DISCRIMINATIVE SALIENCY-POSE-ATTENTION COVARIANCE FOR ACTION RECOGNITION

Jianhai Zhang, Zhiyong Feng, Yong Su, Meng Xing

College of Intelligence and Computing, Tianjin University

MOTIVATION

Most covariance-based representations of actions are focused on statistical features of poses by empirical averaging weighting. Note that the poses have a variety of saliency levels for different actions. Neglecting pose saliency could make the covariance-based feature representations inaccurate and further degrade the performance of recognition tasks.

PROPOSED METHOD

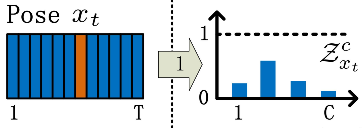
In this paper, we propose a novel saliency weighting covariance feature representation, Saliency-Pose-Attention Covariance(SPA-Cov), which reduces the negative effects from the ambiguous pose samples. Specifically, we utilize a discriminative approach to derive probability distribution of action categories for each pose, which is modeled by the uncertainty of information entropy to obtain the salient weighting.

Empirical Covariance(Empirical-Cov)

$$\Sigma_{\mathbf{D_A}} = \frac{1}{T}\sum_{t=1}^{T}\mathcal{F}(x_t-\mu)$$

Calculating Saliency Weighting

Pose x_t



$y_t \sim \mathbf{Cat}(\mathcal{Z}_{x_t})$

2

Entropy

$$\mathbb{H}(x_t) = -\sum_{c=1}^{C}\mathcal{Z}_{x_t}^{c}\,\mathbf{log}\mathcal{Z}_{x_t}^{c}$$

3

Normalization

$$\mathcal{S}(x_t) = \frac{\mathbb{H}_{\mathbf{max}}-\mathbb{H}(x_t)}{\mathbb{H}_{\mathbf{max}}} \in [0,1]$$

4

Sailency Weighting

$$p(x_t|\mathbf{D_A}) = \frac{\mathcal{S}(x_t)}{\sum_{t=1}^{T}\mathcal{S}(x_t)}$$

Saliency-Pose-Attention Covariance(SPA-Cov)

$$\Sigma_{\mathbf{D_A}}^{\mathcal{S}} = \int_{x_t} p(x_t|\mathbf{D_A})\mathcal{F}(x_t-\mu_{\mathcal{S}})dx_t$$

SINGULARITY ISSUE

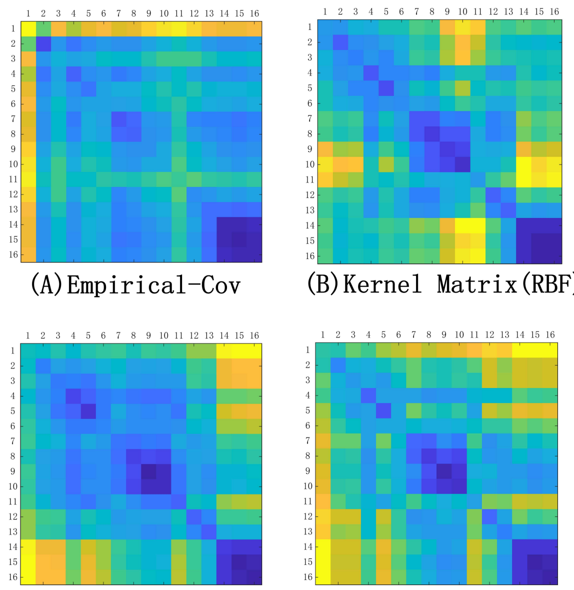
Posterior Distribution

$$p(\Sigma_{\mathbf{D_A}}|\mathbf{D_A},\mu_A) \propto \mathbf{Prior}\times\mathbf{Likelihood} = \mathcal{IW}(\Sigma_{\mathbf{D_A}}|S_T^{-1},\gamma_T)$$

✓ Prior Regularization	× Scaled Regularization
$\hat{\Sigma}_{\mathbf{D_A}} = \pi\Sigma_0+(1-\pi)\Sigma_{\mathbf{D_A}}$	$\hat{\Sigma}_{\mathbf{D_A}} = \Sigma_{\mathbf{D_A}}+\sigma\mathbf{I}$

VISUALIZATION COMPARISON

The average distances between 16 action categories in MSR-DailyActivity3D(S2) database.



(A)Empirical-Cov (B)Kernel Matrix(RBF) (C)Kernelized-Cov (D)SPA-Cov

RESULTS

Recognition accuracy (%) on MSR-Action3D(S1), MSR-DailyActivity3D(S2), MSRC-Kinect12(S3). The sign of ' -/-' means that no results are reported in the paper.

Methods/Databases	S1	S2	S3
Pose Set	90.0	-/-	-/-
Moving Pose	91.7	73.8	-/-
Empirical-Cov	74.0	85.0	89.2
Infinite-Cov	80.4	75.0	89.2
Temporal-Cov	90.5	93.5	91.7
Kernel-RBF	96.2	96.3	92.3
Kernel-POL	**96.9**	96.9	90.5
Kernelized-Cov	96.2	96.3	**95.0**
Proposed Posterior-Cov	96.2	96.3	89.5
Proposed SPA-Cov	96.2	**97.5**	90.5
Proposed Posterior-SPA-Cov	**96.9**	**97.5**	91.5

FUTURE WORK

We will explore more efficient methods to mine the saliency poses and extend it to much more cases, e.g. non-linear kernels, spatio-temporal representations, etc.