

УДК 004.031.42

Вітюк А.Є.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Корнага Я.І.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Барабаш А.О.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ЗАХОПЛЕННЯ НЕВІДОМИХ ОБ'ЄКТІВ МОБІЛЬНИМ РОБОТОМ ІЗ ВИКОРИСТАННЯМ ВІЗУАЛЬНОЇ ІНФОРМАЦІЇ

У статті охарактеризовано проблему захоплення незнайомих об'єктів мобільним роботом, оснащеним однією камерою без додаткових датчиків глибини. Визначено основні властивості такої робототехнічної системи. Розглянуто метод отримання щільної тривимірної моделі об'єкта з карти середовища на основі підходу одночасної локалізації і картографування, що базується на передбаченні карти глибини у режимі реального часу з використанням згорткової нейронної мережі. Отримана тривимірна модель застосована в методі планування стійкого захоплення об'єкту роботизовано рукою, що реалізований на основі аналізу властивостей симетрії серединної вісі моделі об'єкта.

Ключові слова: захоплення об'єктів, мобільний робот, реконструкція середовища, одночасна локалізація і картографування.

1. Вступна частина

Постановка проблеми. У роботі розглядається проблема захоплення невідомих об'єктів мобільним роботом, що орієнтується в навколишньому середовищі та може переміщуватись у ньому. Об'єкти для захоплення є невідомими та сприймаються вперше через систему обробки візуальної інформації. Сучасні роботи можуть бути запрограмовані для виконання багатьох складних завдань по маніпулюванню, починаючи від використання інструментів для збирання складних машин і до проведення високоточних медичних операцій [1]. Але автономне захоплення передчасно невідомого об'єкта з використанням мінімального набору датчиків, що отримують інформацію про стан навколишнього середовища, все ще залишається складною проблемою. При спробі захоплення об'єкту, характеристики якого є завчасно відомими, або якщо відома повна тривимірна модель об'єкта, то існують алгоритми з різними підходами проведення процедури захоплення. Але на практиці отримання повної та точної реконструкції об'єкта, що є невідомим

для робота, є окремою складною задачею. Для пасивних стереосистем тривимірна реконструкція об'єктів з малотекстурованою поверхнею є нерозв'язною задачею. У варіанті, коли стереопсис дає задовільний результат, реконструюються зазвичай лише видимі ділянки об'єкта.

Автоматичне захоплення невідомого об'єкту за допомогою одного зображення є складною проблемою, тому що положення та форма об'єкта є невідомими та можливі конфігурації захоплюючого пристрою, для здійснення жорсткого захоплення, можуть варіюватися. При цьому для рухомого робота, що переміщується в просторі, постановка задачі розрахунку точок захоплення має відмінні умови. На вхід такої системи подається не лише одне зображення сцени з цільовим об'єктом, а послідовність кадрів (відео), за допомогою якого можливо отримати зображення об'єкту з різних ракурсів та розрахувати щільну тривимірну реконструкцію.

Потрібно розглянути систему захоплення невідомого об'єкту мобільним роботом. Ця система для локалізації в навколишньому середовищі та

взаємодії з об'єктом використовує тривимірну реконструкцію отриманих з камери зображень. Система складається з монокамери, роботизованої руки з 7 степенями свободи та захоплюючої кінцівки. На основі відеопотоку з камери відбувається побудова карти навколишнього середовища та реконструкція моделі об'єкта. У процесі сканування, система отримує важливу інформацію щодо структури об'єкта, форми, розміру та орієнтації у просторі. Така обробка та побудова моделі має відбуватися в режимі реального часу.

Аналіз останніх джерел і публікацій. В останні десятиліття проблема автоматичного захоплення невідомих об'єктів стала відігравати все більшу роль в машинному баченні. Більшість розглянутих алгоритмів використовують певний вид навчання для здійснення стійкого захоплення.

Оскільки захоплення передбачає контакт та сили пальців на поверхню об'єкта, то важливо мати достовірну інформацію як про руку, так і про поверхню цільового об'єкта. Необхідно враховувати обмеження руки робота та конфігурації об'єкта, а також обмеження поставленого завдання [2]. Незважаючи на наявність підходу зі здійснення захоплення лише по двовимірним зображенням [3; 4], більшість технік полягають в обробці тривимірних даних. З огляду на складність поставленої задачі багато робіт були спрямовані на спрощення тривимірної форми, зокрема плоских представленнях [5] та компланарності, що комбінується з інформацією про колір [6]. Інші підходи передбачають наявність точної моделі об'єкта або такої, що складається з високорівневих примітивів. Прикладом є використання принципу синергії руки [7; 8; 9]. Також існує підхід із застосуванням машинного навчання, зокрема алгоритму Support Vector Machines (SVM) та використанням примітивів форм високого рівня [10]. Алгоритм з навчанням на основі якості двовимірних захоплень, що базується на нейронних мережах та генетичному алгоритму, представлений у роботі "Real-Time Visual Grasp Synthesis Using Genetic Algorithms and Neural Networks" [11].

Задача побудови реконструкції навколишнього середовища роботом є актуальною та активно досліджується. Оскільки камери глибини не є настільки поширеними, як кольорові камери, багато робіт пов'язані із щільними та напівщільними методами одночасної локалізації та картографування (simultaneous localization and mapping – SLAM) по зображенням з однієї камери [12; 13; 14].

Постановка завдання. Одним з основних обмежень монокулярних підходів до одночасної лока-

лізації та картографування є оцінка абсолютного масштабу. Коли оцінка позиції камери та реконструкція сцени відбувається досить точно, абсолютний масштаб такої реконструкції залишається неоднозначним, обмежуючи використання таких методів. Деякі підходи дозволяють вирішити проблему за допомогою детектування об'єктів шляхом зіставлення сцени із задалегідь визначеним набором тривимірних моделей таким чином, щоб відновити вихідний масштаб на основі розрахованого розміру об'єкта [15], який, тим не менше, не має сенсу за відсутності відомих об'єктів у сцені.

У даній роботі ми поєднуємо метод одночасної локалізації та картографування, що дозволяє позбавитись вказаних проблем та отримати реконструкцію сцени, в якій знаходиться мобільний робот, із методом розрахунку параметрів об'єкта на цій сцені для здійснення стійкого захоплення роботизованою рукою.

2. Отримання моделі об'єкта та розрахунок параметрів захоплення

Кінематика робота. Кінематичне з'єднання [16] складається з набору твердих тіл (що називаються ланками), приєднаних один до одного через механізми (що називаються зв'язками), які обмежують їх рух. Зв'язки мають від 1 до 6 степенів свободи, що визначають обмеження руху приєднаних ланок. Зв'язок, який приєднує ланку А до ланки В та має конфігурацію q_i відображає трансформацію: $T_B^A(q_i) \in SE(3)$, і якщо q_i змінюється, то так само змінюється і трансформація між ланками А та В.

Перетворення будь-якої ланки L_i щодо стійкої системи відліку W може бути розраховано шляхом проходження кінематичним деревом та обчислення загальних перетворень, що ведуть від кожного з'єднання до кореня дерева (цей процес називається прямою задачею кінематики):

$$T_L^W = T_{L_{i-1}}^{L_i} \dots T_{L_2}^{L_1} T_{L_1}^W$$

Конфігурацією робота $q \in \mathbb{R}^N$ є вектор, що поєднує степені свободи всіх зв'язків:

$$q = [q_1 \dots q_N]^T$$

Часткова похідна i -тої системи відліку ланки по q має назву кінематичного Якобіана ланки, та для простих кінематичних ланцюгів може бути ефективно знайдено аналітичне рішення:

$$J_i(q) = \frac{\partial}{\partial q} T_L^W$$

Датчики глибини зображення. Для відображення структури навколишнього середовища за допомогою системи комп'ютерного зору робот повинен бути оснащений датчиками, що дозволять здійснити тривимірну реконструкцію серед-

овища, а не лише двовимірне положення оточуючих об'єктів. Для цього може бути використана стерео камера або глибинна камера. Позначимо глибинне зображення I_D , як функція з областю визначення $\odot \in \mathbb{R}^2$. Відношення між тривимірними точками сцени та двовимірними точками на зображенні може бути змодельовано за допомогою простої пінхольної моделі камери:

$$Proj(x, y, z) = [u, v] = \left[\frac{f_x x}{z} + c_x, \frac{f_y y}{z} + c_y \right]$$

де u, v – де двомірні координати на зображенні; x, y, z – тримірні координати точок в системі координат камери та f_x, f_y, c_x, c_y – внутрішні параметри камери (фокусна відстань і центр камери – f_{max} відповідно). Визначимо також обернену проєкційну модель, яка приймає на вхід координати з камери u, v та вимір глибини точки z , та перетворює їх в тримірний вектор відносно фокусної точки камери. Отримана множина тримірних точок називається хмарою точок за глибинним зображенням. Для конкретного пікселя u, v з глибиною z відповідна точка з хмари точок визначається наступним чином:

$$Proj^{-1}(u, v, z) = z \left[\frac{u - c_x}{f_x}, \frac{v - c_y}{f_y}, 1 \right]$$

Реконструкція середовища мобільним роботом. Для вирішення задачі навігації мобільного робота використовується одночасна локалізація і картографування (ОЛК). Вона полягає в побудові і оновленні мапи невідомого оточуючого середовища з одночасним відстежуванням місцезнаходження під час руху у цьому середовищі.

При заданій послідовності зображень o_t з камери робота в дискретні проміжки часу t ,

задачею ОЛК є розрахунок і визначення положення робота x_t і мапи оточуючого середовища $m_t: P(m_t, x_t | o_{1:t})$.

У даній роботі застосовується реалізація ОЛК методом CNN-SLAM [17]. Перевагою цього алгоритму є використання згорткових нейронних мереж для реконструкції по зображенням з монокамери. Точна та щільна тривимірна реконструкція досягається використанням прогнозованих карт глибини з глибинної нейронної мережі. Такі прогнозовані щільні карти глибини об'єднані зі значеннями глибини, отриманими з прямого монокулярного ОЛК. Схема методу зображена на рис. 1.

Алгоритм надає перевагу передбаченій глибині на частинах зображення, де підхід монокулярного ОЛК, як правило, не працює. Наприклад, вздовж мало текстурованих областей. Крім того, надається оцінка абсолютного масштабу отриманої реконструкції.

Оцінка положення камери відбувається шляхом визначення ключових кадрів $k_i, \supset, k_n \in \mathcal{K}$ як структурних елементів, на яких здійснюється ОЛК. Кожен ключовий кадр k_i пов'язаний з позицією камери T_{k_i} , щільною картою глибини D_{k_i} та картою невизначеності глибини U_{k_i} (ступінь довіри кожному значенню глибини). Поточна позиція камери являє собою трансформацію t -го ключового кадру відносно $t-1$ ключового кадру і складається з поворотної матриці 3×3 $R_t \in \mathbb{SO}(3)$ та тримірного вектора переміщення $t_t \in \mathbb{R}^3$:

$$T_t^{k_i} = [R_t, t_t] \in \mathbb{SE}(3)$$

Ця трансформація оцінюється шляхом мінімізації фотометричної похибки між інтенсивністю

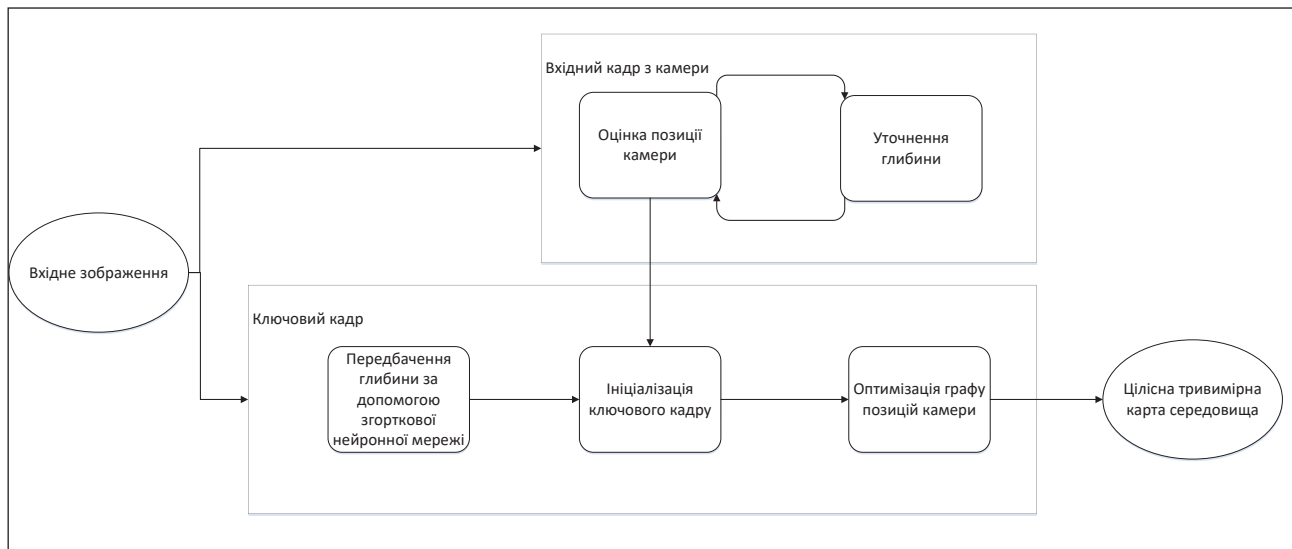


Рис. 1. Алгоритм CNN-SLAM

зображення I_t поточного кадру та інтенсивністю зображення I_{k_i} найближчого ключового кадру k_i , застосовуючи зважену оптимізацію Гауса-Ньютона, що базується на функції:

$$E(T_i^{k_i}) = \sum_{\tilde{u} \in \mathbb{C}} \rho \left(\frac{r(\tilde{u}, T_i^{k_i})}{\sigma(r(\tilde{u}, T_i^{k_i}))} \right)$$

Тут ρ – це норма Хьюбера та σ – це функціональний вимір залишкової невизначеності [17]. Фотометричний залишок r визначається на обмеженій підмножині пікселів, що лежать в областях великого градієнту кольору ($\tilde{u} \subset u \in \mathbb{C}$) та визначається:

$$r(\tilde{u}, T_i^{k_i}) = I_{k_i}(\tilde{u}) - I_t(\text{Proj}(KT_i^{k_i}K^{-1}\tilde{u}D_{k_i}(\tilde{u})))$$

Маючи значення $T_i^{k_i}$, поточну позицію камери у світовій системі координат можна обчислити як $T_t = T_i^{k_i}T_{k_i}$. Далі граф позицій камери на ключових кадрах уточнюються за допомогою методів оптимізації графів [18].

Детектування параметризованих поверхонь у тривимірній карті середовища. Після отримання тривимірної карти середовища у вигляді щільної хмари точок, важливою задачею є розпізнавання твердотілих об'єктів, що може бути здійснена на основі сегментації певної сцени середовища. Цей процес полягає в розподілі хмари точок на підмножини, що відповідають чітким параметризованим моделям, таким як: лінії, площини, кола, циліндричні поверхні, сферичні поверхні, конічні поверхні.

Сегментація хмари точок здійснюється на основі даних про нормалі точок з хмари. Карта, отримана на етапі реконструкції містить лише сукупність точок p_i , для кожної з яких оцінка нормалі представляє собою аналіз власних векторів та власних значень коваріаційної матриці C , що враховує найближчих сусідів заданої точки:

$$C = \frac{1}{k} \sum_{i=1}^k (p_i - \bar{p})(p_i - \bar{p})^T, \quad C * \bar{v}_j = \lambda_j * \bar{v}_j, \quad j \in \{0, 1, 2\}$$

Тут k кількість сусідніх точок, що розміщені у деякому околі точки p ; \bar{p} – центроїд найближчих сусідів; λ_j є j -им власним значенням коваріаційної матриці, а \bar{v}_j – власним вектором. Отриманий вектор нормалі визначається:

$$\sigma = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}$$

Далі, на множині точок p_i та їх нормалей σ_i здійснюється сегментація з використанням методу RANSAC. Алгоритм знаходить різні параметричні моделі та представляє їх у формі класифікованої хмари точок, де всі точки відсегментовані у екземпляри моделей об'єктів, яким вони належать.

Обчислення позиції захоплення об'єкта. На основі цілісної щільної моделі об'єкта здійснюється планування захоплення. Для цього використано метод, що оцінює локальні властивості симетрії об'єкта [19]. Це може бути досягнуто шляхом обчислення серединної вісі, яка представляє тривимірний об'єкт як об'єднання куль. Дані симетрії, що містить серединна вісь, підлягають аналізу. Схему алгоритму представленого методу передбачення захоплення зображено на рис. 2.

Отримавши модель об'єкта довільної форми у вигляді хмари тривимірних точок, обчислюється серединна вісь за алгоритмом, представленим у [20]. Отриманий набір точок, що наближено утворюють серединну вісь, позначимо як $\frac{42122}{(-11-2)}$.

Дані серединної вісі на проекційній площині $\frac{28}{(-11-2)}$ представляємо як зважений повний граф та сегментуємо його на кластери c_i . Кожен утворений кластер c підлягає аналізу та відповідності певному типу серед наступних типів структур: коло, зірка з кільцем, дерево, елемент осі симетрії.

Генерування кандидатів на захоплення здійснюється з використанням набору евристик, що враховують отримані структури у частинах серединної вісі. Основний підхід полягає в розгляді кількох сусідніх площин C та отриманні перспективних цільових напрямків та кутів повороту руки для кожного типу кластерів.

3. Висновки

Показано, що інтеграція ОЛК системи з передбаченням карти глибини за допомогою глибокої нейронної мережі в систему розрахунку позиції руки робота для здійснення захоплення є багато-

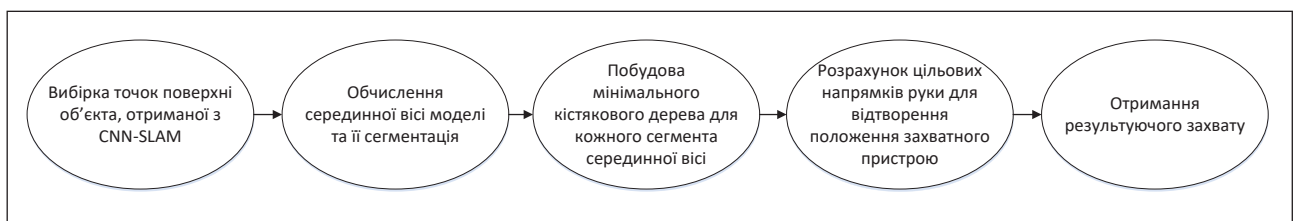


Рис. 2. Планування захоплення на основі серединної вісі моделі об'єкта

обіцяючим напрямком для вирішення проблем розрахунку положень захоплюючого пристрою. Зокрема, вона дозволяє зменшити залежність від типових обмежень традиційної реконструкції по зображенням з монокамери, особливо щодо оцінки абсолютного масштабу, отримання щільної карти глибини на малотекстурованих областях і на кадрах, що мають лише трансформацію повороту. Запропонований підхід до використання

уточнення карти глибини під час реконструкції навколишнього середовища долає ці проблеми, зберігаючи при цьому надійність і точність прямого монокулярного ОЛК. Загальна система здатна виконувати розрахунок положення захоплюючого пристрою на основі загальної реконструйованої сцени навколишнього середовища, її сегментації та виокремлення моделі цільового об'єкта.

Список літератури:

1. Shademan, A., Decker, R., Opfermann, J. D., Leonard, S., Krieger, A., & Kim, P.C. "Supervised autonomous robotic soft tissue surgery", *Science Translational Medicine*, 8 (337), 2016
2. Sancho-Bru JL, Mora MC, Leon BE, Perez-Gonzalez A, Iserte JL, Morales A., "Grasp modeling with a biomechanical model of the hand", *Comput Methods Biomech Biomed Eng* 17(4):297-310, 2014.
3. Han, Yuexing, Bing Wang, Hideki Koike and Masanori Idesawa. "Object Recognition with a Limited Database Using Shape Space Theory." *Image Processing: Concepts, Methodologies, Tools, and Applications*, 2013, pp.181-200.
4. Michael Laskey, Jeffrey Mahler, Zoe McCarthy, Florian T. Pokorny, Sachin Patil, "Multi-Armed Bandit Models for 2D Grasp Planning with Uncertainty.", *IEEE Conference on Automation Science and Engineering*, 2015.
5. Ala R.K., Dong H.K., Shin S.Y., "A 3D-grasp synthesis algorithm to grasp unknown objects based on graspable boundary and convex segments", *Inf. Sci.* 295, 2015, pp.91-106.
6. Popovic M., Kraft D., Bodenhagen L., Baeski E., Pugeault N., Kragic D., Asfour T., Kruger N., "A strategy for grasping unknown object based on co-planarity and colour information", *Robot. Auton. Syst.* 58(5), 2010, pp.551-565.
7. M. Bonilla, D. Resasco, M. Gabiccini, and A. Bicchi, "Grasp Planning with Soft Hands using Bounding Box Object Decomposition", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Hamburg, Germany, pp. 498-503 (2015).
8. M. Bonilla, E. Farnioli, L. Pallottino, A. Bicchi, "Sample-Based Motion Planning for Robot Manipulators with Closed Kinematic Chains", *IEEE ICRA*, Seattle, Washington, USA (2015).
9. Tamim Asfour, Júlia Borrás, Christian Mandery, Peter Kaiser, Eren Erdal Aksoy, and Markus Grotz, "On the dualities between grasping and whole-body loco-manipulation tasks", *International Symposium on Robotics Research (ISRR)* (2016).
10. R. Pelossof, A. Miller, P. Allen, and T. Jebara, "An SVM Learning Approach to Robotic Grasping", in *IEEE International Conference on Robotics and Automation*, 2004, pp. 3512–3518.
11. Huang, Bidan, Sahar El-Khoury, Miao Li, Joanna J. Bryson, and Aude Billard, "Learning a real time grasping strategy" In *IEEE international conference on robotics and automation (ICRA)*, Karlsruhe, 2013, pp.593-600.
12. R. A. Newcombe, S. Lovegrove, and A. J. Davison. Dtm: Dense tracking and mapping in real-time. In *IEEE ICCV*, pages 2320– 2327, 2011
13. J. Engel, T. Schps, and D. Cremers. LSD-SLAM: LargeScale Direct Monocular SLAM. In *European Conference on Computer Vision (ECCV)*, 2014.
14. R. Mur-Artal, J. M. M. Montiel, and J. D. Tards. Orb-slam: A versatile and accurate monocular slam system. *IEEE Trans. Robotics*, 31(5):1147–1163, 2015.
15. D. Galvez-L'opez, M. Salas, J. D. Tard'os, and J. Montiel. Real-time monocular object slam. *Robot. Auton. Syst.*, 75(PB), jan 2016.
16. M. Mason. *Mechanics of Robotic Manipulation*. MIT Press, Cambridge, MA, 2016.
17. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction. K. Tateno, F. Tombari, I. Laina, N. Navab, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
18. R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A General Framework for Graph Optimization. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
19. M. Przybylski, T. Asfour, and R. Dillmann, "Unions of balls for shape approximation in robot grasping," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 1592–1599.
20. J. E. Goodman and J. O'Rourke, "Handbook of Discrete and Computational Geometry, Second Edition", CRC Press LLC, Boca Raton, FL, 2004.

ЗАХВАТ НЕИЗВЕСТНЫХ ОБЪЕКТОВ МОБИЛЬНЫМ РОБОТОМ С ИСПОЛЬЗОВАНИЕМ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ

В статье охарактеризовано проблему захвата незнакомых объектов мобильным роботом, оснащенным одной камерой без дополнительных датчиков глубины. Определены основные свойства такой робототехнической системы. Рассмотрен метод получения плотной трехмерной модели объекта с карты среды на основе подхода одновременной локализации и картографирования, основанный на предвидении карты глубины в режиме реального времени с использованием сверточной нейронной сети. Полученная трехмерная модель применена в методе планирования устойчивого захвата объекта роботизированной рукой, который реализован на основе анализа свойств симметрии срединной оси модели объекта.

Ключевые слова: захват объектов, мобильный робот, реконструкция среды, одновременная локализация и построение карты.

GRASPING OF NOVEL OBJECTS BY MOBILE ROBOT USING VISUAL PROCESSING

The article describes the problem of capturing unseen before objects by a mobile robot, equipped with one camera without additional depth sensors. The basic properties of such robot system are determined. The method of obtaining a dense three-dimensional model of an object from an environmental map based on the simultaneous localization and mapping approach based on the prediction of a real-time depth map using a convolutional neural network is considered. The obtained three-dimensional model is applied in the method of planning of stable grasp of an object rotated by the robotic hand, which uses analysis of the symmetry properties of the median axis of the object model.

Key words: object grasping, mobile robot, environment reconstruction, simultaneous localization and mapping.