A Similarity Measure Between Unordered Vector Sets with Application to Image Categorization

Yan Liu and Florent Perronnin, XRCE, CVPR 2008

Problem Statement

- to compute the similarity between two unordered variable-size vector sets

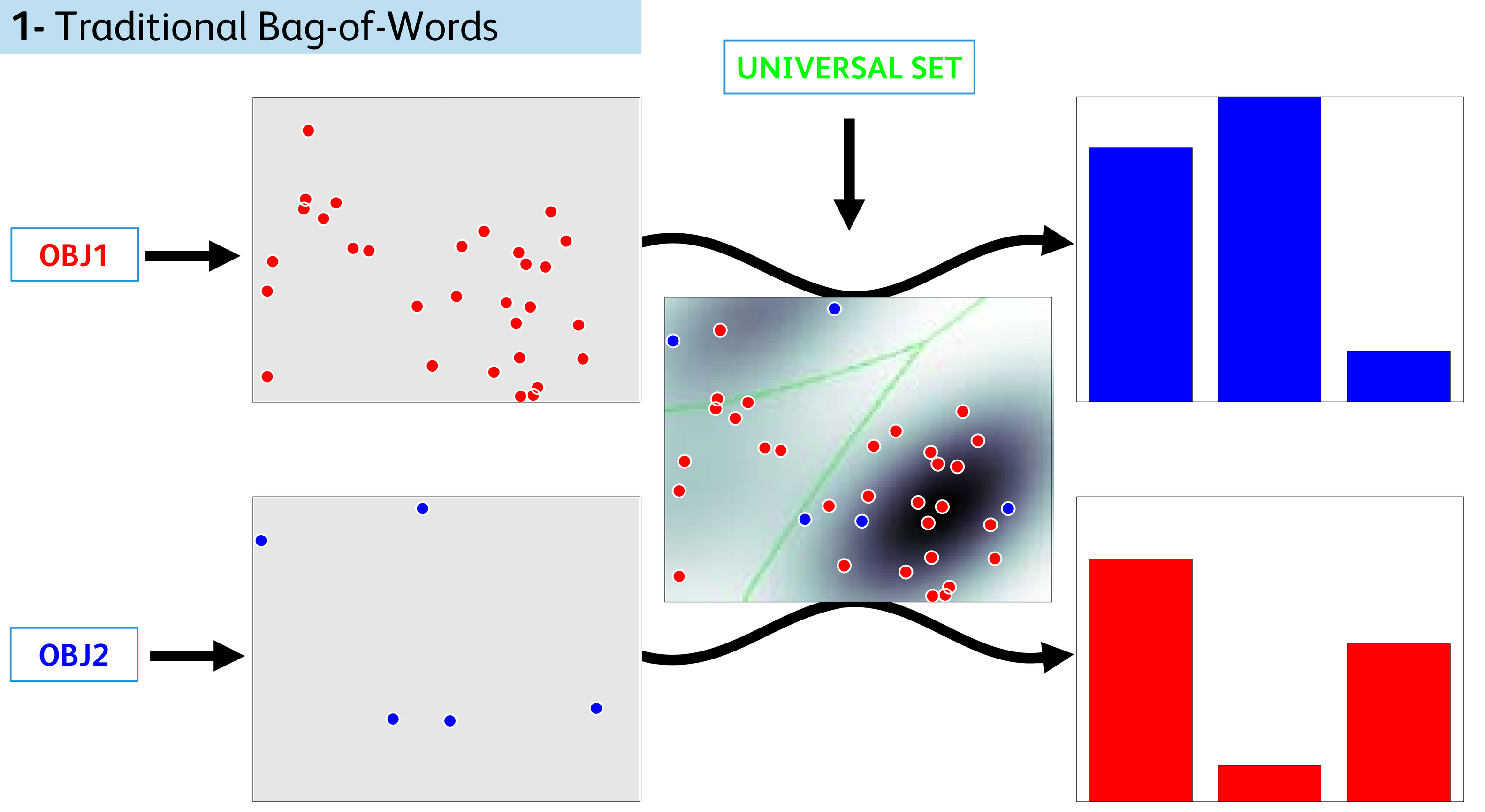
Approach:

- to **model each vector set with its GMM** and then compute a probabilistic measure of similarity
- to **adapt the GMM** from a common *universal* GMM using the maximum a posteriori (MAP) criterion
- to **derive similarity measures** between GMMs, taking advantage of their adapted nature, i.e. 1:1 correspondence between Gaussian components.
  - Kullback-Leibler Kernel (KLK):
  - Probability Product Kernel (PPK):
- use a **kernel classifier** to take advantage of the proposed similarity measure for classification

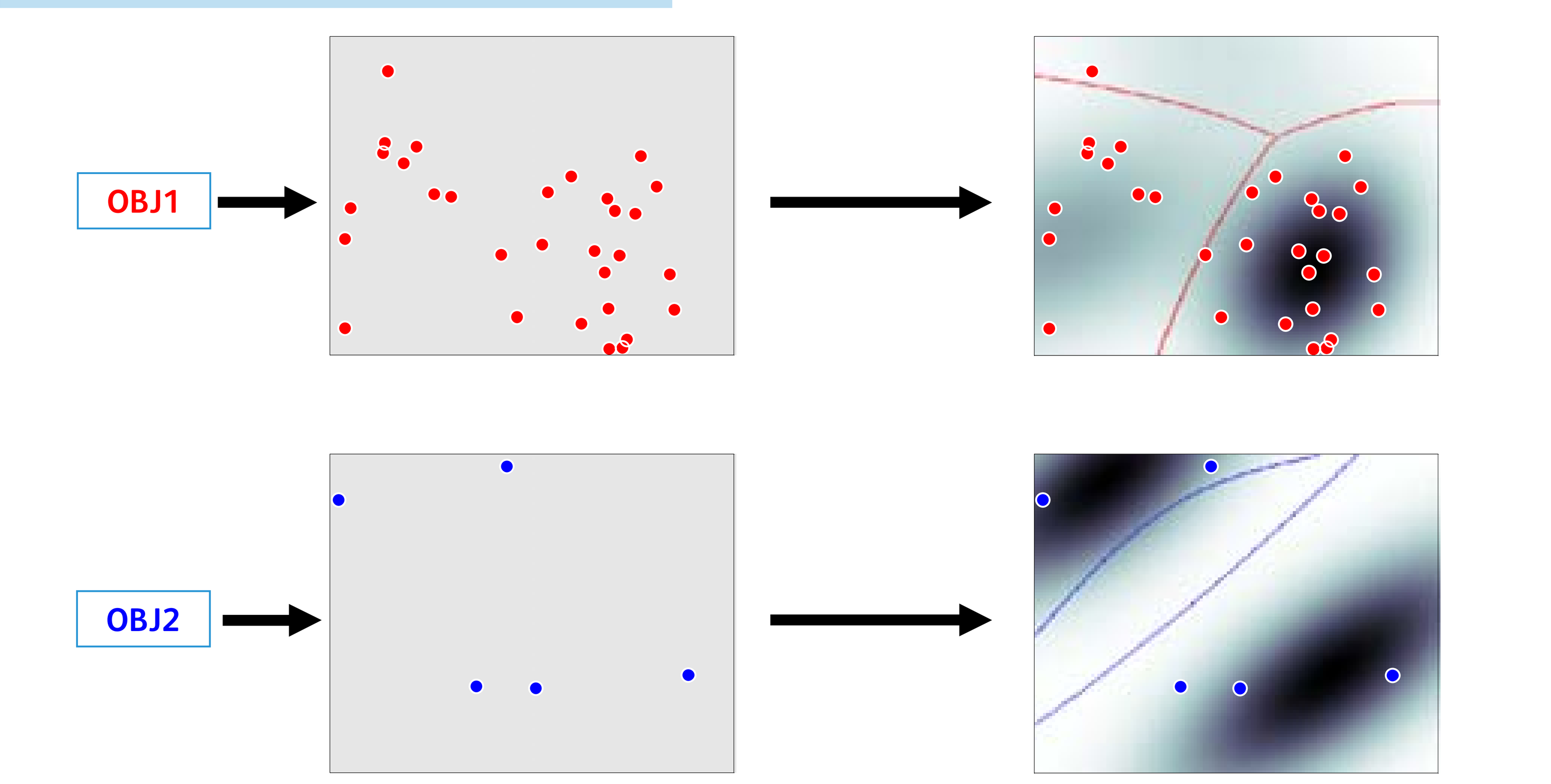
Evaluation

- classification performance on PASCAL VOC 2006/2007 using Kernel Logistic Regression (KLR)
  - comparison of MAP estimation over MLE
  - impact of similarity measure choice
- robustness to variations in common *universal* GMM, through cross-database experiments
- computational cost analysis

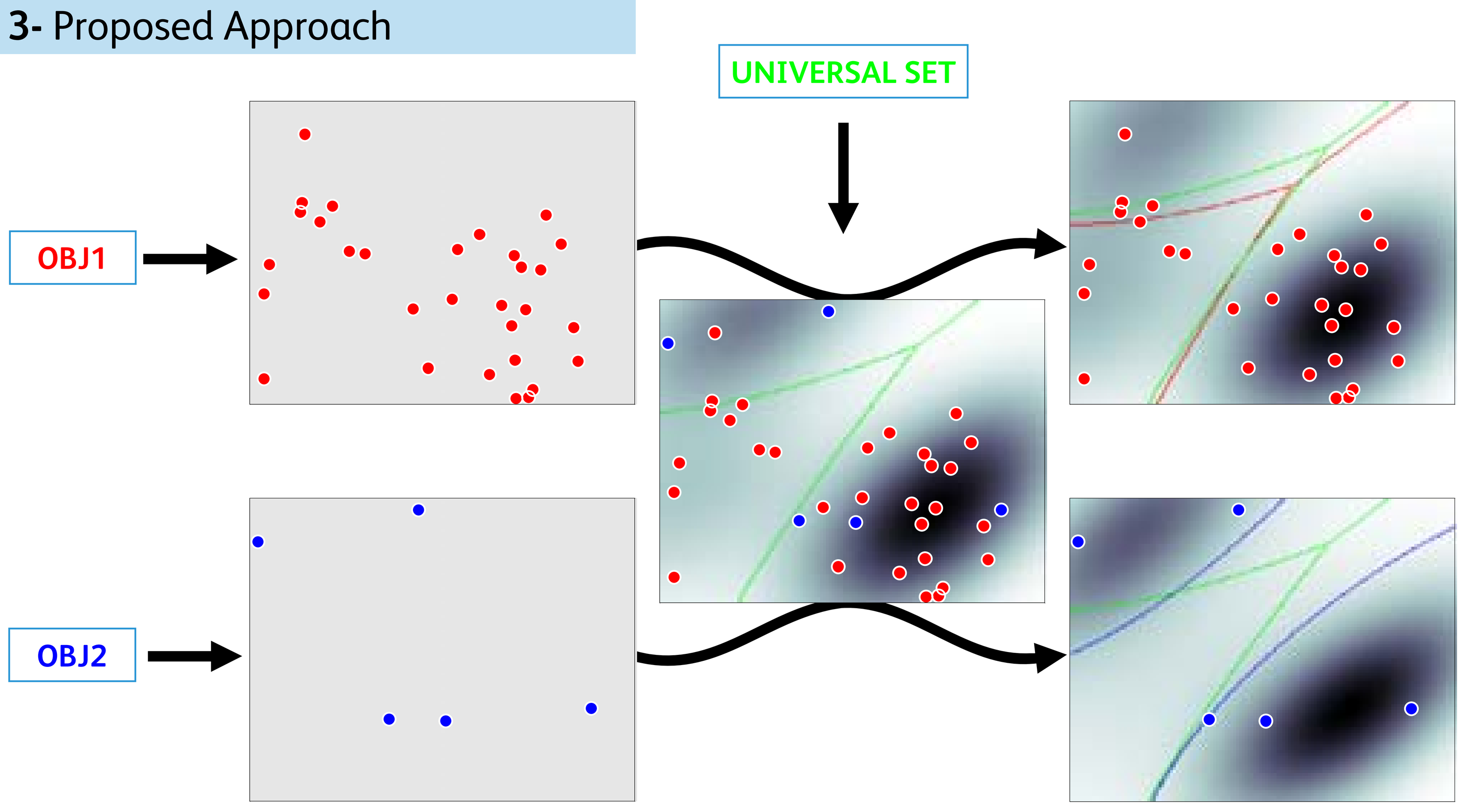
1- Traditional Bag-of-Words



2- Vector Sets as GMMs



3- Proposed Approach



		BOW	Img-GMM	Proposed
Model Estimation	robust to low number of samples	yes	no	yes
	low computational cost	yes	no	yes
Similarity Measure	high precision	no	no	yes
	low computational cost	yes	no	yes
	class-independent representation	no	yes	yes
Supervised Classification	high precision	yes	no	yes
	low computational cost	yes	no	yes

MAP_OTO

M-step for universal model: MLE

$$\hat{w}_i^u = \frac{1}{T}\sum_{t=1}^{T}\gamma_i(x_t),$$

$$\hat{\mu}_i^u = \frac{\sum_{t=1}^{T}\gamma_i(x_t)x_t}{\sum_{t=1}^{T}\gamma_i(x_t)},$$

$$\hat{\Sigma}_i^u = \frac{\sum_{t=1}^{T}\gamma_i(x_t)x_t x_t'}{\sum_{t=1}^{T}\gamma_i(x_t)} - \hat{\mu}_i^u\hat{\mu}_i^{u\prime}.$$

occupancy probability

M-step for adapted image model: MAP

$$\hat{w}_i^a = \frac{\sum_{t=1}^{T}\gamma_i(x_t)+\tau}{T+N\times\tau},$$

$$\hat{\mu}_i^a = \frac{\sum_{t=1}^{T}\gamma_i(x_t)x_t+\tau\mu_i^u}{\sum_{t=1}^{T}\gamma_i(x_t)+\tau},$$

$$\hat{\Sigma}_i^a = \frac{\sum_{t=1}^{T}\gamma_i(x_t)x_t x_t' + \tau[\Sigma_i^u+\mu_i^u\mu_i^{u\prime}]}{\sum_{t=1}^{T}\gamma_i(x_t)+\tau} - \hat{\mu}_i^a\hat{\mu}_i^{a\prime}.$$

relevance factor

Cost: Similar computational cost but considerably lower number of iterations.

One-To-One Gaussian mapping (MAP_OTO): There is a one-to-one correspondence between the i-th Gaussian of two GMMs adapted from the same common universal GMM [RQD00].

PPK_MAP_OTO

Probabilistic Product Kernels: PPK

$$K_{ppk}^{\rho}(p,q) = \int_{x\in\Omega} p(x)^{\rho} q(x)^{\rho} dx.$$

ρ=1, Expected Likelihood Kernel
ρ =0.5, Bhattacharyya Kernel

Closed form solution between two Gaussians

$$K_{ppk}^{\rho}(p,q) = (2\pi)^{(1-2\rho)D/2}\,|\Sigma|^{1/2}|\Sigma_p|^{-\rho/2}|\Sigma_q|^{-\rho/2}\,\exp\left(-\frac{\rho}{2}\mu_p^{\top}\Sigma_p^{-1}\mu_p-\frac{\rho}{2}\mu_q^{\top}\Sigma_q^{-1}\mu_q+\frac{1}{2}\mu^{\top}\Sigma\mu\right).$$

Existing approximation [JK03]

$$K_{ppk}^{\rho}(p,q) \approx \sum_{i=1}^{N}\sum_{j=1}^{M}\alpha_i\beta_j K_{ppk}^{\rho}(p_i,q_j)$$

quadratic

mixture weights

Proposed similarity measure when both GMMs are adapted from the same universal model

$$K_{ppk}^{\rho}(p,q) \approx \sum_{i=1}^{N}\alpha_i\beta_i K_{ppk}^{\rho}(p_i,q_i)$$

linear

KLK_MAP_OTO

Kullback-Leibler Kernel: KLK

$$KL(p||q) = \int_{x\in\Omega} p(x)\log\frac{p(x)}{q(x)}dx$$

Closed form solution between two Gaussians

$$KL(p||q) = \frac{1}{2}\left[\log\frac{|\Sigma_q|}{|\Sigma_p|}+\mathrm{Tr}(\Sigma_q^{-1}\Sigma_p) + (\mu_p-\mu_q)^T\Sigma_q^{-1}(\mu_p-\mu_q)-D\right].$$

Existing approximation [GGG03]

$$KL(p||q) \approx \sum_{i=1}^{N}\alpha_i\left(KL(p_i||q_{\pi(i)})+\log\frac{\alpha_i}{\beta_{\pi(i)}}\right)$$

$$\pi(i) = \arg\min_j\left(KL(p_i||q_j)-\log\beta_j\right)$$

Our approximation

$$KL(p||q) \approx \sum_{i=1}^{N}\alpha_i\left(KL(p_i||q_i)+\log\frac{\alpha_i}{\beta_i}\right)$$

Proposed similarity measure when both GMMs are adapted from the same universal model

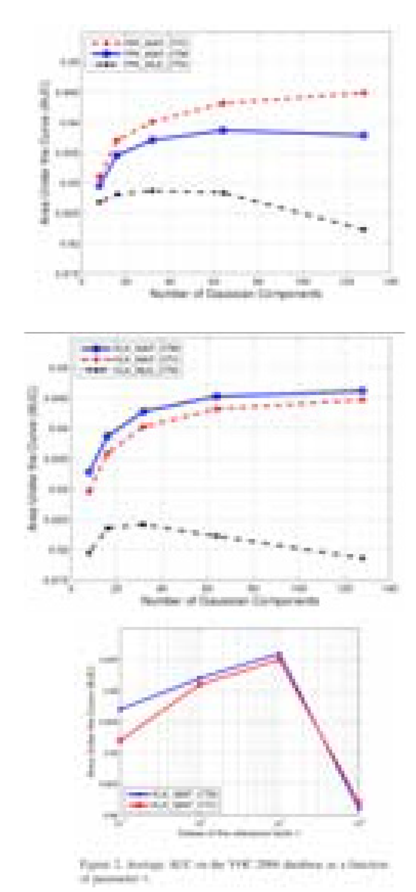
$$SKL(p,q) = KL(p||q)+KL(q||p)$$

$$K_{klk}(p,q) = \exp(-\gamma SKL(p,q)).$$

must ensure positive definite kernel matrix

EXPERIMENTAL RESULTS

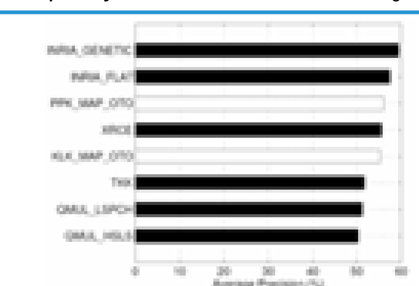
PASCAL VOC 2006



SETUP

- **Low-level feature** vectors extracted on grids at multiple scales (local gradient histograms and RGB stats)
- Iterative splitting and retraining for **universal model GMM**
- **Classification:** Sparse Logistic Regression (SLR) trained in a one-vs-all manner (very similar results with SVM), late fusion of feature-type scores
- **PASCAL VOC2006 dataset**, 10 classes, 2618 training images, 2610 test images, Area Under Curve (AUC) for measuring performance
- **PASCAL VOC2007 dataset**, 20 classes, 5011 training images, 4952 test images, Average Precision (AP) for measuring performance
- **Proprietary dataset**, 120000 unannotated images randomly captured from a photofinishing workflow

PASCAL VOC 2007



COMPUTATIONAL COST

On 2.4 GHz Opteron™ machine

- MLE 850ms per image, MAP 30ms per image
- Classification cost (seconds per image):
  - one to many: 140s for PPK, 30s for KLK
  - one to one: 1.3s for PPK, 0.4 for KLK
  - Montecarlo sampling (PPK/KLK): ~240s

CONCLUSIONS

- MAP estimation outperforms MLE. The relevance factor (τ) can affect performance
- Our one to one approximation of the similarity measures
  - greatly improves the cost for both PPK and KLK (now linear in number of Gaussians)
  - actually improves the one to many approximation in the case of PPK
  - perform similarly to one another (PPK vs. KLK)
- Excellent classification performance on PASCAL VOC 2006/2007 datasets
- No change in performance when external dataset is used for training universal model

Future work

- Application on larger scale database.
- Evaluate other adaptation techniques, e.g. Maximum Likelihood Linear Regression (MLLR), Cluster Adaptive Training (CAT) or "eigenvoices" have proven results in speech recognition.
- Evaluate further approximations of PPK or other probabilistic similarity measures

References

- [CD04] G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray, *Visual categorization with bags of keypoints*, ECCV workshop, 2004.
- [GGG03] J. Goldberger, S. Gordon and H. Greenspan, *An Efficient Image Similarity Measure based on Approximations of KL-Divergence Between Two Gaussian Mixtures*, ICCV, 2003.
- [JK03] T. Jebara and R. Kondor, *Bhattacharyya and expected likelihood kernels*, Proc. of the 16th annual conference on learning theory and seventh kernel workshop, 2003.
- [Lo04] D. Lowe, *Distinctive image features from scale-invariant key-points,* IJCV, 2004.
- [MVH03] P. Moreno, N. Vasconcelos, and P. Ho, *A Kullback-Leibler divergence based kernel for SVM classification in multimedia application*, NIPS, 2003.
- [RQD00] D. Reynolds, T. Quatieri, and R. Dunn, *Speaker verification using adapted Gaussian mixture models*, Digital Signal Processing, 2000.
- [SZ03] J. Sivic and A. Zisserman, *Video Google: a text retrieval approach to object matching in videos*, ICCV, 2003.
- [Va04] N. Vasconcelos, *On the efficient evaluation of probabilistic similarity functions for image retrieval*, IEEE Trans. on PAMI, 2004, vol. 50, no. 7, pp. 1482-1496.
- [VHM04] N. Vasconcelos, P. Ho and P. Moreno, *The Kullback-Leibler kernel as a framework for discriminant and localized representations for visual recognition*, ECCV, 2004.