Pattern Recognition 59 (2016) 1–4

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Compositional models and Structured learning for visual recognition

1. Introduction

Computer vision and pattern recognition has witnessed a fast growth of research on compositional and hierarchical models such as Deformable Part-based Models, Pictorial Structure Models and And-Or Graphs. The virtue of compositional and hierarchical models (CHMs) lies in their expressive power to model diverse and complex visual patterns. Meanwhile, a set of structured learning and optimization methods are intensively studied to handle the training and inference with compositional models, which usually integrate latent structures to specify the task-specific compositional configurations and contextual relations. These methods, such as latent support vector machines, enable inference with rich internal structures and pursue a good mapping between observations and output structured predictions. Moreover, the recently resurgent deep learning methods (e.g., convolutional neural networks) have been studied in the context of combining with CHMs and achieved very promising results on several pattern recognition tasks.

This special issue aims to collect recent state-of-the-art achievements on how to learn CHMs and other effective models from visual data such as images and videos for visual recognition and related problems. 67 submissions have been received from all over the world, and 27 papers based on a strict and comprehensive review process. Each manuscript has been assigned at least three peer reviewers and have at least two round of review. Hopefully, these accepted papers will provide a good reference to the related fields and applications of learning compositional models.

2. Overview of accepted articles

All accepted papers are closely related to the topics of this special issue, making remarkable progresses in several visual recognition tasks. We can overview these works into three categories, according to their representations and learning models: i) representation-based modeling and classification, ii) deep feature learning, and iii) structural SVM and CRF learning.

2.1. Representation-based modeling and classification

Representation-based modeling is an important class of approaches for pursuing and exploiting the intrinsic structure of complex visual patterns. With the success of sparse and low rank models, representation-based methods have been extensively investigated in computer vision and widely adopted in various clustering, classification, detection and matching tasks.

In this special issue, there are 14 articles falling into this category, where sparse representation, low rank representation, covariance descriptors, and Gaussian mixture models have been investigated. In the article “Structure-Constrained Low-Rank and Partial Sparse Representation with Sample Selection for image classification”, Liu et al. first introduce the structure-constrained low-rank model to dictionary learning, and then suggest a low-rank and partial sparse representation algorithm to exploit the correlation between the test sample and the training samples. Experiments on Caltech 101 and Caltech 256 validate its effectiveness in terms of classification accuracy.

In the article “Face Recognition Using Linear Representation Ensembles”, Li et al. present a linear representation ensemble (LRE) method by training an ensemble model over probabilistic patch representation. Generic face confidence and boost-like algorithms are further suggested to tune ensemble weights. The results on four face databases demonstrate the effectiveness of LRE.

In the article “Spectral-Spatial Hyperspectral Image Ensemble Classification via Joint Sparse Representation”, Zhang et al. present a joint sparse representation model to select a compact set of classifier and to assign different weights to the selected classifiers for hyperspectral image ensemble classification. Experiments show that classifier selection is effective in improving the classification accuracy and efficiency in the test stage.

In dictionary-based sparse representation, Zhang et al. in the article “Class Relatedness Oriented-Discriminative Dictionary Learning for Multiclass Image Classification” to explore the relationship between dictionary atoms and the class labels. For sparse representation-based classification, they develop a class relatedness oriented discriminative dictionary learning method by exploiting the $l_{1,\infty}$-norm regularization. This method not only achieves promising classification performance, but also provides an elegant approach to adaptively discover the class relatedness from training data.

For multi-spectral face recognition with multi-view dictionary learning, i.e., MLSDL, Jing et al. in the article “Multi-spectral Low-rank Structured Dictionary Learning for Face Recognition” use both spectrum-common and spectrum-specific dictionaries to exploit the the correlation and complementarity among multiple spectra. Extensive experiments demonstrate the superiority of MLSDL on the HK PolyU, CMU and UWA hyper-spectral face databases.

In the article “Adaptive Noise Dictionary Construction via IRRPCA for Face Recognition”, Chen et al. study the robust regression-based face recognition, and propose an Iteratively Reweighted Robust Principal Component Analysis (IRRPCA) to get an effective noise dictionary. Experiments on several face databases, e.g., AR, Yale B, CMU PIE, CMU

http://dx.doi.org/10.1016/j.patcog.2016.07.030

Multi-PIE, LFW and Pubfig, showed the robustness of the method against occlusion, corruption, illumination, misalignment.

In the article "Sparsity-inducing Dictionaries for Effective Action Classification", Roy, Srinivas and Mohan adopt the features obtained using sparsity-inducing dictionaries for discriminative action representation and classification. Significant performance improvement is achieved on the HMDB51 dataset by the method.

In the article "Transformation Invariant Subspace Clustering", Li et al. study the joint learning of sample alignment and subspace clustering, and suggest a transformation invariant subspace clustering model. On unaligned real data, the method can achieve better clustering results than the state-of-the-art methods.

In the article "Towards Effective Codebookless Model for Image Classification", Wang et al. suggest a codebookless model (CLM) for representing an image with a single Gaussian. Combining with Gaussian embedding and low-rank SVM learning, CLM is both effective and efficient while comparing with the state-of-the-art bag-of-features (BoF) methods.

In the article "Category co-occurrence modeling for large scale scene recognition", Song et al. study the problem of feature representation for scene recognition. Specifically, a co-occurrence model is presented to exploit semantic multi-normial (SMN) patterns jointly in a common semantic space. Experiments further validate its superiority to the other SMN-based representation methods.

In the article "Convex Hull indexed Gaussian Mixture Model (CH-GMM) for 3D Point Set Registration", Fan et al. introduce a convex hull indexed Gaussian mixture model (CH-GMM) for representing 3D point set to preserve the topological structure. Compared with the leading methods, CH-GMM achieves better robustness and accuracy for 3D point set registration.

For background segmentation and extraction, Ge et al. in the article "Dynamic Background Estimation and Complementary Learning for Pixel-wise Foreground/Background Segmentation" present two modifications, i.e., dynamic background estimation and complementary learning, and embed them into three background models, i.e., Gaussian mixture model (GMM), sample-based model, and code book. Extensive experiments demonstrate the effectiveness and efficiency of the methods.

In the article "A Discriminative Representation for Human Action Recognition", Yuan, Zheng and Lu introduce a model to combine the parameterized probabilistic representation and discriminative classifier for human action recognition. Alternating minimization is adopted to obtain the solution. Experimental results on five datasets validate the effectiveness of the method.

In the article "Kinship-Guided Age Progression", Shu et al. suggest a kinship-guided age progression approach. Three major modules are adopted to preserve individual aging characteristics, capture aging tendency, and guide aging direction, respectively. The method is effective in alleviating the adverse effect caused by the non-deterministic characteristics of age progression and the unapparent identity information for people at the tender age.

2.2. Deep feature representation learning

Five articles focus on learning image feature representations by employing neural networks or other learning models. These methods mainly induce their representations in a hierarchical manner, e.g., gradually enriching the representations from bottom to top.

In the article "Learning Structure of Stereoscopic Image for No-Reference Quality Assessment with Convolutional Neural Network", Zhang et al. present an NR IQA metric for stereoscopic images based on convolutional neural networks. The proposed method demonstrates the effectiveness of feature learning compared to handcrafted feature based methods, and achieves a new state-of-the-art on the public stereoscopic benchmarks, e.g., LIVE phase-I, LIVE phase-II, and IVC.

In the article "Scene parsing using inference Embedded Deep Networks", Bu et al. present a novel neural network architecture, named Inference Embedded Deep Networks (IEDNs), for semantic image segmentation. IEDNs is capable of not only learning discriminative image features but also encapsulating spatial relationship information among adjacent objects for enhancing discriminability of the representation. This proposed model is evaluated on SIFT Flow and PASCAL VOC datasets and shows superior accuracy over some other existing methods.

In the article "Human action recognition using genetic algorithms and convolutional neural networks", Ijjina and Chalavadi propose an approach for human action recognition using genetic algorithms (GA) and deep convolutional neural networks. The GA-generated classifier is incorporated into the neural network architecture for improving recognition performance. The experiments and comparisons are conducted on the UCF50 dataset, and the proposed method shows promising results in both accuracy and efficiency.

In the article "Convolutional neural random fields for action recognition", Liu et al. address human action recognition by developing a deep learning framework named a convolutional neural random fields (CNRFs). This framework naturally combines the convolutional neural network model and the conditional random fields for end-to-end training. In the experiments, CNRFs shows its superiority on both segmented and unsegmented action video datasets against some other existing methods.

In the article "Robust lane detection using two-stage feature extraction with curve fitting", Niu et al. develop a well-engineered approach for lane detection. This approach performs two stages of feature extraction by employing the modified HT (Hough Transform) and DBSCAN (Density Based Spatial Clustering of Applications with Noise) cluster algorithms. The experimental results that the proposed method can handle real challenges in lane detection and achieve promising performances.

In the article "Learning to Segment with Image-level Annotations", Wei et al. present a deep learning framework for weakly-supervised semantic image segmentation, e.g., only image-level annotations are available. This framework consists of two following stages. First, reliable hypotheses based localization maps are generated by incorporating the hypotheses-aware classification and cross-image contextual refinement. Second, the segmentation neural network can be trained in a supervised manner by these generated localization maps. The proposed method achieves new state-of-the-art results on PASCAL VOC 2012 benchmark.

2.3. Structural SVM and Graph-based Learning

Structured learning and optimization methods are extensively investigated in recent visual recognition problems. In this special issue, nine articles fall into this category of methods.

In the article "Extended Compressed Tracking via Random Projection Based on MSERs and Online LS-SVM Learning", Cao et al. develop a tracking method based on sparse random projection and online least squares SVM classifier(LS-SVM) learning. An online closed-form LS-SVM is proposed to quickly and robustly predict the target location in tracking. Experimental results on standard benchmark sequences show the stability and robustness of the proposed method compared to some existing trackers.

In the article "Hierarchical Mixing Linear Support Vector Machines for Nonlinear Classification", Wang et al. address nonlinear classification problem by developing a new classifier called HMLSVMs (Hierarchical Mixing Linear Support Vector Machines). This learning framework is built with a hierarchical structure with a mixing linear SVM classifier at each node, and it can predict the sample"s label using only a few hyperplanes. Experimental evaluations show that the proposed learning method outperforms other kernel SVM based methods.