Machine Learning for Visual Concept Recognition and Ranking for Images

Alexander Binder, Wojciech Samek, Klaus-Robert Müller, and Motoaki Kawanabe

Abstract Recognition of a large set of generic visual concepts in Images and Ranking of Images based on visual semantics is one of the unsolved tasks for future multimedia and scientific applications based on image collections. From that perspective improvements of the quality of semantic annotations for image data are well matched to the goals of the THESEUS project with respect to multimedia and scientific services. We will introduce the data-driven and algorithmic challenges inherent in such tasks from a perspective of statistical data analysis and machine learning and discuss approaches relying on kernel-based similarities and discriminative methods which are capable of processing large scale datasets.

1 Introduction

Visual concept recognition in its broadest sense is about the identification of semantic content in images based on the visual information in the pixels of an image. This idea is as old as research on artificial intelligence itself and is pursued since the availability of digitized images and computing machines, yet it started to flourish just in the mid 1990s when computing hardware became affordable and its computing power expanded rapidly. The identification of visual content is one of the

A. Binder and W. Samek
Fraunhofer Institute FIRST, Berlin and Machine Learning Group, TU Berlin, Germany; e-mail: alexander.binder@tu-berlin.de, wojciech.samek@tu-berlin.de

K.-R. Müller
Machine Learning Group, TU Berlin, Germany and Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea e-mail: klaus-robert.mueller@tu-berlin.de

M. Kawanabe
ATR Brain Information Communication Research Laboratory, Kyoto, Japan; e-mail: kawanabe@atr.jp

unsolved tasks for future applications based on digital image collections. With the advent of the internet and consumer digital cameras, digital images have become widely available. This trend is enhanced by the evolution of embedded digital imaging systems in consumer products such as mobile phones and the increased availability in medical applications such as radiology imaging. The collection of image databases like digital archives in medial or medical facilities is facilitated by falling prices for storage devices. All this fuels the demand for systems which are capable of identifying in one or another sense visual contents in a large number of images. We will focus here on semantic concept recognition in images.

Formally, a visual concept can be represented as an indicator function $\mathbb{I}_C$ on the space of all images $\mathscr{X}$ such that $\mathbb{I}_C(x) = 1$ denotes the presence of concept C in an image $x \in \mathscr{X}$.

$$\mathbb{I}_C : \mathscr{X} \longrightarrow \{0,1\} \tag{1}$$

For ambiguous semantic concepts this definition is extended by assigning an image x a score $f_C(x)$ in a bounded interval (e.g. $[0,1]$) which represents a numerical value for the strength of the presence of a semantic concept in an image:

$$f_C : \mathscr{X} \longrightarrow [0,1]\,. \tag{2}$$

This numerical value can be interpreted in a probabilistic manner as the agreement of a set of human annotators with respect to the question whether an image belongs to a semantic concept or not. In the context of statistical methods this is known as label noise. Such ambiguities arise naturally for concepts denoting the emotional impression of an image or concepts related to aesthetic quality.

We set the aim of Semantic Concept Recognition as the prediction of all semantic concepts present in an image. In contrast to multi-class approaches we aim at recognizing multiple concepts in an image, e.g. to find the set of concepts *Outdoor*, *Plants*, *Day*, *NeutralIllumination*, *PartlyBlurred*, *ParkGarden*, *Toy*, *Natural*, *Cute* and *Calm* in one image.

Thus, we are interested in predicting a large set of generic semantic concepts in contrast to a small set of highly specialized concepts as it is the aim of face recognition as an example. One image may show multiple concepts. The prediction output is desired to be a continuous score instead of a binary decision. The continuous score allows to provide hints about uncertainty of the prediction. Such information is highly useful for the common search scenario in which a user is interested to find the top K most likely images for a selected concept. The continuous score may be used for ranking in such a context. Such a ranking-based search scenario has proven to be practical and represents the state of the art for finding images according to their provided filenames by common internet search engines.

The notion of concept recognition may be adapted to more specific concepts: if one considers visual objects with a clearly determinable area, then one can pursue object detection which aims at finding a tight bounding box around the object or segmentation which assigns each pixel in the image a label corresponding for one of the objects or background.

1.1 Why is concept recognition in images a difficult task?

One may ask why common internet search engines use the image search based by filenames as standard tool while search based on visual content appears to be in the beta phase at best. In this section we discuss some issues and challenges of visual concept recognition in images.

Generally speaking one may observe a large variability in the *semantic* structure of visual concepts. This presents a challenge for algorithms capable of predicting semantic concepts and ranking images according to them.

Key factors for the variance in the semantic structure of a semantic concept are the presence and absence of a wide range of all kinds of visual cues, their composition and their contribution to the classification of an image in a non-deterministic manner. The following examples will exemplify the above.

Concepts are defined by the presence of several visual cues in the image. The difference to classical object recognition is that the visual cues may vary highly and may not be classified into one simple object class. Consider the concept *Concert*. Photos showing a small group of people known to be famous music artists on stage are likely to belong to such a concept. At the same time a large group of hobby artists playing in an orchestra also shows a *Concert*.

The composition of cues beyond their mere presence may play an important role: A person holding a guitar in a certain pose may contribute to the classification as a *Concert*. However another pose with a guitar on his back may depict rather a travelling person. Similarly, music at a funeral scene is less likely called a *Concert*. One can think of many setups of musical instruments and people which are more or less likely to be a *Concert*.

This reveals that general semantic concepts are more difficult to recognize compared to classic single object recognition. Another reason besides the wide range of possible cues is that cues contribute in a non-deterministic way (which can be modeled in a probabilistic fashion) to the rating for belonging to semantic concepts. Consider the concept *StreetScene*: the presence of roads and buildings are cues for such a concept however the density and height of buildings, density of roads and the density of parked cars are important for judging whether this is a *StreetScene* or just a lonely road outside a town with some buildings near it. This probabilistic contribution of cues and their composition becomes obvious for concepts related to aesthetic quality or emotional impact such as *Funny* or *Scary*.

We can identify some special cases of the variability of cues which will be mentioned briefly.

- Varying positions and sizes of Regions in an image relevant for a semantic concept:
  When limited to objects one will note that an object can fill a large fraction of the image or a very small region. An smaller object may have a highly varying position within the image. Similarly the appearance of an object may vary with its viewpoint. The same holds for cues contributing to a semantic concept instead of objects.

- Occlusion of Regions in an image relevant for a semantic concept:
  Regions of an image relevant for the recognition of a semantic concept can be occluded.
- Clutter and Complex Scene Compositions:
  Images can have large areas which are irrelevant for the recognition of a visual concept. Consider the example of a living room which contains somewhere an object of class *bottle* which is to be found.

1.2 The Role of Label noise

The informal points discussed above may be interpreted in more formal statistical terms: they have two effects on increasing the difficulty of the classification problem. The first effect in a probabilistic classification setting is an increased irregularity of the optimal decision boundary. In a probabilistic setting the decision boundary for a visual concept would be the set of images x with maximal uncertainty with respect to the question whether a visual concept is present or not: $P(\mathbb{I}_C(x) = 1) = 0.5$.

The second effect is increased label noise. Label noise can be measured as the uncertainty of human annotators in assigning an image to belong to a semantic concept. Mathematically it can be modelled as the probability of an image to belong to a concept $P(\mathbb{I}_C(x) = 1)$.

Label noise has intuitively a deteriorating impact on classification accuracy, and more importantly on model selection. Learning of e.g. a support vector machine corresponds to the choice of a function from a class of functions by via selecting its parameters when solving the optimization problem.

Theorem 6 in [29] provides lower bounds for the expected risk in empirical risk minimization depending on a uniform bound for the label noise which are applicable e.g. to support vector machines with Gaussian kernels on bounded domains for distributions with smoothly differentiable Bayes boundaries.

Theorem 1 (Theorem 6 from [29]) *Let μ be a probability measure on $\mathscr{X}$ and S be some class of classifiers on $\mathscr{X}$ such that for some positive constants K_1, K_2, ε_0 and r*

$$K_2 \varepsilon^{-r} \leq H_1(\varepsilon, S, \mu) \leq K_1 \varepsilon^{-r}$$

for all $0 < \varepsilon \leq \varepsilon_0$, where $H_1(\varepsilon, S, \mu)$ denotes the $\ell_1(\mu)$-metric entropy of S. Then, there exists a positive constant K depending on K_1, K_2, ε_0 and r such that the following bound holds

$$\begin{aligned} R_n(h, S, \mu) &= \inf_{\hat{s} \in S} \sup_{P \in \mathscr{P}(h,S,\mu)} \mathbb{E}[P(Y \neq \hat{s}(X)) - P(Y \neq s^*(X))] \\ &\geq K(1-h)^{\frac{1}{1+r}} \max(h^{-\frac{1-r}{1+r}} n^{-\frac{1}{1+r}}, n^{-\frac{1}{2}}) \end{aligned} \tag{3}$$

whenever $n \geq 2$.

For understanding of the theorem note that $\mathscr{P}(h,S,\mu)$ is the set of distributions on the input-label product space $\mathscr{X} \times \mathscr{Y}$ such that the input space distribution is μ and the label noise is in each point of $\mathscr{X}$ bounded by $1/2 - h/2$: $|P(Y = 1|X = x) - 0.5| \geq h/2$ and s^* is the Bayes classifier. $\mathbb{E}[P(Y \neq \hat{s}(X)) - P(Y \neq s^*(X))]$ is the deviation between the expected errors of the classifier $\hat{s}$ and the a posteriori optimal Bayes classifier s^*. The supremum is taken over a class of distributions followed by selection of the optimal empirical classifier $\hat{s}$ given knowledge of the distribution. Since the underlying distribution of images and their concept labels is unknown this implies that the lower bound is an optimistic formulation and results will be worse in practice.

An increase in the overall label noise corresponds to a decrease of the value of h which yields an increased lower bound in Theorem 1 for the expected deviation between the expected error of an optimistically selected classifier and the statistically best possible classifier within a function class. *The qualitative message is that label noise does have a highly deteriorating influence on model selection.*

1.3 A Machine Learning Approach

As seen in the preceding section, the methodological challenge consists in the ability to deal with a large set of differing visual concepts, ranging from localized objects to overall emotional impressions, the large variability of image appearance present within general visual concepts beyond simple objects like *Partylife, BeachHoliday, Mountains, Indoor, Euphoric, AestheticImpression* and the disagreement of human annotators in labelling images for such concepts.

Statistical learning methods are constructed to be robust against annotator disagreement, labelling errors, varying image qualities and scales of visible cues, differing lighting conditions, occlusion and clutter in images. They are able to extract relevant structures based on *implicit and statistical* definitions such as labelling example images as belonging to the same concept even when there is no way to define deterministic rules for what a concept should be. This makes statistical learning methods a valuable complement to *explicit and deterministic* knowledge modelling via ontologies.

The principle of the imaging solutions is based on countering the large variability of visual concepts by a large set of varying image representations being computed for each image and learning the optimal weighting of these representations for each visual concept based on criteria derived from state of the art statistical machine learning as depicted in Figure 1.

Two key components of the approach will be described in the following: Bag of Words Features and the learning of kernel combinations.

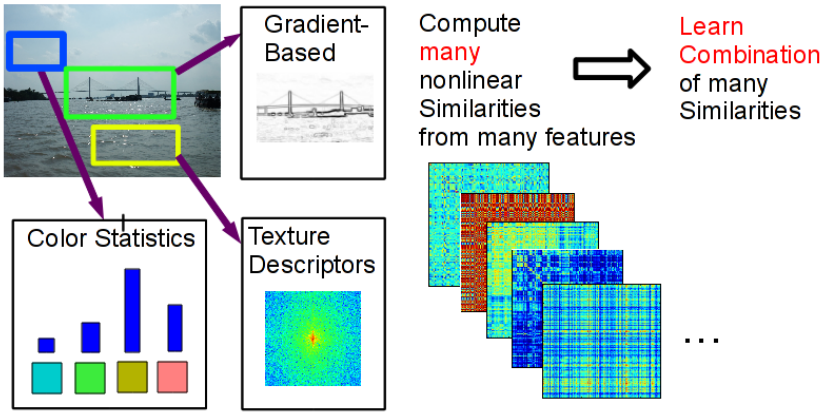

Fig. 1 Graphical depiction of the multiple feature extraction and multiple kernel learning based machine learning approach.

2 Bag of Words Features

In the last decade bag of words features [14] have evolved into the dominating feature extraction approach for visual concept recognition and ranking. They deliver constantly the top-ranked solutions in this field as shown over all the past years in state of the art benchmark challenges like Pascal Visual object categorization [16] and ImageCLEF PhotoAnnotation [32] which deal with images with high visual variance. For image recognition problems on highly specialized narrow domains, other feature extraction methodologies may deliver better results.

The Bag of words feature is a three stage feature extraction principle which is not linked to the SIFT feature itself. In the first stage a set of local features is computed from an image. Formally, a local feature is a vector computed over a localized subset of the image. The SIFT descriptor [28] is the most popular choice. Besides the choice of the local feature, regions for its computation have to be chosen. Typically, local features are computed on small overlapping regions taken from the whole image. Apart from grid sampling, biased random sampling [47, 7] may serve for computation of the corresponding descriptor regions. The number of local features may vary across images, for example by adaptation to image size. The second stage, the computation of the set of visual words, is done once during training time. Formally, a visual word is a point in the space of local features. One possibility to compute the visual words is discretization of the local feature density using k-means. Practically proven alternatives are radius-based clustering [22], Bayesian methods like pLSA [20] and more commonly Fisher vectors based on Gaussian mixture models [13], sparse coding [46]. There has been considerable research on improvements for visual word generation, such as hierarchical clustering [31], class-wise clustering [42], weakly-supervised clustering [9] or optimization of information-theoretic cri-

teria [48]. Finally the last stage is the aggregation of local features into the global Bag of Words feature. The local features are mapped on the visual words, usually with weights based on the distances between the local feature and all the prototypes. Examples are soft codebooks [19] and fast local linear coding [41].

The strength of the bag of words feature lies in its robustness which comes from the following factors:

- the absence of modelling of spatial relations between parts unlike earlier approaches which are susceptible to noise in images with complex sceneries.
- the aggregation of local features into a global feature which implies denoising via averaging of contributions of many local features.
- the choice of robust local features such as SIFT [28] or SURF [3] which are known to be invariant against many changes in lighting conditions. See [35] for an overview from a color theoretic point of view.

Another advantage of bag of words features is their computational scalability. This is an advantage over intuitively more appealing Bayesian approaches which often need to rely on restricted probability models or inference approximations in practice. Computation of bag of words features in real-time is demonstrated in [40]. [36] demonstrates their efficient computation on GPUs.

The work [33] shows by comparing against human performance that Bag of words features yield a similar performance to humans on jumbled images which were cut into square parts and then piecewise randomly permutated and rejoined. The human advantage is our ability to extract spatial relations between parts which requires us, however, to spend years of learning in childhood from millions of examples and some hundred thousand years of brain evolution for the base learning system being operational.

The BoW method is also applied with superior results in competitions in related domains such as semantic indexing for videos in TRECVID [21] or the winning entry in ILSVRC2011 large scale object detection challenge [37].

Despite their robustness for domains with highly variable images Bag of words features are also applied to narrow domains such as concept recognition for medical images [1, 44, 12].

In conclusion, Bag of words features are very well suited for image annotation tasks.

3 Learning Kernel Combinations

Once there are many possible feature representations at hand for each image, the question arises how to combine them. The answer from the point of statistical machine learning is to learn a combination from image features and labels, trying to separate the images belonging to a visual concept from the rest in a high dimensional feature space. The typical dimensionality of single features in the order of several thousands limits the usefulness of classification trees and simple nearest neighbor

algorithms. Several principled machine learning algorithms have been developed like the classic Boosting [17] as the most famous one.

We will focus on kernel-based algorithms [30]. A kernel can be said to be a positive definite matrix of similarity values between data samples. These similarity values have been computed from the features available for each image. Given a loss function, the similarity values contained in a kernel matrix can be used to learn a classifier which predicts the annotation. A well established method for learning such a classifier are support vector machines [11] which have been used with success in a variety of fields such as genome analysis, computational chemistry, finance, brain-computer interfacing, image and natural language processing.

The learning of kernel combinations relies on Multiple kernel learning [24] which extends support vector machines (see formula 4). It learns weights for a linear combination of kernels in addition to the parameters of an ordinary support vector machine for each visual concept. The need for learning kernel combinations in image annotation tasks has been widely recognized in the international community as seen by the application of sparse multiple kernel learning to object recognition [25, 18], object detection [26], or developments of alternative non-sparse Kernel learning algorithms [45] which were influenced by Multiple kernel learning.

$$\begin{aligned} \min_{\beta,\mathbf{w},b,\xi} \quad & \frac{1}{2}\sum_{j=1}^{m}\beta_j\mathbf{w}_j'\mathbf{w}_j + C\|\xi\|_1 \\ \text{s.t.} \quad & \forall i:\ y_i\left(\sum_{j=1}^{m}\beta_j\mathbf{w}_j'\psi_j(\mathbf{x}_i)+b\right) \geq 1-\xi_i \\ & \xi \geq \mathbf{0}\,; \quad \beta \geq \mathbf{0}\,; \quad \|\beta\|_p \leq 1. \end{aligned} \tag{4}$$

The advantage of non-sparse multiple kernel learning lies in the combination of the robustness of the support vector machine with adjustable regularization for the learning of kernel weights. The latter constitutes the essential difference to its predecessor, the sparse multiple kernel learning [27, 2, 38].

The image annotation system developed by the Fraunhofer FIRST and the TU Berlin has been tested in several international benchmark competitions with undisclosed ground truth for the test data. The joint efforts resulted in being the third best group at the ImageCLEF2009 Photo Challenge, the fourth best group at the Pascal VOC 2009 Object Classification Challenge, and finally, submitting the winning entries [7] for two categories of ImageCLEF2011 PhotoAnnotation Challenge [32], namely multi-modal and visual Ranking by the mean average precision measure for which multiple kernel learning was employed.

4 Outlook: Beyond Visual Concept Annotation - Towards Semantic Scene Understanding

This chapter has reviewed on the basis of selected scientific contributions that Machine Learning has proved its high usefulness in practical image annotation [34, 23]. This was – along with excellent publications [9, 42, 43, 6, 8] – also demonstrated in successful submissions to international annotation competitions [32, 16, 4] and most recently the best visual and multi-modal submission entries in the ImageCLEF2011 PhotoAnnotation Challenge [7].

When striving for intelligent Machine Vision a number of hard challenges still lie ahead. Multimodal sources of information and their respective hierarchical status have to fused on firm theoretical grounds, here a lot is still to be done. Machine Learning based Visual Concept Annotation and ranking needs to be combined with semantic knowledge such as taxonomies in order to express semantic relations between visual concept classes [5, 15]. Semantic classes are far from independent, thus a challenge will be to model dependency structure appropriately; Multi-Task Learning will only be a first step into this important direction. Finally, the real world is a complex, structured and highly non-stationary environment. What statistical structure may hold for an image ensemble in 2000 may not be up to date in 2012. Thus it will be important to establish high performing learning methods that work stably despite the non-stationarity of the world around us and all its structural changes [39, 10]. The emerging fields of human action classification and anomaly detection in surveillance videos require solutions for such problems.

Acknowledgements

This work was supported by the Federal Ministry of Economics and Technology of Germany (BMWi) under the project THESEUS (FKZ 01MQ07018). Furthermore it was in part supported by the World Class University Program through the National Research Foundation of Korea funded by the Korean Ministry of Education, Science, and Technology, under Grant R31-10008. We like to express our thanks to Volker Tresp, the work package leader at CTC WP6, Ralf Schäfer from the Fraunhofer HHI and Shinichi Nakajima from the Nikon corporation for the fruitful collaboration.

References

1. André, B., Vercauteren, T., Buchner, A.M., Wallace, M.B., Ayache, N.: Retrieval evaluation and distance learning from perceived similarity between endomicroscopy videos. In: MICCAI, LNCS, p. 8p. Springer, Heidelberg (2011)

2. Bach, F., Lanckriet, G., Jordan, M.: Multiple kernel learning, conic duality, and the smo algorithm. ICML (2004)
3. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer Vision Image Understanding **110**, 346–359 (2008). DOI 10.1016/j.cviu.2007.09.014
4. Binder, A., Kawanabe, M.: Enhancing recognition of visual concepts with primitive color histograms via non-sparse multiple kernel learning. In: CLEF (2), pp. 269–276 (2009)
5. Binder, A., Müller, K.R., Kawanabe, M.: On taxonomies for multi-class image categorization. International Journal of Computer Vision pp. 1–21 (2011). DOI 10.1007/s11263-010-0417-8
6. Binder, A., Nakajima, S., Kloft, M., Müller, C., Samek, W., Brefeld, U., Müller, K.R., Kawanabe, M.: Insights from classifying visual concepts with multiple kernel learning. PLoS ONE Submitted 2011.
7. Binder, A., Samek, W., Kloft, M., Müller, C., Müller, K.R., Kawanabe, M.: The joint submission of the TU Berlin and Fraunhofer FIRST (TUBFI) to the ImageCLEF2011 photo annotation task. In: CLEF (Notebook Papers/Labs/Workshop) (2011)
8. Binder, A., Samek, W., Müller, K.R., Kawanabe, M.: Enhanced representation and multi-task learning for image annotation. Computer Vision and Image Understanding Submitted 2011.
9. Binder, A., Wojcikiewicz, W., Müller, C., Kawanabe, M.: A hybrid supervised-unsupervised vocabulary generation algorithm for visual concept recognition. In: ACCV (3), pp. 95–108 (2010)
10. von Bünau, P., Meinecke, F.C., Király, F.J., Müller, K.R.: Finding stationary subspaces in multivariate time series. Physical Review Letters **103**(21), 214,101 (2009). DOI 10.1103/PhysRevLett.103.214101
11. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3), 273–297 (1995)
12. Cruz-Roa, A., Caicedo, J.C., González, F.A.: Visual pattern mining in histology image collections using bag of features. Artificial Intelligence in Medicine **52**(2), 91–106 (2011)
13. Csurka, G., Perronnin, F., Marchesotti, L., Clinchant, S., Ah-Pine, J.: Fisher kernel representation of images and some of its successful applications. In: VISAPP (1), pp. 21–25 (2010)
14. Dance, C., Willamowski, J., Fan, L., Bray, C., Csurka, G.: Visual categorization with bags of keypoints. In: ECCV International Workshop on Statistical Learning in Computer Vision (2004). URL http://www.xrce.xerox.com/Publications/Attachments/2004-010/2004_010.pdf
15. Deng, J., Berg, A.C., Li, F.F.: Hierarchical semantic indexing for large scale image retrieval. In: CVPR, pp. 785–792 (2011)
16. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision **88**(2), 303–338 (2010)
17. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: EuroCOLT, pp. 23–37 (1995)
18. Gehler, P.V., Nowozin, S.: On feature combination for multiclass object classification. In: ICCV, pp. 221–228 (2009)
19. van Gemert, J., Geusebroek, J.M., Veenman, C.J., Smeulders, A.W.M.: Kernel codebooks for scene categorization. In: ECCV(3), pp. 696–709 (2008)
20. Hofmann, T.: Probabilistic latent semantic analysis. In: UAI, pp. 289–296 (1999)
21. Inoue, N., Kamishima, Y., Wada, T., Shinoda, K., Sato, S.: TokyoTech+Canon at TRECVID 2011. In: 2011 TREC Video Retrieval Evaluation (2011). URL http://www-nlpir.nist.gov/projects/tvpubs/tv11.papers/tokyotechcanon.pdf
22. Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: ICCV, pp. 604–610 (2005)
23. Kawanabe, M., Binder, A., Müller, C., Wojcikiewicz, W.: Multi-modal visual concept classification of images via markov random walk over tags. In: WACV, pp. 396–401 (2011)
24. Kloft, M., Brefeld, U., Sonnenburg, S., Zien, A.: Lp-norm multiple kernel learning. Journal of Machine Learning Research **12**, 953–997 (2011)
25. Kumar, A., Sminchisescu, C.: Support kernel machines for object recognition. In: ICCV (2007)
26. Lampert, C., Blaschko, M.: A multiple kernel learning approach to joint multi-class object detection. In: DAGM, pp. 31–40 (2008)

27. Lanckriet, G.R., Cristianini, N., Bartlett, P., Ghaoui, L.E., Jordan, M.I.: Learning the kernel matrix with semidefinite programming. Journal of Machine Learning Research pp. 27–72 (2004)
28. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
29. Massart, P., Nedelec, E.: Risk bounds for statistical learning (2007). URL http://arxiv.org/abs/math.ST/0702683
30. Müller, K.R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B.: An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks **12**(2), 181–201 (2001)
31. Nistér, D., Stewénius, H.: Scalable recognition with a vocabulary tree. In: CVPR (2), pp. 2161–2168 (2006)
32. Nowak, S., Nagel, K., Liebetrau, J.: The CLEF 2011 photo annotation and concept-based retrieval tasks. In: CLEF (Notebook Papers/Labs/Workshop) (2011)
33. Parikh, D.: Recognizing jumbled images: The role of local and global information in image classification. In: ICCV, pp. 519–526 (2011)
34. Samek, W., Binder, A., Kawanabe, M.: Multi-task learning via non-sparse multiple kernel learning. In: CAIP (1), pp. 335–342 (2011)
35. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(9), 1582–1596 (2010)
36. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Empowering visual categorization with the GPU. IEEE Transactions on Multimedia **13**(1), 60–70 (2011)
37. van de Sande, K.E.A., Uijlings, J.R.R., Gevers, T., Smeulders, A.W.M.: Segmentation as selective search for object recognition. In: ICCV, pp. 1879–1886 (2011)
38. Sonnenburg, S., Rätsch, G., Schäfer, C., Schölkopf, B.: Large Scale Multiple Kernel Learning. Journal of Machine Learning Research **7**, 1531–1565 (2006)
39. Sugiyama, M., Krauledat, M., Müller, K.R.: Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research **8**, 985–1005 (2007)
40. Uijlings, J.R.R., Smeulders, A.W.M., Scha, R.J.H.: Real-time visual concept classification. IEEE Transactions on Multimedia **12**(7), 665–681 (2010)
41. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T.S., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR, pp. 3360–3367 (2010)
42. Wojcikiewicz, W., Binder, A., Kawanabe, M.: Enhancing image classification with class-wise clustered vocabularies. In: ICPR, pp. 1060–1063 (2010)
43. Wojcikiewicz, W., Binder, A., Kawanabe, M.: Shrinking large visual vocabularies using multi-label agglomerative information bottleneck. In: ICIP, pp. 3849–3852 (2010)
44. Xu, R., Hirano, Y., Tachibana, R., Kido, S.: Classification of diffuse lung disease patterns on high-resolution computed tomography by a bag of words approach. In: MICCAI (3), pp. 183–190 (2011)
45. Yan, F., Kittler, J., Mikolajczyk, K., Tahir, A.: Non-sparse multiple kernel learning for fisher discriminant analysis. In: ICDM, pp. 1064–1069. IEEE Computer Society, Washington, DC, USA (2009). DOI 10.1109/ICDM.2009.84
46. Yang, J., Yu, K., Gong, Y., Huang, T.S.: Linear spatial pyramid matching using sparse coding for image classification. In: CVPR, pp. 1794–1801 (2009)
47. Yang, L., Zheng, N., Yang, J., Chen, M., Chen, H.: A biased sampling strategy for object categorization. In: ICCV, pp. 1141–1148 (2009)
48. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: NIPS, pp. 2223–2231 (2009)