

Метод построения HG-LBP дескриптора на основе гистограмм градиентов для детектирования пешеходов*

Разумов¹ И. О, Прошутинский¹ Д. А, Гнеушев^{1,2} А. Н.

razzumoff@gmail.com, gneushev@ccas.ru, dmitriy.proshutinskiy@phystech.edu

¹Московский физико-технический институт, Россия, г. Долгопрудный, Институтский пер., 9

²ФИЦ «Информатика и управление» РАН, Россия, г. Москва, ул. Вавилова, 44/2

Рассматривается подход обобщения алгоритма LBP (Local Binary Patterns) для детектирования пешеходов на изображении на основе учета вклада каждого LBP-кода в пространственную гистограмму дескриптора. Обычный LBP дескриптор включает гистограмму только количества вхождений LBP-кода в локальной области изображения и не учитывает степень выраженности каждого LBP-кода. В статье предлагается использовать вес каждого LBP-кода при построении локальной гистограммы дескриптора. Рассматриваются два варианта оценки веса LBP: на основе модуля градиента и на основе суммы модулей производных по восьми направлениям в точке подсчета LBP-кода для формирования нового дескриптора HG-LPB (Histogram Gradient LBP). Учет локальной выраженности LBP-кода позволяет учитывать больше информации о локальных особенностях изображения объекта в HG-LPB дескрипторе и улучшить качество работы детектора. В статье представлены результаты сравнения работы базовых алгоритмов LBP, HOG-LBP и предлагаемых обобщений HG-LPB на базе изображений пешеходов INRIA. **Ключевые**

слова: *LPB, HOG, гистограмм ориентированных градиентов, Локальные бинарные шаблоны, детектирование пешехода.*

1 Введение

Автоматическое детектирование и распознавание объектов на изображениях является одной из основных задач компьютерного зрения. Задача локализации человека на видео-изображениях широко востребована в таких областях, как мониторинг и анализ дорожных ситуаций, обнаружение дорожно-транспортных происшествий, контроль за соблюдением правил дорожного движения, системы безопасности и следящие системы, беспилотные автомобили, робототехника, системы помощи водителю.

Основные сложности обнаружения человека на изображении связаны с несколькими причинами: изменение структуры изображения вследствие движения человека, неравномерная освещенность изображения, большая вариабельность изображений человека из-за разных ракурсов (поз, размеров, углов поворота), частичные перекрытия фигуры человека другими объектами.

Как правило, задача детектирования объекта разделяется на две подзадачи: выделение характерных свойств изображения объекта и бинарная классификация. Характерные свойства изображения объекта - это набор признаков, приближенно описывающий интересующий объект. Выбранный набор признаков является важнейшим фактором, влияющим на качество классификации и ее устойчивость. Чем лучше признаки обладают разделительной способностью, тем проще устроено признаковое пространство, и работа классификатора наиболее эффективна.

*Работа выполнена при финансовой поддержке РФФИ, проект № 00-00-00000. Научный руководитель: Матвеев И. А. Задачу поставил и консультировал: Гнеушев А. Н.

Признаки можно разбить на два класса: локальные и интегральные. Преимуществом локальных признаков является их универсальность, инвариантность по отношению к неравномерным изменениям яркости и освещенности, но они не уникальны. В тоже время интегральные признаки, характеризующие изображение объекта в целом, не устойчивы к изменению структуры объекта и сложным условиям освещения. Существует комбинированный подход - использование локальных признаков в качестве элементов интегрального описания, когда искомый объект моделируется набором областей, каждая из которых характеризуется своим набором признаков – локальным текстурным дескриптором. Совокупность таких дескрипторов характеризует объект в целом.

Вид локальных признаков определяется характерной структурой объекта на изображении. Изображение человека может быть представлено совокупностью контуров, силуэтов частей тела, которые являются контурными признаками и представляются на изображении как максимальные перепады значений яркости [1]. Для описания локальной контурной структуры изображения пешехода широко применяются подходы основанные на Гистограмме Ориентаций Градиентов (HOG) [4], на Локальных Бинарных Шаблонах [4] и сверточных нейронных сетях (CNN).

Использование сверточных нейронных сетей являются довольно эффективным подходом для одновременного выделения множества локальных признаков и их классификации. В CNN процедура выделения признаков осуществляется в начальных слоях, структура признаков формируется автоматически в процессе его обучения и определяется моделью и архитектурой сети. Чем больше признаков необходимо использовать для характеристики целевых объектов, тем больше параметров требуется для задания модели сети, тем она вычислительно сложнее и требует больше вычислительных ресурсов и объема обучающей выборки изображений. Это связано с тем, что в CNN признаковая модель объекта формируется на основе только той информации, которая содержится в обучающей базе и общих регуляризационных ограничениях. Однако использование априорной, более детальной экспертной информации о характере конкретного класса объектов на изображении для построения пространства признаков и получения дескрипторов позволяет существенно уменьшить объем вычислений при той же точности работы системы. Таким образом, использование дескрипторов, построенных на методах HOG и LPB, остаются актуальными для вычислительно слабых платформ.

Дескрипторы HOG и LBP обладают рядом недостатков, каждый из них учитывает не всю полезную информацию о структуре изображения. Одним из подходов для усложнения пространства признаков является объединение этих дескрипторов в один HOG-LPB [5]. Однако простое объединение не всегда эффективно, есть подходы модификации самих дескрипторов с использованием полезных свойств друг друга. Так, в работе [3] предлагается обобщить HOG дескриптор в HAN-AR и использовать информацию о положении контурных признаков. Обычный LPB дескриптор [6] включает гистограмму только количества вхождений LBP-кода в локальной области изображения и не учитывает степень выраженности каждого LPB-кода, однако подобная выраженность контурных признаков используется в гистограммах HOG дескриптора.

В статье предлагается подход обобщения LPB дескриптора на основе использования веса каждого LBP-кода при построении локальной гистограммы дескриптора. Предлагается два варианта оценки веса LPB: на основе модуля градиента и на основе суммы модулей производных по восьми направлениям в точке подсчета LBP-кода для формирования нового дескриптора HG-LPB (Histogram Gradient LBP). Учет локальной выра-

женности LBP-кода позволяет учитывать больше информации о локальных особенностях изображения объекта в HG-LBP дескрипторе и улучшить качество работы детектора.

2 Существующие подходы

2.1 Используемые модели

Local Binary Pattern (LBP): LBP был представлен в 1994 году [8]. Однако, модель спектра текстуры LBP была предложена еще ранее в 1990 году [9]. Оператор LBP был применен во многих областях после исследований Ojala и Pietikäinen в области подхода с несколькими разрешениями в 2002 году [10], включая идентификацию текста [11] и распознавание лиц [5]. Wang объединил LBP с дескриптором гистограммы направленных градиентов (HOG) для повышения эффективности обнаружения в [12]. К данной статье мы ещё вернёмся. Классический оператор LBP должен следовать определенному алгоритму; Эти действия суммированы на 1 [7]. Есть ряд попыток улучшить метод LBP. Одна недавняя версия находит значения интенсивности точек в круговой окрестности, рассматривая значения, которые имеют маленькие круговые окрестности вокруг центрального пикселя [13].

Histogram of Oriented Gradients (HOG): HOG – дескриптор, основанный на подсчете количества направлений градиента в локальных областях изображения. Традиционный HOG делит изображение на разные области, именуемые ячейками, и вычисляет гистограммы направлений градиентов или направлений краев для пикселей [13]. HOG широко применяется в областях распознавания объектов как распознавание лиц [14]. Процесс вычисления HOG объясняется на 4 [15].

Вектор опорных векторов (SVM): SVM - это алгоритм обучения с учителем, разработанный для бинарных задач классификации с возможным расширением до многоклассовых задач [23]. Кортес и Вапник впервые представили SVM в 1995 году, основная идея которого заключалась в том, что каждый объект данных представляется вектором в k -мерном пространстве и принадлежит одному из двух классов. После ищется такая гиперплоскость размерности $(p-1)$, что расстояние до ближайших к гиперплоскости точек из разных классов максимально. Такая гиперплоскость называется оптимальной. Позже Кортес и Вапник ввели понятие мягких границ, суть которых заключается в возможности классификатора допускать ошибки в случае линейной неразделимости выборки, чтобы в конечном итоге всё таки построить разделяющую гиперплоскость [24]. Некоторыми преимуществами SVM являются обобщение бинарных и регрессионных форм и упрощение обозначений [23]. В SVM можно использовать одно из несколько ядер: полиномиальное, линейное, радиальная базисная функция Гаусса(RBF) [25], [26] и т.д.



Рис. 1 Схема работы алгоритма LBP [12]

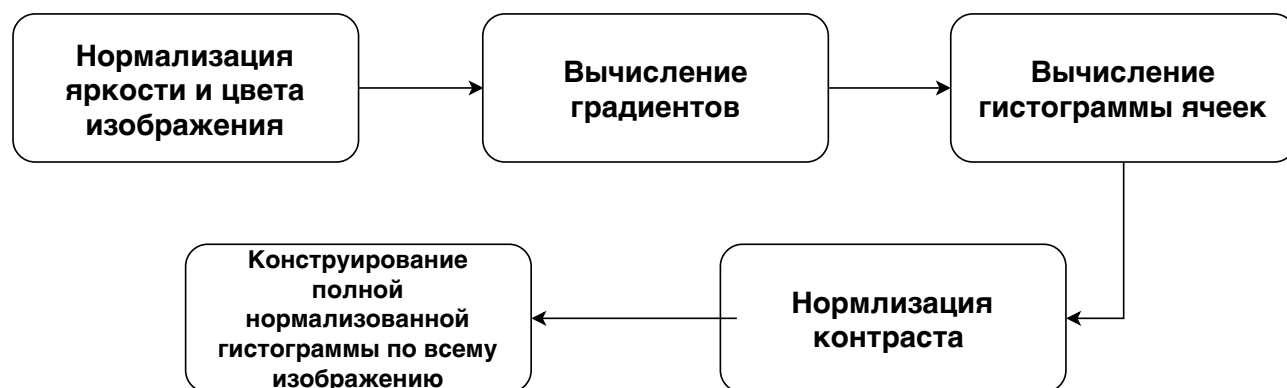


Рис. 2 Схема работы алгоритма HOG

2.2 Референтные работы

Близкой к теме данной работы является статья «An HOG-LBP Human Detector with Partial Occlusion Handling» Xiaoyu Wang [5]. В данной статье комбинируются гистограммы направленных градиентов (HOG) и локальные бинарные шаблоны (LBP) для получения набора признаков. Это делается для улучшенной точности определения человека на изображении даже в ситуации с частично закрытой фигурой человека. Процедура обнаружения человека, основанная на функции HOG-LBP, показана на 3. Похожую схему мы используем в нашей работе.

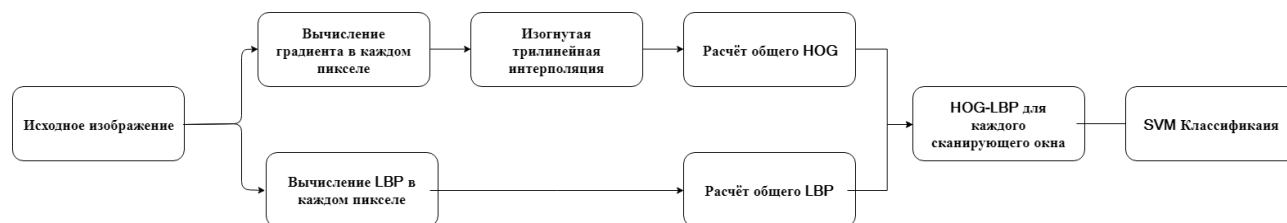


Рис. 3 Алгоритм HOG-LBP дескриптора (без заграждений изображения)

3 Постановка задачи

Рассмотрим структуру интегрального вектора признаков $\mathbf{g}(f) = (g_1, \dots, g_n)^T$ НОГ дескриптора [4] из признакового пространства G для входного изображения f , где n — размерность пространства признаков. В соответствии со схемой НОГ, изображение $f(x, y)$, где x, y — координаты точки изображения размером $X \times Y$, разбивается на смежные области — квадратные ячейки размером $K \times K$ пикселей, $I = X/K$ ячеек по горизонтали и $J = Y/K$ ячеек по вертикали. В каждой ячейке (i, j) , где i, j — индексы ячейки по вертикали и горизонтали соответственно, выделяются локальные признаки, которые определяются дескриптором ячейки — вектором $\mathbf{u}_{i,j} = (u_1, \dots, u_l)^T$, где l — размерность вектора. Четыре смежные ячейки объединяются в пересекающиеся блоки. Таким образом, каждый блок содержит $4K \times 4K$ пикселей и имеет общие ячейки с соседними блоками. Дескриптор блока определяется объединением дескрипторов собственных ячеек, вектором $\mathbf{v}_{i,j} = \mathbf{u}_{i,j} \cup \mathbf{u}_{i+1,j} \cup \mathbf{u}_{i,j+1} \cup \mathbf{u}_{i+1,j+1}$, где i, j — индексы блока. Общее количество блоков — $(I-1)(J-1)$. Под операцией объединения \cup двух векторов \mathbf{u}_1 и \mathbf{u}_2 с размерностью l_1 и l_2 соответственно будем понимать результирующий вектор из пространства размерности $l_1 + l_2$, первые l_1 компонент которого являются компонентами вектора первого аргумента, последние l_2 — компонентами вектора второго аргумента.

Дескриптор блока нормируется с помощью одной из двух норм: L_2 нормы

$$\tilde{\mathbf{v}} = N_{L_2}(\mathbf{v}) = \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \varepsilon^2}}$$

либо L_2 -hys нормы [4]

$$\tilde{\mathbf{v}} = N_{L_2\text{-hys}}(\mathbf{v}) = N_{L_2}(\min(N_{L_2}(\mathbf{v}), h)),$$

где операция \min применяется покомпонентно к вектору первого аргумента; h — пороговое значение, которое используется для ограничения значений компонент вектора в операции \min . В данной работе, как и в работе [4], в качестве порогового значения используется $h = 0,2$.

Интегральный дескриптор определяется объединением всех блочных дескрипторов, т. е. вектором

$$\mathbf{g} = \bigcup_j \bigcup_i \tilde{\mathbf{v}}_{i,j}. \quad (1)$$

Интегральные дескрипторы \mathbf{g} множества изображений формируют признаковое пространство G .

В работе ставится задача преобразования признакового векторного пространства G в трехмерное тензорное пространство с размерностью $2(I-1) \times 2(J-1) \times l$, восстанавливающее пространственную смежность дескрипторов ячеек \mathbf{u} из блоков \mathbf{v} вектора \mathbf{g} . Используя Множество весов и параметров нейросетевого классификатора находятся из процедуры обучения по специально подготовленной обучающей выборке из базы изображений INRIA, CityScapes [9] содержащие два подмножества изображений: с положительными примерами, содержащими пешеходов, и отрицательными примерами, содержащими фон.

Критерием качества классификации на специально подготовленной тестовой выборке из базы изображений INRIA [9] и CityScapes будем считать отношение $MR = FN/(TP + FN)$ — доля неверно отвергнутых классификатором изображений (Miss Rate), $k FPPW = FP/(TN + FP)$ — доля неверно принятых изображений (False Positive Per Window), где

FN — количество неверно отвергнутых классификатором положительных примеров; TP — количество верно классифицированных положительных примеров; TN — количество верно классифицированных отрицательных примеров; FP — количество неверно классифицированных отрицательных примеров.

4 Дескрипторы HOG и HG-LBP

Рассмотрим полученные дескрипторы. Наш HOG дескриптор является классическим дескриптором. На данной выборке он сработал достаточно точно. Можно сравнить результаты с референсной работой [5]. Второй дескриптор же использовал положительные стороны каждого из дескрипторов LBP и HOG. Работа алгоритма была аналогична работе [5].

Алгоритм считает градиент в каждом пикселе в клетке, потом берет направление градиента и раскладывает его в гистограмму с 9 бинами, с каждой клетки одна гистограмма. Предполагается, что качество алгоритма будет походить на данный.

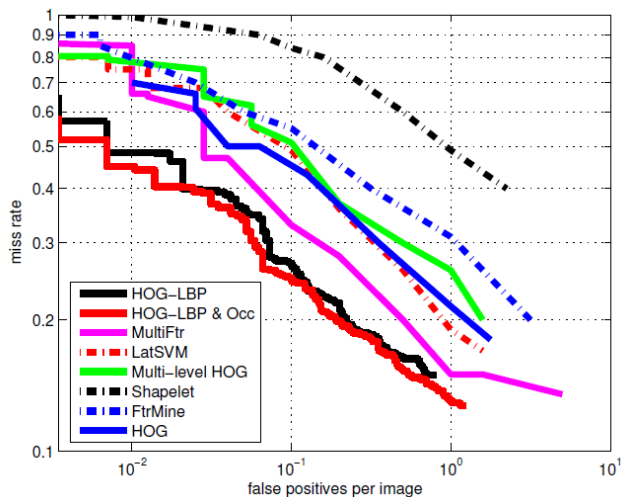


Рис. 4 Сравнение производительности между предлагаемыми детекторами и современные детекторы, основанные на INRIA набор данных [5]

5 Результаты вычислительных экспериментов

В данной работе в качестве набора изображений людей использовался датасет INRIA [9]. Данный набор данных был собран в рамках исследовательской работы по обнаружению вертикально расположенных людей на изображениях и видео [4]. Набор данных делится на 2 формата: (a) исходные изображения с соответствующими файлами аннотаций и (b) положительные изображения в нормализованном формате 64x128 пикселей (как используется в документе CVPR) с исходными отрицательными изображениями.

Мы использовали нормализованные данные, которые распределены в две папки: тест и обучение. Обе папки имеют две подпапки: (a) «pos» (нормализованные положительные тренировочные или тестовые изображения, центрированные на человеке с их левыми отражениями), (b) «neg» (содержащие оригинальные отрицательные тренировочные или тестовые изображения). Положительные изображения в папке обучения имеют размер 96x160 пикселей (с полями в 16 пикселей вокруг каждой стороны), а положительные изображения в папке теста имеют размер 70x134 пикселей (с полями 3 пикселя

вокруг каждой стороны). Это было сделано, чтобы избежать граничных условий (таким образом, чтобы избежать какого-либо конкретного смещения в классификаторе). В обеих папках используйте центрированное окно 64x128 пикселей для исходной задачи обнаружения.

Для создания отрицательных обучающих окон из нормализованных изображений фиксированный набор из 12180 окон (10 окон на негативное изображение) выбирается случайным образом из 1218 отрицательных обучающих фотографий, обеспечивающих исходный отрицательный обучающий набор.

6 Заключение

В данной работе был рассмотрен подход обобщения алгоритма LBP и HOG для определения пешеходов. Обычный LBP дескриптор включает гистограмму только количества вхождений LBP-кода в локальной области изображения и не учитывает степень выраженности каждого LBP-кода. В данной статье использовался вес каждого LBP-кода при построении локальной гистограммы дескриптора. Таким образом было произведено сравнение двух дескрипторов: HOG и построенного нами HG-LBP (Histogram Gradient LBP). Был учтён локальная выраженность LBP-кода, которая позволила учитывать больше информации о локальных особенностях изображения объекта в HG-LBP дескрипторе и улучшить качество работы детектора. Тестирование алгоритма было произведено на базе изображений IRNIA. Однако следует протестировать данный алгоритм на других выборках для более точной оценки качества.

HG-LBP descriptor based on gradient histograms for pedestrian detection

Razumov¹ I. O., Proshutinskiy¹ D. A., Gneushev^{1,2} A. N.

razzumoff@gmail.com, gneushev@ccas.ru, dmitriy.proshutinskiy@phystech.edu

¹Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, Moscow, Russia

²Federal Research Center "Computer Science and Control" of RAS, 44/2 Vavilova Str., Moscow, Russia

In this work we offer the approach is considered to generalize the LBP (Local Binary Patterns) algorithm for detecting pedestrians in an image based on taking into account the contribution of each LPB code to the spatial histogram of the descriptor. A regular LPB descriptor includes a histogram of only the number of occurrences of the LBP code in the local image area and does not take into account the severity of each LPB code. The article proposes to use the weight of each LBP code when constructing a local histogram of the descriptor. Two options for estimating the weight of the LPB are considered: based on the modulus of the gradient and on the basis of the sum of the modules of the derivatives in eight directions at the LBP code counting point to form the Histogram Gradient LBP. Taking into account the local manifestation of the LPB-code allows you to take into account more information about the local features of the image of the object in the HG-LPB descriptor and improve the quality of the detector. The article presents the results of a comparison of the work of the basic LPB, HOG-LBP algorithms and the proposed HG-LPB generalizations based on the images of INRIA pedestrians.

Keywords: *LBP, HOG, Histograms of Oriented Gradients, Local Binary Patterns, pedestrian detection.*

Литература

- [1] Gneushev, A. N., and A. B. Murynin. 2003. Adaptive gradient method for extracting contour features of objects in images of real-world scenes. *J. Comput. Sys. Sci. Int.* 42(6):973–980.

- [2] Гнеушев А.Н. Построение и оптимизация текстурно-геометрической модели изображения лица в пространстве базисных функций Габора. // Известия Академии Наук. Теория и системы управления. – 2007. – № 3. – с. 85-96.
- [3] Samsonov N. A., Gneushev, A. N. 2017. Textural descriptor in the Hough accumulator space of the gradient field for detecting pedestrians. Machine Learning and Data Analysis 3(3):203–215.
- [4] Dalal, N., and B. Triggs. 2005. Histograms of oriented gradients for human detection. IEEE CVPR. San Diego, CA.
- [5] Wang, X., Han X., Yan, S. 2009. An HOG-LBP Human Detector with Partial Occlusion Handling. ICCV.
- [6] Ojala, N., and M. Pietikainen 2002. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Trans on Pattern Analysis and Machine Intelligence. IEEE Trans on Pattern Analysis and Machine Intelligence 7(24).
- [7] Felzenszwalb, P. F., B. R. Girshick, D. McAllester, and D. Ramanan. 2010. Object detection with discriminatively trained part based models. IEEE T. Patt. Anal. 32(9):1627–1645.
- [8] Sun, D., and J. Watanada, 2015. Detecting pedestrians and vehicles in traffic scene based on boosted HOG features and SVM. IEEE 9th Symposium (International) on Intelligent Signal Processing.
- [9] INRIA Person Dataset. Available at: <http://pascal.inrialpes.fr/data/human/> (accessed June 4, 2017).
- [10] Open Source Computer Vision Library. Available at: <http://opencv.org/releases.html> (accessed May 16, 2017).
- [11] Vorontsov, K. V. 2007. Lectures on support vector machine. <http://www.ccas.ru/voron/download/SVM.pdf> (accessed June 25, 2017).
- [12] Alhind, T.J., and S. Kalra, and Ng, K.H. 2018. Comparing LBP, HOG and Deep Features for Classification of Histopathology Images, IEEE WCCI IEEE World Congress on Computational Intelligence 7(24).