Label Consistent Fisher Vectors for Supervised Feature Aggregation

Quan Wang Xin Shen Meng Wang Kim L. Boyer

Rensselaer Polytechnic Institute, Troy, NY 12180, USA

Abstract

Contribution:

- We add supervised information into Fisher vectors to improve classification performance

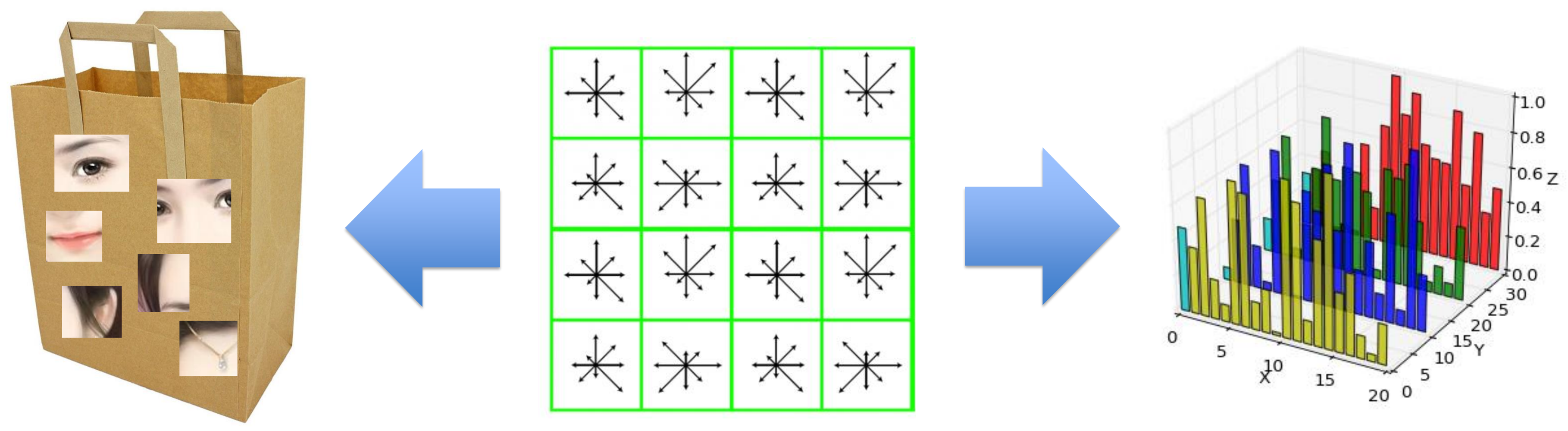
Basic idea:

- In the training process, we add a discriminative *label comparison matrix* to the Fisher kernel, and solve for a transformation matrix to be applied on Fisher vectors

Applications:

- Content-based image retrieval, scene classification, etc.

Feature Aggregation



Bag-of-Words (BoW): only 0th order statistics

Low-level features, e.g. SIFT

Fisher Vector (FV): also 1st order and 2nd order statistics

Fisher Vector

We have extracted T local features such as SIFT: $X = \{x_t\}_{t=1}^{T}$

Assume X follows a probabilistic generative process $p_\theta(X)$

θ is the parameter set (e.g. μ and σ for GMM), and the log-likelihood of the gradient describes the contribution of each parameter: $g_\theta(X) = \frac{1}{T}\nabla_\theta \log p_\theta(X) = \frac{1}{T}\sum_{t=1}^{T}\nabla_\theta \log p_\theta(x_t)$

The Fisher information matrix is $F_\theta = E_X[g_\theta(X) g_\theta(X)^{\mathrm{T}}]$

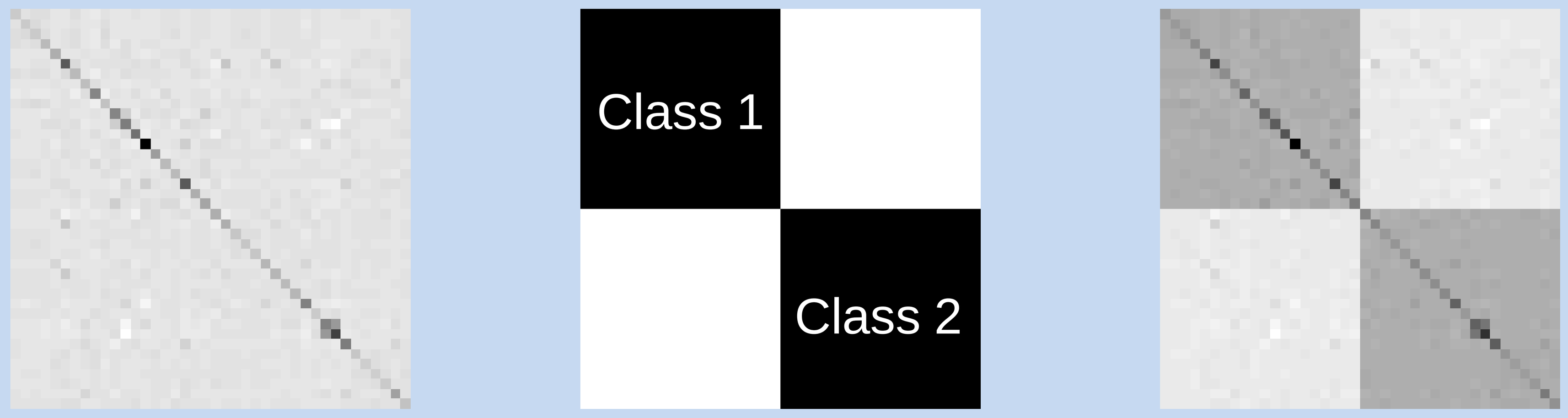
The Fisher kernel: $K(X_1, X_2) = g_\theta(X_1)^{\mathrm{T}} F_\theta^{-1} g_\theta(X_2)$ → theoretically as good as MAP based on the model

Fisher vector $\mathcal{G}_\theta(X) = F_\theta^{-\frac{1}{2}} g_\theta(X)$, then $K(X_1, X_2) = \mathcal{G}_\theta(X_1)^{\mathrm{T}} \mathcal{G}_\theta(X_2)$ → linear classifier (fast) on FV = kernel classifier on FK

Fisher kernels and Fisher vectors are obtained without using any supervised information in the training set

Label Consistent Fisher Vector

We define a label comparison matrix $\mathbf{C} = [C_{i,j}]$, where $C_{i,j} = \delta(c_i = c_j)$, c_i and c_j are class labels



Fisher kernel Label comparison matrix New kernel

$\widetilde{K}(X_i, X_j) = K(X_i, X_j) + \alpha C_{i,j}$ is a better kernel, where $\alpha > 0$

Let $\mathbf{G} = [\mathcal{G}_\theta(X_1), \dots, \mathcal{G}_\theta(X_N)]$, if we can find a transformation matrix $\mathbf{M}$ such that $(\mathbf{MG})^{\mathrm{T}}(\mathbf{MG}) = \mathbf{G}^{\mathrm{T}}\mathbf{G} + \alpha\mathbf{C}$, then we can apply $\mathbf{M}$ on the Fisher vector of a testing image, whose class label is unknown

Two approaches:

- LCFV1: $\mathbf{M}$ is a square matrix, $\mathbf{W} = \mathbf{M}^{\mathrm{T}}\mathbf{M} = \mathbf{I} + \mathbf{B}$
- LCFV2: $\mathbf{M} = \begin{bmatrix}\mathbf{I}\\ \mathbf{E}\end{bmatrix}$ is not square, thus adding new features

Overdetermined: minimize the Frobenius norm of the error

Underdetermined: minimize the Frobenius norm or rank of $\mathbf{B}$ or $\mathbf{E}$

Experiments

We compare LCFV1 and LCFV2 with standard Fisher vectors on three datasets: 15-scene, Graz-02, Corel

α is tuned by cross-validation on training set. We report average classification accuracy

Dataset	PCA dim	# of GMM	FV (%)	LCFV1 (%)	LCFV2 (%)
15-scene	16	16	65.71	66.34	**66.90**
	16	64	70.27	70.46	**70.52**
	16	128	71.76	71.85	**71.89**
	64	16	66.93	67.26	**67.37**
	64	64	72.03	**72.40**	72.39
	64	128	72.86	73.22	**73.22**
Graz-02	64	16	62.61	62.96	**63.08**
	64	64	67.11	67.38	**67.43**
	64	128	68.10	68.28	**68.43**
Corel	64	16	56.27	56.58	**56.70**
	64	64	60.53	60.77	**60.87**
	64	128	61.68	62.03	**62.10**