Robust Late Fusion With Rank Minimization

Guangnan Ye[1], Dong Liu[1], I-Hong Jhuo[1,2], Shih-Fu Chang[1]

[1]Dept.of Electrical Engineering, Columbia University
[2]Dept.of Computer Science and Information Engineering, National Taiwan University

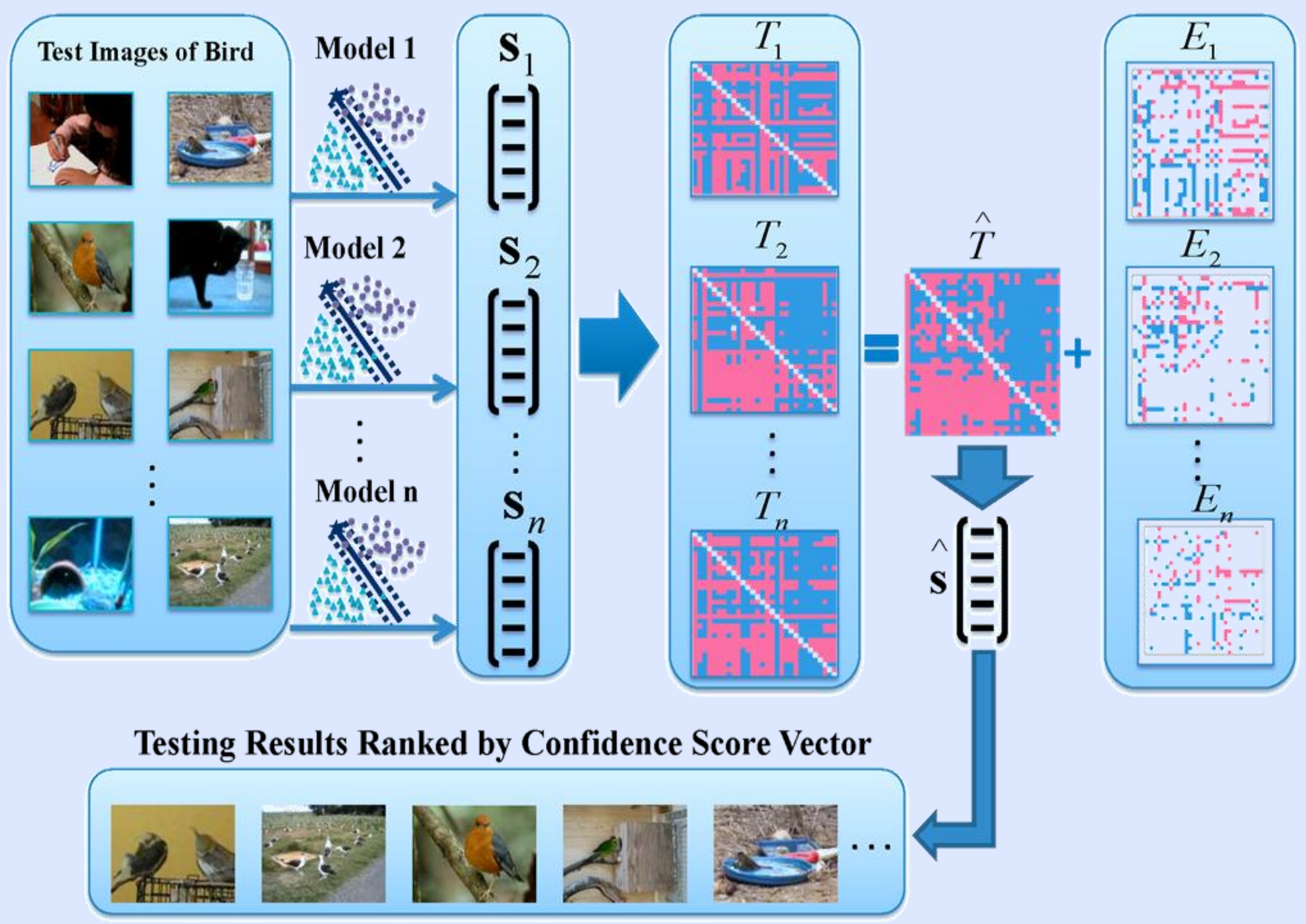
Motivation

- Late Fusion : Combine the prediction scores of multiple models.
- Issues: (1) Scales of scores from the individual models may vary a lot
  (2) Scores from each model may contain noise and outliers

Approach Overview

Observation:

- Preserve rank order relationship among scores instead of absolute values;
- If we have a real-value matrix $\hat{T}$ such that $\hat{T}_{jk} = \hat{s}_j - \hat{s}_k$, we can find a rank-2 factorization of $\hat{T}$ such that $\hat{T} = \hat{\mathbf{s}}\mathbf{e}^\top - \mathbf{e}\hat{\mathbf{s}}^\top$.



Steps:

- Convert each confidence score vector into a pairwise rank relationship matrix to address the scale variance issue;
- Seek a shared rank-2 pairwise matrix based on which each score matrix can be decomposed into the consistent rank-2 matrix and sparse errors;
- A robust score vector is then extracted to fit the recovered low rank score rank relation matrix.

Problem Formulation

$$\min_{\hat{T},E_i} \|\hat{T}\|_* + \lambda \sum_{i=1}^{n} \|E_i\|_1, \quad \text{s.t. } T_i = \hat{T} + E_i,\ i = 1,\dots,n, \quad \hat{T} = -\hat{T}^\top.$$

$$T_{ij} = \begin{cases} -1, & \text{if } \mathbf{s}_i < \mathbf{s}_j, \\ 0, & \text{if } \mathbf{s}_i = \mathbf{s}_j, \\ 1, & \text{if } \mathbf{s}_i > \mathbf{s}_j. \end{cases}$$

Optimization and Score Recovery

Theorem: Given a set of skew-symmetric matrices, the solution from SVT solver is a skew-symmetric matrix if the spectrums between the dominant singular values are separated.
The skew-symmetric constraint can be ignored.

Equivalent Form:

$$\min_{\hat{T},E_i} \|\hat{T}\|_* + \lambda \sum_{i=1}^{n} \|E_i\|_1 + \sum_{i=1}^{n} \langle Y_i, T_i - \hat{T} - E_i \rangle + \frac{\mu}{2} \sum_{i=1}^{n} \|T_i - \hat{T} - E_i\|_F^2,$$

Algorithm 1 Solving Problem by Inexact ALM

1: **Input:** Comparative relationship matrix T_i, $i = 1, 2, \dots, n$, parameter λ, number of samples m.
2: **Initialize:** $\hat{T} = 0$, $E_i = 0$, $Y_i = 0$, $i = 1, \dots, n$, $\mu = 10^{-6}$, $max_\mu = 10^{10}$, $\rho = 1.1$, $\varepsilon = 10^{-8}$.
3: **repeat**
4: Fix the other term and update $\hat{T}$ by $(U, \Lambda, V) = SVD(\frac{1}{n\mu}\sum_{i=1}^n Y_i + \frac{1}{n}\sum_{i=1}^n T_i - \frac{1}{n}\sum_{i=1}^n E_i))$, $\hat{T} = U\mathcal{S}_{\frac{1}{\mu}}[\Lambda]V^\top$, where $\mathcal{S}$ is a shrinkage operator for singular value truncating defined as:

$$\mathcal{S}_\varepsilon[x] = \begin{cases} x - \varepsilon, & \text{if } x > \varepsilon, \\ x + \varepsilon, & \text{if } x < -\varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

5: Fix the other term and update E_i by $E_i = \mathcal{S}_{\frac{\lambda}{\mu}}[T_i + \frac{Y_i}{\mu} - \hat{T}]$.
6: Update the multipliers $Y_i = Y_i + \mu(T_i - \hat{T} - E_i)$.
7: Update the parameter μ by $\mu = \min(\rho\mu, max_\mu)$.
8: **until** $\max_i \|T_i - \hat{T} - E_i\|_\infty < \varepsilon$ and $\mathrm{rank}(\hat{T}) = 2$.
9: **Output:** $\hat{T}$.

Score Recovery:

$$(1/m)\hat{T}\mathbf{e} = \arg\min_{\hat{\mathbf{s}}} \|\hat{T}^\top - (\hat{\mathbf{s}}\mathbf{e}^\top - \mathbf{e}\hat{\mathbf{s}}^\top)\|_F^2$$

Extension with Graph Laplacian:

$$\min_{\hat{T},E_i} \|\hat{T}\|_* + \lambda \sum_{i=1}^{n} \|E_i\|_1 + \gamma \sum_{i=1}^{n} \Psi^i(\hat{T}),$$

$$s.t.\ T_i = \hat{T} + E_i,\ i = 1, \dots, n, \hat{T} = -\hat{T}^\top.$$

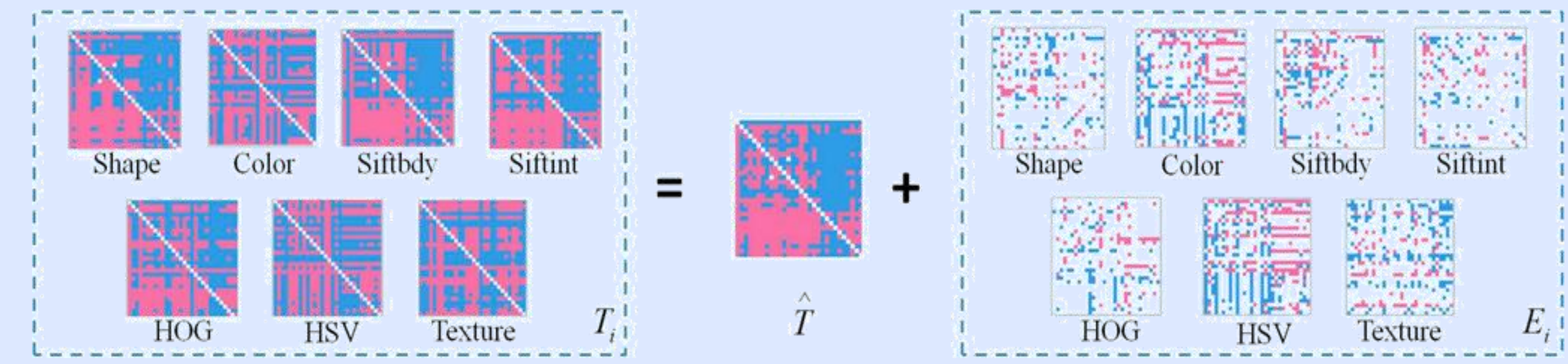
$$\Psi^i(\hat{T}) = \frac{1}{2} \sum_{j,k=1}^{m} P^i_{jk} \|\hat{\mathbf{t}}_j - \hat{\mathbf{t}}_k\|_2^2 = tr(\hat{T}^\top L^i \hat{T})$$

Experiments

- **Oxford Flower 17**

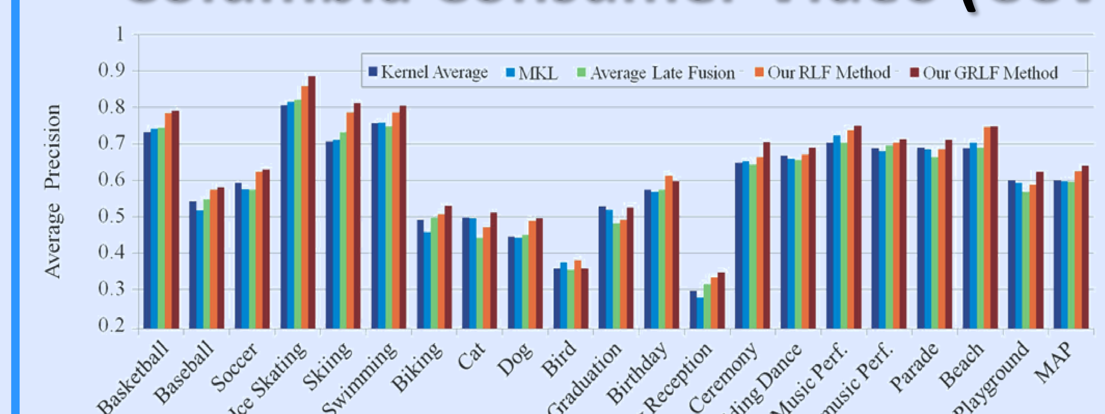
Method	MAP
SIFTint	0.749 ± 0.013
Kernel Average	0.860 ± 0.017
MKL	0.863 ± 0.021
Average Late Fusion	0.869 ± 0.021
Our RLF Method	0.898 ± 0.019
Our GRLF Method	0.917 ± 0.017

Table 1: MAP comparison, the proposed method achieves 5.5% gain over the best baseline

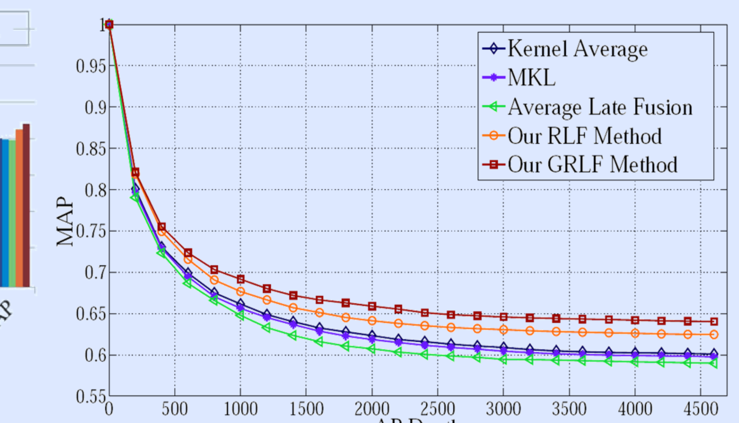


Visualization of the low rank and sparse matrices

- **Columbia Consumer Video (CCV) Dataset**

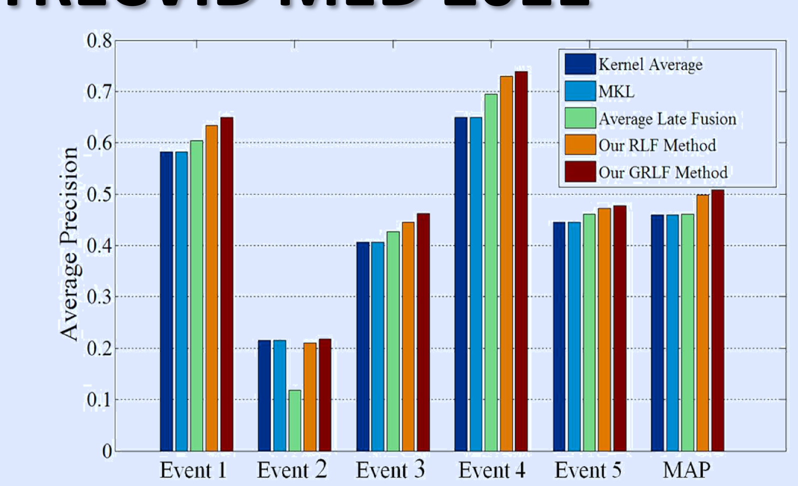


6.6% gain over the best baseline

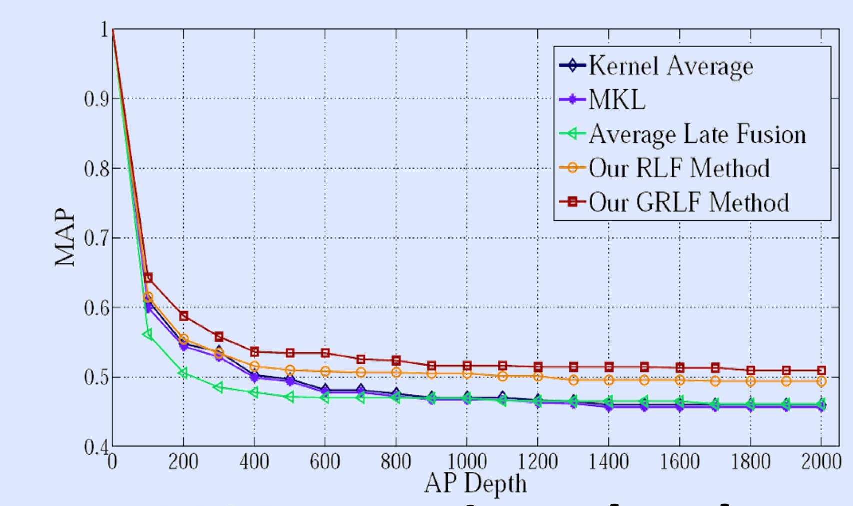


MAP at variant depths

- **TRECVID MED 2011**



10.4% gain over the best baseline



MAP at variant depths

Conclusions

- Robust late fusion discovers a consistent pattern shared among models while solving the score scale variation and noise issues.
- Experiments confirm that the proposed method can robustly extract a rank-2 skew-symmetric matrix and sparse errors.
- Robust late fusion achieves 5.5%, 6.6%, and 10.4% improvement in Oxford Flower 17, CCV, and TRECVID.