Chapter 29
Context-Based Scene Understanding

Esfandiar Zolghadr
Florida Atlantic University, USA

Borko Furht
Florida Atlantic University, USA

ABSTRACT

Context plays an important role in performance of object detection. There are two popular considerations in building context models for computer vision applications; type of context (semantic, spatial, scale) and scope of the relations (pairwise, high-order). In this paper, a new unified framework is presented that combines multiple sources of context in high-order relations to encode semantical coherence and consistency of the scenes. This framework introduces a new descriptor called context relevance score to model context-based distribution of the response variables and apply it to two distributions. First model incorporates context descriptor along with annotation response into a supervised Latent Dirichlet Allocation (LDA) built on multi-variate Bernoulli distribution called Context-Based LDA (CBLDA). The second model is based on multi-variate Wallenius' non-central Hyper-geometric distribution and is called Wallenius LDA (WLDA). WLDA incorporates context knowledge as bias parameter. Scene context is modeled as a graph and effectively used in object detection framework to maximize semantical consistency of the scene. The graph can also be used in recognition of out-of-context objects. Annotation metadata of Sun397 dataset is used to construct the context model. Performance of the proposed approaches was evaluated on ImageNet dataset. Comparison between proposed approaches and state-of-art multi-class object annotation algorithm shows superiority of presented approach in labeling of scene content.

INTRODUCTION

Current state-of-the-art image retrieval applications rely on metadata to recognize the content of images. Source of metadata information usually is manual annotation, a process that is prone to inaccuracy and cannot scale with the continually growing size of today's multimedia content. Automated annotation requires scene analysis and object recognition using information obtained from acquisition system.

DOI: 10.4018/978-1-5225-5204-8.ch029

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Scene understanding has been studied for decades in specialized areas such as surveillance and sporting event recognition. Main objective of this field is to provide more precise and accurate description of the scene to better serve the user queries. A successful solution for this task must be able to recognize the scene elements and to be able to interpret various interactions among them such as events.

Image visual features and the contextual information are two main sources of information used by visual cognitive system in humans which performs scene recognition. Variations in image quality or acquisition conditions can significantly affect the quality of the visual features. To mitigate variations and improve robustness of features, algorithms such as Scale-Invariant Feature Transform (SIFT) (Lowe, 2004), Speed-Up Robust Features (SURF) (Bay, Ess, Tuytelaars, & Van Gool, 2008), Binary Robust Invariant Scalable Keypoints (BRISK) (Stefan Leutenegger & Siegwart, 2011) and Oriented Fast and Rotated Brief (ORB) (Rublee, Rabaud, Konolige, & Bradski, 2011) have been developed. Starting with SIFT as first generation of gradient based methods, these algorithms detect best keypoints in a scale pyramid type and extract orientation using the directed gradients or moments. The computational costs of the feature point calculation in earlier methods have been the motivation for continuous development of more efficient and robust algorithms. But even most superior algorithms fail to produce high quality features in extremely poor and degraded image conditions. Study on human visual cognitive system show superior recognition ability even in children when images of various qualities are presented with a familiar “context”. In computer vision, the context is referred to any information that can be used in accurate semantical understanding of a scene and recognition of its element.

Previous works show evidence that performance of recognition systems can be improved when context is exploited (Wang, Chen, & Wu, 2011; Zhang, Kalashnikov, Mehrotra, & Vaisenberg, 2014; Galleguillos & Belongie, 2010). Contextual models can capture properties, relationships and interactions among image components to infer more accurate meaning of a scene (Choi, Torralba, & Willsky, 2011; Zhu, Nayak, & Roy-Chowdhury, 2013).

One of the most effective context taxonomies is semantic context which is expressed as co-occurrence statistics of image components (Shotton, Winn, Rother, & Criminisi, 2009). In natural language analysis, “Latent Dirichlet Allocation” (LDA) is a generative process that uses latent variables in discovery of intrinsic structure of document and maximizes words co-appearance probability. Given a set of documents as collections of words, LDA decomposes distribution of the words in each document into the mixture of K topics such that a document is viewed as the mixture of topics which itself is multinomial distribution of words. Supervised LDA learns a small set of topics that account for most of the correlation and reveals semantic structure of the document. These models explore various associations between the image features and the corresponding annotation words (Wang, Blei, & Fei-Fei, 2009). LDA topic modeling approach has been widely used in computer vision to discover latent semantical correlation of image visual patterns for clustering and classification applications (Huang, Zhou, & Zhang, 2014; Bahmanyar, Cui, & Datcu, 2015).

In this paper two new models are presented. First model is extension of a supervised LDA-bin probabilistic model (Putthividhya1, Attias, & Nagarajan, 2010) that learns statistical distributions of the inter-object coherency and intra-class relationships influenced by context. This model is named “Context-Based LDA” (*CBLDA*). This approach uses a multi-variate Bernoulli distribution in prediction of the response variable in the LDA process. The second model is a new model build on a multivariate Wallenius distribution (Fog, 2008; Chesson, 1976). This model is incorporated with the contextual descriptor as bias weight in the multi-class classification application.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/chapter/context-based-scene-understanding/196977?camid=4v1

This title is available in InfoSci-Books, Communications, Social Science, and Healthcare, InfoSci-Computer Science and Information Technology, InfoSci-Media and Communications, Science, Engineering, and Information Technology, InfoSci-Select, InfoSci-Select. Recommend this product to your librarian:
www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Edge Detection by Maximum Entropy: Application to Omnidirectional and Perspective Images
Ibrahim Guelzim, Ahmed Hammouch, El Mustapha Mouaddib and Driss Aboutajdine (2013). *Intelligent Computer Vision and Image Processing: Innovation, Application, and Design (pp. 146-159).*
www.igi-global.com/chapter/edge-detection-maximum-entropy/77038?camid=4v1a

Digital Image Matting: A Review
Sweta Singh and Anand Singh Jalal (2013). *International Journal of Computer Vision and Image Processing (pp. 16-36).*
www.igi-global.com/article/digital-image-matting/103956?camid=4v1a

Creating Sound Glyph Database for Video Subtitling
Chitralekha Ganapati Bhat and Sunil Kumar Kopparapu (2017). *Multi-Core Computer Vision and Image Processing for Intelligent Applications (pp. 136-154).*
www.igi-global.com/chapter/creating-sound-glyph-database-for-video-subtitling/163729?camid=4v1a

Educational Data Mining and Indian Technical Education System: A Review
Nancy Kansal, Vijender Kumar Solanki and Vineet Kansal (2017). *Feature Detectors and Motion Detection in Video Processing (pp. 18-34).*
www.igi-global.com/chapter/educational-data-mining-and-indian-technical-education-system/170210?camid=4v1a