Object Localization by Subspace Clustering of Local Descriptors

C. Bouveyron[1], J. Kannala[2], C. Schmid[1], and S. Girard[1]

[1] INRIA Rhône-Alpes, 655 avenue de l'Europe, 38300 Saint-Ismier, France
[2] Machine Vision Group, dept. of Electrical and Information Engineering,
90014 University of Oulu, Finland

Abstract. This paper presents a probabilistic approach for object localization which combines subspace clustering with the selection of discriminative clusters. Clustering is often a key step in object recognition and is penalized by the high dimensionality of the descriptors. Indeed, local descriptors, such as SIFT, which have shown excellent results in recognition, are high-dimensional and live in different low-dimensional subspaces. We therefore use a subspace clustering method called High-Dimensional Data Clustering (HDDC) which overcomes the curse of dimensionality. Furthermore, in many cases only a few of the clusters are useful to discriminate the object. We, thus, evaluate the discriminative capacity of clusters and use it to compute the probability that a local descriptor belongs to the object. Experimental results demonstrate the effectiveness of our probabilistic approach for object localization and show that subspace clustering gives better results compared to standard clustering methods. Furthermore, our approach outperforms existing results for the Pascal 2005 dataset.

1 Introduction

Object localization is one of the most challenging problems in computer vision. Earlier approaches characterize the objects by their global appearance and are not robust to occlusion, clutter and geometric transformations. To avoid these problems, recent methods use local image descriptors. Many of these approaches form clusters of local descriptors as an initial step; in most cases clustering is achieved with k-means or EM-based clustering methods. Agarwal and Roth [1] determine the spatial relations between clusters and use a Sparse Network of Windows classifier. Dorko and Schmid [2] select discriminant clusters based on the likelihood ratio and use the most discriminative ones for recognition. Leibe and Schiele [3] learn the spatial distribution of the clusters and use voting for recognition. Bag-of-keypoint methods [4,5] represent an image by a histogram of cluster labels and learn a Support Vector Machine classifier. Sivic *et al.* [6] combine a bag-of-keypoint representation with probabilistic latent semantic analysis to discover topics in an unlabeled dataset. Opelt *et al.* [7] use AdaBoost to select the most discriminant features.

However, visual descriptors used in object recognition are often high-dimensional and this penalizes classification methods and consequently recognition. Indeed, clustering methods based on the Gaussian Mixture Model (GMM) [8] show a disappointing behavior when the size of the training dataset is too small compared to the number

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 457–467, 2006.

of parameters to estimate. To avoid overfitting, it is therefore necessary to find a balance between the number of parameters to estimate and the generality of the model. Many methods use global dimensionality reduction and then apply a standard clustering method. Dimension reduction techniques are either based on *feature extraction* or *feature selection*. Feature extraction builds new variables which carry a large part of the global information. The most popular method is Principal Component Analysis (PCA) [9], a linear technique. Recently, many non-linear methods have been proposed, such as Kernel PCA [10]. Feature selection, on the other hand, finds an appropriate subset of the original variables to represent the data [11]. Global dimension reduction is often advantageous in terms of performance, but loses information which could be discriminant, *i.e.*, clusters often lie in different subspaces of the original feature space and a global approach cannot capture this. It is also possible to use a parsimonious model [12] which reduces the number of parameters to estimate by fixing some parameters to be common within or between classes. These methods do not solve the problem of high dimensionality because clusters usually lie in different subspaces and many dimensions are irrelevant. Recent methods determine the subspaces for each cluster. Many subspace clustering methods use heuristic search techniques to find the subspaces. They are usually based on grid search methods and find dense clusterable subspaces [13]. The approach "mixture of Probabilistic Principal Component Analyzers" [14] proposes a latent variable model and derives an EM based method to cluster high-dimensional data. A similar model is used in [15] in the supervised framework. The model of these methods can be viewed as a mixture of constrained Gaussian densities with class-specific subspaces. An unified approach for subspace clustering in the Gaussian mixture model framework was proposed in [16]. This method, called High Dimensional Data Clustering (HDDC), includes the previous approaches and involves additional regularizations as in parsimonious models.

In this paper, we propose a probabilistic framework for object localization combining subspace clustering with the selection of the discriminative clusters. The first step of our approach is to cluster the local descriptors using HDDC [16] which is not penalized by the high-dimensionality of the descriptors. Since only a few of the learned clusters are useful to discriminate the object, we then determine the discriminative score of each cluster with positive and negative examples of the category. This score is based on a maximum likelihood formulation. By combining this information with the posterior probabilities of the clusters, we finally compute the object probability for each visual descriptor. These probabilities are then used for object localization, *i.e.*, localization assumes that points with higher probabilities are more likely to belong to the object. We evaluate our approach on two recently proposed object datasets [7,17]. We first compare HDDC to standard clustering methods within our probabilistic recognition framework. Experiments show that results with HDDC are consistently better than with other clustering methods. We then compare our probabilistic approach to the state of the art results and show that it outperforms existing results for object localization.

This paper is organized as follows. Section 2 presents the EM-based clustering method HDDC, *i.e.*, the estimation of the parameters and of the intrinsic dimensions of the subspaces. In Section 3, we describe the probabilistic object localization framework.

Experimental results for our approach are presented in Section 4. We conclude the paper in Section 5.

2 High-Dimensional Data Clustering

This section presents the clustering method HDDC [16]. Clustering divides a given dataset $\{x_1, ..., x_n\}$ of n data points into k homogeneous groups. Popular clustering techniques use Gaussian Mixture Models (GMM). The data $\{x_1, ..., x_n\} \in \mathbb{R}^p$ are then modeled with the density $f(x, \theta) = \sum_{i=1}^{k} \pi_i \phi(x, \theta_i)$, where ϕ is a multi-variate normal density with parameter $\theta_i = \{\mu_i, \Sigma_i\}$ and π_i are mixing proportions. This model estimates the full covariance matrices and therefore the number of parameters is very large in high dimensions. However, due to the *empty space* phenomenon we can assume that high-dimensional data live in subspaces with a dimensionality lower than the dimensionality of the original space. We therefore propose to work in low-dimensional class-specific subspaces in order to adapt classification to high-dimensional data and to limit the number of parameters to estimate. Here, we will present the parameterization of GMM designed for high-dimensional data and then detail the EM-based technique HDDC.

2.1 Gaussian Mixture Models for High-Dimensional Data

We assume that class conditional densities are Gaussian $\mathcal{N}(\mu_i, \Sigma_i)$ with means μ_i and covariance matrices Σ_i, $i = 1, ..., k$. Let Q_i be the orthogonal matrix of eigenvectors of Σ_i, then $\Delta_i = Q_i^t \, \Sigma_i \, Q_i$ is a diagonal matrix containing the eigenvalues of Σ_i. We further assume that Δ_i is divided into two blocks:

$$\Delta_i = \left(\begin{array}{cc} \boxed{\begin{array}{ccc} a_{i1} & & 0 \\ & \ddots & \\ 0 & & a_{id_i} \end{array}} & \mathbf{0} \\ \mathbf{0} & \boxed{\begin{array}{ccc} b_i & & 0 \\ & \ddots & \\ 0 & & b_i \end{array}} \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{array}{c} a \\ \ddots \\ a \end{array}}\right\} \; d_i \\ \left.\vphantom{\begin{array}{c} b \\ \ddots \\ b \end{array}}\right\} \; (p - d_i) \end{array}$$

where $a_{ij} > b_i, \forall j = 1, ..., d_i$. The class specific subspace $\mathbb{E}_i$ is generated by the d_i first eigenvectors corresponding to the eigenvalues a_{ij} with $\mu_i \in \mathbb{E}_i$. Outside this subspace, the variance is modeled by a single parameter b_i. Finally, let $P_i(x) = \tilde{Q}_i \tilde{Q}_i^{\,t}(x - \mu_i) + \mu_i$ be the projection of x on $\mathbb{E}_i$, where $\tilde{Q}_i$ is made of the d_i first columns of Q_i supplemented by zeros. Figure 1 summarizes these notations.

The mixture model presented above will be in the following referred to by $[a_{ij} b_i \, Q_i d_i]$. By fixing some parameters to be common within or between classes, we obtain particular models which correspond to different regularizations. For example, if we fix the first d_i eigenvalues to be common within each class, we obtain the more restricted model $[a_i b_i Q_i d_i]$. This model is in many cases more robust, *i.e.*, the assumption that the matrix Δ_i contains only two eigenvalues a_i and b_i seems to be an efficient way to regularize the estimation of Δ_i. In this paper, we focus on the models $[a_{ij} b_i Q_i d_i]$, $[a_{ij} b Q_i d_i]$, $[a_i b_i Q_i d_i]$, $[a_i b Q_i d_i]$ and $[a b Q_i d_i]$.

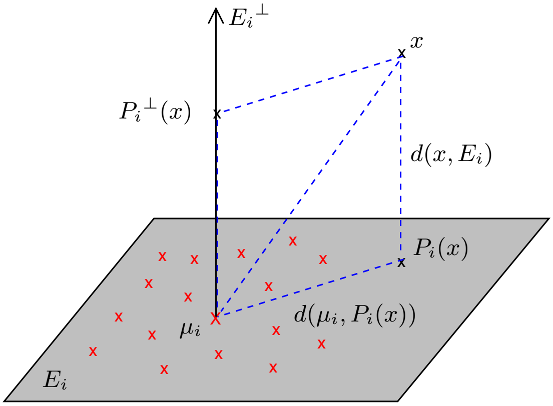


Fig. 1. The specific subspace $\mathbb{E}_i$ of the ith mixture component

2.2 EM Estimation of the Model Parameters

The parameters of a GMM are usually estimated by the EM algorithm which repeats iteratively expectation (E) and maximization (M) steps. In this section, we present the EM estimation of the parameters for the subspace GMM.

The **E-step** computes, at iteration q, for each component $i = 1, ..., k$ and for each data point $j = 1, ..., n$, the conditional probability $t_{ij}^{(q)} = P(x_j \in C_i^{(q-1)}|x_j)$. Using the Bayes formula and the parameterization of the model $[a_{ij}b_iQ_id_i]$, the probability $t_{ij}^{(q)}$ can be expressed as follows (the proof of the following result is available in [16]):

$$t_{ij}^{(q)} = \frac{\pi_i^{(q-1)}\phi(x_j, \theta_i^{(q-1)})}{\sum_{\ell=1}^{k}\pi_\ell^{(q-1)}\phi(x_j, \theta_\ell^{(q-1)})} = 1/\sum_{\ell=1}^{k}\exp\left(\frac{1}{2}(K_i(x_j) - K_\ell(x_j))\right),$$

where $K_i(x) = -2\log(\pi_i\phi(x, \theta_i))$ is called the cost function and is defined by:

$$K_i(x) = \|\mu_i - P_i(x)\|_{\mathcal{A}_i}^2 + \frac{1}{b_i}\|x - P_i(x)\|^2 + \sum_{j=1}^{d_i}\log(a_{ij}) + (p - d_i)\log(b_i) - 2\log(\pi_i),$$

where $\|.\|_{\mathcal{A}_i}$ is a norm on $\mathbb{E}_i$ such that $\|x\|_{\mathcal{A}_i}^2 = x^t\mathcal{A}_ix$ with $\mathcal{A}_i = \tilde{Q}_i\Delta_i^{-1}\tilde{Q}_i^{\,t}$. We can observe that $K_i(x)$ is mainly based on two distances: the distance between the projection of x on $\mathbb{E}_i$ and the mean of the class and the distance between the observation and the subspace $\mathbb{E}_i$. This cost function favours the assignment of a new observation to the class for which it is close to the subspace and for which its projection on the class subspace is close to the mean of the class. The variance terms a_{ij} and b_i balance the importance of both distances. For example, if the data are very noisy, *i.e.*, b_i is large, it is natural to weight the distance $\|x - P_i(x)\|^2$ by $1/b_i$ in order to take into account the large variance in $\mathbb{E}_i^{\perp}$.

The **M-step** maximizes at iteration q the conditional likelihood and uses the following update formulas. The proportions, the means and the covariance matrices of the mixture are classically estimated by:

$$\hat{\pi}_i^{(q)} = \frac{n_i^{(q)}}{n}, \; \hat{\mu}_i^{(q)} = \frac{\sum_{j=1}^n t_{ij}^{(q)} x_j}{n_i^{(q)}}, \; \hat{\Sigma}_i^{(q)} = \frac{1}{n_i^{(q)}} \sum_{j=1}^n t_{ij}^{(q)} (x_j - \hat{\mu}_i^{(q)})(x_j - \hat{\mu}_i^{(q)})^t.$$

where $n_i^{(q)} = \sum_{j=1}^n t_{ij}^{(q)}$. The ML estimators of model parameters are in closed form for the models considered in this paper. Proofs of the following results are given in [16].

– Subspace $\mathbb{E}_i$: the d_i first columns of Q_i are estimated by the eigenvectors associated with the d_i largest eigenvalues λ_{ij} of $\hat{\Sigma}_i$.
– Model $[a_{ij} b_i Q_i d_i]$: the estimator of a_{ij} is $\hat{a}_{ij} = \lambda_{ij}$ and the estimator of b_i is:

$$\hat{b}_i = \frac{1}{(p - d_i)} \left(\operatorname{Tr}(\hat{\Sigma}_i) - \sum_{j=1}^{d_i} \lambda_{ij} \right). \tag{1}$$

– Model $[a_{ij} b Q_i d_i]$: the estimator of a_{ij} is $\hat{a}_{ij} = \lambda_{ij}$ and the estimator of b is:

$$\hat{b} = \frac{1}{(p - \xi)} \left(\operatorname{Tr}(\hat{W}) - \sum_{i=1}^{k} \hat{\pi}_i \sum_{j=1}^{d_i} \lambda_{ij} \right), \tag{2}$$

where $\xi = \sum_{i=1}^k \hat{\pi}_i d_i$ and $\hat{W} = \sum_{i=1}^k \hat{\pi}_i \hat{\Sigma}_i$ is the estimated within-covariance matrix.
– Model $[a_i b_i Q_i d_i]$: the estimator of b_i is given by (1) and the estimator of a_i is:

$$\hat{a}_i = \frac{1}{d_i} \sum_{j=1}^{d_i} \lambda_{ij}. \tag{3}$$

– Model $[a_i b Q_i d_i]$: the estimators of a_i and b are respectively given by (3) and (2).
– Model $[a b Q_i d_i]$: the estimator of b is given by (2) and the estimator of a is:

$$\hat{a} = \frac{1}{\xi} \sum_{i=1}^{k} \hat{\pi}_i \sum_{j=1}^{d_i} \lambda_{ij}. \tag{4}$$

2.3 Intrinsic Dimension Estimation

Within the M step, we also have to estimate the intrinsic dimension of each class-specific subspace. This is a difficult problem with no exact solution. Our approach is based on the eigenvalues of the class conditional covariance matrix Σ_i of the class C_i. The jth eigenvalue of Σ_i corresponds to the fraction of the full variance carried by the jth eigenvector of Σ_i. We estimate the class specific dimension d_i, $i = 1, ..., k$, with the empirical method scree-test of Cattell [18] which analyzes the differences between successive eigenvalues in order to find a break in the scree. The selected dimension is the one for which the subsequent differences are smaller than a threshold. In our experiments the value used for this threshold was 0.2 times the maximum difference. The resulting average value for dimensions d_i was approximately 10 in the experiments presented in Section 4.

3 A Probabilistic Framework for Object Localization

In this section, we present a probabilistic framework for object localization which computes for each local descriptor x_j of an image the probability $P(x_j \in O|x_j)$ that x_j belongs to a given object O. It is then easy to precisely locate the object by considering only the local descriptors with high probabilities $P(x_j \in O|x_j)$. We first extract a set of local invariant descriptors using the Harris-Laplace detector [19] and the SIFT descriptor [20]. The dimension of the obtained SIFT features is 128. An interest point and its corresponding descriptor are in the following referred to by x_j.

3.1 Training

During training we determine the discriminative clusters of local descriptors. We first cluster local features and then identify discriminative clusters. Training can be either supervised or weakly supervised. In the weakly supervised scenario the positive descriptors include descriptors from the background, as only the image is labeled as positive.

Clustering. Descriptors of the training images are organized in k groups using the clustering method HDDC. From a theoretical point of view, the descriptors x_j of an image are realizations of a random variable $X \in \mathbb{R}^p$ with the following density $f(x) = \sum_{i=1}^k \pi_i \phi(x, \theta_i) = \tau f^O(x) + (1-\tau) f^B(x)$, where f^O and f^B are respectively the densities of descriptors of the object and of the background and τ denotes the prior probability $P(O)$. The parameter τ is equal to $\sum_{i=1}^k R_i \pi_i$, where $R_i = P(C_i \in O)$. The density f can thus be rewritten as follows:

$$f(x) = \underbrace{\sum_{i=1}^{k} R_i \pi_i \phi(x, \theta_i)}_{\text{Object}} + \underbrace{\sum_{i=1}^{k} (1 - R_i) \pi_i \phi(x, \theta_i)}_{\text{Background}} .$$

The clustering method HDDC provides the estimators of parameters π_i and θ_i, $\forall i = 1, ..., k$ and it thus remains to estimate parameters R_i, $\forall i = 1, ..., k$.

Identification of discriminative clusters. This step aims to identify discriminative clusters by computing estimators of parameters R_i. Positive descriptors are denoted by P and negative ones by N. The conditional ML estimate of $R = \{R_1, ..., R_k\}$ satisfies:

$$\hat{R} = \underset{R}{\text{argmax}} \left\{ \prod_{x_j \in P} P(x_j \in O|x_j) \prod_{x_j \in N} P(x_j \in B|x_j) \right\} .$$

The expression of the gradient is:

$$\nabla_R = \sum_{x_j \in P} \frac{\Psi_j}{< R, \Psi_j >} - \sum_{x_j \in N} \frac{\Psi_j}{1- < R, \Psi_j >},$$

where $\Psi_j = \{\Psi_{ji}\}_{i=1,...,k}$ and $\Psi_{ji} = P(x_j \in C_i|x_j)$ which are provided by HDDC. The ML estimate of R does not have an explicit formulation and it requires an iterative optimization method to find $\hat{R}$. We observed that the classical gradient method converges towards a solution very close to the least square estimator $\hat{R}_{LS} = (\Psi^t\Psi)^{-1}\Psi^t\Phi$, where $\Phi_j = P(x_j \in O|x_j)$. In our experiments, we use this least square estimator of R in order to reduce computation time. We assume for this estimation that $\forall x_j \in P$, $P(x_j \in O|x_j) = 1$ and $\forall x_j \in N$, $P(x_j \in O|x_j) = 0$. Thus, R_i is a measure for the discriminative capacity of the class C_i for the object O.

3.2 Object Localization

During recognition we compute the probability for each local descriptor of a test image to belong to the object. Using these probabilities, it is then possible to locate the object in a test image, *i.e.*, the descriptors of an image with a high probability to belong to the object give a strong indication for the presence of an object. Using the Bayes formula we obtain the posterior probability of an descriptor x_j to belongs to the object O:

$$P(x_j \in O|x_j) = \sum_{i=1}^{k} R_i P(x_j \in C_i|x_j), \tag{5}$$

where the posterior probability $P(x_j \in C_i|x_j)$ is given by HDDC. The object can then be located in a test image by using the points with the highest probabilities $P(x_j \in O|x_j)$. For comparison with existing methods we determine the bounding box with a very simple technique. We compute the mean and variance of the point coordinates weighted by their posterior probabilities given by (5). The mean is then the center of the box and a default bounding box is scaled by the variance.

4 Experiments and Comparisons

In this section, we first compare HDDC to standard clustering techniques within our probabilistic localization framework on the *Graz* dataset [7]. We then compare our approach to the results on the *Pascal 2005* dataset [17].

4.1 Evaluation of the Clustering Approach

In the following, we compare HDDC to the several standard clustering methods within our probabilistic localization framework: diagonal Gaussian mixture model (Diagonal GMM), spherical Gaussian mixture model (Spherical GMM), and data reduction with PCA combined with a diagonal Gaussian mixture model (PCA + diag. GMM). The diagonal GMM has a covariance matrix defined by $\Sigma_i = \text{diag}(\sigma_{i1}, ..., \sigma_{ip})$ and the spherical GMM is characterized by $\Sigma_i = \sigma_i Id$. In all cases, the parameters are estimated with the EM algorithm. The initialization of the EM estimation was obtained using k-means and was exactly the same for both HDDC and the standard methods. For this evaluation, we use the bicycle category of the *Graz* dataset which is consists of 200

Table 1. Object localization on *Graz*: comparison between HDDC and other methods. Precision is computed on segmented images with on average 10 detections per image (*i.e.*, detections such that $P(x_j \in O|x_j) > 0.9$).

Clustering	HDDC $[**Q_id_i]$					Classical GMM			Result
method	$[a_{ij}b_i]$	$[a_{ij}b]$	$[a_ib_i]$	$[a_ib]$	$[ab]$	PCA+diag	Diag.	Sphe.	of [2]
Precision	0.85	0.83	**0.92**	0.89	0.88	0.63	0.70	0.76	0.62

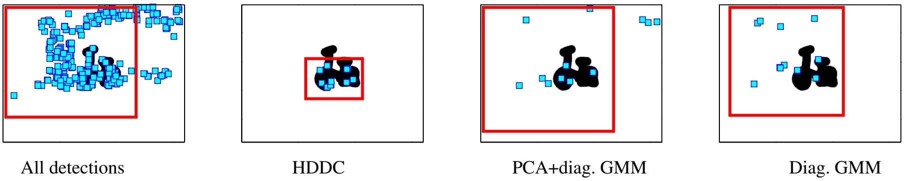

All detections
HDDC
PCA+diag. GMM
Diag. GMM

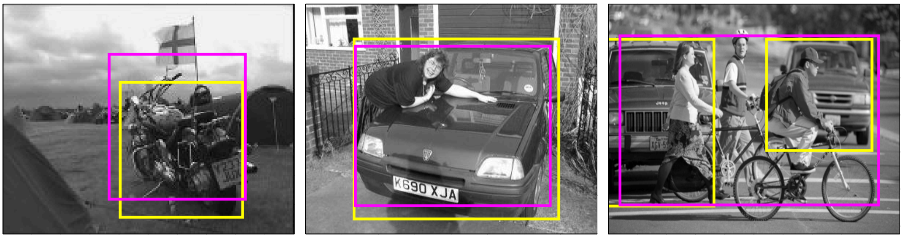
Fig. 2. Object localization on *Graz*: localization results displayed on groundtruth segmentations. We display the points with highest probabilities $P(x_j \in O|x_j)$. The same number of points is displayed for all models (5% of all detections which is equal to 12 detections per image).

training images and 100 test images. We determined 40 clusters with each clustering method in a weakly supervised setting.

The localization performance was evaluated using segmented images [7]. Table 1 summarizes localization performance of the compared methods as well as results presented in [2]. Precision is the number of points within the object region with respect to the total number of selected points. We can observe that the HDDC models give better localization results than the other methods. In particular, the model $[a_ib_iQ_id_i]$ obtains best results, *i.e.*, a precision of 92% when considering points with $P(x_j \in O|x_j) > 0.9$. We also observe that a global dimension reduction with PCA does not improve the results compared to diagonal GMM. This confirms our initial assumption that data of different clusters live in different low-dimensional subspaces and that a global dimension reduction technique is not able to take this into account. Figure 2 shows localization results on segmented test images with the different methods. The left image shows all interest points detected on the test images. The bounding boxes are computed with the displayed points, *i.e.*, the points with the highest probabilities in the case of the three right most images. It appears that our localization method identifies precisely the points belonging to the object and consequently is able to locate small objects in different positions, poses and scales whereas other methods do not give an efficient localization.

Table 2. Average precision (AP) for supervised and weakly-supervised localization on *Pascal test2*. The result in italic is the average result of the best method of the Pascal challenge [17].

Clustering	Supervised					Weakly-supervised				
method	Moto	Bike	People	Car	Aver.	Moto	Bike	People	Car	Aver.
HDDC	0.315	**0.172**	**0.091**	0.155	**0.183**	0.311	**0.161**	**0.046**	0.049	**0.142**
Best of [17]	**0.341**	0.113	0.021	**0.304**	*0.112*	/	/	/	/	/



(a) motorbike (b) car (c) two cars

Fig. 3. Supervised localization on *Pascal test2*: predicted bounding boxes are in magenta and true boxes in yellow

4.2 Comparison to the State of the Art

For this second experiment, we compare our approach to the results on the Pascal visual object class 2005 dataset [17]. It contains four categories: motorbikes, bicycles, people and cars. It is made of 684 training images and two test sets: test1 and test2. We chose to evaluate our method on the set test2, which is the more difficult one and contains 956 images. Since the bounding boxes of the objects are available for all categories we evaluate our method with supervised as well as a weakly supervised training data. In the supervised case only the descriptors located inside the bounding boxes are labeled as positive during training. Here we use 50 clusters for each of the four categories. We use the model $[a_i b_i Q_i d_i]$ for HDDC, since the previous experiment has shown that it is the most efficient model. To compare with the results of Pascal Challenge [17], we use the localization measure "average precision" (AP) which is the arithmetic mean of 11 values on the precision-recall curves computed with ground-truth bounding boxes (see [17] for more details).

The localization results on *Pascal test2* are presented in Table 2 for supervised and weakly supervised training data. In the supervised case, Table 2 shows that our probabilistic recognition approach performs well compared to the results in the Pascal competition. In particular, our approach wins two "competitions" (bicycle and people) and is on average more efficient than the methods of the Pascal challenge. This is despite the fact that our approach detects only one bounding box per image for each category and this reduces the performance when multiple objects are present, as shown in the right part of Figure 3. Notice that our approach has the best overall performance although we do not have any model for the spatial relationships of the local features.

We can also observe that our weakly-supervised localization results are only slightly lower than the ones in the supervised case and on average better than the Pascal results in the supervised case. This means that our approach efficiently identifies discriminative clusters of each object category and this even in the case of weak supervision. There are no corresponding results for the Pascal Challenge, since all competing methods used supervised data. It is promising that the weakly supervised approach obtains good localization results because the manual annotation of training images is time consuming.

5 Conclusion

The main contribution of this paper is the introduction of a probabilistic approach for object localization which combines subspace clustering with the selection of discriminative clusters. This approach has the advantage of using posterior probabilities to weight interest points. We proposed to use the subspace clustering method called HDDC designed for high-dimensional data. Experimental results show that HDDC performs better than other Gaussian models for locating objects in natural images. This is due to the fact that HDDC correctly models the groups in their subspaces and thus forms more homogeneous groups. In addition, our method performs well also in the weakly-supervised framework which is promising. Finally, our approach provides better results than the state of the art methods and that using only one type of detector and descriptor (Harris-Laplace+Sift). We believe that the results could be further improved using a combination of descriptors as in [2,5]. Also, the localization results presented here are based on a very simple spatial model which can be easily improved to further increase the performance of our approach.

References

1. Agarwal, S., Roth, D.: Learning a sparse representation for object detection. In: 7th European Conference on Computer Vision. Volume 4. (2002) 113–130
2. Dorko, G., Schmid, C.: Object class recognition using discriminative local features. Technical Report 5497, INRIA (2004)
3. Leibe, B., Schiele, B.: Interleaved object categorization and segmentation. In: British Machine Vision Conference, Norwich, England (2003)
4. Willamowski, J., Arregui, D., Csurka, G., Dance, C., Fan, L.: Coategorizing nine visual classes using local appareance descriptors. In: International Workshop on Learning for Adaptable Visual Systems, Cambridge, UK (2004)
5. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories. Technical report, INRIA (2005)
6. Sivic, J., Russell, B., Efros, A., Zisserman, A., Freeman, W.: Discovering objects and their location in images. In: International Conference on Computer Vision. (2005)
7. Opelt, A., Fussenegger, M., Pinz, A., Auer, P.: Weak hypotheses and boosting for generic object detection and recognition. In: European Conference on Computer Vision. Volume 2. (2004) 71–84
8. McLachlan, G., Peel, D.: Finite Mixture Models. Wiley Interscience, New York (2000)
9. Jolliffe, I.: Principal Component Analysis. Springer-Verlag, New York (1986)
10. Schölkopf, B., Smola, A., Müller, K.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation **10** (1998) 1299–1319
11. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research **3** (2003) 1157–1182
12. Fraley, C., Raftery, A.: Model-based clustering, discriminant analysis and density estimation. Journal of American Statistical Association **97** (2002) 611–631
13. Parsons, L., Haque, E., Liu, H.: Subspace clustering for high dimensional data: a review. SIGKDD Explor. Newsl. **6** (2004) 90–105
14. Tipping, M., Bishop, C.: Mixtures of probabilistic principal component analysers. Neural Computation **11** (1999) 443–482

15. Moghaddam, B.: Principal Manifolds and Probabilistic Subspaces for Visual Recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence **24** (2002) 780–788
16. Bouveyron, C., Girard, S., Schmid, C.: High-Dimensional Data Clustering. Technical Report 1083M, LMC-IMAG, Université J. Fourier Grenoble 1 (2006)
17. Everingham, M., Zisserman, A., Williams, C., Gool, L.V., et al.: The 2005 PASCAL visual object classes challenge. In: First PASCAL Challenge Workshop. Springer (2006)
18. Cattell, R.: The scree test for the number of factors. Multivariate Behavioral Research **1** (1966) 245–276
19. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal of Computer Vision **60** (2004) 63–86
20. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110