

СЕЛЕКТИВНО-КОВАРИАЦИОННЫЙ МЕТОД ЛОКАЛИЗАЦИИ, КЛАССИФИКАЦИИ И ОТСЛЕЖИВАНИЯ ЛЮДЕЙ В ВИДЕОПОТОКАХ ОТ МНОЖЕСТВА ВИДЕОКАМЕР

В.В. Девятков

deviatkov@comtv.ru

А.Н. Алфимцев

alfim@bmstu.ru

А.Р. Таранян

taranyan.bmstu@gmail.com

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Приведен селективно-ковариационный метод локализации, классификации и отслеживания людей в видеопотоках от нескольких камер. Подобные методы являются ключевыми при разработке систем безопасности и видеонаблюдения, интеллектуального окружения и робототехники. Метод назван селективно-ковариационным, поскольку прежде, чем относить объект к тому или иному классу (в нашем случае к тому или иному классу людей) на основе применения ковариационных дескрипторов, осуществляется выделение (селекция) специфической области, характерной для рассматриваемого класса объектов (людей) — это области головы и плеч человека. Предложены новые функции признаков для построения ковариационных дескрипторов, проведено сравнение эффективности применения этих функций с базовыми функциями признаков. Предложена и апробирована маска, позволяющая выделять наиболее информативные пиксели области интереса. Использование предложенных функций признаков и маски для выделения информативной области интереса позволило значительно улучшить достоверность классификации людей (с 75 % при применении базовых функций признаков до 94,6 % по предложенному методу) при невысокой вычислительной сложности

Ключевые слова

Распознавание образов, компьютерное зрение, отслеживание людей, ковариационная матрица, ковариационный дескриптор, селективная локализация

Поступила в редакцию 21.04.2016

© МГТУ им. Н.Э. Баумана, 2016

Устойчивое отслеживание подвижных объектов в реальном времени с использованием множества камер, каждая из которых имеет сравнительно небольшое поле зрения, является непростой задачей, решение которой становится особенно проблематичным при увеличении скорости перемещения объекта и его временном исчезновении. Отслеживание подвижных объектов рекурсивно: суть его состоит в предварительной локализации (выделении) объекта на начальном кадре видеопотока, нахождении его характерных признаков и последующем отслеживании этого объекта по вычисленным для начального кадра признакам.

Прежде чем переходить к постановке задачи настоящей статьи и сравнительному анализу ее с известными работами, введем ряд понятий.

Плоское изображение (в нашем случае прямоугольное), формируемое цифровой видеокамерой в момент времени t , будем называть *кадром*. Кадр, получаемый в момент времени t видеокамерой c , координаты пикселей которого в декартовой системе координат принадлежат соответственно множествам V по горизонтали и W по вертикали, обозначим $I_t^c(V, W)$. Отдельный пиксель в кадре будем обозначать $I_t^c(v, w)$, $v \in V$, $w \in W$. *Видеопотоком* будем называть последовательность кадров $I_t^c(V, W)$, $I_{t+1}^c(V, W)$, ..., $I_{t+k}^c(V, W)$. Записями $I_{Rt}^c(v, w)$, $I_{Gt}^c(v, w)$ и $I_{Bt}^c(v, w)$ будем обозначать соответственно компоненты R , G и B пикселя $I_t^c(v, w)$ кадра $I_t^c(V, W)$.

Множество пикселей кадра $I_t^c(V, W)$, содержащих интересующий на объект (например, лицо человека), назовем областью интереса. Область интереса кадра $I_t^c(V, W)$ обозначим $R_t^c(V_R, W_R)$, где $V_R \subseteq V$, $W_R \subseteq W$. Множество пикселей кадра, не принадлежащих области интереса, назовем *фоном*.

Локализацией объекта называется нахождение области интереса, содержащей этот объект, в кадре в момент времени t .

Отслеживанием объекта называется процесс последовательной локализации в кадрах $I_t^c, I_{t+1}^c, \dots, I_{t+k}^c$ областей интереса $R_t^c, R_{t+1}^c, \dots, R_{t+k}^c$, содержащих отслеживаемый объект.

Классификацией областей интереса назовем неявное разбиение множества всех этих областей на подмножества, называемые классами. Для классификации используется дескриптор класса, изначально вычисляемый на основе определенных признаков выбранного объекта, принадлежащего классу, и используемый в дальнейшем для локализации и классификации как эталон, с которым сравнивается вычисляемый по тем же признакам и алгоритму дескриптор, называемый образцом. Сравнение осуществляется путем вычисления значения критерия близости эталона к образцу. Если значение критерия близости не превышает определенного порога, то считается, что объект принадлежит классу дескриптора класса. Классификации области интересов посвящено много различных методов, основные отличия которых связаны с типами используемых дескрипторов. Все эти методы обычно подразделяют на две больших группы.

Методы первой группы при формировании дескриптора используют нахождение ключевых точек, выделение специфических признаков объекта, наилучшим образом характеризующих его геометрические особенности (углы, дуги, линии, выпуклости, вогнутости, формы контура и т. д.). Среди методов данной группы можно выделить использование дескрипторов SIFT в работах [1, 2], основанных на выделении признаков, инвариантных к повороту и масштабированию, метод *shape context* [3], нахождение характерных частей (в частности, частей тела человека [4]), силуэтов [5] и т. п. Существенным недостатком данных

методов является требование высокого разрешения классифицируемых объектов, тогда как разработанный в настоящей работе метод может эффективно функционировать на видеопотоках относительно небольшого разрешения.

Дескрипторы методов второй группы для классификации области интересов используют низкоуровневые признаки (градиенты, цвета, интенсивности, положения и т. д.), вычисляемые по всей области интересов. Наиболее популярными дескрипторами данной группы являются гистограммы [6–8] и ковариационные матрицы [9–12]. Например, в работе [6] в качестве дескриптора используется нечеткая гистограмма. Каждый интервал этой гистограммы соответствует определенному цвету, а высота столбца на каждом интервале пропорциональна количеству пикселей соответствующего интервала цвета. Количество этих пикселей находится путем нечеткой кластеризации пикселей по цвету. Недостатком гистограмм как дескрипторов, по крайней мере, в известных работах, является отбрасывание пространственной информации при построении гистограмм, что делает невозможным классификацию объектов схожих цветов, но с разной геометрией. Ковариационная матрица в отличие от гистограмм, в принципе, может быть построена для любого количества признаков как цветовых, так и пространственных, обеспечивая при этом невысокую вычислительную сложность. В настоящей статье вследствие указанных достоинств именно ковариационная матрица выбрана как дескриптор для создания метода локализации, классификации и отслеживания людей несколькими камерами.

Отметим, что предлагаемый метод должен уметь без предварительного обучения отслеживать одного и того же человека с разных камер и в разных местоположениях, под разными углами, с разным фоном. В отличие от отслеживания объектов с одной камеры, значительно ограничивается применение методов, основанных на предположении, что отслеживаемый объект на соседних кадрах видеопотока должен иметь близкие координаты [5]. Если при отслеживании с одной камеры известные методы позволяют получать хорошие результаты, то в нашем случае применение подобных методов возможно лишь в одном отдельно взятом видеопотоке. В нашем случае также представляется ограниченным применение адаптивного обновления дескрипторов объектов, которое используется в некоторых методах при небольших изменениях таких параметров отслеживаемого объекта, как положение относительно камеры, освещение и фон, поскольку указанные параметры могут сильно различаться для изображений объекта, полученных с разных видеокамер.

Предлагаемый метод основан на применении классификатора Виолы — Джонса для локализации всех областей головы и плеч людей и дальнейшем сопоставлении (классификации) локализованных областей с ранее обнаруженными людьми посредством ковариационного дескриптора. Голова и плечи выбраны как область, достаточная информативная для классификации людей, но в то же время в наименьшей степени перекрываемая другими объектами.

Алгоритм ковариационного метода отслеживания несколькими камерами. Алгоритм включает следующие шаги.

1. Инициализация пустого списка D ковариационных дескрипторов отслеживаемых людей.

2. Считывание кадров $I_t^1, I_t^2, \dots, I_t^p$ со всех камер в момент времени t .

3. Локализация всех областей головы и плеч R_1, R_2, \dots, R_m посредством алгоритма Виолы — Джонса на кадрах $I_t^1, I_t^2, \dots, I_t^p$.

4. Построение ковариационных дескрипторов d_1, d_2, \dots, d_m для областей R_1, R_2, \dots, R_m , локализованных в момент времени t .

5. Сопоставление построенных в момент времени t дескрипторов с дескрипторами, внесенными в список D в моменты времени $1, 2, \dots, t-1$, обновление списка D , а именно, для каждого дескриптора $d_i, i \in 1, t$:

5.1. Нахождение дескриптора d^* из списка D , расстояние $Dist(d^*, d_i)$ которого от d_i минимально;

5.2. Если это расстояние меньше определенного порога, то фиксируется обнаружение в кадре человека, соответствующего дескриптору d^* , если же оно больше или равно этому порогу, то фиксируется обнаружение нового человека, и соответствующий ему дескриптор d_i добавляется в список дескрипторов обнаруженных людей D .

На рис. 1 приведена блок-схема алгоритма считывания и обработки кадров с видеочамер в момент времени t .

В предложенном методе можно выделить два модуля — модуль локализации областей головы и плеч и модуль классификации (далее — классификатор). Модуль локализации отвечает за третий шаг представленного алгоритма, а классификатор — за четвертый и пятый шаги. Классификатор, по сути, является совокупностью выбранного метода построения дескриптора и способа сопоставления дескрипторов.

Рассмотрим подробнее ключевые этапы предложенного алгоритма ковариационного метода отслеживания людей в видеопотоках с нескольких камер и детально обсудим предложенный классификатор.

Локализация областей головы и плеч. Локализация областей головы и плеч (областей интереса) в момент времени t производится для каждого видеопотока по-отдельности, затем найденные области склеиваются в один общий набор областей R_1, R_2, \dots, R_m , который и является результатом работы данного этапа. Локализация областей интереса в каждом отдельно взятом видеопотоке выполняется в два этапа. На первом этапе посредством алгоритма вычитания фона на основе смеси гауссовых распределений [13] (метод *BackgroundSubtractorMOG2* библиотеки OpenCV) отбрасываются статические области кадра, после чего с помощью алгоритма Виолы — Джонса [14] на оставшихся областях кадра локализуются области головы и плеч (используется каскад *haarcascade_mcs_upperbody* из библиотеки OpenCV). Алгоритм Виолы — Джонса был выбран ввиду его высокой производительности и хорошего качества работы, благодаря чему он является одним из наиболее широко применяемых алгоритмов в задачах локализации объектов.

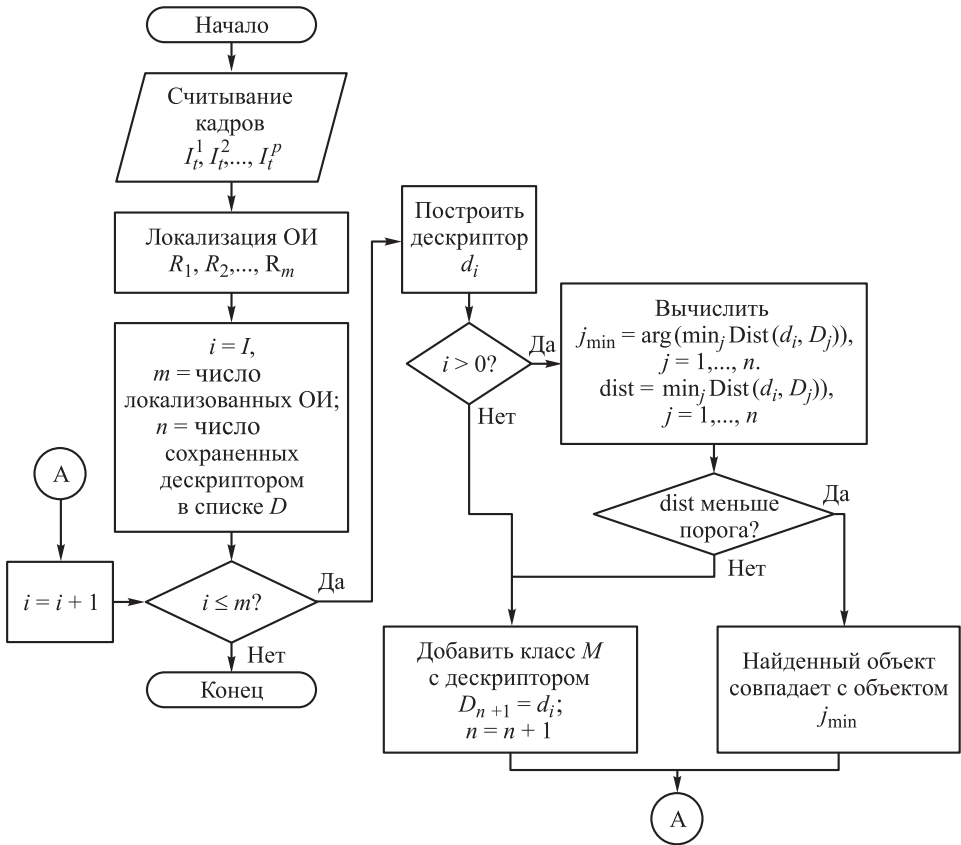


Рис. 1. Блок-схема алгоритма локализации и сопоставления людей на кадрах видеопотоков

Применение вычитания фона позволяет повысить качество локализации благодаря уменьшению числа ложно-положительных срабатываний на статических областях кадра (на которых либо нет людей, либо те уже были локализованы при входе в кадр), а также в некоторых сценариях увеличить скорость локализации за счет сканирования классификатором лишь части кадра. Отметим, что для работы алгоритма вычитания фона ему на вход необходимо передавать последние $N > 1$ кадров видеопотока.

Классификатор. Построение ковариационных дескрипторов. В настоящей работе в качестве дескриптора области интереса предлагается использование ковариационной матрицы, описанной в работе [9]. Ковариационная матрица, как было отмечено ранее, позволяет кодировать взаимосвязи различных признаков области интереса, среди которых основными, безусловно, являются цветовая и пространственная информация.

Ковариационная матрица для области интереса R строится следующим образом.

1. Для каждого пикселя $p_i \in R$ посредством заранее predetermined функции признаков F вычисляется вектор признаков $f_i = F(p_i)$ размерности d , кото-

рый содержит информацию о пикселе i , возможно, об определенной области вокруг рассматриваемого пикселя. Функция F может иметь следующий вид:

$$F(x, y) = [xy I_R(x, y) I_G(x, y) I_B(x, y) \partial_x(x, y) \partial_y(x, y)],$$

где x, y — координаты пикселя; $I_R(x, y), I_G(x, y), I_B(x, y)$ — значения компонент R, G и B пикселя с координатами (x, y) ; ∂_x и ∂_y — первые производные интенсивности в точке (x, y) по горизонтали и по вертикали соответственно.

2. Для полученного множества F_R векторов признаков области R вычисляется средний вектор f_{mean} и строится ковариационная матрица размера $d \times d$:

$$C_R = \frac{1}{n-1} \sum_{i=1}^n (f_i - f_{mean})(f_i - f_{mean})^T,$$

где n — количество пикселей в области R .

На рис. 2 представлены ковариационные матрицы, построенные для двух различных изображений с использованием функции признаков (1). Оба изображения имеют высоту в 60 пикселей и ширину в 55 пикселей.

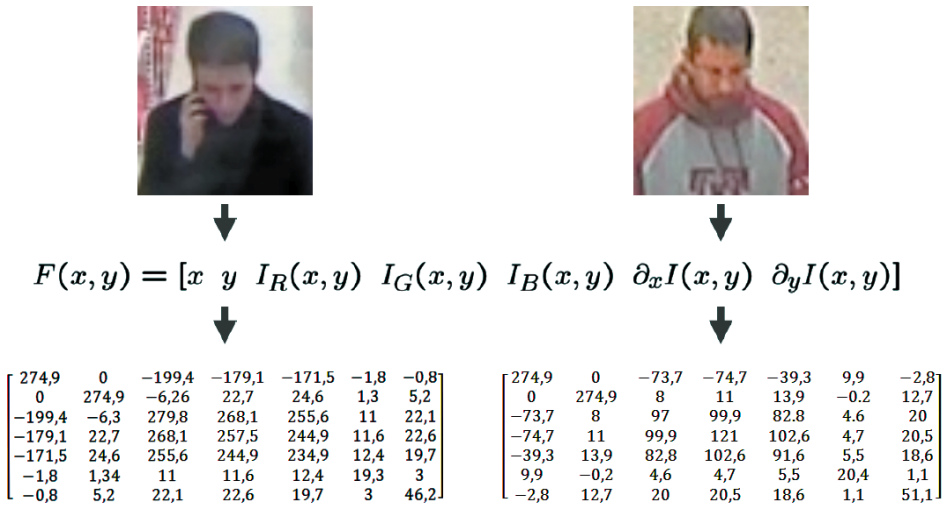


Рис. 2. Ковариационные матрицы для двух изображений

Очевидно, что ключевым аспектом при построении ковариационной матрицы является выбор функции признаков F . Вопросу выбора функции F в настоящей работе посвящен отдельный раздел, следующий далее.

Предложенный дескриптор можно улучшить, проанализировав значимость отдельных пикселей области интереса при построении ковариационной матрицы. Поскольку прямоугольная область (рис. 3), которую находит алгоритм Виолы — Джонса при поиске головы и плеч человека, содержит достаточно большие области, являющиеся, по сути, фоном, целесообразно строить ковариационную матрицу лишь для тех пикселей, которые принадлежат областям головы и плеч человека.

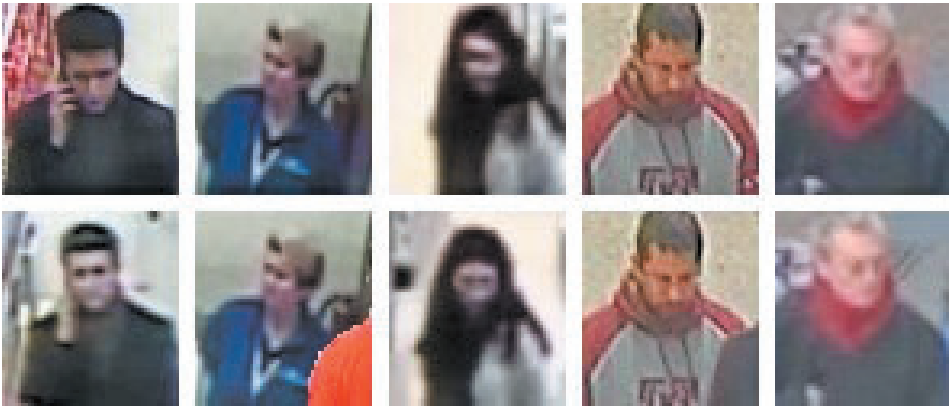


Рис. 3. Области, локализованные алгоритмом Виолы — Джонса



Рис. 4. Маска изображения. Черные пиксели соответствуют нулям, а белые — единицам

Для этого посредством разработанного жадного алгоритма на обучающей выборке была вычислена бинарная маска M размера 60×55 пикселей, нули в которой обозначают пиксели, которые не репрезентативны при построении дескриптора, а единицы — те пиксели, которые следует использовать при построении ковариационной матрицы (рис. 4).

Таким образом, предлагается строить ковариационную матрицу C_R области интереса R на основе векторов признаков лишь тех пикселей, которым соответствует единица в маске M .

Классификатор. Сопоставление дескрипторов. Для определения «похожести» дескрипторов двух областей интереса необходимо определить метрику для ковариационных матриц. В настоящей работе была проанализирована эффективность применения евклидовой меры D_{eucl} и меры D_{eigen} , основанной на вычислении обобщенных собственных значений для сравниваемых ковариационных матриц [9].

Для ковариационных матриц C_1 и C_2 мера D_{eucl} определяется следующим образом:

$$D_{eucl}(C_1, C_2) = \sqrt{\sum_{i=1}^d \sum_{j=1}^d (C_1(j, i) - C_2(j, i))^2}.$$

Вычисление метрики D_{eucl} имеет сложность $O(d^2)$.

Метрика D_{eigen} вычисляется следующим образом:

$$D_{eigen}(C_1, C_2) = \sqrt{\sum_{i=1}^n \ln^2 \lambda_i(C_1, C_2)},$$

где $\{\lambda_i(C_1, C_2)\}_{i=1, \dots, n}$ — ненулевые обобщенные собственные значения для матриц C_1 и C_2 , вычисленные для уравнения

$$\lambda_i C_1 x_i - C_2 x_i = 0, \quad i = 1, \dots, d.$$

Отметим, что неотрицательность обобщенных собственных значений $\{\lambda_i(C_1, C_2)\}_{i=1, \dots, n}$ следует из того, что ковариационные матрицы — положительно полуопределенные матрицы.

Вычисление метрики D_{eigen} имеет сложность $O(d^3)$, обусловленную вычислением обобщенных собственных значений.

Классификатор. Выбор функции признаков. Поскольку репрезентативность ковариационной матрицы напрямую определяется выбором функции признаков $F(x, y)$, было проведено подробное сравнение эффективности применения раннее предложенных функций признаков, а также были разработаны и апробированы новые функции признаков.

Для удобства компоненты вектора, определяемого функцией признаков $F(x, y)$, были разделены на два подмножества:

$$F(x, y) = [a_1(x, y), \dots, a_t(x, y)b_1(x, y), \dots, b_p(x, y)] = [A(x, y)B(x, y)],$$

где $a_i(x, y) (i = 1, \dots, t)$ представляет информацию о цвете, а $b_i(x, y) (i = 1 \dots p)$ — пространственную информацию, причем $t + p = d$.

Для цветового компонента A были рассмотрены две тривиальные схемы, представляющие собой информацию о цвете пикселя в системах RGB и HSV :

$$A_{RGB}(x, y) = [(x, y)I_R(x, y)I_G(x, y)I_B(x, y)];$$

$$A_{HSV}(x, y) = [I_H(x, y)I_S(x, y)I_V(x, y)].$$

Однако тестирование данных схем показало их недостаточную эффективность. Для повышения информативности кодируемых данных о цветовых характеристиках области интереса была разработана схема $A_{RGBHistN}$, основанная на гистограммах. Для построения векторов признаков данной схемы область интереса разбивается на сетку с пятью клетками по горизонтали и пятью клетками по вертикали. Далее, для каждой клетки данной сетки по компонентам R , G и B строится нечеткая гистограмма с N ячейками, после чего строятся сами векторы признаков:

$$A_{RGBHistN}(x, y) = [I_R(x, y)I_G(x, y)I_B(x, y) \\ HR_1(x, y) \dots HR_N(x, y) \\ HG_1(x, y) \dots HG_N(x, y) \\ HB_1(x, y) \dots HB_N(x, y)],$$

где $HR_i(x, y)$, $HG_i(x, y)$ и $HB_i(x, y)$ — значения i -х ячеек гистограмм, содержащих пиксель (x, y) по компонентам R , G и B соответственно.

На рис. 5 приведена гистограмма для одной из ячеек области интереса (представляющей собой голову и плечи человека в зеленой верхней одежде), построенная для вектора признаков $A_{RGBHist5}$.

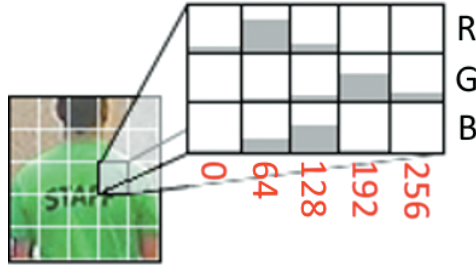


Рис. 5. Гистограмма одной из ячеек области интереса для вектора признаков

В качестве пространственного компонента B функции признаков F были рассмотрены схемы B_{Radial} , B_{Coord} , B_{Deriv} и предложена новая схема $B_{RadialGrid}$.

Значение вектора B_{Radial} для пикселя представляет собой удаление этого пикселя от центра области интереса

$$B_{Radial}(x, y) = \left[\sqrt{(x - x_0)^2 + (y - y_0)^2} \right],$$

где x_0 и y_0 координаты x и y центра области интереса.

Очевидно, что данная схема инвариантна к поворотам области интереса.

Схема B_{Coord} кодирует информацию о координатах x и y

$$B_{Coord}(x, y) = [x \ y].$$

Схема B_{Deriv} , кроме координат x и y , также кодирует информацию о первых производных интенсивности пикселей области интереса в точке (x, y) по осям x и y

$$B_{Deriv}(x, y) = [x \ y \ \partial_x(x, y) \ \partial_y(x, y)].$$

Разработанная схема $B_{RadialGrid}$ рассчитана на кодирование информации о зависимости цветовой информации от области интереса.

Для этого области интереса разбивается сеткой на пять горизонтальных и пять вертикальных областей, и схема $B_{RadialGrid}$ определяется следующим образом:

$$B_{RadialGrid}(x, y) = \left[\sqrt{(x - x_0)^2 + (y - y_0)^2} \right. \\ \left. H_1(x, y)H_2(x, y)H_3(x, y)H_4(x, y)H_5(x, y) \right. \\ \left. V_1(x, y)V_2(x, y)V_3(x, y)V_4(x, y)V_5(x, y) \right],$$

где $H_i(x, y) = 1$, если пиксель (x, y) принадлежит i -й горизонтальной области, и $H_i(x, y) = 0$ в обратном случае, и аналогично $V_j(x, y) = 1$, если пиксель (x, y) принадлежит j -й вертикальной области, и $V_j(x, y) = 0$, если он не принадлежит j -й вертикальной области.

Таким образом, комбинируя рассмотренные цветовые и пространственные схемы, было сформировано 16 функций признаков для построения ковариационного дескриптора области интереса.

Как было отмечено ранее, для построения классификатора необходимо выбрать метод построения дескриптора области интереса и метод сопоставления построенных дескрипторов. В настоящей работе было подробно описано построение ковариационного дескриптора, предложена маска, получено 16 различных функций признаков для построения ковариационного дескриптора. Также были рассмотрены две различные метрики для сопоставления ковариационных дескрипторов. Комбинируя функции признаков, применение либо не применение маски, а также используемую метрику, мы получаем 64 различных классификатора для классификации областей головы и плеч. В следующем разделе приведены результаты тестирования данных классификаторов, на основании которых выбран наиболее эффективный классификатор для решаемой задачи.

Экспериментальные результаты. Для тестирования разработанного метода классификации людей на основе изображений их головы и плеч была сформирована тестовая выборка из 413 изображений головы и плеч 93 различных людей (далее 93 классов). Данная выборка была создана на основе видеозаписей,



Рис. 6. Примеры четырех различных классов тестовой выборки

полученных авторами, тестовой базы изображений людей PETS 2006 Университета Рединга, а также из изображений из сети интернет. На рис. 6 представлены тестовые изображения четырех различных людей.

Элементы всех 93 классов сравнивались внутри классов между собой, после чего строились всевозможные пары классов и сравнивались элементы между классами этих пар.

Классификаторы тестировались следующим образом. Для каждого из классификаторов подбирался порог $dist$, для которого ошибка классификатора минимальна, причем при фиксированном пороге ошибка вычислялась как среднее значение процента ложно-положительных срабатываний и процента ложно-негативных срабатываний.

На рис. 7 приведено сравнение работы классификатора с функцией признаков $A_{RGB}B_{Radial}$ с применением метрики D_{eigen} при использовании предложенной маски, и без ее использования.

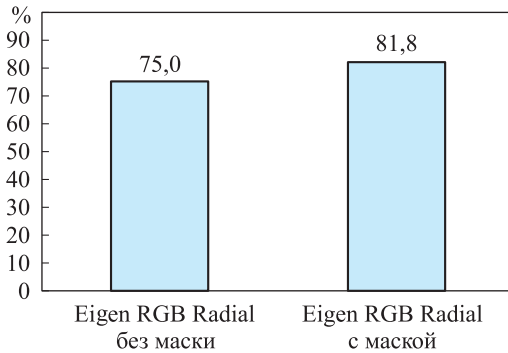


Рис. 7. Сравнение качества классификации с применением маски и без нее

Тестирование показало значительный прирост эффективности классификатора при использовании предложенной маски. Все дальнейшие тесты проводились с использованием маски.

Далее было проведено сравнение эффективности использования метрик D_{eucl} и D_{eigen} (рис. 8).

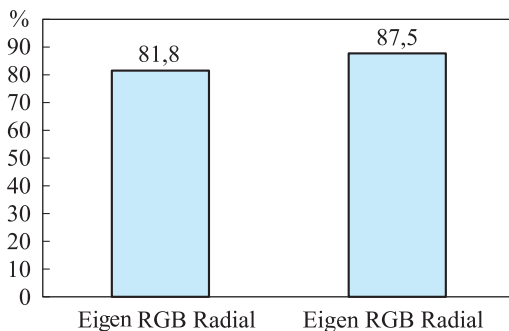


Рис. 8. Сравнение качества классификации при использовании метрик D_{eucl} и D_{eigen}

Метрика D_{eucl} являясь более быстрой, также показала лучшее качество классификации. При использовании этой метрики удалось правильно классифицировать 87,5 % пар. Дальнейшие тесты проводились с использованием метрики D_{eucl} .

На рис. 9 приведены результаты сравнения 16 классификаторов, основанных на функциях признаков, представляющих собой все возможные комбинации четырех цветовых схем A и четырех пространственных схем B . Результаты тестирования классификаторов на приведенном рисунке сгруппированы по цветовой схеме A .

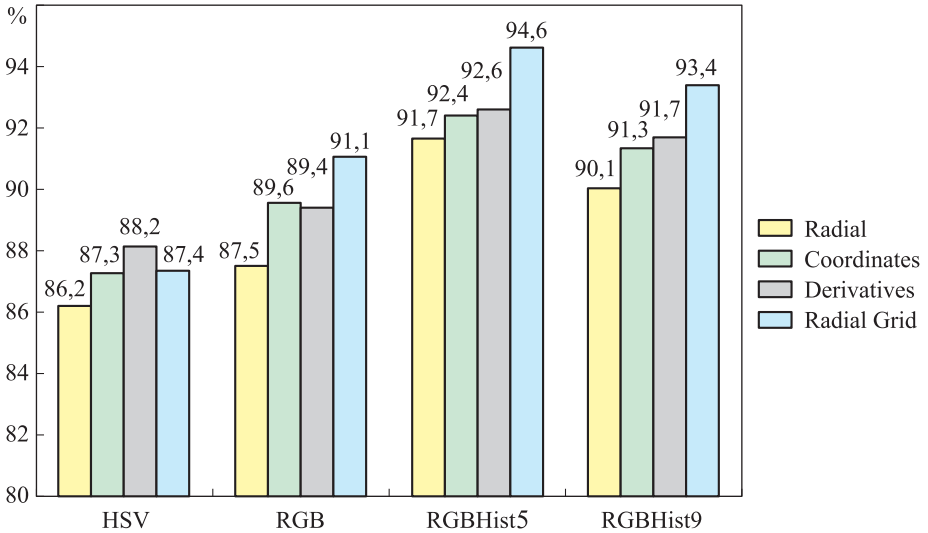


Рис. 9. Сравнение качества классификации для рассматриваемых функций признаков

Тестирование показало, что среди функций признаков, основанных на цветовой схеме A_{HSV} , наилучший результат у функции $A_{HSV}B_{Deriv}$ — 88,2 % верно классифицированных пар.

Наилучший результат из функций признаков, основанных на A_{RGB} , позволила получить схема $A_{RGB}B_{RadialGrid}$ — 91,1 % верно классифицированных пар.

Среди функций признаков на основе $A_{RGBHist5}$ и $A_{RGBHist9}$ лучший результат у функции признаков $A_{RGBHist5}B_{Radial}$ — 94,6 %. Отметим, что результаты тестирования показали нецелесообразность увеличения размерности более 5 в схеме $A_{RGBHistN}$.

На рис. 10 приведены сводные результаты сравнения функций признаков, где для каждой цветовой схемы указан наилучший результат, который удалось получить при ее применении.

Применение предложенной маски, выбор подходящей метрики и применение разработанной в настоящей работе функции признаков позволили получить качество классификации в 94,6 % на достаточно сложной тестовой выборке, что почти на 20 % процентов лучше изначального классификатора (75 %),

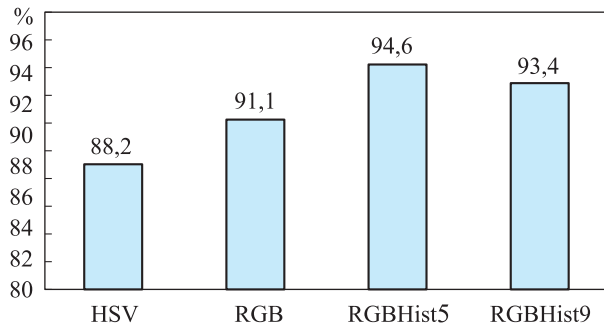


Рис. 10. Сводная таблица функций признаков

в котором маска не применялась, использовались метрика D_{eigen} и функция признаков $A_{RGB}B_{Radial}$.

Заключение. В работе рассмотрена задача отслеживания, локализации и классификации людей в видеопотоках с нескольких камер, решение которой является ключевым при разработке систем видеонаблюдения и обеспечения безопасности [15], интеллектуальных мультимодальных интерфейсов [16] и роботов, взаимодействующих с людьми [17]. Предложен метод локализации людей на основе селективной маски, классификации на основе ковариационных дескрипторов, отслеживании в видеопотоках с нескольких камер, в различных местоположениях, под различными углами, с разным «фоном», позволивший повысить от 75 до 94,6 % эффективность классификации людей, что является весьма хорошим результатом, учитывая сложность тестовых данных. Ключевой особенностью предложенного метода является возможность классифицировать людей на основе ковариационного дескриптора, минуя этап его обучения.

В процессе разработки метода, изложенного в настоящей статье, была апробирована маска, позволяющая отфильтровывать значительную часть тех пикселей ОИ, которые являются фоновыми, была выбрана наиболее подходящая для данной задачи метрика для ковариационных дескрипторов, рассмотрены популярные функции признаков, разработаны новые и проведен детальный экспериментальный анализ эффективности их применения.

Однако вычислительная сложность и вероятность ложной локализации объекта резко возрастают при увеличении скорости движения объекта, попадании его в поле зрения нескольких камер в различных местоположениях и под разными углами с разным фоном, увеличении размерности кадра и числа камер. Учитывая это, на основе метода пространственно-временного анализа ситуаций на множестве движущихся объектов, изложенного в работе [18], в целях более полного анализа поведения людей и снижения вычислительных затрат предполагается дальнейшее развитие методов множественной локализации областей интересов отдельных объектов, формирования дескрипторов для этих областей, введения отношений между классифицированными объектами и моделями их динамического поведения.

ЛИТЕРАТУРА

1. Lowe D.G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. 2004. Vol. 60. No. 2. P. 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94
2. Fazli S., Pour H.M., Bouzari H. Particle filter based object tracking with sift and color feature // International Conference on Machine Vision. 2009. P. 89–93. DOI: 10.1109/ICMV.2009.47
3. Belongie S., Malik J., Puzicha J. Shape matching and object recognition using shape contexts // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. Vol. 24. No. 4. P. 509–522. DOI: 10.1109/34.993558
4. Ioffe S., Forsyth D.A. Probabilistic methods for finding people // International Journal of Computer Vision. 2001. Vol. 43. No. 1. P. 45–68. DOI: 10.1023/A:1011179004708
5. Elzein H., Lakshmanan S., Watta P. A motion and shape-based pedestrian detection algorithm // IEEE Intelligent Vehicles Symposium. Proc. 2003. P. 500–504. DOI: 10.1109/IVS.2003.1212962
6. Liu H., Wang L., Sun F. Mean-shift tracking using fuzzy coding histogram // International Journal of Fuzzy Systems. 2014. Vol. 16. No. 4. P. 457–467.
7. Comaniciu D., Ramesh V., Meer P. Kernel-based object tracking // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003. Vol. 25. No. 5. P. 564–577. DOI: 10.1109/TPAMI.2003.1195991
8. Perez P., Hue C., Vermaak J., Gangnet M. Color-Based Probabilistic Tracking // European Conference on Computer Vision. 2002. P. 661–675.
9. Tuzel O., Porikli F., Meer P. Region covariance: A fast descriptor for detection and classification // European Conference on Computer Vision. 2006. P. 589–600.
10. Porikli F., Tuzel O., Meer P. Covariance tracking using model update based on Lie algebra // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006. Vol. 1. P. 728–735. DOI: 10.1109/CVPR.2006.94
11. Wu Y., Cheng J., Wang J., Lu H. Real-time visual tracking via incremental covariance tensor learning // Proceedings of IEEE International Conference on Computer Vision. 2009. P. 1631–1638. DOI: 10.1109/ICCV.2009.5459369
12. Wu Y., Wang J., Lu H. Robust Bayesian tracking on Riemannian manifolds via fragments-based representation // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2009. P. 765–768. DOI: 10.1109/ICASSP.2009.4959696
13. Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction // Proceedings of International Conference on Pattern Recognition. 2004. Vol. 2. P. 28–31. DOI: 10.1109/ICPR.2004.1333992
14. Viola P., Jones M.J. Robust real-time face detection // International Journal of Computer Vision. 2004. Vol. 57. No. 2. P. 137–154. DOI: 10.1023/B:VISI.0000013087.49260.fb
15. Watada J., MUSAAND Z.B. Tracking human motions for security system // SICE Annual Conference. 2008. P. 3344–3349. DOI: 10.1109/SICE.2008.4655242
16. Алфимцев А.Н. Нечеткое агрегирование мультимодальной информации в интеллектуальном интерфейсе // Программные продукты и системы. 2011. № 3. С. 10.

17. *Bellotto N., Hu H.* Multisensor-based human detection and tracking for mobile service robots // *IEEE Transactions on Systems, Man, and Cybernetics.* 2009. Vol. 39. No. 1. P. 167–181. DOI: 10.1109/TSMCB.2008.2004050

18. *Девятков В.В.* Пространственно-временной анализ ситуаций на множестве движущихся объектов // *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение.* 2011. № 3. С. 88–112.

Девятков Владимир Валентинович — д-р техн. наук, профессор, зав. кафедрой «Информационные системы и телекоммуникации» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Алфимцев Александр Николаевич — канд. техн. наук, доцент кафедры «Информационные системы и телекоммуникации» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Таранян Арам Робертович — аспирант кафедры «Информационные системы и телекоммуникации» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Просьба ссылаться на эту статью следующим образом:

Девятков В.В., Алфимцев А.Н., Таранян А.Р. Селективно-ковариационный метод локализации, классификации и отслеживания людей в видеопотоках от множества видеокамер // *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение.* 2016. № 6. С. 54–70. DOI: 10.18698/0236-3933-2016-6-54-70

**SELECTIVE COVARIANCE-BASED LOCALIZATION,
CLASSIFICATION AND TRACKING IN VIDEO STREAMS
FROM MULTIPLE CAMERAS**

V.V. Devyatkov
A.N. Alfimtsev
A.R. Taranyan

deviatkov@comtv.ru
alfim@bmstu.ru
taranyan.bmstu@gmail.com

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

This paper proposes a novel selective covariance-based method for human localization, classification and tracking in video streams from multiple cameras. Such methods are crucial for security and surveillance systems, smart environments and robots. The method is called selective covariance-based because before classifying the object into this or that class (in this case the classes are the different people being tracked) we use covariance descriptors and sort out (select) definite regions, which are typical for the class of objects we deal with (people). In our case, the region being sorted out is the human head and shoulders. We develop and describe new feature functions for covariance region descriptors and compare the efficiency of their application to that of basic feature functions. Moreover, we propose and evaluate a mask, filtering out the most of the

Keywords

Pattern recognition, computer vision, human tracking, covariance matrix, covariance region descriptor, selective localization

background information from the target area. The use of the proposed feature functions and mask significantly improved the human classification performance (from 75% when using basic feature functions to 94.6% accuracy with the proposed method) while keeping computational complexity moderat

REFERENCES

- [1] Lowe D.G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, vol. 60, no. 2, pp. 91–110. DOI: 10.1023/B:VISI.0000029664.99615.94
- [2] Fazli S., Pour H.M., Bouzari H. Particle filter based object tracking with sift and color feature. *International Conference on Machine Vision*, 2009, pp. 89–93. DOI: 10.1109/ICMV.2009.47
- [3] Belongie S., Malik J., Puzicha J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, vol. 24, no. 4, pp. 509–522. DOI: 10.1109/34.993558
- [4] Ioffe S., Forsyth D.A. Probabilistic methods for finding people. *International Journal of Computer Vision*, 2001, vol. 43, no. 1, pp. 45–68. DOI: 10.1023/A:1011179004708
- [5] Elzein H., Lakshmanan S., Watta P. A motion and shape-based pedestrian detection algorithm. *IEEE Intelligent Vehicles Symposium. Proc.*, 2003, pp. 500–504. DOI: 10.1109/IVS.2003.1212962
- [6] Liu H., Wang L., Sun F. Mean-shift tracking using fuzzy coding histogram. *International Journal of Fuzzy Systems*, 2014, vol. 16, no. 4, pp. 457–467.
- [7] Comaniciu D., Ramesh V., Meer P. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, no. 5, pp. 564–577. DOI: 10.1109/TPAMI.2003.1195991
- [8] Perez P., Hue C., Vermaak J., Gangnet M. Color-based probabilistic tracking. *European Conference on Computer Vision*, 2002, pp. 661–675.
- [9] Tuzel O., Porikli F., Meer P. Region covariance: A fast descriptor for detection and classification. *European Conference on Computer Vision*, 2006, pp. 589–600.
- [10] Porikli F., Tuzel O., Meer P. Covariance tracking using model update based on Lie algebra. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 1, pp. 728–735. DOI: 10.1109/CVPR.2006.94
- [11] Wu Y., Cheng J., Wang J., Lu H. Real-time visual tracking via incremental covariance tensor learning. *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 1631–1638. DOI: 10.1109/ICCV.2009.5459369
- [12] Wu Y., Wang J., Lu H. Robust Bayesian tracking on Riemannian manifolds via fragments-based representation. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 765–768. DOI: 10.1109/ICASSP.2009.4959696
- [13] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction. *Proceedings of International Conference on Pattern Recognition*, 2004, vol. 2, pp. 28–31. DOI: 10.1109/ICPR.2004.1333992

- [14] Viola P., Jones M.J. Robust real-time face detection. *International Journal of Computer Vision*, 2004, vol. 57, no. 2, pp. 137–154. DOI: 10.1023/B:VISI.0000013087.49260.fb
- [15] Watada J., Musaand Z.B. Tracking human motions for security system. *SICE Annual Conference*, 2008, pp. 3344–3349. DOI: 10.1109/SICE.2008.4655242
- [16] Alfimtsev A.N. Fuzzy aggregation of multimodal information in an intelligent interface. *Programmnye produkty i sistemy* [Software & Systems], 2011, no. 3, pp. 10 (in Russ.).
- [17] Bellotto N., Hu H. Multisensor-based human detection and tracking for mobile service robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 2009, vol. 39, no. 1, pp. 167–181. DOI: 10.1109/TSMCB.2008.2004050

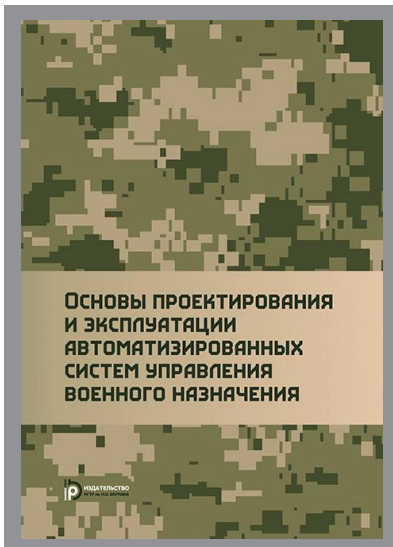
Devyatkov V.V. — Dr. Sci. (Eng.), Professor, Head of Information Systems and Telecommunications Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Alfimtsev A.N. — Cand. Sci. (Eng.), Assoc. Professor of Information Systems and Telecommunications Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Taranyan A.R. — postg-graduate student of Information Systems and Telecommunications Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Devyatkov V.V., Alfimtsev A.N., Taranyan A.R. Selective Covariance-Based Localization, Classification and Tracking in Video Streams from Multiple Cameras. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Priborostr.* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.], 2016, no. 6, pp. 54–70. DOI: 10.18698/0236-3933-2016-6-54-70



В Издательстве МГТУ им. Н.Э. Баумана
вышло в свет учебное пособие автора
В.Л. Лясковского

**«Основы проектирования и эксплуатации
автоматизированных систем управления
военного назначения»**

Рассмотрены вопросы, связанные с общей характеристикой систем обработки информации и управления автоматизированных систем управления военного назначения, приведено подробное описание процесса их проектирования и эксплуатации.

По вопросам приобретения обращайтесь:

105005, Москва, 2-я Бауманская ул., д. 5, стр. 1
+7 (499) 263-60-45
press@bmstu.ru
www.baumanpress.ru