УДК 528.88

В.Е. Дементьев, Д.С. Кондратьев, А.С. Стрельцова

ТРАЕКТОРНАЯ ОБРАБОТКА НА ОСНОВЕ КОМБИНАЦИИ БАЙЕСОВСКИХ НЕЙРОННЫХ СЕТЕЙ И ПРОЦЕДУР НЕЛИНЕЙНОЙ ФИЛЬТРАЦИИ¹

Дементьев Виталий Евгеньевич, кандидат технических наук, окончил радиотехнический факультет Ульяновского государственного технического университета, заведующий кафедрой «Телекоммуникации» УлГТУ. Имеет монографии, статьи и патенты в областях статистической обработки многомерных сигналов и защиты информации. [e-mail: dve@ulntc.ru].

Кондратьев Дмитрий Сергеевич, аспирант кафедры «Телекоммуникации» УлГТУ, окончил радиотехнический факультет УлГТУ. Имеет работы в области искусственного интеллекта и обработки изображений. [e-mail: kondratev.dmitriy@gmail.com].

Стрельцова Анастасия Сергеевна, аспирант кафедры «Телекоммуникации» УлГТУ, окончила факультет математики, информатики и авиационных технологий Ульяновского государственного университета. Инженер-программист 2 категории ФНПЦ АО «НПО «Марс». Имеет научные труды в области тематического картографирования многозональных спутниковых изображений, а также в области защиты информации при помощи стеганографических методов. [e-mail: nastya94strel@mail.ru].

Аннотация

В работе ставится задача по разработке программно-алгоритмических решений, позволяющих производить автоматическое обнаружение и сопровождение объектов на видеопоследовательностях, получаемых с беспилотных летательных аппаратов. Для этого исследуются наиболее производительные нейросетевые алгоритмы, ориентированные на сегментацию изображений и выделение на них объектов разного рода. Делается вывод о важных преимуществах байесовских нейросетевых процедур, позволяющих выполнять объединение результатов их работы с другими алгоритмами. Определяются количественные характеристики эффективности таких комбинаций. Особое внимание уделяется вопросам построения оценки возможного расположения объекта на последующих кадрах с учетом особенностей его движения. Для повышения качества траекторного слежения за объектами предлагается объединять результаты работы байесовских нейросетевых алгоритмов и процедур нелинейной калмановской фильтрации, ориентированных на сопровождение объектов, которые перемещаются в том числе с переменным ускорением. Рассматриваются вопросы программной реализации предложенной комбинации алгоритмов. Выполняется сравнительный анализ полученного алгоритма траекторного слежения с известными решениями, которые реализованы в общедоступных библиотеках OpenCV. Результаты такого сравнения позволяют сделать вывод о существенных преимуществах предлагаемого алгоритма, что дает возможность рекомендовать его в реальных системах обработки видеоданных.

Ключевые слова: сегментация изображений, нейронные сети, нелинейный фильтр, обнаружение объектов, траекторное слежение.

doi:

THE OBJECT-TRAJECTORY TRACKING ALGORITHM BASED ON THE COMBINATION OF BAYESIAN NEURAL NETWORKS AND NONLINEAR KALMAN FILTRATION PROCEDURES

Vitalii Evgenevich Dementev, Candidate of Sciences in Engineering; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; Head of the Department of Telecommunications of UISTU; an author of articles and patents in the field of statistical processing of multidimensional signals and information protection. e-mail: dve@ulntc.ru.

Dmitrii Sergeevich Kondratev, Postgraduate Student at the Department of Telecommunications of UISTU; graduated from Radioengineering Faculty of UISTU; an author of articles in the field of artificial intelligence and image processing. e-mail: kondratev.dmitriy@gmail.com.

¹ Работа поддержана грантами РФФИ №19-29-09048 и №18-47-730009.

Anastasiia Sergeevna Streltsova, *Postgraduate Student at the Department of Telecommunications of UISTU; graduated from the Faculty of Mathematics, Informatics, and Aviation Technologies; Software Engineer at FRPC JSC 'RPA Mars', an author of articles in the field of thematic mapping the multi-zone satellite imagery as well as in information protection by applying the steganographic methods. e-mail: nastya94strel@mail.ru.*

Abstract

The paper aims to develop software and algorithmic decisions allowing to detect automatically and track objects on video sequences obtained from unmanned aerial vehicles. For this purpose, authors studied the most productive neural network algorithms focused on the image segmentation and selection of various objects shown in these images. The article concluded that the advantages of Bayesian neural network procedures are important, which allow combining their operation results with other algorithms. The effectiveness of such combinations is quantified. Special attention is paid to the assessment of the possible object location on subsequent frames, taking into considerations its motion characteristics. To improve the object trajectory tracking, it is proposed to combine the results of Bayesian neural network algorithms and nonlinear Kalman filtering procedures aimed to track the objects moved including with variable acceleration. The article also deals with problems of software implementation of the proposed combination of algorithms. A comparative analysis of the obtained trajectory tracking algorithm with known solutions implemented in public OpenCV libraries is performed. The results of this comparison allow us to conclude that the proposed algorithm has significant advantages, so it opens us the possibility to recommend its application in real image-data processing systems.

Key words: image segmentation, neural networks, nonlinear filter, object detecting, trajectory tracking.

ВВЕДЕНИЕ

Среди задач обработки статистической информации особое место занимает задача обработки изображений и видеопоследовательностей. Это связано с широкой областью применения методов обработки изображений. Эти методы в настоящее время используются при обработке данных дистанционного зондирования земли, медицинских изображений, видеоданных и т. д. Кроме этого, элементы алгоритмов машинного зрения используются при обработке многомерных статистических данных, в радиотехнических и эконометрических приложениях. В последние годы наряду с классическими методами обработки изображений, связанными с модельно-ориентированным подходом [1, 2], контурным [3] и спектральным анализом [4], широкое распространение получили подходы, ориентированные на использование обучаемых нейросетевых конструкций. Наиболее впечатляющих успехов подобные алгоритмы достигли в задачах распознавания образов и сегментации изображений [5]. Считается [6], что наиболее эффективные сверточные нейронные сети [7, 8] идентифицируют объекты на плоских изображениях с точностью, превышающей возможности среднестатистического человека. В то же время подобные сети обладают рядом недостатков, препятствующих их широкому применению в иных задачах. Среди этих недостатков можно выделить ориентированность на конкретный тип обрабатываемых данных и сложности при определении ошибок обработки. Первый недостаток связан с исключительной сложностью обобщения имеющегося нейросетевого алгоритма обработки, например, плоского изображения, для случая обработки видеопоследовательности подобных изображений. Фактически на первый взгляд тривиальное увеличение размерности обрабатываемых данных приводит к не-

обходимости кардинального переосмысления структуры используемой нейронной сети, переформированию процедур обучения и существенному изменению характеристик работы алгоритмов. Вторым недостатком определяется известной категоричностью современных нейронных сетей, особенно их сверточных вариаций. При использовании подобных сетей для идентификации сложных объектов результатом, как правило, является вероятность наличия объекта на текущем изображении, близкая к единице или нулю. Подобное отсутствие сомнения приводит к почти непреодолимым сложностям при оценке эффективности работы нейросетевых алгоритмов на конкретных изображениях, что может иметь критическое значение при решении задач, например, автоматического управления подвижными объектами.

В последние годы одним из важных направлений работы исследователей являются глубокие байесовские нейросетевые конструкции [9–12]. Такой подход предполагает направленный синтез и использование ансамбля нейронных сетей, каждая из которых формирует собственный результат обработки. Интегрирование этих результатов позволяет получить совокупности оценок и характеристик достоверности этих оценок. Это в свою очередь позволяет формировать объективные представления о наблюдаемом сигнале и составляющих его объектах.

В настоящей работе рассматривается использование комбинации байесовских сверточных нейронных сетей Bayesian SegNet [11] и YOLO v3 [12] для решения задачи обнаружения и сопровождения подвижных объектов на последовательностях аэрокосмических изображений. При этом для повышения качества обработки предлагается использовать комбинацию результатов обработки и априорного прогноза поведения объекта, учитывающего в том числе особенности динамики этого объекта.

ОБЪЕДИНЕНИЕ РЕЗУЛЬТАТОВ БАЙЕСОВСКИХ НЕЙРОННЫХ СЕТЕЙ ПРИ ОБРАБОТКЕ ВИДЕОИЗОБРАЖЕНИЯ

Рассмотрим следующую задачу. Предположим, что имеется видеопоследовательность, состоящая из n отдельных видеокадров, зарегистрированных с интервалом времени T . Пусть каждый видеокадр представляет собой массив значений яркости размером $N_1 \times N_2$. Выполним обработку каждого из этих кадров байесовской сверточной нейронной сетью. Пусть в результате этой обработки было выявлено S_t объектов, соответствующих одному из наперед заданных M типов, где t – момент регистрации обрабатываемого кадра. Кроме этого, особенностью байесовских нейронных сетей типа SegNet и YOLO v3 является возможность получения оценки вероятности принадлежности конкретного элемента (пикселя) изображения к тому или иному объекту. В работе [11] показана адекватность и состоятельность подобных оценок для изображений разного вида. Эти же оценки позволяют простым образом [12] объединять результаты работы двух нейросетевых алгоритмов, позволяя получать совместные оценки вероятности принадлежности элемента кадра тому или иному объекту. При этом важным моментом является тот факт, что при такой постановке предполагается, что один и тот же элемент изображения может принадлежать разным объектам. Обратим внимание, что наличие в качестве результата обработки и имеющихся на изображении объектов и информации о вероятности принадлежности элементов изображения к этим объектам позволяет потенциально получить принципиально новые алгоритмы, позволяющие, в том числе, решать задачи траекторного слежения за объектами на видеопоследовательностях.

Для этого будем считать, что для каждого из выделенных на изображении объектов можно определить функционал $F_i(J, t)$, где i – номер типа объекта; J – текущее множество элементов, составляющих изображение этого объекта; t – время, прошедшее с момента регистрации объекта; возвращающий массив времени размером $N_1 \times N_2$, значения которого равны априорной вероятности того, что соответствующий элемент массива будет принадлежать i -му объекту. Для иллю-

Рис. 1. Изображение объекта и поле вероятностей его нахождения

страции особенностей функционала $F_i(J, t)$ на рисунке 1а приведен фрагмент изображения, содержащего автомобиль (1а) и приведенный по уровню яркости массив $F_i(J, t)$, где t соответствует одной секунде (1б).

Для получения $F_i(J, t)$ с помощью описанной выше комбинации нейронных сетей типа SegNet и YOLO v3 и последующей ручной валидации была выполнена обработка представительного объема (2 тыс. отдельных кадров) видеоизображений, полученных с беспилотных летательных аппаратов с целью выделения и сопоставления изображений движущегося автомобиля.

Пусть теперь в результате обработки первого кадра байесовским ансамблем нейронных сетей выявлено S_1 объектов, которые можно представить в виде вектора $\bar{o} = \{o_1 \ o_2 \ \dots \ o_{s_1-1} \ o_{s_1}\}^T$, элементами которого являются номера соответствующих типов объектов. Например, если в результате обработки на изображении выявлено четыре объекта: поле (тип объекта 1), дорога (тип объекта 2) и два автомобиля (тип объекта 3), то вектор $\bar{o}$ примет вид $\bar{o} = \{1 \ 2 \ 3 \ 3\}^T$. Сам результат обработки представим в виде двумерного массива размером $N_1 \times N_2$, элементами которого являются векторы $\bar{p}_{ij}^{(1)}$ длиной s_1 . Индекс каждого элемента такого вектора соответствует индексу одного из выделенных элементов в векторе $\bar{o}$, а значение k -го элемента будет определяться оценкой вероятности того, что данная точка принадлежит k -му объекту. Так, например, в точке с координатами (30,30) для изображения получен вектор $\bar{p}_{30,30} = (0.01, 0.01, 0.97, 0.01)$. Непосредственная обработка такого вектора позволяет сделать вывод о том, что соответствующий пиксель с вероятностью 97% принадлежит первому экземпляру объекта «автомобиль».

Далее пусть по истечении промежутка времени в Δt зарегистрирован второй кадр изображения, который обработан аналогичным образом. При этом получено множество векторов $\{\bar{p}_{ij}^{(2)}\}$, каждый длиной s_2 , определяющих вероятность принадлежности точки с координатами (i, j) к одному из s_2 выделенных объектов. Кроме этого, для выделенных на первом этапе объектов можно применить $F_i(J, \Delta t)$ и получить соответственное множество векторов $\{\bar{q}_{ij}^{(2)}\}$, каждый из s_1 отсчетов которого определяется как $F_i(J_k, \Delta t)$, где $k=1, \dots, s_1$.

Такая постановка предполагает необходимость решения двух задач.

Во-первых, возникает вопрос относительно соответствия объектов, выделенных на основе построения априорного прогноза и непосредственной обработки кадра. Для решения этой задачи для каждого выделенного на втором кадре объекта построим характеристики

$$R_{m,k}^{(2)} = \sum_{i,j \in J_k} \bar{q}_{ij}^{(2)} m, \quad m \in S_1^k, \quad (1)$$

где S_1^k – множество индексов объектов, выделенных на первом кадре, такого же типа как и k -й объект, вы-

деленный на втором кадре. Среди всех характеристик $\{R_m^{(2)}\}$ выберем максимальную. Тогда соответствующий этой характеристике индекс обозначим как $m_{max,k}^{(2)}$. По смыслу $m_{max,k}^{(2)}$ является индексом объекта, выделенного на первом кадре, которому с наибольшей вероятностью соответствует объект с индексом k , выделенный на втором кадре. Введенная операция позволяет осуществлять покадровое траекторное слежение за объектом. На практике удобно выполнять переиндексирование объектов, так чтобы одному и тому же объекту соответствовал один и тот же индекс на протяжении процесса обработки последовательности из большого количества кадров. Кроме этого, разумным является дополнить векторы $\{\bar{p}_{ij}^{(2)}\}$ и $\{\bar{q}_{ij}^{(2)}\}$ нулевыми элементами, соответствующими недостающим объектам на первом и втором кадрах соответственно. То есть, например, если на втором кадре выявлен новый объект и, соответственно, $s_2 = s_1 + 1$, то можно увеличить длину вектора $\{\bar{q}_{ij}^{(2)}\}$ до s_2 , присвоив всем значениям $\bar{q}_{ij}^{(2)}$ нулевую величину. По смыслу эта операция означает нулевую вероятность присутствия на первом кадре нового объекта, выявленного только на втором кадре.

Во-вторых, необходимо определить способ объединения фактических наблюдений $\{\bar{p}_{ij}^{(2)}\}$ и соответствующего априорного прогноза $\{\bar{q}_{ij}^{(2)}\}$. Введенные обозначения позволяют выполнить это объединение на основе байесовской формулы:

$$\bar{p}_{ij}^{(2)} = \frac{\bar{p}_{ij}^{(2)} \bar{q}_{ij}^{(2)}}{\sum_{l=1}^{s_2} \bar{p}_{ij}^{(2)} \bar{q}_{ij}^{(2)}}. \quad (2)$$

Множество векторов $\{\bar{p}_{ij}^{(2)}\}$ может быть использовано в качестве опорного для построения априорного прогноза при обработке третьего кадра.

Представленное решение обладает существенным недостатком, связанным с невозможностью учитывать динамику наблюдаемых объектов. Действительно, если, например, автомобиль перемещается с определенной скоростью и ускорением, то разумным является строить прогноз с учетом этих скорости и ускорения. Для решения этой задачи для каждого из выделенных на изображении объектов определим геометрический центр. Для этого можно использовать средние значения для соответствующих относительных координат элементов, входящих в объект:

$$z_{xk}^t = \text{avg}_{D_k^t}(i), \quad z_{yk}^t = \text{avg}_{D_k^t}(j),$$

где D_k^t – область, соответствующая объекту с индексом k в момент времени t . Для ее получения можно, например, выбрать все элементы изображения, для которых $\bar{p}_{ij}^{(t)}$ принимает наибольшее значение по сравнению со всеми остальными $\bar{p}_{ij}^{(t)}$, $l \neq k$, которые с наибольшей вероятностью принадлежат k -му объекту.

Ясно, что координаты (z_{xk}^t, z_{yk}^t) являются не точными координатами объекта, а наблюдениями, искаженными шумами. Причиной искажающих шумов могут быть ошибки дискретизации при регистрации цифровых изображений, ошибки обработки и т. д. Обычно эти ошибки могут быть описаны белым шумом. Тогда имеют место следующие соотношения:

$$z_{xk}^t = z_k^t + n_{xk}^t, \quad z_{yk}^t = y_k^t + n_{yk}^t,$$

где (z_k^t, y_k^t) – истинные координаты центра объекта; n_{xk}^t и n_{yk}^t – отсчеты белого шума.

Будем считать, что наблюдаемые на видеоизображениях объекты могут перемещаться с некоторым случайным ускорением. Тогда имеет место следующие соотношения, определяющие динамику объекта:

$$\begin{aligned} x_k^t &= x_k^{t-1} + v_{xk}^{t-1} + a_{xk}^t (x_k^{t-1} - x_k^{t-2}) = \\ &= 2x_k^{t-1} - x_k^{t-2} + a_{xk}^t (x_k^{t-1} - x_k^{t-2}), \\ y_k^t &= y_k^{t-1} + v_{yk}^{t-1} + a_{yk}^t (y_k^{t-1} - y_k^{t-2}) = \\ &= 2y_k^{t-1} - y_k^{t-2} + a_{yk}^t (y_k^{t-1} - y_k^{t-2}), \\ a_{xk}^t &= r_{ax} a_{xk}^{t-1} + \xi_{axk}^t, \\ a_{yk}^t &= r_{ay} a_{yk}^{t-1} + \xi_{ayk}^t, \end{aligned} \quad (3)$$

где r_{ax}, r_{ay} – скалярные параметры, соответствующие скорости изменения ускорений a_{xk}^t и a_{yk}^t ; v_{xk}^t и v_{yk}^t – компоненты скорости t -го объекта в момент времени t ; ξ_{axk}^t, ξ_{ayk}^t – независимые нормальные случайные величины с нулевым математическим ожиданием и дисперсией σ_ξ^2 . Модель (3) является дважды стохастической (ДС) моделью и позволяет описывать существенно нестационарные процессы [14, 15]. Обозначим через $\bar{X}_k^t = (x_k^t \quad v_{xk}^t \quad a_{xk}^t)$, $\bar{Y}_k^t = (y_k^t \quad v_{yk}^t \quad a_{yk}^t)$.

Тогда модель (3) можно переписать в виде:

$$\bar{X}_k^t = \phi_{xk}^t \bar{X}_k^{t-1} + \xi_{xk}^t, \quad \bar{Y}_k^t = \phi_{yk}^t \bar{Y}_k^{t-1} + \xi_{yk}^t,$$

$$\begin{aligned} \phi_{xk}^t &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 + a_{xk}^{t-1} & 0 \\ 0 & 0 & r_{ax} \end{pmatrix}, \\ \phi_{yk}^t &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 + a_{yk}^{t-1} & 0 \\ 0 & 0 & r_{ay} \end{pmatrix}, \\ \xi_{xk}^t &= \begin{pmatrix} 0 \\ 0 \\ \xi_{axk}^t \end{pmatrix}, \quad \xi_{yk}^t = \begin{pmatrix} 0 \\ 0 \\ \xi_{ayk}^t \end{pmatrix}. \end{aligned}$$

Другим вариантом представленных ДС-моделей (3) могут быть следующие нелинейные векторные стохастические выражения, описывающие независимое изменение координат:

$$\bar{X}_k^t = \varphi_{xk}^t(\bar{X}_k^{t-1}) + \bar{\xi}_{xk}^t, \quad \bar{Y}_k^t = \varphi_{yk}^t(\bar{Y}_k^{t-1}) + \bar{\xi}_{yk}^t. \quad (4)$$

Тогда $\varphi_{xk}^{t'}(\bar{X}_k^{t-1}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & a_{xk}^{t-1} & v_{xk}^{t-1} \\ 0 & 0 & r_{ax} \end{pmatrix}$ и $\varphi_{yk}^{t'}(\bar{Y}_k^{t-1}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & a_{yk}^{t-1} & v_{yk}^{t-1} \\ 0 & 0 & r_{ay} \end{pmatrix}$

При этом модель наблюдений можно переписать в виде:

$$z_{xk}^t = C_x \bar{X}_k^t + n_{xk}^t, \quad z_{yk}^t = C_y \bar{Y}_k^t + n_{yk}^t, \quad k = 1, \dots, s_t,$$

где $C_x = C_y = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$.

Определим $\bar{X}_{\Delta k}^t = \varphi_{xk}^t(\bar{X}_k^{t-1})$ и $\bar{Y}_{\Delta k}^t = \varphi_{yk}^t(\bar{Y}_k^{t-1})$ – экстраполированные прогнозы координат центра k -го объекта в момент времени t по предыдущим наблюдениям. Ковариационные матрицы ошибок такой экстраполяции составят:

$$\begin{aligned} P_{\Delta xk}^t &= M \left\{ \left(\hat{X}_{\Delta k}^t - \bar{X}_{\Delta k}^t \right) \left(\hat{X}_{\Delta k}^t - \bar{X}_{\Delta k}^t \right)^T \right\} = \\ &= \varphi_{xk}^{t'}(\bar{X}_k^{t-1}) P_{xk}^{t-1} \varphi_{xk}^{t'}(\bar{X}_k^{t-1})^T + V_{x\bar{\xi}k}^t, \\ P_{\Delta yk}^t &= M \left\{ \left(\hat{Y}_{\Delta k}^t - \bar{Y}_{\Delta k}^t \right) \left(\hat{Y}_{\Delta k}^t - \bar{Y}_{\Delta k}^t \right)^T \right\} = \\ &= \varphi_{yk}^{t'}(\bar{Y}_k^{t-1}) P_{yk}^{t-1} \varphi_{yk}^{t'}(\bar{Y}_k^{t-1})^T + V_{y\bar{\xi}k}^t, \end{aligned}$$

где P_{xk}^{t-1} , P_{yk}^{t-1} – ковариационные матрицы ошибок фильтрации в момент времени $(t-1)$, $V_{x\bar{\xi}k}^t = M \left\{ \bar{\xi}_{xk}^t \bar{\xi}_{xk}^{tT} \right\}$, $V_{y\bar{\xi}k}^t = M \left\{ \bar{\xi}_{yk}^t \bar{\xi}_{yk}^{tT} \right\}$ – диагональные ковариационные матрицы случайных добавок $\bar{\xi}_{xk}^t$ и $\bar{\xi}_{yk}^t$.

Тогда можно записать следующие соотношения для дважды стохастических нелинейных координатных фильтров:

$$\begin{aligned} \hat{X}_k^t &= \hat{X}_{\Delta k}^t + B_{xk}^t (z_{xk}^t - \hat{x}_{\Delta k}^t), \\ \hat{Y}_k^t &= \hat{Y}_{\Delta k}^t + B_{yk}^t (z_{yk}^t - \hat{y}_{\Delta k}^t), \end{aligned} \quad (5)$$

где $\hat{x}_{\Delta k}^t$, $\hat{y}_{\Delta k}^t$ – первые элементы векторов $\hat{X}_{\Delta k}^t$ и $\hat{Y}_{\Delta k}^t$; $B_{xk}^t = P_{\Delta xk}^t C_x^T D_{xk}^{t-1}$; $B_{yk}^t = P_{\Delta yk}^t C_y^T D_{yk}^{t-1}$; $D_{xk}^t = C P_{\Delta xk}^t C^T + \sigma_n^2$; $D_{yk}^t = C P_{\Delta yk}^t C^T + \sigma_n^2$. Дисперсия ошибки фильтрации на каждом шаге определяется матрицами $P_{xk}^t = (E - B_{xk}^t) P_{\Delta xk}^t$, $P_{yk}^t = (E - B_{yk}^t) P_{\Delta yk}^t$.

Анализ полученных соотношений позволяет получить следующий алгоритм обработки видеопоследовательности. Пусть к текущему моменту времени выполнена обработка t кадров.

1. Тогда для каждого k -го из выделенных s_t объектов на кадре, зарегистрированном в момент времени $t+1$, выполняется построение прогноза $(\bar{X}_{\Delta k}^{t+1}, \bar{Y}_{\Delta k}^{t+1})$ координат, скорости и ускорения на основе информации с предыдущих кадров.

2. С учетом полученного прогноза и особенностей поведения объекта, которые могут быть определены функционалом $F_i(J, t)$, строится совокупность векторов $\{\bar{q}_{ij}^{(t+1)}\}$. Практика показывает, что для этого в большинстве случаев достаточно выполнить преобразование $F_i(\bar{J}, t)$, где $\bar{J}$ – совокупность элементов изображения объекта с центром в точке $(\bar{X}_{\Delta k}^{t+1}, \bar{Y}_{\Delta k}^{t+1})$.

3. Выполняется обработка кадра, зарегистрированного в момент времени $t+1$, с помощью байесовской сверточной нейронной сети. В результате получим совокупность векторов $\{\bar{p}_{ij}^{(t+1)}\}$.

4. Производится объединение априорной информации, содержащейся в массиве $\{\bar{q}_{ij}^{(t+1)}\}$, и результатов обработки наблюдений $\{\bar{p}_{ij}^{(t+1)}\}$. Результатом такого объединения будет совокупность векторов $\{\bar{p}_{ij}^{(t+1)}\}$, для которых процедура обработки повторяется.

Приведем практические результаты работы данного алгоритма. Для этого была выполнена обработка трех тестовых видеопоследовательностей, отдельные кадры которых представлены на рисунке 2.

Была выполнена оценка качества детектирования объектов на изображениях, составляющих видеопос-

Рис. 2. Кадры тестовых видеопоследовательностей

следовательность, с помощью отдельной обработки нейросетевыми алгоритмами Bayesian SegNet [12] и YOLO v3 [13] и представленным в работе алгоритмом (табл. 1).

Анализ представленных данных показывает, что эффективность работы YOLO v3 в целом выше, чем Bayesian SegNet при меньшем количестве выделенных объектов. Однако представленный в работе алгоритм позволяет существенно улучшить качество обработки (до 30% по количеству верно детектированных объектов) продолжительных видеопоследовательностей.

Был выполнен сравнительный анализ (табл. 2) представленного алгоритма траекторного слежения с процедурой, реализованной в библиотеке OpenCV, на основе последовательного применения алгоритмов Хорна-Шанка (Horn-Schunck) [13], Лукаса-Канаде (Lucas-Kanade) [14] и метода блочного сопоставления [15]. Для этого выполнялась оценка вероятности того, что выделенный на некотором кадре объект будет корректно идентифицирован и соотнесен с оригиналом на K последующих кадрах. Для повышения достоверности результатов эксперимент проводился 1000 раз.

ЗАКЛЮЧЕНИЕ

Анализ представленных данных показывает существенный выигрыш, получаемый с помощью предлагаемого подхода, в сравнении с реализованным в OpenCV алгоритмом. Так реализованная процедура траекторного слежения позволяет корректно (с вероятностью 0,97) производить последовательное соотнесение объектов на протяжении 500 кадров (более 20 секунд видео). В то же время описанный аналог в этих условиях обеспечивает вероятность корректного сопровождения только 0,68. Более того, произведенные оценки получены только для хорошо наблюдаемых контрастных объектов. Практика показывает, что первичное обнаружение малоконтрастных объектов с помощью нейросетевых алгоритмов выполняется значительно более увереннее, чем на основе классических подходов, поэтому выигрыши в этом случае будут еще более значительными.

СПИСОК ЛИТЕРАТУРЫ

1. Васильев К.К. Оптимальная обработка сигналов в дискретном времени. – М. : Радиотехника, 2016. – 288 с.

Таблица 1

Эффективность детектирования объектов на видеопоследовательности

Номер кадра	Bayesian SegNet, число корректно детектированных объектов / общее число объектов / число ложно выделенных объектов	YOLO v3, число корректно детектированных объектов / общее число объектов / число ложно выделенных объектов	Алгоритм траекторной обработки, число корректно детектированных объектов / общее число объектов / число ложно выделенных объектов
1	5/9/1	6/9/0	5/9/1
10	5/9/1	6/9/0	6/9/1
30	5/9/1	5/9/0	7/9/0
50	5/10/1	6/10/0	7/10/0
100	4/10/1	5/10/1	7/10/1
200	4/8/0	5/8/0	6/8/0
500	5/11/1	6/11/0	8/11/0

Таблица 2

Сравнение эффективности траекторного слежения

	Предлагаемый алгоритм на основе комбинации байесовских нейронных сетей и нелинейной калмановской оценки	Алгоритм на основе последовательного применения алгоритмов Хорна-Шанка, Лукаса-Канаде и метода блочного сопоставления
$K=1$	1	0,998
$K=2$	1	0,995
$K=5$	1	0,989
$K=10$	0,999	0,972
$K=50$	0,996	0,921
$K=100$	0,992	0,843
$K=500$	0,968	0,683

2. Васильев К.К., Дементьев В.Е. Представление и обработка спутниковых многозональных изображений. – Ульяновск, 2017. – 247 с.

3. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. – М.: ДМК-Пресс, 2018. – 652 с.

4. Девайкин П.А., Шикуть А.В. Вычисление оптического потока для выделения транспортных средств на видеоизображении // Инженерный журнал: наука и инновации. – 2013. – № 2 (14). – URL: engjournal.ru/articles/543/543.pdf (дата обращения: 01.09.2019).

5. Методы компьютерной обработки изображений / В.А. Сойфер, В.В. Сергеев [и др.]. – М.: Физматлит, 2003. – 784 с.

6. Обухова Н.А. Обнаружение и сопровождение движущихся объектов методом сопоставления блоков // Информационно-управляющие системы. – 2004. – № 1. – С. 30–37.

7. Фурман Я.А. Введение в контурный анализ. – М.: ФИЗМАТЛИТ, 2003. – 598 с.

8. Badrinarayanan V., Kendall A., Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv:1511.00561, 2015. – URL: <https://arxiv.org/pdf/1511.00561.pdf> (дата обращения: 01.09.2019).

9. Noh H., Hong S., Han B. Learning deconvolution network for semantic segmentation. arXiv:1505.04366, 2015. – URL: <https://arxiv.org/pdf/1505.04366.pdf> (дата обращения: 01.09.2019).

10. Christopher M. Bishop. Bayesian Neural Networks. arXiv: 1801.07710, 1997. – URL: <https://arxiv.org/ftp/arxiv/papers/1801/1801.07710.pdf> (дата обращения: 01.09.2019).

11. Kendall A., Badrinarayanan V., Cipolla R. Bayesian SegNet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv:1511.02680v2, 2016. – URL: <https://arxiv.org/abs/1511.02680> (дата обращения: 01.09.2019).

12. Li Y., Gal Y. Dropout Inference in Bayesian Neural Networks with Alpha-divergences. arXiv:1703.02914, 2017. – URL: <https://arxiv.org/pdf/1703.02914.pdf> (дата обращения: 01.09.2019).

13. Bouman C. A. Model Based Imaging Processing. Purdue University, 2013. 414 p.

14. Lucas B., Kanade T. An iterative image registration technique with an application to stereo vision // Proceedings of Imaging Understanding Workshop. 1983. pp. 121–130.

15. Redmon J., Farhad A. YOLOv3: An Incremental Improvement. – URL: <https://pjreddie.com/media/files/papers/YOLOv3.pdf> (дата обращения: 01.09.2019).

REFERENCES

1. Vasilev K.K. *Optimalnaia obrabotka signalov v diskretnom vremeni* [Optimal Signal Processing in Discrete Time]. Moscow, Radiotekhnika Publ., 2016. 288 p.

2. Vasilev K.K., Dementev V.E. *Predstavlenie i obrabotka sputnikovyykh mnogoazonalnykh izobrazhenii* [Representation and Processing of Multizone Satellite Imagery]. Ulyanovsk, 2017. 247 p.

3. Gudfellow Ja., Bengio I., Courville A. *Glubokoe obuchenie* [Deep Learning]. Moscow, DMK-Press, 2018. 652 p.

4. Devaikin P.A., Shikut A.V. Vychislenie opticheskogo potoka dlia vydeleniia transportnykh sredstv na videoizobrazhenii [Optical Flow Calculation for Vehicle Detection on Video Stream]. *Inzhenernyi zhurnal: nauka i innovatsii* [Engineering Journal: Science and Innovation], 2013, no. 2 (14). Available at: engjournal.ru/articles/543/543.pdf (accessed: 01.09.2019).

5. Soifer V.A., Sergeev V.V., et al. *Metody kompiuternoi obrabotki izobrazhenii* [Methods of Image Processing]. Moscow, Fizmatlit Publ., 2003. 784 p.

6. Obukhova N.A. Obnaruzhenie i soprovozhdenie dvizhushchikhsia obektov metodom sopostavleniia blokov [Detection and Tracking Moving Object Based on Block Matching Algorithm]. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2004, no. 1, pp. 30–37.

7. Furman Ia.A. *Vvedenie v konturnyi analiz* [Introduction to Contour Analysis]. Moscow, Fizmatlit Publ., 2003. 598 p.

8. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv:1511.00561, 2015. Available at: <https://arxiv.org/pdf/1511.00561.pdf> (accessed: 01.09.2019).

9. Noh H., Hong S., Han B. *Learning Deconvolution Network for Semantic Segmentation*. arXiv:1505.04366, 2015. Available at: <https://arxiv.org/pdf/1505.04366.pdf> (accessed: 01.09.2019).

10. Christopher M. Bishop. *Bayesian Neural Networks*. arXiv: 1801.07710, 1997. Available at: <https://arxiv.org/ftp/arxiv/papers/1801/1801.07710.pdf> (accessed: 01.09.2019).

11. Kendall A., Badrinarayanan V., Cipolla R. *Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding*. arXiv:1511.02680v2, 2016. Available at: <https://arxiv.org/abs/1511.02680> (accessed: 01.09.2019).

12. Li Y., Gal Y. *Dropout Inference in Bayesian Neural Networks with Alpha-divergences*. arXiv:1703.02914, 2017. Available at: <https://arxiv.org/pdf/1703.02914.pdf> (accessed: 01.09.2019).

13. Bouman S. A. *Model Based Imaging Processing*. Purdue University Publ., 2013. 414 p.

14. Lucas B., Kanade T. An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings of Imaging Understanding Workshop*, 1983, pp. 121–130.

15. Redmon J., Farhad A. *YOLOv3: An Incremental Improvement*. Available at: <https://pjreddie.com/media/files/papers/YOLOv3.pdf> (accessed: 01.09.2019).