

Двухуровневый структурный подход к детектированию объектов

Николай Сергиевский, Александр Харламов
Кафедра интеллектуальных информационных систем и технологий
Московского Физико-Технического Института
derevly@gmail.com, kharlamov@analyst.ru

Abstract

The article describes a method for creating an effective set of bounding boxes. To solve this problem a two level approach applies. At the first level a stable structure which is a subset of the object is searching. It should be noted that the probability of detection of such structure on the background (or random location of another object) is small. At the second level the keypoints are searching. After that it is created a model of interaction between the elements of the first and second level, which is the result of a set of hypotheses about the location of the object, i.e. bounding boxes.

Keywords: object detection, object recognition, SIFT, SURF

Аннотация

В статье описывается метод построения эффективного набора детектирующих окон. Для решения этой задачи применяется двухуровневый подход. На первом уровне осуществляется поиск стабильной структуры, такого подмножества объекта, для которого вероятность обнаружения на фоне (случайном месте на сцене) мала. На втором уровне осуществляется поиск характерных точек. После этого строится модель взаимодействия элементов первого и второго уровня, результатом которой оказывается набор гипотез о расположении объекта, т.е. детектирующих окон.

1. Введение.

В направлении распознавания образов в настоящее время наблюдается большой интерес к детектированию объектов на сцене, как с практической стороны, так и со стороны фундаментальных исследований, направленных на понимание и моделирование процессов зрительного восприятия. В данный момент разработано много баз для оценки алгоритмов детектирования, например, конкурс Pascal [1]. Анализируя данные, полученные в рамках этого конкурса (таблица 1), можно заметить, что объекты с большой внутриклассовой вариативностью дают небольшой процент правильного детектирования.

Отчасти это связано с ошибками классификации (таблица 2), но так же большую роль играет выбор метода определения границ объекта и параметров оценки, например размера сканирующего окна.

2. Поиск особенностей на изображении.

В существующих подходах к автоматическому определению границ объектов [2,3] используют методы детектирования и распознавания, основанные на одноуровневых алгоритмах извлечения признаков, таких как SIFT, SURF, MSER. Несколько работ было посвящено определению распределения взаимного расположения характерных точек [4]. На первом этапе работы алгоритма из работы [4] на изображении находят все характерные точки, затем выбирают

характерные точки, для которых геометрический центр фигуры, их включающей, лежит внутри объекта, после чего характерные точки классифицируются по дескрипторам. Затем выявляется статистика взаимного распределения расстояний между точками принадлежащим разным классам. Модель строится на основании вычисления степени правдоподобия полученных распределений. При этом возникает проблема комбинаторного взрыва, поскольку максимальное количество характерных точек, которое реально можно соотнести не превышает шести. Подобные проблемы возникают и при использовании подхода, основанного на взаимодействии суперпикселей [5]. Возможно снижение комбинаторной нагрузки и повышение эффективности алгоритмов за счет использования структурного подхода.

В этом случае решаются две задачи разных уровней:

1. поиск стабильных структур;
2. поиск характерных точек.

К детектору первого уровня (поиск стабильной структуры) предъявляют требования, чтобы данная структура почти всегда принадлежала объекту, а вероятность обнаружения объекта на фоне (случайном месте на сцене) была мала. Это не значит, что данная структура не может принадлежать другим классам объектов как подмножество их собственной структуры. Такое детектирование не должно происходить в случайных точках объекта или сцены.

На втором уровне ищутся характерные точки с помощью одного из стандартных методов (Harris, DOG).

Далее должно быть принято компромиссное решение между детектированием объекта целиком и детектированием объекта по характерным точкам. Повысить качество данного компромиссного решения может анализ расстояний между классифицированными характерными точками и стабильной структурой.

3. Построение детектирующего окна.

Стабильные структуры могут нести контекстную информацию о расположении объекта. Эта информация извлекается из взаимного распределения положения структур и границ объектов (Рис 1).

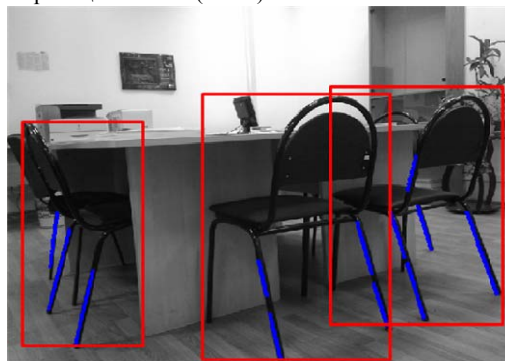


Рис 1. Пример взаимного распределения стабильных структур (ножек стульев) и рамок объектов.

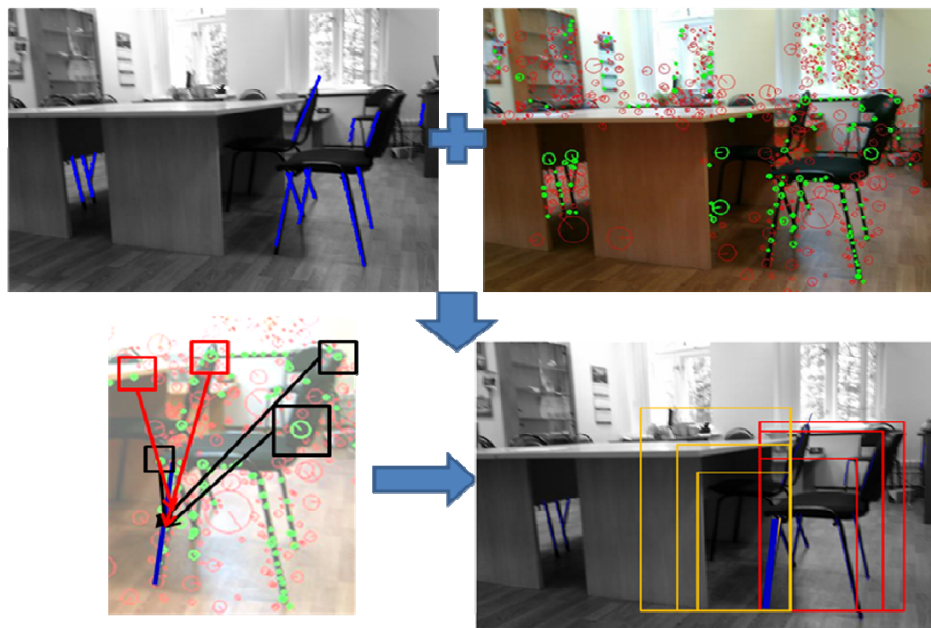


Рис 2. Диаграмма построения гипотез детектирующего окна. Верхний левый рисунок: результат поиска стабильных структур (ножек). Верхний правый: характерные точки SURF, где зелеными кругами обозначаются характерные точки, классифицированные как часть объекта. Нижний левый: взаимодействие стабильной структуры и характерных точек, черным отмечено взаимодействие внутри объекта, красным – ложное срабатывание. Нижний правый: порождаемые гипотезы.

После обучения происходит поиск стабильных структур на каждом изображении из обучающей и контрольной выборки. Далее вычисляются распределения положения объекта и детектирующего окна объекта относительно стабильной структуры, при этом учитываются размеры и положение самой стабильной структуры. Таким образом каждая структура порождает гипотезы о местоположении объекта (Рис. 2). Гипотезы уточняются за счет характерных точек и их взаимодействия со стабильными структурами. Стабильные структуры могут образовывать связи друг с другом и формировать более сильные гипотезы, для этого вычисляются распределения местоположения объекта относительно нескольких опорных точек (опорными точками могут быть как характерные точки, так и стабильные структуры). С использованием базы детектируемых объектов для каждого изображения ищутся характерные точки, после чего на основании полученного дескриптора производится обучение классификатора точки на принадлежность объекту (Рис 2). Точки, принадлежащие объекту, образуют словарь для алгоритма «мешок-слов», и далее на основании этого словаря идет повторная внутриклассовая классификация (без учителя).

4. Детектирование стульев.

Небольшой эксперимент, проведенный в рамках описанного подхода, показал точность 17.3% на проверочной выборке соревнования паскаль. Для получения результата было взято тренировочная выборка паскаль и дополнительно 600 объектов из LabelMe. После фазы генерации детектирующих окон использовалась классификация квазилинейным SVM по мешку слов, построенному на плотном SIFT (PHOW), размер мешка слов составил 300 элементов.

5. Выводы.

В работе представлен подход к детектированию сложных объектов на основе использования двух уровней поиска разнородных элементов на изображении с их взаимодействием, и построения эффективного множества границ объекта для его последующей классификации. Предлагаемый подход направлен на уменьшение вариативности результатов для подходов, основанных на взаимном распределении данных, за счет введения однонаправленных связей типа «звездочка» (элемент верхнего уровня – соотнесенные с ним элементы нижнего уровня) между разными уровнями. Исследование может быть расширено на решение задачи семантической классификации объектов путем замены характерных точек на сегменты объекта.

6. Список литературы.

- [1] <http://pascallin.ecs.soton.ac.uk/challenges/VOC/>
- [2] P. Felzenszwalb, D. McAllester, D. Ramanan. A Discriminatively Trained, Multiscale, Deformable Part Model, CVPR 2008
- [3] S.Vijayanarasimhan, K. Grauman. Efficient Region Search for Object Detection, CVPR 2011.
- [4] R. Fergus, P. Perona, R. Fergus. Object Class Recognition by Unsupervised Scale-Invariant Learning CVPR 2003
- [5] B. Fulkerson, A. Vedaldi, S Soatto. Class Segmentation and Object Localization with Superpixel Neighborhoods, CVPR 2009

Таблица 1. Часть таблицы результатов Pascal Voc2010 по детектированию объектов.

aero								dining	tv/
plane	bicycle	bird	boat	bottle	car	Cat	chair	table	monitor
50,5	24,4	17,1	13,3	10,9	32,9	36,5	5,6	6,6	27,2
54,2	48,5	15,7	19,2	29,2	43,5	41,7	16,9	26,7	40,8
53,3	55,3	19,2	21	30	46,7	41,2	20	20,7	40,3
49,1	52,4	17,8	12	30,6	32,8	37,3	17,7	27,7	38,4
46,1	52,6	13,8	15,5	28,3	44,5	26,6	17,6	16,1	-
40,4	34,7	2,7	8,4	26	33,8	17,2	11,2	14,4	35,7
37,9	33,7	2,7	6,5	25,3	33,1	15,5	10,9	12,5	34,1
52,4	54,3	13	15,6	35,1	49,1	31,8	15,5	13,5	38
56,7	39,8	16,8	12,2	13,8	36,9	47,7	12,1	26,5	41,7
58,4	39,6	18	13,3	11,1	37,8	43,9	10,3	20,8	41,9

Таблица 2. Часть таблицы результатов Pascal Voc2010 по классификации объектов.

bird	53,1	39,5	42,9	47,3	64,5	63	66,7	66,8	49,9	54,6	46,1	61,7	69,9	69,9
chair	71,1	46,5	58,7	49,1	72	66,9	72,2	71,2	59,9	60	59,4	67,5	79,3	79,4