

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВОЛОДИМИРА ДАЛЯ

СІРЯК РОСТИСЛАВ ВІКТОРОВИЧ



УДК 004.5+004.93

**МОДЕЛІ ТА МЕТОД ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ЛЮДИНО-
МАШИННОЇ ВЗАЄМОДІЇ З ВИКОРИСТАННЯМ ЖЕСТІВ**

05.13.06 – інформаційні технології

Автореферат
дисертації на здобуття наукового ступеню
кандидата технічних наук

Сєвєродонецьк – 2021

Дисертацією є рукопис.

Робота виконана на кафедрі комп'ютерних наук та інженерії Східноукраїнського національного університету імені Володимира Даля Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор
Скарга-Бандурова Інна Сергіївна,
Східноукраїнський національний університет ім. В. Даля,
професор кафедри комп'ютерних наук та інженерії.

Офіційні опоненти:
доктор технічних наук, професор
Філатов Валентин Олександрович,
Харківський національний університет радіоелектроніки,
завідувач кафедри штучного інтелекту;

доктор технічних наук, професор
Леонов Сергій Юрійович,
Національний технічний університет «Харківський
політехнічний інститут»,
професор кафедри обчислювальної техніки та
програмування.

Захист відбудеться «13» травня 2021 р. об 11⁰⁰ годині на засіданні спеціалізованої вченої ради К 29.051.16 у Східноукраїнському національному університеті імені Володимира Даля за адресою: 93400, м. Сєвєродонецьк, вул. Донецька, 43.

З дисертацією можна ознайомитись у науково-технічній бібліотеці Східноукраїнського національного університету імені Володимира Даля за адресою: 93400, м. Сєвєродонецьк, вул. Донецька, 43.

Автореферат розісланий «12» квітня 2021 р.

Вчений секретар
спеціалізованої вченої ради К 29.051.16,
кандидат технічних наук, доцент



С.О. Сафонова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми дослідження. Розпізнавання жестів рук на основі комп'ютерного зору є активною сферою досліджень людино-машинної взаємодії (ЛМВ) оскільки жести є природним засобом спілкування людей між собою, а нещодавно і з пристроями в інтелектуальному середовищі. Інтелектуальні системи розпізнавання жестів відкривають нову еру взаємодії людина-комп'ютер, тенденція ЛМВ рухається до відстеження та розпізнавання жестів рук у режимі реального часу для використання у різноманітних сферах застосування. Усвідомлюючи можливості безконтактної взаємодії, ми знаходимось у дуже захоплюючий час розвитку цих технологій. Численні ініціативи у всьому світі свідчать про надзвичайний прогрес, досягнутий за останні роки. Разом з тим, останні дослідження також висвітлюють різноманітні способи, за допомогою яких можливо підійти до цього проблемного простору. Інтерфейси, засновані на сприйнятті жестів та комп'ютерному зорі, повинні відповідати вимогам взаємодії на основі щоденної діяльності й підтримувати неформальну та неструктуровану інформацію. Останнє означає відсутність потреби мати при собі спеціальне обладнання для реалізації ЛМВ. На даний момент мета полягає не лише в демонстрації можливостей безконтактного контролю за допомогою систем комп'ютерного зору, але й у важливих технологічних та дизайнерських викликах щодо конкретних способів досягнення ЛМВ, починаючи від словникового запасу використовуваних жестів, умов та способів введення жестових команд та закінчуючи конкретними технічними проблемами вибору, проектування та використання механізмів їх розпізнавання. Насьогодні, методи машинного навчання, зокрема глибоке навчання, стали золотим стандартом у сфері комп'ютерного зору.

Питаннями розробки нових моделей і методів машинного навчання для розпізнавання жестів рук займалися S. Alashhab, V. Bheda, C. Cadoz, G. Devineau, J. Donahue, S. Hussain, P. Molchanov, N. Neverova, N. Nishida, L. Pigou, K. Simonyan, N. Tavarı та ін. В напрямках розпізнавання жестів рук, зокрема в сфері розвитку інформаційних технологій моделювання, перекладу та навчання української жестової мови активно працюють провідні вітчизняні вчені О.В. Бармак, М.В. Давидов, Ю.В. Крак, Ю.П. Кондратенко, В.В. Пасічник, І.В. Сергієнко та ін.

За останні кілька десятиліть розроблені різні інструменти і алгоритми для реалізації людино-машинної взаємодії для керування роботами, відеоспостереження, комунікації, ігор та інших областей, де рухи, що здійснюються людиною, можуть бути інтерпретовані як певні команди або елементи комунікаційної взаємодії.

Разом з тим, незважаючи на значні досягнення та кількість розробок в галузі глибокого навчання та розпізнавання жестів рук, все ще існує цілий ряд невирішених завдань. Наявні техніки показують високі показники, але переважно в спеціально створених для цього умовах. У реальному житті продуктивність значно знижується внаслідок різних умов освітлення, шумів зображення, неоднорідності фону, оклюзій і руху камери. Крім цього, в більшості систем не вирішена проблема положення руки

по відношенню до камери: рука повинна перебувати в зумовленому місці і не ухилятися далі заданого кута. Проблемою також залишається визначення початку і кінця жесту в послідовності кадрів. Все це обумовлює необхідність проведення подальших досліджень, спрямованих на розробку комплексних систем, що використовують сильні сторони різних підходів, взаємно перекривають недоліки. Існує широкий вибір алгоритмів оптимізації та обчислювання втрат, однак знаходження оптимального шляху до цільової точки потребує подальшого покращення, що в кінцевому підсумку підвищить ефективність обробки даних та результативність розпізнавання образів у цілому.

Таким чином, *актуальним науковим завданням* є розроблення моделей та методу інформаційної технології обробки і розпізнавання жестів рук у режимі реального часу для покращення людино-машинної взаємодії з використанням жестів.

Зв'язок роботи з науковими програмами, планами, темами. Дослідження, результати яких викладені в дисертації, проводилися у Східноукраїнському національному університеті імені Володимира Даля відповідно до державних програм і планів науково-дослідних робіт (НДР), а також міжнародних проектів і програм:

- НДР “Система управління медичною інформацією” (Східноукраїнський національний університет імені Володимира Даля, 0111U001748, 2011-2015 pp.);
- НДР “Дослідження методів аналізу даних в медицині” (Східноукраїнський національний університет імені Володимира Даля, 0116U008700, 2016-2018 pp.);
- НДР “Інтегрована система віддаленого моніторингу стану здоров'я” (Східноукраїнський національний університет імені Володимира Даля, 0120U102758, 2020-2022 pp.);
- Проект Європейського Союзу ERASMUS+ ALIOT 573818-EPP-1-2016-1-UK-EPPKA2-CBHE-JP “Internet of Things: Emerging Curriculum for Industry and Human Applications”, <https://aliot.eu.org>, (2016-2020 pp.).

Роль автора у зазначених науково-дослідних роботах і проектах, у яких дисертант був безпосереднім виконавцем, полягає в розробленні і реалізації моделей, методів та алгоритмів розпізнавання візуальних образів з використанням неймережових архітектур.

Мета й задачі дослідження. Метою дисертації є покращення характеристик автоматичного розпізнавання статичних та динамічних жестів рук за рахунок розробки та практичного використання моделей і методу інформаційної технології людино-машинної взаємодії з використанням жестів.

Для досягнення поставленої мети розв'язуються такі **задачі**:

- 1) провести аналіз предметної області, вимог і передумов до розробки і використання систем ЛМВ за допомогою жестів;
- 2) розробити методологію проектування і спільного використання візуальних методів розпізнавання жестів рук;
- 3) розробити моделі комп'ютерного зору на основі глибоких нейронних мереж, що забезпечують відстеження та розпізнавання заданих жестів рук, здатних

функціонувати в режимі реального часу, інваріантних до афінних перетворень і зміни освітлення;

- 4) удосконалити модель статичного розпізнавання жестів;
- 5) удосконалити технологію розпізнавання та прогнозування динамічних жестів на основі моделі генерації послідовностей з використанням ConvLSTM;
- 6) створити новий датасет і виконати концептуальну розробку інтуїтивного словникового запасу динамічних жестів, адаптованих до особливостей хірургічного контексту;
- 7) удосконалити технологію взаємодії між хірургом та системою перегляду зображень в операційній залі яка використовує функції просторового домену для відстеження потенційних напрямків руху рук за допомогою скінченного автомату;
- 8) розробити елементи інформаційної технології для проектування системи ЛМВ з використанням жестів;
- 9) виконати тестування та практичне впровадження розроблених моделей, засобів і технологій.

Об'єкт дослідження – процеси людино-машинної взаємодії, що здійснюють ідентифікацію, відстеження та класифікацію жестів рук або інших об'єктів на зображенні та у відеопослідовності.

Предмет дослідження – моделі та метод інформаційної технології ідентифікації та класифікації статичних і динамічних жестів рук для вирішення задач людино-машинної взаємодії у реальному часу.

Методи дослідження. Проведені в роботі дослідження ґрунтуються на наступних методах і технологічних підходах:

- для аналізу предметної області та вимог до розробки і використання систем ЛМВ за допомогою жестів використано пошук в академічних БД, патентний пошук, технології окреслення, коментування, конспектування, схематизація, анатомування, конденсаування, аналіз, синтез, агрегування;
- при розробленні методології проектування і спільного використання візуальних методів розпізнавання жестів рук використано методи загальної теорії систем, системного аналізу, методи структурного синтезу, аналізу та моделювання процесів;
- при розробленні моделей комп'ютерного зору на основі глибоких нейронних мереж застосовано методи машинного навчання, технології збільшення даних (data augmentation), технології попередньої обробки даних (очищення, фільтрація розмиття за Гауссом, перетворення кольорів, адаптивна порогова обробка, алгоритм Otsu, детектор контурів Canny, для відділення зон зображення, які містять обличчя та долоні, використано модель дифузного освітлення, для визначення форми долоні використані моделі деформації зображень, порівняння зображень із спотвореннями)
- при розробленні технології розпізнавання та прогнозування динамічних жестів використано моделі генерації послідовностей з використанням ConvLSTM, для навчання нейромережевого класифікатора використано методи зворотного поширення похибки та спряжених градієнтів, для стеження за долонями використані методи статичної та динамічної кластеризації;
- при розробленні технології взаємодії між хірургом та системою перегляду зображень в операційній залі використано методи математичної статистики, функції

просторового домену, теорія автоматів, теорія подібності.

Наукова новизна результатів досліджень полягає у наступному.

1) *вперше розроблено* методологію розробки і спільного використання візуальних методів розпізнавання жестів рук, яка дозволяє створювати і досліджувати моделі глибокого навчання для розпізнавання статичних та динамічних жестів, здатних працювати в режимі реального часу і забезпечує розуміння способів їх налаштування для різних інтерфейсів управління жестами та потенційних застосувань в системах ЛМВ;

2) *удосконалено* модель статичного розпізнавання жестів, побудовану згідно запропонованої методології на основі згорткової нейронної мережі, шляхом штучного збільшення даних і використання контурів. Завдяки використанню контурів, модель є стійкою до відносно широких кутів обертання рук і незалежною від освітлення. Модель із доповненими даними досягла точності 97,12%, що майже на 4% перевищує модель без доповнень (92,87%), при цьому, для ефективної роботи достатньо стандартної веб-камери.

3) *дістала подальшого розвитку* технологія розпізнавання та прогнозування жестів на основі моделі генерації послідовностей з використанням ConvLSTM2D і Conv3D. Для подальшого розширення можливостей ConvLSTM для вирішення проблеми прогнозування жестів було виконано наступні модифікації: (1) на етапі попередньої обробки введено виявлення контура, реалізованого у вигляді фільтру. Фільтрація на вхідному рівні дозволила фіксувати часові залежності та зменшити шум вхідного сигналу; (2) проведена заміна 3 шарів batch normalization на 3 шари dropout regularization що дозволило знизити перенавчання та мінімізувати помилки передбачення при прогнозуванні жестів. Отримана точність склала 90%, що на 30% краще ніж у моделі ConvLSTM з batch normalization (60%).

4) *удосконалено* модель скінченного автомату для безконтактного управління переглядом медичних зображень за допомогою жестів яка, на відміну від існуючих, використовує дані прогнозованих кадрів відеопослідовностей, що дозволяє зменшити час відгуку системи;

5) Для тестування запропонованих моделей і методу інформаційної технології розв'язування задач розпізнавання та прогнозування жестів в операційній залі створено *новий датасет* і виконано концептуальну розробку інтуїтивного словникового запасу динамічних жестів, що дозволяє реалізувати ефективну безконтактну інтерактивну систему, адаптовану до особливостей хірургічного контексту;

6) *удосконалено* структурну модель інформаційної технології ЛМВ з використанням жестів, за рахунок визначення основних етапів та інформаційних потоків створення та інтеграції моделей глибокого навчання, яка забезпечує прийняття рішень щодо застосування розроблених методів, засобів і технологій.

Практичне значення одержаних результатів полягає в тому, що розроблені методологія, моделі та метод створюють концептуальну основу для розробки інформаційної технології людино-машинної взаємодії з використанням жестів в режимі реального часу. Основні наукові положення дисертації реалізовано в вигляді розрахункових моделей, інженерних алгоритмів, а також відповідних програмних засобів, які утворюють прикладну інформаційну технологію ідентифікації та

класифікації статичних і динамічних жестів рук на основі глибоких нейронних мереж.

В результаті виконаного дисертаційного дослідження розроблено програмні реалізації моделей системи розпізнавання, яка демонструє високу точність і швидкодію з низкою чутливістю до умов освітлення. Результати дисертаційної роботи використовуються в якості базових елементів системи ЛМВ для навігації та перегляду медичних зображень в операційній залі.

Реалізація результатів та впровадження. Результати дисертаційної роботи впроваджено

– у Східноукраїнському національному університеті імені Володимира Даля при розробці навчальних курсів “Вступ до інтерактивного проектування”, “Машинне навчання” на кафедрі комп’ютерних наук та інженерії, виконанні міжнародного проекту ERASMUS+ ALIOT 573818-EPP-1-2016-1-UK-EPPKA2-CBHE-JP “Internet of Things: Emerging Curriculum for Industry and Human Applications” (акт впровадження від 04.01.2021 р.);

– у медичному центрі “Твоя лікарня” при розробці проекту клінічної системи безконтактного перегляду медичних зображень, м. Сєвєродонецьк (акт впровадження від 21.11.2019 р.);

Особистий внесок здобувача полягає в розробці нових і удосконаленні існуючих моделей та інструментальних засобів, які забезпечують вирішення поставлених у дисертації завдань. Всі положення, які виносяться на захист, належать автору особисто. Роботи [7, 15] опубліковані без співавторів. У роботах, опублікованих у співавторстві, здобувачеві належать такі результати: результати аналізу вимог та моделювання систем нейромережевої архітектури [2, 6, 14], розроблені інструментальні засоби для автоматизації процесів автоматичного розпізнавання статичних та динамічних жестів рук [3, 9], аналіз критеріїв оптимальності для якісної оцінки вихідних даних моделі [4, 11, 12], модель статичного розпізнавання жестів [7, 16], технологія розпізнавання та прогнозування жестів на основі моделі генерації послідовностей з використанням ConvLSTM2D і Conv3D [1], моделі розпізнавання динамічних жестів за допомогою мережі CNN-LSTM [8, 10, 13].

Апробація результатів наукових досліджень проводилася на 10 міжнародних і національних науково-практичних конференціях, а саме: 10th IEEE International Conference on Dependable Systems, Services and Technologies (DESSERT), UK, Leeds, 2019; 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Metz, France, 2019; II International Conference on Theoretical and Applied Computer Science and Information Technology (TACSIT), Сєвєродонецьк, 2019; XI Всеукраїнській науково практичній WEB конференції аспірантів, студентів та молодих вчених “Комп’ютерні інтелектуальні системи та мережі”, Кривий Ріг, 2019; XIX міжнародній конференції “Проблеми інформатики і моделювання”, Харків, 2019; “Сучасні технології в освіті та науці”, 2018; “ІТ-Ідея”, Сєвєродонецьк, 2018; XXVI міжнародній науково-

практичній конференції “Інформаційні технології: наука, техніка, технологія, освіта, здоров’я” MicroCAD-2018, Харків, 2018; міжнародній конференції студентів, аспірантів та молодих вчених “Технологія-2012”, Сєверодонецьк, 2012.

Публікації. За темою дисертації з викладенням її основних результатів опубліковано 16 наукових праць, серед яких 6 статей у наукових фахових виданнях України [2-7], 1 стаття у фаховому виданні, індексованому у міжнародній наукометричній базі Scopus [1]; 2 публікації в працях міжнародних конференцій, які включено до бази даних Scopus [8,9], 7 публікацій за матеріалами наукових конференцій [10-16].

Структура та обсяг дисертації. Дисертація складається з анотацій, вступу, 5 розділів, висновків, додатків та списку використаних джерел з 125 найменувань на 16 сторінках. Загальний обсяг дисертації становить 172 сторінки, з них 148 сторінок основного тексту, 40 рисунків, 5 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Вступ дисертації надає огляд предметної області, окреслює основні положення роботи й містить обґрунтування актуальності теми й наукових задач; мету й задачі дослідження; об’єкт, предмет і методи дослідження; характеристику наукової новизни й практичного значення результатів, а також особистого внеску здобувача; дані щодо реалізації, апробації та публікації результатів.

У першому розділі проведено дослідження сучасного стану проблеми, проаналізовано методи та підходи до розроблення систем розпізнавання статичних і динамічних жестів рук. Розглянуто стан розвитку інформаційних технологій в цій сфері та області її застосування. Показані та докладно описані етапи, створення системи розпізнавання жестів для ЛМВ, а саме: збір даних, моделювання жесту, витяг ознак, класифікація, формування інструкцій. Встановлено, що незважаючи на те, що жест є природним способом спілкування, при використанні жестів для ЛМВ необхідно враховувати існуючі обмеження технології та людини. Так, наприклад, для безпосередньої взаємодії людина-робот в спільному просторі використання динамічних жестів заборонено, оскільки на виконання динамічного жесту зазвичай потрібно понад півсекунди, що стоїть на заваді реакції системи взаємодії. Що стосується жестів для дистанційного зчитування з екрана, навпаки, динамічний жест дозволяє досягти необхідного рівня взаємодії з мінімальною затримкою, неможливого при застосуванні лише статичних жестів. Для побудови системи з динамічним розпізнаванням образів є необхідність в поєднанні двох мережевих архітектур – згорткових нейронних мереж (CNN) та мереж з довгою короткостроковою пам’яттю (LSTM).

Проведено аналіз нормативної бази та вимог до систем людино-машинної взаємодії із застосуванням жестів. Сформульовані задачі досліджень, пов’язаних з аналізом та обробкою зображень, приведенням даних до вигляду, який був би

найбільш прийнятним для подальшої роботи з урахуванням особливостей використовуваних архітектур нейронної мережі. Обґрунтована методика і математичний апарат досліджень.

У другому розділі пропонується методологія розробки моделей розпізнавання статичних та динамічних жестів рук, побудованих на основі CNN та CNN+LSTM архітектур.

Завдання класифікації статичного жесту сформульовано у вигляді задачі пошуку приналежності зображення X_t до певного класу жестів S_i , для якого розподіл ймовірності $P(X_t, S_i)$ має максимальне значення $\forall i$:

$$P(X_i, S_i) \rightarrow \max. \quad (1)$$

Для системи ЛМВ, яка отримує серію розпізнаних жестів $\mathcal{G} = \{g_1, g_2, \dots, g_n\}$ від CNN класифікатора, де g_i - це i -й жест, розпізнаний системою, дискримінантна функція для виконання команди k_i , яка відповідає жесту g_i , може бути представлена у вигляді ймовірнісної моделі (2):

$$f_i(\mathcal{G}) = Pr(k_i|\mathcal{G}). \quad (2)$$

У загальному випадку, для знаходження f_i можна використовувати різні підходи - моделі Маркова, оцінку ризику Байєса та ін., але, оскільки, в реальному середовищі, статистичні дані про ризик і частоту появи кожного жесту та команди можуть бути недоступні, дискримінантну функцію (2) перетворено до моделі (3):

$$f_i(\mathcal{G}) = \frac{\sum_j w_{i,j}}{|\mathcal{G}|}, \quad (3)$$

де

$$w_{i,j} = \begin{cases} 1, & g_j = k_i \\ 0, & g_j \neq k_i \end{cases} \quad (4)$$

де $j = 1, \dots, n$ та $|\cdot|$ є операцією потужності.

Модель (3) використовує поняття “період відповіді”, що означає певну затримку – система ЛМВ не виводить відповідь одразу після того, як отримує жест, розпізнаний класифікатором, а виконує команду відповідно до дискримінантної функції, яка обчислюється за допомогою послідовних жестів, що з'являються протягом певного періоду. На відміну від (3), задача знаходження дискримінантної функції у вигляді (4) здатна ефективно покращити надійність системи ЛМВ, але, з іншого боку, може призвести до затримки реакції системи. Спосіб вирішення цієї проблеми полягає в тому, що систему налаштовують продовжувати виконувати “затриману” дію, як тільки дія буде підтверджена, і поки нова дія не буде підтверджена схемою відповідей, обговореною вище.

Завдання класифікації динамічного жесту сформульовано аналогічно задачі класифікації статичного жесту, але з урахуванням змін у часі.

Для динамічного жесту d_i з набору класів динамічних жестів $\mathbb{C} = \{C_l\}_{l=1}^N$, де N визначає кількість класифікованих жестів, зміни d_i в часі визначено як:

$$D_i = \{\mathfrak{I}_i^t\}_{t=1}^{T_i}, \quad (5)$$

де $t \in [1, T_i]$ визначає певний момент у часовому вікні розміром T_i , а $\mathfrak{I}_i^t$ являє собою відповідний кадр d_i в момент часу t .

Типи жестів визначаються на основі конкретного набору даних. Дискримінантна функція для виконання команди k_i така сама як і у випадку роботи зі статичними жєстами, тобто (2) або (3).

Виходячи зі спільності постановок задач класифікації статичних і динамічних жестів, у розділі запропоновано нову методологію розробки і спільного використання візуальних методів розпізнавання жестів рук. Пропонована методологія дозволяє створювати та досліджувати моделі для розпізнавання статичних і динамічних жестів і складається з наступних етапів (рис. 1): отримання наборів даних; збільшення даних (на випадок використання власного відео та/або зображень); підготовка та попередня обробка даних; навчання класифікаторів і вилучення ознак за допомогою глибокої нейронної мережі класу CNN і, безпосередньо, розпізнавання жестів, яке може виконуватися за допомогою мереж CNN, RNN та їх комбінацій.

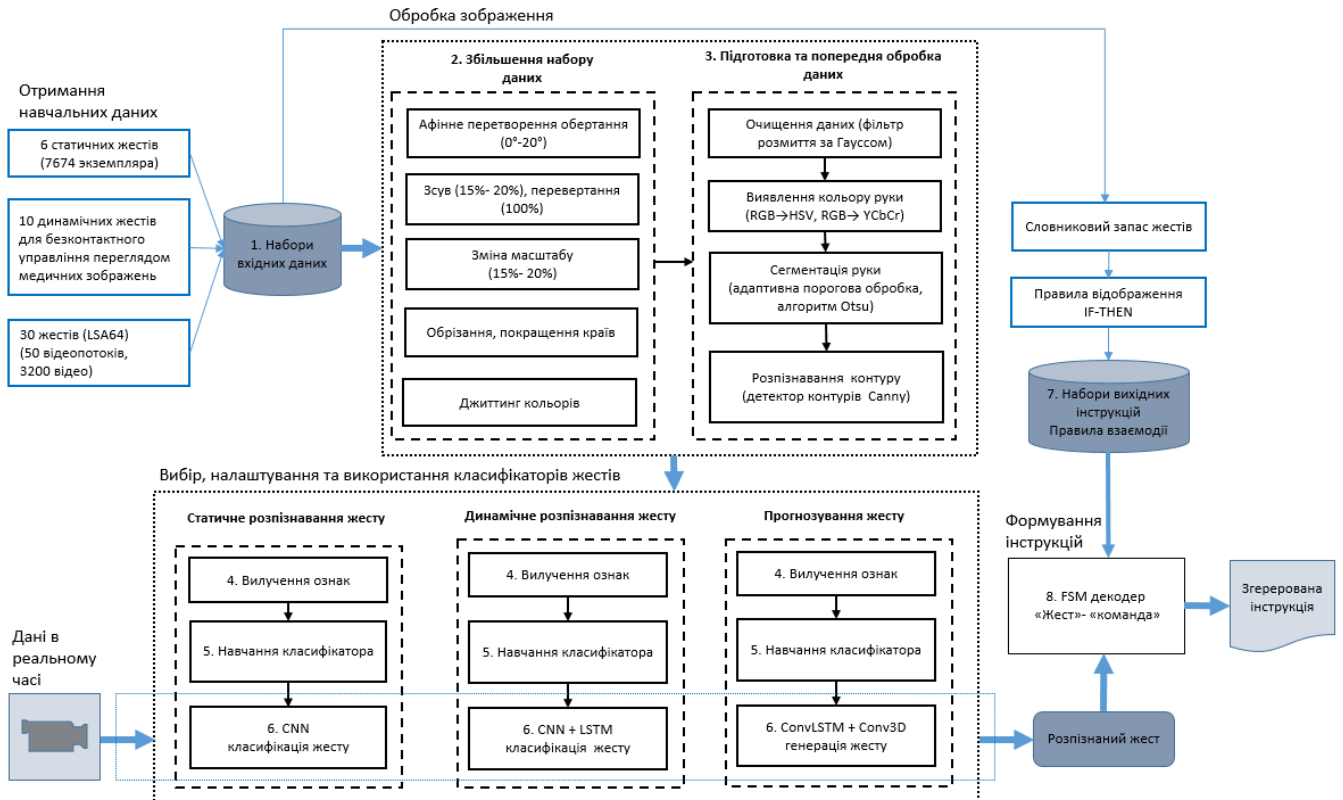


Рисунок 1 – Методологія розробки і спільного використання візуальних методів розпізнавання жестів рук

Далі в розділі розглянуто основні етапи розробки кожної з моделей розпізнавання жестів, процедури збільшення набору даних, очищення даних, зміни колірної моделі, сегментації та виділення контуру під час попередньої обробки, що сприяє кращому вилученню ознак, застосування операції згортки квадратною матрицею з непарною розмірністю. Розкрито математичний апарат формування ознак та класифікації у згортковій нейронній мережі. Основна увага приділяється операціям формування ознак, притаманних тому чи іншому жесту, що здійснює згорткова мережа. Аналізується вибір гіперпараметрів, оптимізація нейромережі та класифікація зображень.

Для вирішення задачі (1), в дисертаційній роботі використовується CNN яка дозволяє визначити ознаки об'єктів і ймовірності приналежності об'єкта (зображення X_i) до того чи іншого класу жестів S_i . Модель CNN для статичного розпізнавання жестів, була побудована згідно пропонованої методології, з використанням штучного збільшення даних і повного циклу процедур підготовки та попередньої обробки даних для виділення контурів. Модель пропонованої нейронної мережі надано на рис. 2.

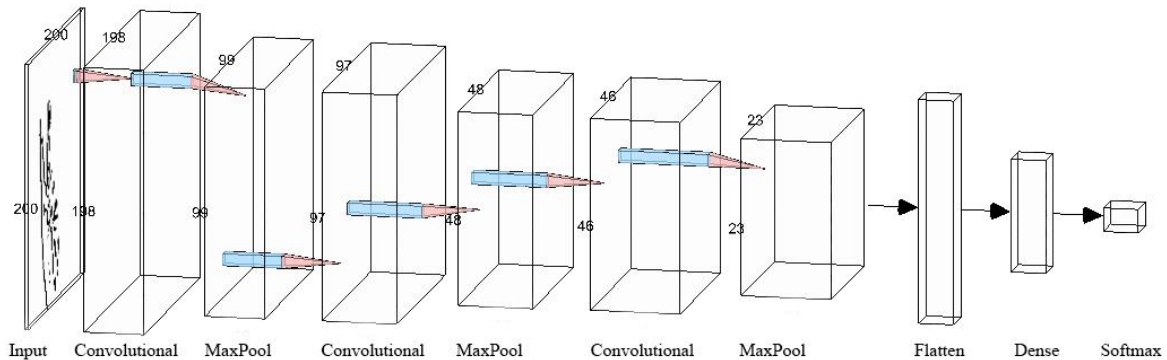


Рисунок 2 – Архітектура CNN для розпізнавання статичних жестів

Розроблена мережа має три згорткових шари з функцією активації ReLU, за кожним з яких розташовується субдискретизаційний шар максимізаційного агрегування max pooling. Після чого, дані в згладжувальному шарі (flatten) перетворюються з 2D-представлення в одновимірний вектор, і в кінці проходять через два повноз'єднаних шари. Другий повноз'єднаний шар з функцією softmax, що перетворює вектор дійсних чисел в вектор ймовірностей, є вихідним і містить softmax-класифікатор з трьома класами.

Мережа навчалася з використанням функції втрат категорійної крос-ентропії на основі алгоритму зворотного поширення помилки. При виборі оптимізатора були випробувані два адаптивних методи: Adadelata і Adam. Проведені експерименти показали високу ефективність Adadelata, однак, при порівнянні з Adam навчання тривало більш ніж в два рази довше.

Завдяки використанню контурів, отримано модель, яка є стійкою до відносно широких кутів обертання рук і незалежною від освітлення. Модель із доповненими даними досягла точності 97,12%, що майже на 4% перевищує модель без доповнень (92,87%), при цьому, для ефективної роботи достатньо стандартної веб-камери. Враховуючи той факт, що дослідження проводилось на двоядерному i3-7100

процесорі з використанням простої веб-камери, отримано високий ступінь вірних прогнозів. Серед недоліків моделі варто відзначити, що вона підходить лише до статичних жестів, без виявлення та відстеження рук або випадків оклюзії.

Для розпізнавання динамічних жестів, CNN доповнюється мережею LSTM. Перший компонент системи, CNN, призначений для вилучення ознак із зображення суб'єкта, який виконує жест рукою. Другий компонент дозволяє виконувати вилучення часових ознак і, тим самим проводити динамічне розпізнавання жестів яке залежить від упорядкованої послідовності зображень. Загальна архітектура мережі CNN-LSTM наведена на рис. 3.

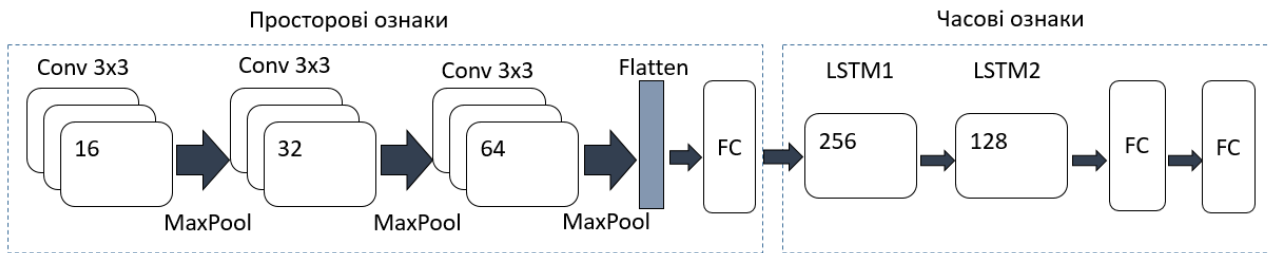


Рисунок 3 - Архітектура нейронної мережі CNN-LSTM для динамічного розпізнавання жестів

Частина мережі, що відповідає за виявлення просторових ознак, містить три згорткові шари (CNN), за якими йдуть шари агрегування (MaxPool), згладжувальний (flatten), шари виключення (Dropout), і один вихідний повноз'єднаний шар (FC). Вхідними даними у LSTM частину є просторові ознаки, виявлені на попередніх згорткових шарах, сплосчених в одновимірний вектор. Шари LSTM з 256-а та 128-ю вентильними вузлами визначають часові патерни на 60 часових кроках.

В результаті експериментів, на тестовій підмножині розміром в 1535 зображень була досягнута точність розпізнавання 94%. В ході тестування з розпізнаванням жестів на веб-камеру в режимі реального часу була отримана висока ступінь вірних відповідей мережі, в тому числі, на людях, які не брали участі у створенні бази даних зображень. За результатами тестування зроблено висновок, що запропонована методологія дозволяє створювати моделі для розпізнавання двох типів жестів і забезпечує розуміння способів їх налаштування для різних інтерфейсів управління жестами та потенційних застосувань в системах ЛМВ.

Третій розділ дисертації присвячений розробленню і дослідженню технологій розпізнавання та прогнозування жестів на основі моделі генерації послідовностей жестів з використанням ConvLSTM. На відміну від розпізнавання, модель прогнозування жестів у реальному часі дозволяє визначити намір руху руки та передбачати кінцевий жест до кінця руху руки. Модель прогнозування жестів може бути використана для оцінки жестів людини до фактичної дії, щоб зменшити затримки в інтерактивних системах ЛМВ.

У цій частині роботи ставилась задача навчити модель прогнозувати наступні кадри по набору вже переглянутих протягом виділеного часу. Часовим інтервалом виступає частота кадрів у відеопослідовності довжиною t . Ознаки кадрів з 1 до $t-1$ використовуються для прогнозування ознак на наступному кадрі t .

Задача прогнозування жесту формулюється наступним чином. Нехай відома множина кадрів S , представлених сіткою розміром $M \times N$, що складається з M рядків і N стовбців. Всередині кожної клітинки сітки є P вимірювань, які змінюються з часом. X – тензор ознак кадру $X \in \mathbb{R}^{P \times M \times N}$, де $\mathbb{R}$ означає простір спостережуваних ознак. $T \times S$ – просторово-часові послідовності, де $S = \{s_1, s_2, \dots, s_i\}$ – просторова сітка, та $T = \{t_1, t_2, \dots, t_i\}$ – тривалість послідовності навчальних кадрів. C – тензор прогнозування ознак кадру C , де $C(s, t)$ – ознаки кадру в сітці $s \in S$ в часі $t \in T$. X – список тензорів ознак $X = \{x_1, x_2, \dots, x_T\}$, де x_i являє собою тривимірний тензор, що записує відповідний атрибут в кожному сітку s_i на кожному кадрі. Для набору даних D знайти модель, що прогнозує $C(S, t)$, для кадрів $t \in T_{test}$ таку, що дозволяє знизити перенавчання та мінімізувати помилку прогнозування за умов, коли: 1) залежність між C і X змінюється в просторі, 2) $X(S, t_j)$ недоступний для прогнозування $C(S, t_i)$, $t_i \in T_{test}$.

Основою структури мережі для прогнозування жестів є ConvLSTM з 8 шарами. Її завдання полягає в обробці кожного зображення, зробленого з камери, зі швидкістю 10 кадрів в секунду для отримання значущих функцій. Крім того, вона являє собою одну з небагатьох структур, здатних масштабуватися відповідно до даних, тобто її глибину можна легко збільшити або зменшити залежно від складності сцени, без небезпеки перенавчання моделі. Розмір ядра (3, 3) підтримується на всіх шарах для поліпшення виявлення ознак у різних масштабах.

Для подальшого удосконалення можливостей ConvLSTM для вирішення задачі прогнозування жестів в моделі було виконано наступні модифікації:

- на етапі попередньої обробки введено виявлення контуру, реалізованого у вигляді фільтру, який дозволяє виділити контури руки. Фільтрація на входному рівні дозволяє фіксувати часові залежності та зменшувати шум вхідного сигналу;
- проведена заміна 3 шарів Batch Normalization на 3 шари Dropout regularization.

Структура ConvLSTM складається з чотирьох шарів ConvLSTM2D, одного шару Conv3D та трьох шарів регуляризації виключенням. Архітектуру запропонованої просторово-часової ConvLSTM надано на рис.4. Модель формується шляхом багаторазового накладення однієї нейронної мережі поверх іншої. Це означає, що вихідні дані одного шару є вхідними даними для наступного шару. Мережа ConvLSTM, розширює повністю підключений рівень LSTM, щоб використовувати згорткові структури як при переході між входами і станами, так і між станами, що надає можливість одночасно вивчати просторові і часові відносини.

Для вирішення завдання прогнозування просторово-часової послідовності використовується структура, яка складається з двох мереж, мережі кодування і мережі прогнозування. Початкові стани і вихідні дані ділянок мережі прогнозування копіюються з останнього стану мережі кодування. Обидві мережі формуються шляхом об'єднання кількох шарів ConvLSTM. До першої згорткової частини нейронної мережі застосовувалася оболонка TimeDistributed перед тим, як дані будуть подані на вхід LSTM. Ця обгортка дозволяє застосовувати обгорнені шари до кожного тимчасового фрагменту вхідних значень. Оскільки прогнозовані дані мають таку саму розмірність, що і вхідні дані, всі стани в мережі прогнозування об'єднуються і подаються в згортковий шар 1×1 для генерації остаточного прогнозу.

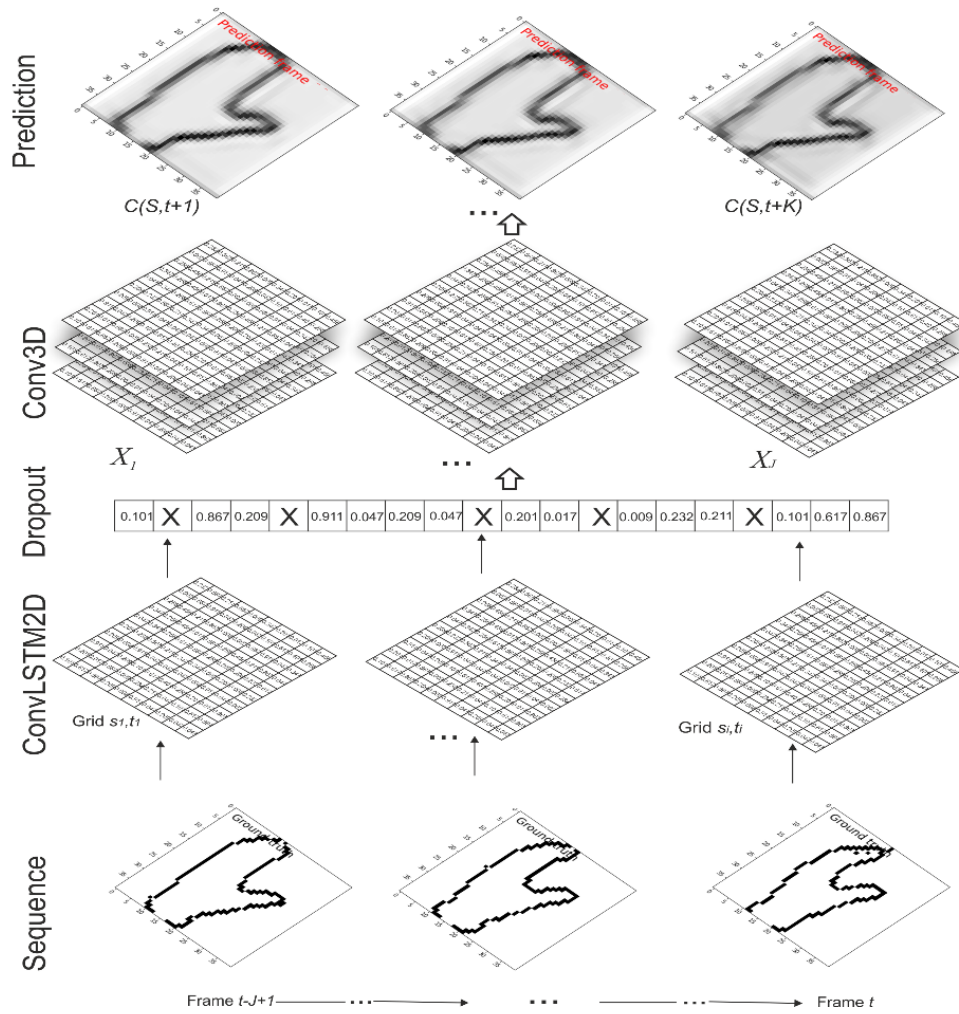


Рисунок 4 - Архітектура запропонованої мережі

Кодуюча LSTM стискає всю вхідну послідовність в тензор прихованого стану, і LSTM прогнозування розкриває цей прихований стан, щоб дати остаточний прогноз.

$$\begin{aligned}
 \tilde{X}_{t+1}, \dots, \tilde{X}_{t+K} &= \arg \max_{X_{t+1}, \dots, X_{t+K}} p(X_{t+1}, \dots, X_{t+K} \mid \hat{X}_{t-J+1}, \hat{X}_{t-J+2}, \dots, \hat{X}_t) \\
 &\approx \arg \max_{X_{t+1}, \dots, X_{t+K}} p(X_{t+1}, \dots, X_{t+K} \mid f_{\text{encoding}}(\hat{X}_{t-J+1}, \hat{X}_{t-J+2}, \dots, \hat{X}_t)) \\
 &\approx g_{\text{encoding}}(f_{\text{encoding}}(\hat{X}_{t-J+1}, \hat{X}_{t-J+2}, \dots, \hat{X}_t))
 \end{aligned}$$

Всі вхідні і вихідні елементи є тривимірними тензорами, які зберігають всю просторову інформацію. Оцінка моделі прогнозування проводилася з використанням перехресної бінарної ентропії, точності, Евклідової, манхеттенської відстаней, нормалізованої перехресної кореляції.

Результати тестування моделі показали, що фільтрація на вхідному рівні дозволяє фіксувати часові залежності та зменшувати шум вхідного сигналу; проведена заміна 3 шарів batch normalization на 3 шари dropout regularization дозволила знизити перенавчання та мінімізувати помилки передбачення при

прогнозуванні жестів. Отримана точність склала 90%, що на 30% краще ніж у базовій моделі ConvLSTM з batch normalization (60%).

У четвертому розділі представлено новий набір даних динамічних жестів і визначено лексикон для навігації медичних зображень в операційній залі. Розроблено технологію взаємодії між хірургом та системою перегляду зображень в операційній залі на основі скінченного автомату (FSM). Надано опис процесу розробки FSM для системи ЛМВ з урахуванням контекстної інформації - результатами моделювання розпізнавання та прогнозування жестів, які представлені інформацією про схожість прояву того чи іншого жесту.

Метою нового датасету було створити просту, виразну структуру, яку хірурги можуть легко інтерпретувати та прийняти для взаємодії зі спеціалізованим безконтактним хірургічним монітором в ОЗ без запам'ятовування складних мовних правил. Задля визначення найбільш інтуїтивних та легких у виконання жестів, хірургам був запропонований базовий набір жестів, за допомогою якого пропонувалося дати опис найпоширеніших дій, які вони виконують з медичними зображеннями під час операцій для навігації зображень, зміни яскравості, орієнтації та управління масштабом тощо. Всі жести оцінювалися за чотирибальною шкалою (0 – «категорично не прийнятне», 1 – «не прийнятне», 2 – «нейтрально», 3 – «прийнятне») за наступними критеріями: інтуїтивність (жест інтуїтивно зрозумілий і відповідає виконуваний команді); простота використання (жест легко і зручно виконувати); легкість запам'ятовування. Після аналізу результатів було створено власний набір даних який містить 10 видів динамічних жестів рук для безконтактного управління переглядом медичних зображень. Розпізнаючи ці жести рук, програмний засіб дозволяє реалізувати безконтактний перегляд необхідних медичних зображень за умов збереження стерильності рук лікаря.

Для ефективного перетворення жестів в інструкції в роботі запропонована детермінована модель на основі FSM. Напрямок руху рук хірурга моделюється як послідовність кадрів зображення. Ознаки прогнозованих кадрів відеопослідовностей зображень, що отримуються з використанням моделей розпізнавання та прогнозування кадрів відеопослідовностей жестів, використовуються в якості вхідних даних для моделі FSM. Взаємодія з системою безконтактного управління переглядом медичних зображень в ОЗ відбувається в режимі реального часу з урахуванням даних, що отримуються в поточний момент часу, та на підставі яких генеруються прогнозовані дані. Для дослідження даних, що отримуються в поточний момент часу використовується контекстна інформація, що зберігається в FSM.

Контекстною інформацією є результати моделювання розпізнавання та прогнозування жестів, представлені інформацією про схожість прояву того чи іншого жесту. Міра схожості жесту отримується з використанням прогнозованих ознак кадру.

Для вимірювання подібності між двома ключовими кадрами використана міра відстані Хаусдорфа, яка визначається як максимум між двома наборами точок O і I :

$$H(O, I) = \max \{h(O, I), h(I, O)\}, \quad (6)$$

де $h(O, I) = \max_{o \in O} \min_{i \in I} \|o - i\|$ і $h(I, O) = \max_{i \in I} \min_{o \in O} \|i - o\|$.

Точки ознак позначаються $o_1, \dots, o_m$ для множини O та $i_1, \dots, i_n$ для множини I .

Обчислення відстані Хаусдорфа виконується за допомогою алгоритму перетворення відстані.

Схему взаємодії розроблених моделей розпізнавання та прогнозування відеопослідовностей зображень жестів та моделі скінченного автомату представлено на рис. 5.

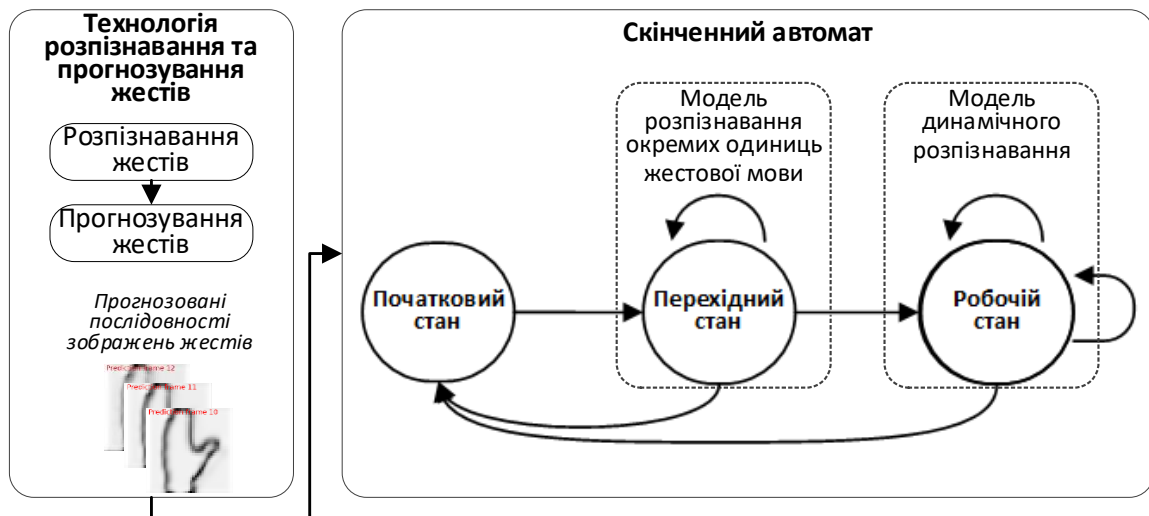


Рисунок 5 – Схема взаємодії розроблених компонент ЛМВ

Дані відео жестів поступають на вхід моделі прогнозування. Дані проходять всі описані вище етапи та прогнозуються послідовності зображень жестів, які є вхідними даними для скінченного автомату. Для запобігання помилкових спрацьовувань на перехідному стані моделі скінченного автомату використовується модель розпізнавання окремих одиниць жестової мови, а на робочому стані використовується модель динамічного розпізнавання жестів, розробка яких представлена в Розділі 2.

У п'ятому розділі представлено елементи інформаційної технології людино-машинної взаємодії в умовах операційної зали з використанням жестів. Розкрито етапи розробки та реалізації інформаційної технології у вигляді системи безконтактного перегляду зображень, представлено процес взаємодії з системою в режимі реального часу, наведено результати реалізації та оцінки ефективності запропонованих в роботі моделей і елементів інформаційної технології.

Структура інформаційної технології ЛМВ з використанням жестів складається з наступних процесів (рис.6): визначення мети ЛМВ (управління переглядом зображень, взаємодія з роботом, віртуальна реальність), інтерфейсів управління жестами (веб-камера, Microsoft Kinect, рукавички); об'єктів взаємодії (робот, монітор), типів взаємодії (одностороння - монітори та дисплеї, двостороння - робот), словникового запасу жестів; формування вимог ТЗ до ЛМВ на основі жестів; визначення базових компонент ЛМВ на основі жестів, важливих для конкретного

застосування (розпізнавання статичних жестів, розпізнавання динамічних жестів, прогнозування жестів); визначення моделей і нейромережових архітектур, використовуваних для розпізнавання жестів; визначення правил взаємодії ЛМВ на основі жестів; визначення стратегії спільного використання моделей розпізнавання жестів, та їх налаштування для відповідних інтерфейсів управління жестами та потенційних застосувань в системах ЛМВ.

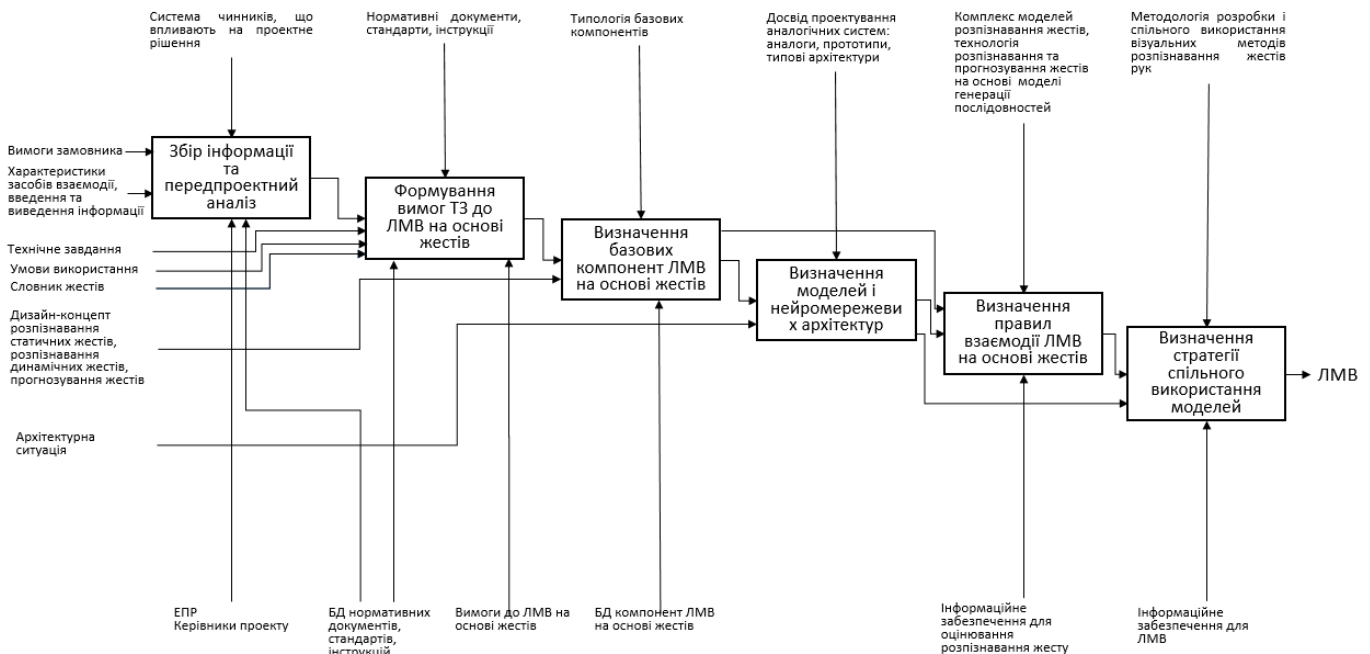


Рисунок 6 - Схема розробки інформаційної технології ЛМВ з використанням жестів

В результаті реалізації запропонованої інформаційної технології було створено систему ЛМВ, яка пройшла тестування в наступних умовах. Початок жесту, який використовувався та на підставі якого здійснювалося прогнозування, складав 10 кадрів. На підставі цих 10 кадрів отримуються прогнозовані наступні кадри, які поступають на вхід моделі FSM. На підставі записаного набору жестів, визначено середню тривалість жесту, яка склала 1,5 секунд. Відео має наступні характеристики: розподільна здатність дорівнює 720x720, 30 кадрів на секунду.

Без використання прогнозування (використовувалась середня тривалість відео одного жесту) час розпізнавання та відповідної реакції системи в режимі реального часу залежить від тривалості жесту, та в середньому, при ідеальних умовах розпізнавання, мав затримку в 1,5 секунди, що дорівнює 45 кадрам.

З використанням методу прогнозування час розпізнавання та відповідної реакції системи в режимі реального часу не залежить від тривалості жесту, та, в середньому, при ідеальних умовах розпізнавання, мав затримку в 0,75 секунди, або 25 кадрів.

Таким чином, удосконалена технологія безконтактної взаємодії для управління медичними зображеннями з використанням жестів рук дозволила зменшити затримку відклику системи на 0,75 секунди, що в 2 рази краще за час відклику без використання запропонованої інформаційної технології. Разом з тим, варто зазначити, в режимі реального часу система повинна мати можливість завершити конвеєр обробки не

більше ніж 250 мілісекунд, оскільки середній час реакції на зорові подразники коливається від 250 до 350 мілісекунд. Більший діапазон призводить до візуально помітного відставання. Отже, в майбутньому необхідні додаткові експерименти для отримання найкращого результату, що гарантуватиме ЛМВ без помітного відставання.

ВИСНОВКИ

У дисертації поставлено і розв'язано актуальну науково-прикладну задачу розроблення моделей та методу інформаційної технології для систем комп'ютерного зору як основи людино-машинної взаємодії з використанням жестів рук.

Отримано нові теоретичні та практичні результати:

1. На основі проведеного аналізу існуючих підходів і методів ЛМВ встановлено, що існує об'єктивне протиріччя між зростаючою кількістю впроваджень високотехнологічних систем комп'ютерного зору і рівнем адаптації цих технологій в системах ЛМВ за допомогою жестів. Головною вимогою візуального розпізнавання жестів до їх використання в інтерфейсах ЛМВ є безперервне розпізнавання в режимі реального часу з мінімальною затримкою на виході розпізнавання, але досягти такої продуктивності системи надзвичайно складно з наступних причин: (1) початкова і кінцева точка жесту невідомі; (2) захоплений кадр має негайно пройти попередню обробку, після чого всі елементи повинні бути витягнуті та використані для класифікації, перш ніж буде захоплено наступний кадр; (3) необхідно встановити спеціальні умови щодо освітлення, фону, розташування і людини і камери.

2. Попри значні дослідницькі досягнення в сфері безконтактного управління жестами, у багатьох процесах розробки все ще бракує врахування ергономічних вимог, що може мати наслідки для здоров'я та досвіду користувачів при тривалому використанні.

3. З метою забезпечення ефективного процесу проектування систем ЛМВ у дисертації пропонується нова методологія розробки моделей розпізнавання статичних та динамічних жестів рук за допомогою CNN та CNN+LSTM, що дозволило отримати дорожню карту з розробки моделей розпізнавання жестів, способів їх налаштування для різних інтерфейсів управління жестами та потенційних застосувань в системах ЛМВ.

4. Проведено розробку моделей комп'ютерного зору на основі глибоких нейронних мереж, що забезпечують відстеження та розпізнавання статичних і динамічних жестів рук, здатних функціонувати в режимі реального часу, інваріантних до афінних перетворень і зміни освітлення. Зокрема удосконалено модель статичного розпізнавання жестів, яку побудовано згідно пропонованої методології на основі згорткової нейронної мережі, шляхом штучного збільшення даних і використання контурів. Завдяки використанню контурів, модель є стійкою до відносно широких кутів обертання рук і незалежною від освітлення. Штучне збільшення даних дозволило покращити точність розпізнавання жестів на 4% по відношенню з моделлю без аугментації.

5. Дістала подальшого розвитку технологія розпізнавання та прогнозування жестів на основі моделі генерації послідовностей з використанням ConvLSTM

шляхом введення контурів на етапі попередньої обробки, реалізованих у вигляді фільтру, що дозволило фіксувати часові залежності та зменшити шум вхідного сигналу. В базовій моделі була проведена заміна 3 шарів batch normalization на 3 слоя dropout regularization що дозволило знизити перенавчання та мінімізувати помилки передбачення при прогнозуванні жестів.

6. Проведено удосконалення моделі скінченного автомату для безконтактного управління переглядом медичних зображень шляхом її адаптації до використання даних прогнозованих кадрів відеопослідовностей з моделі ConvLSTM, що дозволило зменшити час відгуку системи;

7. Створено новий набір даних і виконано концептуальну розробку інтуїтивного словникового запасу динамічних жестів, що дозволило реалізувати ефективну безконтактну інтерактивну систему, адаптовану до особливостей хірургічного контексту.

8. Удосконалено структурну модель інформаційної технології ЛМВ з використанням жестів шляхом визначення основних інформаційних потоків створення та інтеграції моделей глибокого навчання, яка забезпечує прийняття рішень щодо застосування розроблених методів, засобів і технологій.

9. Усі теоретичні положення дисертації доведено до відповідних інженерних рішень із застосуванням запропонованої інформаційної технології. Практичні результати дисертаційної роботи апробовано та впроваджено в медичному центрі “Твоя лікарня” при розробці проекту клінічної системи безконтактного перегляду медичних зображень у м. Сєвєродонецьк, Луганській області.

10. Матеріали дисертації використані і впроваджені в навчальний процес на кафедрі комп'ютерних наук та інженерії Східноукраїнського національного університету ім. В. Даля.

Тим самим доведено, що поставлені завдання виконано, мету роботи досягнуто.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Праці, які відображають основні наукові результати дисертації

1. Skarga-Bandurova I., Siriak R., Bilodorodova T., Cuzzolin F., Bawa V.S., Mohamed M.I., Charles R.D. “Surgical Hand Gesture Prediction for the Operating Room”, *Studies in Health Technology and Informatics* (Eds. Bernd Blobel, Lenka Lhotska, Peter Pharow, Filipe Sousa). 2020, Vol. 273, pp. 97-103. (Індексується в SCOPUS).

2. Сіряк Р.В., Скарга-Бандурова І.С., Шумова Л.О. “Особливості технології обробки даних для розпізнавання жестів”, *Вісник Національного технічного університету “ХПІ”*. Збірник наукових праць. Серія: Інформатика та моделювання. Харків: НТУ “ХПІ”. 2019, №.13(1338). С. 112-122.

3. Білобородова Т.О., Сіряк Р.В., Скарга-Бандурова І.С., Давіденко М.О., Сіроштан І.В., Приймак С.О. “Розпізнавання взіємодії інструментів з тканиною на медичних відеозображеннях”, *Наукові вісті Давідовського університету : електронне наукове фахове видання*. 2019, № 17. Режим доступу: http://nvdu.snu.edu.ua/wp-content/uploads/2020/03/2019_17_5.pdf

4. Сіряк Р.В., Скарга-Бандурова І.С., Білобородова Т.О. "Towards an empirical hyper parameters optimization in CNN", *Вісник Східноукраїнського національного університету ім. В. Даля*. Сєверодонецьк: СХУ ім. В. Даля. 2019, № 5(253). С. 87-91.
5. Сіряк Р.В., Скарга-Бандурова І.С. "Модель обробки поточкових даних для розпізнавання окремих одиниць жестової мови", *Вісник Національного технічного університету "ХПІ". Збірник наукових праць. Серія: Інформатика та моделювання*. Харків: НТУ "ХПІ". 2018, № 42(1318). С. 73-81.
6. Болтов Є.В., Сіряк Р.В., Щербаков Є.В., Лифар О.К. "Тестування продуктивності операційних систем реального часу для комп'ютерного зору і обробки зображень на одноплатних комп'ютерах", *Вісник Східноукраїнського національного університету ім. В. Даля*, Сєверодонецьк: СХУ ім. В. Даля. 2018, № 6(247). С. 23-26.
7. Сіряк Р.В. "Технологии идентификации и распознавания жестов", *Вісник Східноукраїнського національного університету ім. В. Даля*, Сєверодонецьк: СХУ ім. В. Даля. 2017, № 8(238). С.79-85.

Наукові праці, які засвідчують апробацію матеріалів дисертації

8. Siriak R., Skarga- Bandurova I., Boltov Y. "Deep Convolutional Network with Long Short-Term Memory Layers for Dynamic Gesture Recognition", *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2019 10th IEEE International Conference*, 2019. Vol. 1, pp. 158-162. (Індексується в SCOPUS).
9. Boltov Y., Hrushka M., Kotsiuba I., Skarga-Bandurova I., Krivoulya G., Siriak R. "Performance Evaluation of Real-Time System for Vision-Based Navigation of Small Autonomous Mobile Robots", *2019 IEEE 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, UK, Leeds, June 5-7, 2019. pp. 218-222. (Індексується в SCOPUS).
10. Сіряк Р.В., Скарга-Бандурова І.С. Система розпізнавання динамічних жестів за допомогою мережі CNN-LSTM. *Комп'ютерні інтелектуальні системи та мережі. Матеріали XI Всеукраїнської науково практичної WEB конференції аспірантів, студентів та молодих вчених (21-23 березня 2019 р.)*. – Кривий Ріг: ДВНЗ «Криворізький національний університет», 2019. – С. 94-95.
11. Siriak R., Skarga-Bandurova I., Biloborodova T. Hyperparameters Optimization in CNN for Image Recognition. *Theoretical and Applied Computer Science and Information Technology: Proceedings of the III International Conference TACSIT-2019*, May 8-9 2019. Severodonetsk: Volodymyr Dahl East Ukrainian National University, pp. 21-22.
12. Сіряк Р.В., Скарга-Бандурова І.С. Аналіз ефективності адаптивних методів оптимізації гіперпараметрів згорткових нейронних мереж. *Проблеми інформатики і моделювання. Тезиси дев'ятнадцятої міжнар. наук.-техн. конф.* – Харків: НТУ "ХПІ", 2019. С. 76.
13. Сіряк Р.В., Скарга-Бандурова І.С. Використання конволюційної нейронної мережі для розпізнавання динамічних образів. *Сучасні технології в освіті та науці: Матеріали міжнар. наук.-практ. конф.*; 19 – 22 лютого 2018 р., м. Сєверодонецьк / гол. ред. О.І. Рязанцев – Сєверодонецьк: вид-во СХУ ім. В. Даля, 2018. С. 45-47.

14. Сіряк Р.В., Скарга-Бандурова І.С. Вибір архітектури глибокого навчання для системи розпізнавання жестів рук. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: тези доповідей XXVI міжнародної науково-практичної конференції MicroCAD-2018*, 16-18 травня 2018р.: у 4 ч. Ч. IV. / за ред. проф. Сокола Є.І. – Харків: НТУ «ХПІ». С. 203.

15. Сіряк Р.В. Средства создания отчетной медицинской документации в медицинской информационной системе. *Збірник тез доповідей міжнародної конференції студентів, аспірантів та молодих вчених "Технологія-2012"*, Сєверодонецьк: Технол. ін-т Східноукр. Нац. ун-ту ім. В. Даля (м. Сєверодонецьк), 2012. С. 94-95.

16. Сіряк Р.В., Висоцький Д.В. Система розпізнавання жестів рук через реалізацію моделі нейронної мережі. *IT-Ідея – 2018: збірник науково-практичних праць*. Сєверодонецьк : Вид-во Східноукр. ун-ту ім. В. Даля, 2018. С. 80-81.

АНОТАЦІЇ

Сіряк Ростислав Вікторович. Моделі та метод інформаційної технології людино-машинної взаємодії з використанням жестів. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. Східноукраїнський національний університет імені Володимира Даля, Сєверодонецьк, 2021.

Дисертацію присвячено вирішенню актуального науково-прикладного завдання покращення характеристик автоматичного розпізнавання статичних та динамічних жестів рук за рахунок розробки та практичного використання моделей і методу інформаційної технології людино-машинної взаємодії з використанням жестів.

Об'єктом дослідження є процеси людино-машинної взаємодії, що здійснюють ідентифікацію, відстеження та класифікацію жестів рук або інших об'єктів на зображенні та у відеопослідовності, а предметом досліджень є моделі та метод інформаційної технології ідентифікації та класифікації статичних і динамічних жестів рук для вирішення задач людино-машинної взаємодії у реальному часу.

У роботі виконано аналітичний огляд сучасних методів та підходи до розроблення систем розпізнавання статичних і динамічних жестів рук. Встановлено, що: незважаючи на те, що жест є природним способом спілкування, при використанні жестів для ЛМВ необхідно враховувати існуючі обмеження технології та людини. За результатами аналізу сучасного стану та тенденцій розвитку систем людино-машинної взаємодії показано необхідність розробки ефективних моделей і методів інформаційних технологій з використанням жестів. Проведено аналіз нормативної бази та вимог до систем людино-машинної взаємодії із застосуванням жестів. Сформульовані задачі досліджень, пов'язаних з аналізом та обробкою зображень, приведенням даних до вигляду, який був би найбільш прийнятним для подальшої роботи з урахуванням особливостей використовуваних архітектур нейронної мережі. Обґрунтована методика і математичний апарат досліджень.

У роботі поставлене та вирішене актуальне науково-прикладне завдання покращення характеристик автоматичного розпізнавання статичних та динамічних

жестів рук за рахунок розробки та практичного використання моделей і методу інформаційної технології людино-машинної взаємодії з використанням жестів.

Методологічну основу дисертаційного дослідження становлять методи загальної теорії систем, системного аналізу, методи структурного синтезу, аналізу та моделювання процесів, що використовувались для розроблення методології проектування і спільного використання візуальних методів розпізнавання жестів рук; методи машинного навчання, технології збільшення даних (data augmentation), технології попередньої обробки даних (очищення, фільтрація розмиття за Гауссом, перетворення кольорів, адаптивна порогова обробка, алгоритм Otsu, детектор контурів Canny, для відділення зон зображення, які містять обличчя та долоні, використано модель дифузного освітлення, для визначення форми долоні використані моделі деформації зображень, порівняння зображень із спотвореннями), що використовувались при розробленні моделей комп'ютерного зору на основі глибоких нейронних мереж; методи зворотного поширення похибки та спряжених градієнтів, для стеження за долонями використані методи статичної та динамічної кластеризації, що використовувались для при розробленні технології розпізнавання та прогнозування динамічних жестів використано моделі генерації послідовностей з використанням ConvLSTM; методи математичної статистики, функції просторового домену, теорія автоматів, теорія подібності, що використовувались при розробленні при розробленні технології взаємодії між хірургом та системою перегляду зображень в операційній залі.

Вперше розроблено методологію розробки і спільного використання візуальних методів розпізнавання жестів рук, яка дозволяє створювати і досліджувати моделі глибокого навчання для розпізнавання статичних та динамічних жестів, здатних працювати в режимі реального часу і забезпечує розуміння способів їх налаштування для різних інтерфейсів управління жестами та потенційних застосувань в системах ЛМВ.

Удосконалено модель статичного розпізнавання жестів, побудовану згідно запропонованої методології на основі згорткової нейронної мережі, шляхом штучного збільшення даних і використання контурів. Завдяки використанню контурів, модель є стійкою до відносно широких кутів обертання рук і незалежною від освітлення.

Дістала подальшого розвитку технологія розпізнавання та прогнозування жестів на основі моделі генерації послідовностей з використанням ConvLSTM2D і Conv3D.

Удосконалено модель скінченного автомату для безконтактного управління переглядом медичних зображень за допомогою жестів яка, на відміну від існуючих, використовує дані прогнозованих кадрів відеопослідовностей, що дозволяє зменшити час відгуку системи.

Для тестування запропонованих моделей і методу інформаційної технології розв'язування задач розпізнавання та прогнозування жестів в операційній залі створено новий датасет і виконано концептуальну розробку інтуїтивного словникового запасу динамічних жестів, що дозволяє реалізувати ефективну безконтактну інтерактивну систему, адаптовану до особливостей хірургічного контексту.

Удосконалено структурну модель інформаційної технології ЛМВ з використанням жестів, за рахунок визначення основних етапів та інформаційних

потоків створення та інтеграції моделей глибокого навчання, яка забезпечує прийняття рішень щодо застосування розроблених методів, засобів і технологій.

Усі теоретичні положення дисертації доведено до конкретних інженерних рішень із застосуванням запропонованої інформаційної технології людино-машинної взаємодії з використанням жестів. Використання моделей та інформаційної технології людино-машинної взаємодії з використанням жестів в системах людино-машинної взаємодії, зокрема для безконтактного перегляду медичних зображень в медичній установі, дозволило зробити висновок, про їх ефективність в частині підвищення точності розпізнавання жестів, зменшення часу на обробку даних жестів у відеопослідовностях та зменшення часу на відгук системи та виконання інструкцій у відповідь на жести.

Ключові слова: комп'ютерний зір, людино-машинна взаємодія, нейромережева архітектура, статичне та динамічне розпізнавання жестів, скінчений автомат.

Rostyslav Siriak. Models and Information Technology for Human-Machine Interaction Using Hand Gestures. – Manuscript copyright.

The dissertation on competition of a scientific degree of the candidate of technical sciences on a specialty 05.13.06 "Information technology". - Volodymyr Dahl East Ukrainian National University, Severodonetsk, 2021.

The dissertation is devoted to solving the current scientific and applied problem of improving the characteristics of static and dynamic hand gestures recognition through the development and practical use of models and methods of information technology of human-machine interaction using gestures.

The object of research is the processes of human-machine interaction that identify, track and classify hand gestures or other objects in the image and video sequence, and the subject of research are models and methods of information technology identification and classification of static and dynamic hand gestures to solve problems of human-machine interaction in real time.

The paper presents an analytical review of modern methods and approaches to the development of systems for recognizing static and dynamic hand gestures. It is established that: despite the fact that gesture is a natural way of communication, when using gestures for HMI it is necessary to take into account the existing limitations of technology and man. According to the results of the analysis of the current state and trends in the development of human-machine interaction systems, the need to develop effective models and methods of information technology using gestures is shown. The analysis of the normative base and requirements to the systems of human-machine interaction with the use of gestures is carried out. The tasks of researches connected with the analysis and processing of images, reduction of data to a kind which would be the most acceptable for the further work taking into account features of the used architectures of a neural network are formulated. Substantiated methodology and mathematical apparatus of research.

The actual scientific and applied task of improving the characteristics of automatic recognition of static and dynamic hand gestures due to the development and practical use of models and methods of information technology of human-machine interaction with the use of gestures is set and solved in the work.

For the first time, a methodology for developing and sharing visual hand gesture recognition techniques has been developed to create and explore deep learning models for recognizing static and dynamic gestures capable of real-time operation and to understand how to configure them for different gesture control interfaces and potential applications in HMI systems.

The model of static gesture recognition, built according to the proposed methodology based on a convolutional neural network, by artificially enlarging data and using contours, has been improved. Thanks to the use of contours, the model is resistant to relatively wide angles of rotation of the hands and independent of lighting.

Gesture recognition and prediction technology based on the sequence generation model using ConvLSTM2D and Conv3D has been further developed.

The model of the finite automatic machine for contactless control of viewing of medical images by means of gestures which, unlike existing, uses the data of the predicted frames of video sequences that allows to reduce response time of system is improved.

A new dataset was created to test the proposed models and method of information technology for solving problems of gesture recognition and prediction in the operating room, and conceptual development of an intuitive vocabulary of dynamic gestures was created, which allows to implement an effective contactless interactive system adapted to the surgical context.

The structural model of HMI information technology using gestures has been improved by identifying the main stages and information flows of creating and integrating deep learning models, which provides decisions on the application of developed methods, tools and technologies.

All theoretical provisions of the dissertation are brought to concrete engineering decisions with application of the offered information technology of human-machine interaction using gestures. The use of models and information technology of human-machine interaction using gestures in human-machine interaction systems, in particular for contactless viewing of medical images in a medical departments, allowed us to conclude that they are effective in improving the accuracy of gesture recognition. video sequences and reduce system response time and execution of instructions in response to a gesture.

Key words: computer vision, human-machine interaction, neural network architecture, static and dynamic gesture recognition, finite state machine.

Підписано до друку 09.04.2021 р. Формат 60*84 1/16. Цифровий друк.
Умов. друк. арк. 1,4. Обл.-вид. арк. – 1,36.
Наклад 100 прим. Замовлення № 138, 2021

Віддруковано ФОП Семенова Т.В.
Свідоцтво В02 №112665 видане виконавчим комітетом
Сєверодонецької міської ради Луганської обл. 26.02.2007р.