



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G06T 7/20 (2024.01); G06V 20/58 (2024.01)

(21)(22) Заявка: 2023113896, 29.05.2023

(24) Дата начала отсчета срока действия патента:
29.05.2023Дата регистрации:
21.06.2024

Приоритет(ы):

(22) Дата подачи заявки: 29.05.2023

(45) Опубликовано: 21.06.2024 Бюл. № 18

Адрес для переписки:

620002, г. Екатеринбург, ул. Мира, 19, УрФУ,
Центр интеллектуальной собственности,
Маркс Т.В.

(72) Автор(ы):

Кашин Илья Владимирович (RU),
Галина Екатерина Маратовна (RU)

(73) Патентообладатель(и):

Федеральное государственное автономное
образовательное учреждение высшего
образования "Уральский федеральный
университет имени первого Президента
России Б.Н. Ельцина" (RU)

(56) Список документов, цитированных в отчете

о поиске: RU 2008114730 А, 20.10.2009. RU
2628172 С1, 15.08.2017. RU 2634225 С1,
24.10.2017. RU 2395787 С2, 27.07.2010. RU
2656711 С2, 06.06.2018. WO 2011054971 А2,
12.05.2011. WO 2012145819 А1, 01.11.2012. US
20180061063 А1, 01.03.2018.

(54) Способ отслеживания объектов на видео

(57) Реферат:

Изобретение относится к способу отслеживания объектов на видео. Техническим результатом является повышение точности отслеживания объектов на видео. В способе выполняют регистрацию последовательных цифровых изображений и дальнейшую их обработку для выделения движущихся объектов, при этом для обработки одного кадра, зарегистрированного в текущий момент времени стационарной камерой, представленного в виде изображения в градациях серого, загруженного в оперативную память компьютера как двумерный числовой массив \tilde{n} размерами $W \times H$, осуществляют итерационный процесс, в ходе которого производят преобразования над массивом n^i , содержащим элементы целочисленного типа, и размеры которого совпадают с размерами изображения, а также массивом D размерами $w \times h$, элементы которого

равны нулю и $w \leq W, h \leq H$, а именно: для каждого произвольно выбранного на начальном этапе стартового элемента n_{xy}^i рассматривают

четыре смежных элемента - пробных ячеек $n_{x_p y_p}^i$,

$$x_p = x + \Delta x, \quad y_p = y + \Delta y, \quad \text{где}$$

$\Delta y, \Delta x = -1, 0, 1$, не учитывая случаи

$|\Delta y| = |\Delta x|$, и вычисляют величину локального отличия в соответствующей ячейке

$$\alpha_i^p = \tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} n_{x_p y_p}^i, \text{ где } i - \text{ номер итерации, } p$$

- номер пробной ячейки, \tilde{N} - сумма всех элементов

массива \tilde{n} , N_i - сумма всех элементов массива n^i ;

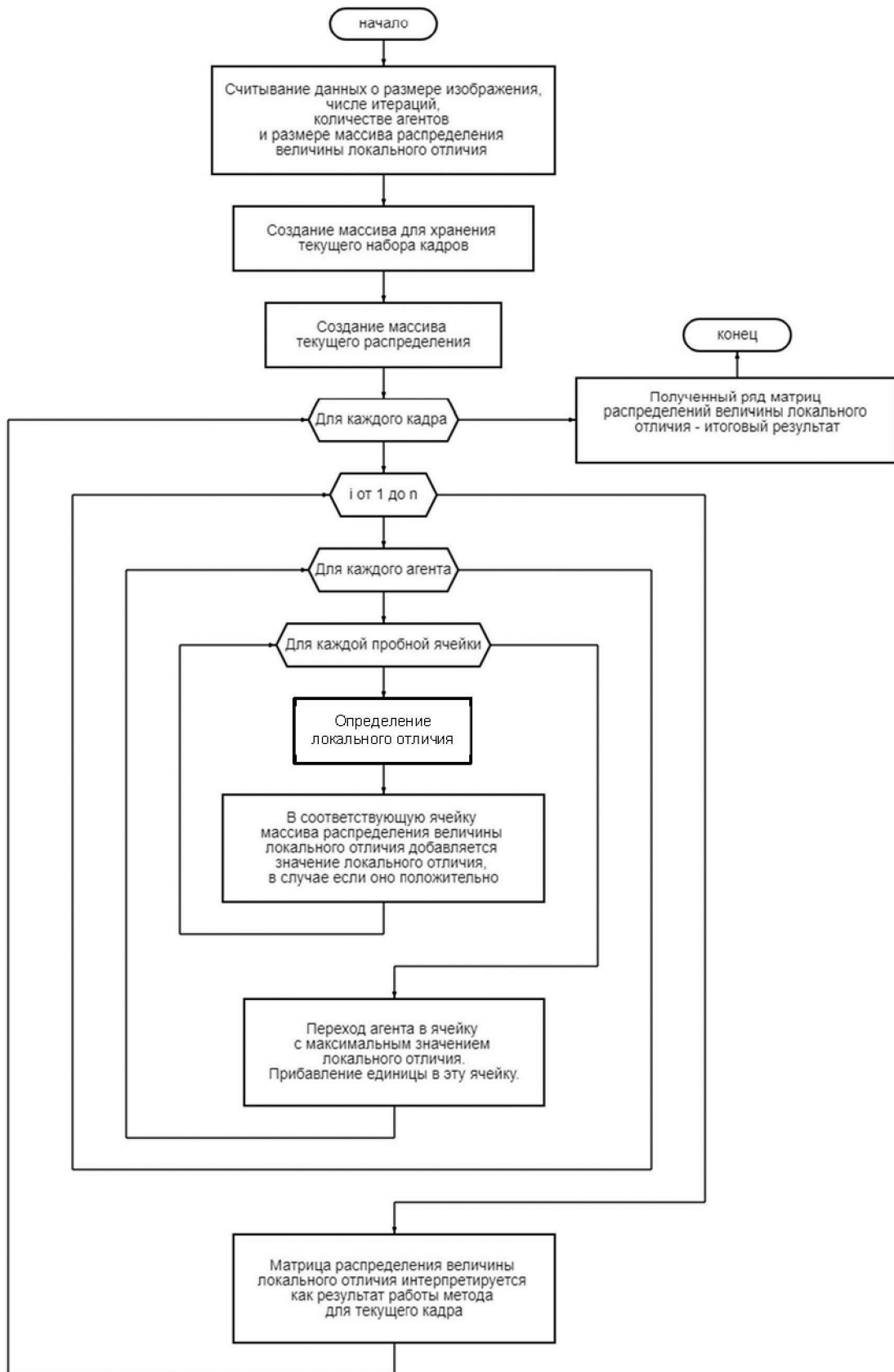
к элементу массива D , соответствующему пробной ячейке, добавляют значение локального

отличия, в случае если оно положительно $D_{x_p/\Delta W, y_p/\Delta H}^{new} = D_{x_p/\Delta W, y_p/\Delta H}^{old} + \max(a_i^p, 0)$ где $\Delta W = W/w, \Delta H = H/h$; в качестве направления следующего шага выбирают пробную ячейку с наибольшим значением a_i^p , если таких ячеек несколько, то среди них выбирают случайную; к выбранному элементу массива n^i добавляют единицу и значение N_i так же увеличивают на единицу; после завершения итерационного

процесса все элементы массива n^i умножают на отношение $\frac{N}{N_i}$; полученную матрицу D интерпретируют как итоговый результат обработки текущего кадра посредством масштабирования и наложения матрицы D на само изображение, при этом определяют максимумы сформированного распределения матрицы D , указывающие на движущиеся объекты, изображенные на кадре; массив n^i используют для обработки следующего кадра. 1 з.п. ф-лы, 2 ил.

RU 2 8 2 1 3 7 7 C 1

RU 2 8 2 1 3 7 7 C 1



Фиг. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC

G06T 7/20 (2024.01); G06V 20/58 (2024.01)(21)(22) Application: **2023113896, 29.05.2023**(24) Effective date for property rights:
29.05.2023Registration date:
21.06.2024

Priority:

(22) Date of filing: **29.05.2023**(45) Date of publication: **21.06.2024** Bull. № 18

Mail address:

**620002, g. Ekaterinburg, ul. Mira, 19, UrFU, Tsentr
intellektualnoj sobstvennosti, Marks T.V.**

(72) Inventor(s):

**Kashin Ilia Vladimirovich (RU),
Galina Ekaterina Maratovna (RU)**

(73) Proprietor(s):

**Federal State Autonomous Educational
Institution of Higher Education Ural Federal
University named after the first President of
Russia B.N.Yeltsin (RU)**(54) **METHOD OF TRACKING OBJECTS ON VIDEO**

(57) Abstract:

FIELD: physics.

SUBSTANCE: invention relates to a method of tracking objects on video. Method comprises registration of successive digital images and their further processing to select moving objects, at that, to process one frame, recorded at the current moment of time by a stationary camera, presented in the form of an image in grayscale, loaded into the computer random access memory as a two-dimensional numerical array \tilde{n} with dimensions $W \times H$, performing an iterative process, during which transformations are performed on array n^i , containing elements of integer type, and dimensions of which coincide with dimensions of image, as well as array D with dimensions $w \times h$, which elements are equal to zero and $w \leq W, h \leq H$, namely: for each randomly selected starting element n_{xy}^i , at the initial stage, four adjacent elements are considered - trial cells $n_{x_p y_p}^i$, $x_p = x + \Delta x$, $y_p = y + \Delta y$, where $\Delta y, \Delta x = -1, 0, 1$, not taking into account the cases

$|\Delta y| = |\Delta x|$, and calculating the value of local difference in the corresponding cell

$\alpha_i^p = \tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} n_{x_p y_p}^i$, where i is the iteration

number, p is the number of the test cell, \tilde{N} is the sum of all elements of the array \tilde{n} , N_i is the sum of all elements of the array n^i ; to the array element D corresponding to the test cell, the local difference value is added, if it is positive

$D_{x_p/\Delta W, y_p/\Delta H}^{new} = D_{x_p/\Delta W, y_p/\Delta H}^{old} + \max(\alpha_i^p, 0)$,

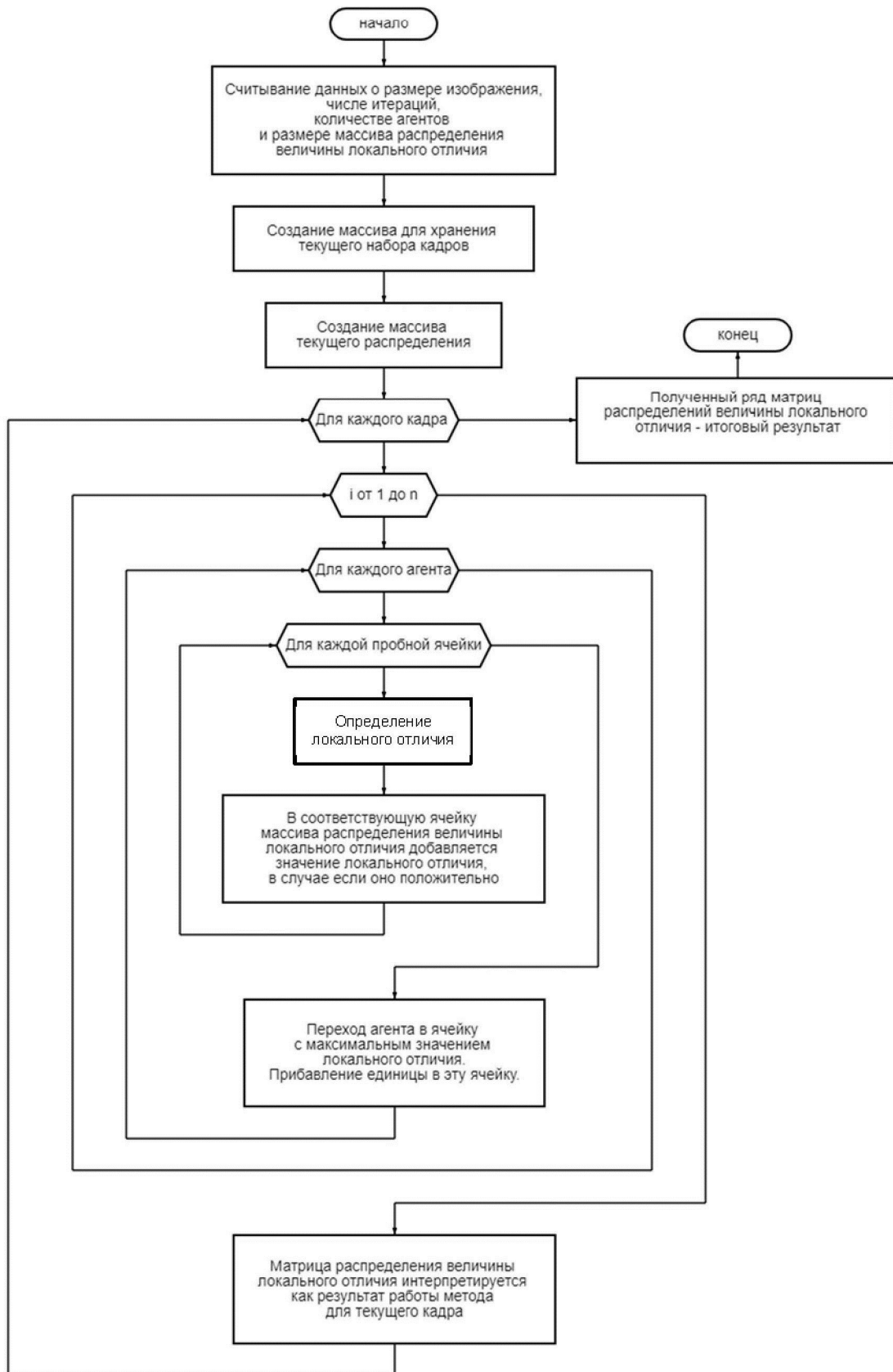
where $\Delta W = W/w$, $\Delta H = H/h$; test cell with maximum value α_i^p is selected as the direction of the next step, if there are several such cells, then random is selected among them; one is added to the selected element of the array n^i and the value N_i is also increased by one; after completion of iterative process all elements of array n^i are multiplied by ratio $\frac{\tilde{N}}{N_i}$;

obtained matrix D is interpreted as the final result of processing the current frame by scaling and superimposing the matrix D on the image itself, at that, determining the maxima of the formed distribution of the matrix D, indicating the moving objects displayed

on the frame; array n^i is used to process the next frame.

EFFECT: high accuracy of tracking objects on video.

1 cl, 2 dwg



Фиг. 1

Изобретение относится к области компьютерного зрения и может быть использовано для обработки кадров видео и применено в тех сферах, где требуется обнаружение и отслеживание каких-либо объектов. Например, задача отслеживания напрямую связана с определением подозрительного поведения человека, находящегося в магазине, парке, и в других общественных местах. Также актуальной задачей является определение скорости перемещения автомобилей по автомагистралям для контроля безопасности в области дорожного движения.

Известен широко используемый способ обнаружения движущихся объектов - алгоритм Лукаса-Канаде, заключающийся в решении уравнения оптического потока методом наименьших квадратов [1]. Метод достаточно точен, хорошо определяет маленькие сдвиги и менее чувствителен к шуму, чем многие другие подходы, однако из-за ряда допущений (смещение в окрестности произвольной точки считается постоянным и цвет пикселей объекта не меняется), сделанных при составлении и решении уравнения оптического потока, его рабочая область оказывается значительно ограниченной. Помимо этого, существенным недостатком метода является неспособность определять сдвиги более чем на 2-3 пикселя.

Известен один из самых простых и распространенных способов отслеживания изменений на видео со стационарных камер - метод вычитания фона [2]. Из текущего кадра вычитают фоновое изображение, для создания которого могут быть применены разные подходы.

Например, одним из подходов является метод медианной фильтрации [3]. В этом случае для нахождения фоновой модели используют предшествующие кадры, с помощью которых вычисляют медианные значения каждого пикселя. Недостатком этого метода является необходимость хранения нескольких предыдущих кадров, а так же возможные вычислительные затраты на поиск медианы.

Так же одним из методов вычитания фона является нахождение разности текущего и предыдущего кадра [3]. Способ прост в реализации, но его эффективность сильно зависит от скорости движения объектов и частоты кадров. Например, при большом сдвиге между кадрами не получится определить точное местонахождение объекта и оценить его размеры.

Задача отслеживания объектов на видео тесно связана с задачей нахождения контуров объектов. Известен способ [Патент RU 2016124833] определения контуров объектов на изображении, в котором путем применения различных ядер свертки выбирают затравочные пиксели на границах яркости и итеративно выстраивают контуры объекта. Способ подходит для обработки статического изображения, однако требует временных затрат и предполагает подбор множества параметров для управления определением контуров, что может привести к проблемам при обработке видеоряда.

Близким аналогом изобретения является способ селекции движущегося объекта [Патент RU 2008114730/28], в котором обрабатывают три последовательных кадра. Из каждого кадра предварительно формируют контурные изображения. Путем вычитания из последнего кадра предыдущих двух, получают промежуточные разностные изображения, которые затем используют для вычисления итоговых разностных изображений, содержащих только контуры движущегося объекта. Способ подходит для обработки видео, однако его недостатком является необходимость обработки трёх кадров в каждый момент времени. Кроме того предварительное формирование контурных изображений требует больших временных и вычислительных затрат, так как для получения разностных изображений контуры должны быть достаточно точными и не прерываться (в частотности контуры фона должны совпадать на всех

изображениях).

В основу изобретения лег метод моделирования динамики механической системы с предварительным распределением (Predetermined equilibrium driven dynamics [4]). Суть данного метода заключается в восстановлении произвольного наперёд заданного распределения $w_Q(\vec{r})$ итеративными модификациями (i - номер итерации) другого, текущего распределения $w_i(\vec{r})$. Это реализуют путем выстраивания в фазовом пространстве квазинепрерывной траектории, на каждой итерации выражаемой своим распределением $w_i(\vec{r})$. Для осуществления каждого нового шага рассматривают актуальную точку на траектории и возможные варианты следующего её положения. Выбор между этими вариантами осуществляют путем приведения системы в пробные состояния и вычисления для них степени отличия текущего $w_i(\vec{r})$ и равновесного $w_Q(\vec{r})$ распределений:

$$H = \sum_{\vec{r}} [w_Q(\vec{r}) - w_i(\vec{r})] \cdot \theta [w_Q(\vec{r}) - w_i(\vec{r})]. \quad (1)$$

Где θ - функция Хевисайда.

Переход системы осуществляют в точку с наименьшим значением H .

При создании изобретения в качестве распределений рассматривались черно-белые графические изображения, а свойство перемещения актуальной точки траектории в сторону наиболее интенсивных изменений при смене кадра, чаще всего выражающееся в движении объектов, стало основополагающим в заявленном способе. Поскольку технологический интерес представляет именно актуальная точка, а не вся траектория, для этой точки вводится название «агент», который характеризуется текущей позицией на изображении.

Техническим результатом изобретения является (по сравнению с прототипом, см. выше) повышение производительности, уменьшение затрачиваемых ресурсов и повышение точности отслеживания объектов на видео.

Исходные данные представляют в виде двух двумерных массивов размерами $W \times H$, для одного из которых вводят название «эталонное распределение» (которое воспроизводится), а для другого «текущее распределение» (где воспроизводится эталонное распределение). Эталонное распределение задают матрицей значений яркости пикселей текущего кадра из видеопотока.

Вводится понятие нормы для эталонного и текущего распределения как суммы всех их элементов:

$$\tilde{N} = \sum_{xy} \tilde{n}_{xy}, \text{ где } \tilde{n}_{xy} \in N - \text{ элемент массива эталонного распределения в точке } (x, y).$$

$$N_i = \sum_{xy} n_{xy}^i, \text{ где } n_{xy}^i \in N - \text{ элемент массива текущего распределения в точке } (x, y).$$

Где i - номер итерации.

Из ранее записанного выражения (1) с учетом приближения $N_{i+1} \approx N_i$ может быть выведена формула для оценки изменения степени отличия текущего и эталонного распределений в результате пробного шага:

$$\Delta K_i^p = - \left(\tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} n_{x_p y_p}^i \right)_+ + \left(\tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} (n_{x_p y_p}^i + 1) \right)_+$$

Где операция $(x)_+ = \max(0, x)$.

Наименьшему значению N в оригинальном методе [4] будет соответствовать минимальное значение ΔK_i .

Пусть $a_i^p = \tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} n_{x_p y_p}^i$, тогда $\Delta K_i^p = -(a_i^p)_+ + \left(a_i^p - \frac{\tilde{N}}{N_i}\right)_+$

Таким образом, параметр a_i^p однозначно определяет ΔK_i и может быть использован в качестве критерия выбора пробного шага - переход осуществляют в точку с максимальным значением a_i^p . Для данного параметра вводится название «величина локального отличия».

Обе схемы (с параметром ΔK_i и a_i^p) могут быть успешно применены для обработки изображений, однако вторая (с параметром a_i^p) позволяет избежать ситуаций, в которых выбор между несколькими вариантами направлений шага происходит случайно. В настоящем способе будет использоваться вторая схема.

Для каждого кадра в видеопотоке (кадр задаётся соответствующим эталонным распределением) итерационно строят массив D размерами $w \times h$ ($w \leq W, h \leq H$) - распределение величины локального отличия. Начальный массив D для одного кадра представляют в виде матрицы, все элементы которой равны нулю. Итоговый массив будет отмасштабирован и наложен на само изображение.

Процедура обработки одного кадра в видеопотоке происходит следующим образом:

1. Выбирают несколько стартовых точек текущего распределения $(x_0^i, y_0^i)_{i=0}$.

На каждой итерации:

2. Для каждой точки (агента) рассматривают четыре смежные ячейки, для которых вводится название «пробные ячейки»: $x_p^i = x_0^i + \Delta x$, $y_p^i = y_0^i + \Delta y$, где

$\Delta y, \Delta x = -1, 0, 1$, не учитывая случаи $|\Delta y| = |\Delta x|$. Для каждой пробной ячейки

вычисляют величину локального отличия a_i^p ($p = 0, 1, 2, 3$):

$$a_i^p = \tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} n_{x_p y_p}^i$$

В случае если значение a_i^p положительно, в соответствующий элемент массива распределения величины локального отличия прибавляют a_i^p :

$$D_{x_p/\Delta W, y_p/\Delta H}^{new} = D_{x_p/\Delta W, y_p/\Delta H}^{old} + \max(a_i^p, 0)$$

Где $\Delta W = W/w$, $\Delta H = H/h$.

3. Выбирают пробную ячейку с максимальным значением a_i^p , в которую осуществляют переход агента - обновляют его координаты $x_0^{i+1} = x_0^i + \Delta x$,

$y_0^{i+1} = y_0^i + \Delta y$ и прибавляют единицу в эту точку $n_{x_p y_p}^{i+1} = n_{x_p y_p}^i + 1$ (соответственно норма увеличивается на единицу $N_{i+1} = N_i + 1$).

4. Шаг 2-3 повторяют для каждого агента.

5. Шаг 2-4 повторяют до получения сходимости.

6. Извлеченные данные из массива D интерпретируют как итоговый результат обработки текущего кадра: максимумы формируемого распределения указывают на движущиеся объекты, изображенные на кадре.

Таким образом, покадровая развёртка результата работы способа - ряд распределений в массиве D, получаемых для каждого кадра отдельно.

Заявленный способ обладает множеством преимуществ. Способ отличается тем, что его производительность не зависит от размера кадра, так как в рамках одной итерации вычисления производят не над всей матрицей изображения, а только с локальными её областями, в которых находятся агенты. Обработав всего два кадра, можно гарантированно выделить движущиеся объекты. В этом состоит принципиальное отличие заявленного способа от многих известных аналогов, требующих анализа как минимум трёх последовательных кадров для обнаружения объекта. Способ достаточно прост в реализации и легко может быть распараллелен за счет аддитивности массива распределения локального отличия. Кроме того, не требуется предварительная обработка изображений, что является преимуществом перед многими другими существующими подходами.

Способ осуществляют следующим образом:

Из текстового файла, хранящегося на жестком диске компьютера, загружаются основные параметры программы, реализующей описанный способ (данные о размере изображения, числе итераций, количестве агентов и размере массива распределения величины локального отличия) в оперативную память компьютера.

Текущий набор кадров, представленный числовыми массивами значений яркости пикселей, загружается для дальнейшего анализа в оперативную память компьютера.

Текущее распределение задают массивом с элементами целочисленного типа.

Для каждого кадра производят вычисления, описанные выше. Полученные массивы распределения величины локального отличия интерпретируют как итоговый результат.

В качестве примера, подтверждающего возможность достижения технического результата, приведен результат работы изобретения на фиг.2. Способ был протестирован на видео с камеры на Ленинском проспекте в городе Йошкар-Ола [5], и взяты кадры разрешением 1080×720. Для наглядности матрица распределения величины локального отличия (сетка размерами 108×72) наложена на само изображение.

На Фиг.1. приведена блок-схема способа.

На Фиг.2. приведены результаты обработки кадров.

Используемые источники

1. Сакович И.О., Белов Ю.С. Применение метода Лукаса - Канаде для вычисления оптического потока. Инженерный журнал: наука и инновации, 2014, вып. 7. URL: <http://engjournal.ru/catalog/pribor/optica/1275.html>

2. Piccardi, M. (2004). Background subtraction techniques: A review. URL: http://profs.sci.univr.it/~cristanm/teaching/sar_files/lezione4/Piccardi.pdf (PDF).

3. Tamersoy, B. (September 29, 2009). Background Subtraction - Lecture Notes. University of Texas at Austin. URL: https://www.cs.utexas.edu/~grauman/courses/fall2009/slides/lecture9_background.pdf (PDF).

4. I.V. Kashin, AIP Conf. Proc. 2313, 030048 (2020). URL: <https://doi.org/10.1063/5.0032348>

5. Камера на Ленинском проспекте в городе Йошкар-Ола. URL: <https://www.marimedia.ru/video/livecams/kamera-na-leninskom-prospekte/>

6. David J. Fleet, Yair Weiss. Optical Flow Estimation. URL: <https://www.cs.toronto.edu/~fleet/research/Papers/flowChapter05.pdf> (PDF)

5

(57) Формула изобретения

1. Способ отслеживания объектов на видео, включающий регистрацию последовательных цифровых изображений и дальнейшую их обработку для выделения движущихся объектов, отличающийся тем, что для обработки одного кадра, зарегистрированного в текущий момент времени стационарной камерой, представленного в виде изображения в градациях серого, загруженного в оперативную память компьютера как двумерный числовой массив \tilde{n} размерами $W \times H$, осуществляют итерационный процесс, в ходе которого производят преобразования над массивом n^i , содержащим элементы целочисленного типа, и размеры которого совпадают с размерами изображения, а также массивом D размерами $w \times h$, элементы которого равны нулю и $w \leq W, h \leq H$, а именно:

для каждого произвольно выбранного на начальном этапе стартового элемента n_{xy}^i рассматривают четыре смежных элемента - пробных ячейки $n_{x_p y_p}^i, x_p = x + \Delta x, y_p = y + \Delta y$, где $\Delta y, \Delta x = -1, 0, 1$, не учитывая случаи $|\Delta y| = |\Delta x|$, и вычисляют величину локального отличия в соответствующей ячейке

$$a_i^p = \tilde{n}_{x_p y_p} - \frac{\tilde{N}}{N_i} n_{x_p y_p}^i$$

где i - номер итерации, p - номер пробной ячейки, \tilde{N} - сумма всех элементов массива \tilde{n} , N_i - сумма всех элементов массива n^i ;

к элементу массива D , соответствующему пробной ячейке, добавляют значение локального отличия, в случае если оно положительно

$$D_{x_p/\Delta W, y_p/\Delta H}^{new} = D_{x_p/\Delta W, y_p/\Delta H}^{old} + \max(a_i^p, 0)$$

где $\Delta W = W/w, \Delta H = H/h$;

в качестве направления следующего шага выбирают пробную ячейку с наибольшим значением a_i^p , если таких ячеек несколько, то среди них выбирают случайную; к выбранному элементу массива n^i добавляют единицу и значение N_i так же увеличивают на единицу;

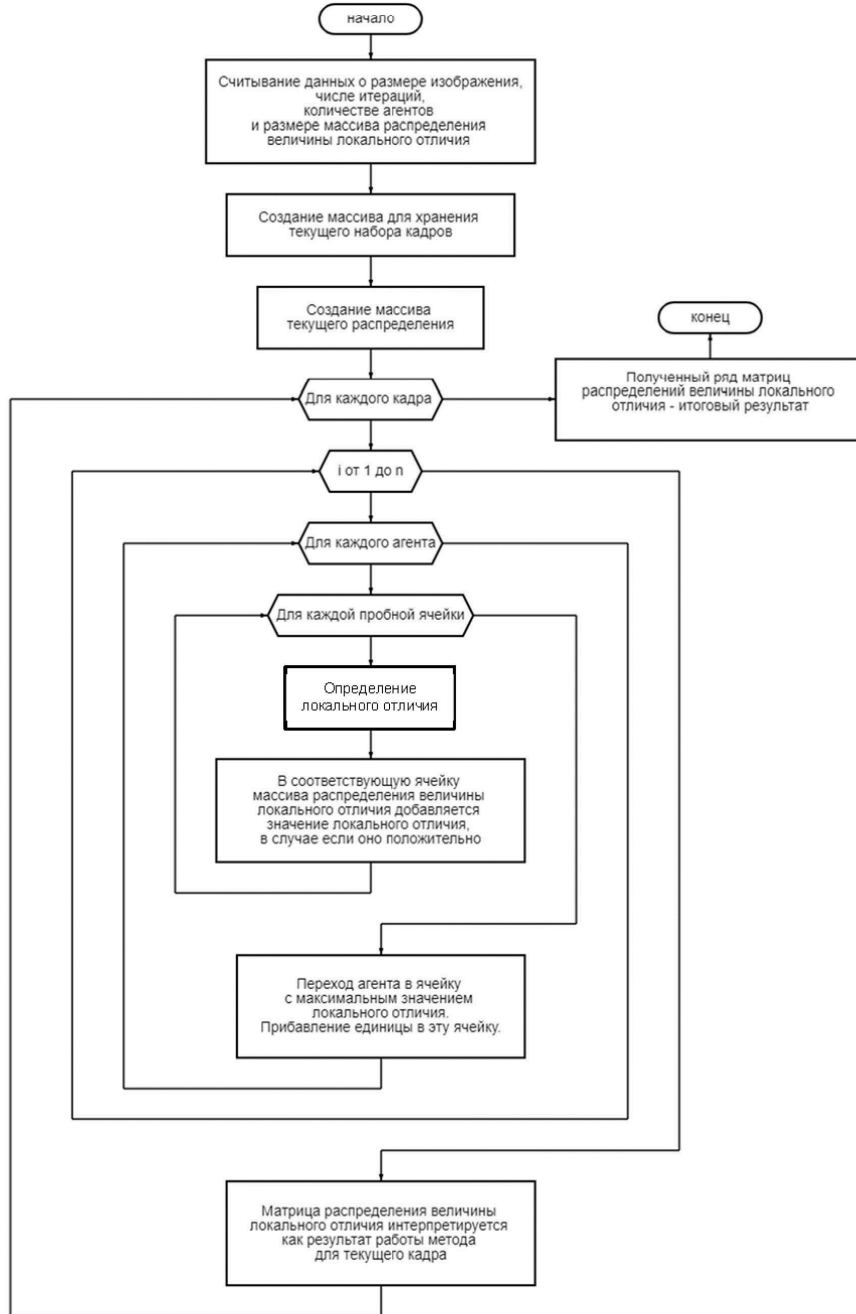
после завершения итерационного процесса все элементы массива n^i умножают на отношение $\frac{\tilde{N}}{N_i}$;

полученную матрицу D интерпретируют как итоговый результат обработки текущего кадра посредством масштабирования и наложения матрицы D на само изображение, при этом определяют максимумы сформированного распределения матрицы D , указывающие на движущиеся объекты, изображенные на кадре;

массив n^i используют для обработки следующего кадра.

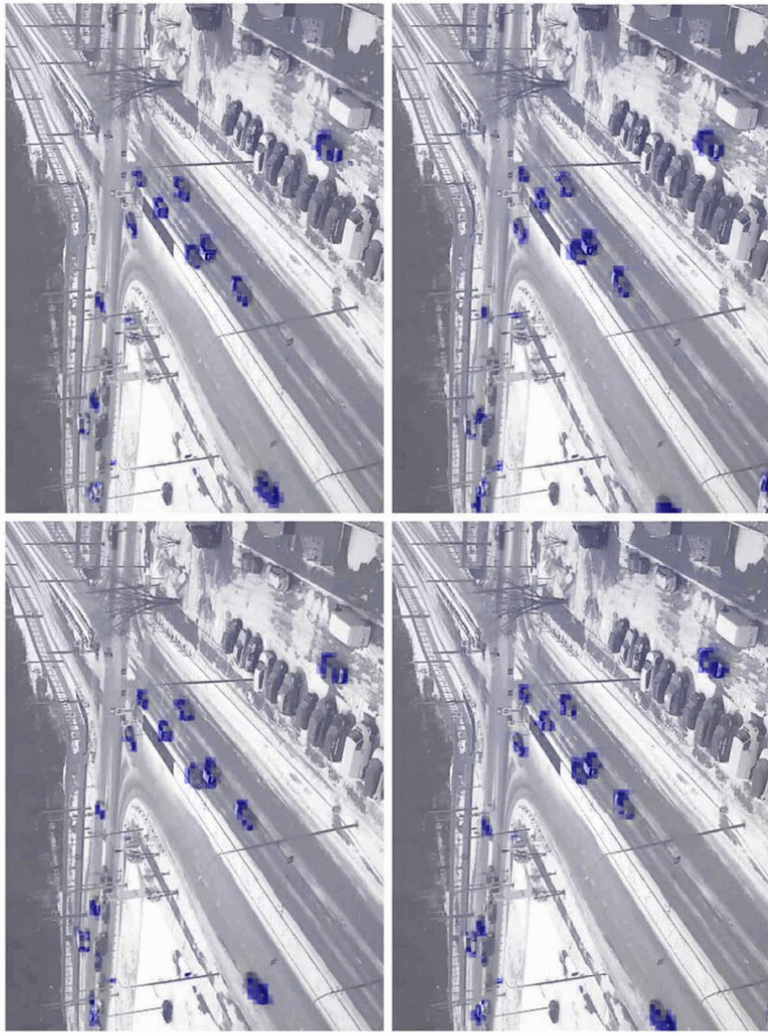
2. Способ отслеживания объектов на видео по п.1, отличающийся тем, что регулируют размеры массива D в зависимости от желаемой точности выделения объектов.

1



Фиг. 1

2



Фиг. 2