

ЛОКАЛИЗАЦИЯ ЧЕЛОВЕКА В КАДРЕ ВИДЕОПОТОКА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА НА ОСНОВЕ РАСТУЩЕГО НЕЙРОННОГО ГАЗА И НЕЧЁТКОГО ВЫВОДА

О.С. Амосов¹, Ю.С. Иванов¹, С.В. Жиганов¹

¹ФГБОУ ВО «Комсомольский-на-Амуре государственный технический университет»,
Комсомольск-на-Амуре, Россия

Аннотация

Решается задача локализации человека в кадре видеопотока с помощью алгоритма расширяющегося нейронного газа и признакового описания на основе гистограмм ориентированных градиентов. В основе алгоритма используется оригинальная нейро-нечеткая модель растущего нейронного газа для обучения с подкреплением (GNG-FIS). Также предложена модификация алгоритма GNG-FIS с использованием двухпроходного обучения с нечеткой перемаркировкой классов и построением тепловой карты.

В результате экспериментов показатель правильности локализаций разработанного классификатора составил от 90,5 % до 93,2 %, в зависимости от условий сцены, что позволяет применять алгоритм в реальных системах ситуационной видеоаналитики.

Ключевые слова: локализация человека, растущий нейронный газ, кластеризация, нечеткий вывод.

Цитирование: Амосов, О.С. Локализация человека в кадре видеопотока с использованием алгоритма на основе растущего нейронного газа и нечеткого вывода / О.С. Амосов, Ю.С. Иванов, С.В. Жиганов // Компьютерная оптика. – 2017. – Т. 41, № 1. – С. 46-58. – DOI: 10.18287/2412-6179-2017-41-1-46-58.

Введение

Распознавание нестандартных ситуаций особенно актуально в системах, где требуется незамедлительное автоматическое принятие решения. Так, например, в системах безопасности к нестандартным ситуациям можно отнести нестандартное поведение людей, связанное с хулиганством, грабежом, потасовкой, праздничными, бег, появление в кадре посторонних предметов.

Задачи распознавания и реагирования на такие ситуации решаются алгоритмами интеллектуальной (ситуационной) видеоаналитики, которые включают в себя: распознавание номерных знаков [1], детекторы движения [2], распознавание лиц [3], обеспечение безопасности на транспорте [4] и т.д. При этом основным объектом, который необходимо локализовать и распознать в кадре видеопоследовательности, является человек.

Одним из классических подходов для решения задачи локализации объектов в кадре видеопотока является алгоритм Виолы–Джонса [3], успешно показавший себя при локализации лиц. Алгоритм основан на выделении локальных признаков и последующем обучении классификатора. В качестве признаков описания используются признаки Хаара [3].

Алгоритм реализован в виде загружаемых классификаторов для локализации следующих объектов: лицо человека (frontalface), верхняя часть тела человека (upperbody) [5], нижняя часть тела человека (lowerbody), тело человека (fullbody), глаза (eye) и др.

Тем не менее использование признаков Хаара оказалось малоэффективно в ситуациях, когда большинство (в том числе и фон) объектов на изображении обладают схожими цветовыми характеристиками [6]. Для решения этой проблемы Далал и Триггс [7] предложили использовать информацию о контурах объекта на основе гистограммы направленных градиентов (Histogram of Oriented Gradients, HOG).

Для распознавания людей на выделенных дескрипторах изображения Далал и Триггс [7] предложили использовать метод опорных векторов (Support Vector Machine, SVM) [8]. Несмотря на достоинства метода [8], также существует и ряд недостатков [9]:

- метод чувствителен к шумам и нормализации данных, т.е. изображение должно быть приведено к предопределенному стандарту;
- не существует общего подхода к автоматическому выбору ядра в случае линейной неразделимости классов.

Подход, основанный на вычислении признаков HOG и последующей классификации с использованием алгоритма SVM, получил название HOG SVM [7].

В статье [10] предлагается модификация HOG SVM, основанная на использовании в качестве слабого классификатора деревьев решений заданной высоты, что повышает скорость вычисления HOG.

В работе [11] применен метод деформируемых моделей. Он использует анализ геометрических ограничений, позволяя производить поиск в кадре человека с уменьшенным масштабом.

В работе [12] предлагается иной подход, основанный на использовании пространственно-временных признаков. В статье описываются признаки движения с отдельными пространственными и временными гистограммами, что позволяет определить человека в сложных ситуациях.

Однако, несмотря на различные модификации HOG SVM, существенным ограничением метода опорных векторов является неустойчивость по отношению к шуму в исходных данных. Шумовые выбросы обучающей выборки будут учтены при построении разделяющей гиперплоскости [13, 14], т.е. построенное на них решающее правило будет опираться на шум.

При этом все цифровые изображения, как правило, подвержены тем или иным шумовым искажениям, связанным с условиями съемки или помехами и влияющим на качество работы классификатора в целом.

Таким образом, возникает необходимость использования классификаторов устойчивых как к шумам, возникающим на поступающих изображениях, так и к шумовым выбросам в обучающей выборке.

В последнее время наилучшие результаты в оптическом распознавании образов достигаются при помощи использования искусственных нейронных сетей (ИНС) [15, 16, 17]. В работе [18] приводится сравнение различных алгоритмов для задачи распознавания рукописных цифр.

Наиболее перспективным представляется использование глубоких нейросетей (Deep Neural Network) [19, 20]. В работе приведено объединение классических подходов и сверточной нейронной сети. В работе [19] показано, что использование классических методов вместе со сверточной нейронной сетью значительно улучшает качество локализации человека. Основными недостатками таких сетей являются низкая скорость обучения и необходимость большого объема обучающих данных.

Для задачи локализации человека в видеопотоке особый интерес представляют нейронные сети, обучаемые без учителя [21, 22], так как они используют модель обучения, наиболее близкую к биологической [23]. Такой подход позволяет обнаружить скрытые зависимости и закономерности, существующие между объектами, тем самым разделяя множество образов на различные классы, в том числе и на шумовые выбросы.

Среди существующих самоорганизующихся сетей наилучшую самоорганизацию показал алгоритм, предложенный М. Мартинесом, С. Берковичем и К. Шультенем в работе [24] и названный растущий нейронный газ (Growing Neural Gas, GNG). Алгоритм отличается высокой скоростью обучения.

Таким образом, при построении алгоритма для решения задачи локализации человека в кадре видеопотока перспективными являются следующие подходы:

- использование признакового пространства НОГ;
- использование самоорганизующихся нейронных сетей в качестве классификатора из-за их устойчивости к шумовым выбросам.

Научная новизна предлагаемой работы состоит в применении комбинированного алгоритма на основе нейро-нечеткой модели растущего нейронного газа для обучения с подкреплением в задаче локализации человека в видеопотоке. Новым в алгоритме является применение двухпроходного обучения с последующей перемаркировкой классов, осуществляемой блоком нечеткого вывода.

1. Постановка задачи локализации человека в кадре видеопотока

Пусть имеется непрерывный видеопоток S , состоящий из отдельных кадров. Каждый кадр видеопоследовательности представляет собой цифровое

изображение. В памяти компьютера изображение представляется в виде двумерной матрицы пикселей I_t размером, $width \times height$, где t – номер кадра в видеопоследовательности. Каждый пиксель в матрице представлен некоторым неотрицательным числом:

- от 0 до 255, если изображение черно-белое;
- от 0 до 255^3 , если изображение в модели RGB (Red – красный, Green – зеленый, Blue – синий).

По кадру видеопоследовательности требуется определить, если он есть, список прямоугольников, в которых присутствует локализуемый объект – человек:

$$\mathit{rectangle}_\eta = \{x, y, w, h\}, \quad (1)$$

где x, y – координаты центра η -го прямоугольника, w – ширина, h – высота.

Необходимо заметить, что в кадре может присутствовать несколько объектов.

Поиск объекта будет выполняться «сканирующим окном» C_r , где r – номер сканирующего окна (шаг сканирования), которое представляет собой матрицу пикселей. Окно проходит последовательно по изображению с шагом 1 или более пикселей (слева направо, сверху вниз) с изменением масштаба при каждом проходе. Размер окна определяется его шириной w_c и высотой h_c .

При прохождении сканирующего окна по изображению для каждого положения окна необходимо определить, содержится ли в окне человек или нет. Это может быть выполнено путем отнесения анализируемой части изображения к одному из двух классов. Первый класс – наличие человека, второй класс – его отсутствие. Причем, если в окне C_r содержится искомым объект, то $\mathit{rectangle} = \{x_r, y_r, w_r, h_r\}$, где x_r, y_r – координаты центра сканирующего окна, а w_r, h_r – его высота и ширина.

Так как существует всего два класса {«не-человек»; «человек»} = {−1; +1}, то задача локализации человека в кадре видеопотока сводится к задаче классификации. Рассмотрим классическую постановку задачи классификации [14].

Дано множество X , в котором хранится описание объектов o . Y – конечное множество классов. Решающей функцией (классификатором) F является отображение множества X в множество Y , т.е. $F: X \rightarrow Y$. Признак f объекта o – это отображение $f: o \rightarrow D_f$, где D_f – множество допустимых значений признака f . Если задан набор признаков f_1, \dots, f_m для некоторого объекта o , то вектор признаков x объекта $o \in X$ может быть определен так: $x = (f_1(o), \dots, f_m(o))$.

Классификатор F должен быть способен классифицировать произвольный объект $o \in X$. Приведенная постановка задачи иллюстрируется рис. 1.

Оптимальным считается классификатор, который дает наименьшую вероятность ошибки $P(x)$ при всех допустимых значениях x . Тогда критерием оптимальности будет $P(x) \rightarrow \min_{x \in X}$. Следует отметить, что ошибки разделяются на «ошибки 1-го рода» (ложно-

положительные) – ложное срабатывание, «ошибки 2-го рода» (ложноотрицательные) – пропуск события.

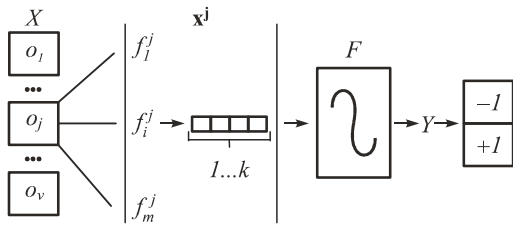


Рис. 1. Постановка задачи бинарной классификации

В оптическом распознавании образов для оценки бинарного классификатора применяется ROC-анализ (*receiver operating characteristic*, рабочая характеристика приёмника) [25]. Количественную интерпретацию ROC дает показатель AUC (*area under ROC curve*, площадь под ROC-кривой) — площадь, ограниченная ROC-кривой и осью абсцисс.

Для обучения классификатора F используется обучающая выборка, заданная множеством $D = \{(x_1, y_1), \dots, (x_v, y_v)\}$, где $y_j \in Y = \{-1; +1\}, j = 1 \dots v$.

В качестве признакового пространства нами используется метод HOG [7], как наиболее зарекомендовавший себя для решения задачи локализации человека. Тогда алгоритм перевода в пространство HOG выглядит следующим образом:

$$\Phi: o_j \rightarrow x^j. \tag{2}$$

Пример изображения в пространстве HOG представлен на рис. 2.

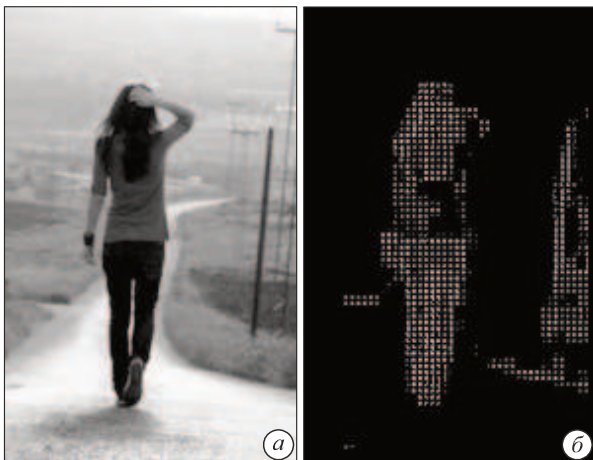


Рис. 2. Пример изображения в формате: а) оригинальный кадр; б) визуализация HOG

Использование метода HOG имеет следующие преимущества:

1. Устойчивость к геометрическим преобразованиям изображения.
2. Высокая устойчивость к шуму, связанному с движением за счет глубокого разбиения пространства малыми ядрами.

2. Решение задачи локализации человека в кадре видеопотока с использованием метода нейронного газа и последующей перемаркировкой

Решение задачи локализации человека в кадре видеопотока разбивается на решение ряда подзадач:

1. Выполняется предобработка с выделением информации о яркости изображения и перевод изображения в признаковое пространство HOG.
2. Выполняется классификация объекта, выделенного сканирующим окном, алгоритмом растущего нейронного газа.
3. Выполняется переход к бинарной классификации путем маркировки с использованием нейро-нечеткой модели растущего нейронного газа.
4. Выполняется нечеткая перемаркировка и построение тепловой карты для уменьшения ошибки 2-го рода.

Алгоритм представлен на рис. 3, где F_{GNG} – классификатор нейронного газа, l – номер класса, D – обучающее множество.

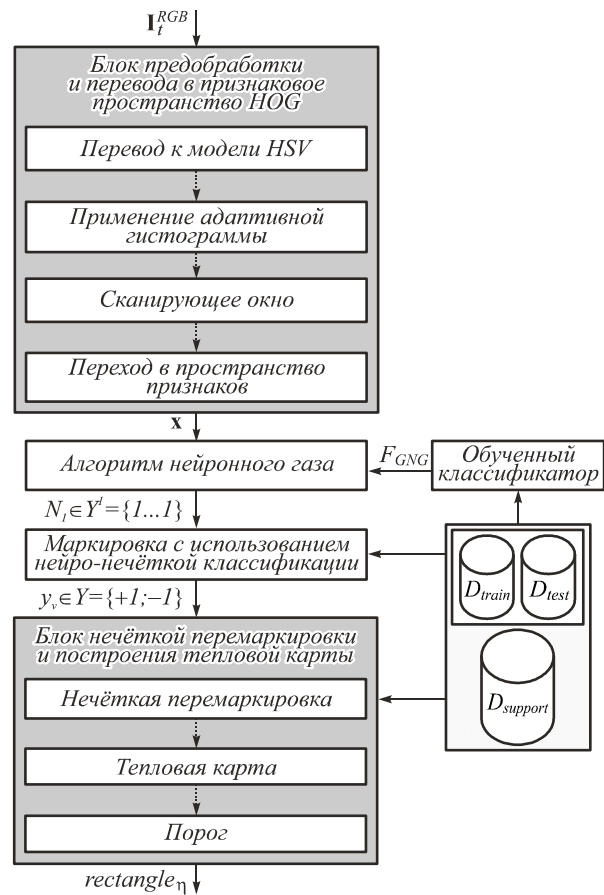


Рис. 3. Решение задачи локализации человека

2.1. Предобработка изображения с выделением информации о яркости

Пусть имеется кадр видеопотока I_t в модели RGB. Необходимо извлечь информацию о яркостных характеристиках изображения.

Классический алгоритм использует изображение в модели RGB, но, как показывает практика [26], модель HSV (Hue – тон, Saturation – насыщенность, Value – значение) [27] более информативна для локализации человека [26].

В отличие от распространенной цветовой модели RGB, каналы HSV содержат информацию о цветовом тоне, насыщенности и яркости. На рис. 4 приведены

изображения в моделях RGB (а) и HSV (б), что подтверждает предположение о высокой информативности модели HSV.

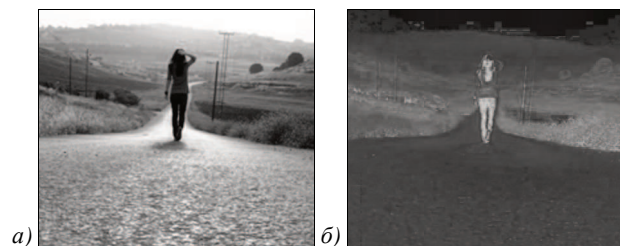


Рис. 4. Изображение в различных цветовых моделях: а) RGB, б) HSV

Для повышения контрастности изображения используется метод адаптивной эквализации (выравнивания) гистограммы, подробно описанный в работе [28], заключающийся в вычислении нескольких гистограмм, каждая из которых соответствует отдельной части изображения.

Для программной реализации данного метода можно использовать готовую функцию *adapthisted* из библиотеки компьютерного зрения OpenCV [27].

После обработки изображения I_7^{HSV} методом адаптивного выравнивания гистограммы *adapthisted* (I) по нему проходит сканирующее окно, результатом которого становится некая область-объект o , которую необходимо отнести к одному из множества классов.

2.2. Классификация объекта алгоритмом растущего нейронного газа

Пусть имеется объект o , представленный вектором признаков x , а также обучающая выборка, заданная конечным множеством D . Необходимо построить оптимальный классификатор F , дающий наименьшую вероятность ошибки распознавания при различных значениях x .

В качестве классификатора предлагается использовать алгоритм растущего нейронного газа [24]. Алгоритм применялся для задач оптического распознавания образов, анализа временных рядов, кластеризации и др. [24, 30, 31].

В работах [32, 33] также предлагались различные модификации, основанные на изменении механизма роста, онлайн-обучении и др.

При этом, как правило, решалась либо задача кластеризации [34], либо задача многомерной классификации. Таким образом, можно утверждать, что алгоритм GNG может быть применен для построения классификатора, удовлетворяющего условиям приведенной ранее задачи.

Алгоритм обучения нейронного газа выглядит следующим образом [24]:

1. Установить параметры растущего нейронного газа:
 - age_{max} – максимальный возраст ребра;
 - ϵ_x, ϵ_z – коэффициенты обучения нейрона-победителя и его соседей;
 - λ – количество итераций для вставки нового узла (нейрона);

- τ – количество итераций до остановки обучения;
 - α, β – коэффициенты для адаптации ошибок.
2. Задаются два начальных узла u и v в случайных точках ω_u и ω_v , u и v соединяются ребром, возраст age которого равен нулю. Ошибка err в узлах u и v принимается равной нулю.
 3. Выбирается очередной объект o , который представлен вектором x из входного множества D .
 4. Определяются узлы ψ_1 и ψ_2 – первый и второй по критерию близости к o . В качестве меры близости используется стандартное евклидово расстояние $E = \|\omega_{\psi} - x\|$, $\psi = \psi_1, \psi_2$.
 5. Увеличивается на 1 возраст дуг, исходящих от ψ_1 .
 6. Счетчик ошибки нейрона ψ_1 увеличивается на величину квадрата расстояния до o : $E(\psi_1) = \|\omega_{\psi_1} - x\|^2$.
 7. ψ_1 и узлы, соединённые с ним, смещаются по направлению к o на расстояния $\Delta\omega_{\psi_1} = \epsilon_x(x - \omega_{\psi_1})$ и $\Delta\omega_{\psi_n} = \epsilon_z(x - \omega_{\psi_n})$ соответственно, где $0 < \epsilon_x < 1$ и $0 < \epsilon_z < \epsilon_x$.
 8. Если ψ_1 и ψ_2 соединены ребром, то возраст ребра обнуляется; в противном случае между ψ_1 и ψ_2 создается новое ребро с возрастом, равным нулю.
 9. Все ребра в графе с возрастом более age_{max} удаляются. В случае, если после этого некоторые узлы не имеют ребер (оказываются изолированы), эти узлы также удаляются.
 10. Если номер текущей итерации кратен величине λ , осуществляется вставка нового узла в точке $\omega_p = 0,5(\omega_u - \omega_v)$, где u – узел с наибольшей накопленной ошибкой; v – узел, соединенный ребром u с наибольшей накопленной ошибкой. Ребро между u и v удаляется, вместо него добавляется новый узел – p и добавляются ребра между u и p , а также между p и v . Ошибка err в узлах u и v уменьшается умножением на константу $\alpha < 1$; значение ошибки нового нейрона p инициализируется равным значению ошибки u .
 11. Ошибка err во всех узлах уменьшается умножением на константу $\beta < 1$.
 12. Если условие остановки не выполнено, перейти к шагу 2. Стандартным условием остановки является выполнение определенного количества итераций обучения τ . Пример структуры GNG приведен на рис. 5.

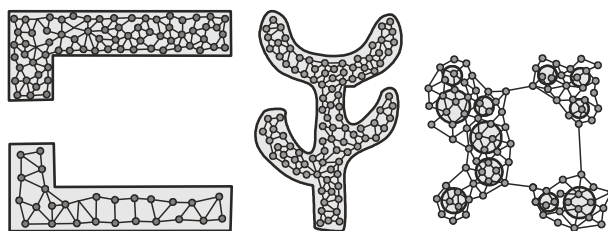


Рис. 5. Пример структуры GNG

Для решения задачи локализации человека на изображении с использованием алгоритма GNG была составлена обучающая выборка D . Выборка была со-

ставлена на основе классических наборов открытых баз данных [35, 36], а также данных, собранных вручную. В выборке представлены позитивные изображения (с человеком) и негативные (без человека) в равных количествах. Для ускорения сходимости и повышения качества кластеризации при обучении данные подавались потоком с чередованием негативных и позитивных изображений [14].

Тестирующая выборка была составлена на основании первоначального набора данных случайным разбиением выборки D на 2 части в соотношении обучающая/тестирующая – 50/50 %.

Изображения в выборке были нормированы к формату 128×64 пикселей. Обучающая выборка D_{train} содержит 1500 позитивных и 1500 негативных изображений. Тестирующая выборка D_{test} содержит 3000 изображений с тем же соотношением позитивных/негативных, что и обучающая выборка. Примеры изображений из обучающей и тестовой выборки приведены на рис. 6а – негативные и рис. 6б – позитивные.

Результатом обучения растущего нейронного газа является разбиение обучающего множества на $N_i \geq 2$ классов, т.е. решается задача кластеризации. Важно отметить, что состав и количество кластеров формируется в ходе самоорганизации и неизвестно заранее. В отличие от линейных классификаторов, разделяющих многомерное пространство на 2 класса, алгоритм GNG создает отдельные кластеры для зашумленных объектов.

При получении на вход GNG нового объекта o , отсутствующего в обучающей выборке D_{train} , алгоритм будет относить его к одному из известных классов. Таким образом, решается задача многоклассовой классификации, нам же необходимо решить задачу бинарной классификации, т.е. отнесения объекта o к одному из двух классов: «человек»/«не человек».



Рис. 6. Примеры изображений, содержащихся в обучающей и тестирующей выборках: а) негативные; б) позитивные

2.3. Маркировка с использованием нейро-нечеткой модели растущего нейронного газа

Пусть имеется обученный классификатор F_{GNG} , относящий объект o , т.е. его признаковое описание x , к одному из множества N_i классов. Необходимо при-

вести работу классификатора F_{GNG} к решению задачи бинарной (двух классовой) классификации, где один класс является позитивным (с присутствием человека), а другой – негативным (без присутствия человека), т.е. отнесение к классам $y_j \in Y = \{+1; -1\}$.

Казалось бы, очевидным способом приведения к задаче бинарной классификации является искусственное ограничение числа кластеров в сети растущего нейронного газа на этапе обучения. Но при этом теряется суть самоорганизации, а результат работы алгоритма ухудшается с увеличением количества разнородных данных в выборке, что делает такой подход неприемлемым.

Для перехода к задаче бинарной классификации, т.е. понижения размерности, нами предлагается использование заранее размеченного и разбитого на 2 класса обучающего множества. После обучения классификатора F_{GNG} необходимо выполнить маркировку (изменение метки класса) полученных классов на позитивные и негативные. Выполнение такой маркировки является трудноформализуемой задачей, которая решается с использованием блока нечеткого вывода.

Поэтому нами предлагается объединение нейронных сетей и нечеткой логики. Такой подход показал свою эффективность во многих задачах [37, 38]. Так, например, одной из известных нейро-нечетких структур является ANFIS [39].

Нами предлагается структура нейро-нечеткой модели (GNG-FIS), принцип работы которой представлен на рис. 7. В предлагаемом алгоритме входные данные подаются на нейронную сеть GNG, а выход из нейронной сети обрабатывается при помощи нечеткой системы FIS.

В этом случае нейронная сеть действует как пре-процессор нечетких выходов, что позволяет привести классификатор к бинарному виду.

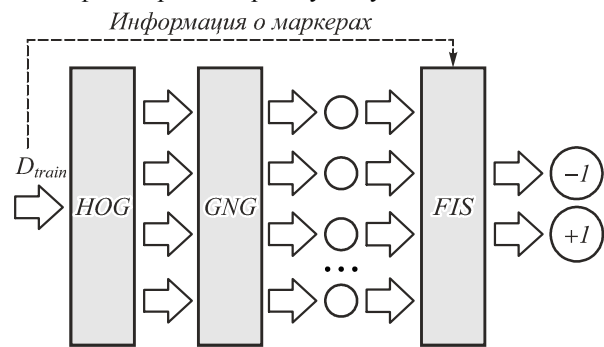


Рис. 7. Принцип работы предлагаемого алгоритма GNG-FIS

Маркировка выполняется в 2 этапа: на первом этапе используется классический алгоритм, на втором этапе используются нечеткие правила.

Первый этап маркировки выглядит следующим образом:

1. Необходимо создать массив объектов R , размером равным количеству классов N_i . Каждый объект R содержит следующие параметры, характеризующие состав кластера (класса):

- n_{pos} – количество позитивных образов;

- n_{neg} – количество негативных образов.
2. Необходимо классифицировать объект o из обучающего множества D_{train} с помощью обученной сети GNG.
 3. Необходимо выбрать объект $R_l \in R$, где l – полученный номер класса из сети GNG. Используя заранее установленную маркировку из D_{train} , требуется увеличить один из параметров R_l следующим образом:
 - если образ негативный – увеличиваем значение n_{neg} на 1;
 - если образ позитивный – увеличиваем значение n_{pos} на 1.

После завершения 1-го этапа маркировки следующие классы маркируются как негативные, если:

- количество n_{neg} больше, чем n_{pos} ($n_{neg} > n_{pos}$);
- количество n_{neg} и n_{pos} равно ($n_{neg} = n_{pos}$);
- количество n_{neg} равно 1 ($n_{pos} = 1$).

На 2-м этапе для всех кластеров, удовлетворяющих условию $0 < n_{neg} < n_{pos}$, необходимо вычислить:

- долю позитивных:
- $$h_l(n_{pos}) = n_{pos} / (n_{pos} + n_{neg}) 100, \tag{3}$$

- долю негативных:
- $$h_l(n_{neg}) = n_{neg} / (n_{pos} + n_{neg}) 100, \tag{4}$$

- среднее евклидово расстояние для позитивных образов до центра кластера:
- $$m_l(n_{pos}) = (\sum E_{pos}^l) / n_{pos}, \tag{5}$$

- среднее евклидово расстояние для негативных образов до центра кластера:
- $$m_l(n_{neg}) = (\sum E_{neg}^l) / n_{neg}. \tag{6}$$

Переход к нечеткой модели доверия к кластерам выглядит следующим образом.

Пусть имеются входные переменные $h_l(n_{pos})$, $h_l(n_{neg})$, $m_l(n_{pos})$, $m_l(n_{neg})$ и выход q , характеризующий степень доверия к кластерам. Введем термножества:

1. $T_h(h_l) = \{ \text{маленькая, средняя, большая} \}$ – для доли позитивных/негативных образов,
2. $T_m(m_l) = \{ \text{маленькая, средняя, большая} \}$ – для среднего евклидового расстояния образов до центра кластеров.
3. $T_w = \{ \text{положительный, сомнительный, отрицательный} \}$ – для выхода q .

Табл. 1. Результаты эксперимента, D_{test} – 3000 тестирующих образов

| Название метода | Количество не распознанных (ошибка 2-го рода) | Количество ложных распознаваний (ошибка 1-го рода) | Общая ошибка | Время обработки | Время обучения |
|-----------------|---|--|--------------|-----------------|----------------|
| HOG SVM | 481 | 93 | 19,13 % | 1 мс | 3–5 часов |
| Viola-Jones | 1119 | 91 | 40,33 % | 24 мс | 3–5 дней |
| GNG-FIS | 149 | 123 | 9,06 % | 10 мс | 5 часов |

Для повышения качества работы алгоритма нами предлагается использование двухпроходного обучения с дополнительной перемаркировкой, т.е. изменением метки класса.

Процесс последующей перемаркировки предполагает наличие отдельного подкрепляющего множества $D_{support}$.

Алгоритм вывода Мамдани, состоящий из 16 правил, был реализован в системе Matlab.

В качестве функции принадлежности используется симметричная гауссовская кривая (gaussmf).

Правило вывода для нечеткой маркировки в общем виде записывается так:

$$\begin{aligned} \text{if } h_l(n_{pos}) = T_h(h_l) \& m_l(n_{pos}) = T_m(m_l) \& \\ \& h_l(n_{neg}) = T_h(h_l) \& m_l(n_{neg}) = T_m(m_l), \tag{7} \\ \text{then } q_l = T_q. \end{aligned}$$

В результате работы алгоритма нечеткой маркировки «положительные» и «сомнительные» кластеры маркируются как +1 (позитивные), «отрицательные» – как -1 (негативные).

Таким образом, результатом работы предложенного алгоритма GNG-FIS является $y_i \in Y = \{-1; +1\}$.

2.4. Нечеткая перемаркировка и построение тепловой карты

Нами был проведен промежуточный эксперимент для выявления недостатков предложенного алгоритма. Сравнение проводилось с классическими алгоритмами: HOG SVM, Viola Jones с классификатором fullbody [29].

Для проведения вычислительного эксперимента нами была разработана программа на объектно-ориентированном языке программирования C#, в которой были реализованы алгоритмы GNG-FIS.

Программа была протестирована на компьютере со следующими характеристиками: ЦПУ – Intel Core i5-2500 CPU @ 3.30GHz; ОЗУ – 8 Gb; ОС – Microsoft Windows 10.

Для обучения алгоритма GNG-FIS использовались следующие параметры:

1. Гистограмма направленных градиентов: размер ячейки 16×16, размер блока 32×32, шаг сдвига блока 16×16, размер изображения 64×128.
2. Растущий нейронный газ: $\epsilon_g = 0,2$, $\epsilon_g = 0,006$, $\alpha = 0,5$, $\lambda = 100$, $\beta = 0,995$, $\tau = 6000$, $age_{max} = 50$.

Результаты проведенного промежуточного эксперимента (табл. 1) на описанной ранее выборке D_{test} показывают высокий уровень ошибки 2-го рода, что обуславливает необходимость модификации предложенного алгоритма.

Была проведена серия дополнительных экспериментов с последовательным уменьшением тестирующего множества D_{test} до 2100, с шагом 100 образов. В качестве критерия сравнения использовалось значение процента общей ошибки каждого из алгоритмов, причем допустимым колебанием можно считать значение, не превышающее 0,5 %.

Для выборки D_{test} размером 2100/2400/2700 образцов значения процентов общей ошибки по результатам тестирования следующие:

- HOG SVM – 19,05 / 19,33 / 18,81;
- Viola-Jones – 40,29 / 40,17 / 9,96;
- GNG-FIS – 8,95 / 9,46 / 9,22.

Таким образом, уменьшение тестирующей выборки на 900 изображениях является допустимым и позволяет сформировать подкрепляющее множество $D_{support}$, которое будет использовано для корректировки работы алгоритма.

Тогда, в отличие от классического подхода к построению обучающей и тестирующей выборки, нами предлагается сформировать подкрепляющую выборку $D_{support}$ путем случайного разбиения обучающего множества D на 3 части ($D_{train} / D_{test} / D_{support}$) в пропорциях 50 % / 35 % / 15 %.

Табл. 2. Результаты эксперимента, D_{test} – 2100 тестирующих образов

| Название метода | Количество не распознанных (ошибка 2-го рода) | Количество ложных распознаваний (ошибка 1-го рода) | Общая ошибка |
|---|---|--|--------------|
| HOG SVM | 358 | 42 | 19,05 |
| Viola-Jones | 801 | 45 | 40,29 |
| GNG-FIS | 76 | 112 | 8,95 |
| GNG-FIS с использованием двухпроходной перемаркировки | 120 | 52 | 8,19 |

Применение двухпроходной перемаркировки несколько увеличивает ошибку 2-го рода, но за счёт снижения количества ложных срабатываний (ошибки 1-го рода) общая эффективность алгоритма увеличивается.

Принцип работы предлагаемого модифицированного алгоритма GNG-FIS с использованием двухпроходной перемаркировки представлен на рис. 8.

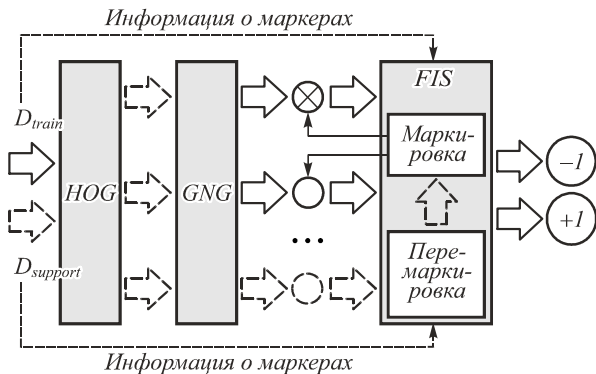


Рис. 8. Принцип работы модифицированного алгоритма GNG-FIS с использованием двухпроходной перемаркировки

Так как работа алгоритма подразумевает использование «сканирующего окна» с настраиваемым шагом сканирования и изменяющимся масштабом при многократных итерациях, то возникает эффект перекрывающихся областей. То есть при прохождении окна возможна многократная классификация, когда один и тот же объект будет распознан классификатором с небольшими сдвигами.

Такой эффект позволяет построить «тепловые карты» классификации. Подобный подход применяется во многих задачах, в том числе и для локализации пешеходов [40].

В случае появления ошибок 1-го или 2-го рода классы должны быть перемаркированы по правилам:

- На вход GNG подается изображение из подкрепляющего множества $D_{support}$, на выходе будет получен номер класса l , к которому относится изображение. Выбираем из массива объект с полученным номером класса R_l .
- Если объект из $D_{support}$ не содержит человека и евклидово расстояние от объекта до центра кластера «маленькое», а класс маркирован как положительный, то класс маркируется как негативный (не содержит человека).

В табл. 2 представлены результаты сравнения алгоритмов HOG SVM, Viola-Jones, GNG-FIS и GNG-FIS с использованием двухпроходной перемаркировки. Эксперимент проводился на тестирующей выборке 2100 образов (0,35 % от множества D).

Алгоритм построения таких карт сводится к следующему:

- 1) Пусть $I(width, height)$ – изображение.
- 2) Создается пустая матрица весов G размером $width \times height$.
- 3) Выбирается коэффициент масштаба тепловой карты d .
- 4) На каждом шаге сканирования, количество которых зависит от размера изображения:
 - а) При прохождении сканирующего окна S_r размером $w_c \times h_c$, по изображению I :
 - если окно отнесено к положительному классу, в матрице G выделяется область Q , соответствующая координатам и размеру окна S_r на изображении I . Значения всех Q элементов увеличиваются на d ;
 - если окно отнесено к отрицательному классу, в матрице G выделяется область Q , соответствующая координатам и размеру окна S_r на изображении I . Значения всех Q элементов уменьшаются на d .
 - б) После завершения работы сканирующего окна области с высокими значениями характеризуют позитивные классы, а области с отрицательными значениями характеризуют негативные.

На рис. 9 представлена матрица G , а также ее визуализация в виде поверхности и карты уровней.

Таким образом, если количество негативных срабатываний в области больше позитивных, то вероятность отнесения ее к позитивному классу снижается. Применение такого подхода позволяет бороться с эффектом «единичных выбросов», т.е. единичных ошибочных срабатываний классификатора. При наложении матрицы G на изображение I мы получаем тепловую карту, представленную на рис. 10.

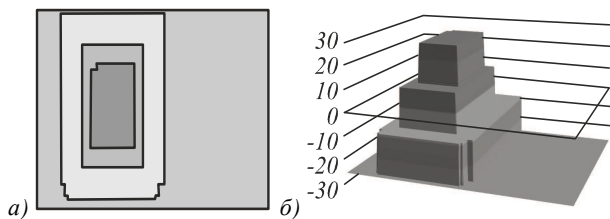


Рис. 9. Визуализация матрицы G



Рис. 10. Визуализация «тепловой карты»

Применение настраиваемого порогового фильтра позволяет выбрать области с наибольшей вероятностью нахождения в ней человека.

Тогда конечным результатом работы предложенного модифицированного нейро-нечеткого алгоритма расширяющегося нейронного газа является список прямоугольников, в которых присутствует локализуемый объект – человек $rectangle_{\eta}$.

3. Эксперимент с перемаркировкой на данных, полученных с камер видеонаблюдения

Разработанная нами программа была доработана с учетом предложенных модификаций.

При обучении используется выборка, полученная случайным разбиением первоначального набора данных на 3 части:

- D_{train} – содержит 3000 изображений;
- D_{test} – содержит 2100 изображений;
- $D_{support}$ – содержит 900 изображений.

Параметры обучения не изменялись.

Существенным отличием второго эксперимента является источник изображений для тестирования – видеокмеры действующей системы видеонаблюдения Университета.

Видеопоток, получаемый с камеры 1 (рис. 11а), обладает следующими параметрами: угол наклона $\sim 45^\circ$, уровень освещения – 300 Лк (утро), разрешение – 640×480 пикселей, помещение (длинный коридор).

Видеопоток, получаемый с камеры 2 (рис. 11б), обладает следующими параметрами: угол наклона $\sim 60^\circ$, уровень освещения – 250 Лк (день), помещение (узкий оживленный холл), разрешение – 640×480 пикселей.

Видеопоток, получаемый с камеры 3 (рис. 11в), обладает следующими параметрами: угол наклона $\sim 35^\circ$, уровень освещения – более 1000 Лк (пасмурный день), улица (вход и тротуар), разрешение – 640×480 пикселей.

Для экспериментов были вручную промаркированы видеопоследовательности продолжительностью по 5 минут с частотой кадров (Frame Per Second, FPS) 30 кадров/с.



Рис. 11. Видеокмеры действующей системы видеонаблюдения Университета: а) камера 1, длинный коридор; б) камера 2, узкий оживленный холл; в) камера 3, улица (вход и тротуар)

Сравнение классификаторов выполнялось путем построения ROC-кривых с вычислением показателя AUC, характеризующего качество классификации.

В табл. 3 и на рис. 12 приведены результаты сравнения алгоритмов: Viola-Jones, HOG-SVM, GNG-FIS без маркировки, GNG-FIS с перемаркировкой для камеры 1.

Табл. 3. Результаты второго эксперимента на видеопотоке с камеры наблюдения 1

| Название метода | Количество верных локализаций | Количество ложных локализаций (ошибка 1-го рода) | Процент правильных локализаций (AUC) |
|--|-------------------------------|--|--------------------------------------|
| HOG SVM | 66 | 35 | 92,3 % |
| Viola-Jones | 56 | 17 | 82,9 % |
| GNG-FIS | 66 | 20 | 87,5 % |
| GNG-FIS с применением двухпроходного обучения и построением тепловой карты | 66 | 11 | 93,2 % |

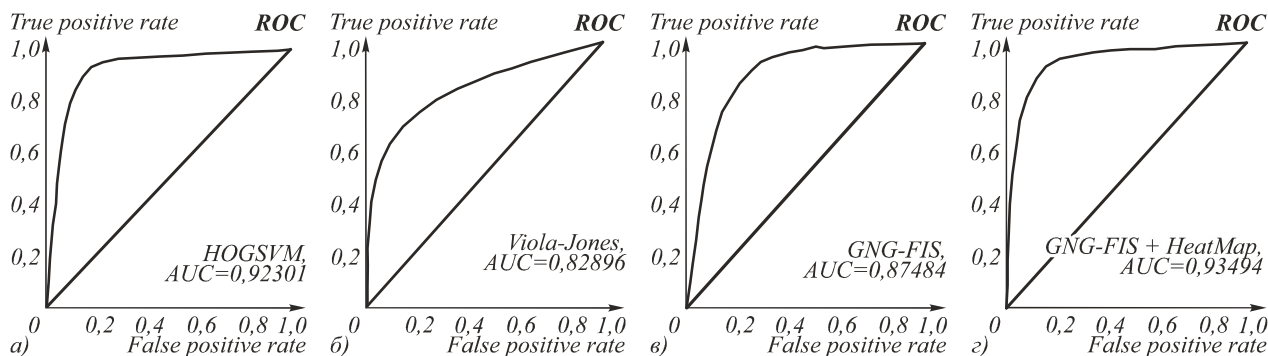


Рис. 12. ROC-кривые для камеры 1: а) HOG SVM, б) Viola-Jones, в) GNGFIS, г) GNGFIS (с двухпроходным обучением и тепловой картой)

Следует отметить, что при попадании в поле зрения камеры каждый объект (человек) должен детектироваться на каждом кадре. При этом в колонках «количество верных локализаций» и «количество ложных локализаций» указано количество человек и количество ложных объектов, локализованных хотя бы один раз.

На рис. 13 показана работа тестируемых алгоритмов и возникающие ошибки.

4. Обсуждение результатов экспериментов

Проверка алгоритма осуществляется как на открытых контрольных наборах изображений, так и на реальной задаче охранного видеонаблюдения при сложных условиях.

Как показывает практика, особый интерес вызывают именно экспериментальные данные, полученные на действующих камерах, а не на синтезированных выборках данных, пригодных для обучения и предварительного контроля. Важно отметить, что во втором вычислительном эксперименте в одном кадре могло присутствовать одновременно несколько человек, что не влияло на качество и скорость работы алгоритмов, так как в основе всех рассмотренных подходов лежит принцип сканирующего окна.

Применение «тепловой карты» снижает количество ложных срабатываний, что видно из форм ROC-кривых (кривая смещается влево или вверх под влиянием ошибок того или иного рода), приведенных на рис. 12.

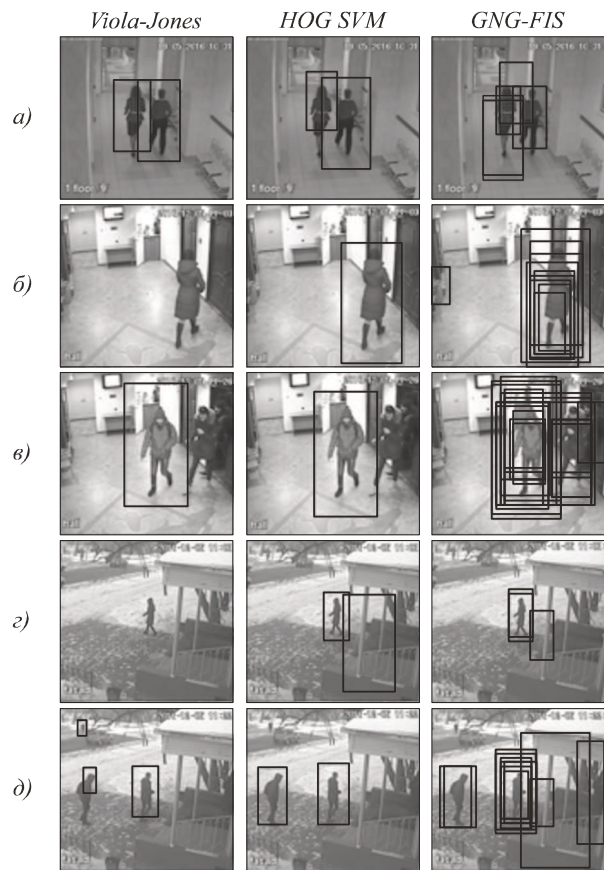


Рис. 13. Работа тестируемых алгоритмов

Табл. 4. Результаты сравнения всех классификаторов для 3 камер

| Название метода | Процент правильных локализаций на камере 1 (AUC) | Процент правильных локализаций на камере 2 (AUC) | Процент правильных локализаций на камере 3 (AUC) |
|--|--|--|--|
| HOG SVM | 92,3 % | 91,1 % | 93,1 % |
| Viola-Jones | 82,9 % | 81,5 % | 85,5 % |
| GNG-FIS | 87,5 % | 86,7 % | 81,7 % |
| GNG-FIS с применением двухпроходного обучения и построением тепловой карты | 93,2 % | 92,2 % | 90,5 % |

Анализируя табл. 4 и рис. 13, можно утверждать, что классические методы мало устойчивы к возможным перекрытиям объектов на изображении и показывают большую ошибку в помещениях с малой площадью (рис. 13в). Предлагаемый нами алгоритм показывает более высокий процент правильных локализаций для внутренних камер (рис. 13а-в) и чуть меньший процент на изображениях, полученных с уличных камер видеонаблюдения, в особенности для удаленных объектов (рис. 13г). Однако этот алгоритм обладает возможностью «тонкой» подстройки, что позволит адаптировать его под конкретные условия съемки.

Кроме того, рис. 13 демонстрирует применения тепловой карты (множественной классификации) для снижения ошибок 1-го рода.

Результаты доказывают, что предложенный алгоритм с модификацией имеет следующие преимущества перед классическими алгоритмами:

- высокий процент распознавания за счет снижения количества ложных срабатываний (ошибки 1-го рода);
 - возможность адаптивной корректировки результатов кластеризации, что позволяет провести более «тонкую» настройку алгоритма;
 - высокая скорость обучения алгоритма.
- Однако для предложенного алгоритма остается трудность при подборе параметров обучения.

Заключение

В настоящей статье решена задача локализации человека в кадре видеопотока с помощью расширяющегося нейронного газа, нечеткого вывода и признакового описания на основе гистограмм ориентированных градиентов.

Научная новизна состоит в применении комбинированного алгоритма на основе растущего нейронного газа и нечеткого вывода для обучения с подкреплением в задаче локализации человека в видеопотоке.

Новым в алгоритме является применение двухпроходного обучения и последующей перемаркировкой классов, осуществляемой блоком нечеткого вывода.

Классические алгоритмы настроены на высокий порог срабатывания, что приводит к пропуску объектов. В отличие от известных алгоритмов локализации человека в кадре видеопотока, предлагаемый подход ориентирован на ситуационную видеоаналитику, в которой главное – не пропустить тревожное событие.

По сравнению с классическими алгоритмами предлагаемый модифицированный подход устойчив к ошибкам 2-го рода, обладает высокой скоростью обучения, а также обладает возможностями адаптивной настройки за счет корректировки правил.

Таким образом, указанные преимущества позволяют применять предложенный алгоритм в реальных системах локализации образов, что подтверждается результатами экспериментов.

Благодарности

Работа выполнена при поддержке Минобрнауки России научного проекта – госзадания в рамках проектной части № 2.1898.2017/ПЧ «Создание математического и алгоритмического обеспечения интеллектуальной информационно-телекоммуникационной системы безопасности вуза».

Литература

1. **Амосов, О.С.** Модифицированный алгоритм локализации номерных знаков транспортных средств на основе метода Виолы-Джонса / О.С. Амосов, Ю.С. Иванов // Информатика и системы управления. – 2014. – Т. 39, № 1. – С. 127-140.
2. **Мельников, И.И.** Детектор движения на основе импульсных нейронных сетей / И.И. Мельников, К.А. Демиденков, И.А. Емельянов, И.А. Евсеенко // Информационные технологии. – 2013. – № 7. – С. 57-60.
3. **Viola, P.** Robust real-time face detection / P. Viola, M.J. Jones // International Journal of Computer Vision. – 2004. – Vol. 57(2). – P. 137-154. – DOI: 10.1023/B:VISI.0000013087.49260.fb.
4. **Минаев, Е.Ю.** Детектирование и распознавание объектов в системах автомобильной безопасности на основе фрактального анализа / Е.Ю. Минаев, А.В. Никоноров // Компьютерная оптика. – 2012. – Т. 36, № 1. – С. 124-130.
5. **Viola, P.** Detecting Pedestrians Using Patterns of Motion and Appearance / P. Viola, M.J. Jones, D. Snow // International Journal of Computer Vision. – 2005. – Vol. 63(2). – P. 153-161. – DOI: 10.1007/s11263-005-6644-8.
6. **Enzweiler, M.** Monocular Pedestrian Detection: Survey and Experiments / M. Enzweiler, G.M. Darius // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2009. Vol. 31(12). – P. 2179-2195. – DOI: 10.1109/TPAMI.2008.260.
7. **Dalal, N.** Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2005. – P. 886-893. – DOI: 10.1109/CVPR.2005.177.
8. **Cristianini, N.** An introduction to support Vector Machines and other kernel-based learning methods / N. Cristianini, J. Shawe-Taylor. – Cambridge: Cambridge University Press, 2000. – P. 204. – ISBN: 978-0521780193.
9. Лекции по методу опорных векторов [Электронный ресурс] / К.В. Воронцов. – URL: <http://www.ccas.ru/voron/download/SVM.pdf> (дата обращения 10.06.2016).
10. **Казаков, А.** Быстрый алгоритм обнаружения пешеходов по видеоданным / А. Казаков, А. Бовырин // Proceedings of the 22nd International Conference on Computer Graphics and Vision. – 2012. – С. 144-148.
11. **Cho, H.** Real-time pedestrian detection with deformable part models / H. Cho, P.E. Rybski, A. Bar-Hillel, W. Zhang // Intelligent Vehicles Symposium (IV), 2012 IEEE, Alcalá de Henares. – 2012. – P. 1035-1042. – DOI: 10.1109/IVS.2012.6232264.
12. **Hua, Ch.** Pedestrian detection by using spatio temporal histogram of oriented gradients / Ch. Hua, Y. Makihara, Y. Yagi // IEICE Transactions on Information and Systems. – 2013. – Vol. E96-D(6). – P. 1376-1386. – DOI: 10.1587/transinf.E96.D.1376.
13. **Vapnik, V.N.** An overview of statistical learning theory / V.N. Vapnik // IEEE Transactions on Neural Networks. – 1999. – Vol. 10(5). – P. 988-999. – DOI: 10.1109/72.788640.
14. **Воронцов, К.В.** Математические методы обучения по прецедентам (теория обучения машин) [Электронный ресурс] / К.В. Воронцов. – 2011. – URL: <http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf> (дата обращения 21.10.16).
15. **Cireşan, D.** Multi-column deep neural networks for image classification / D. Cireşan, U. Meier, J. Schmidhuber // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. – 2012. – P. 3642-3649. – DOI: 10.1109/CVPR.2012.6248110.
16. **Cireşan, D.** Flexible, high performance convolutional neural networks for image classification / D. Cireşan, U. Meier, J. Masci, L.M. Gambardella, J. Schmidhuber // Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. – 2011. – Vol. 2. – P. 1237-1242. – DOI: 10.5591/978-1-57735-516-8/IJCAI11-210.
17. **Karungaru, S.G.** Face recognition in colour images using neural networks and genetic algorithms / S.G. Karungaru, M. Fukumi, N. Akamatsu // International Journal of Computational Intelligence and Applications. – 2005. – Vol. 5(1). – P. 55-67. – DOI: 10.1142/S1469026805001477.
18. **Солдатова, О.П.** Применение свёрточной нейронной сети для распознавания рукописных цифр / О.П. Солдатова, А.А. Гаршин // Компьютерная оптика. – 2010. – Т. 34, № 2. – С. 252-259.
19. **Verma, A.** Pedestrian detection via mixture of CNN experts and thresholded aggregated channel features / A. Verma, R. Hebbalaguppe, L. Vig, S. Kumar, E. Hassan // Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW '15), Santiago. – 2015. – P. 555-563. – DOI: 10.1109/ICCVW.2015.78.
20. **Ouyang, W.** Joint Deep Learning for Pedestrian Detection / W. Ouyang, X. Wang // Proceedings of the 2013 IEEE International Conference on Computer Vision. – 2013. – P. 2056-2063. – DOI: 10.1109/ICCV.2013.257.
21. **Хайкин, С.** Нейронные сети: Полный курс: Пер. с англ. / С. Хайкин.; пер. с англ. – М.: Вильямс, 2008. – 1103 с. – ISBN: 5-8459-0890-6.
22. **Кохонен, Т.** Самоорганизующиеся карты / Т. Кохонен; пер. с англ. – М.: БИНОМ. Лаборатория знаний, 2008. – 655 с. – ISBN: 5-94774-352-3.
23. **Уоссермен, Ф.** Нейрокомпьютерная техника: Теория и практика / Ф. Уоссермен; пер. с англ. – М.: Мир. 1992. – 240 с.
24. **Martinetz, T.M.** "Neural-gas" network for vector quantization and its application to time series prediction / T.M. Martinetz, S.G. Berkovich, K.J. Schulten // IEEE Transactions on Neural Networks. – 1993. – Vol. 4(4). – P. 558-569. – DOI: 10.1109/72.238311.

25. **Fawcett, T.** An introduction to ROC analysis / T. Fawcett // Pattern Recognition Letters – Special Issue: ROC Analysis in Pattern Recognition. – 2006. – Vol. 27(8). – P. 861-874. – DOI: 10.1016/j.patrec.2005.10.010.
26. **Goto, Y.** CS-HOG: Color similarity-based HOG / Yuhi Goto, Y. Yamauchi, H. Fujiyoshi // Proceedings of 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision. – 2013. – P. 266-271. – DOI: 10.1109/FCV.2013.6485502.
27. **Agoston, M.K.** Computer graphics and geometric modeling: Implementation and algorithms // M.K. Agoston. – London: Springer, 2005. – P. 300-306. – ISBN: 978-1-85233-818-3. – DOI: 10.1007/b138805.
28. **Pizer, S.M.** Adaptive histogram equalization and its variations / S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.T.H. Romeny, J.B. Zimmerman // Computer Vision, Graphics, and Image Processing. – 1987. – Vol. 39(3). – P. 355-368. – DOI: 10.1016/S0734-189X(87)80186-X.
29. Библиотека компьютерного зрения OpenCV [Электронный ресурс]. – URL: <http://docs.opencv.org/> (дата обращения: 10.07.2016).
30. **Beyer, O.** Online semi-supervised growing neural gas / O. Beyer, Ph. Cimiano // International Journal of Neural Systems. – 2012. – P. 425-435. – DOI: 10.1142/S0129065712500232.
31. **Qin, A.K.** Robust growing neural gas algorithm with application in cluster analysis / A.K. Qin, P.K. Suganthan // Neural Networks. – 2004. – Vol. 17(8). – P. 1135-1148. – DOI: 10.1016/j.neunet.2004.06.013.
32. **Beyer, O.** Online labelling strategies for growing neural gas / O. Beyer, P. Cimiano // Proceedings of 12th International Conference on Intelligent Data Engineering and Automated Learning. – 2011. – P. 76-83. – DOI: 10.1007/978-3-642-23878-9_10.
33. **Муравьев, А.С.** Модифицированный алгоритм растущего нейронного газа применительно к задаче классификации / А.С. Муравьев, А.А. Белоусов // Вестник науки Сибири. – 2014. – Т. 14, № 4. – С. 105-111.
34. **Айвазян, С.А.** Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с. – ISBN: 5-279-00054-X.
35. INRIA Person Dataset [Электронный ресурс]. – URL: <http://pascal.inrialpes.fr/data/human/> (дата обращения 10.07.2016).
36. Caltech Pedestrian Detection Benchmark [Электронный ресурс]. – URL: http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/ (дата обращения 10.07.2016).
37. **Amosov, O.S.** High-speed neurofuzzy algorithms for filtering the mobile object trajectory parameters / O.S. Amosov, E.A. Malashevskaya, S.G. Baena // 23rd Saint Petersburg International Conference on Integrated Navigation Systems, ICINS 2016. – 2016. – P. 389-392.
38. **Park, Y.M.** A self-organizing fuzzy logic controller for dynamic systems using a fuzzy auto-regressive moving average model / Y.M. Park, U.C. Moon, K.Y. Lee // IEEE Transactions on Fuzzy Systems. – 1995. – Vol. 3(1). – P. 75-82. – DOI: 10.1109/91.366563.
39. **Jang, J.-S.R.** ANFIS: adaptive-network-based fuzzy inference system / J.-S.R. Jang // IEEE Transactions on Systems, Man and Cybernetics. – 1993. – Vol. 23(3) – P. 665-685. – DOI: 10.1109/21.256541.
40. **Dollar P.** Pedestrian detection: An evaluation of the state of the art / P. Dollar, C. Wojek, B. Schiele, P. Perona // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 34(4). – P. 743-761. – DOI: 10.1109/TPAMI.2011.155.

Сведения об авторах

Амосов Олег Семенович, 1957 года рождения, доктор технических наук, профессор, работает заведующим кафедрой промышленной электроники в ФГБОУ ВО «Комсомольский-на-Амуре государственный технический университет». Область научных интересов: фильтрация и оценивание состояния стохастических систем, оптимальное управление, искусственные нейронные сети, нечеткие системы, навигация и управление движением. E-mail: osa18@yandex.ru.

Иванов Юрий Сергеевич, 1987 года рождения, кандидат технических наук, работает доцентом кафедры промышленной электроники в ФГБОУ ВО «Комсомольский-на-Амуре государственный технический университет». Область научных интересов: программирование, распознавание образов, машинное обучение, электроника, нечеткие системы, искусственные нейронные сети, E-mail: ivanov_ys@garmvis.ru.

Жиганов Сергей Викторович, 1991 года рождения, в 2012 году окончил ФГБОУ ВО «Комсомольский-на-Амуре государственный технический университет» по магистерской программе подготовки 231000 «Программная инженерия», работает системным администратором ИТ-управления ФГБОУ ВО «Комсомольский-на-Амуре государственный технический университет». Область научных интересов: программирование, проектирование программного обеспечения, машинное обучение, искусственные нейронные сети, распознавание образов, нечеткие системы. E-mail: id_zero@list.ru.

ГРНТИ: 28.23.15

Поступила в редакцию 21 июля 2016 г. Окончательный вариант – 20 января 2017 г.

HUMAN LOCALIZATION IN VIDEO FRAMES USING A GROWING NEURAL GAS ALGORITHM AND FUZZY INFERENCE

O.S. Amosov¹, Y.S. Ivanov¹, S.V. Zhiganov¹

¹ Komsomolsk-on-Amur State Technical University, Komsomolsk-on-Amur, Russia

Abstract

A problem of human body localization in video frames using growing neural gas and feature description based on the Histograms of Oriented Gradients is solved. The original neuro-fuzzy

model of growing neural gas for reinforcement learning (GNG-FIS) is used as a basis of the algorithm. A modification of the GNG-FIS algorithm using a two-pass training with fuzzy remarking of classes and building of a heat map is also proposed.

As follows from the experiments, the index of the correct localizations of the developed classifier from 90.5% to 93.2%, depending on the conditions of the scene, that allows the use of the algorithm in real systems of situational video analytics.

Keywords: human localization, growing neural gas, clustering, fuzzy inference.

Citation: Amosov OS, Ivanov YS, Zhiganov SV. Human localization in video frames using a growing neural gas algorithm and fuzzy inference. *Computer Optics* 2017; 41(1): 46-58. DOI: 10.18287/2412-6179-2017-41-1-46-58.

Acknowledgements: The work was supported by the Russian Ministry of Education re-search project - state task in the framework of the project № 2.1898.2017 / PCH "Creating Mathematical and algorithmic support of intellectual information and telecommunication high school security".

References

- [1] Amosov OS, Ivanov YS. Modified algorithm of localization of license plates of vehicles based on the method of Viola-Jones [in Russian]. *Informatics and Control Systems* 2014; 39(1): 127-140.
- [2] Melnikov II, Demidenko SV, Evseenko IA, Emelyanov IA. Motion detection based on pulsed neural networks [In Russian]. *Information Technology* 2013; 7: 57-60.
- [3] Viola P, Jones M. Robust real-time face detection. *International Journal of Computer Vision* 2004; 57(2): 137-154. DOI: 10.1023/B:VISI.0000013087.49260.fb.
- [4] Minaev EY, Nikonov AV. Object detection and recognition in the driver assistance system based on the fractal analysis [in Russian]. *Computer Optics* 2012; 36(1): 124-130.
- [5] Viola P, Jones MJ, Snow D. Detecting Pedestrians Using Patterns of Motion and Appearance. *Int J Comput Vision* 2005; 63(2): 153-161. DOI: 10.1007/s11263-005-6644-8.
- [6] Enzweiler M, Dariu MG. Monocular Pedestrian Detection: Survey and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009; 31(12): 2179-2195. DOI: 10.1109/TPAMI.2008.260.
- [7] Dalal N, Triggs B. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2005: 886-893. DOI: 10.1109/CVPR.2005.177.
- [8] Cristianini N, Shawe-Taylor J. *An introduction to support Vector Machines and other kernel-based learning methods*. Cambridge: Cambridge University Press; 2000. ISBN: 978-0521780193.
- [9] Lectures on the method of support vector machine [In Russian]. Source: <<http://www.ccas.ru/voron/download/SVM.pdf>>.
- [10] Kazakov A, Bovyryn A. Fast algorithm for the detection of pedestrians on the video data [in Russian]. *The 22nd International Conference on Computer Graphics and Vision* 2012; 144-148.
- [11] Cho H, Rybski PE, Bar-Hillel A and Zhang W. Real-time pedestrian detection with deformable part models. *Intelligent Vehicles Symposium (IV)*, 2012 IEEE, Alcalá de Henares 2012: 1035-1042. DOI: 10.1109/IVS.2012.6232264.
- [12] Hua C, Makihara Y, Yagi Y. Pedestrian detection by using spatio temporal histogram of oriented gradients. *IEICE Transactions on Information and Systems* 2013; E96-D(6): P. 1376-1386. DOI: 10.1587/transinf.E96.D.1376.
- [13] Vapnik VN. An overview of statistical learning theory. *IEEE Transactions on Neural Networks* 1999; 10(5): 988-999. DOI: 10.1109/72.788640.
- [14] Vorontsov KV. Mathematical methods of training on precedents (machine learning theory) [In Russian]. Source: <<http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf>>.
- [15] Cireşan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. *CVPR '12* 2012: 3642-49. DOI: 10.1109/CVPR.2012.6248110.
- [16] Cireşan D, Meier U, Masci J, Gambardella L, Schmidhuber J. Flexible, high performance convolutional neural networks for image classification. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI '11)* 2011; 2: 1237-1242. DOI: 10.5591/978-1-57735-516-8/IJCAI11-210.
- [17] Karungaru SG, Fukumi M, Akamatsu N. Face recognition in colour images using neural networks and genetic algorithms. *International Journal of Computational Intelligence and Applications* 2005; 5(1): 55-67; DOI: 10.1142/S1469026805001477.
- [18] Soldatova OP, Garshin AA. Convolutional neural network applied to handwritten digits recognition [In Russian]. *Computer Optics* 2010; 34(2): 252-259.
- [19] Verma A, Hebbalaguppe R, Vig L, Kumar S, Hassan E. Pedestrian detection via mixture of CNN experts and thresholded aggregated channel features. *ICCVW '15* 2015: 555-563. DOI: 10.1109/ICCVW.2015.78.
- [20] Ouyang W, Wang X. Joint deep learning for pedestrian detection. *ICCV '13* 2013: 2056-2063. DOI: 10.1109/ICCV.2013.257.
- [21] Haykin S. *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice Hall PTR; 1998. ISBN: 0132733501.
- [22] Kohonen T. *Self-organizing maps*. Berlin, Heidelberg: Springer-Verlag; 2001. ISBN: 978-3-540-67921-9.
- [23] Wasserman PD. *Neural computing: theory and practice*. New York, NY, USA: Van Nostrand Reinhold Co.; 1989. ISBN: 0-442-20743-3.
- [24] Martintez TM, Berkovich SG, Schulten KJ. "Neural-gas" network for vector quantization and its application to time series prediction. *IEEE Transactions on Neural Networks* 1993; 4(4): 558-569. DOI: 10.1109/72.238311.
- [25] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters – Special issue: ROC analysis in pattern recognition* 2006; 27(8): 861-874. DOI: 10.1016/j.patrec.2005.10.010.
- [26] Goto Y, Yamauchi Y, Fujiyoshi H. CS-HOG: Color similarity-based HOG. *FCV* 2013: 266-271. DOI: 10.1109/FCV.2013.6485502.
- [27] Agoston MK. *Computer graphics and geometric modeling: Implementation and algorithms*. London: Springer; 2005. ISBN: 978-1-85233-818-3. DOI: 10.1007/b138805.
- [28] Pizer SM, Amburn EP, Austin JD, et al. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* 1987; 39(3): 355-368. DOI: 10.1016/S0734-189X(87)80186-X.

- [29] The image processing library OpenCV. Source: <<http://docs.opencv.org/>>.
- [30] Beyer O, Cimiano P. Online semi-supervised growing neural gas. *International Journal of Neural Systems* 2012; 22(5): 425-435. DOI: 10.1142/S0129065712500232.
- [31] Qin AK, Suganthan PK. Robust growing neural gas algorithm with application in cluster analysis. *Neural Networks* 2004; 17(8): 1135-1148. DOI: 10.1016/j.neunet.2004.06.013.
- [32] Beyer O, Cimiano P. Online labelling strategies for growing neural gas. *IDEAL* 2011: 76-83. DOI: 10.1007/978-3-642-23878-9_10.
- [33] Muravev AS, Belousov AA. Modified algorithm of growing neural gas applied to the problem of classification [In Russian]. *Siberia Science Bulletin* 2014; 4(14): 105-111.
- [34] Ayvazyan SA, Buchstaber VM, Enyukov IS, Meshalkin LD. Applied Statistics: Classification and reduction of dimension [In Russian]. Moscow: "Finansy i Statistika" Publisher; 1989.
- [35] INRIA Person Dataset. Source: <<http://pascal.inrialpes.fr/data/human/>>.
- [36] Caltech Pedestrian Detection Benchmark. Source: <http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/>.
- [37] Amosov OS, Malashevskaya EA, Baena SG. High-speed neurofuzzy algorithms for filtering the mobile object trajectory parameters. 23rd Saint Petersburg International Conference on Integrated Navigation Systems, ICINS 2016: 389-392.
- [38] Park YM, Moon UC, Lee KY. A self-organizing fuzzy logic controller for dynamic systems using a fuzzy autoregressive moving average model. *IEEE Transactions on Fuzzy Systems* 1995; 3(1): 75-82. DOI: 10.1109/91.366563.
- [39] Jang J.-S.R. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics* 1993; 23(3): 665-685. DOI: 10.1109/21.256541.
- [40] Dollar P, Wojek C, Schiele B, Perona P. Pedestrian detection: An evaluation of the state of the art. *IEEE PAMI* 2012; 34(4): 743-761. DOI: 10.1109/TPAMI.2011.155.

Authors' information

Oleg Semenovich Amosov, (b. 1957), doctor of Technical Sciences, professor, working head of Industrial Electronics department in Komsomolsk-on-Amur State Technical University. Research interests: filtering and estimation of the state of stochastic systems, optimal control, artificial neural networks, fuzzy systems, navigation and motion control. E-mail: osa18@yandex.ru.

Yuri Sergeevich Ivanov, (b. 1987), PhD in Technical Sciences, work assistant professor of Industrial Electronics department in Komsomolsk-on-Amur State Technical University. Research interests: programming, pattern recognition, machine learning, electronics, fuzzy systems, artificial neural networks. E-mail: ivanov_ys@garmvis.ru.

Sergey Viktorovich Zhiganov, (b. 1991), graduated from Komsomolsk-on-Amur State Technical University for the Master's program of training 231000 "Software Engineering" in 2012 year, working a system administrator IT Management Komsomolsk-on-Amur State Technical University. Research interests: programming, software design, machine learning, artificial neural networks, recognition of images, fuzzy systems. E-mail: id_zero@list.ru.

Received July 21, 2016. The final version – January 20, 2017.