Jia-geng Feng, Jun Xiao, 2015. View-invariant human action recognition via robust locally adaptive multi-view learning. *Frontiers of Information Technology & Electronic Engineering*, **16**(11):917-929. [doi:10.1631/FITEE.1500080]

View-invariant human action recognition via robust locally adaptive multi-view learning

Key words: View-invariant, Action recognition, Multi-view learning, L1-norm, Local learning

Contact: Jia-geng Feng
E-mail: fengjiageng@126.com

ORCID: http://orcid.org/0000-0003-4577-4520

Motivation/Main ideas

- **Motivation**

  Some extrinsic factors are barriers for the development of action recognition; e.g., human actions may be observed from arbitrary camera viewpoints in realistic scene. Thus, view-invariant analysis becomes important for action recognition algorithms, and a number of researchers have paid much attention to this issue.

- **Main ideas**

  We present a multi-view learning approach to recognize human actions from different views.

  A robust locally adaptive multi-view learning algorithm based on learning multiple local L1-graphs is proposed.

  An efficient iterative optimization method is proposed to solve the proposed objective function.

Method (I)

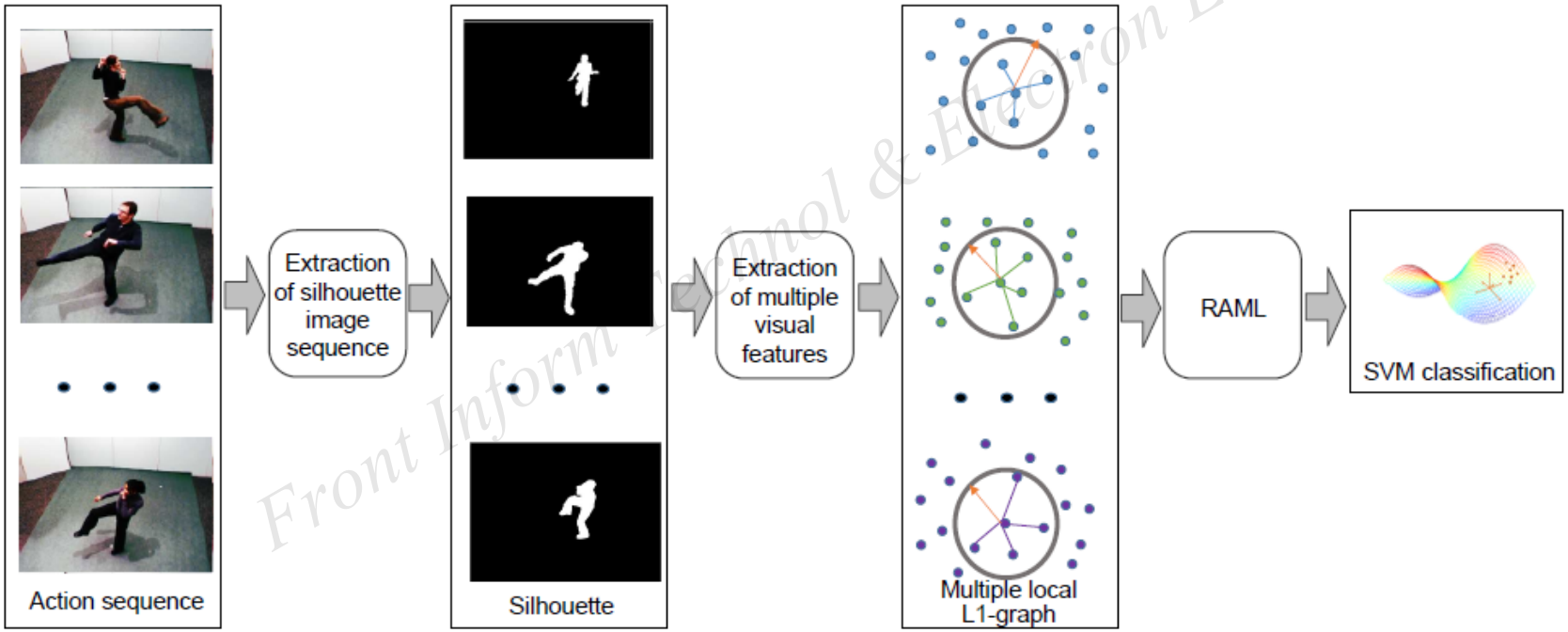


Fig. 1 Flowchart of our proposed RAML algorithm

Method (II)

Single-view robust local L1-graph

$$\min_{w} \| \boldsymbol{w}_p \|_1 \quad \text{s.t.} \quad \boldsymbol{x}_p^i = \boldsymbol{X}_M^i \boldsymbol{w}_p,$$

where $\boldsymbol{X}_M^i = [\boldsymbol{x}_{i_1}^i, \boldsymbol{x}_{i_2}^i, \ldots, \boldsymbol{x}_{i_M}^i] \in \mathbb{R}^{d_i \times M}$ ($i_1, i_2, \ldots, i_M \leq N$, $M \leq N$) represents the M neighbors of $\boldsymbol{x}_p^i$ under the ith feature representation.

$$W_{jk}^i = \begin{cases} w_{jk}^i, & j > k, \\ w_{j(k-1)}^i, & j < k, \\ 0, & j = k. \end{cases}$$

Method (II)

- Objective function

$$\arg\min_{Y,\alpha} \sum_{i=1}^{m} \alpha_i^r \operatorname{tr}(\boldsymbol{Y}\boldsymbol{L}_n^i\boldsymbol{Y}^{\mathrm{T}})$$

$$\text{s.t. } \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{I}, \sum_{i=1}^{m} \alpha_i = 1, \ \alpha_i \geq 0.$$

$$\boldsymbol{L}_j^i = \begin{bmatrix} -\boldsymbol{e}_k^{\mathrm{T}} \\ \boldsymbol{I}_k \end{bmatrix} \operatorname{diag}(\boldsymbol{w}_j^i)[-\boldsymbol{e}_k, \boldsymbol{I}_k] = \begin{bmatrix} \sum_{l=1}^{k} (w_j^i)_l, & -(\boldsymbol{w}_j^i)^{\mathrm{T}} \\ -\boldsymbol{w}_j^i, & \operatorname{diag}(\boldsymbol{w}_j^i) \end{bmatrix}.$$

Method (III)

Algorithm 1 Robust locally adaptive multi-feature learning

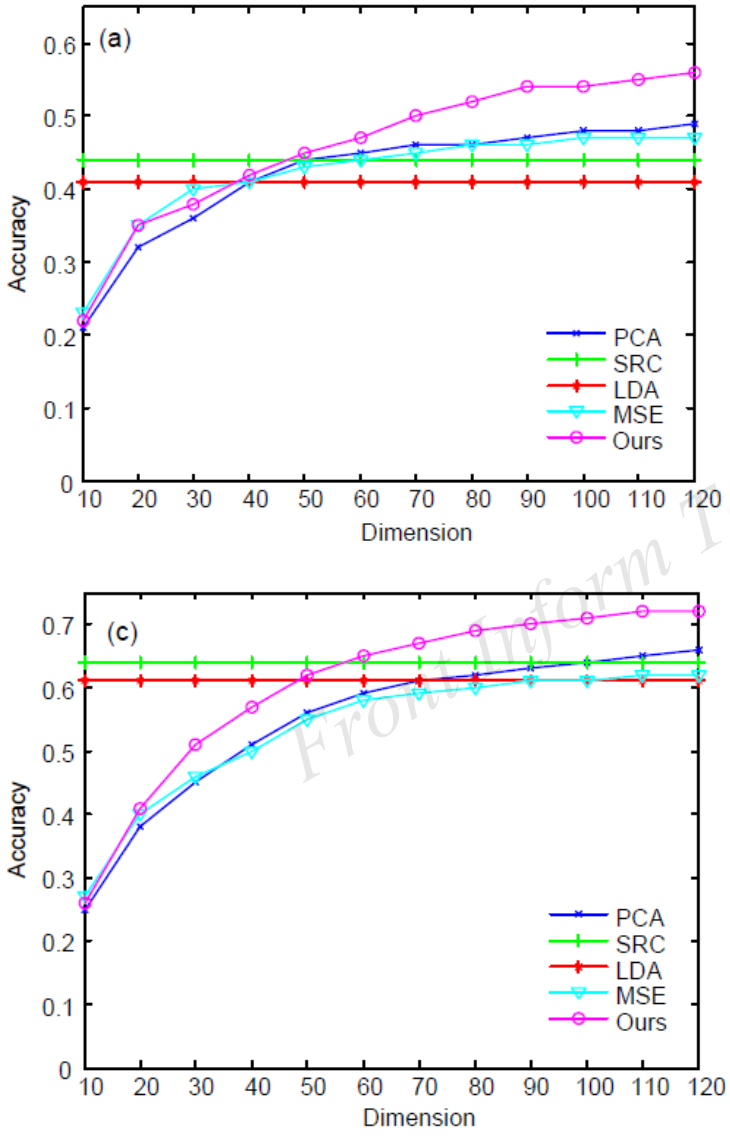
Input: the input dataset $X = \{\boldsymbol{X}^i \in \mathbb{R}^{d_i \times N}\}_{i=1}^{m}$, visual feature dimension d, and power value r.

Output: $\boldsymbol{Y} \in \mathbb{R}^{d \times n}$, where $d_i \geq d$ $(1 \leq i \leq m)$.

Optimization procedure:

1. Search M neighbors for each datum in X.
2. Construct local L1-graph by solving Eq. (2).
3. Calculate $\boldsymbol{L}$ according to Eq. (6).
4. Calculate $\boldsymbol{G}^i$ $(1 \leq i \leq m)$ for each feature (initially $\alpha = [1/m, 1/m, \ldots, 1/m]$).
5. Do the following iteration until reaches convergence:
   (1) $\boldsymbol{Y} = \boldsymbol{U}^{\mathrm{T}}$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d]$ ($\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_d$ are the corresponding eigenvectors of d smallest eigenvalues of matrix $\boldsymbol{G}$).
   (2) Calculate α_i (Eq. (16)).

Major results (I)



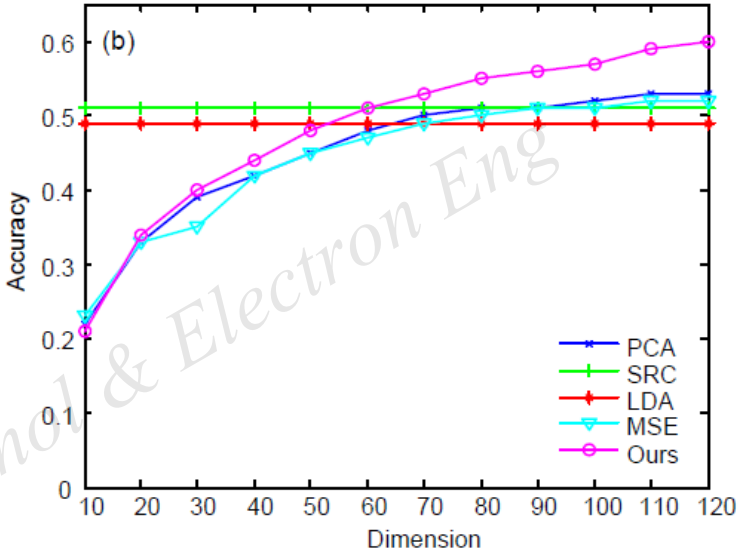
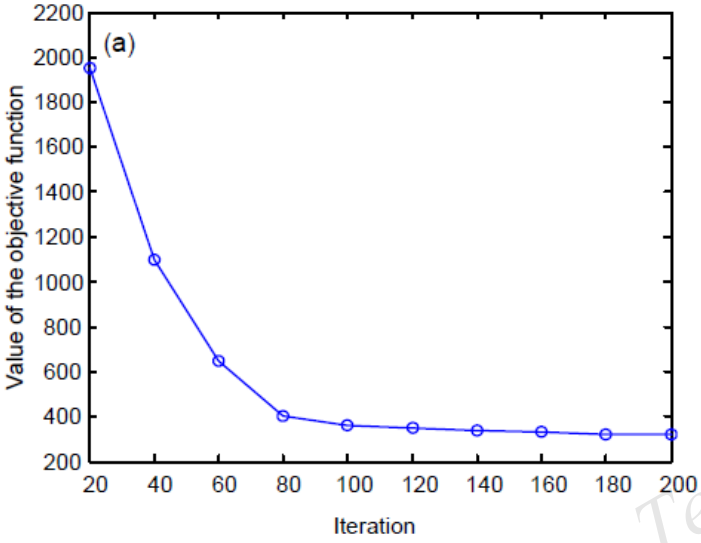
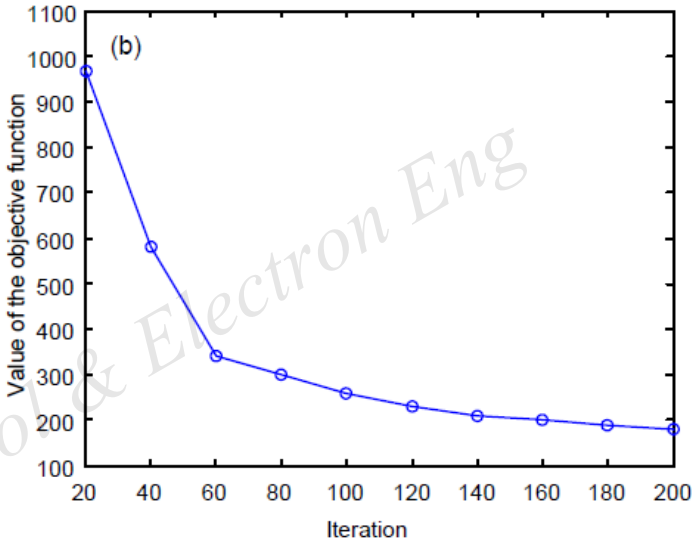


Fig. 4 Performance comparison results of different algorithms on WVU (a), ViHASi (b), and IXMAS (c)

Major results (II)





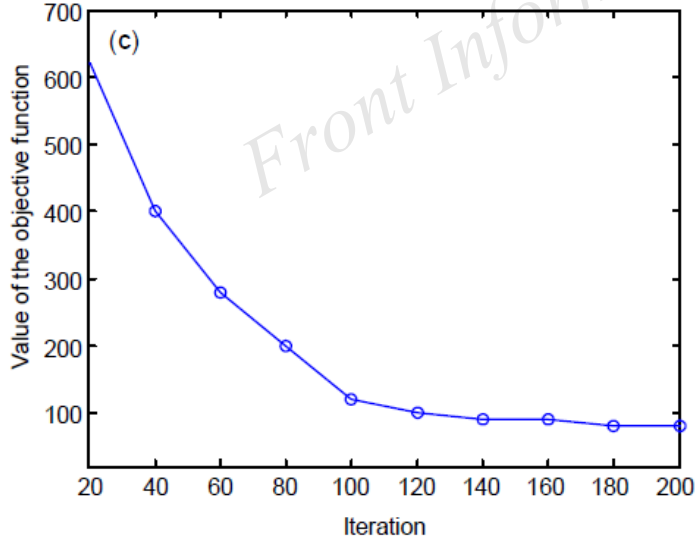


Fig. 5 Convergence curves on WVU (a), ViHASi (b), and IXMAS (c)

Conclusions

- We have applied the proposed RAML algorithm on three public datasets to solve the view-invariant human action recognition problem. Compared with the other methods, the RAML algorithm consistently outperforms the other methods if the selected feature dimension is higher than 60. Meanwhile, we notice that the RAML algorithm involves solving eigenvalue decomposition problem, which is time-consuming. So, there are two key issues that should be carefully considered in the further work. One is how to reduce the computation cost, and the other one is how to apply the RAML algorithm to deal with large-scale real-world problems.