

УДК 004.891.032.26:629.7.014-519.066

doi: 10.32620/reks.2019.2.10

В. В. МОСКАЛЕНКО, М. О. ЗАРЕЦЬКИЙ, А. С. МОСКАЛЕНКО

Сумський державний університет, Україна

МОДЕЛЬ ТА МЕТОД НАВЧАННЯ СИСТЕМИ КЛАСИФІКАЦІЇ РУХОМИХ ОБ'ЄКТІВ ДЛЯ МАЛОГАБАРИТНОГО БЕЗПІЛОТНОГО АПАРАТУ

У роботі розроблено модель класифікаційного аналізу, яка складається з детектора руху, трекера об'єктів, згорткового розріджено кодуючого екстрактора ознак і стекованого інформаційно-екстремального класифікатора. Пропонується побудова детектора руху на основі визначення різниці двох послідовних вирівняних кадрів, де вирівнювання здійснюється шляхом співставлення ключових точок, оцінки гомографії та проєктивної трансформації. Використання детектора руху спрямоване на спрощення задачі класифікаційного аналізу об'єктів, оскільки дозволяє знизити варіативність вхідних образів і зекономити ресурси необхідні для синтезу без навчання моделі пошуку області руху. Запропонована модель характеризується не високою обчислювальною складністю і може бути використана як інструмент збору розмічених даних для глибокої моделі детектора рухомих об'єктів. Крім того, розроблено метод навчання детектора рухомих об'єктів. Метод навчання моделі детектора включає етап навчання екстрактора ознак без вчителя на основі розріджено кодуючого нейронного газу, етап навчання з учителем і тонкої настройки стекованого інформаційно-екстремального класифікатора. Також пропонується автоматизована розмітка нових вхідних даних шляхом трекінгу об'єктів з відомою міткою, яка формується на основі закріплення прогнозу як мітки у випадку його високої достовірності або на основі ручної розмітки для зразків з малою достовірністю розпізнавання. При цьому балансування класів здійснюється на основі техніки андерсамплінгу в рамках стратегії «один проти всіх». Результати моделювання на відкритих наборах даних підтверджують придатність запропонованих моделі та методу навчання для практичного використання. Ключові слова: класифікація, відстеження руху, детектування об'єктів, згорткова нейронна мережа, розріджено кодуючий нейронний газ, інформаційно-екстремальне навчання, активне навчання.

Ключові слова: класифікація; відстеження руху; детектування об'єктів; згорткова нейронна мережа; розріджено кодуючий нейронний газ; інформаційно-екстремальне навчання; активне навчання

Вступ

Малогабаритні безпілотні апарати часто використовуються у задачах моніторингу території та об'єктів. При цьому безпілотний апарат часто розглядається як мобільний сенсор для збору даних для їх подальшої обробки в хмарному сервісі [1]. Однак існує тенденція до розвитку крайових обчислень з метою розвантаження комунікаційних каналів інфокомунікаційного середовища і захисту інформації [2]. Тому багато сучасних розробок пов'язані з розширенням функціональних можливостей та підвищенням рівня автономності безпілотних апаратів.

Однією з затребуваних функціональних можливостей безпілотних апаратів є пошук і класифікація рухомих об'єктів. Рухомі об'єкти часто представляють інтерес, оскільки стан об'єктів і причини їх руху можуть бути малопередбачуваними, що обумовлює потребу в контролі. Детектування рухомих об'єктів і їх класифікаційний аналіз можуть бути корисними в таких практичних задачах, як відеоаналітика систем рятувально-пошукових заходів,

охорони периметру, прицілювання і наведення зброї.

У найпростішому випадку алгоритм детектування руху оснований лише на визначенні різниці двох послідовних кадрів [3]. Тому в сучасних бібліотеках машинного зору доступні лише реалізації детектор руху для стаціонарних камер. Однак у випадку нестационарної камери перед порівнянням кадрів необхідно здійснювати їх вирівнювання один відносно одного шляхом пошуку матриці афінного або перспективного перетворення. У працях пропонується [4, 5] реалізація вирівнювання на основі обчислення оптичного потоку або пошуку ключових точок, що описуються SIFT-дескрипторами. Проте ці алгоритми є ресурсомісткими і тому є придатними лише для обробки відеопотоку низької роздільної здатності. У праці [6] були запропоновані можливі підходи до підвищення оперативності вирівнювання кадрів.

Використання детектора руху значно спрощує задачу класифікаційного аналізу детектованих об'єктів, оскільки це дозволяє знизити варіативність

вхідних образів. Крім того пошук області руху не потребує машинного навчання і є менш ресурсозатратним порівняно зі скануванням зображення на основі штучних нейронних мереж.

Класифікаційний аналіз детектованих об'єктів дозволяє автоматично визначати категорію, до якої вони відносяться. При цьому модель класифікаційного аналізу візуальних образів прийнято поділяти на екстрактор ознак та вирішальні правила. Серед екстракторів ознак набули поширення гістограми орієнтованих градієнтів, гістограми візуальних слів та згорткові нейронні мережі [7]. Однак спосіб подання спостережень з використанням низькорівневих та середньорівневих ознак характеризується невисокими інформаційною ємністю та завадозахищеністю. Згорткові нейронні мережі є більш гнучкими у використанні, і мають високу інформаційну ємність за рахунок використання ієрархічного ознакового подання. Проте навчання згорткових мереж, як правило, відбувається на основі алгоритму зворотного поширення помилки, що потребує великого обсягу розмічених навчальних даних і ресурсів для успішного навчання.

Оперативність машинного навчання визначає швидкість розширення функціональних можливостей під час адаптації до нових умов функціонування. Незважаючи на досягнення в галузі апаратних прискорювачів нейронних обчислень для мобільних пристроїв, досі процес машинного навчання глибоких мереж не може бути реалізований в реальному темпі часу. Крім того, існує дефіцит ефективних алгоритмів навчання за умов обмеженого обсягу розмічених даних [8]. При цьому процес розмітки навчальних даних є дорогим і рутинним. Тому розробка і дослідження нових підходів до зменшення трудомісткості розмітки навчальних даних та процесу машинного навчання є затребуваними в області крайових обчислень.

У працях [9, 10] було запропоновано алгоритм інкрементального пошарового навчання згорткової мережі без вчителя на основі модифікації нейронного газу і тонкої настройки на основі алгоритму симуляції відпалу. Отримані результати показали, що даний підхід дозволяє підвищити оперативність навчання. У працях [11, 12] було розглянуто використання інформаційно-екстремального класифікатора для аналізу високорівневого ознакового подання. Цей класифікатор оснований на двійковому кодуванні ознакового подання на основі дерев рішень і побудові радіально-базисних вирішальних правил в двійковому просторі Хеммінга. Тобто класифікатор оснований на обчислювально простих операціях порівняння та розрахунку відстані Хеммінга між двійковими кодами. Такий підхід є перспективним з точки зору завадозахищеності і оперативності фун-

кціонування. Однак досі не було розглянуто використання методики бустінгу та стекінгу для підвищення ефективності апроксимації складних нелінійних залежностей [13].

У статті пропонуються модель і метод навчання детектора рухомих об'єктів для підвищення оперативності навчання, зменшення трудомісткості розмітки даних та вимог до обсягу навчальних вибірок. В основу запропонованого рішення планується покласти ідеї і методи детектування руху на зображенні нестационарної камери, навчання без вчителя згорткової мережі, інформаційно-екстремальне навчання класифікаційних вирішальних правил, активне навчання та самонавчання.

Постановка задачі

Нехай $D_{\text{train}} = \{ I_k, B_k \mid k = \overline{1, K_1} \}$ та $D_{\text{test}} = \{ I_k, B_k \mid k = \overline{1, K_2} \}$ є набором навчальних та тестових даних, які неперервно поповнюються, де I_k – k -та пара послідовних зображень з кольоровою схемою RGB з роздільною здатністю 512×512 , B_k – набір правильно заданих обмежувальних рамок, які відповідають об'єктам інтересу на k -му зображенні, K_1, K_2 – обсяги навчального та тестового наборів даних відповідно, які неперервно можуть зростати. Нехай алфавіт класів розпізнавання $\{ X_z^o \mid z = \overline{1, Z} \}$ характеризує рухомі об'єкти інтересу. В даному випадку початкова кількість правильно заданих обмежувальних рамок об'єктів інтересу на клас розпізнавання становить не більше 200.

Дано структурований вектор просторово-часових параметрів функціонування системи детектування об'єктів інтересу, який у загальному випадку має структуру

$$g = \langle e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}, f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle, \quad (1)$$

$$\Xi_1 + \Xi_2 = \Xi,$$

де $e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}$ – генотипні параметри функціонування, які впливають на параметри алгоритмів екстракції ознак розпізнавання;

$f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2}$ – фенотипні параметри функціонування, які впливають на геометрію контейнерів класів розпізнавання.

При цьому відомі обмеження на параметри функціонування:

$$R_{\xi_1}(e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}) \leq 0, \quad (2)$$

$$R_{\xi_2}(f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2}) \leq 0.$$

Необхідно в процесі машинного навчання визначити оптимальні значення координат вектора (1), які забезпечують максимум усередненого за алфаві-

том класів розпізнавання комплексного критерію

$$g^* = \arg \max_G \{J(g)\}, \quad (3)$$

де J – критерій функціональної ефективності класифікаційного аналізу.

Модель та метод навчання детектора

У загальному випадку система детектування рухомих об'єктів, окрім режиму онлайн навчання, може мати режим офлайн навчання, під час якого відбувається узагальнення накопиченого досвіду. При цьому режим онлайн навчання та режим розпізнавання можуть слугувати механізмами збору розмічених навчальних даних для офлайн навчання глибокої моделі детектування об'єктів (рис. 1).

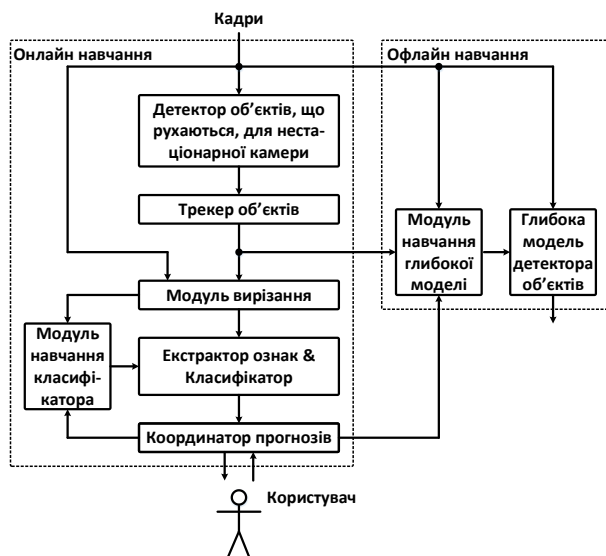


Рис. 1. Узагальнена структурна схема здатного навчатися детектора рухомих об'єктів

Як показано на рис. 1, важливим компонентом системи, що пропонується є детектор руху. Пропонується побудова детектора руху на основі визначення різниці двох послідовних вирівняних кадрів. При цьому вирівнювання пропонується здійснювати на основі співставлення ключових точок, оцінки гомографії та проєктивної трансформації (рис. 2). Для підвищення оперативності обчислення дескрипторів ключових точок при збереженні їх інваріантності до повороту і масштабу можна використати алгоритм SURF замість SIFT [14].

Різницеве зображення, утворене після віднімання вирівняних кадрів, містить шум. Фільтрацію найпростіше виконати на основі швидких морфологічних операцій, таких як ерозія (Erode) та дилатація (Dilate) [15]. Ядра морфологічних фільтрів обираються з огляду на мінімальний розмір об'єктів інтересу на зображенні.

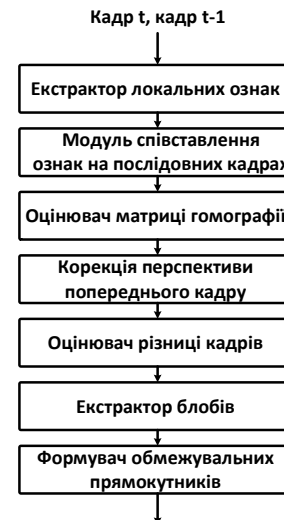


Рис. 2. Функціональна схема детектора руху

Трекер детектованого об'єкта дозволяє продовжити збір навчальних даних навіть при зупинці руху об'єкта інтересу відносно камери. Як показано в праці [16] найбільш ефективними з точки зору стабільності і обчислювальної складності є алгоритм Predator, ще відомий як TLD (Tracking-Learning-Detection). Вихід детектора може бути поданий для просторового пулінгу в глибоку мережу детектування. Тоді детектор руху і трекер виконуватимуть функцію мережі RPN (Region proposal network). Однак для невеликого алфавіту класів і невеликої кількості об'єктів інтересу в області спостереження цілком достатньо виконати вирізання і уніфікацію розміру зображень області інтересу для їх подальшого класифікаційного аналізу. В режимі онлайн навчання можна усунути з розгляду кадри, що містять забагато об'єктів інтересу для підвищення оперативності функціонування.

Важливим компонентом запропонованого детектора є екстрактор ознак, який пропонується реалізувати у вигляді тришарової згорткової нейронної мережі з субдискретизуючим шаром типу max-pooling між ними. Вхідна роздільна здатність згорткової мережі обирається рівною 64x64. Ядра фільтрів встановлюються рівними 3x3. Перший шар містить 64 фільтри, а другий і третій – по 32 фільтри.

Зони інтересу кодується згортковою нейронною мережею і надходять на вхід класифікатора. Інформаційно-екстремальний класифікатор здійснює двійкове кодування вхідних спостережень, використовуючи дерева рішень. Прийняття рішень відбувається на основі радіально-базисних вирішальних правил, що розбивають двійковий простір на області класів (рис. 3).

На рис. 3 показано, що кодування спостережень відбувається ансамблем дерев рішень. Вузли дерев рішень пронумеровані. Номери ненульових

бітів результуючого двійкового коду відповідають номерам вузлів, через які проходить шлях прийняття рішень в деревах рішень.

Інформаційно-екстремальний класифікатор в режимі екзамену приймає рішення щодо належності вхідних даних x з відповідним бінарним представленням b до одного з класів повороту з множини $\{X_z^0 | z = \overline{1, Z}\}$ у відповідності з максимальним значенням функції приналежності $\mu_z(b)$ через вираз $\arg \max_z \{\mu_z(b)\}$.

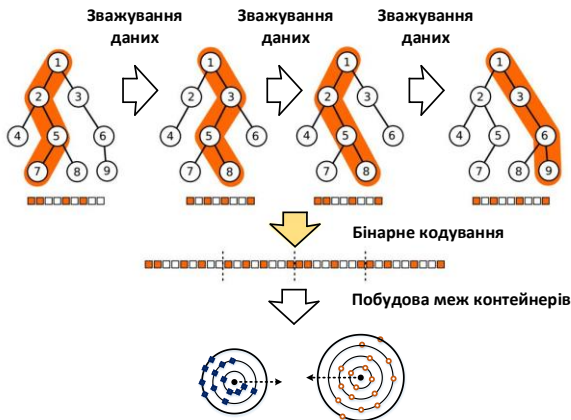


Рис. 3. Архітектура класифікатора

Таким чином, функція належності $\mu_z(b)$, оптимальний контейнер якої має опорний вектор b_z^* та радіус d_z^* , виводяться з формули

$$\mu_z(b) = \exp\left(-\sum_{i=1}^{N_2} b_i \oplus b_{z,i}^* / d_z^*\right). \quad (4)$$

Базова модель інформаційно-екстремального класифікатора, показана на рис. 3, є досить простою і може бути недостатньою для апроксимації складних залежностей. Одним з ефективних шляхів підвищення точності простих моделей є їх стекінг. Стекінг полягає у доповненні вибіркового даних складними ознаками, які є виходами простих моделей, побудованих на даних з попереднім словником ознак (рис. 4). Остання модель використовує як початковий словник ознак, так і утворені нові ознаки.

Навчання моделі детектора пропонується здійснювати в декілька етапів.

- попереднє навчання без вчителя згорткового екстрактора ознак на великому наборі даних ImageNet64;

- початкове навчання класифікатора за малою вибіркою об'єктів інтересу;

- донавчання класифікатора з використанням техніки самонавчання та активного навчання;

- навчання глибокої моделі детектора на зібраних розмічених навчальних зразках.

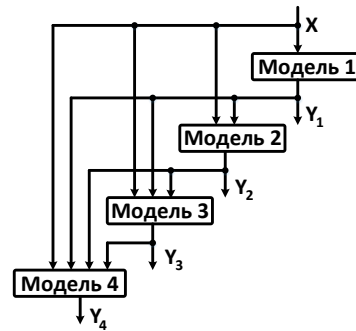


Рис. 4. Структурна схема складної моделі, побудованої на основі стекінгу простих моделей

Для реалізації першого етапу пропонується використати модифікацію нейронного газу, описану в працях [9]. Цей алгоритм збігається близько до глобального оптимуму завдяки м'якій конкуруючій схемі навчання. При цьому завдяки використанню правила Ойя глобальний оптимум відповідає такому словнику фільтрів, що дозволяє найкраще реконструювати вхідні дані [12].

Для формування навчальної вибірки згорткового шару кожен патч зображення (або карти ознак) перетворюється до одновимірного вигляду. Навчання відбувається пошарово. Вхідними даними для модифікованого нейронного газу є потужність словника базисних векторів M , розмірність простору ознак N , $\lambda_0, \lambda_{\text{final}}$ – початкове та кінцеве значення коефіцієнта розміру околу сусідів, $\eta_0, \eta_{\text{final}}$ – початкове та кінцеве значення коефіцієнта швидкості навчання. Розглянемо основні кроки алгоритму.

- 1) ініціалізація словника базисних векторів (фільтрів) $D = (d_1, \dots, d_M)$ випадковими числами з рівномірного розподілу;

- 2) ініціалізація лічильника навчальних векторів $t := 1$;

- 3) вибір випадкового вектора x з множини навчальних векторів X ;

- 4) нормалізація векторів із словника $D = (d_1, \dots, d_M)$ шляхом приведення до одиничної довжини (L2-нормалізація);

- 5) обчислення поточних значень коефіцієнта розміру околу сусідів λ_t та швидкості навчання η_t :

$$\lambda_t := \lambda_0 (\lambda_{\text{final}} / \lambda_0)^{t/t_{\text{max}}};$$

$$\eta_t := \eta_0 (\eta_{\text{final}} / \eta_0)^{t/t_{\text{max}}};$$

- 6) обчислення міри схожості вхідного вектора x до базисних векторів $d_{1k} \in D$ для їх сортування

$$-(d_{10}^T x)^2 \leq \dots \leq -(d_{1k}^T x)^2 \leq \dots \leq -(d_{1M-1}^T x)^2;$$

- 7) оновлення координат базисних векторів

$d_{1_k} \in D$ за правилом Ойа [9]:

$$d_{1_k} := d_{1_k} + \eta_t \exp(-k / \lambda_t) y(x - y d_{1_k}), \quad y := c_{1_k}^T x,$$

$k = 0, M-1$; якщо $t < t_{\max}$, то інкремент лічильника $t := t+1$ та перехід до кроку 3.

У праці [17] було показано, що використання неповного базису для розрідженого кодування спостережень в задачах класифікаційного аналізу так само ефективне, як і використання надповного базису. Тому для економії ресурсів, кількість фільтрів менша за розмірність патчів. Розріджене кодування пропонується здійснювати з використанням алгоритму ортогонального узгодженого переслідування (Orthogonal Matching Pursuit) з кількістю активних атомів рівній $K = M/3$. Для зниження розмірності і підвищення інваріантності до зміщень пропонується використання шару субдискретизації – max-пулінг. Після шару субдискретизації перед початком наступного згорткового шару знаходиться шар локальної нормалізації контрасту (Local Contrast Normalization). Шар локальної нормалізації підсилює інформативні ознаки і послаблює решту пікселів карти ознак.

Дані цільової області використання будуть використані для тонкої настройки фільтрів згорткової мережі з використанням малих постійних значень $\lambda_t = \lambda_0 = \lambda_{\text{final}} = 10^{-4}$, $\eta_t = \eta_0 = \eta_{\text{final}} = M/3$.

Інформаційно-екстремальний класифікатор, який оцінює приналежність j -го набору даних x_j з N_1 ознаками до одного з класів Z , виконує кодування ознак за допомогою композиції дерев рішень (boosted trees) і вирішальних правил, побудованих в радіальному базисі двійкового простору Хеммінга. У цьому випадку є навчальний набір $D = \{x_j, y_j \mid j = \overline{1, n}\}$ де n розмір датасету, а y_j це мітка j -го набору даних, яка належить до одного класу з набору класів $\{X_z^\circ \mid z = \overline{1, Z}\}$. Навчання інформаційно-екстремального класифікатора з використанням бустінгу виконується наступним чином

1) Ініціалізація вагових коефіцієнтів $w_j = 1/n$.

2) Для $k = 1, \dots, K$ виконуються наступні кроки

3) Генерація D_k набору навчальних даних з усього набору D з використанням функції розподілу ймовірності $P(X = x_j) = w_j$.

4) Навчання дерева рішень T_k на наборі D_k із застосуванням ентропійного критерію для оцінки якості розбиття.

5) Бінарне кодування вектору x_j з набору да-

них D шляхом об'єднання шляхів рішення в деревах T_1, \dots, T_k . Результатом роботи цього етапу є бінарна матриця $\{b_{z,s,i} \mid i = \overline{1, N_2}; s = \overline{1, n_z}; z = \overline{1, Z}\}$, де N_2 кількість індукованих бінарних ознак та n_z кількість реалізацій відповідного класу X_z° , що задовольняє рівність $n = \sum_z n_z$

б) Побудова інформаційно-екстремальних вирішальних правил в радіальному базисі бінарного простору Хемінга та обчислення інформаційного критерію:

$$E_z^* = \max_{\{d\}} E_z(d), \quad (5)$$

де $\{d\} = \left\{0, 1, \dots, \left(\sum_i b_{z,i} \oplus b_{c,i} - 1\right)\right\}$ – набір концентричних радіусів з центром b_z ;

b_z – опорний вектор розподілу даних в класі X_z° , який розраховується за правилом

$$b_{z,i} = \begin{cases} 1, & \text{if } \frac{1}{n_z} \sum_{s=1}^{n_z} b_{z,s,i} > \frac{1}{Z} \sum_{c=1}^Z \frac{1}{n_c} \sum_{s=1}^{n_c} b_{c,s,i}; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

де E_z – критерій ефективності побудови вирішальних правил в режимі навчання для класу X_z° , який розраховується як нормалізована модифікація інформаційної міри Кульбака [14]:

$$E_z = \frac{1 - (\alpha_z + \beta_z)}{\log_2(2 + \zeta) - \log_2 \zeta} \cdot \log_2 \left[\frac{2 - (\alpha_z + \beta_z) + \zeta}{(\alpha_z + \beta_z) + \zeta} \right], \quad (7)$$

де α_z, β_z – частота хибно позитивних та хибно негативних спрацювань вирішальних правил класу X_z° ;

ζ – будь-яке мале додатне число, яке вводиться, для уникнення невизначеності при діленні на нуль.

Задля підвищення ефективності навчання, загальноприйнятим шляхом є зведення проблеми багатокласової класифікації до серії двокласових класифікацій за принципом «один проти всіх». Для уникнення проблеми незбалансованості класів розпізнавання, обумовленої переважанням у навчальному наборі даних негативних зразків, здійснюється введення синтетичного класу, який є альтернативним для X_z° . Синтетичний клас представлено n_z векторами з інших класів, які найбільш близькі до опорного вектора b_z , де n_z – обсяг навчального набору даних класу X_z° .

7) Тестування одержаних інформаційно-екстремальних вирішальних правил на наборі даних

D та розрахунок коефіцієнту помилок для кожної реалізації з D . При цьому, в режимі екзамєну прийняття рішення про належність вектора b до одного з класів розпізнавання з алфавіту $\{X_z^\circ | z = \overline{1, Z}\}$ здійснюється за максимальним значенням функції належності $\mu_b(b)$ відповідно до виразу $\arg \max_z \{\mu_z(b)\}$. У цьому випадку функція належності $\mu_z(b)$ бінарного подання b вхідного вектора даних x до класу X_z° , оптимальний контейнер якого має опорний вектор b_z^* та радіус d_z^* , розраховується за формулою (4).

8) Оновлення вагових коефіцієнтів $\{w_j\}$ пропорційно до отриманих коефіцієнтів помилок для вектора x_j .

9) Якщо $|E_k^* - E_{k-1}^*| < \epsilon$ то вихід з циклу.

Під час роботи системи накопичуються нові розмічені навчальні зразки. Вони з'являються в результаті стратегії активного навчання: початковий зразок з двійковим поданням b надходить на розмітку користувачу, якщо виконується умова $\max_z \mu_z(b) > \mu_{th}$, де μ_{th1} – порогове значення функції належності, значення нижче якого інтерпретуються як невпевнене розпізнавання; z – номер класу об'єкта інтересу. При цьому навчальний зразок з двійковим поданням b з міткою $z^* = \arg \max_z \mu_z(b)$ додається в навчальну вибірку з цією міткою, якщо $\mu_{z^*} > \mu_{th2}$, де μ_{th2} – порогове значення функції належності, перевищення якого є ознакою впевненості розпізнавання. В обох випадках після отримання ручної чи автоматичної мітки, ця мітка зберігатиметься для всіх зразків, що формуватимуться під час трекінгу розпізнаних об'єктів. Зразки з максимальним значенням функції належності з діапазону $[\mu_{th1}, \mu_{th2}]$ ігноруються.

Класифікатор після першого навчання на малій вибірці залишає 80% дерев фіксованими. Решта дерев будуються заново при поповненні навчальної вибірки.

3. Результати фізичного моделювання

Для початкового навчання класифікатора сформовано алфавіт з семи класів: X_1° – велосипед (bicycle); X_2° – автобус (bus); X_3° – легковий авто-

мобіль (car), X_4° – мотоцикл (motorcycle), X_5° – пікап (pickup truck), X_6° – вантажівка (articulated truck), X_7° – фон. Для цього було використано відкриті набори даних MIO-TCD dataset [20]. У тестовій. У тестовій і навчальній вибірці мітиться 100 зразків на клас. Навчальні зразки приводяться до уніфікованої роздільної здатності 64x64 пікселя. Залежність нормованого значення інформаційного критерію (7) за навчальною вибіркою та точність за тестовою вибіркою від кількості дерев для двійкового кодування показано на рис. 5. При цьому глибина дерев рішень була рівна 5, а глибина стекінгу дорівнює 3.

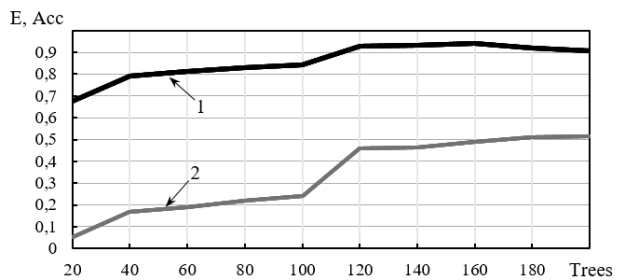


Рис. 5. Графік залежності критерію ефективності від кількості дерев рішень для двійкового кодування: 1 – нормований інформаційний критерій (7), обчислений за навчальною вибіркою; 2 – точність, обчислена за тестовою вибіркою

Аналіз рис. 5 показує, що після побудови 160 дерев точність класифікатора починає зменшуватися внаслідок перенавчання. Тому обирається оптимальна кількість дерев, що забезпечує точність 0.94%. Подальше збільшення точності може здійснюватися за рахунок збільшення кількості інформативних навчальних даних. Збір даних здійснюватиметься в інтерактивному режимі з відеофайлів набору даних VIRAT Video Dataset [21].

На рис. 6 показано залежність нормованого значення інформаційного критерію (7) за навчальною вибіркою та точність за тестовою вибіркою від кількості зразків, які були розмічені користувачем в режимі активного навчання зі значення параметра $\mu_{th1} = 0.1$. Самонавчання при цьому не застосовувалося.

Аналіз рис. 6 показує, що розмітка і додавання зразків, які розпізнаються з низькою впевненістю, до навчальної вибірки дозволяє підвищити точність класифікатора на тестовому наборі даних. При цьому було додано всього 6 зразків, проте вдалося підвищити точність до 97%. Тобто додатково 15 зразків тестової вибірки стали розпізнаватися правильно.

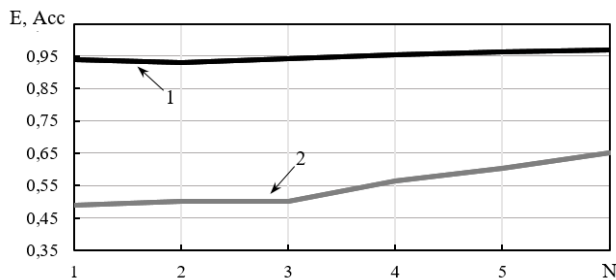


Рис. 6. Графік залежності критерію ефективності від кількості зразків, які були розмічені користувачем в режимі активного навчання:
1 – нормований інформаційний критерій (7), обчислений за навчальною вибіркою;
2 – точність, обчислена за тестовою вибіркою

Для підвищення репрезентативності навчальних даних додаватимемо зразки, що виникають під час трекінгу об'єктів, розмічених користувачем та тих, що розпізнаються з високим рівнем впевненості. На рис. 7 показано залежності нормованого інформаційного критерію (7) за навчальною вибіркою та точність за тестовою вибіркою від кількості зразків, що додаються за результатами трекінгу. При цьому встановлено таке порогове значення $\mu_{th2} = 0,7$.

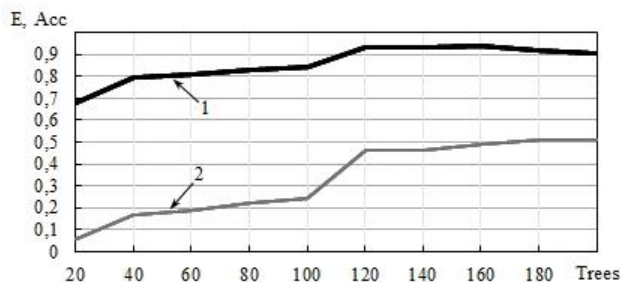


Рис. 7. Графік залежності критерію ефективності від кількості зразків, які були розмічені користувачем в режимі активного навчання:
1 – нормований інформаційний критерій (7), обчислений за навчальною вибіркою; 2 – точність, обчислена за тестовою вибіркою

Аналіз рис. 7 показує, що збільшення варіації навчальних зразків, отриманих внаслідок трекінгу дозволив додатково підвищити точність вирішальних правил навіть без втручання користувача. Було додано 178 навчальних зразків і отримано точність за тестовою вибіркою до 98,5 %.

Таким чином, запропоновані алгоритми дозволяють зменшити трудомісткість розмітки навчальних даних та підвищити репрезентативність вибірок і точність вирішальних правил. При цьому сформовані під час функціонування кадри зображень з розміченими обмежувальними прямокутниками можуть бути використані для навчання глибокого детектора об'єктів в режимі офлайн.

Висновки

1. Наукова новизна одержаних результатів:
– розроблено модель класифікаційного аналізу рухомих об'єктів, що має не високу обчислювальну складність, може бути використана для полегшення збору розмічених навчальних даних, яка складається з детектора руху, трекера об'єктів, згорткового розріджено кодуючого екстрактора ознак та стекованого інформаційно-екстремального класифікатора;

– розроблено метод навчання детектора рухомих об'єктів, що полягає в попередньому навчанні без вчителя екстрактора ознак на основі розріджено кодуючого нейронного газу, початковому навчанні з учителем стекованого інформаційно-екстремального класифікатора, в основі якого лежить двійкове кодування спостережень деревами рішень за схемою бустінгу, а також перенавчанні шляхом перебудови 20% останніх дерев рішень під час надходження нових розмічених даних;

– запропоновано формування нових розмічених даних шляхом саморозмітки зразків, які розпізнаються з високим рівнем впевненості, та шляхом ручної розмітки зразків, які розпізнаються з низьким рівнем впевненості, з наступним трекінгом цих зразків і балансуванням вибірок шляхом андерсмплінгу в рамках дихотомічної схеми “кожен клас проти всіх”.

2. Практична цінність отриманих результатів для безпілотної авіації полягає у формуванні сучасної науково-методологічної основи проектування здатних навчатися автономних систем детектування об'єктів інтересу на місцевості, що функціонують за умов ресурсних та інформаційних обмежень. Запропоновані моделі і методи навчання дозволяють також знизити трудомісткість формування розміченого навчального набору даних для навчання традиційних глибоких моделей детектування об'єктів на зображенні. При цьому результати імітаційного моделювання підтверджують високу ефективність моделі та методу навчання.

Робота виконана на базі лабораторії інтелектуальних систем кафедри комп'ютерних наук Сумського державного університету при фінансовій підтримці МОН України в рамках держбюджетної науково-дослідної роботи ДР № 0117U003934.

Література

1. A UAV-Cloud System for Disaster Sensing Applications [Text] / C. Luo, J. Nightingale, E. Asemota, C. Grecos // *IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, 2015.* – P. 1–5. DOI: 10.1109/VTCspring.2015.7145656.

2. Bandwidth-Efficient Live Video Analytics for Drones Via Edge Computing [Text] / J. Wang, Z. Feng, Z. Chen, S. George, M. Bala, P. Pillai, S. Yang, M. Satyanarayanan // *IEEE/ACM Symposium on Edge Computing (SEC)*, Bellevue, WA, 2018. – P. 159–173. DOI:10.1109/SEC.2018.00019.
3. Savitha, C. Motion detection in video surveillance: A systematic survey [Text] / C. Savitha, D. Ramesh // *2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, 2018. – P. 51–54. DOI: 10.1109/ICISC.2018.8398880.
4. Moving object detection and tracking from video captured by moving camera [Text] / Wu-ChihHu, Chao-HoChen, Tsong-YiChen, Deng-YuanHuang, Zong-CheWu // *Journal of Visual Communication and Image Representation*. – 2015. – Vol. 30. – P. 164–180. DOI:10.1016/j.jvcir.2015.03.003.
5. Detection of moving objects with a moving camera using non-panoramic background model [Text] / S. W. Kim, K. Yun, K. M. Yi, S. J. Kim, J. Y. Choi // *Machine Vision and Applications*. – 2013. – Vol. 24, Iss. 5. – P. 1015–1028. DOI:10.1007/s00138-012-0448-y.
6. Logoglu, K. Feature-based efficient moving object detection for low-altitude aerial platforms [Text] / K. Logoglu, H. Lezki, M. Yucel // *IEEE International Conference on Computer Vision Workshop (ICCVW)*, Venice, Italy, 2017. – P. 2119–2128. DOI: 10.1109/ICCVW.2017.248.
7. Okafor, E. Comparative study between deep learning and bag of visual words for wild-animal recognition [Text] / E. Okafor, P. Pawara // *IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, 2016. – P. 1–8. DOI: 10.1109/SSCI.2016.7850111.
8. Lemley, J. Smart augmentation learning an optimal data augmentation strategy [Text] / J. Lemley, S. Bazrafkan, P. Corcoran // *IEEE Access*. – 2017. – Vol. 5. – P. 5858–5869. DOI: 10.1109/ACCESS.2017.2696121.
9. Labusch, K. Sparse coding neural gas: learning of overcomplete data representations [Text] / K. Labusch, E. Barth, T. Martinetz // *Neurocomputing*. – 2009. – Vol. 72, Iss. 7–9. – P. 1547–1555. DOI:10.1016/j.neucom.2008.11.027.
10. Optimization of convolutional neural network using microcanonical annealing algorithm [Text] / V. Ayumi, L. M. R. Rere, M. I. Fanany, A. M. Arymurthy // *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Malang, 2016, P. 506–511. DOI: 10.1109/ICACSIS.2016.7872787.
11. The model and training algorithm of compact drone autonomous visual navigation system [Text] / V. Moskalenko, A. Moskalenko, A. Korobov, V. Semashko // *Data*. – 2019. – Vol. 4, Iss. 1. – P. 1–14. DOI: 10.3390/data4010004.
12. Improving the effectiveness of training the on-board object detection system for a compact unmanned aerial vehicle [Text] / V. Moskalenko, S. Dovbysh, I. Naumenko, A. Moskalenko, A. Korobov // *Eastern-European Journal of Enterprise Technologies*. – 2018. – Vol. 4/9, Iss. (94). – P. 19–26. DOI: 10.15587/1729-4061.2018.139923.
13. Montoya-Catalá, M. Experiments in combining boosting and deep stacked networks [Text] / M. Montoya-Catalá, R. F. Alvear-Sandoval, A. R. Figueiras-Vidal // *IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, Vietri sul Mare, 2016. – P. 1–6. DOI: 10.1109/MLSP.2016.7738874.
14. Tareen, S. A. K. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK [Text] / S. A. K. Tareen, Z. Saleem // *International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, 2018. – P. 1–10. DOI: 10.1109/ICOMET.2018.8346440.
15. Image restoration based on morphological operations [Text] / A. M. Raid, W. Khedr, M. El-dosuky, M. I. Aoud // *International Journal of Computer Science, Engineering and Information Technology*. – 2014. – Vol. 4. – P. 9–21. DOI: 10.5121/ijcseit.2014.4302.
16. An improved TLD target tracking algorithm [Text] / T. Xu, C. Huang, Q. He, G. Guan, Y. Zhang // *IEEE International Conference on Information and Automation (ICIA)*, Ningbo, 2016. – P. 2051–2055. DOI: 10.1109/ICInfA.2016.7832157.
17. Wang, D.-w. Undercomplete dictionary-based feature extraction for radar target identification [Text] / D.-w. Wang, X. Ma, Y. Su // *Progress in Electromagnetics Research M*. – 2008. – Vol. 1. – P. 1–19. DOI: 10.2528/PIERM08012805.
18. Luo, Z. MIO-TCD: A new benchmark dataset for vehicle classification and localization [Text] / Z. Luo // *IEEE Transactions on Image Processing*. – 2018. – Vol. 27, No. 10. – P. 5129–5141. DOI: 10.1109/TIP.2018.2848705.
19. A large-scale benchmark dataset for event recognition in surveillance video [Text] / S. Oh, A. Hoogs, A. G. Amitha Perera, N. Cuntoor // *IEEE Computer Vision and Pattern Recognition (CVPR)*. – Colorado Springs, CO, USA, 2011. – P. 3153–3160. DOI:10.1109/CVPR.2011.5995586.
20. MIO-TCD dataset [Electronic resource]. – Access mode: <http://podoce.dinf.usherbrooke.ca/challenge/dataset/>. – Access date 24.05.2019. – Title by screen.
21. VIRAT Video Dataset [Electronic resource]. – Access mode: <http://www.viratdata.org/> – Access date 24.05.2019. – Title by screen.

References

1. Luo, C., Nightingale, J., Asemota, E., Grecos, C. A UAV-Cloud System for Disaster Sensing Applications, 2015 *IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, 2015, pp. 1–5. DOI: 10.1109/VTCspring.2015.7145656.
2. Wang, J., Feng, Z., Chen, Z., George, S., Bala, M., Pillai, P., Yang, S., Satyanarayanan, M.

Bandwidth-Efficient Live Video Analytics for Drones Via Edge Computing, *2018 IEEE/ACM Symposium on Edge Computing (SEC)*, Bellevue, WA, 2018, pp. 159–173. DOI:10.1109/SEC.2018.00019.

3. Savitha, C., Ramesh, D. Motion detection in video surveillance: A systematic survey, *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, 2018, pp. 51–54. DOI: 10.1109/ICISC.2018.8398880.

4. Wu-ChihHu, Chao-HoChen, Tsong-YiChen, Deng-YuanHuang, Zong-CheWu, Moving object detection and tracking from video captured by moving camera *Journal of Visual Communication and Image Representation, Elsevier*, vol. 30, pp. 164–180. DOI:10.1016/j.jvcir.2015.03.003.

5. Kim, S. W., Yun, K., Yi, K. M., Kim, S. J., Choi, J. Y. Detection of moving objects with a moving camera using non-panoramic background model, *Machine Vision and Applications*, 2013, vol. 24, iss. 5, pp 1015–1028. DOI:10.1007/s00138-012-0448-y.

6. Logoglu, K., Lezki, H., Yucel, M. Feature-based efficient moving object detection for low-altitude aerial platforms, *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Venice, Italy, 2017, pp. 2119–2128. DOI: 10.1109/ICCVW.2017.248.

7. Okafor, E., Pawara, P. Comparative study between deep learning and bag of visual words for wild-animal recognition, *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, 2016, pp. 1–8. DOI: 10.1109/SSCI.2016.7850111.

8. Lemley, J., Bazrafkan, S., Corcoran, P. Smart augmentation learning an optimal data augmentation strategy, *IEEE Access*, 2017, vol. 5, pp. 5858–5869. DOI: 10.1109/ACCESS.2017.2696121.

9. Labusch, K., Barth, E., Martinetz, T. Sparse coding neural gas: learning of overcomplete data representations, *Neurocomputing*, 2009, vol. 72, iss. 7–9, pp. 1547–1555. DOI:10.1016/j.neucom.2008.11.027.

10. Ayumi, V., Rere, L. M. R., Fanany, M. I., Arymurthy, A. M. Optimization of convolutional neural network using microcanonical annealing algorithm, *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Malang, 2016, pp. 506 – 511. DOI: 10.1109/ICACSIS.2016.7872787.

11. Moskalenko, V., Moskalenko, A., Korobov, A., Semashko, V. The model and training algorithm of compact drone autonomous visual navigation system, *Data*, 2019, vol. 4, iss. 1, DOI: 10.3390/data4010004.

12. Moskalenko, V., Dovbysh, S., Naumenko, I.,

Moskalenko, A., Korobov, A. Improving the effectiveness of training the on-board object detection system for a compact unmanned aerial vehicle, *Eastern-European Journal of Enterprise Technologies*, 2018, vol. 4, no.9 (94), pp. 19–26. DOI: 10.15587/1729-4061.2018.139923.

13. Montoya-Catalá, M., Alvear-Sandoval, R. F., Figueiras-Vidal, A. R. Experiments in combining boosting and deep stacked networks, *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, Vietri sul Mare, 2016, pp. 1–6. DOI: 10.1109/MLSP.2016.7738874.

14. Tareen, S. A. K., Saleem, Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK, *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Sukkur, 2018, pp. 1–10. DOI: 10.1109/ICOMET.2018.8346440.

15. Raid, A. M., Khedr, W., El-dosuky, M., Aoud, M. Image restoration based on morphological operations. *International Journal of Computer Science, Engineering and Information Technology*, vol. 4, pp. 9–21, 2014. DOI: 10.5121/ijcseit.2014.4302.

16. Xu, T., Huang, C., He, Q., Guan, G., Zhang, Y. An improved TLD target tracking algorithm, *2016 IEEE International Conference on Information and Automation (ICIA)*, Ningbo, 2016, pp. 2051–2055. DOI: 10.1109/ICInfA.2016.7832157.

17. Wang, D.-w., Ma, X., Su, Y. Undercomplete dictionary-based feature extraction for radar target identification, *Progress in Electromagnetics Research M*, vol. 1, 2008, pp. 1–19. DOI: 10.2528/PIERM08012805.

18. Luo Z. MIO-TCD: A new benchmark dataset for vehicle classification and localization, in *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5129–5141, Oct. 2018. DOI: 10.1109/TIP.2018.2848705.

19. Oh, S., Hoogs, A., Amitha Perera, A. G., Cuntoor, N. A large-scale benchmark dataset for event recognition in surveillance video, *2011 Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2011, pp. 3153–3160. DOI:10.1109/CVPR.2011.5995586.

20. MIO-TCD dataset. Available at: <http://podoce.dinf.usherbrooke.ca/challenge/dataset/> (accessed 24.05.2019).

21. VIRAT Video Dataset. Available at: <http://www.viratdata.org/> (accessed 24.05.2019).

Надійшла до редколегії 5.06.2019 розглянута на редколегії 12.06.2019

МОДЕЛЬ И МЕТОД ОБУЧЕНИЯ СИСТЕМЫ КЛАССИФИКАЦИИ ПОДВИЖНЫХ ОБЪЕКТОВ ДЛЯ МАЛОГАБАРИТНЫХ БЕСПИЛОТНЫХ АППАРАТОВ

В. В. Москаленко, Н. А. Зарецкий, А. С. Москаленко

В работе разработана модель классификации, которая состоит из детектора движения, трекера объекта, сверточного разреженного кодированного экстрактора признаков и стекированного информационно-

экстремального классификатора. Предложенная модель характеризуется низкой вычислительной сложностью и может быть использована в качестве инструмента сбора размеченных данных для глубокой модели детектора подвижных объектов.

Кроме того, разработан метод обучения детектора движущихся объектов. Метод включает обучение экстрактора признаков без учителя, методом с использованием разреженно-кодированного нейронного газа, обучении с учителем и дальнейшей тонкой настройке стекированного информационно-экстремального классификатора. Также предлагается автоматическая разметка новых полученных данных для образцов с высокой достоверностью прогноза и ручная разметка для образцов с малой достоверностью, с последующим трекингом объекта. При этом балансировка классов осуществляется на основе техники андерсэмплинга в рамках стратегии «один против всех». Результаты моделирования на открытых наборах данных подтверждают работоспособность предложенной модели и метода обучения для практического использования.

Ключевые слова: классификация; отслеживание движения; детектирование объектов; свёрточная нейронная сеть; разреженно кодированный нейронный газ; информационно-экстремальное обучение; активное обучение; самообучение.

MODEL AND TRAINING METHOD OF MOVING OBJECT CLASSIFICATION SYSTEM FOR A COMPACT DRONE

V. V. Moskalenko, M. O. Zaretskyi, A.S. Moskalenko

The classification model which consists of the motion detector, object tracker, convolutional sparse coded feature extractor and stacked information-extreme classifier is developed. It is proposed to build a motion detector based on the difference of consecutive aligned frames where alignment is performed via keypoints matching, homography estimation, and projective transformations. Motion detector seeks to simplify object classification task through reduction of input data variations and resource savings for motion region search model synthesis without training. The proposed model is characterized by low computational complexity and it can be used as labeling dataset gathering tool for deep moveable object detector. Furthermore, the training method for moving object detector is developed. The method consisting in unsupervised pretraining feature extractor based on sparse coding neural gas, supervised pretraining and following fine-tuning of stacked information-extreme classifier. Using soft-competitive learning scheme in sparse coding neural gas facilitates robust convergence to close to optimal distributions of the neurons over the data. Sparse coding neural gas reduces the requirements for the volume of labeled observations and computational resource. As a criterion for the effectiveness of classifier's machine training, the normalized modification of S. Kullback's information measure is considered. Labeling new emerging data through self-labeling for high prediction score cases and manual labeling for low prediction score cases, and following labeled object tracking are also offered. In this case, class balancing using undersampling within dichotomous strategy "one-against-all". The set of classes include bicycle, bus, car, motorcycle, pickup truck, articulated truck, and background. Simulation results on MIO-TCD dataset confirm the suitability of the proposed model and training method for practical usage.

Keywords: classification; motion detection; object detection; convolutional neural network; sparse coding neural gas; information-extreme learning; active learning; self-taught learning.

Москаленко Вячеслав Васильович – канд. техн. наук, доцент каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Зарецький Микола Олександрович – аспірант каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Москаленко Альона Сергіївна – канд. техн. наук, асистент каф. комп'ютерних наук Сумського державного університету, Суми, Україна.

Moskalenko Viacheslav Vasilyovich – PhD, associate professor of Computer Sciences Department of Sumy State University, Sumy, Ukraine, e-mail: v.moskalenko@cs.sumdu.edu.ua, ORCID Author ID: 0000-0001-6275-9803.

Zaretsky Nikolay Olexandrovich – PhD student of Computer Sciences Department of Sumy State University, Sumy, Ukraine, e-mail: n.zaretskij@gmail.com, ORCID Author ID: 0000-0001-9117-5604.

Moskalenko Alona Sergiyvna – PhD, teaching assistant of Computer Sciences Department of Sumy State University, Sumy, Ukraine, e-mail: a.moskalenko@cs.sumdu.edu.ua, alenarizhova@gmail.com, ORCID Author ID: 0000-0003-3443-3990.