論文の内容の要旨
Abstract of Dissertation

論文題目　Object-level visual mining
（オブジェクトレベルのビジュアルマイニング）

氏　　名　　張　拳石

Name of Author　　Zhang Quanshi

Abstract

In this thesis, the author focuses on one of ultimate dreams and biggest challenges in the field of computer vision, *i.e.*, object-level precise visual mining. The author aims to propose a general platform that can directly mine category models from large, unaligned, and cluttered images without labeling "what is where" for many visual tasks, rather than a technique for a specific visual application.

This will be a breakthrough of computer vision in the era of big data. Based on the proposed method, people can avoid obtaining a conventional style of training category models from a large number of well prepared of training samples. Instead, the category model will be automatically mined from big visual data without much cost of human labeling. In addition, this visual mining idea can be extended from mining category models for object detection to mining category knowledge for many other visual applications, such as object tracking, recognition, segmentation, and 3D reconstruction, Therefore, the proposed object-level visual mining can be also understood as a plausible way that efficiently build a comprehensive knowledge base that contains a huge number of object categories and can provide the object-level knowledge to guide different kinds of visual tasks.

Conventional studies of model training usually developed algorithms considering the terms of improving performance or increasing accuracy, and ignored problems with the choice of the training data. By contrast, the object-level visual mining pays more attention to loosening the requirement for training data, as without sufficient human labeling, the big visual data in the internet contains all typical kinds of challenges in the field of computer vision. For example, the mining process would probably simultaneously suffer from the intra-category variations in texture, rotation, scale, illumination, pose, and structure. To some extent, this is the biggest challenge in the field of computer vision, and to the best of the knowledge, no approaches can automatically overcome all of these intra-category variations without sufficient human labeling. Some methods related to the concept of visual mining escape from this challenge by ignoring some of these variations, *e.g.*, the techniques of object discovery and co-segmentation mainly focus on textural knowledge and ignore the information of structure and scale. However, in this research, the author aims to develop visual-mining methods that can make a full use of these kinds of information, rather than simply ignore some of them. It is because that the author refers to achieve high-level robustness in model mining, *i.e.*, the mined model should have an ability of dealing with different kinds of variations, although most model learning

methods only seek for low-level model robustness by applying X-independent methods or X-invariant features.

Therefore, the author has proposed a set of algorithms for visual mining, including “attributed graph mining” and “unsupervised learning for graph matching”. In particular, attributed graph mining is the core technology. The author first defines the problem of “maximal-size graph mining” in graph domain of attributed relational graphs (ARGs), as the attributed graph mining. This extends the boundary of graph theory from previous graph domain of “labeled graphs”. Attributed graph mining bridges the two main fields of artificial intelligence, *i.e.*, graph theory and computer vision. The author uses the ARG to represent an image, and in this case, the model for small common objects inside a number of unaligned images corresponds to the common subgraph pattern among a set of ARGs. The typical variations in texture, rotation, scale, illumination, pose, and structure can be modeled as the attribute variations of the subgraph pattern. In addition, the task of identifying target objects in unlabeled images can be formulated as a graph matching problem. In this way, the attributed graph mining can be regarded a general tool for category discovery and modeling from ubiquitous images

In addition, the proposed technique has a large number of extended applications besides the basic application of category modeling. Because attributed graph mining can automatically discover the target object in unlabeled images, label the object parts, and determine inter-image part correspondences, the author uses the part correspondences as input for further model training for other visual tasks. For example, this method has been successfully applied to training models for object tracking and single-view 3D reconstruction. The proposed method can also be used to recover the model for the whole object from object fragments.

Furthermore, theoretically speaking, the applications of proposed techniques in this thesis are not limited to mining from ordinary images. They can be applied to other kinds of visual data, such as RGB-D images, videos, 3D point clouds, and so on. Their applications can be even extended to the other fields, for example, mining structural patterns of protein. These techniques can be used, as long as the people want to mine knowledge from fuzzy data that can be represented using ARGs.

In this thesis, the author introduces the algorithms of attributed graph mining and unsupervised learning for graph matching, as well as some technical extensions and applications of them.