

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.931

ВЫЧИСЛЕНИЕ И АНАЛИЗ ПРИЗНАКОВ ДВИЖУЩИХСЯ ОБЪЕКТОВ ДЛЯ СОПРОВОЖДЕНИЯ НА ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ

канд. техн. наук, доц. Р.П. БОГУШ¹, академик, д-р техн. наук, проф. С.В. АБЛАМЕЙКО^{2,3},
И.Ю. ЗАХАРОВА¹

¹(Полоцкий государственный университет)

²(Белорусский государственный университет)

³(Объединенный институт проблем информатики Национальной академии наук Беларуси)

Рассмотрена задача формирования признаков объектов на видеопоследовательностях. Представлены основные типы движения одиночного объекта и группы. Предложена следующая классификация основных признаков, характеризующих движение объекта на видеопоследовательности: смещение, траектория, скорость и ускорение, время движения. С использованием предложенных обобщений для вычисления признаков динамических объектов описана модификация алгоритма сопровождения множества людей на видеопоследовательностях за счет использования фильтра Калмана для видеонаблюдения вне помещений. Первый этап алгоритма требует обнаружения всех людей во входных кадрах с использованием сверточной нейронной сети YOLOv4. Для сопоставления изображений людей на кадрах выполняется анализ их признаков в пространственной области кадров и во временной области на видеопоследовательности. При этом используются нейросетевые и гистограммные признаки, вычисленные для последнего правильного обнаружения человека в кадре; координаты центра выделенной области человека в кадре; смещение в текущем кадре относительно предыдущего; ширина и высота области на предыдущем кадре; траектория движения; время движения. Представлены результаты экспериментов для видеопоследовательностей, полученных с использованием стационарной и движущейся видеокамеры.

Ключевые слова: видеопоследовательность, анализ объектов, признаки движения, сверточная нейронная сеть.

Введение. Анализ объектов на видеопоследовательностях требуется при решении многих прикладных задач, включая обнаружение людей и их распознавание, сопровождение объектов для установления их перемещения на одной динамической сцене, повторную идентификацию (реидентификацию) людей в мультикамерных системах видеонаблюдения, определение нехарактерного поведения различных классов объектов. Задача анализа изображений объектов на видеопоследовательностях заключается в том, что требуется извлечение различной информации из последовательности входных кадров изображений объекта, однако для обработки он представляется в виде двумерного массива значений уровней яркостей пикселей. Таким образом, сложность заключается в том, что в отличие от статических изображений в данном случае необходимо учитывать наличие движения. Известно, что временные переходы объектов на трехмерной сцене могут приводить к таким преобразованиям на двумерном изображении, которые невозможно описать существующими аналитическими зависимостями [1], в связи с чем одни и те же признаки одного объекта будут характеризовать разные объекты. Поэтому при практической реализации алгоритмов обработки видео на первом шаге выполняется обнаружение объектов и их локализация или же детектирование областей-кандидатов, которые могут быть отнесены к объектам интереса. Следующий этап требует вычисления признаков выделенных фрагментов, на основе которых выполняется анализ и конечная их классификация. Подход с использованием областей-кандидатов позволяет сократить временные затраты за счет детальной классификации только более схожих объектов с заданными и повысить точность правильного обнаружения в целом.

Для анализа выделенных фрагментов кадров, содержащих объекты, на видеопоследовательности необходимо сформировать комплексный дескриптор, состав которого определяется решаемой прикладной задачей. Следует отметить, что в него необходимо включать признаки, позволяющие описать объект на основе данных не только текущего кадра, но и на последовательности предыдущих, которая позволит оценить особенности движения объекта. Для выделения признаков объектов на изображениях существуют различные подходы [2]. Однако, несмотря на имеющиеся результаты в теории распознавания образов и анализе статических изображений, пока не выработано четких определений и терминов для классификации и выбора признаков объектов, учитывающих их движение на видеопоследовательностях, а лишь присутствуют общие подходы для решения поставленной задачи.

Целью статьи является обобщение и классификация признаков движения объектов на видеопоследовательностях и их применение для задачи сопровождения множества людей с использованием сверточных нейронных сетей (СНС).

Постановка задачи. Ключевыми этапами процесса автоматической обработки видеопоследовательностей являются обнаружение объектов, вычисление и анализ их признаков для классификации. Под объектом (Ob) понимаем локальную область, которая отличается от окружающего фона и отображает некоторую часть признаков объекта из реального мира.

Видеопоследовательность (V), которая получена со стационарно установленной видеокамеры, состоит из набора изображений (кадров) одной и той же сцены, полученных с определенным временным интервалом (частотой кадров): $V = \{F_k\}$, k – номер изображения в последовательности.

По критерию движения объект ($Ob_q^{F_k}$), который присутствует на кадрах, может быть стационарным или движущимся.

Стационарный объект описывается множеством признаков ($Ft_{Ob_q^S}$) и не изменяющимися в течение интервала времени (t) координатами (x_{Ob_q}, y_{Ob_q}):

$$Ob_q^S = \{Ft_{Ob_q^S}, x_{Ob_q}, y_{Ob_q}, Ns_{Ob_q^{F_k}}\},$$

где $(Ft_{Ob_q}, x_{Ob_q}, y_{Ob_q}) = const \forall F_k, k \in t$;

$Ns_{Ob_q^{F_k}}$ – множество шумовых воздействий на объект.

Для движущегося объекта на видеопоследовательности наблюдается изменение одного или нескольких параметров: координат ($x_{Ob_q^{F_k}}, y_{Ob_q^{F_k}}$), размеров ($SZ_{Ob_q^{F_k}}$), формы ($FR_{Ob_q^{F_k}}$) на интервале времени (t).

Трансформация его формы и (или) размеров изменяет его признаки на кадрах ($ft_{Ob_q^{F_k}}$). Соответственно, движущийся объект описывается как

$$Ob_q^D = \{ft_{Ob_q^{F_k}}, x_{Ob_q^{F_k}}, y_{Ob_q^{F_k}}, Ns_{Ob_q^{F_k}}\},$$

где $Ft_{Ob_q^D}$ – множество признаков динамического объекта, $Ft_{Ob_q^D} \supseteq ft_{Ob_q^{F_k}}, \forall k \in t$. Как правило, для движущегося объекта характерно изменение его признаков на кадрах, однако $ft_{Ob_q^{F_k}} \cap ft_{Ob_q^{F_{k+1}}}$.

На изображении обнаруженный объект выделяется, как правило, прямоугольной рамкой и для дальнейших итераций обработки используется анализ данной области кадра.

Анализ объектов предполагает преобразование выделенного фрагмента изображения из его исходного представления в конечное, в качестве которого обычно используется вектор признаков, характеризующий объект в целом:

$$Ob = \{ob_{ij}\} \rightarrow P_{Ob} = \{p_l\}, l = 1, \dots, L,$$

где P_{Ob} – множество признаков (p) изображения объекта.

Результат анализа фрагмента изображения в кадре на основе выделенных признаков должен содержать информацию о наличии или отсутствии объекта, о месте его положения в кадре и о классе объекта, т.е. объект в дальнейшем должен быть отнесен к определенному классу (распознан). В отличие от изображений, для видеопоследовательности требуется выполнять анализ не только в пространственной области, но и во временной. Для описания объектов необходимо применять признаки, которые инвариантны относительно яркостных и геометрических преобразований с учетом их временных изменений на видеопоследовательности: сдвиг, поворот и изменение масштаба. Для общего случая задача определения наиболее эффективного набора признаков движущихся объектов теоретически и технически не решена.

Типы движения объектов. Движение одиночного объекта, которое наблюдается на последовательности изображений, обычно представляет собой один из четырех основных элементов или их комбинацию [3]: смещение в сторону (рисунок 1, а) приводит к изменению его координат, смещение по оптической оси видеокамеры (рисунок 1, б) приводит к увеличению размеров либо к их уменьшению, вращение на постоянном расстоянии от видеокамеры вокруг оптической оси (рисунок 1, в) приводит к повороту объекта

на изображении, вращение вокруг собственной оси перпендикулярно оптической оси на постоянном расстоянии от видеокамеры (рисунок 1, *з*) приводит к визуальному изменению внешних признаков объекта, координаты и размер не изменяются.

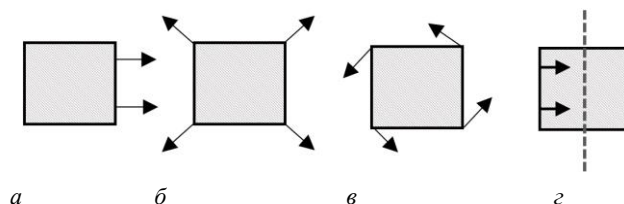


Рисунок 1. – Основные типы движения объектов на видеопоследовательностях

За время наблюдения объект может изменять свой класс по критерию движения. Он переходит из класса «стационарный» в класс «движущийся» на k -кадре, если выполняется хотя бы одно из условий:

$$\left(x_{Ob_q}^{F_{k-1}}, y_{Ob_q}^{F_{k-1}}\right) \neq \left(x_{Ob_q}^{F_k}, y_{Ob_q}^{F_k}\right), \left(FR_{Ob_q}^{F_{k-1}}\right) \neq \left(FR_{Ob_q}^{F_k}\right), \left(SZ_{Ob_q}^{F_{k-1}}\right) \neq \left(SZ_{Ob_q}^{F_k}\right).$$

Переход движущегося объекта в класс «стационарный» происходит после прекращения его движения, соответственно, координаты, форма и размеры объекта на соседних кадрах должны быть идентичны. За время наблюдения возможны неоднократные изменения классов для одного и того же объекта по критерию движения, тогда за этот период движение объекта будет иметь дискретный характер.

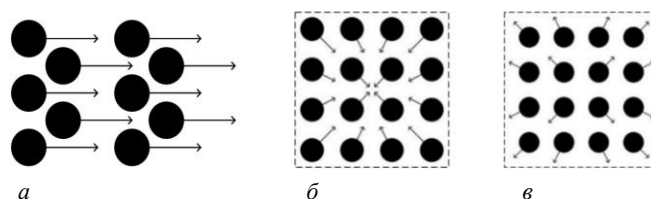
На каждом кадре последовательности может присутствовать множество движущихся объектов, поэтому при видеомониторинге часто возникает задача обнаружения и отслеживания поведения не единичного, а группы объектов (множества или толпы людей, клеточной популяции, движения автомобилей и т.п.).

При исследовании движения группы объектов на видео можно выделить три типа движения [4]: направленное движение объектов, агрегация объектов или их групп (движение к общему центру), рассеивание объектов (движение от центра).

Направленное движение определяется тогда, когда несколько объектов движутся в одном и том же направлении (рисунок 2, *а*). Основными признаками такого движения являются: одновременное перемещение нескольких объектов из одной области изображения в другую, скорость движения объектов, превышающая скорость движения объектов фона, совпадающие направления движения объектов.

Под агрегацией понимают движение объектов по направлению к общему центру. Движение объектов может быть симметричным или могут преобладать два направления (рисунок 2, *б*). К признакам агрегации относятся: перемещение объектов в одну область изображения, большая скорость движения объектов по сравнению со скоростью их хаотического движения, возможность выделения двух и более преобладающих направлений движения.

При рассеивании движение объектов осуществляется от общей точки на изображении (рисунок 2, *в*). К признакам рассеивания относятся: перемещение объектов от их общего центра на другие участки изображения, скорость движения объектов выше скорости хаотического движения, возможность выделения двух и более преобладающих направлений движения.



а – направленное движение; *б* – агрегация; *в* – рассеивание

Рисунок 2. – Основные типы движения группы объектов

Анализ рассмотренных типов движения показывает, что для эффективного анализа движущихся объектов на видеопоследовательностях следует использовать различные наборы признаков, включая пространственные характеристики объекта в кадре, и признаки, описывающие особенности движения объектов.

Классификация признаков движущихся объектов на видео. Можно выделить следующие классификации признаков и соответствующие им классы:

- по типу объекта (области) на видео: пиксели, блоки пикселей, фрагменты объекта, объект, группа объектов;
- по модели представления: в спектральной и временной области;

- по типу движения: для одиночного объекта и для группы объектов;
- по «уровню» признака: первичные и вторичные.

К признакам, характеризующим движение объекта, на видеопоследовательности целесообразно отнести:

- 1) смещение;
- 2) траектория движения;
- 3) скорость и ускорение;
- 4) время движения.

Можно выделить следующие параметры признака смещения объекта: длина вектора смещения объекта от кадра к кадру, направление вектора смещения объекта, плотность векторов смещения, однородность векторов смещения, карты движения, гистограмма оптического потока.

Первичными признаками смещения являются длина вектора и его направление, которые могут применяться для оценки этих величин представленных уровней декомпозиции объекта. На их основе вычисляются вторичные признаки: плотность и однородность векторов, карты движения, гистограмма оптического потока. Максимально плотным считается оптический поток, если он найден для всех точек объекта интереса. Однородность векторов максимальна, если все они по модулю и направлению одинаковы.

Гистограммы оптического потока (HOF) [5] позволяют извлечь информацию об относительном движении небольших фрагментов соседних кадров и оценить их связность в движении. Вычисление HOF основано на разбиении векторного поля на равные прямоугольные части (рисунок 3, а) и нахождении гистограмм количества векторов и их суммарной длины по диапазону направлений. Гистограммы состоят из столбцов, каждый из которых отражает количество или суммарную длину векторов, попавших в заданный диапазон значений направлений, и позволяют классифицировать движение объектов определенных областей. Визуально такие гистограммы отображают в виде векторов, исходящих из одной точки и имеющих направления, равные среднему значению диапазона направлений столбца, и длины, пропорциональные столбцам (рисунок 3, б). Итоговая гистограмма формируется на основе объединения гистограмм, рассчитанных для каждой части (рисунок 3, в).

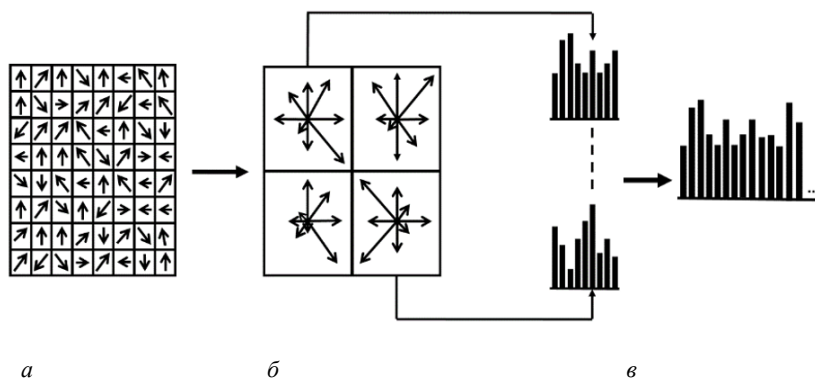


Рисунок 3. – Принцип получения гистограммы оптического потока

Интегральный оптический поток применяется для анализа движения групп объектов, например, людей [6]. Для каждого пиксела или блока изображения он формируется как результат интегрирования значений оптического потока по заданному количеству кадров видеопоследовательности:

$$VOF_t^N(p) = \sum_{i=0}^{N-1} OF_{t+i}(p_{t+i}),$$

- где OF_t – классический оптический поток для видеопоследовательности V ;
- VOF_t^N – интегральный оптический поток фрагмента видеопоследовательности для N кадров;
- p_{t+i} – положение пиксела $I_t(p)$ на $t+1$ кадре видеопоследовательности.

Для определения типа движения совокупности объектов в работе [6] построены четыре вида карт движения, использующие анализ интегрального оптического потока: карта сходимости пикселов, определяющая для каждого узла число пикселов, перемещающихся по направлению к нему; карта расходимости пикселов, определяющая число пикселов, движущихся по направлению от данного узла; результирующая карта сходимости пикселов, отображающая результирующее движение всех пикселов по направлению к данному узлу; результирующая карта расходимости пикселов, отображающая результирующее движение всех пикселов по направлению от данного узла.

Рассмотренные вторичные признаки позволяют выполнять локализацию объектов и оценку их поведения на динамической сцене на основе анализа видеопоследовательности.

Вторым признаком, который важен при анализе движения, является *траектория*, отображающая движение объекта на кадрах видеопоследовательности в виде тонкой линии или линии единичной ширины поперечного сечения на кадре. Данная линия может быть построена на основе координат объекта на предыдущих кадрах видеопоследовательности. Могут быть разные варианты их определения, но наиболее распространенный подход предполагает нахождение центра объекта (один пиксел на кадре) с координатами $(x_{Ob_q}^{F_i}, y_{Ob_q}^{F_i})$. Тогда траектория движения объекта на видео описывается последовательностью в виде набора координат центра объекта на каждом кадре:

$$Tr(Ob^D) = (Ob_{F_i}^D) = \{(x_{Ob_q}^{F_1}, y_{Ob_q}^{F_1}), (x_{Ob_q}^{F_2}, y_{Ob_q}^{F_2}), (x_{Ob_q}^{F_3}, y_{Ob_q}^{F_3}), \dots, (x_{Ob_q}^{F_n}, y_{Ob_q}^{F_n})\}.$$

Следует учитывать, что объекты могут появляться на видеопоследовательности не с первого кадра и выходить за пределы сцены в некоторый момент времени на n -м кадре, а не на последнем кадре. Тогда траектория движения может быть построена с кадра m , на котором обнаружен объект и до n -го кадра. Таким образом, траектория создается при появлении нового объекта в кадре и, как правило, удаляется при выходе его за пределы кадра. На время кратковременного скрывания движущегося объекта за другими или при невозможности его обнаружения по другим причинам, траектория прерывается, но восстанавливается через несколько кадров, т.е. наблюдается фрагментация траектории движения. Поэтому для оценки состояния необходим учет особенностей движения объекта, что может быть обеспечено использованием некоторых априорных сведений о перемещении объекта на предыдущих кадрах и его динамических возможностях, на основе которых может быть предварительно построена модель траектории, которая может быть линейной или нелинейной, периодической или непериодической. Исходя из траектории, может быть вычислено перемещение объекта на видео для заданного интервала времени, граничные значения которого определяются номерами кадров, на которых необходимо определить координаты положения объекта и найти между ними кратчайшее расстояние в координатах кадра.

Скорость и ускорение движения объекта в координатах кадра являются следующей группой важных признаков объекта, которые можно рассчитать путем вычисления первой и второй производных по одному из 4 или 8 направлений в зависимости от используемого типа связности пикселов.

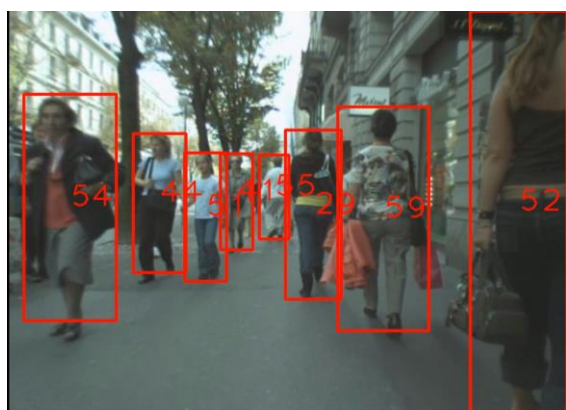
Четвертым признаком, который необходим при анализе движения, является *время движения объекта* (временной признак). Основными параметрами этого признака являются время начала и окончания движения объекта, вторичными – интервалы времени, характеризующие непрерывное движение объекта и его стационарность за весь период наблюдения.

Сопровождение множества людей на видеопоследовательностях на основе анализа комплекса признаков. Для сопровождения множества людей эффективным является принцип сопровождения через обнаружение. В работе [7] предложен алгоритм, который использовал СНС YOLOv3 для обнаружения людей. Этап детектирования является одним из ключевых в алгоритмах трекинга, построенных на основе сопровождения с использованием обнаружения, так как точность его работы в значительной мере определяет и качество сопровождения. Поэтому для дальнейшего развития алгоритма будет использоваться более точная СНС YOLOv4, преимущества которой указаны в [8]. После обнаружения фигуры человека на кадре вычисляются и анализируются следующие признаки выделенного фрагмента в пространственной области кадров и во временной области на видеопоследовательности: СНС и гистограммные признаки канала цветного тона пространства HSV человека при последнем правильном его обнаружении на кадре, координаты центра выделенной области человека в кадре, смещение в текущем кадре относительно предыдущего, ширина и высота области на предыдущем кадре, траектория движения, время движения. Для всех сопровождаемых и обнаруженных на текущем кадре объектов вычисляются значения схожести, на основе которых устанавливается соответствие между обнаруженными и сопровождаемыми объектами. Установление соответствия для детектированных людей и их траекторий выполняется на основе венгерского алгоритма. Траектория создается при первом обнаружении человека, а удаляется если данный человек на протяжении определенного числа последовательных кадров не детектируется и для него отсутствует сопоставление с предыдущими кадрами, т.е. считается, что он вышел из сцены, которую снимает видеочамера.

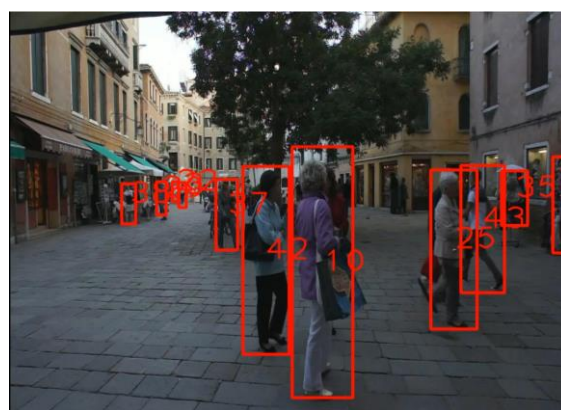
При сопровождении людей в помещениях траектория их движения имеет нелинейный характер и является труднопредсказуемой, поэтому построение и использование модели траектории для прогнозирования нового положения объекта не применяется в алгоритме [7]. При наблюдении вне помещений человек, как правило, перемещается в одном направлении на значительно большем количестве кадров, чем в помещении. Значит, траектория движения человека при внешнем видеонаблюдении ближе к линейной на более значительных временных интервалах. Следовательно, для данной задачи предлагается модификация алгоритма для предсказания положения человека на текущем кадре относительно предыдущих

с использованием линейного фильтра Калмана [9]. Первый шаг алгоритма Калмана выполняет прогноз значения возможного положения человека на текущем кадре на основе его траектории движения, а на втором шаге выполняется корректировка с учетом полученного прогнозируемого положения и сопоставления с ранее найденными изображениями людей на последовательности кадров. Следует отметить, что линейный фильтр Калмана чувствителен к единичным выбросам и нелинейности траектории движения.

Для оценки возможности практического использования алгоритма важной задачей является определение эффективности сопровождения с учетом совместной работы этапов обнаружения и трекинга, поэтому тестирование алгоритма сопровождения для наблюдения в помещениях и модифицированного алгоритма для применения вне помещений, который использует фильтр Калмана, выполнено с учетом результатов обнаружения людей СНС YOLOv4. В этом случае ошибки работы данной СНС приводят к ухудшению критериев оценки сопровождения алгоритмов, но эксперименты отражают реальную результативность алгоритмов сопровождения, что очень важно для принятия решения об их применении на практике. Тестирование проведено по методике MOT16 из [10] и определены критерии оценки сопровождения алгоритма. Основные из этих критериев представлены в таблице 1: обобщающий точность и полноту критерий путем вычисления их гармонического среднего за период (IDF1), точность локализации сопровождаемых людей (MOTP), количество отслеживаемых траекторий (MT), количество потерянных траекторий (ML), общее количество срывов за счет перезахвата ложного объекта при сопровождении (IDS), точность сопровождения множества объектов (MOTA). Для оценки эффективности сопровождения людей вне помещений использованы видеопоследовательности из базы данных MOT16, которые получены с использованием нестационарной видеокамеры при естественном и искусственном уличном освещении в ночное время (примеры показаны на рисунке 4).



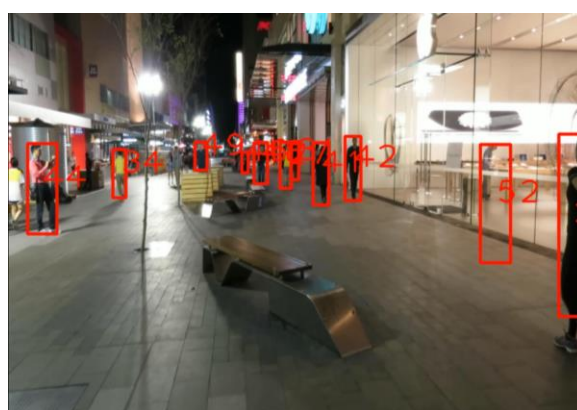
а



б



в



г

Рисунок 4. – Примеры сопровождения множества людей вне помещения с использованием движущейся видеокамеры

Эксперименты для видеопоследовательностей, полученных внутри помещений, проведены на шести видео со стационарной камеры наблюдения (рисунок 5). Общее количество кадров, для которых характерно изменение освещения, нелинейная траектория движения людей, их перекрытие объектами фона или пересечение их траекторий, схожие характеристики и др., составляет 11890.



Рисунок 5. – Примеры сопровождения множества людей с нелинейной траекторией движения в помещении с использованием стационарной видеокамеры

Таблица 1. – Результаты экспериментов по оценке результативности сопровождения множества людей

Видеоряд	IDF1	МОТР	MT	ML	IDS	МОТА
MOT16-02 [10]	24,4223	75,6913	7	30	184	21,4266
MOT16-04 [10]	29,3523	77,4336	7	44	253	27,1758
MOT16-05 [10]	34,0707	73,0136	27	27	179	44,7199
MOT16-09 [10]	37,9400	78,9493	9	2	100	48,2024
MOT16-10 [10]	30,1604	70,9951	10	23	239	30,2565
MOT16-11 [10]	38,8816	79,7906	17	29	139	49,2697
MOT16-13 [10]	28,3641	69,5090	16	55	235	23,0742
Рисунок 5, а	66,6366	81,1972	2	0	8	83,5165
Рисунок 5, б	74,1494	80,9533	2	0	1	90,2609
Рисунок 5, в	75,3374	83,1033	2	0	13	87,1513
Рисунок 5, г	85,2735	85,9602	3	0	11	98,6632
Рисунок 5, д	68,8444	79,1090	3	0	3	92,3880
Рисунок 5, е	95,1529	80,9356	3	0	3	90,7566

Анализ таблицы 1 и рисунков 4, 5 свидетельствует, что на точность работы алгоритмов значительно влияет результативность обнаружения людей. На видеопоследовательностях, полученных с помощью движущейся камеры из-за значительных шумовых воздействий, работа детектора достаточно долгое время характеризуется пропуском многих людей, что приводит к потере текущего индекса и присвоению нового при возобновлении обнаружения. Также из-за шумовых факторов во многих случаях наблюдается неточное выделение области человека в кадрах. Например, область человека с индексом 38 в кадре на рисунке 4, в обнаружена с высокой степенью точности, а для того же человека на рисунке 4, г детектор включил в область значительно больший фрагмент кадра из-за тени. Это приводит в дальнейшем к значительному отличию пространственных признаков фрагментов и переназначению индекса для сопровождаемого человека (в данном случае – индекс 52 на рисунке 4, з). Кроме этого, в некоторых видео из [10], получаемых переносимой камерой в потоке идущих людей, достаточно продолжительное время присутствуют движущиеся люди, снимаемые на близком расстоянии, и лишь часть их фигур отображается в кадрах. Применяемая СНС для выделения признаков людей обучена с использованием баз данных PRID [11] и iLIDS [12], которые не содержат полные изображения фигур людей, поэтому для повышения точности сопровождения в таких условиях требуется доработка базы данных для обучения СНС, вычисляющей признаки обнаруженных людей на видеопоследовательности. Однако предложенный алгоритм улучшает точность сопровождения на тестовых видеопоследовательностях, полученных вне помещения, по сравнению с алгоритмом из [13], так как для его реализации из [14] обеспечивается точность сопровождения для всех видео из [10] МОТА = 0,288306, а для предложенного алгоритма МОТА = 0,300860. Для видеопоследовательностей со стационарной камеры

наблюдения в помещении также обеспечивается улучшение точности: для алгоритма из [13] MOTA = 0,8793, для предложенного алгоритма с использованием СНС YOLOv4 MOTA = 0,9266.

Пример применения алгоритма для видеонаблюдения в помещении с изменяющимся уровнем освещения и построением траекторий движения людей показан на рисунке 6.

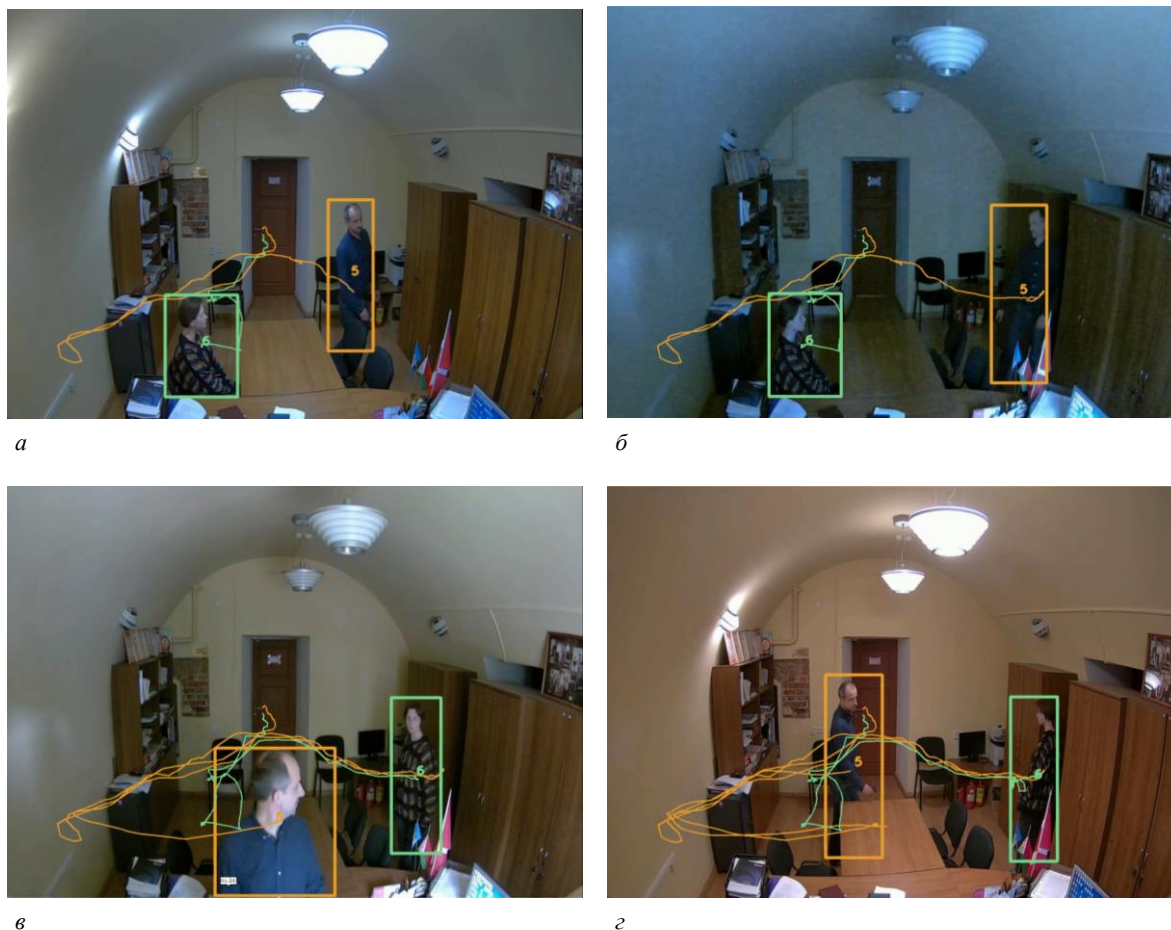


Рисунок 6. – Пример сопровождения людей внутри помещений с использованием стационарной видекамеры с отображением их траекторий движения на кадрах

Анализ траекторий на рисунке 6, *а* показывает, что в помещение вошли два человека, один из них, с индексом 6, подошел к столу и сел на стул, а человек с индексом 5 шел с левой стороны стола к видеокамере, затем развернулся и, обойдя стол, пошел в правую сторону по направлению к шкафам. Рисунок 6, *б* свидетельствует о том, что яркость изображения кадра значительно уменьшилась, т.к. было выключено освещение человеком с индексом 5, что подтверждается сравнением яркости светильников на рисунках 6, *а* и 6, *б*. После этого возникло большое количество шумов, однако сопровождение выполняется корректно (см. рисунок 6, *б*), что подтверждает эффективность разработанного алгоритма. Через незначительное время средствами аппаратной автоподстройки камеры яркость изображения кадра увеличилась, количество шумов стало меньше, обнаружение и сопровождение людей выполнялось корректно (см. рисунок 6, *в*). Траектории движения на рисунке 6, *г* свидетельствуют о том, что освещение в помещении включил человек с индексом 6. Таким образом, анализ траекторий на выбранном кадре позволяет определить, как передвигался каждый человек на динамической сцене, что очень важно для ряда прикладных задач.

Заключение. Анализ объектов на видеопоследовательности предполагает преобразование выделенного фрагмента изображения из его исходного представления в вектор признаков, который в дальнейшем используется для классификации и распознавания. Для последовательностей изображений, в отличие от изображений статических, требуется не только выполнять анализ объекта в пространственной области кадра, но и учитывать признаки движения, поэтому в статье рассмотрены основные типы и особенности движения одного объекта и группы объектов. Предложена следующая классификация основных признаков, характеризующих движение объекта на видеопоследовательности: смещение, траектория, скорость и ускорение, время движения. Приведены описания алгоритмов сопровождения людей на видеопоследовательностях, полученных в помещении и при наружном наблюдении, и выполнено

их тестирование по методике MOT16 по оценке критериев качества работы. Представлены примеры результатов обнаружения людей на видеопоследовательностях, а также кадры с визуализацией траекторий движения людей в условиях значительного изменения освещенности сцены.

ЛИТЕРАТУРА

1. Клетте, Р. Компьютерное зрение. Теория и алгоритмы / Р. Клетте ; пер. с англ. А.А. Слинкин. – М. : ДМК Пресс, 2019. – 506 с.
2. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. –3-е изд. – М. : Техносфера, 2012. – 1104 с.
3. Sonka, M. Image Processing, Analysis, and Machine Vision / M. Sonka, V. Hlavac, R Boyle. – 4th ed. – Cengage Learning, 2015. – 930 p.
4. Application of Integral Optical Flow for Determining Crowd Movement from Video Images Obtained Using Video Surveillance Systems / H. Chen [et al.] // J. of Appl. Spectrosc. – 2018. – Vol. 85, iss. 1. – P. 126–133.
5. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions / R. Chaudhry [et al.] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2009. – 1932–1939.
6. Motion Maps and Their Applications for Dynamic Object Monitoring / H. Chen [et al.] // Pattern Recogn. Image Anal. – 2019. – 29. – P. 131–143.
7. Богуш, Р.П. Алгоритм сопровождения людей на видеопоследовательностях с использованием сверточных нейронных сетей для видеонаблюдения внутри помещений / Р.П. Богуш, И.Ю. Захарова // Компьютерная оптика. – 2020. – Т. 44, № 1. – С.109–116.
8. Bochkovskiy, A. YOLOv4: Optimal Speed and Accuracy of Object Detection [Electronic resource] / A. Bochkovskiy, Ch.-Y. Wang, H.-Y. M. Liao. – Mode of access: <https://arxiv.org/abs/2004.10934>. – Date of access: 12.08.2020.
9. Simon, D. Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches / D. Simon. – New Jersey : John Wiley & Sons, 2006.
10. MOTChallenge: The Multiple Object Tracking Benchmark [Electronic resource]. – Mode of access: <https://motchallenge.net>. – Date of access: 16.06.2020.
11. Person Re-ID (PRID) Dataset [Electronic resource]. – Mode of access: <https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/prid11/>. – Date of access: 12.04.2019.
12. iLIDS Video re-Identification (iLIDS-VID) Dataset [Electronic resource]. – Mode of access: http://www.eecs.qmul.ac.uk/~xiatian/downloads_qmul_iLIDS-VID_ReID_dataset.html. – Date of access: 12.04.2019.
13. Wojke, N. Simple online and real time tracking with a deep association metric / N. Wojke, A. Bewley, D. Paulus // IEEE International Conference on Image Processing 2017: 3645-3649. – DOI: 10.1109/ICIP.2017.8296962.
14. Real-time Multi-person tracker using YOLOv3 and deep_sort with tensorflow [Electronic resource]. – Mode of access: https://github.com/Qidian213/deep_sort_yolov3. – Date of access: 16.06.2019.

Поступила 05.03.2021

CALCULATION AND ANALYSIS OF MOVING OBJECT FEATURES FOR TRACKING IN VIDEO SEQUENCES

R. BOHUSH, S. ABLAMEYKO, I. ZAHKARAVA

This paper discusses the formation of object features considering the peculiarities of their presentation in video sequences. The main types of movement of a single object and object group are presented. We propose a classification of features that characterize the movement of objects in a video sequence. A modification of the algorithm for tracking multiple people on video sequences using the Kalman filter for outdoor video surveillance is described. The first stage requires detecting person in the input frames by YOLOv4 convolutional neural network. For assignment problem solving of person we store information about individual object in spatial domain of frames and in the time domain on a video sequence. For person description feature set is used: neural network and histogram features, center coordinates of a person in the frame, offset in the current frame relative to the previous one, person width and height in the previous frame, trajectory and time of movement. The results of experiments for video sequences obtained using a stationary and moving video camera are presented.

Keywords: video sequence, object analysis, motion features, convolutional neural network.