Special issue on weakly supervised learning

Zhang, L.; Ji, R.; Yi, Z.; Lin, W.; Snoek, C.

DOI
10.1016/j.jvcir.2016.02.012

Publication date
2016

Document Version
Final published version

Published in
Journal of Visual Communication and Image Representation

J. Vis. Commun. Image R. 37 (2016) 1–2

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Special issue on weakly supervised learning

Most modern computer vision systems involve models learned from human-labeled image examples. For instance, an object detector is typically trained on a large collection of images manually annotated with masks or bounding boxes denoting the location of the object of interest in each photo. The reliance on time-consuming human labeling poses a significant limitation to practical applications of these methods. Weakly supervised learning aims at reducing the amount of human intervention needed to train the models by making use of examples that are only partially labeled.

Submissions came from an open call for papers and finally 9 papers are selected after rigorous reviews. These papers cover wide applications based on weakly supervised learning, such as Image binarization, augmented reality, image retrieval, image recognition, 3D object retrieval, visual tracking, and other applications. We divide the whole special issue into four parts according to the topics of the papers.

There are three papers related to visual tracking. In "Network in Network (NIN) based Weakly Supervised Learning for Visual Tracking," Chen et al. propose a tracking method which constructs a robust object appearance model via learning and transferring mid-level image representations using a deep network. First, they design a simple yet effective method to transfer the mid-level features learned from NIN with large scale training data for tracking tasks with limited labeled data. Then, to address the drifting problem, they simultaneously utilize the samples collected in the initial and most previous frames. Finally, a heuristic schema is used to judge whether to update the object appearance model or not. Extensive experiments show the robustness of their method. The second paper is "Adaptive Visual Target Detection and Tracking using Weakly Supervised Incremental Appearance Learning and RGM-PHD Tracker," in which Dehkordi et al. propose an adaptive appearance model and develop an incremental appearance learning algorithm to learn the target appearances in time. The proposed method employs the background information and the defined keypoints' mismatching history to adapt the target appearances within different frames. Furthermore, they combine the Refined Gaussian Mixture Probability Hypothesis Density (RGM-PHD) tracker with the detectors to keep target trajectories and handle uncertainties. The experiments conducted on several video datasets show the effectiveness of their proposed method. The last paper in this part is "APPOS: An Adaptive Partial Occlusion Segmentation Method for Multiple Vehicles Tracking," in which Zhao et al. propose an adaptive partial occlusion segmentation (APPOS) method for multi-vehicle tracking. In this method, the occlusion detection process is firstly conducted to discover the occlusion. After that, the candidate regions of the respective occluded vehicles are roughly evaluated by the contour's optical flow. Finally, the line scanning which uses color contrast among regions is adopted to accurately locate the vehicles. They evaluate the effectiveness and accuracy of APPOS by the experiments on both real-world and simulated videos.

There are four papers related to image retrieval, categorization, 3D model retrieval, sparse learning and robust recovery. The first paper is "Local consistent hierarchical Hough match for image re-ranking," in which Cai et al. propose an accurate yet efficient image re-ranking algorithm specific for a small vocabulary. Their idea is inspired by Hough voting in the transformation space, where votes come from local feature matches. Most notably, this geometry re-ranking can easily be aggregated to cutting-edge image based retrieval yielding superior performance with a small vocabulary and being able to store in the mobile end facilitating mobile visual search systems. They further prove that its time complexity is linear with the re-ranking instance, and this is a significant advantage over the existing scheme. In terms of mean average precision, they show that its performance is comparable or in some cases better than the state-of-the-art re-ranking schemes. The second paper in this part is "3D Object Retrieval Based on Sparse Coding in Weak Supervision," in which Nie et al. apply sparse coding in a weakly supervision manner to address 3D model retrieval. First, each 3D object, which is represented by a set of 2D images, is used to learn a dictionary. Then, sparse coding is used to compute the reconstruction residual for each query object. Finally, the residual between the query model and the candidate model is used for 3D model retrieval. In the experiment, ETH, NTU and ALOL dataset are used to evaluate the performance of the proposed method. The results demonstrate the superiority of the proposed method. The third paper is "Towards robust subspace recovery via sparsity-constrained latent low-rank representation," in which Li et al. present the Sparse Latent Low-rank representation (SLL) method, which explicitly imposes the sparsity constraint on Latent LRR to encourage a sparse representation. In this way, each data point can be represented by only selecting a few points from the same subspace. Its objective function is solved by the linearized Augmented Lagrangian Multiplier method. Favorable experimental results on subspace clustering, salient feature extraction and outlier detection have verified promising performances of their method. In the last paper "An Aerial Image Recognition Framework using Discrimination and Redundancy Quality Measure," Hu et al. propose a new aerial image categorization framework, focusing on organizing the local patches of each aerial image into multiple discriminative subgraphs. These meaningful

http://dx.doi.org/10.1016/j.jvcir.2016.02.012
1047-3203/

subgraphs reflect both the geometric property and the color distribution of an aerial image. Firstly, each aerial image is decomposed into a collection of regions in terms of their color intensities. Thereby the region connected graph (RCG), which models the connection between the spatial neighboring regions, is constructed to encode the spatial context of an aerial image. Secondly, a novel subgraph mining technique is adopted to discover the frequent structures in the RCGs constructed from the training aerial images. Thereafter, a set of refined structures is selected among the frequent ones toward being highly discriminative and low redundant. Lastly, given a new aerial image, its sub-RCGs corresponding to the refined structures are extracted. They are further quantized into a discriminative vector for SVM classification. Thorough experimental results validate the effectiveness of the proposed method. In addition, the visualized mined subgraphs show that the discriminative topologies of each aerial image are discovered.

There is one paper related to augmented reality: "An Improved Augmented Reality System Based on AndAR," in which Chen et al. propose a 3D registration method based on Oriented fast and Rotated Brief (ORB) and optical flow. ORB is used for feature point matching and RANSAC is used to choose good matches, called inliers, from all the matches. When the ratio of inliers is more than 50% in a video frame, inliers tracking based on optical flow is used to calculate the homography matrix in the later frames and when the number of inliers successfully tracked is small, it goes back to ORB feature point matching again. The result shows that the improved AndAR can augment not only reality based on markers but also reality based on planar natural features in near real time and the hybrid approach can not only improve speed but also extend the usable tracking range.

The last paper is related to image binarization. It is "Genetic Algorithm and Mathematical Morphology Based Binarization Method for Strip Steel Defect Image with Non-uniform Illumination," by Liu et al., which proposes an enhancement operator based on mathematical morphology (EOBMM). A binarization method based on genetic algorithm (BMBGA) is applied to the binarization of the strip steel defect image processed by EOBMM. The experiment results show that their method is effective and efficient in the strip steel defect image binarization, and outperforms the traditional image binarization method by Otsu and Bernsen.

To conclude, the papers in this special issue cover different techniques and applications of weakly supervised learning. It is hoped that this special issue will benefit researchers and practitioners in this emerging area.

Acknowledgments

We would like to thank Drs. Ming-Ting Sun, Zicheng Liu and Yehoshua Y. Zeevi for providing the chance for us to organize this special issue. We also thank the reviewers for their efforts to ensure the high quality of this special issue. Finally, we would like to thank all the authors who have contributed to this special issue. Thanks also go to all the other colleagues in JVCI who help us to make this special issue a reality.

Luming Zhang
Rongrong Ji
Zhen Yi
Weisi Lin
Cees Snoek