

Robust Object Detection by Voting in Multiple Feature Spaces  
複数の特徴量空間における投票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.

Since the performance of the first method in complex scenes 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

物体検出は人間の基本の知覚能力うちの1つで、マシン・ビジョンに重要な役割を果たす。効果的な検出方法はビデオ監視、運転補助などに役立つ。研究者は、よりよい表現モデル、よりよいクラシファイヤーあるいは解空間をより効率的に探索する方法の提案により検知方法のパフォーマンスを改善する。

本論文では、異なるチャンネルからの情報を組み合わせることにより検出方法のパフォーマンスを改善する。本論文では、新しい方法で情報融合を探索する。2種類の努力がある: 1) 外観情報と運動情報を組み合わせること 2) 同じ物体のローカルの画像特徴の中でコード化された外観と位置情報を組み合わせること。3つの検出方法が提案する。最も重要なコンポーネントは3つの方法の投票システムである。

最初に、実時間検出を主眼とした方法を作成する。階層的に、この方法の時間かかるステップは少数の実例を処理する。この方法は外観情報と運動情報を非常に効率的に組み合わせる。時間ディメンションに沿ったローカルの外観運動パターンの投票から検出を実施する。これは強健な方法およびリアルタイム方法である。

前の方法は複雑なシーンにおいて強健ではないので、2番目の方法が提案する。この方法は、運動情報を組込むために Implicit Shape Model(それはローカルな外観情報の投票システムである)を拡張する。2つのデータセットで、この方法は最先端技術の方法より強健ことが確認する。この方法は近接する物体と相似な物体をよく識別する。

2番目の方法の投票システムの高速化するために、PMS(Pyramid Match Score)を開発する。この方法は、効率のためにトレーニングと検出の間にピラミッドマッチングを行う。この方法は同じ物体のローカルの画像特徴の中でコード化された外観と位置情報を完全使用する。また、検出結果は強健である。