

ВЫЧИСЛИТЕЛЬНЫЕ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ

УДК 004.032.26

ОТСЛЕЖИВАНИЕ ОБЪЕКТА В ВИДЕОПОТОКЕ С ПОМОЩЬЮ СВЁРТОЧНОЙ НЕЙРОННОЙ СЕТИ

© Ю. Н. Золотухин, К. Ю. Котов, А. А. Нестеров, Е. Д. Семенюк

*Институт автоматизации и электрометрии СО РАН,
630090, г. Новосибирск, просп. Академика Коптюга, 1
E-mail: kotov@idisys.iae.nsk.su*

Предложен алгоритм 6-координатного трекинга подвижного объекта по последовательности RGB-изображений на основе свёрточной нейронной сети. Обучение сети выполняется на синтезированных данных объекта с динамической моделью движения. В обратную связь от выхода ко входу сети включен фильтр Калмана для получения сглаженной оценки координат объекта. Предварительные результаты трекинга объекта на синтезированных изображениях демонстрируют эффективность предложенного подхода.

Ключевые слова: трекинг по отслеживанию, видеопоток, свёрточная нейронная сеть, фильтр Калмана.

DOI: 10.15372/AUT20200611

Введение. Задача распознавания и отслеживания объектов на последовательности кадров видеопотока востребована во многих областях робототехники, связанных с использованием систем технического зрения. В зарубежной литературе принято говорить о задаче трекинга. Различают трекинг одного или нескольких объектов, 2–6-координатный, трекинг по распознаванию и трекинг по отслеживанию [1–4].

Развитие микроэлектроники и связанное с этим появление компактных вычислительных ресурсов — микрокомпьютеров — позволяют говорить о создании мобильных систем трекинга на основе нейронных сетей, функционирующих в режимах, близких к режимам реального времени. Активно развивающейся областью практического применения систем трекинга является трекинг мобильных роботов — летательных аппаратов, колёсных платформ, захватов манипуляторов и др. [5–8].

Коммерческие системы трекинга, например VICON и Optitrack, имеют высокую стоимость [9]. Доступные решения на основе визуальных маркеров, такие как SwisTrack или библиотека `ar_track_alvar` пакета ROS, накладывают ограничения на условия освещения и возможного перемещения объекта.

Способность нейронных структур к обучению позволяет учесть априорные факторы, связанные с условиями освещения, динамикой движения и особенностями получаемого изображения объекта. Кроме того, применение нейронных сетей при распознавании объекта позволяет отказаться от использования специальных маркеров.

Одной из ключевых проблем применения нейронных сетей в задачах трекинга является проблема формирования обучающего набора данных. В настоящее время можно выделить два основных подхода к решению этой задачи: первый — использование баз данных изображений реальных сцен и объектов [10], второй — синтезированных последовательностей изображений. Как правило, второй подход используется только в задачах трекинга по распознаванию или для обучения сети шумовым искажениям изображений в задачах трекинга по отслеживанию [3]. По мнению авторов, значительный потенциал в применении

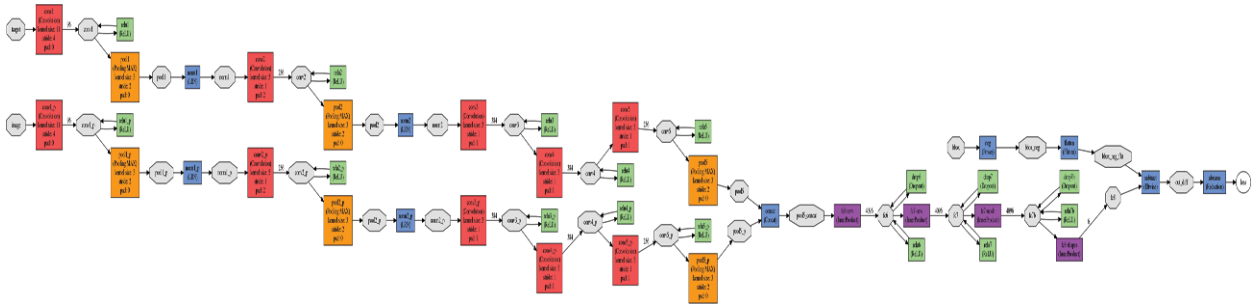


Рис. 1. Структурная схема нейронной сети для трекинга

синтезированных наборов данных состоит в возможности формирования движения синтезированного объекта на основе математического описания динамики реального объекта.

Основным методом решения задачи трекинга является детектирование объектов в кадре и их сопоставление с детекциями, полученными на предыдущих кадрах. Реализация двухэтапного метода трекинга осуществляется непосредственно с помощью нейронной сети или с использованием дополнительных методов, например определения степени близости объектов на последовательных кадрах. Наличие обратной связи от выхода ко входу сети позволяет построить регрессионную систему оценки координат объекта. В большинстве работ получения изображения объектов используются RGBD-сенсоры, область применения которых ограничена дальностью и скоростью перемещения объекта [11, 3]. Перспективным направлением исследований является изучение возможности получения полного набора координат объекта от монокулярного RGB-сенсора.

Цель данной работы — представление предварительных результатов работы предложенного алгоритма 6-координатного трекинга подвижного объекта по последовательности RGB-изображений на основе свёрточной нейронной сети. Обучение сети выполняется на синтезированных данных объекта с динамической моделью движения. В обратную связь от выхода ко входу сети включен фильтр Калмана для получения сглаженной оценки координат объекта. Предварительные результаты трекинга объекта на синтезированных изображениях демонстрируют эффективность предложенного подхода.

Описание нейронной сети и модели объекта. Для трекинга объектов на последовательности изображений используется свёрточная нейронная сеть из алгоритма GOTURN [10], реализованная в библиотеке Caffe. Размерность выходного слоя сети равна 6, что соответствует трём координатам и трём углам поворота центра масс объекта. На вход сети подаётся два изображения в текущий и предыдущий моменты времени. Сеть состоит из восьми слоёв, по пять слоёв в каждом из двух каналов обработки изображений и трёх полносвязных слоёв (рис. 1).

Для повышения устойчивости процедуры трекинга для нейронной сети, в том числе в случае наличия на изображении нескольких объектов, может быть предоставлена информация о предполагаемом местоположении объекта [10]. Как правило, это реализуется кадрированием исходного изображения по полученной оценке координат объекта на выходе сети и получением регрессионной оценки по фрагментам изображений. В качестве оценки координат объекта может быть использована оценка координат в фильтре Калмана с учётом модели движения [12].

В качестве модели движения для оценки эффективности сети примем динамическую модель вида

$$\begin{cases} \dot{V} = U; \\ \dot{X} = V \end{cases} \quad (1)$$

с законом управления

$$U = a(V - V_{ref}). \quad (2)$$

Здесь $X = [x, y, z, \psi, \phi, \theta]^T$, $V = [v_x, v_y, v_z, \omega_\psi, \omega_\phi, \omega_\theta]^T$ — вектор линейных и угловых скоростей, $a = \text{diag} \{a_x, a_y, a_z, a_\psi, a_\phi, a_\theta\}$ — матрица постоянных коэффициентов. Положение объекта в пространстве характеризуется координатами x, y, z центра масс в неподвижной декартовой системе координат и углами Эйлера ψ, ϕ, θ , определяющими ориентацию связанной с осями объекта системы координат по отношению к неподвижной системе координат xyz .

Для формирования набора данных обучения необходимо определить вектор V_{ref} . Введём ограничения на координаты объекта и будем менять скорость движения на скорость с противоположным знаком при достижении заданных границ:

$$V_{ref} = \begin{cases} v_{x_ref} = -\text{sign}(v_{x_ref})\gamma_x, & \text{если } x \leq x_{\min} \text{ или } x \geq x_{\max}, \\ v_{y_ref} = -\text{sign}(v_{y_ref})\gamma_y, & \text{если } y \leq y_{\min} \text{ или } y \geq y_{\max}, \\ v_{z_ref} = -\text{sign}(v_{z_ref})\gamma_z, & \text{если } z \leq z_{\min} \text{ или } z \geq z_{\max}, \\ \omega_{\psi_ref} = -\text{sign}(\omega_{\psi_ref})\gamma_\psi, & \text{если } \psi \leq -\pi \text{ или } \psi \geq \pi, \\ \omega_{\phi_ref} = -\text{sign}(\omega_{\phi_ref})\gamma_\phi, & \text{если } \phi \leq -\pi/2 \text{ или } \phi \geq \pi/2, \\ \omega_{\theta_ref} = -\text{sign}(\omega_{\theta_ref})\gamma_\theta, & \text{если } \theta \leq -\pi/2 \text{ или } \theta \geq \pi/2. \end{cases}$$

Здесь $\gamma_{x,y,z} \sim U(v_{\min}, v_{\max})$, $\gamma_{\psi,\phi,\theta} \sim U(\omega_{\min}, \omega_{\max})$.

Отрисовка объекта во время движения, формирование и сохранение последовательности изображений выполняется с помощью средств библиотек OpenGL и OpenCV. За основу программной реализации взят исходный код [3]. Изображение объекта помещается в точку начала координат и определяется трёхмерной моделью в формате PLY. Движение реализуется непрерывным смещением и вращением вектора, совпадающего с оптической осью камеры, направленной на объект.

Настройка сети выполнялась на обучающей выборке из 100 тысяч изображений. Аналогично работе [3] из исходных данных обучения случайным образом выбирались два последовательных кадра, которые затем поступали на вход сети. Такой подход обеспечивает уменьшение зависимости результатов обучения от вида сгенерированного в модели (1) движения. Для обучения в среде Caffe использовался метод стохастического градиентного спуска. Параметр *base_lr*, определяющий начальную скорость обучения, устанавливался 10^{-6} и каждые 50 тысяч шагов уменьшался в 10 раз. Изменение параметра *loss*, характеризующего степень приближения выходов сети к требуемым значениям, приведено на рис. 2.

Описание метода трекинга. Структура предложенной системы оценки координат подвижного объекта представлена на рис. 3.

Изображение I_k в k -й момент времени поступает на вход блока RBV, в котором выполняется вычисление прямоугольной области, ограничивающей исходное изображение, посредством рендеринга трёхмерной модели объекта в координатах X_k^{kf} полученной оценки фильтра Калмана с выхода блока АЕКФ. Изображение сохраняется в выделенной области памяти для формирования кадра в момент времени $k - 1$. Полученные области последовательных изображений I'_k и I'_{k-1} поступают на вход свёрточной нейронной сети CNN.

Построение системы трекинга с использованием фильтра Калмана позволяет формировать выделенную область с интересующим объектом на основании известной модели движения (1) и повышать точность трекинга, что и показано далее. Совместно с фильтром

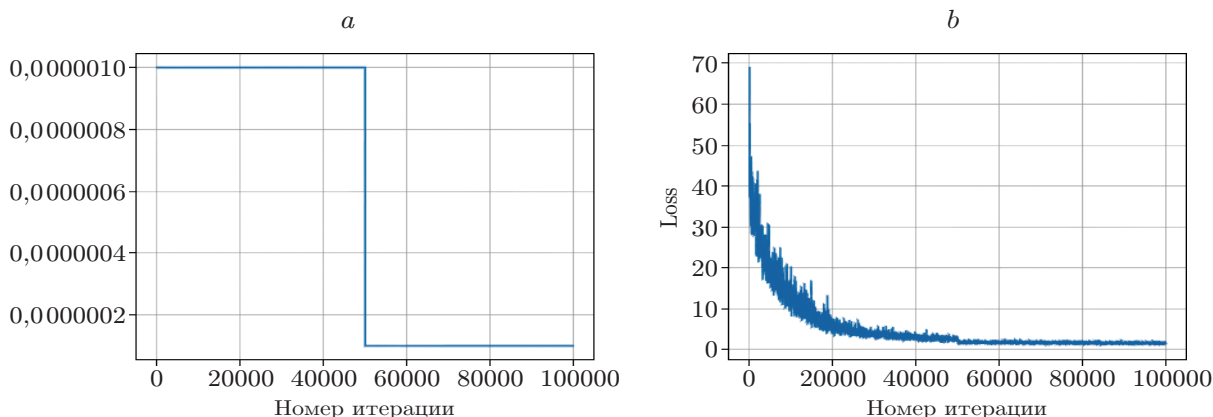


Рис. 2. Параметры обучения сети в зависимости от номера итерации: *a* — величина шага градиентного спуска *base_lr*; *b* — величина параметра *loss*

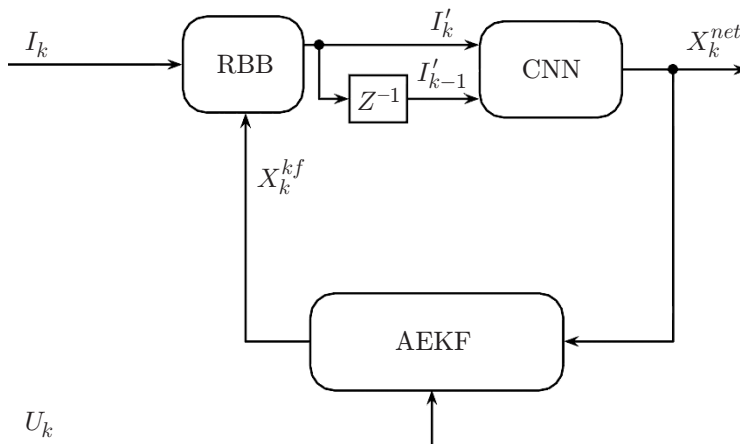


Рис. 3. Структурная схема системы трекинга

Калмана в блоке КФ применяется схема коррекции транспортного запаздывания в координатах на основе аналитического экстраполятора, как и в работе [13]. Необходимость такой коррекции вызвана тем, что предсказание по текущей координате для выделенной области должно быть сделано на основании предыдущих отсчётов. Чем меньше на практике частота следования кадров, тем больше величина запаздывания между координатой на выходе сети и требуемой оценкой. Величина коррекции определяется количеством шагов интегрирования N математической модели (1) [13].

Исходную модель (1) представим в виде

$$\begin{cases} q_{k+1} = F_k q_k + B_k U_k + w_k; \\ \hat{q}_k = C_k q_k + v_k. \end{cases} \quad (3)$$

Здесь w_k и v_k — случайные процессы, описывающие погрешности моделирования и шумы измерений с ковариационными матрицами Q_k и R_k соответственно; $q_k = [V_k \mid X_k]$; \hat{q}_k — вектор измеряемых компонент состояния q_k объекта;

$$F_k = \begin{pmatrix} I & 0 \\ I \cdot \Delta t & I \end{pmatrix}; \quad B_k = \begin{pmatrix} I \cdot \Delta t \\ 0 \end{pmatrix}; \quad C_k = (0 \ I); \quad (4)$$

Δt — шаг дискретизации по времени. Соотношения для получения оценки \hat{q}_k^{kf} текущего

состояния системы с помощью линейного фильтра Калмана имеют вид

$$\hat{q}_k^{kf} = F_k \hat{q}_{k-1}^{kf} + B_k U_k, \quad (5)$$

$$P_k = F_k P_{k-1} F_k^\top + Q_k, \quad (6)$$

$$K_k = \frac{P_k H_k^\top}{H_k P_k H_k^\top + R_k}, \quad (7)$$

$$\hat{q}_k^{kf} = \hat{q}_k^{kf} + K_k (\hat{q}_k - \hat{q}_k^{kf}), \quad (8)$$

$$P_k = (I - K_k H_k) P_k. \quad (9)$$

Отметим, что в случае нелинейной модели движения (1) необходимо использовать модификации фильтра Калмана, например расширенный или ансамбльный фильтр Калмана. Обоснование выбора типа фильтра в случае нелинейной модели движения является предметом дальнейших исследований.

Результаты моделирования. Для оценки эффективности предложенного алгоритма трекинга рассмотрим два режима работы. Первый — в обратной связи отсутствует блок АЕКФ и координаты

$$X_k^{net} = [x_k^{net}, y_k^{net}, z_k^{net}, \psi_k^{net}, \phi_k^{net}, \theta_k^{net}]^\top, \quad (10)$$

полученные на выходе сети, используются напрямую в блоке RBV. Это соответствует простейшей стационарной модели движения, в которой смещение объекта между двумя последовательными изображениями определяется только случайными факторами [10]. Второй режим соответствует полной схеме на рис. 3.

Введём метрики, характеризующие степень близости выходов сети (индекс *net*) или оценки координат $x_k^{kf}, y_k^{kf}, z_k^{kf}, \psi_k^{kf}, \phi_k^{kf}, \theta_k^{kf}$ на выходе фильтра Калмана (индекс *kf*), к координатам эталонной модели (1). В качестве метрик примем евклидовы нормы поступательной (индекс *translation*) и вращательной (индекс *rotation*) компонент соответствующих векторов:

$$E_{net,kf}^{translation} = \sqrt{((x_k^{net,kf} - x_k^{gt})^2 + (y_k^{net,kf} - y_k^{gt})^2 + (z_k^{net,kf} - z_k^{gt})^2)/3}, \quad (11)$$

$$E_{net,kf}^{rotation} = \sqrt{((\psi_k^{net,kf} - \psi_k^{gt})^2 + (\phi_k^{net,kf} - \phi_k^{gt})^2 + (\theta_k^{net,kf} - \theta_k^{gt})^2)/3}. \quad (12)$$

В качестве тестового набора данных использовалась сгенерированная выборка из 10 тысяч изображений. Рассчитанные значения среднего и среднеквадратичного отклонения по метрикам (11) и (12) сведены в табл. 1 и 2. Метрики указаны соответствующими индексами. Сравнение двух режимов работы показывает значительное улучшение точности в оценке вращательной компоненты координаты после применения фильтра Калмана (см. табл. 1), что объясняется сглаживанием выбросов по углам ψ, ϕ, θ .

В табл. 2 приведены результаты оценки влияния величины N коррекции транспортно-го запаздывания на качество трекинга. Наилучшее качество достигается при количестве шагов коррекции, равном 2. Это вызвано наличием задержки в один кадр между выходом и

Таблица 1

Сравнение режимов работы схемы на рис. 3

Метрика	Режим без АЕКФ, <i>net</i>	Режим полный, <i>kf</i>
<i>translation</i> , м	$0,010 \pm 0,008$	$0,005 \pm 0,004$
<i>rotation</i> , рад	$0,182 \pm 0,137$	$0,063 \pm 0,050$

Таблица 2

Оценка влияния количества шагов коррекции запаздывания

Метрика	$N = 0$	$N = 1$	$N = 2$	$N = 3$
<i>net, translation</i> , м	$0,007 \pm 0,007$	$0,006 \pm 0,006$	$0,005 \pm 0,004$	$0,007 \pm 0,009$
<i>net, rotation</i> , рад	$0,112 \pm 0,080$	$0,104 \pm 0,076$	$0,101 \pm 0,075$	$0,114 \pm 0,105$
<i>kf, translation</i> , м	$0,009 \pm 0,009$	$0,006 \pm 0,008$	$0,005 \pm 0,004$	$0,010 \pm 0,012$
<i>kf, rotation</i> , рад	$0,070 \pm 0,054$	$0,060 \pm 0,054$	$0,063 \pm 0,050$	$0,093 \pm 0,100$

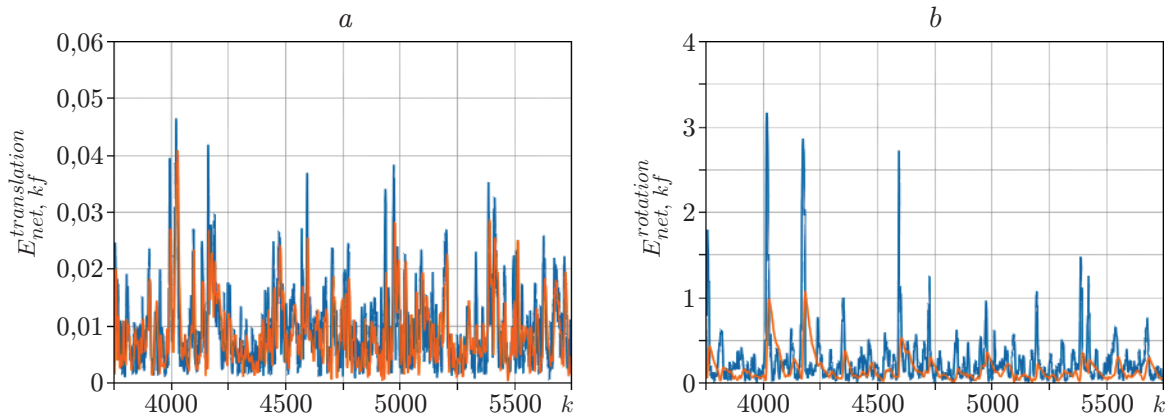


Рис. 4. Изменение метрик (11), (12) в зависимости от номера кадра: *a* — компонента *translation* (E_{net} — синяя кривая, E_{kf} — коричневая кривая); *b* — компонента *rotation* (E_{net} — синяя кривая, E_{kf} — коричневая кривая)

входом сети и ускоренным движением объекта, которое требует задания ограничивающей области на текущем кадре с упреждением.

На рис. 4 приведено изменение метрик (11) и (12) в зависимости от номера кадра при $N = 2$. Разработанная система трекинга позволяет получить точность 0,005 м и 0,063 рад по координатам и углам ориентации подвижного объекта соответственно.

В дальнейшей работе предполагается получить результаты на синтезированных зашумлённых изображениях со случайным фоном и на реальных изображениях летательного аппарата в стендовом комплексе [12].

Заключение. В данной работе предложен алгоритм 6-координатного трекинга подвижного объекта по последовательности RGB-изображений на основе свёрточной нейронной сети. Предварительное обучение сети выполняется на синтезированных данных объекта с динамической моделью движения. В обратную связь от выхода ко входу сети включен фильтр Калмана для получения сглаженной оценки координат объекта, обеспечивающий повышение точности трекинга в 1,5–2 раза. Результаты трекинга объекта на синтезированных изображениях демонстрируют эффективность предложенного подхода.

Финансирование. Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (грант № 18-58-76003 ЭРА_а) и Министерства науки и высшего образования РФ (государственная регистрация № АААА-А17-117060610006-6).

СПИСОК ЛИТЕРАТУРЫ

1. **Siam M., Singh A., Perez C., Jägersand M.** 4-dof tracking for robot fine manipulation tasks // Proc. of the 14th Conference on Computer and Robot Vision (CRV 2017). Edmonton, Canada, 16–19 May, 2017. P. 329–336.
2. **Gallego G., Lund J., Mueggler E. et al.** Event-based, 6-dof camera tracking from photometric depth maps // IEEE Trans. Pattern Analysis and Machine Intell. 2017. **40**, N 11. P. 2402–2412.
3. **Garon M., Lalonde J.-F.** Deep 6-dof tracking // IEEE Trans. Vis. Computer Graph. Cornell Univers., 2017. **23**. P. 2410–2418. URL: <http://arxiv.org/abs/1703.09771> (дата обращения: 03.03.2020).
4. **Doering A., Iqbal U., Gall J.** Joint flow: Temporal flow fields for multi person tracking // Computer Vis. and Pattern Recognition. Cornell Univers., 2018. URL: <http://arxiv.org/abs/1805.04596> (дата обращения: 03.03.2020).
5. **Zheng W., Zhou F., Wang Z.** External vision based robust pose estimation system for a quadrotor in outdoor environments // Proc. of the 3rd Intern. Conference on Pattern Recognition Applications and Methods (ICPRAM 2014). Angers, France, 6–8 March, 2014. P. 718–723.
6. **Faessler M., Mueggler E., Schwabe K., Scaramuzza D.** A monocular pose estimation system based on infrared leds // Proc. of the IEEE Intern. Conference on Robotics and Automation (ICRA 2014). Hong Kong, China, 31 May – 7 June, 2014. P. 907–913.
7. **Censi A., Strubel J., Brandli Ch. et al.** Low-latency localization by active led markers tracking using a dynamic vision sensor // Proc. of the IEEE/RSJ Intern. Conference on Intelligent Robots and Systems (IROS 2013). Tokyo, Japan, 3–7 Nov., 2013. P. 891–898.
8. **Mueggler E., Faessler M., Fontana F., Scaramuzza D.** Aerial-guided navigation of a ground robot among movable obstacles // Proc. of the 12th IEEE Intern. Symposium on Safety, Security and Rescue Robotics (SSRR 2014). Hokkaido, Japan, 27–30 Oct., 2014. P. 1–8.
9. **Millard A., Hilder J., Timmis J., Winfield A.** A low-cost real-time tracking infrastructure for ground-based robot swarms // Proc. of the 9th Intern. Conference (ANTS 2014). Brussels, Belgium, 10–12 Sept., 2014. Vol. 8667. P. 278–279.
10. **Held D., Thrun S., Savarese S.** Learning to track at 100 FPS with deep regression networks // Computer Vis. and Pattern Recognition. Cornell Univers., 2016. URL: <http://arxiv.org/abs/1604.01802> (дата обращения: 03.03.2020).
11. **Tan D. J., Navab N., Tombari F.** 6d object pose estimation with depth images: A seamless approach for robotic interaction and augmented reality // Computer Vis. and Pattern Recognition. Cornell Univers., 2017. URL: <http://arxiv.org/abs/1709.01459> (дата обращения: 03.03.2020).
12. **Котов К. Ю., Мальцев А. С., Соболев М. А. и др.** Алгоритмы и архитектура системы управления траекторным движением мультироторного летательного аппарата // Автометрия. 2020. **53**, № 3. С. 20–28. DOI: 10.15372/AUT20200303.
13. **Золотухин Ю. Н., Котов К. Ю., Мальцев А. С. и др.** Коррекция транспортного запаздывания в системе управления мобильным роботом // Автометрия. 2011. № 2. С. 46–57.

Поступила в редакцию 03.03.2020

После доработки 19.06.2020

Принята к публикации 29.06.2020