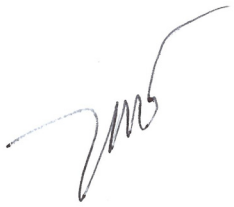
Міністерство освіти і науки України
Одеський національний політехнічний університет

НГУЄН ТХІ КХАНЬ ТІЄН



УДК 004.932

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ЛОКАЛІЗАЦІЇ ТЕКСТОВИХ ОБЛАСТЕЙ НА ЗОБРАЖЕННЯХ ЗІ СКЛАДНИМ ФОНОМ

05.13.06 – Інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Одеса – 2015

Дисертацією є рукопис.

Робота виконана в Одеському національному політехнічному університеті Міністерства освіти і науки України.

Науковий керівник — кандидат технічних наук, доцент **Ніколенко Анатолій Олександрович,** Одеський національний політехнічний університет, доцент кафедри інформаційних систем

Офіційні опоненти:

доктор технічних наук, професор **Мещеряков Володимир Іванович**, Одеський державний екологічний університет, завідувач кафедри інформатики

доктор технічних наук, професор **Романюк Олександр Никифорович,** Вінницький національний технічний університет, перший проректор

Захист відбудеться 28 серпня 2015 р. о 13-30 на засіданні спеціалізованої вченої ради Д 41.052.01 Одеського національного політехнічного університету за адресою: 65044, м. Одеса, пр. Шевченка, 1, ауд. 400а.

З дисертацією можна ознайомитись у науково-технічній бібліотеці Одеського національного політехнічного університету за адресою: 65044, м. Одеса, пр. Шевченка, 1.

Автореферат розісланий "28" липня 2015 р.

Вчений секретар
спеціалізованої вченої ради — О.Є. Колесніков

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Однією з задач, що вирішуються інформаційними системами переробки графічної інформації в умовах постійного зростання її обсягів, є задача розпізнавання тексту на зображеннях (фотографіях, відео-файлах, сканованих документах тощо). Ця задача вимагає розробки і використання сучасних прикладних інформаційних технологій (ІТ) розпізнавання тексту для автоматизації документообігу, запровадження паспортно-візової інформації, ведення обліку рухомого транспорту, ідентифікації виробів, пошуку текстової інформації в мультимедійних колекціях тощо. Однією з базових складових ІТ розпізнавання тексту є процедура локалізації текстових областей (ТО) на зображеннях, застосування якої дозволяє скоротити обсяг оброблюваної інформації та визначає оперативність і якість інформаційних систем переробки графічної інформації. В результаті виконання цієї процедури визначаються координати меж ТО. Однак, на практиці якість локалізації ТО на зображенні суттєво знижується в залежності від складності і неоднорідності фону та впливу різноманітних факторів: нерівномірності освітлення зображення, різного розташування рядків тексту по відношенню до фото- або відеореєструючої камери тощо.

Проведений аналіз показав, що при наявності складного і неоднорідного фону на зображеннях найкращі результати при оцінюванні якості локалізації ТО належать методам на основі кореляційно-екстремального підходу. Проте, їх недоліки – велика кількість моделей, великі витрати часу як у процесі розробки моделей, так і в процесі локалізації ТО, відсутність інваріантності до масштабу зображень символів – знижують оперативність систем переробки графічної інформації, що використовують ці методи.

Таким чином виникає протиріччя між якістю локалізації ТО та оперативністю цієї процедури, особливо при наявності складного фону на оброблювальних зображеннях. Тому створення інформаційної технології локалізації текстових областей на зображеннях зі складним фоном на основі нових моделей та методів, що дозволить усунути наведене протиріччя або зменшити його негативний вплив, є важливою науково-технічною задачею, що вирішується в даній роботі.

Зв'язок роботи з науковими програмами, планами, темами. Роботу виконано відповідно до пріоритетних напрямів досліджень Одеського національного політехнічного університету (ОНПУ), відповідно до координаційних планів Міністерства освіти і науки України, зокрема, в рамках наукових досліджень за держбюджетними науково-дослідними роботами кафедри інформаційних систем ОНПУ – «Аналіз та розпізнавання зображень на базі субградієнтних ітеративних алгоритмів у просторі вейвлет-перетворення» (№ держ. реєстрації 0109U008468), «Дослідження та розробка методів та алгоритмів розпізнавання візуальної інформації та аналізу сцен» (№ держ. реєстрації 0111U010458) – при безпосередній участі автора.

Мета і задачі дослідження. Метою дисертаційної роботи є підвищення якості та оперативності локалізації текстових областей на зображеннях зі складним фоном в інформаційних технологіях переробки графічної інформації.

Для досягнення поставленої мети вирішені такі задачі:

– проведено аналіз інформаційних технологій локалізації текстових областей на зображеннях зі складним фоном, що використовуються при переробці графічної інформації;

– створено модель представлення символу тексту на зображенні;

– розроблено метод формування текстових областей на зображеннях шляхом використання ймовірнісної моделі представлення символу тексту при формуванні навчальної вибірки для навчання класифікатора;

– розроблено метод локалізації текстових областей на зображеннях зі складним фоном на основі згорткової нейронної мережі.

– розроблено на основі запропонованих моделі та методів інформаційну технологію локалізації текстових областей на зображеннях для автоматичної переробки графічної інформації.

Об'єкт дослідження – процес локалізації текстових областей на зображеннях в системах переробки графічної інформації.

Предмет дослідження – моделі та методи локалізації текстових областей на зображеннях зі складним фоном.

Методи дослідження. Теорія ймовірностей і методи цифрової обробки зображень використані при створенні моделі представлення символу та методу формування текстових областей на зображенні; теорія штучних нейронних мереж, методи вейвлет-аналізу використані при розробці кореляційно-екстремального методу виявлення текстових областей та методу локалізації текстових областей на зображеннях зі складним фоном на основі згорткової нейронної мережі; методи теорії ймовірності, математичної статистики та імітаційного моделювання використані при розробці та апробації ІТ локалізації ТО.

Наукова новизна одержаних результатів полягає в наступному:

– отримала подальший розвиток ймовірнісна модель представлення символу тексту шляхом урахування розподілу ймовірності інтенсивності в двовимірній області зображення символу, що дозволило формувати зображення псевдосимволів тексту з ймовірнісними характеристиками, які відповідають реальним символам;

– вперше розроблено метод формування текстових областей на зображеннях за рахунок використання ймовірнісної моделі представлення символу тексту при формуванні навчальної вибірки, який дозволив автоматизувати процес створення представницької вибірки, що усунуло її суб'єктивність;

– отримав подальший розвиток кореляційно-екстремальний метод виявлення текстових областей на зображеннях, що реалізований на базі згорткової нейронної мережі, структура якої враховує багатомасштабне представлення зображення на основі двовимірного дискретного вейвлет-перетворення, що дозволило забезпечити високу якість та оперативність процедури виявлення ТО і зменшити залежність результатів виявлення ТО від розмірів і типів шрифтів тексту;

– удосконалено метод локалізації текстових областей на зображенні на основі розробленого кореляційно-екстремального методу виявлення ТО та їх морфологічної обробки, що дозволило забезпечити високу якість та оперативність процедури локалізації текстових областей на зображеннях зі складним фоном в системах

переробки графічної інформації.

Практичне значення одержаних результатів полягає в тому, що розроблені модель та методи реалізовані в інформаційній технології локалізації текстових областей на зображеннях зі складним фоном. Розроблена інформаційна технологія дає змогу підвищити ефективність локалізації текстових областей на зображеннях зі складним фоном. Результати тестування показали, що якість локалізації текстових областей за допомогою розробленої ІТ на 10-15% краще, ніж якість локалізації ТО такими програмними продуктами як FreeOCR (onlineOCR.net) та FineReader11.

Метод локалізації текстових областей на зображеннях зі складним фоном на основі згорткової нейронної мережі та інформаційна технологія локалізації текстових областей були використані для моделювання та розробки програмно-алгоритмічного забезпечення при автоматизації обліку накопиченої графічної та відеоінформації в науково-виробничому об'єднанні «Діскрет» (м. Одеса), а саме модифікації системи автоматизації перевірки якості упаковки продукції та правильності розподілення її при формуванні партій товару, що дозволило підвищити ефективність функціонування складу продукції на підприємстві.

Основні положення дисертаційного дослідження впроваджені в навчальний процес Одеського національного політехнічного університету на кафедрі інформаційних систем при викладанні дисциплін «Методи та засоби комп'ютерних інформаційних технологій», «Методи і засоби штучного інтелекту», «Обчислювальний інтелект» та при виконанні курсового проектування.

Особистий внесок здобувача. Усі наукові положення та практичні результати, що подані в дисертаційній роботі, були отримані автором особисто [1].

В роботах, опублікованих у співавторстві, особистий внесок здобувача такий: [2, 7, 8, 9, 10, 11] – проведено аналітичний огляд існуючих методів, моделей та підходів до вирішення задачі локалізації текстових областей на зображеннях зі складним фоном; [3, 12, 14] – розроблено та досліджено метод виявлення та локалізації текстових областей та виділення символів на зображеннях з неоднорідним фоном [4, 13, 15] – розроблені метод локалізації текстових областей на зображеннях зі складним фоном на основі згорткової мережі та структура згорткової нейронної мережі для локалізації текстових областей; [5] – розроблена ймовірнісна модель представлення символу тексту та метод формування текстових областей на основі ймовірнісної моделі представлення символу; [6] – розроблена інформаційна технологія локалізації текстових областей на зображеннях зі складним фоном.

Апробація результатів дисертації. Основні положення дисертаційної роботи доповідалися та обговорювалися на наукових семінарах кафедри інформаційних систем ОНПУ і наукових конференціях: міжнародних науково-практичних конференціях «Сучасні інформаційні та електронні технології» (СІЕТ-2010, 2012, 2013, 2014, м. Одеса); XI Всеукраїнській міжнародній конференції «Оброблення сигналів і зображень та розпізнавання образів» (УкрОБРАЗ’2012, м. Київ), VIII міжнародній науково-практичній конференції «Штучний інтелект. Інтелектуальні системи – 2012» (с. Кацівелі), XIII міжнародній науково-технічній конференції «Проблеми інформатики та моделювання» (Харків – Ялта, 2013).

Публікації. За темою дисертації опубліковано 15 робіт, у тому числі 6 статей у

наукових журналах з переліку фахових видань України (з яких 3 статті у виданнях, які входять у наукометричні бази), 9 у працях або тезах доповідей міжнародних та всеукраїнських конференцій.

Структура дисертаційної роботи. Дисертація складається зі вступу, 4 розділів, висновків, списку використаних джерел із 110 найменувань і 2 додатків. Загальний обсяг дисертації становить 115 сторінок. Робота містить 50 рисунків і 12 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету, задачі, об'єкт, предмет і методи дослідження, визначено наукову новизну та практичне значення одержаних результатів, наведено відомості про публікації та апробацію результатів дисертаційної роботи.

У першому розділі проведено аналіз інформаційних технологій локалізації текстових областей на зображеннях зі складним фоном, що використовуються при переробці графічної інформації.

Показано, що важливою задачею переробки графічної інформації (спеціалізовані зображення (медичні, супутникові та інші), фотографії, скановані документи, відео-файли, тощо) для широкого кола застосовань є розпізнавання тексту на зображеннях зі складним фоном, ефективність вирішення якої залежить від якості локалізації текстових областей. Фон вважається складним у випадку, коли текст на зображеннях не відокремлений від іншої інформації, а є частиною цієї інформації (текст на рекламних щитах, номери будинків, написи на одязі, автомобілях, тощо).

Показано, що процес локалізації текстових областей передбачає виконання двох основних процедур: виявлення текстової області та її верифікацію. Проведений аналіз інформаційних технологій локалізації текстових областей дозволив виділити три основні підходи до реалізації процедури локалізації ТО: кореляційно-екстремальний, на основі сегментації спектральних текстур та структурний. При використанні кореляційно-екстремального підходу для виявлення ТО для кожної області зображення та набору моделей ТО обчислюються взаємокореляційні функції, визначається максимальне значення та приймається класифікаційне рішення – «текстова область» або «не текстова область». Переваги даного підходу: висока якість локалізації, врахування просторового розташування символів в ТО, недоліки – невизначеність вибору моделі в разі складного фону, відсутність інваріантності до масштабу зображення символів, значні обчислювальні витрати, і як наслідок – великі витрати часу як у процесі розробки моделей, так і в процесі локалізації ТО. При підході на основі сегментації спектральних текстур напис розглядається як квазіперіодична послідовність імпульсів. Недоліки – не враховується просторове розташування символів, не забезпечується завадостійкість методу в умовах складного і неоднорідного фону. При структурному підході виявляються структурні примітиви (літери, символи) і оцінюється порядок їх просторового розташування (тобто перевіряється, чи утворюють вони напис). Недоліки – низька завадостійкість

в умовах складного і неоднорідного фону, відсутність інваріантності до масштабу зображення символів.

Визначено, що наведені підходи до локалізації ТО, за виключенням кореляційно-екстремального підходу, не враховують просторові співвідношення між символами тексту в ТО, що негативно позначається на якості локалізації ТО. Тому для подальшого дослідження в якості базового методу для реалізації процесу локалізації ТО на зображеннях зі складним фоном обрано кореляційно-екстремальний метод.

Визначена можливість підвищення оперативності цього методу шляхом його реалізації згортковими нейронними мережами.

Показано, що якість локалізації ТО на зображеннях зі складним фоном визначається етапом класифікації областей зображення на два класи: «текстова область» або «не текстова область», якість та оперативність якого залежить від кількості та виду зображень, що входять до навчальної вибірки. Для створення навчальних вибірок використовують спеціально підібрані бази зображень, наприклад, база даних конкурсу алгоритмів штучного інтелекту ICDAR 2011 Robust Reading Competition (http://www.cvc.uab.es/icdar2011competition/). Слід відзначити, що підбір таких баз та створення навчальних вибірок, як правило, визначається досвідом розробника та є суб'єктивним.

Зауважено, що при оцінюванні процедури локалізації потрібно враховувати показники достовірності виявлення ТО (на основі помилок I-го та II-го роду) і показники якості виділення меж ТО (по критерію Претта).

В розділі визначені основні задачі дослідження, які пов’язані зі створенням інформаційної технології локалізації текстових областей на зображеннях зі складним фоном, зокрема, задачі формування навчальної вибірки зображень з ТО; зменшення впливу суб'єктивного фактора при підготовці навчальної вибірки; підвищення оперативності кореляційно-екстремального методу виявлення ТО; створення методу локалізації на зображеннях зі складним фоном на основі згорткової нейронної мережі.

У другому розділі розроблена ймовірнісна модель представлення символу тексту з отриманням її параметрів на основі статистичної обробки реальних текстових областей і метод формування текстових областей на основі використання запропонованої моделі.

Зображення символу розглядається як двовимірний масив пікселів, в якому певним чином (залежно від шрифту, розмірів та ін.) розташовані інформативні та фонові пікселі (ФП). Зображення текстової області – це впорядкований певним чином набір зображень символів.

Запропоновано зображення символу представляти ймовірнісною моделлю представлення символу у вигляді кортежу

$$S = \langle \mathbf{H}, \mathbf{Wh}, \mathbf{P} \rangle, \qquad (1)$$

де **H** – вектор можливих значень висоти h_i i-го текстового символу (в пікселях); **Wh** – вектор можливих значень ширини wh_i i-го текстового символу (в пікселях);

P – множина матриць **p**, що відповідають ймовірності появи інформативного пікселя в зображеннях символу тексту розмірами $h_i \times wh_i$, $h_i \in \mathbf{H}, wh_i \in \mathbf{Wh}$, $i = \overline{1,n}$, n – кількість символів в алфавіті.

Для отримання ймовірнісної моделі представлення символу проведено аналіз бінарних зображень текстових символів англійських букв ('a', 'b', .., 'z') і десяткових цифр ('0', '1', .., '9') шрифтів різних типів (Arial, Times New Roman, Calibri, Cambria, Courier New, Adobe Garamond Pro, тощо) і розмірів (14,15, .., 20).

Запропоновано зображення всіх символів розділити на 4 групи залежно від характеру заповнення інформативними пікселями форми (рис. 1), що містить три зони (I_1, I_2, I_3).

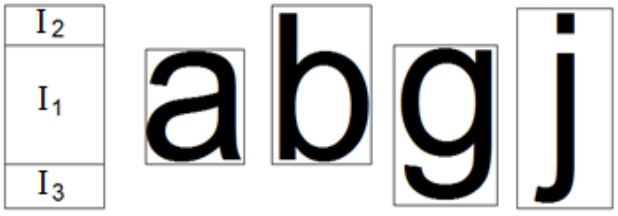
У першій групі містяться символи букв 'a', 'c', 'e', 'm', 'n', 'o', 'r', 's', 'u', 'v', 'w', 'x', 'z', інформативні пікселі (ІП) яких розташовані в зоні I_1.

У другій – 'b', 'd', 'f', 'h', 'i', 'k', 'l', 't', '0', '1', '2', '3', '4', '5', '6', '7', '8', '9', ІП розташовані одночасно в зонах I_1 і I_2.

У третій – 'g', 'p', 'q', 'i', ІП займають зони I_1 і I_3.

У четвертій – 'j', ІП займають всі три зони.

Для кожної групи символів визначені параметри **H**, **Wh**, **P** моделі (1). Для отримання матриці **p** з множини **P** для кожної групи символів визначається ймовірність $p_{i,j} = L_{i,j} / N_{gr}$ появи ІП в кожній точці зображення символу, де $L_{i,j}$ – сумарна кількість ІП у позиції з координатами (i, j) для всіх зображень символів групи, N_{gr} – кількість символів в групі. Приклад матриці ймовірності появи ІП в кожній точці зображення (**p**) для символів першої групи наведено в табл. 1. Одержані таким чином матриці складають базу ймовірнісних моделей представлення символів і використовуються для створення зображень псевдосимволів при формуванні навчальної вибірки.



1 група	I_1
2 група	$I_1 U I_2$
3 група	$I_1 U I_3$
4 група	$I_1 U I_2 U I_3$

Рисунок 1 – Розбиття символів на групи

Таблиця 1 – Матриця ймовірності появи інформаційного пікселя в кожній точці зображення для символів першої групи розміру 13×9

0,194	0,300	0,377	0,467	0,550	0,358	0,141	0,030	0,000
0,203	0,609	0,505	0,386	0,436	0,332	0,201	0,030	0,000
0,256	0,415	0,259	0,274	0,368	0,273	0,274	0,077	0,000
0,252	0,415	0,259	0,274	0,368	0,272	0,274	0,077	0,000
0,223	0,403	0,308	0,329	0,358	0,371	0,236	0,076	0,000
0,188	0,352	0,349	0,259	0,455	0,339	0,277	0,087	0,017
0,209	0,335	0,247	0,279	0,370	0,306	0,285	0,111	0,000
0,242	0,332	0,271	0,294	0,444	0,226	0,296	0,162	0,017
0,209	0,288	0,244	0,292	0,385	0,192	0,238	0,173	0,017
0,223	0,320	0,189	0,265	0,402	0,341	0,242	0,133	0,000
0,227	0,509	0,415	0,406	0,471	0,506	0,409	0,206	0,017
0,139	0,379	0,365	0,289	0,342	0,43	0,311	0,147	0,000
0,120	0,149	0,168	0,117	0,126	0,192	0,123	0,106	0,050

Запропоновано метод формування ТО на зображенні на основі застосування

ймовірнісної моделі представлення символу, реалізований наступним алгоритмом.

Крок 1. Проводиться вибір номера групи (i=1,..,4) та ініціалізація розмірів псевдосимволу (висоти h_i, і ширини wh_i) згідно з наявною базою ймовірнісних моделей представлення символів.

Крок 2. Генерується матриця випадкових чисел $\mathbf{TR_{rand}}$ розміру $h_i \times wh_i$, розподілених по рівномірному закону в діапазоні $[0 \ldots p_{imax}]$, де p_{imax} – максимальне значення ймовірності появи ІП в зображеннях символів i-ої групи.

Крок 3. Формується матриця **SM** бінарного зображення псевдосимволу шляхом порівняння матриць $\mathbf{TR_{rand}}$ та **p** з бази ймовірнісних моделей представлення символів за наступним правилом:

$$IF\ TR_{rand}(i,j) \geq p(i,j),\ THEN\ \ SM(i,j)=ІП,\ ELSE\ SM(i,j)=ФП.$$

Крок 4. Додається зображення псевдосимволу до фрагменту зображення текстової області .

Крок 5. Повторюються кроки 1-4 до досягнення необхідного розміру текстової області на зображенні.

У разі необхідності формування текстової області для напівтонового зображення отримане бінарне зображення оброблюється з допомогою фільтра Гауса.

Для створення кольорового зображення текстової області в кольоровому просторі *RGB* напівтонове зображення приймають за *R*-складову, а складові *G* і *B* отримують поелементним множенням елемента r_{ij} *R*-складової на величину $(1+\varepsilon_{ij})$, де ε_{ij} – випадкові числа, розподілені по рівномірному закону в діапазоні [-0,2 ... 0,2]. Тобто зображення текстової області формується як упорядкований певним чином набір зображень символів (рис. 2).

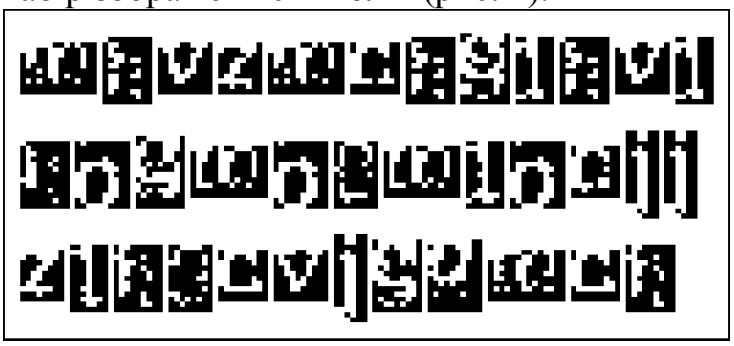


Рисунок 2 – Приклад сформованого зображення ТО

Для символів кожної групи визначена ймовірність появи інформаційного пікселя

$$p_{zs} = N_i /(h_i \cdot wh_i),$$

де N_i – загальна кількість інформаційних пікселів у зображенні символу.

Дослідження показали, що значення цієї ймовірності лежать в певному діапазоні та не залежать від розміру шрифту:

група 1 – p_{zs} = [0,330…0,543]; група 2 – p_{zs} = [0,296…0,606];

група 3 – p_{zs} = [0,314…0,482]; група 4 – p_{zs} = [0,393…0,397].

Одержані результати довели інваріантність запропонованої ймовірнісної моделі представлення символів (1) до змін масштабу.

Запропонована модель та метод формування ТО дозволили формувати зображення псевдосимволів тексту з ймовірнісними характеристиками, які відповідають реальним символам, та автоматизувати процес створення представницької вибірки, що усунуло суб'єктивний вплив на навчання різних класифікаторів при локалізації ТО.

В третьому розділі отримав подальший розвиток кореляційно-екстремальний метод виявлення текстових областей на зображеннях.

Для отримання ТО-кандидатів запропоновано удосконалений кореляційно-екстремальний метод виявлення ТО на основі ймовірнісної моделі представлення символу та двовимірного дискретного вейвлет-перетворення. Метод реалізується наступною послідовністю дій.

Крок 1. Для забезпечення інваріантності до масштабу та врахування як інтегральних так і диференціальних характеристик символів та ТО запропоновано використовувати ДВП-представлення зображення – багатомасштабне розкладання за вейвлет-базисом (застосовується дискретне вейвлет-перетворення на два рівні) еталонного зображення текстової області $ET(x,y)$ розміром $M \times N$ пікселей

$$WET^{A}_{k\varphi}(j_0,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} ET_{k-1}(x,y)\varphi_{j_0,m,n}(x,y),$$
$$WET^{i}_{k\psi}(j,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} ET_{k-1}(x,y)\psi^{i}_{j,m,n}(x,y),\ i=\{H,V,D\}, \quad (2)$$

де $k = 1, 2$ – поточний номер еталонного зображення в розкладанні, $ET_0(x,y) = ET(x,y)$, $ET_k(x,y) = WET^{A}_{k\varphi}(j_0,m,n)$ для $k = 1, 2$, $j_0 = 0, 1$ – початковий масштаб, $WET^{A}_{k\varphi}(j_0,m,n)$ – апроксимуючі коефіцієнти, $WET^{i}_{k\psi}(j,m,n)$ – деталізуючі коефіцієнти (горизонтальні H, вертикальні V, діагональні D) для масштабу $j>j_0$, $n = 0, 1,\ldots, N-1$, $m=0, 1, \ldots, M-1$ – координати зсуву вейвлетів $\varphi_{j_0,m,n}$ та $\psi^{i}_{j,m,n}$.

Крок 2. Виконується двовимірне дискретне вейвлет-перетворення на два рівні розкладання для фрагменту $FR(x,y)$ вхідного зображення того ж розміру, що й еталонне (отримуємо $FR^{A}_{k\varphi}(j_0,m,n)$, $FR^{i}_{k\psi}(j,m,n)$ аналогічно п.1).

Крок 3. Для отриманого на кожному рівні розкладання фрагменту вхідного зображення знаходиться коефіцієнт кореляції з відповідними еталонами, отриманими в п.1.

$$KOR^{i}_{k\alpha}(j,m,n) = WET^{i}_{k\alpha}(j,m,n) * FR^{i}_{k\alpha}(j,m,n),\ \alpha=\{\varphi,\psi\},\ i=\{A,H,V,D\}.$$

Крок 4. Приймається кваліфікаційне рішення про виявлення ТО на основі наступного вирішального правила

$IF\ \exists KOR^{i}_{k\alpha}(j,m,n):(KOR^{i}_{k\alpha}(j,m,n) > TR)$, $THEN\ FR(x,y) \in \text{ТО}$, $ELSE\ FR(x,y) \notin \text{ТО}$,

де TR – порогове значення коефіцієнта кореляції.

Обґрунтовано вибір в якості базису дискретного вейвлет-перетворення вейвлет-функцію Хаара завдяки меншим обчислювальним затратам при реалізації вейвлет-перетворення.

Проведені дослідження показали, що запропонований метод виявлення має високу якість виявлення ТО, але низьку оперативність, оскільки для кожного фрагменту вхідного напівтонового зображення потрібно обчислити $9 \times N_m$ коефіцієнтів, а для кольорового $27 \times N_m$, де N_m – кількість еталонних зображень. Для підвищення оперативності даного методу запропоновано реалізувати його за допомогою нейронних мереж, що дозволяють реалізувати процес паралельних обчислень функцій кореляції.

В роботі обґрунтовано вибір згорткової нейронної мережі для реалізації запропонованого кореляційно-екстремального методу виявлення текстових областей.

Запропонована структура згорткової нейронної мережі для виявлення текстових областей на напівтоновому зображенні (рис. 3).

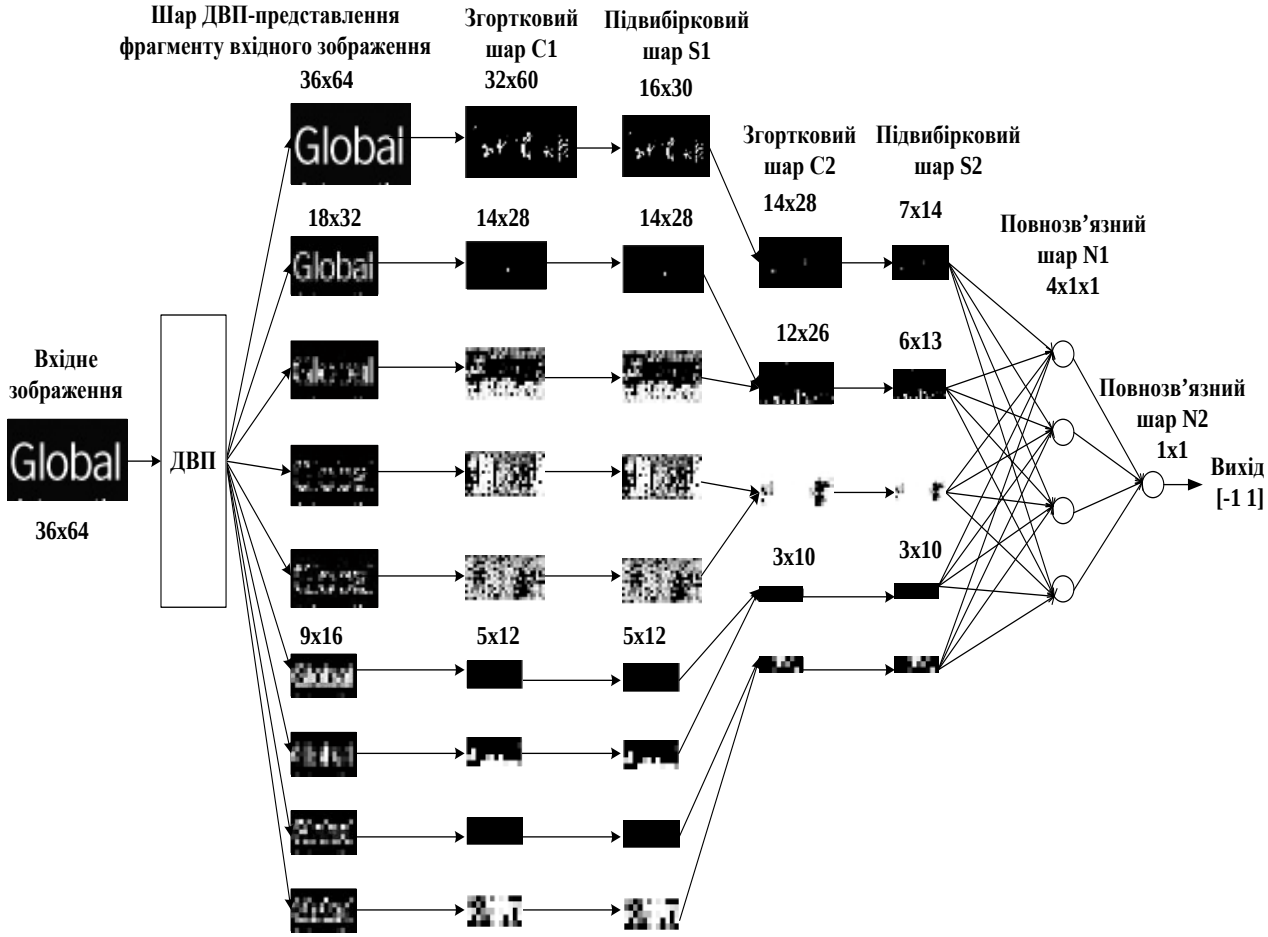


Рисунок 3 – Структура запропонованої згорткової нейронної мережі для виявлення ТО на напівтонових зображеннях

Згорткова нейронна мережа складається з шару ДВП-представлення фрагменту вхідного зображення, двох згорткових шарів *C*1 і *C*2 (для їх створення використовуються рецептивні матриці 5×5 і 3×3 пікселя відповідно), двох шарів підвибірки *S*1 і *S*2 (розмір рецептивної матриці 2×2) і двох повнозв'язних шарів *N*1 (4 нейрони) і *N*2 (один нейрон).

Шар ДВП-представлення фрагменту зображення реалізує дискретне вейвлет-перетворення згідно з виразами (2). На виході шару вхідне зображення представляється одночасно декількома зображеннями різних розмірів у вигляді піраміди вхідних зображень (сформовано 9 зображень – два набори по 4 зображення з розмірами 18×32 і 9×16 пікселей та вхідне зображення). Значення інтенсивності всіх зображень перетворюються в діапазон [-1 , 1].

Як функція активації нейронів використовується гіперболічний тангенс.

Шар *C*1 має 9 згорткових карт, кожна з яких є результатом обробки піраміди вхідних зображень за допомогою згортки з маскою 5×5.

Особливістю шару підвибірки *S*1 є те, що він зменшує в 2 рази тільки одну

карту шару *C*1 розміром 32×60 пікселів, інші залишаються незмінними. Шар *C*2 містить 5 згорткових карт, кожна з яких отримана за допомогою згортки з маскою 3×3 деяких карт шару *S*1. У шарі підвибірки *S*2 карти шару *C*2 розміром 3×10 дублюються, решта – зменшуються в 2 рази. Кожен нейрон в шарі *N*1 з'єднується тільки з однією картою шару підвибірки *S*2.

В разі обробки кольорових зображень нейронна мережа має аналогічну структуру для кожного каналу *R, G, B* (колірний простір *RGB*).

Запропонований кореляційно-екстремальний метод виявлення для кольорового зображення реалізується наступним алгоритмом.

Крок 1. З вхідного кольорового зображення (колірний простір *RGB*) виділяється прямокутний фрагмент розміром 36×64 пікселя.

Крок 2. Формується ДВП-представлення, тобто піраміда зображень для виділеного фрагмента. Для цього здійснюється перетворення значень інтенсивності фрагмента зображення до діапазону [-1, 1] і виконується дискретне вейвлет-перетворення з розкладанням зображення на два рівня за кожною складовою каналів кольору *R, G, B*.

Крок 3. Отриманий набір пірамід зображень надходить на попередньо навчену згорткову нейронну мережу, яка класифікує дану ділянку зображення на два класи: «текстова область» і «не текстова область» – проводиться вибір ТО-кандидатів.

Крок 4. Повторюються кроки алгоритму 1 - 3 (для всього зображення).

Крок 5. Формується список координат ТО-кандидатів.

Для навчання мережі був обраний алгоритм зворотного поширення помилки з врахуванням особливостей структури згорткової мережі. Значення початкових синаптичних ваг для всіх згорткових шарів мережі генерувалися відповідно до закону рівномірного розподілу з нульовим математичним очікуванням і дисперсією, обернено пропорційною квадратному кореню з кількості синаптичних зв'язків нейрона.

Навчальна вибірка сформована за допомогою розроблених ймовірнісної моделі представлення символу тексту та методу формування текстових областей на зображеннях. Автоматизація процесу створення навчальної вибірки дозволила усунути вплив суб'єктивних факторів при її формуванні та підвищити рівень її представництва. В ході навчання зображення сформованих ТО послідовно подавались на вхід згорткової нейронної мережі. В результаті навчання мережа формувала відгуки зі значеннями -1 для фрагмента зображення, що не містить текст, і +1 для фрагмента зображення з текстом. Перевірка якості навчання мережі здійснювалась окремо на контрольних вибірках (по 500 зображень) кольорових та напівтонових зображень. Кожна контрольна вибірка включала 50% зображень ТО, що сформовані за допомогою запропонованого методу формування текстових областей на зображеннях, і 50% реальних зображень.

Для оцінювання якості виявлення текстових областей (табл. 2) використано такі оцінки, як помилки першого та другого роду, міри *Precision* (*Pr*) та *Recall* (*Rc*):

$$Pr = (OTP / OP) \times 100\%,$$
$$Rc = (OTP / (OTP + HOTP)) \times 100\%,$$

де *OTP*– вірно виявлені ТО; *OP* – області, які виявлені як ТО, хоча серед них є і нетекстові, *HOTP* – ТО, що не виявлені.

Таблиця 2 – Результати оцінювання якості виявлення текстових областей

Умови формування навчальної вибірки		*Pr*, %	*Rc*, %	Помилка I роду, %	Помилка II роду, %
без використання ймовірнісної моделі представлення символу тексту		77,7	80,0	22,3	20,0
з використанням ймовірнісної моделі представлення символу	напівтонові зображення	87,7	86,6	12,1	12,9
	кольорові зображення	87,3	87,2	12,4	12,7

Таким чином, якість виявлення текстових областей контрольної вибірки (500 зображень на складному фоні) з використанням запропонованого кореляційно-екстремального методу та його реалізації на основі згорткової нейронної мережі з ДВП-представленням зображення на основі дискретного вейвлет-перетворення і навчальним набором на основі створених текстових областей склала 87,7 %, що на 10% вище якості виявлення ТО мережею без використання навчальних наборів на основі створених текстових областей (77,7%).

За рахунок автоматизації процесу створення навчальної вибірки суттєво зменшився час її формування (до хвилин порівняно з годинами і днями, необхідними для отримання та обробки достатньої кількості реальних зображень ТО), що підвищило оперативність виявлення ТО, а використання згорткової нейронної мережі з ДВП-представленням зображення дозволило зменшити залежність результатів виявлення ТО від розмірів та типів шрифтів тексту.

Запропоновано удосконалити метод локалізації текстових областей за рахунок використання розробленого кореляційно-екстремального методу виявлення ТО та морфологічної обробки зображень текстових областей.

Удосконалений метод локалізації реалізується наступним алгоритмом.

Крок 1. ТО-кандидат, що отримана з допомогою розробленого методу виявлення ТО, піддається дискретному двовимірному вейвлет-перетворенню з вейвлет-функцією Хаара. В результаті формується піраміда з п'яти (для напівтонових) або 15 (для кольорових, простір *RGB*) зображень.

Крок 2. Для кожного з зображень, що візуалізують деталізуючі коефіцієнти, виконується морфологічна операція – дилатація зі структурними елементами розміру 3×5, 3×3 і 7×5 пікселей (для горизонтальних, діагональних та вертикальних зображень деталізуючих коефіцієнтів відповідно).

Крок 3. Оброблені зображення разом з вхідним та двома зображеннями апроксимуючих коефіцієнтів подаються на вхід згорткової нейронної мережі, яка остаточно класифікує ТО-кандидат на два класи: «текстова область» або «не текстова область».

Крок 4. Уточнюються координати текстової області і визначаються її межі у

вигляді відрізків прямих ліній.

Для тестування роботи нейронної мережі при локалізації ТО оцінювалися точність класифікації фрагментів зображень з навчальної та контрольної вибірок, які містили текстові області. Для порівняння точності локалізації текстових областей виконано локалізацію набору зображень зі складним фоном з телевізійних роликів, фільмів, рекламних банерів за допомогою запропонованої ІТ локалізації ТО і для порівняння – програм FineReader11 та FreeOCR.

Показником, що характеризує точність локалізації, обрано метричний показник близькості між множиною координат еталонних меж I_{et} ТО тестового зображення і множиною координат меж I_t, отриманими в результаті застосування запропонованого методу локалізації (критерій FOM (Figure of Merit), запропонований Преттом)

$$F_{loc} = \frac{1}{\max(N_t, N_{et})} \sum_{i=1}^{N_t} \frac{1}{1 + d^2(i)},$$

де N_{et} – кількість пікселей у множині I_{et}; N_t – кількість пікселей у множині I_t; $d^2(i) = \min\left(I_t(i) - I_{et}(j)\right)^2$, $j = \overline{1, N_t}$ – квадрат відстані від *i*-го пікселя множини I_t до найближчого пікселя множини I_{et}.

Використання запропонованого методу локалізації дозволило підвищити якість локалізації на 10,5 %.

Таким чином, в розділі розроблено кореляційно-екстремальний метод виявлення текстових областей на зображеннях, що реалізований на базі згорткової нейронної мережі, та метод локалізації текстових областей на основі розробленого кореляційно-екстремального методу виявлення ТО та морфологічної обробки зображень текстових областей, що послужили основою для створення інформаційної технології локалізації текстових областей на зображеннях зі складним фоном для автоматичної переробки графічної інформації.

У четвертому розділі розроблена інформаційна технологія локалізації текстових областей на зображеннях.

Розроблено структуру інформаційної технології локалізації ТО (табл. 3) на основі запропонованих методів виявлення та локалізації текстових областей на зображенні. Інформаційна технологія локалізації ТО на зображеннях зі складним фоном містить два етапи: підготовчий (навчання) та основний (локалізація).

На підготовчому етапі за допомогою методу формування текстових областей на зображенні на основі застосування ймовірнісної моделі представлення символу формується навчальна вибірка та навчаються згорткові нейронні мережі для реалізації запропонованих методів виявлення та локалізації ТО на зображеннях.

На основному етапі використовуються попередньо навчені згорткові нейронні мережі, які з допомогою кореляційно-екстремального методу виявлення ТО та удосконаленого методу локалізації ТО виконують остаточну класифікацію ТО-кандидата на два класи: «текстова область» і «не текстова область».

Таблиця 3 – Структура інформаційної технології локалізації текстових областей на зображенні на основі згорткової нейронної мережі

Назва етапу	Процедура, що виконується	Модель або метод, що запропоновано використати
I. Підготовчий (етап навчання)	1.1.Створення навчальної вибірки	
	1.1.1 Створення бази інформаційних ймовірнісних моделей представлення символів тексту	Ймовірнісна модель представлення символу тексту
	1.1.2 Формування зображення ТО на основі застосування ймовірнісної моделі представлення символу	Метод формування зображення ТО на основі застосування ймовірнісної моделі представлення символу
	1.2 Навчання згорткової нейронної мережі для виявлення ТО-кандидатів	Кореляційно-екстремальний метод виявлення текстових областей (КЕМВ ТО), що використовує багатомасштабне представлення зображення на основі двовимірного дискретного вейвлет-перетворення.
	1.3 Навчання згорткової нейронної мережі для локалізації ТО	Удосконалений метод локалізації текстових областей (УМЛ ТО) на основі розробленого кореляційно-екстремального методу виявлення ТО та морфологічної обробки зображень текстових областей.
II. Основний (етап локалізації ТО)	2.1 Виявлення ТО-кандидатів	КЕМВ ТО
	2.2 Формування піраміди зображень ТО-кандидатів з використанням вейвлет-перетворення	УМЛ ТО
	2.3 Верифікація та остаточна локалізація ТО	УМЛ ТО

Якість локалізації зображень зі складним фоном (телевізійних роликів, фільмів, рекламних банерів) за допомогою запропонованої інформаційної технології склала 87,7% (*Precision*), 86,6% (*Recall*), що перевершує результат, отриманий за допомогою програм FineReader11 і FreeOCR (рис. 4).

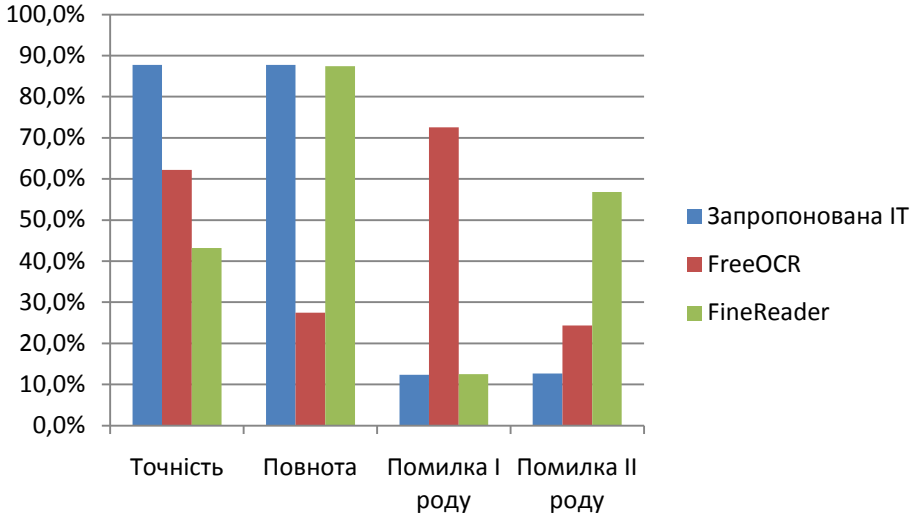


Рисунок 4 – Порівняльна оцінка достовірності локалізації текстових областей

Приклади локалізації текстових областей на реальних зображеннях, отримані з використанням запропонованої інформаційної технології, наведені на рис. 5-6.

а) б)

Рисунок 5 – Приклад локалізації ТО на зображенні з застосуванням запропонованої технології: а – вхідне кольорове зображення; б – результат локалізації ТО

а) б)

Рисунок 6 – Приклад локалізації ТО на зображенні з застосуванням запропонованої технології: а – вхідне напівтонове зображення; б – результат локалізації ТО.

Метод локалізації текстових областей на зображеннях зі складним фоном на основі згорткової нейронної мережі та інформаційна технологія локалізації текстових областей для автоматичної переробки графічної інформації були використані для моделювання та розробки програмно-алгоритмічного забезпечення при автоматизації обліку накопиченої графічної та відеоінформації в науково-виробничому об'єднанні «Діскрет», (м. Одеса), а саме модифікації системи автоматизації перевірки якості упаковки продукції та правильності розподілення її при формуванні партій товару, що дозволило підвищити ефективність функціонування складу продукції на підприємстві.

Проведення автоматизованої перевірки якості упаковки продукції та правильності розподілення її при формуванні партій товару дозволило підвищити ефективність функціонування складу продукції на підприємстві, що підтверджується відповідним актом впровадження.

ВИСНОВКИ

У дисертаційній роботі вирішена актуальна науково-практична задача розробки інформаційної технології локалізації текстових областей на зображеннях зі складним фоном з метою підвищення якості та оперативності локалізації текстових областей. При цьому отримано такі основні результати:

1. Аналіз існуючих інформаційних технологій, моделей і методів локалізації текстових областей на зображеннях зі складним фоном показав, що якість локалізації ТО на зображенні суттєво знижується в залежності від складності і неоднорідності фону та впливу різноманітних факторів. Запропоновано локалізацію текстових областей розглядати як класифікацію деякої області зображення на два класи: «текстова область» або «не текстова область». Обрано кореляційно-екстремальний підхід як базовий для виявлення текстових областей на зображенні завдяки його найкращій достовірності виявлення ТО.

2. Удосконалено ймовірнісну модель представлення символу тексту і запропоновано метод формування зображень текстових областей для створення представницької навчальної вибірки для навчання класифікаторів. Це дозволило автоматизувати процес створення представницької вибірки, що усунуло суб'єктивний вплив на навчання різних класифікаторів для виявлення текстових областей на зображеннях. Показано, що використання запропонованого методу формування текстових областей на зображеннях на основі використання ймовірнісної моделі представлення символу тексту при формуванні навчальної вибірки за рахунок автоматизації процесу створення навчальної вибірки суттєво зменшило час її формування (у середньому на два порядки), що підвищило оперативність виявлення та локалізації текстових областей на зображеннях.

3. Запропоновано кореляційно-екстремальний метод виявлення текстових областей на зображеннях на основі двовимірного дискретного вейвлет-перетворення. Це дозволило при збереженні високої достовірності виявлення ТО підвищити оперативність локалізації та зменшити залежність результатів виявлення ТО від розмірів і типів шрифтів тексту.

4. Для реалізації запропонованого кореляційно-екстремального методу виявлення текстових областей розроблена структура згорткової нейронної мережі з багатомасштабним представленням зображення на основі двовимірного дискретного вейвлет-перетворення, що дозволило зменшити залежність результатів локалізації текстової області від розмірів і типів шрифтів тексту. Забезпечена якість виявлення текстових областей контрольної вибірки (500 зображень на складному фоні) на 10% вище якості виявлення ТО мережею без використання навчальних наборів на основі створених текстових областей.

5. Удосконалено метод локалізації текстових областей з використанням згорткової нейронної мережі, дискретного вейвлет-перетворення та морфологічних операцій, що дозволило забезпечити локалізацію текстових областей для різних розмірів шрифтів. Використання запропонованого методу локалізації дозволили підвищити якість локалізації на 10,5% порівняно з результатами після кореляційно-екстремального методу виявлення ТО.

6. Розроблено інформаційну технологію локалізації ТО на основі запропонованих методів виявлення та локалізації текстових областей на зображенні. Якість локалізації на зображеннях зі складним фоном телевізійних роликів, фільмів, рекламних банерів за допомогою запропонованої інформаційної технології склала 87,7% (*Precision*), 86,6% (*Recall*), що перевершує результат, отриманий за допомогою програм FineReader11 і FreeOCR.

7. Отримані в роботі наукові результати та програмні засоби використані для моделювання та розробки програмно-алгоритмічного забезпечення в науково-виробничому об'єднанні «Діскрет» (м. Одеса) та впроваджені в навчальний процес кафедри інформаційних систем Одеського національного політехнічного університету.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ:

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Нгуен, Т. Т. К. Обнаружение и распознавание текстов на изображениях сложных графических сцен с помощью сверточной нейронной сети / Т. Т. К. Нгуен // Електротехнічні та комп’ютерні системи. – 2014.– №13(89). – С. 125–130. [*Видання включене до наукометричних баз: Ulrich's Periodicals Directory, eLIBRARY.RU, Google Scholar, Index Copernicus, каталог національної бібліотеки України ім. В.І. Вернадського, РЖ ВІНІТІ*].

2. Николенко, А. А. Обнаружение текстовых областей в видеопоследовательностях / А. А. Николенко, Т. Т. К. Нгуен // Искусственный интеллект. – 2012. – № 4. – С. 227 – 234.

3. Николенко, А. А. Обнаружение текстовых областей и выделение символов на изображениях с неоднородным фоном /А. А. Николенко, А. В. Котляр, Т. Т. К. Нгуен // Праці Одеського політехнічного університету. – 2013. – Вип. 1(40).– С.55–60. [*Видання включене до міжнародних наукометричних баз ULRICHSWEB, FreeFullPD, eLIBRARY*].

4. Николенко, А. А. Локализация текстовых областей на изображениях с использованием сверточной нейронной сети / А. А. Николенко, О. Ю. Бабилунга, Т. Т. К. Нгуен // Вісник Національного технічного університету "Харківський політехнічний інститут". Збірник наукових праць. Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2013. – № 19 (992). – С. 121–127.

5. Антощук, С. Г. Метод локализации текста на изображении с обучением на основе вероятностной модели символов / С. Г. Антощук, А. А. Николенко, Т. Т. К. Нгуен // Вимірювальна та обчислювальна техніка в технологічних процесах. – 2014. – № 3. – С. 79 – 84.

6. Антощук, С. Г. Информационная технология локализации текстовых областей на сложном фоне / С. Г. Антощук, А. А. Николенко, Т. Т. К. Нгуен // Електротехнічні та комп'ютерні системи. – 2014. – № 16 (92). – С. 114-121. [*Видання включене до наукометричних баз: Ulrich's Periodicals Directory, eLIBRARY.RU, Google Scholar, Index Copernicus, каталог національної бібліотеки України ім. В.І. Вернадського, РЖ ВІНІТІ*].

Опубліковані праці апробаційного характеру

7. Антощук, С. Г. Локализация квазипериодических последовательностей на изображениях / С. Г. Антощук, А. А. Николенко, Т. Т. К. Нгуен // – Компьютерные науки и технологии . – №2 – 2009.– С. 35 – 40.

8. Николенко, А. А. Выбор масштаба вейвлет-преобразования при локализации квазипериодических последовательностей / А. А. Николенко, Е. В. Ткаченко, Т. Т. К. Нгуен // Современные информационные и электронные технологии: труды XI Междунар. научно-практ. конф.: (24 – 28 мая 2010 г.). – Одесса, 2010. – Т. 1. – С. 54.

9. Николенко, А. А. Обнаружение текстовых фрагментов на видеоизображениях / А. А. Николенко, Т. Т. К. Нгуен // Современные информационные и электронные технологии: труды XIII Междунар. научно-практ. конф.: (4– 8 июня 2012 г.). – Одесса, 2012. – С. 114.

10. Ніколенко, А. Метод виявлення текстових областей в відеопослідовності з використанням вейвлет-перетворення / [А. Ніколенко, О. Бабілунга, Т. Т. К. Нгуєн, В. Зайковський] / – Оброблення сигналів і зображень та розпізнавання образів: XI Всеукраїнська міжнар. конф., 12 – 15 жовтня 2012 р. – Збірник наукових праць. – Київ, МННЦІТС, 2012.– С. 99 – 102.

11. Николенко, А. А. Обнаружение текстовых областей в видеопоследовательностях / А. А. Николенко, Т. Т. К. Нгуен // Искусственный интеллект. Интеллектуальные системы («ИИ-2012»): материалы Междунар. научно-технической конференции: (1–5 октября 2012 г.) / АР Крым, 2012. – С.115–118.

12. Николенко, А. А. Обнаружение и выделение символов текста на изображениях с неоднородным фоном / А. А. Николенко, А. В. Котляр, Т. Т. К. Нгуен // Современные информационные и электронные технологии: труды XIV Междун. научно-практ. конф.: 27 – 31 мая 2013 г. / Одесса, 2013. – С. 55–57.

13. Николенко, А. А. Локализация текстовых областей на изображениях с

использованием сверточной нейронной сети / А. А. Николенко, О. Ю. Бабилунга, Т. Т. К. Нгуен // Проблемы информатики и моделирования: тезисы тринадцатой междунар. научно-технической конф.: 23 – 29 сентября 2013 г. / Харьков: НТУ "ХПИ", 2013. – С.52.

14. Kotlyar, A. V. Text detection and characters extraction in images with non-uniform background / A. V. Kotlyar, T. T. K. Nguyen // Сучасні інформаційні технології – 2013: матеріали III Міжнар. конф. студентів і молодих науковців: 25–26 квітня 2013 р. / ОНПУ. – Одеса, 2013. – С. 110 – 111.

15. Николенко, А. А. Распознавание текстов на изображениях сложных графических сцен / А. А. Николенко, Т. Т. К. Нгуен, П. В. Гордиенко // Современные информационные и электронные технологии: труды XV Междун. научно-практ. конф.: 26-30 мая 2014 г. / Одесса, 2014. – С. 46–47.

АНОТАЦІЯ

Нгуєн Т. К. Т. Інформаційна технологія локалізації текстових областей на зображеннях зі складним фоном. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – «Інформаційні технології». – Одеський національний політехнічний університет Міністерства освіти і науки України, Одеса, 2015.

Дисертація присвячена підвищенню якості та оперативності локалізації текстових областей на зображеннях зі складним фоном в інформаційних технологіях переробки графічної інформації. У роботі запропонована ймовірнісна модель представлення символу тексту, на основі якої розроблено метод формування текстових областей на зображеннях при формуванні навчальної вибірки.

Запропоновано кореляційно-екстремальний метод виявлення текстових областей на зображеннях та удосконалено метод локалізації текстових областей. Кореляційно-екстремальний метод реалізовано на базі згорткової нейронної мережі, структура якої враховує багатомасштабне представлення зображення на основі двовимірного дискретного вейвлет-перетворення. На основі запропонованих методів розроблено інформаційну технологію локалізації текстових областей, що дозволило підвищити оперативність локалізації та зменшити залежність результатів від розмірів і типів шрифтів тексту на зображеннях зі складним фоном в системах переробки графічної інформації.

Ключові слова: інформаційна технологія, переробка графічної інформації, локалізація текстових областей, ймовірнісна модель, згорткова нейронна мережа, дискретне вейвлет-перетворення.

ABSTRACT

Nguyen T. T. K. Information technology of text areas localization on a complex background. – Manuscript.

The thesis for the degree of candidate of technical sciences, specialty 05.13.06 – Information Technologies. – Odessa National Polytechnic University, Ministry of

education and science of Ukraine, Odessa, 2015.

The dissertation is devoted to improving quality and efficiency of the localization of text areas in images with complex background in information technology processing image information. The paper proposed probabilistic model representing a character of text. The paper proposed probabilistic model representing a character of text. On basis of this model the method of forming text areas on the images in the training set formation was developed.

A correlation-extreme method of detecting text regions in images and improved method of localization of text areas were proposed. A correlation-extreme method was implemented based on convolutional neural network structure which takes into account the representation of multiple image based on two-dimensional discrete wavelet transform. On the basis of the proposed methods developed information technology for localization of text areas, thus improving the efficiency of localization and reduce the dependence of the results on the size and type font text in images with complex background in graphics processing systems.

Keywords: information technology, processing of graphics, text localization areas, probabilistic model, convolutional neural network, discrete wavelet transform.

АННОТАЦИЯ

Нгуен Т. К. Т. Информационная технология локализации текстовых областей на изображениях со сложным фоном. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – «Информационные технологии». – Одесский национальный политехнический университет Министерства образования и науки Украины, Одесса, 2015.

Диссертация посвящена повышению качества и оперативности локализации текстовых областей на изображениях со сложным фоном для повышения эффективности автоматической переработки графической информации.

Анализ существующих информационных технологий, моделей и методов локализации текстовых областей на изображениях показал, что существует противоречие между качеством локализации текстовой области и оперативностью этой процедуры, особенно при наличии сложного фона на обрабатывающих изображениях. Поэтому создание информационной технологии локализации текстовых областей на изображениях со сложным фоном на основе новых моделей и методов, позволяющей устранить указанное противоречие или уменьшить его негативное влияние, является важной научно-технической задачей, которая решается в данной работе.

В работе получила дальнейшее развитие вероятностная модель представления символа текста, позволившая формировать изображение псевдосимволов текста с вероятностными характеристиками, соответствующими реальным символам. На основе вероятностной модели представления символа текста впервые разработан метод формирования текстовых областей на изображениях при формировании обучающей выборки, что позволило уменьшить влияние субъективных факторов и

автоматизировать процесс создания представительной выборки для обучения различных классификаторов для локализации текстовых областей.

Для уменьшения зависимости результатов локализации текстовой области от размеров и типов шрифтов текста получил дальнейшее развитие корреляционно-экстремальный метод обнаружения текстовых областей на изображениях, реализованный на базе сверточной нейронной сети, структура которой учитывает многомасштабное представление изображения на основе двумерного дискретного вейвлет-преобразования. Качество обнаружения текстовых областей контрольной выборки (500 изображений на сложном фоне) с использованием предложенного корреляционно-экстремального метода и его реализации на основе сверточной нейронной сети с ДВП-представлением изображения на основе дискретного вейвлет-преобразования и учебным набором на основе созданных текстовых областей на 10% выше качества обнаружения текстовых областей сетью без использования учебных наборов на основе созданных текстовых областей.

Усовершенствован метод локализации текстовых областей на основе сверточной нейронной сети, дискретного вейвлет-преобразования и морфологических операций, что позволило обеспечить локализацию текстовых областей для разных размеров шрифтов. Использование предложенного метода локализации позволило повысить качество локализации на 10,5% по сравнению с результатами после кореляционно-экстремального метода обнаружения ТО.

Качество локализации изображений со сложным фоном (телевизионных роликов, фильмов, рекламных баннеров) с помощью предложенной информационной технологии составила 87,7% (*Precision*), 86,6% (*Recall*), что превосходит результат, полученный с применением программ FineReader11 и FreeOCR.

На основе предложенных модели и методов разработана информационная технология локализации текстовых областей для автоматической переработки графической информации, что позволило повысить качество и оперативность локализации текстовых областей на изображениях со сложным фоном в системах переработки графической информации.

Полученные в работе научные разработки и программные средства использованы для моделирования и разработки программно-алгоритмического обеспечения в научно-производственном объединении «Дискрет» (г. Одесса) и внедрены в учебный процесс кафедры информационных систем Одесского национального политехнического университета.

Ключевые слова: информационная технология, переработка графической информации, локализация текстовых областей, вероятностная модель, сверточная нейронная сеть, дискретное вейвлет-преобразование.