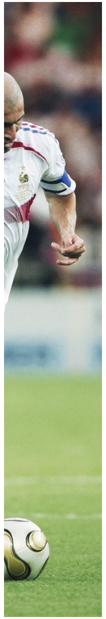
Object tracking with Robust PCA initialization and sparse recovery

Chenye Yang (cy2540)
Yanchen Liu (yl4189)

Outline

- Introduction
- Robust PCA
- L1 APG tracker
    1) Math Model
    2) Solver: modified accelerated proximal gradient
- Image super-resolution
    1) Math Model
    2) Methods Description
- Limitation & Summary
- Reference

Introducion: Problem



Introduction: Solution

Object detection:

- Robust PCA - background segmentation as initialization
- L1 tracker - object detection and tracking

Image Super-Resolution:

- Sparse representation - image super-resolution reconstruction

Robust PCA

PCA: Observation is sum of ROI and i.d.d. Gaussian Noise matrix

Robust PCA: Observation is sum of low rank matrix and sparse ‘noise’ matrix

Rank of matrix → Nuclear norm
Sparsity of matrix → L0 norm → L1 norm

minimize ||L||*+||S||_1
s.t. O = L+S

Experiment: Robust PCA

L1 Tracker

A method based on particle filter.

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}, \quad (1)$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}. \quad (2)$$

$$\mathbf{x}_t^* = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}_{1:t}).$$

$\mathbf{T}_t = [\mathbf{t}_t^1, \mathbf{t}_t^2, \cdots, \mathbf{t}_t^n]$, Target template $\mathbf{S}_t = \{\mathbf{x}_t^1, \mathbf{x}_t^2, \cdots, \mathbf{x}_t^N\}$ Sample $\mathbf{O}_t = \{\mathbf{y}_t^1, \mathbf{y}_t^2, \cdots, \mathbf{y}_t^N\}$ Observation

$$\mathbf{y}_t^i = \mathbf{T}_t\mathbf{a}_T^i + I\mathbf{a}_I^i, \quad \forall \mathbf{y}_t^i \in \mathbf{O}_t, \quad (4)$$

$$\mathbf{a}_t^i = [\mathbf{a}_T^i; \mathbf{a}_I^i]$$

$$\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{y}_t^i - A\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1, \ \mathbf{a} \succcurlyeq 0, \quad (5)$$

$$A = [\mathbf{T}_t, I, -I].$$

$$p(\mathbf{z}_t|\mathbf{x}_t^i) = \frac{1}{\Gamma}\exp\{-\alpha\|\mathbf{y}_t^i - \mathbf{T}_t\mathbf{c}_T^i\|_2^2\}, \qquad (6)$$

$$\mathbf{x}_t^* = \arg\max_{\mathbf{x}_t^i \in \mathbf{S}_t} p(\mathbf{z}_t|\mathbf{x}_t^i).$$

$$\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{y} - A'\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1 + \frac{\mu_t}{2}\|\mathbf{a}_I\|_2^2, \quad \text{s.t. } \mathbf{a}_T \succcurlyeq 0, \qquad A' = [\mathbf{T}_t, I], \mathbf{a} = [\mathbf{a}_T; \mathbf{a}_I]$$

APG method is for optimization problem with no constraints, to transfer the formula above, import indicator function

$$\mathbf{1}_{\mathbb{R}_+^N}(\mathbf{a}) = \begin{cases} 0, & \mathbf{a} \succeq 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

$$\arg\min_{\mathbf{a}} \frac{1}{2}\|\mathbf{y} - A'\mathbf{a}\|_2^2 + \lambda\mathbf{1}_T^\top\mathbf{a}_T + \|\mathbf{a}_I\|_1 + \frac{\mu_t}{2}\|\mathbf{a}_I\|_2^2 + \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{a}_T).$$

$$F(\mathbf{a}) = \frac{1}{2}\|\mathbf{y} - A'\mathbf{a}\|_2^2 + \lambda\mathbf{1}_T^\top\mathbf{a}_T + \frac{\mu_t}{2}\|\mathbf{a}_I\|_2^2,$$
$$G(\mathbf{a}) = \|\mathbf{a}_I\|_1 + \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{a}_T).$$

L1 Tracker

(i) Set $\alpha_0 = \alpha_{-1} = \mathbf{0} \in \mathbb{R}^N$ and set $t_0 = t_{-1} = 1$.

(ii) For $k = 0, 1, \ldots,$ iterate until convergence

$$\begin{cases} \beta_{k+1} := \alpha_k + \frac{t_{k-1}-1}{t_k}(\alpha_k - \alpha_{k-1}); \\ \alpha_{k+1} := \underset{\mathbf{a}}{\arg\min}\ \frac{L}{2}\|\mathbf{a} - \beta_{k+1} + \frac{\nabla F(\beta_{k+1})}{L}\|_2^2 + G(\mathbf{a}); \\ t_{k+1} := \frac{1+\sqrt{1+4t_k^2}}{2}. \end{cases}$$

$$\hat{\mathbf{a}} = \arg\min \|\mathbf{T}_t\mathbf{a} - \mathbf{y}\|_2^2. \qquad (9)$$

$$p(\mathbf{z}_t|\mathbf{x}_t^i) \leq \frac{1}{\Gamma}\exp\{-\alpha\|\mathbf{T}_t\hat{\mathbf{a}} - \mathbf{y}_t^i\|_2^2\} \triangleq q(\mathbf{z}_t|\mathbf{x}_t^i), \qquad (10)$$

Input:
Current frame F_t;
Sample Set $\mathbf{S}_{t-1} = \{\mathbf{x}_{t-1}^i\}_{i=1}^N$;
Template set $\mathbf{T} = \{\mathbf{t}_i\}_{i=1}^n$.
for $i = 1$ to N **do**
  Drawing the new sample $\mathbf{x}_t^i$ from $\mathbf{x}_{t-1}^i$;
  Preparing the candidate patch $\mathbf{y}_t^i$ in template space;
  Solving the least square problem (9);
  Computing q_i according to (10);
end for
Sorting the samples in descent order according to q;
Setting $i = 1$ and $\tau = 0$.
while $i < N$ and $q_i \geq \tau$ **do**
  Solving the minimization (11) via Algorithm 2;
  Computing the observation likelihood p_i in (6);
  $\tau = \tau + \frac{1}{2N}p_i$;
  $i = i + 1$;
end while
Set $p_j = 0, \forall j \geq i$.
Output:
Finding the $\mathbf{x}_t^*$ according to (7);
Detecting the occlusion [16] and update μ in (11);
Updating the template set $\mathbf{T}_{t-1}$ [16];
Updating the sample set $\mathbf{S}_{t-1}$ with p.

Experiment: L1 Tracker

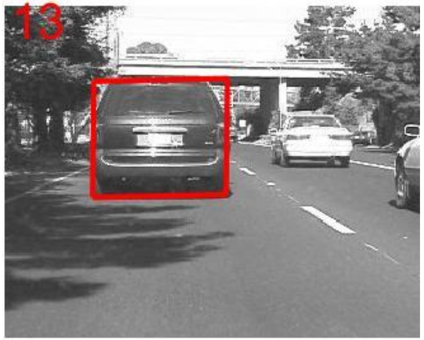


Image Super-Resolution

x is the image with high-resolution, y is the image with low-resolution.

$$y = SHx$$

H blurring filter, S downsampling operator

Images can be represented as a sparse linear combination in a over-complete dictionary

$$x = D_h \times \alpha_h, \;\; y = D_l \times \alpha_l$$

D over-complete dictionary, α sparse representation coefficient

The idea is to train the dictionaries which lead to same coefficient for both high and low-resolution images. Once we get the dictionaries and testing low-resolution image, coefficient can be calculated. Then according to high-resolution dictionary, high-resolution image can be recovered.

$$y = D_l \times \alpha, \;\; x = D_h \times \alpha$$

Training:

Local (dictionary matches image)

$$\min \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \begin{aligned} &\|F\boldsymbol{D}_l\boldsymbol{\alpha} - F\boldsymbol{y}\|_2^2 \le \epsilon_1, \\ &\|P\boldsymbol{D}_h\boldsymbol{\alpha} - \boldsymbol{w}\|_2^2 \le \epsilon_2, \end{aligned}$$

Global (recovery matches original)

$$\boldsymbol{X}^* = \arg\min_{\boldsymbol{X}} \|SH\boldsymbol{X} - \boldsymbol{Y}\|_2^2 + c\|\boldsymbol{X} - \boldsymbol{X}_0\|_2^2$$

Experiment: Image Super-Resolution

Original
360x240

Low-res
90x60

Recovery
360x240

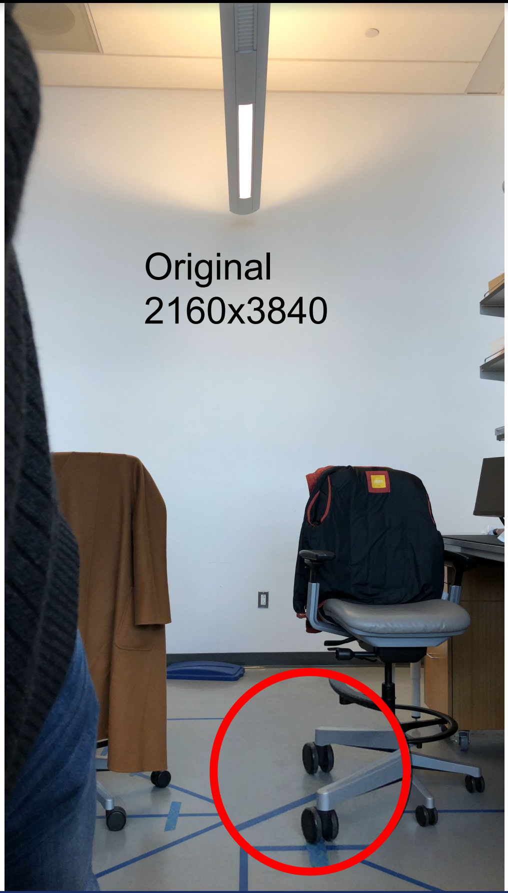
Experiment: Image Super-Resolution

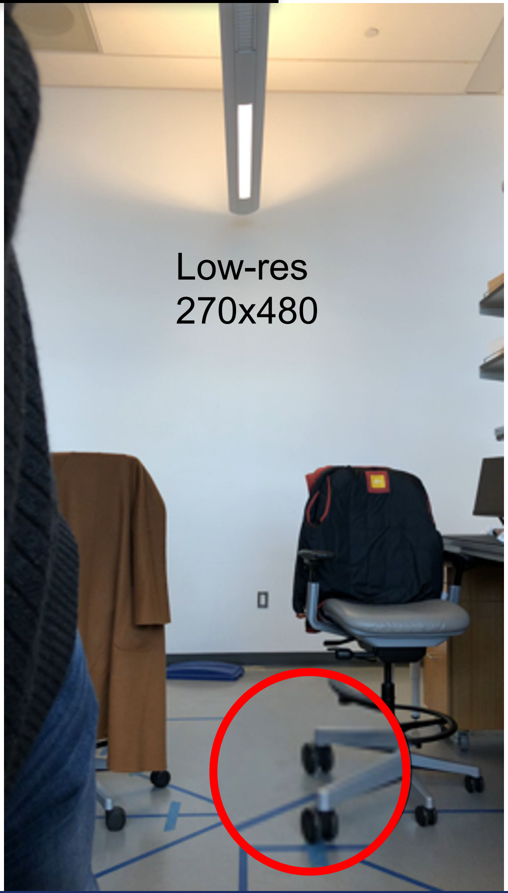
Original
640x360

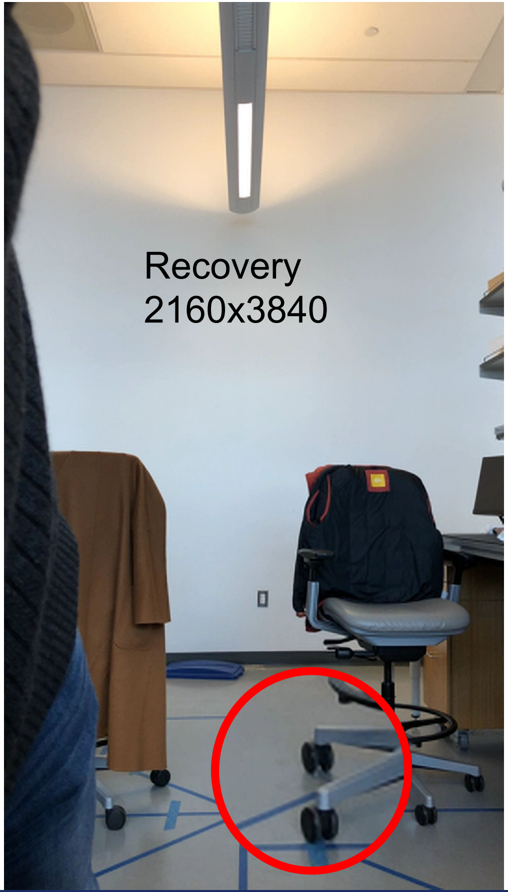
Low-res
160x90

Recovery
640x360

Experiment: Image Super-Resolution







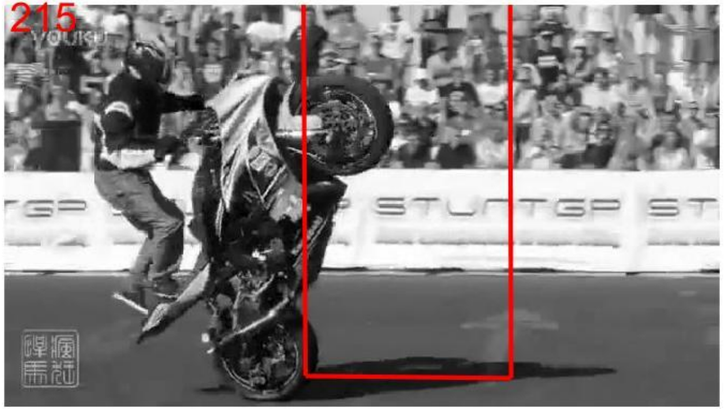
Limitation & Summary

For Robust PCA initialization, if the background is not that stable in the first few seconds and the object we want to track is not that obvious, we may not get the object box we need for L1 Tracker.

For L1 Tracker, it works not that well when the tracking object rotating.

For Image Super-Resolution via sparse representation, if the training images are clear/vague (not equals high resolution), the recovered images are clear/vague (recall 4k image and motorbike image). Moreover, training and recovering takes a long time (highly related to resolution).

Limitation & Summary



Reference

Mei, Xue, and Haibin Ling. "Robust visual tracking using ℓ 1 minimization." *2009 IEEE 12th International conference on computer vision*. IEEE, 2009.

Yang, Jianchao, et al. "Image super-resolution via sparse representation." *IEEE transactions on image processing* 19.11 (2010): 2861-2873.

C. Bao, Y. Wu, H. Ling, and H. Ji. Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach. In CVPR, 2012

E. Cands, X. Li, Y. Ma and J. Wright, "Robust Principal Component Analysis?", *J. ACM*, 2009

Thank you