Robust Object Detection by Voting in Multiple Feature Spaces
複数の特徴量空間における Voting システムを用いたロバストな物体検出手法

by

Zhipeng Wang
王志鵬

A Doctor Thesis
博士論文

Submitted to
the Graduate School of the University of Tokyo
on December 14, 2013
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Information Science and Technology
in Computer Science

Thesis Supervisor: Katsushi Ikeuchi　池内　克史
Professor of Computer Science

ABSTRACT

Object detection is a fundamental perceptual skill in human, and plays an important role in machine vision area. Effective object detection methods can help with video surveillance, driving assistance, etc. Researchers improve performance of detection methods mainly by proposing better representative models, better classifiers, or more efficient methods for solution space exploring.

In this work, the performance of detection methods is improved from an aspect of combining information from different channels. The work explores information fusion in novel ways. And the efforts are two-fold: 1) utilizing of motion information by combining it with appearance information, and 2) combining visual and spatial information encoded among the local image features of the same object. Three detection methods are proposed accordingly, and the most important component is a voting system in each of the methods.

The first detection method is developed for real-time applications. In a hierarchical way, this method makes time-consuming steps in its pipeline deal with fewer instances. It combines motion information with appearance information very efficiently, and gets the final results from the voting of local appearance-motion patterns along the temporal dimension. This method gives promising detection results in real time, and gives 100% detection rate and 0% false alarm rate in one of the experiments.

Since the performance of the first method in complex scene is not promising, the second method is proposed. This method extends the Implicit Shape Model, which is a voting system of local appearance information, to incorporate motion information, and outperforms the state-of-the-art method on two datasets. This method also performs well in distinguishing near objects and similar different-class objects.

To improve the efficiency of the voting system in the second method, the third method is proposed, which is Pyramid Match Score for detection. The method does pyramid matching during training and detection for efficiency, and makes full use of the visual and spatial information encoded among the image features of the same object, which improves detection performance.

論文要旨

本論文では、動き、見え、位置を組み合わせた局所特徴量の投票システムによる物体認識手法を提案する。最初に、実時間検出を主眼としたシステムを作成した。候補領域が絞れる場合であるが、時間軸方法の投票により高速かつ正確に動作する。次に複雑なシーンでも動作するようにより多くの空間特徴量を導入したシステムを拡張し、近接する物体同士でも正しく認識するにした。最後に、2 番目の手法を高速化するために PMS(Pyramid Match Score) を開発した。PMS は 2 番目の手法より 100 倍程度高速に動作することを確認した。