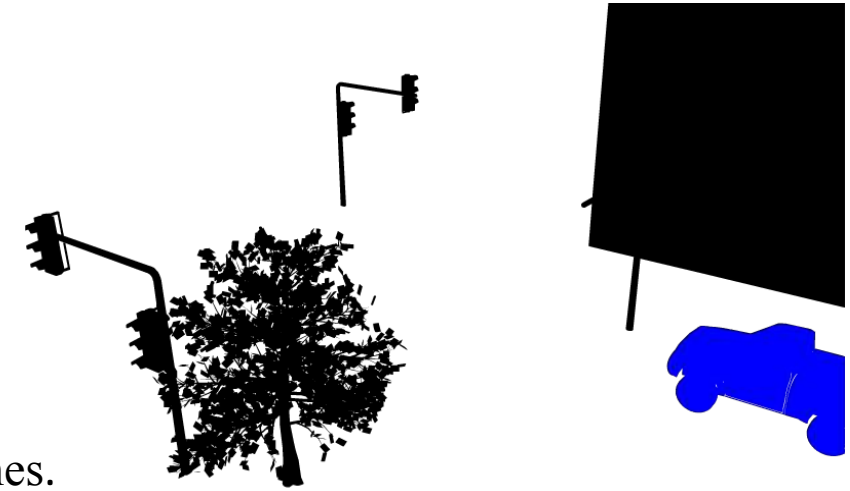
(a) (b) (c)

Figure 2: Phase transition diagram for Relative error for RMC problems: (a) $|\Omega^C| = 0.5(m.n)$, (b) $|\Omega^C| = 0.75(m.n)$, (c) $|\Omega^C| = 0.9(m.n)$. Here, $\rho_r = \text{rank}(L)/m$ and α is the sparsity measure. We have $(\rho_r, \alpha) \in (0.025, 1] \times (0, 1)$ with $r = 5 : 25 : 200$ and $\alpha = \texttt{linspace}(0, 0.99, 8)$.

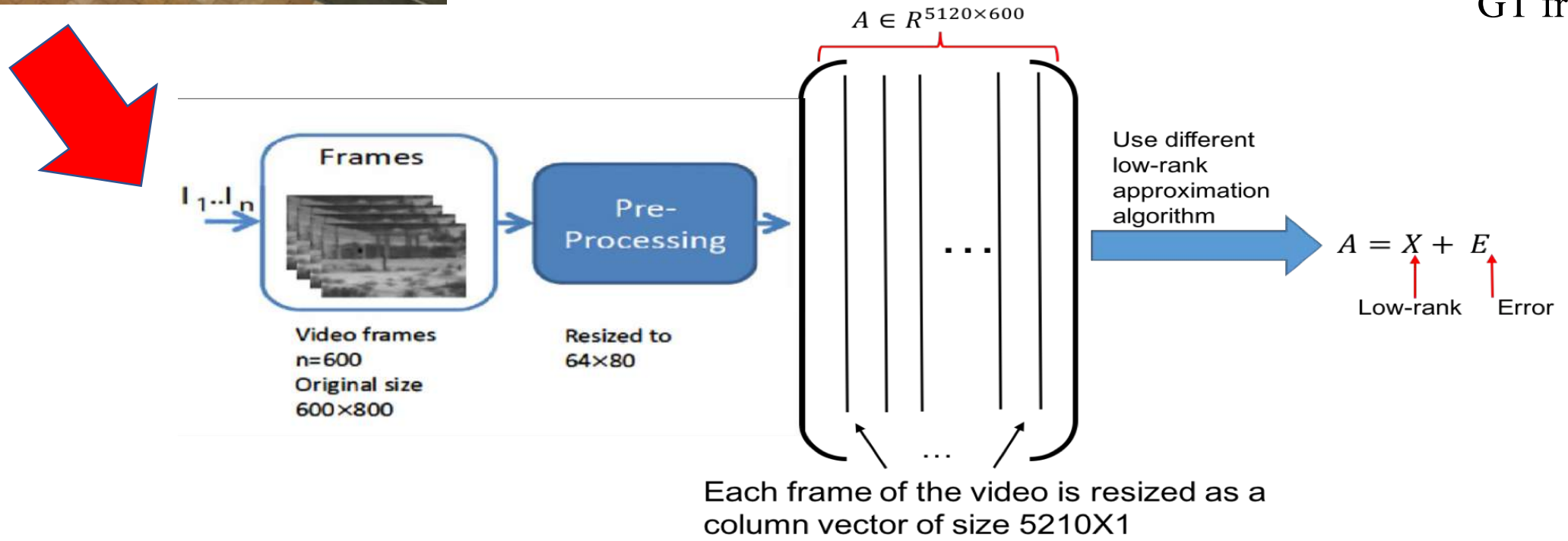
Numerical Experiments: On Stuttgart Synthetic Video Sequence

It is a computer generated sequence from the vantage point of a static camera located on the side of a building viewing a city intersection.

We choose the first 600 frames of the BASIC sequence to capture the changing illumination and foreground object. Correspondingly, we have 600 ground truth frames.



GT frame



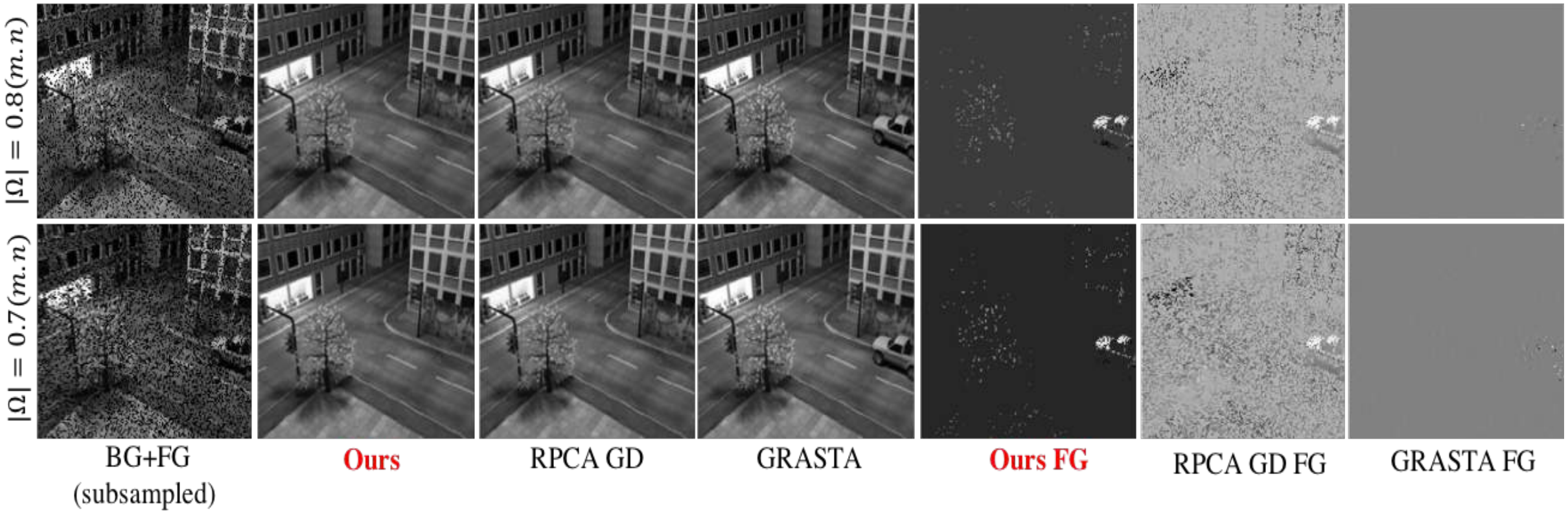
Video Demo on Background Estimation

BG+FG | **Low-rank BG (Ours)** | RPCA GD | GoDec | iEALM | APG | **Sparse (Ours)** | RPCA GD Sparse | GoDec Sparse | iEALM Sparse | APG Sparse

Background

Foreground
(Frames 553-600
static foreground)

Background Estimation on Subsampled Data: Application of Algorithm 3



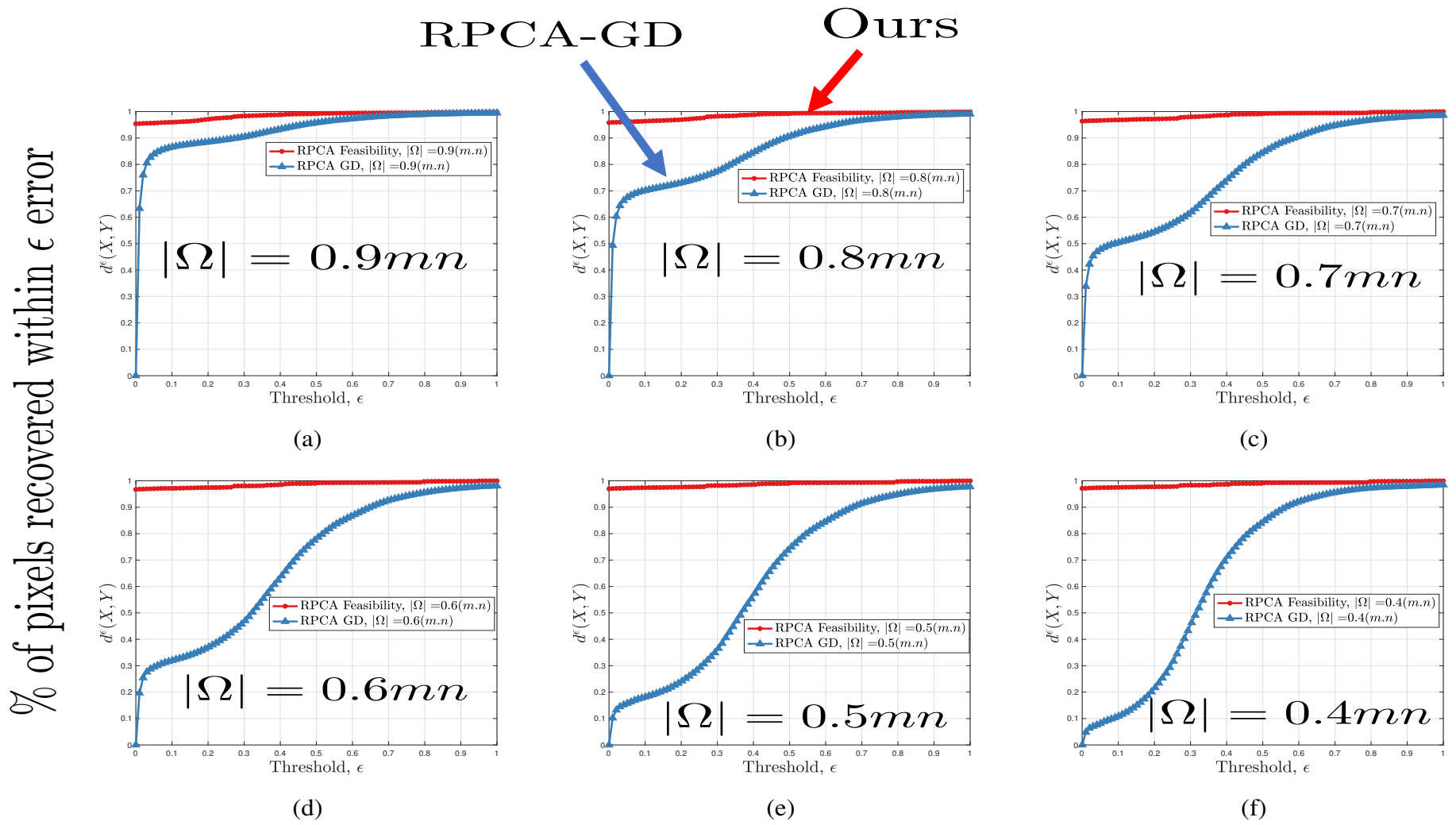


Figure 12: Quantitative comparison of foreground recovered by RPCA GD and RPCA F on `Basic` video, frame size 144×176 with observable entries: (a) $|\Omega| = 0.9(m.n)$, (b) $|\Omega| = 0.8(m.n)$, (c) $|\Omega| = 0.7(m.n)$, (d) $|\Omega| = 0.6(m.n)$, (e) $|\Omega| = 0.5(m.n)$, and (f) $|\Omega| = 0.4(m.n)$. The performance of RPCA GD drops significantly as $|\Omega|$ decreases. In contrast, the performance of RPCA F stays stable irrespective of the size of $|\Omega|$.

Removal of Shadows

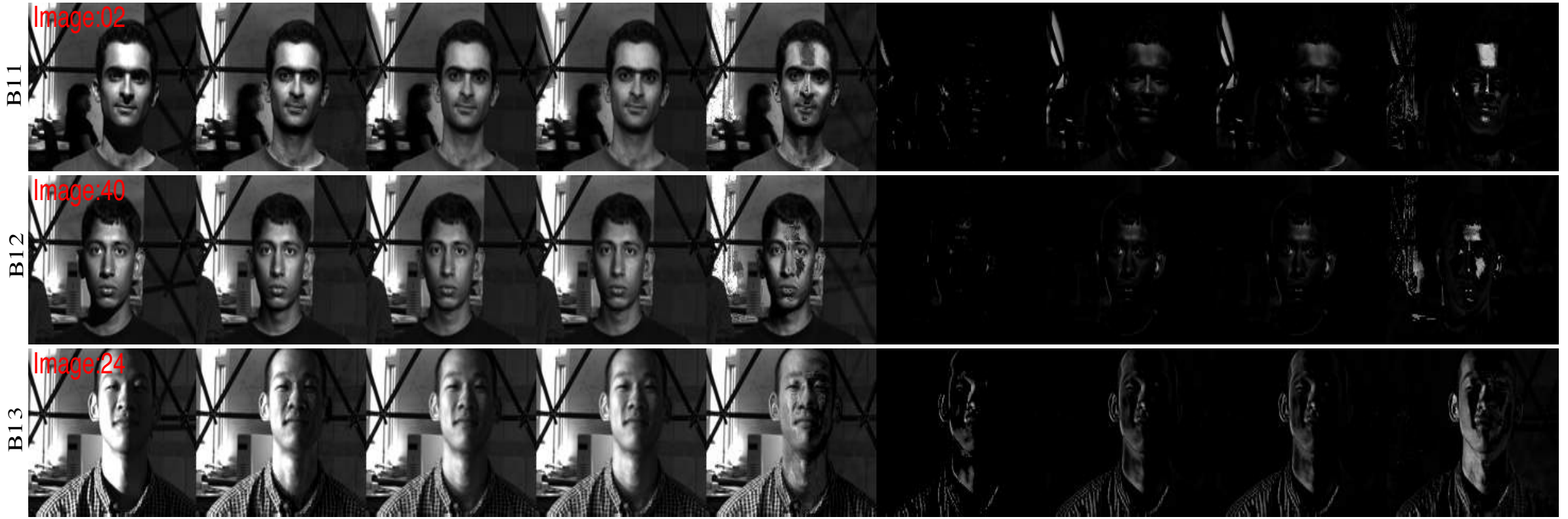


Figure 5: Shadow and specularities removal from face images captured under varying illumination and camera position. Our feasibility approach provides comparable reconstruction to that of iEALM and APG.

Inlier Detection

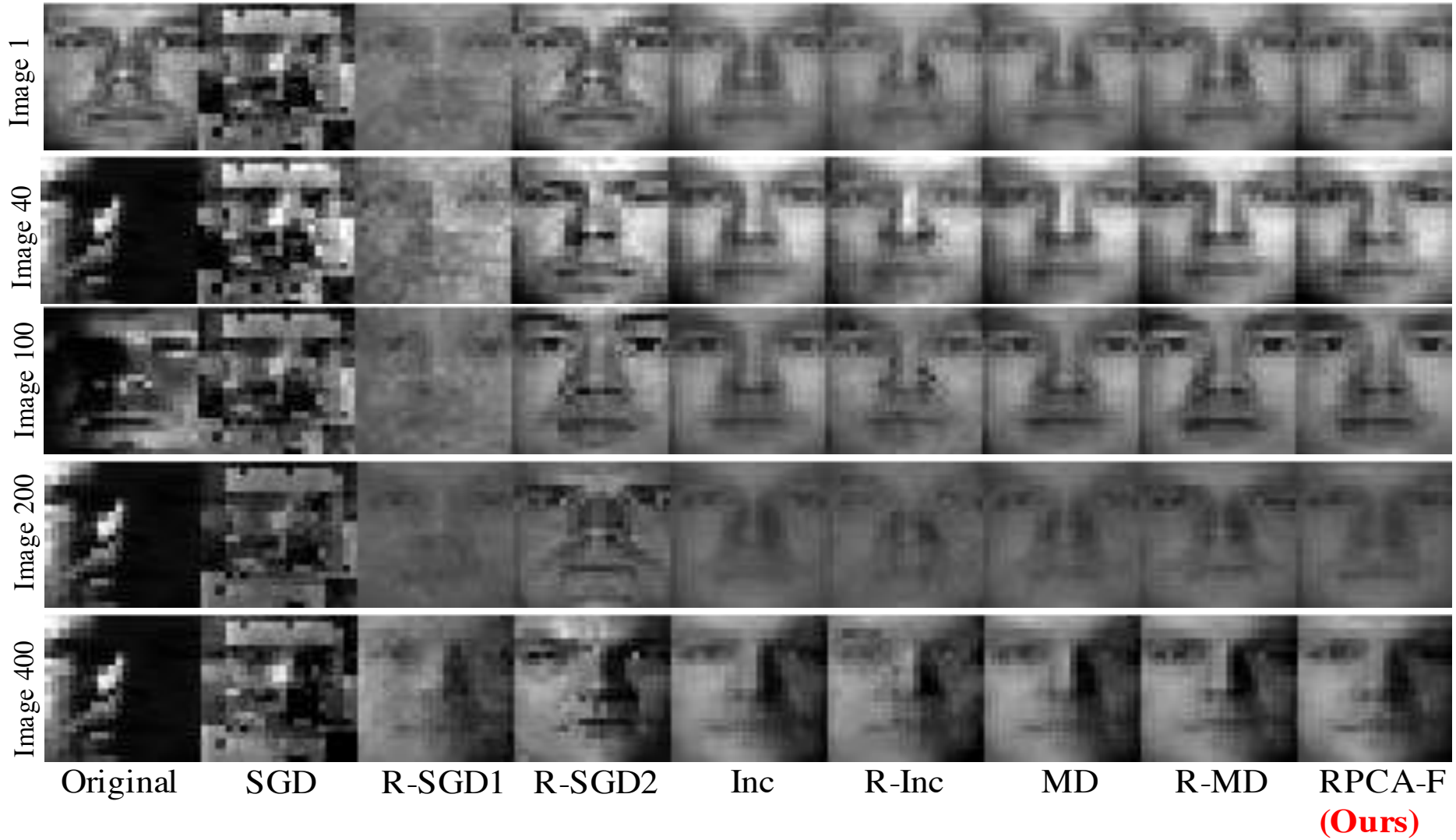


Figure 13: Inliers and outliers detection. Face images captured in different lighting conditions are inliers. We project different faces to 9 dimensional subspaces found by different methods.

Please come to our poster session **this evening**. Paper ID-5833.

Also see the paper online at:
https://www.aaai.org/Papers/AAAI/2019/AAAI-DuttaA.5833.pdf

Questions?