

МЕТОД ОБНАРУЖЕНИЯ ОБЪЕКТА В ВИДЕОПОТОКЕ В РЕАЛЬНОМ ВРЕМЕНИ

А.Н. Алфимцев, И.И. Лычков

*Кафедра «Информационные системы и телекоммуникации»,
ГОУ ВПО «Московский государственный технический университет
им. Н.Э. Баумана», г. Москва; alfim@iu3.bmstu.ru, lychkovi@gmail.com*

Представлена членом редколлегии профессором В.И. Коноваловым

Ключевые слова и фразы: обнаружение объекта; процедура обучения; программируемая логическая интегральная схема; распознавание образов; цифровое изображение.

Аннотация: Предложен вычислительно эффективный метод обнаружения объекта в видеопотоке с упрощенной процедурой обучения. Представлен аналитический обзор популярных методов обнаружения объектов, выделены их преимущества и недостатки. Приведены примеры работы предложенного метода на тестовой выборке изображений статического жеста открытой ладони. Рассмотрена аппаратная реализация метода в виде периферийного модуля системы на кристалле на базе программируемой логической интегральной схемы Altera Stratix 1S10.

1. Введение

Компьютерное зрение является динамично развивающимся направлением современной науки, востребованным в различных областях, начиная с интеллектуальных человеко-машинных интерфейсов, принятия решений роботами и заканчивая системами автоматического контроля на производстве. Неотъемлемой частью компьютерного зрения является распознавание образов, решающее задачу определения принадлежности входного изображения к одному из хранимых эталонных изображений объектов. При создании интеллектуальных систем также часто требуется отслеживать положение подвижных объектов в реальном времени на основе зрительной информации, полученной от видеокамеры. Располагая рядом последовательных по времени цифровых изображений, можно выделить специальную информацию об объекте и затем использовать ее для обнаружения текущего положения объекта и отслеживания его перемещений.

Цифровое изображение, получаемое видеокамерой в момент времени t и имеющее по вертикали h , а по горизонтали w пикселей, обозначим $I_t(h, w)$. *Видеопоток* – это последовательность цифровых изображений (кадров) $I_t(h, w)$, $I_{t+1}(h, w)$, ..., $I_{t+k}(h, w)$. Под прямоугольной областью интересов $Ob_i(X, Y)$ понимается множество пикселей цифрового изображения $I_t(h, w)$, очерчивающих искомым объект, содержащее X пикселей по вертикали и Y по горизонтали. *Обнаружением объекта* называется выделение области интересов $Ob_i(X, Y)$ на цифровом изображении $I_t(h, w)$ в момент времени t . Под обнаружением объекта *в реальном времени* понимается обработка видеопотока с частотой не менее 10 кадров в секунду. *Обучением* называется предварительная настройка и задание параметров метода для обнаружения интересующего объекта. Обучение может выполняться

на основе обучающей выборки, состоящей из одного или нескольких обучающих изображений. Под *обучающим цифровым изображением* понимается изображение $I_t(h, w)$ с дополнительной информацией о наличии или отсутствии на нем искомого объекта. Обучающее изображение называется *позитивным*, если оно содержит объект, который нужно обнаружить. В противном случае обучающее изображение называется *негативным*.

На сегодняшний день существуют различные методы обнаружения объектов на цифровом изображении $I_t(h, w)$: каскадный классификатор Виолы и Джонса, обобщенное преобразование Хафа, метод Капура–Винна и другие. Однако методы обнаружения объекта, имеющие высокие показатели надежности и устойчивости, требуют значительных временных и машинных ресурсов для обучения на новых объектах.

Разработаем и реализуем вычислительно эффективный метод обнаружения объекта на цифровом изображении $I_t(h, w)$, способный обрабатывать видеопоток с частотой 25 кадров в секунду и обучаться на одном позитивном изображении.

2. Обзор существующих методов

Методы обнаружения объекта на цифровом изображении $I_t(h, w)$, по способу отнесения участка изображения к области интересов $Ob_f(X, Y)$, можно разделить на две большие группы [1]: обобщающие (generative) и различающие (discriminative).

Обобщающие методы составляют абстрактное, идеализированное представление о структуре объекта на основе позитивных обучающих изображений. Это представление называют моделью, понимая под этим понятием: схему, математическую формулу, набор данных и алгоритмов, воспроизводящих в специально оговоренном виде строение или свойства объекта. При анализе нового изображения обобщающие методы оценивают, насколько точно это изображение соответствует построенной модели. Кроме анализа имеющихся изображений обобщающие методы могут использовать полученную модель для построения новых изображений объекта. При обучении обобщающих методов негативные обучающие изображения не требуются.

Наиболее популярными моделями обобщающих методов являются [2]: модель случайного поля, модель неявной формы, модель констелляции. В *модели случайного поля* изображение разбивают на небольшие участки. Каждому участку приписывается метка, которая представляет его смысловое значение, например: «вода», «небо», «земля», «объект». Кроме того, оцениваются вероятности смежного размещения двух разных меток. Совокупность назначенных меток и их вероятностей составляет случайное поле. При этом метки подбираются таким образом, чтобы получилось наиболее правдоподобное случайное поле. *Модель неявной формы* использует преобразование Хафа [3]. В процессе обучения на изображении объекта выделяют характерные точки. Для каждой точки определяют радиус-вектор, который соединяет ее с геометрическим центром объекта. При обнаружении характерной точки на исследуемом изображении соответствующий радиус-вектор указывает ожидаемую позицию центра объекта. Пикселю, расположенному в этой позиции, добавляют один голос. Пиксель, набравший наибольшее количество голосов после сканирования всего изображения, будет соответствовать наиболее вероятному положению центра объекта. *Модель констелляции* рассматривает объект в виде совокупности элементов. На исследуемом изображении выделяют характерные участки и ставят их в соответствие элементам объекта или фона. Затем с помощью перебора отыскивают наиболее правдоподобное соответствие. Характеристика методов Капура–Винна [4], Феррари [5] и Фергюса-Пероны [6], базирующихся соответственно на моделях случайного поля, неявной формы и констелляции, представлены в таблице.

Сравнение методов обнаружения объекта

Методы		Объем обучающей выборки, шт.	Время обработки одного изображения, мс	Верхняя оценка вычислительной сложности
Обобщающие	Капура–Винна	35	3000	Квадратичная
	Феррари	48	1000	
	Фергюса–Пероны	400	2500	
Различающие	Виолы–Джонса	14416	67	Линейная
	Лекуна	24300	100	
	Папагеоргиу	27159	100	

Различающие методы используют специальную функцию-классификатор для обнаружения объекта на цифровом изображении $I_d(h, w)$. В процессе обучения они выделяют различия между позитивными и негативными обучающими изображениями. На основе этих различий они подбирают параметры функции-классификатора. Полученная функция применяется для разделения новых входных изображений на изображения, содержащие и не содержащие область интересов $Ob_d(X, Y)$. Характеристики наиболее популярных различающих методов Виолы–Джонса [7], Лекуна [8] и Папагеоргиу [9] представлены в таблице.

Таким образом, различающие методы сосредоточены на различиях между позитивными и негативными изображениями конкретной обучающей выборки, в то время как обобщающие методы пытаются воссоздать структуру самого объекта по его изображениям. Поэтому при обучении обобщающих методов негативные изображения не требуются, а необходимое количество позитивных изображений существенно меньше, чем при обучении различающих методов. Вследствие чего для упрощенного обучения следует использовать подход обобщающих методов.

При выборе модели обобщающего метода будем руководствоваться следующими замечаниями. Модель случайного поля требует больших вычислительных затрат, что неприемлемо для обработки видеопотока в реальном времени. Учет возможных вариаций формы объекта в модели неявной формы требует значительного увеличения количества обучающих изображений. Но модель констелляции изначально представляет объект в виде совокупности отдельных элементов. Именно в этой модели проще учитывать различные искажения геометрической формы объекта за счет изменения взаимного положения его элементов.

Предлагаемый метод будет основываться на модели констелляции, которая в качестве элементов объекта использует линии границ, выделенные на изображении объекта, совокупность которых называется *каркасом изображения*.

3. Описание предлагаемого метода

Рассмотрим метод обнаружения объекта в видеопотоке на примере нахождения статического жеста открытой ладони [10]. Метод включает в себя упрощенную процедуру обучения и процедуру обнаружения объекта.

Упрощенная процедура обучения использует эталонное изображение, содержащее искомый объект крупным планом на контрастном фоне, и состоит из следующих шагов.

1. Построение эталонного каркаса с помощью процедуры выделения границ.
2. Задание параметров метода.

Процедура обнаружения объекта получает на вход полутоновое изображение $I_t(h, w)$, эталонный каркас и параметры, заданные при обучении, и состоит из следующих основных шагов:

- 1) построение исследуемого каркаса с помощью процедуры выделения границ;
- 2) поиск наилучшего совмещения исследуемого и эталонного каркасов;
- 3) принятие решения по обнаружению объекта путем сравнения количества совпавших линий в наилучшем совмещении с заданным порогом;
- 4) обрисовка области интересов $Ob_t(X, Y)$.

3.1. Выделение границ

Как на этапе обучения, так и в процессе работы цифровые изображения подвергаются предварительной обработке. При этом извлекается информация о расположении границ между темными и светлыми областями изображения.

Процедура выделения границ строит каркас цифрового изображения $I_t(h, w)$ и состоит из следующих шагов:

- 1) обработка цифрового изображения $I_t(h, w)$ фильтром Хаара [7];
- 2) выделение точек границ на обработанном изображении $I_t(h, w)$;
- 3) объединение найденных точек в линии;
- 4) сохранение интересующих свойств линий в каркасе изображения.

На первом шаге процедуры входное изображение (рис. 1, а) обрабатывается фильтром Хаара (рис. 4). В результате получается обработанное изображение (рис. 1, б) того же размера, отражающее величину местных перепадов яркости исходного изображения. Для вычисления пикселя обработанного изображения необходимо найти соответствующий пиксель на исходном изображении, очертить прямоугольную окрестность этого пикселя и разделить эту окрестность пополам (рис. 2). Значение пикселя обработанного изображения будет равно алгебраической сумме пикселей выделенной окрестности исходного изображения, в которой пиксели правой части берутся со знаком «плюс», а пиксели левой части – со знаком «минус». Если использовать вычислительный прием интегрального изображения [7], то для расчета одного пикселя обработанного изображения достаточно выполнить всего пять операций сложения, независимо от количества пикселей в его прямоугольной окрестности. Это существенно повышает быстродействие метода, что важно для обработки видеопотока.

На третьем шаге процедуры найденные граничные точки объединяются в линии. Две точки относятся к одной линии, если расстояние между ними не превышает заданного расстояния, которое назовем *радиусом связности*. При этом точки максимумов никогда не объединяются с точками минимумов. Таким образом, на этом шаге выделяются вертикальные линии границ переходов яркости с темного на светлое и со светлого на темное (рис. 1, в). Для выделения горизонтальных границ достаточно развернуть прямоугольную окрестность фильтра на 90° и сканировать производное изображение в поисках экстремумов уже в вертикальном направлении.

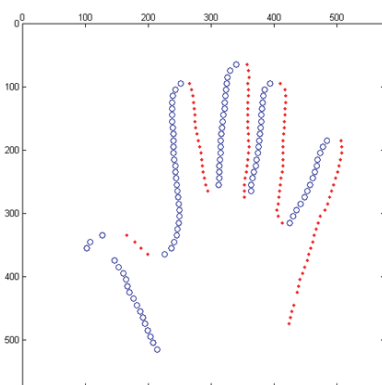
На заключительном шаге процедуры параметры линий границ, выделенных на изображении, сохраняются в каркасе этого изображения. Каждая линия каркаса задается координатами крайних точек и типом границы, которой она соответствует.



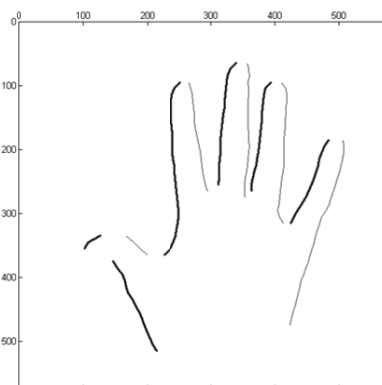
а)



б)



в)



г)

Рис. 1. Выделение вертикальных границ:
a – исходное изображение; *б* – обработанное изображение;
в – точки границ; *г* – линии границ

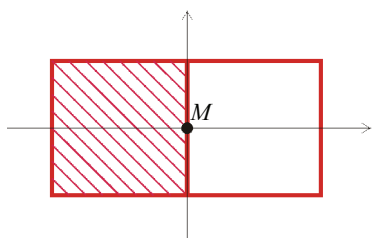


Рис. 2. Прямоугольная окрестность некоторой точки *M*

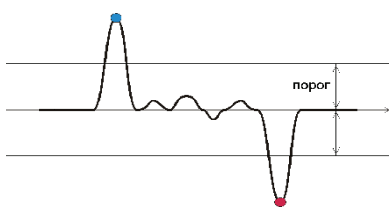


Рис. 3. Применение порога чувствительности

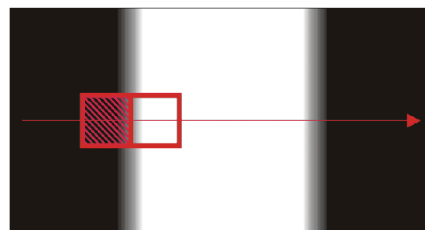


Рис. 4. Применение фильтра Хаара

3.2. Обучение

Для работы метода необходимо задать значения восьми параметров. Первые три параметра используются в процедуре выделения границ: размер прямоугольной окрестности, порог чувствительности (рис. 3) и радиус связности. Еще четыре параметра используются при наложении каркасов: характерная линия, радиус зон поиска, коэффициенты допустимого растяжения и сжатия линии. Восьмым параметром является количество линий исследуемого каркаса, которые должны совпасть с линиями эталонного каркаса для принятия решения по обнаружению объекта. Рассмотрим упрощенную процедуру обучения.

На первом шаге процедуры обучения с помощью описанной процедуры выделения границ из эталонного изображения получается эталонный каркас. Чтобы не загромождать рисунки, будем рассматривать только вертикальные границы переходов с темного на светлое, отмеченные жирным на рис. 1, 2. Тогда эталонный каркас будет иметь вид, представленный на рис. 5, б.

На втором шаге процедуры обучения задаются значения параметров метода. Рассмотрим порядок выбора наиболее важных параметров: характерной линии и радиуса зон поиска. Характерная линия эталонного каркаса выбирается исходя из наибольшей длины (рис. 6, а). По ней будет выполняться совмещение каркасов. Положение центров остальных линий эталонного каркаса определяется относительно этой линии. Зоны поиска позволяют учесть возможные отклонения центров линий от расчетных точек при изменении положения пальцев и пропорций руки. Радиус зон поиска принимается равным четверти длины характерной линии.

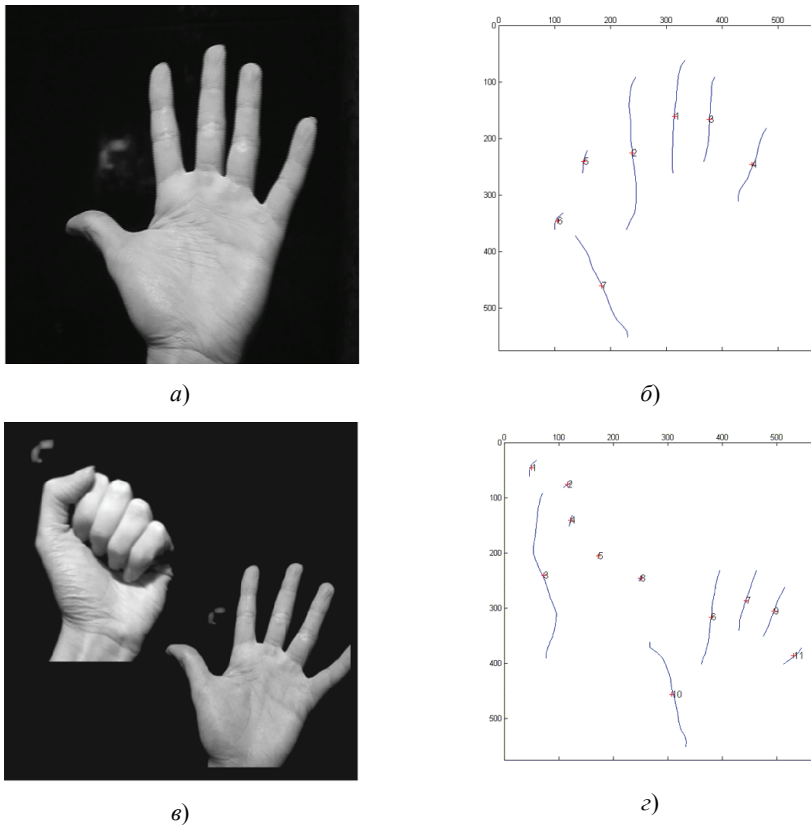


Рис. 5. Каркасы изображений:
а – эталонное изображение; б – эталонный каркас;
в – исследуемое изображение; г – исследуемый каркас

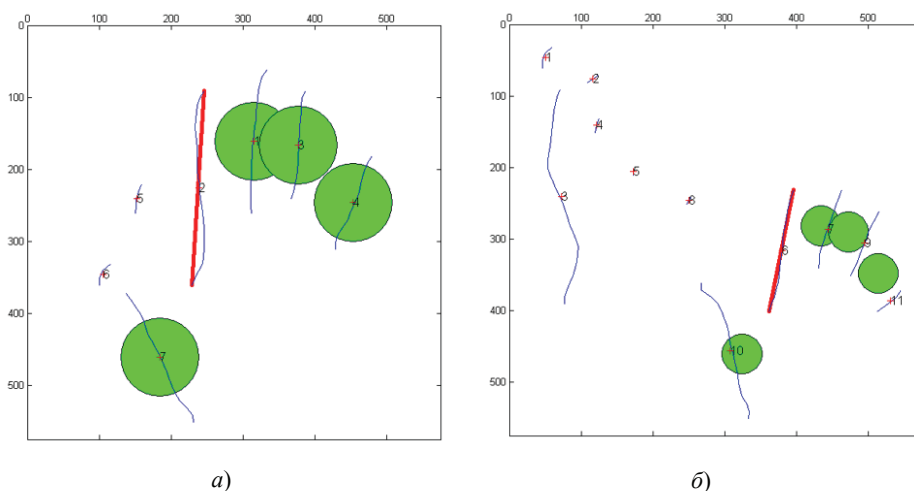


Рис. 6. Характерная линия и зоны поиска:
 а – на эталонном каркасе; б – после наложения на исследуемый каркас

3.3. Обнаружение объекта

На первом шаге на вход подается исследуемое изображение $I_i(h, w)$ и из него получается исследуемый каркас (рис. 5, з). На втором шаге по очереди перебираются линии исследуемого каркаса. На каждой итерации эталонный каркас масштабируется и совмещается с исследуемым каркасом так, чтобы его характерная линия совпала с выбранной линией исследуемого каркаса. При этом круговые зоны поиска преобразуются вместе с эталонным каркасом и переносятся на исследуемый каркас (рис. 6, б). Затем подсчитывается количество линий исследуемого каркаса, центры которых попали в соответствующие зоны поиска. В результате перебора находится наилучшее совмещение каркасов, которое дает наибольшее число совпавших линий. На третьем шаге количество совпадений сравнивается с установленным порогом, принимается решение по обнаружению области интересов $Ob_f(X, Y)$ на цифровом изображении $I_i(h, w)$, и выполняется обрисовка захваченного объекта.

3.4. Оценка вычислительной сложности

Проведем оценку вычислительной сложности разработанного метода обнаружения объекта на цифровых изображениях $I_i(h, w)$.

Пусть n – количество линий в исследуемом каркасе; m – количество линий в эталонном каркасе. Исходя из количества повторений циклов (рис. 7), верхняя оценка вычислительной сложности будет эквивалентна сумме двух бесконечно больших величин: $O(nhw) + O(mn^2)$. Количество линий каркаса не превышает количества пикселей изображения, поэтому $O(n) \sim O(hw)$. Количество m линий эталонного каркаса определяется на этапе обучения и может быть принято за константу на этапе обнаружения объекта. Преобразуем выражение $O(hwn) + O(n^2) \sim O(hw \cdot hw) + O(h^2w^2) \sim O(hw \cdot hw) \sim O(nN)$, где $N = hw$ – количество пикселей исследуемого изображения $I_i(h, w)$. Таким образом, метод имеет линейную оценку сложности, что позволяет проводить обнаружение объекта в реальном времени.

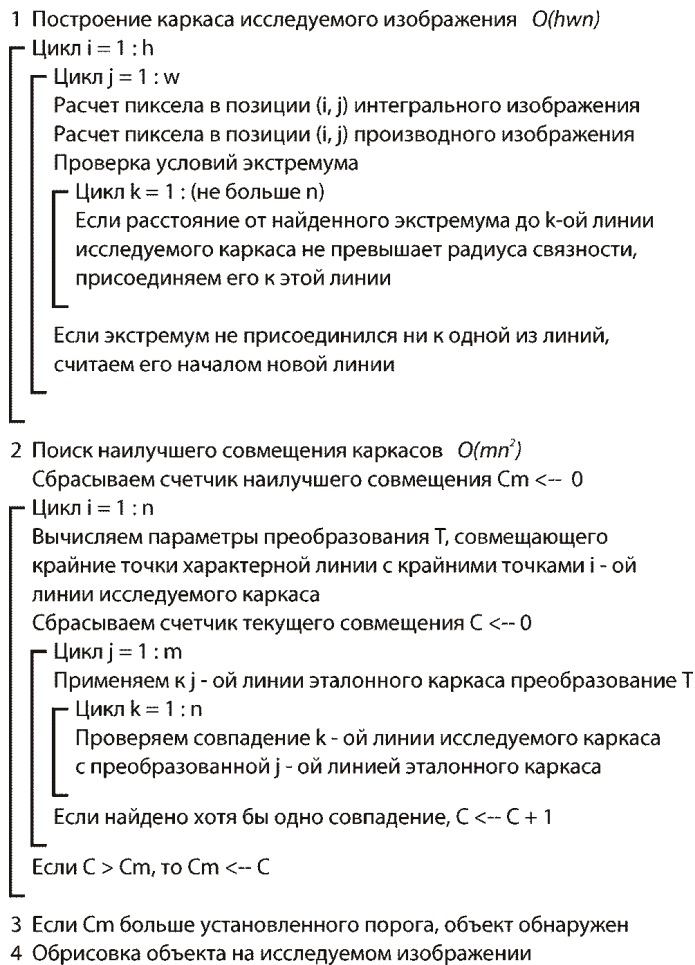


Рис. 7. Структура программной реализации метода в псевдокоде

4. Результаты экспериментов

Метод реализован программно и протестирован на стандартном персональном компьютере с процессором Intel Pentium IV 3,0 GHz видеокамерой Logitech 5000pro, передающей 25 цифровых изображений в секунду с разрешением 640×480 пикселей. Обучающая выборка состояла из одного изображения (см. рис. 1, а). Тестовая выборка состояла из 52 позитивных и 48 негативных изображений. Примеры позитивных тестовых изображений представлены на рис. 8. Освещение на месте проведения экспериментов имело следующие характеристики: освещенность 100...250 лк, коэффициент естественной освещенности 1,0; коэффициент пульсации 20 %.

Позитивное тестовое изображение считалось правильно обработанным, если метод обнаруживал объект на цифровом изображении. Негативное тестовое изображение считалось правильно обработанным, если метод не обнаруживал объект на цифровом изображении. Точность обнаружения рассчитывалась как отношение количества правильно обработанных изображений к общему количеству изображений в тестовой выборке. Точность обнаружения области интересов $Ob_f(X, Y)$ на цифровом изображении $I_f(h, w)$ составила 89 %. Среднее время обработки одного цифрового изображения 34 мс.



Рис. 8. Результаты обработки позитивных тестовых изображений

5. Аппаратная реализация метода

Компактные устройства типа смартфонов, карманных компьютеров и цифровых фотоаппаратов в последнее время стали товарами массового спроса. Аппаратная реализация предложенного метода позволит переносить его на компактные устройства, имеющие низкую тактовую частоту.

Метод был реализован аппаратно в виде периферийного модуля системы на кристалле на базе программируемой логической интегральной схемы (ПЛИС) Altera Stratix 1S10. Структурно модуль обнаружения объекта состоит из блока построения каркаса и модуля совмещения каркасов (рис. 9).

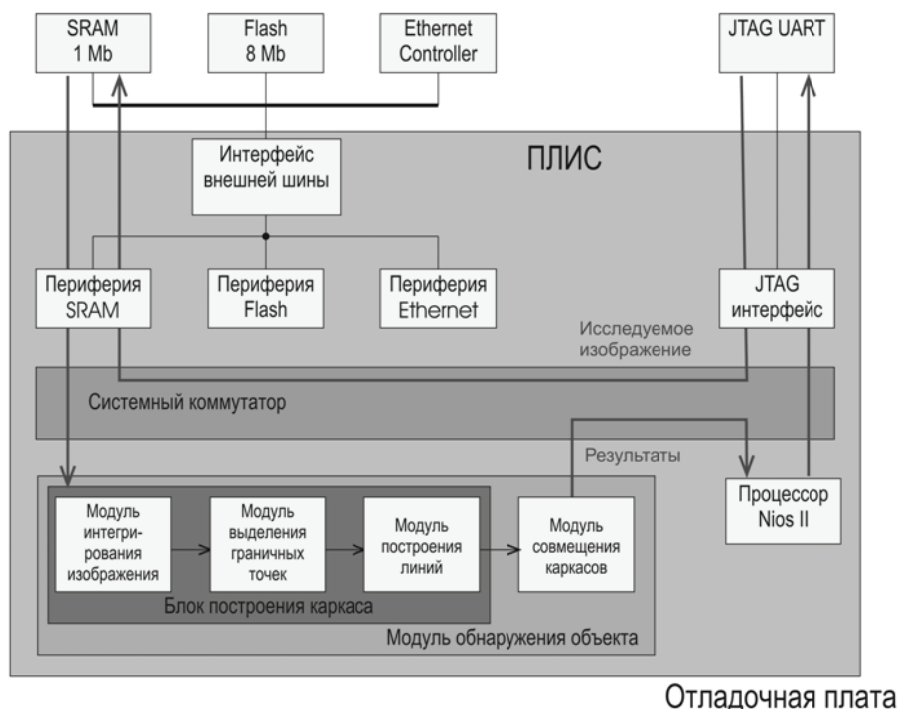


Рис. 9. Функциональная схема системы на кристалле

Обработка изображения в блоке построения каркаса организована по потоковому принципу. Изображение передается на вход блока фрагментами по несколько пикселей, которые обрабатываются на конвейере. После передачи всех фрагментов рассчитывается каркас изображения. Блок построения каркаса включает модули интегрирования, выделения граничных точек и построения линий, каждый из которых реализует отдельный этап конвейера. Ускорение вычислений достигается трехэтапной конвейерной обработкой и одновременным продвижением пикселей по нескольким параллельным каналам обработки. Модуль совмещения каркасов содержит матрицу детекторов совпадения, которая ускоряет подсчет количества совпадающих линий каркасов. Разработанные модули имеют масштабируемую архитектуру, то есть допускают наращивание количества каналов параллельной обработки без существенных изменений во внутренней структуре модулей.

Испытания аппаратной реализации проводились на отладочной плате Nios Development Board Stratix Edition, оснащенной ПЛИС Altera Stratix EP1S10F780C6 с тактовой частотой 50 МГц. Восьмибитное полутоновое изображение размером 640×480 пикселей передавалось на вход блока построения каркаса фрагментами по 8 пикселей за такт. Объем блока составил 2534 логических элемента, а время построения каркаса изображения – 1,04 мс против 34 мс для программной реализации метода на процессоре Intel Pentium IV 3,0 GHz.

6. Заключение

Рассмотрен метод обнаружения объекта на цифровых изображениях с упрощенной процедурой обучения, способный обрабатывать видеопоток в реальном времени (25 кадров в секунду). Для обучения метода достаточно предоставить единственное изображение объекта и задать значения восьми параметров. В сравнении с существующими методами обнаружения, предлагаемый метод демонстрирует приемлемую точность (89 %) при упрощенном обучении, а также обладает инвариантностью к повороту и масштабированию объекта.

Аппаратная реализация метода позволяет переносить его на компактные устройства с низкой тактовой частотой.

В дальнейшем планируется развить процедуру обучения метода путем автоматического задания параметров, используя агрегирование статистического и экспертного подходов анализа эталонного изображения. Кроме того, планируется дополнить аппаратную реализацию метода функциями декодирования видеопотока, поступающего с видекамеры по интерфейсу S-Video, и преобразования видеопотока для отображения на внешнем дисплее.

Работа выполнена при финансовой поддержке гранта Президента РФ МК2010 (проект № МК-3013.2010.9).

Список литературы

1. Roth, P.M. Survey of Appearance-Based Methods for Object Recognition / P.M. Roth, M. Winter // Technical Report ICG-TR-01/08, Institute for Computer Graphics and Vision, Graz University of Technology, Austria, January, 2008. – 68 p.
2. Murphy, K.P. Models for Generic Visual Object Detection / K.P. Murphy // Technical report, Department of Computer Science, University of British Columbia, Vancouver, Canada, May, 2005. – 8 p.

3. Greenspan, M. Efficient Tracking with the Bounded Hough Transform / M. Greenspan, L. Shang, P. Jasiobedzki // Proceedings of Computer Vision and Pattern Recognition. – 2004. – Vol. 1. – P. 520–527.
 4. Kapoor, A. Located Hidden Random Fields: Learning Discriminative Parts for Object Detection / A. Kapoor, J. Winn // Proceedings of the European Conference on Computer Vision. – 2006. – Vol. 3. – P. 302–315.
 5. Ferrari, V. From Images to Shape Models for Object Detection / V. Ferrari, F. Jurie, C. Schmid // Intl. J. Computer Vision. – 2010. – Vol. 87(3). – P. 284–303.
 6. Fergus, R. Object Class Recognition by Unsupervised Scale-Invariant Learning / R. Fergus, P. Perona, A. Zisserman // Proceedings of Computer Vision and Pattern Recognition. – 2003. – Vol. 2, Sec. II. – P. 264–271.
 7. Viola, P. Robust Real-Time Object Detection / P. Viola, M. Jones // Intl. J. Computer Vision. – 2004. – Vol. 57(2). – P. 137–154.
 8. LeCun, Y. Learning Methods for Generic Object Recognition with Invariance to Pose and Lighting / Y. LeCun, F.-J. Huang, L. Bottou // Proceedings of Computer Vision and Pattern Recognition. – 2004. – Vol. 2. – P. 97–104.
 9. Papageorgiou, C. A Trainable System for Object Detection / C. Papageorgiou, T. Poggio // Intl. J. Computer Vision. – 2000. – Vol. 38(1). – P. 15–33.
 10. Девятков, В.В. Распознавание манипулятивных жестов / В.В. Девятков, А.Н. Алфимцев // Вест. Моск. гос. техн. ун-та им. Н.Э. Баумана. Сер. Приборостроение. – 2007. – Т. 68, № 3. – С. 56–75.
-

The Technique of Real-Time Object Detection in the Video Stream

A.N. Alfimtsev, I.I. Lychkov

*Department “Information Systems and Telecommunications”,
Bauman Moscow State Technical University;
alfim@iu3.bmstu.ru, lychkovi@gmail.com*

Key words and phrases: digital image; FPGA; object recognition; object detection; training procedure.

Abstract: The paper presents a computationally effective method for object detection on digital images with simplified training procedure. Analytical survey of popular existing methods for object detection is presented; their advantages and disadvantages are identified. Object detection examples for test images containing open palm static gesture are presented. The hardware implementation of the method as a peripheral unit of the system on the programmable chip FPGA Altera Stratix 1S10 is also considered.

Методы обнаружения объектов в видеопотоке в реальном времени

Zusammenfassung: Es ist die recheneffektive Methode des Objektentdeckens im Videostrom mit der vereinfachten Prozedure der Ausbildung vorgeschlagen. Es ist die analytische Übersicht der populären Methoden des Objektenentdeckens dargelegt, es

sind ihre Vorteile und Nachteile ausgezeichnet. Es sind die Beispiele des Funktionierens der vorgeschlagenen Methode an der Testenauswahl der Darstellungen der statischen Geste der offenen Hand angeführt. Es ist die Apparatenrealisierung der Methode als peripherischer Modul des Systems am Kristall auf Grund von Altera Stratix 1S10 betrachtet.

Méthode de la révélation de l'objet dans un flux vidéo dans le temps réel

Résumé: Est proposée une méthode efficace par le calcul de la révélation de l'objet dans un flux vidéo avec une procédure simplifiée de l'apprentissage. Est présentée une revue des méthodes de la révélation des objets, sont montrés leurs plus et leurs minus. Sont cités les fonctionnements de la méthode sur un extrait des textes des images d'un geste statique d'une paume ouverte. Est examinée la réalisation matérielle de la méthode en vue d'un module périphérique du système à la base de Altera Stratix 1S10 FPGA.

Авторы: *Алфимцев Александр Николаевич* – кандидат технических наук, доцент кафедры «Информационные системы и телекоммуникации»; *Лычков Игорь Игоревич* – аспирант кафедры «Информационные системы и телекоммуникации», ГОУ ВПО «Московский государственный технический университет им. Н.Э. Баумана», г. Москва.

Рецензент: *Павлов Юрий Николаевич* – доктор технических наук, профессор кафедры «Информационные системы и телекоммуникации», ГОУ ВПО «Московский государственный технический университет им. Н.Э. Баумана», г. Москва.
