

О.Ю. АКСЕНОВ

НТЦ «Модуль», г.Москва

e-mail: aks@module.ru

МЕТОДИКА ФОРМИРОВАНИЯ ОБУЧАЮЩИХ ВЫБОРОК ДЛЯ РАСПОЗНАЮЩЕЙ СИСТЕМЫ

Аннотация

Рассматривается задача обнаружения статических и подвижных объектов на последовательности изображений естественных сцен, поступающих от подвижной видеокамеры, наблюдающей земную поверхность и объекты на ней «сверху вниз». Обнаружение основано на использовании образа объекта. Предлагаемая методика формирования обучающей выборки для распознающей системы обеспечивает существенное снижение объема используемых изображений-примеров, минимизирует и формализует действия оператора, готовящего выборку. Работоспособность методики иллюстрируется на примере аппаратно-программного комплекса обработки видеоизображений на базе 4-х процессоров Л1879ВМ1.

O.Yu. AKSENOV

RC «Module», Moscow

e-mail: aks@module.ru

METHOD OF TRAINING SET GENERATION FOR A RECOGNITION SYSTEM

Abstract

The task of the static and non-static objects detecting on a sequence of images of real scenes is considering. Images are produced by a moving TV camera, which observes earth surface and objects on it from above. Detecting is based on the object shape analysis. The suggested method of training set generation provides essential decreasing of the used number of pattern images, minimizes and formalizes the actions of the operator who creates a training set. Working capacity of the suggested method is illustrated by a hardware-software complex based on four NM6403 (Л1879ВМ1) processors.

Один из путей обнаружения объектов на видеоизображениях естественных сцен в реальном масштабе времени базируется на принципе обнаружения–распознавания [1], использующем распознаватель *образов*¹ объектов, обучаемый на примерах. Реализация подобного обнаружителя связана, в том числе, с необходимостью формирования представительной выборки *образов* объектов для обучения распознающей системы. Ниже предлагается методика формирования такой выборки. Предлагаемая методика призвана снизить трудоемкость процедуры формирования выборки и уменьшить зависимость ее результатов от субъективного фактора – человека, осуществляющего формирование выборки.

Формируемая выборка предназначена для обучения распознающей системы комплекса обработки видеоизображений. Входная информация комплекса представляет собой последовательность полутоновых изображений – кадров, полученных как вид «сверху вниз» при помощи телевизионной камеры, осуществляющей обзор местности. Это обуславливает относительно небольшие размеры изображений обнаруживаемых объектов по сравнению с размерами кадра, а также то, что наблюдаемый ракурс объекта хоть и близок к вертикальному, но все же отличается от него. Под ракурсом в данном случае понимается угол между линией визирования объекта и вертикалью.

Схема обработки видеоизображений представлена на Рисунке 1. Обнаружение тут происходит следующим образом. По отдельному исходному изображению – кадру – формируются гипотезы объектов. Для каждой из них определяется некоторый набор признаков – *образ*. Под *образом* тут понимается набор и форма представления параметров, описывающих объект с точностью до класса. Применительно к рассматриваемой задаче *образ* характеризует видимую форму объекта и формируется по изображению объекта.

Образы гипотез объектов сравниваются с эталонными *образами* – проводится их распознавание. Если *образ* гипотезы совпадает с эталонным *образом* одного из классов объектов, принимается решение об обнаружении объекта данного класса. Обнаруженные объекты сопровождаются на последовательности изображений. При этом результаты распознавания накапливаются и уточняются.

¹ Здесь и далее *курсивом* выделены характерные термины.

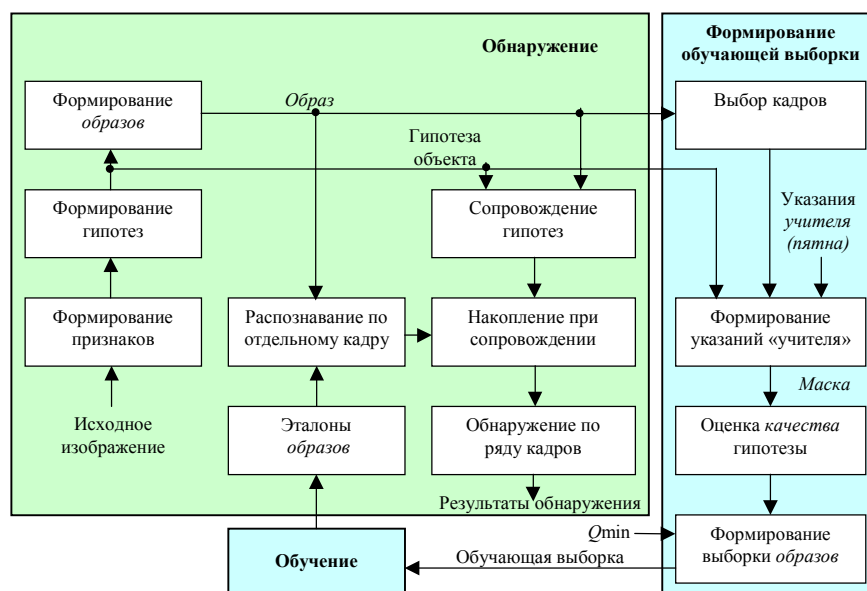


Рисунок 1. Схема обработки информации.

Обучение распознающей системы существенным образом влияет на результаты обнаружения. Обучение, в данном случае, проводится по примерам, при участии человека – *учителя*. Наиболее трудно поддающаяся формализации и автоматизации задача *учителя* – сформировать представительную выборку *образов* заданных классов объектов. Такая выборка должна быть, с одной стороны, компактной, чтобы облегчить обучение, а с другой стороны, подробной, чтобы включить в себя требуемое многообразие *образов* объектов каждого класса. Последнее требование дополнительно усложняется тем, что в рассматриваемом комплексе инвариантность результатов распознавания к ракурсу объекта достигается за счет соответствующего обучения распознающей системы.

В работе рассматриваемого обнаружителя можно выделить два шага. На первом – вне реального времени – производится обучение системы распознавания. За это отвечают блоки «Формирование обучающей выборки» и «Обучение» на Рисунке 1. Второй шаг – собственно обнаружение. Этому соответствует блок «Обнаружение» на Рисунке 1. На втором шаге обработка входных изображений ведется с темпом их поступления.

Процесс обучения системы распознавания также можно разделить на 2 части, как это и показано на Рисунке 1: формирование обучающей выборки и собственно обучение распознающей системы – генерация по обучающей выборке эталонных *образов* объектов.

Предмет настоящего рассмотрения – формирование выборок *образов* объектов. Исходная информация – последовательность входных изображений. Дополнительная информация – указания *учителя*. В качестве распознавателя используется искусственная нейронная сеть (ИНС).

Процесс формирования выборки идет в несколько этапов:

- *выбор кадров*, которые могут дать наибольший вклад в обучение;
- *формирование инструкций учителя*;
- *формирование образов* объектов и *оценка их качества* для выбранных кадров;
- собственно *формирование выборки образов*.

Выбор кадров происходит в ходе обработки последовательности исходных изображений без участия человека. При этом, также как и при обнаружении, для каждого кадра формируются гипотезы объектов, и на последовательности кадров осуществляется сопровождение каждой из них. Для каждой гипотезы на каждом кадре вычисляется ее *образ*. Критерии *выбора* текущего кадра для дальнейшего использования – формирование на этом кадре новой гипотезы объекта или существенное отличие реализовавшегося на текущем кадре *образа* одной из сопровождаемых гипотез от *образа*, характерного для данной гипотезы. Последний критерий использует Евклидово расстояние и может быть записан в следующем виде:

$$p \geq \sum_i (a_{ijk} - b_{ij})^2,$$

где: p - значение порога;

a_{ijk} - реализовавшееся на k – *ом* кадре сопровождения j – *ой* гипотезы значение i – *ой* компоненты ее *образа*;

b_{ij} - характерное для j – *ой* гипотезы значение i – *ой* компоненты ее *образа*.

Целесообразно принять $b_{ij} = a_{ij0}$, что означает использование в качестве характерного *образа* j – *ой* гипотезы ее *образа*, реализовавшегося на том кадре, на котором эта гипотеза была сформирована впервые. Результат этого этапа – переход от n кадров, составляющих исходный *сюжет* к

автоматически выбранным m кадрам – *фрагменту*. При этом $n \gg m$. Схематически это отражает Рисунок 2.

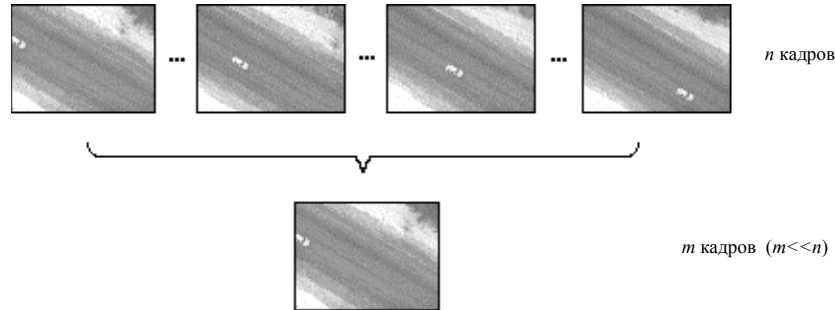


Рисунок 2. Автоматический выбор кадров.

Второй этап – *формирование инструкций учителя*. Этот этап требует участия человека, что предполагает введение человеком инструкций о классе объекта(ов), присутствующих на кадрах, составляющих *фрагмент*.

После опробования на практике различных способов введения таких инструкций был выбран следующий. Для каждого исходного кадра фрагмента создается соответствующий ему кадр – *маска*. На *маске учитель* закрашивает каждый визуально обнаруженный им объект, создавая на изображении *пятно*, закрывающее объект. Цвет *пятна* говорит о классе объекта. *Пятна*, помечающие объекты на *маске*, могут и не соответствовать гипотезам, формируемым автоматически в режиме обнаружения. Это иллюстрирует Рисунок 3 на левой части которого белым цветом в серых прямоугольниках показаны сформированные гипотезы, а на правой – соответствующая этому кадру *маска*. При этом пятно на *маске* может не вполне точно соответствовать объекту. Так в реализованной версии программного обеспечения несоответствие *пятна* объекту на величину ~20% от площади объекта практически не оказывают влияние на результат обучения. Это свойство существенно упрощает подготовку *масок*.

Маска, таким образом, содержит информацию, как о классе объекта, так и о его расположении на изображении. Если число кадров во *фрагменте* невелико, *маски* могут быть созданы вручную. Однако на практике использовалось специальное программное обеспечение, упрощающее создание *масок* для похожих кадров на основе корреляционного слежения за объектами, присутствующими на таких кадрах фрагмента.

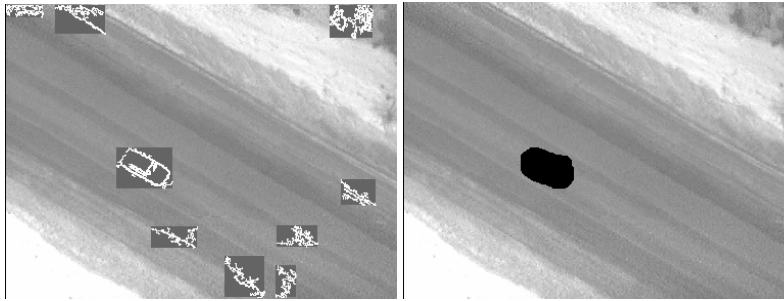


Рисунок 3. Гипотезы объектов (слева) и маска (справа).

Третий этап – *формирование образов объектов и оценка их качества* – проводится автоматически. Его суть – сравнение каждой гипотезы, сформированной по данному кадру *фрагмента*, с пятнами на *маске*, соответствующей этому кадру. При этом рассчитывается площадь перекрытия гипотезы и *пятна*. Отношение площади перекрытия гипотезы и соответствующего *пятна маски* к площади гипотезы интерпретируется как параметр – Q – *качество гипотезы* ($Q = 0 \dots 1$). При $Q = 0$ гипотеза ошибочна – она соответствует фону. При $Q > 0$ гипотеза соответствует объекту. Цвет пятна *маски* несет информацию о классе объекта. Если гипотеза перекрывает несколько пятен, то в рассмотрение принимается пятно, дающее наибольшее значение Q . Одновременно, как и в случае проведения обнаружения, для каждой гипотезы формируется образ $\{a_i\}_{jk}$. В итоге для каждого из t кадров *фрагмента* автоматически формируются l записей ($l \geq 0$). Каждая запись содержит следующую информацию:

- класс объекта, соответствующего гипотезе (фону присваивается класс «0»);
- *качество гипотезы* – Q ;
- *образ гипотезы* $\{a_i\}_{jk}$.

Рассмотренные этапы по набору исходных *сюжетов* формируют набор *записей*, в которых *образы* объектов различных классов представлены в различном количестве и с разным *качеством*. Это и обуславливает необходимость следующего этапа – *формирования выборки образов*.

На этом этапе *учитель* задает набор используемых *записей* и минимально допустимое *качество* гипотез, которые можно использовать при *формировании выборки образов*. Все дальнейшие операции проводятся автоматически. Представительство каждого класса объектов в *формируемой выборке образов* выравнивается на основании анализа гистограмм распределения *качества* гипотез. При этом отбираются *образы*, характеризующиеся наибольшим *качеством*.

Сформированная выборка образов, содержит *образы* объектов и информацию о классе объекта, которому соответствует каждый *образ*. Эта выборка используется для обучения системы распознавания. В данном случае – ИНС. В значительной мере необходимость этапа *формирования выборки образов* обусловлена свойствами используемой для распознавания нейросетевой парадигмы.

Рассмотренная методика формирования обучающих выборок была реализована в аппаратно-программном комплексе обнаружения объектов по видеоинформации. Схему обработки видеоинформации в рассматриваемом комплексе отображает Рисунок 1. Аппаратная часть рассматриваемого комплекса представляет собой устройство видеобработки VM1 [2] на базе 4 процессоров L1879VM1 [3]. Устройство VM1 выполнено в виде сборки плат, вставляемой в PCI слот персонального компьютера.

Часть алгоритмов, призванная вести обработку информации в темпе поступления кадров от видеокамеры и обозначенная на этом рисунке как «Обнаружение», реализована непосредственно на VM1. При этом такие этапы обработки, как формирование признаков и *образов*, распознавание *образов* реализованы на векторном узле используемого процессора. Часть алгоритмов комплекса, не требующая обработки информации со столь высоким темпом и обозначенная на этом рисунке как «Формирование обучающей выборки» и «Обучение», реализована на персональном компьютере.

Таблица 1 содержит некоторые количественные данные, характеризующие, как параметры, заложенные при создании комплекса, так и величины, оцененные в ходе его тестирования.

Таблица 1

Параметр	Значение
Размер кадра	384×288 пикселей
Глубина яркости входного кадра	256 уровней серого

Темп поступления и обработки кадров при обнаружении	25 Гц
Размерность <i>образа</i>	~20 элементов
Размерность представления отдельного элемента <i>образа</i>	8 бит
Число эталонов классов объектов	до 10
Объем файла эталонов <i>образов</i> для 10 классов объектов	~20 Кбайт
Сокращение числа кадров за счет <i>выбора кадров</i> (n/m)	~ 20
Число записей (l), формируемых по 1 кадру	2...20

Рисунок 4 иллюстрирует работу комплекса с использованием описанной методики формирования обучающей выборки. На представленном изображении, смонтированном из частей различных кадров объекты, обнаруженные и сопровождаемые комплексом, показаны светлыми окружностями.

Таким образом, рассмотренная методика формирования обучающих выборок обеспечивает:

- существенное прореживание исходной последовательности кадров при подготовке данных для обучения;
- формализованную процедуру формирования указаний *учителя*, снижающую влияние субъективного фактора;

Принятый подход к формированию обучающей выборки, обеспечивает, в конечном итоге, эффективное обучение распознающей системы. Использование предложенной методики при работе по изображениям естественных сцен подтверждает ее работоспособность и эффективность.

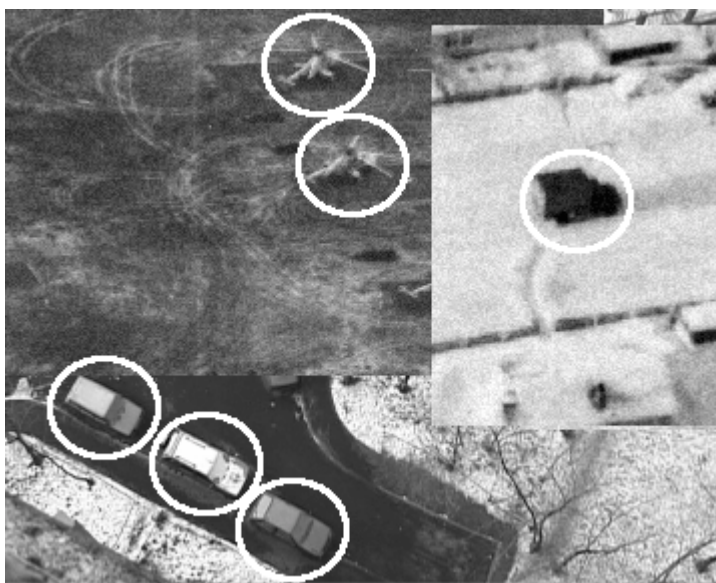


Рисунок 4. Результаты обнаружения.

Список литературы

1. Аксенов О.Ю. Обнаружение и распознавание объектов на изображениях с использованием искусственных нейросетей. // Всероссийская научно-техническая конференция “Нейроинформатика-99”. 1999. Часть 3. М.: МИФИ. С. 131-137.
2. <http://www.module.ru/ruproducts/dspmod/bm1-r.shtml>
3. <http://www.module.ru/ruproducts/proc/nm6403.shtml>