# Discovering Class-Specific Visual Patterns for Visual Recognition

Similar to frequent patterns in data mining, visual pattern refers to a recurring composition of visual contents in images or videos, such as repetitive texture regions, common objects among images, or similar actions among videos. Such visual patterns capture the recurrence nature of visual data and can represent the essence of the visual data. Finding such visual patterns is critical to image and video data analysis.

In spite of the recent successes of unsupervised mining of representative visual patterns in unlabeled visual data, for visual recognition tasks, the unsupervised mined visual patterns are often not discriminative enough to distinguish among different classes. One natural way to overcome this limitation is to leverage supervised learning and discover class-specific visual patterns, which is the focus of this thesis. Particularly, we target at discovering the following three types of visual patterns of different structures: (1) class-specific visual pattern of local spatial and feature structure, e.g., local texture structure that can help differentiate different visual classes; (2) class-specific spatial layout patterns, e.g., spatial layout patterns that can help differentiate different visual scenes; (3) class-specific visual pattern of compositional structures, e.g., conjunction (AND) and disjunction (OR) forms of individual visual features that can help differentiate different visual classes.

To discover the above mentioned three types of class-specific visual patterns, this thesis is composed by three technical works. In the first work, we propose to mine mid-level visual phrases from low-level visual primitives, e.g., local image patches or regions, by leveraging local spatial context of visual primitives, multi-feature structure of visual primitives, and also the weakly-supervised image label information. Experiments show that our proposed algorithm can learn more representative and discriminative visual phrases for visual recognition tasks, such as texture pattern discovery, scene clustering and object recognition.

In the second work, we propose to discover class-specific spatial layouts for scene recognition by casting a max-margin optimization problem. Unlike previous methods that either use class-generic spatial layouts or use pre-defined spatial layouts, our joint learning of class-specific spatial layouts and image classifier can achieve superior performance for scene recognition problem, by leveraging the recent deep learning features.

In the third work, we propose a novel branch-and-bound based co-occurrence pattern mining algorithm that can directly mine both optimal conjunctions (AND) and disjunctions (OR) of individual features at arbitrary orders simultaneously. This pattern mining process is integrated into boosting framework such that the weighted error is minimized by the discovered co-occurrence pattern in each boosting step. Experiments

on versatile benchmark datasets show that our proposed algorithm achieves superior performances than algorithms using raw individual features directly.

Compared with unsupervised visual pattern discovery, which usually separates the step of pattern discovery and classification, our method can provide a joint learning of visual pattern discovery and visual recognition. Also, different from conventional visual recognition which emphasize purely on the classification performance, our class-specific visual patterns target more on capturing the essence of difference visual classes, such that we not only can recognize the visual classes, but also can explain and understand why they are different visual classes, thanks to the discovered class-specific visual patterns.