УДК 004.032.26:004.357]:37.018.43
DOI: 10.31866/2617-796X.5.1.2022.261293

Ковалюк Тетяна,
к. т. н., доцент кафедри програмних систем і технологій,
Київський національний університет імені Тараса Шевченка,
Київ, Україна
tetyana.kovalyuk@gmail.com
https://orcid.org/0000-0002-1383-1589

Шевченко Анастасія,
магістр, кафедра програмних систем і технологій,
Київський національний університет імені Тараса Шевченка,
Київ, Україна
nastyashev99@gmail.com
https://orcid.org/0000-0001-5230-8339

Кобець Наталія,
інженер, UNITY-BARS LLC,
Київ, Україна
nmkobets@gmail.com
https://orcid.org/0000-0003-4266-9741

МУЛЬТИБІОМЕТРИЧНА ІДЕНТИФІКАЦІЯ СТУДЕНТА ЗА ЙОГО ГОЛОСОВИМИ ТА ВІЗУАЛЬНИМИ БІОМЕТРИЧНИМИ ПОКАЗНИКАМИ В ПРОЦЕСІ ДИСТАНЦІЙНОЇ ОСВІТИ

Мета дослідження – розкрити сутність мультибіометричної ідентифікації студента й обґрунтувати доцільність її застосування для покращення якості, мінімізувати похибки в процесі його розпізнавання із застосуванням голосових і візуальних біометричних ідентифікаторів, що зберігаються в аудіофайлах, відео- та фотозображеннях.

Методи дослідження. Застосовано системний підхід щодо визначення вимог до програмного забезпечення системи мультибіометричної ідентифікації людини, методи обробки звуку, моделі нейронних мереж як класифікатори, що ідентифікують особу за вектором голосових ознак, методи візуальної ідентифікації особи за відеопотоком і за фотозображеннями.

Наукова новизна. Набули подальшого розвитку методи виявлення голосових ознак диктора, методи ідентифікації та реєстрації особи за її голосовими ознаками, алгоритми візуальної ідентифікації особи за її зображеннями у відеопотоці та за фотозображеннями на базі алгоритмів Віоли-Джонса, Eigenface і FisherFace; розроблено архітектуру системи мультибіометричної ідентифікації людини.

Висновки. Запропоновано мультибіометричну ідентифікацію студента за його голосовими та візуальними біометричними показниками для системи дистанційної освіти. Система передбачає витягнення акустичних характеристик із запису мови людини та подальше

віднесення отриманих даних до одного з наперед заданих класів (дикторів). У ролі класифікатора застосовано багатошарову нейронну мережу (БШНМ). Класифікатор навчений на наборі даних з 43832 аудіофайлів від 108 дикторів. БШНМ на тестовій вибірці продемонструвала точність у 91 %. На етапі обробки кадрів відеопотоку здійснено виявлення обличчя в кадрі та розпізнавання виявленого обличчя. Розпізнавання облич у системі проводилося на основі пошуку найбільш відповідного шаблону базових зображень, що зберігаються в базі даних. Розроблено програмну систему для розпізнавання та індексації людей на відео одночасно з ідентифікацією особи за голосовими ознаками, щоб використовувати її в освітньому процесі для обліку відвідування дистанційних занять.

Ключові слова: машинне навчання; штучні нейронні мережі; ідентифікація диктора; біометрія; розпізнавання голосу; розпізнавання облич.

Вступ. З розвитком і впровадженням у повсякденне життя інформаційних систем важливим завданням є забезпечення контролю доступу до даних користувача та їх збереження. Епідеміологічна ситуація у світі, що пов'язана з Covid-19, привела до неминучого переходу освітніх установ на дистанційну форму навчання. Дистанційні освітні технології сприяють диференціації навчання, розвитку колективної творчості та креативності студента, зростанню навчальної мотивації, академічної мобільності, гнучкості освітнього процесу (Distance Learning in 2021: How to make the most of this school year). Проте найбільш складним завданням у процесі впровадження дистанційних технологій освіти є проведення дистанційної атестації з ідентифікацією особи студента, який складає іспит, залік, пише тест, контрольну або захищає есе, курсову чи кваліфікаційну роботу. Ідентифікація особи потрібна й під час доступу до навчальних онлайн-ресурсів і конфіденційної інформації, зокрема журналів обліку успішності та відвідування занять студентами, індивідуальних освітніх траєкторій навчання тощо. Ідентифікація особи передбачає отримання деякого набору унікальних даних, що її характеризують і визначають у процесі перевірки. Як альтернатива парольній верифікації та ідентифікації може застосовуватися біометрична ідентифікація людини за унікальними, властивими тільки їй біологічними ознаками. До таких ознак зараховують геометричну будову руки, відбитки пальців, особливості малюнка сітківки ока, райдужну оболонку ока, портрет (наприклад, інфрачервону карту людини), характеристики й особливості мови, рукописний почерк, клавіатурний і комп'ютерний почерк, інші фізіологічні особливості людини, що роблять її унікальною. Біометричні параметри завжди в наявності, їх не можна забути, втратити, передати іншій людині, вкрасти й досить важко відтворити. У цій роботі розглянуто голосову та візуальну біометрію як дві підсистеми розпізнавання користувачів у системі дистанційної освіти.

Голосова біометрія заснована на аналізі унікальних характеристик промови диктора. Вона охоплює процедури ідентифікації та верифікації особи за голосом (Юдін та Зюбіна, 2017), є однією з найбільш перспективних технологій через широке поширення засобів зв'язку. Перевагою технології голосової біометрії є можливість ідентифікації користувача на відстані, обмеженій тільки каналом зв'язку,

та її дешевизна через легкість отримання даних, що особливо актуально для віддаленої взаємодії студентів і викладачів у системі дистанційної освіти.

Візуальна біометрія – це технологія, що здатна ідентифікувати або верифікувати особу на цифровому зображенні або відеокадрі. Технологія передбачає зіставлення людського обличчя з цифрового зображення або відеокадру з базою даних облич, яка працює за допомогою точного визначення та вимірювання рис обличчя на поданому зображенні. Системи візуального розпізнавання дають змогу розпізнати емоції, стать, вік, етнічну приналежність, здійснювати пошук обличчя в наявній базі тощо.

Огляд останніх публікацій і досліджень з теми. У процесі роботи проаналізовано низку наукових праць з обраної теми статті та визначено, що дослідження в галузі голосової біометрії направлені на розробку алгоритмів за такими напрямами:

– отримання зразка голосу диктора й обробка мовного сигналу з метою отримання ознак для розпізнавання диктора;

– побудова моделі диктора на основі ознак, що витягнуті зі зразка його голосу;

– методи ухвалення рішень щодо результатів ідентифікації та верифікації особи.

Проблеми розпізнавання та ідентифікації обличчя людини:

– розпізнавання для виявлення вузлових точок і вимірювання відстані між певними точками на обличчі;

– ідентифікація для розпізнавання індивідуального екземпляра об'єкта;

– виявлення з перевіркою відеоданих на наявність визначеної умови.

Більшість систем обробки аудіосигналів поєднує два етапи: виділення ознак і вибір класифікатора. Для цього використовуються різноманітні характеристики сигналу, такі як частота переходів через нуль, смуга пропускання сигналу, спектральний центроїд та енергія сигналу, мел-кепстральні коефіцієнти. Залежно від зони обробки мовного сигналу методи можна розділити на три групи: ті, що працюють в частотній зоні, часовій і частотно-часовій (Алимурадов и Чураков, 2015). Найпоширенішими методами аналізу мовного сигналу, на основі яких проводиться розпізнавання диктора, є: швидке перетворення Фур'є (Ernawan, Abu and Suryana, 2011), вейвлет-перетворення (Pandiaraj and Kumar, 2015; Nair and Shah, 2015), перетворення Гільберта-Хуанга (Huang, Acero and Hon, 2001), кепстральний аналіз (Lokesh and Devi, 2019), лінійне передбачення (Wu and Lin, 2009), кореляційний аналіз (Pramanik and Raha, 2012; Gupta, Raibagkar and Palsokar, 2017), нейронні мережі (Kydyrbekova et el., 2020; Ye and Yang, 2021), приховані марківські моделі (Das and Nahar, 2016; Uchat, 2006). Аналіз стану справ у галузі розпізнавання дикторів з метою визначення найбільш перспективних напрямів дослідження подано в аналітичному огляді (Сорокин, 2012).

Проблемі виявлення рис обличчя та його розпізнавання присвячено багато наукових праць вітчизняних і зарубіжних дослідників. Описано також практичну систему видобування рис обличчя (Tin and Htake, 2012). Метод аналізу основних компонентів розглянуто в системі розпізнавання обличчя (Javed, 2013). Серед відомих методів ідентифікації людини за зображенням її обличчя на фото та відео можна виділити метод Віоли-Джонса, який будується за допомогою машинного навчання (Viola and Jones, 2001, 2004). Алгоритм базується на ідеях, що пра-

цюють у режимі реального часу, зокрема функції Хаара, цілісному зображенні, AdaBoost та каскадній структурі (Wang, 2014). Розглянуто алгоритм AdaBoost для адаптивного покращення класифікації через побудову «сильного» класифікатора як лінійної комбінації «слабких» класифікаторів (Sochman and Matas, 2010). Порушено питання застосування візуальної ідентифікації студентів для обліку відвідування занять (Kobets and Kovaliuk, 2020).

Мета статті – розкрити сутність мультибіометричної ідентифікації студента й обґрунтувати доцільність її застосування для покращення якості, мінімізувати похибки в процесі його розпізнавання із застосуванням голосових і візуальних біометричних ідентифікаторів, що зберігаються в аудіофайлах, відео- та фотозображеннях.

Для досягнення мети необхідно розв'язати такі завдання:

– сформувати перелік значущих характеристик аудіосигналу для подальшої обробки та датасет;

– провести попередню обробку даних;

– розробити та застосувати алгоритми видобування характеристик з даних для формування векторів ознак;

– розробити архітектуру класифікатора за векторами ознак і навчити його, оцінити точність, зробити висновки;

– провести дослідження ефективності розробленого алгоритму.

Результатом роботи є програмна система, яка має демонструвати точність вище 90 % на тестовій вибірці під час розпізнавання наперед обраних класів.

Результати дослідження. Бізнес-логіка системи голосової та відеоідентифікації й верифікації студента в системі дистанційної освіти охоплює такі етапи:

– підготовку наборів даних, що містять транскрибовані записи мови дикторів з різними акцентами та якістю аудіо і зображення людей з різними варіаціями повороту голови й емоцій, освітленням, виразами обличчя;

– попередню обробку аудіофайлів (дискредитація, нормалізація, знешумлення, видалення тиші) і відеозображень (вибір усіх можливих зображень облич);

– формування вектора акустичних ознак, що характеризують мовця;

– класифікацію за акустичними ознаками для ухвалення рішення щодо належності певному диктору вхідного аудіо;

– кластеризацію зображень для їх угруповання за рівнем схожості між собою;

– обробки відеозображень для індексації присутніх на них людей;

– формування звітних документів щодо ідентифікації мовця та особи за її зображенням.

На першому етапі підготовки даних для голосової ідентифікації студентів використовувався набір даних VoxForge (http://www.voxforge.org/) з транскрибованими записами мови дикторів з різними акцентами та якістю аудіо. База даних VoxForge складається з опублікованих і стандартизованих користувачами записів читань англомовних текстів. До записів також додавалися такі метадані, як транскрипція тексту, властивості wav-файлу, вид мікрофона, на який записане аудіо, ім'я, стать, акцент, вікова група мовця. Для навчання класифікатора було відібрано 3348 файлів від 14 дикторів, а з метаданих для описової статистики обрано стать й акцент мовця. Для описової статистики (для більшого різноманіття даних)

було також відібрано ще 18 мовців. Сумарно для голосової аналітики використано 4346 файлів від 32 дикторів.

Для навчання класифікатора з розпізнавання обличчя студента за його зображенням у відеопотоці використовувався набір даних AT&T (https://www.kaggle.com/datasets/kasikrit/att-database-of-faces?resource=downloa), який містить зображення 40 людей з використанням 10 різних варіацій повороту голови й емоцій. Для деяких об'єктів зображення зроблені в різний час, зі зміною освітлення, виразу обличчя (відкриті/закриті очі, посміхається / не посміхається) і деталей обличчя (окуляри / без окулярів). Усі зображення зроблені на темному однорідному фоні з об'єктами у вертикальному, фронтальному положенні (з допуском до деяких рухів убік). Розмір кожного зображення становить 92x112 пікселів з 256 рівнями сірого рівня на один піксель.

Попередня обробка даних, що є другим етапом голосової та відеоідентифікації, охоплює такі операції, як дискредитація, нормалізація, знешумлення, видалення тиші. Частота дискретизації кожного аудіофайлу була приведена до 8000 семплів за секунду, що достатньо для отримання інформативних ознак, а в разі більшої частоти процес підрахунків сильно сповільнюється. Отримані wav-форми сигналу нормалізувалися в межах [-1;1]. Для усунення непотрібних звуків, шумів та ізоляції голосу використовувався алгоритм зменшення шумів, оснований на методі спектрального стробування. Він працює через обчислення спектрограми сигналу й оцінки порогу шуму (або вентиля) для кожної смуги частот цього сигналу/шуму. Цей поріг використовується для обчислення маски, яка блокує шум нижче порогу, який може змінюватися. Дані нормалізуються ще раз, і запускається процес приведення їх до одної тривалості – 4 с. Екземпляри, коротші 4 с, розширювалися зацикленням цього аудіо. Від аудіо, довшого 4 с, з початку та з кінця відсікалися ділянки з тишею, за необхідності екземпляр розбивався на декілька сигналів, що охоплювали 4 с, яким присвоювалися ті самі мітки класу та метадані. Після цієї обробки дані нормалізувалися востаннє. Попередня обробка відеозображень охоплює відбір усіх можливих зображень обличчя.

Обчислення акустичних характеристик мовця передбачає видобуток з аудіофайлів чотирьох видів ознак: частоти основного тону (або фундаментальної частоти F0), спектрального центроїда, спектральної пропускної здатності та мел-кепстральних коефіцієнтів (MFCC). Вектор індивідуальних ознак мовця складається з мел-кепстральних коефіцієнтів. Для отримання MFCC вхідний сигнал представлявся в частотному просторі у вигляді спектрограми сигналу на основі дискретного перетворення Фур'є. Алгоритм розрахунку мел-кепстральних коефіцієнтів складається з п'яти основних кроків (Lavrynenko, Kocherhin and Konakhovych, 2018):

– розбиття сигналу на фрейми та застосування віконної функції;
– отримання модулів коефіцієнтів дискретного перетворення Фур'є;
– перехід до мел-простору частот;
– застосування банку мел-фільтрів;
– застосування дискретного косинусного перетворення.

Частоту основного тону F0 отримували за допомогою алгоритму PYIN (Mauch and Dixon, 2014). F0 приймає вигляд вектора частот залежно від проміжків часу. З цього

відбираються для опису максимальне, мінімальне та середнє значення в проміжку 80 Гц – 450 Гц (коливання голосових зв'язок людини). Формантні частоти F1 та F2 знаходилися за допомогою бібліотеки обробки аудіо Parselmouth (Jadoul, Thompson and Boer, 2018). Для кластеризації використовувався метод k-середніх. Спектральний центроїд розраховувався за допомогою бібліотеки обробки аудіо Librosa (https://librosa.org/). З нього відбиралися максимальні, мінімальні та середні значення.

Ідентифікація дикторів за векторами ознак аудіофайлів їх мовлення є завданням класифікації. Класифікатор обрав багатошарову повнозв'язну штучну нейронну мережу (ШНМ) прямого поширення. Обрана мережа має два приховані шари по 128 та 256 нейронів відповідно до сигмоїдної функції активації. Усього параметрів мережі – 66436, з яких навчаються – 66104. Вхідні дані нормалізувалися на шарах батч-нормалізації, а для запобігання перенавчання використано шари прорідження – присвоєння нулів випадково обраним ознакам у процесі навчання. Коефіцієнт прорідження – це доля ознак, які обнуляються (у цьому разі 0,2 та 0,3).

На етапі обробки кадрів відеопотоку відбувається виявлення обличчя в кадрі, після чого здійснюється безпосередньо розпізнавання та ідентифікація виявленого обличчя. Під час розробки системи використано метод виявлення осіб Віоли-Джонса, а для розпізнавання – методи Eigenface і FisherFace (Pissarenko, 2003; Belhumeur, Hespanha and Kriegman, 1997.). Основні принципи, на яких заснована робота методу Віоли-Джонса:

– представлення зображення в інтегральному вигляді;
– пошук осіб за допомогою ознак Хаара;
– каскадна класифікація;
– навчання системи розпізнавання об'єктів на основі методу AdaBoost.

Пошук облич відбувається за допомогою сканувального вікна, яке послідовно рухається по зображенню з кроком в 1 осередок вікна. Під час сканування зображення в кожному вікні обчислюється приблизно 200000 варіантів розташування ознак. Усі знайдені ознаки передаються класифікатору, який визначає за їх значенням, чи є ділянка зображення, що відповідає вікну, обличчям, чи ні. Оскільки для опису об'єкта з достатньою точністю необхідна велика кількість ознак Хаара, вони не дуже підходять для навчання або класифікації. У зв'язку з цим у методі Віоли-Джонса використовується каскадний класифікатор, який дає змогу прискорити виявлення осіб, фокусуючи роботу на найбільш цікавих ділянках зображення. Для вирішення проблеми навчання застосовується технологія бустингу, що є процедурою послідовної побудови композиції алгоритмів машинного навчання, коли кожен наступний алгоритм прагне компенсувати недоліки композиції всіх попередніх алгоритмів.

У бібліотеці OpenCV реалізовано алгоритми розпізнавання осіб Eigenface (метод головних компонент для розпізнавання осіб) та Fisherface (лінійний дискримінантний аналіз). Для алгоритму Eigenface використовується база даних осіб, де зображення мають розмір N x N пікселів. Кожне зображення з бази даних є крапкою в просторі розмірністю N * N. Щоб знайти вектор у просторі осіб, що відповідає цьому зображенню, розкладаємо зображення по кожному з M власних векторів, обчисливши скалярний добуток. Набір M значення утворює вектор у просторі осіб. Для розпізнавання особи на зображенні потрібно знайти відповідний цьому

зображенню вектор у просторі осіб і визначити, до якого вектора з навчальної вибірки він найближчий. Для оцінки відстані доцільно використати дистанцію Махаланобіса. Алгоритм Fisherface передбачає наявність багатьох фотографій за різних умов освітленості у кожної особи в базі даних. В алгоритмі передбачається пошук такого базису, який дав би змогу максимізувати дисперсію між множинами зображень осіб й одночасно мінімізувати дисперсію всередині кожної множини.

Програмна система для мультибіометричної ідентифікації студента складається з шести підсистем:

– підсистема первинної обробки аудіо та відео;
– підсистема формування вектора акустичних ознак;
– підсистема класифікації для ідентифікації мовця (студента);
– підсистема кластеризації зображень;
– підсистема ідентифікації осіб за розпізнаванням облич у відеопотоці;
– підсистема формування звітів.

Результати експериментів досліджували на наборі даних, який був розбитий на тренувальну (70 %), валідаційну (15 %) та тестову (15 %) вибірки. Для навчання ШНМ у ролі функції втрат обрано категоріальну перехресну ентропію. Оптимізатором був метод Adam. Побудована ШНМ навчалася протягом 100 епох, кожна епоха складалася з 480 кроків. На кожному кроці подавалися набори по 64 екземпляри. Після навчання модель разом з вагами збережено у форматі файлу h5 для подальшого використання. У процесі розробки протестовано декілька архітектур ШНМ, результати тестування наведено в табл. 1. Обрано архітектуру № 3 для використання у фінальній версії системи.

Таблиця 1

Результати тестування нейронних мереж

№	Кількість прихованих шарів	Функції активації	Запобігання перенавчанню	Кількість параметрів, що навчаються	Точність на тренувальній та валідаційній вибірках	Точність на тестовій вибірці
1	3	relu, softmax	2 шари проріджження (0.4, 0.5)	71,928	92 %	88 %
2	3	sigmoid, relu, softmax	2 шари проріджження (0.3, 0.4)	132,408	97 %	90 %
3	**2**	**sigmoid, softmax**	**2 шари проріджження (0.2, 0.3)**	**66,104**	**95 %**	**91 %**
4	2	sigmoid, relu, softmax	2 шари проріджження (0.1, 0.3); регуляризатор	13,752	88 %	85 %
5	1	relu, softmax	шар проріджження (0.2); регуляризатор	9,592	88 %	86 %

ШНМ № 3 має меншу різницю між точністю на тренувальній і тестовій вибірках та найкращі результати на тестовій серед усіх ШНМ, тому є найбільш придатною до використання.

Для розпізнавання облич з метою навчання системи використано 960 зображень, з яких 320 зображень обличчя, 640 зображень без обличчя. Для тестування випадковим чином відібрано по 2 зображення з кожного набору, тож у тестовій вибірці було 80 зображень людей, яких немає в навчальній вибірці, і 160 зображень без облич. Для прямого виявлення використано повну бібліотеку з 400 зображень. Вибірок без осіб – 800. Оскільки алгоритм навчання AdaBoost спрямований на створення найкращого поділу даних на два класи, збільшення кількості операцій дає змогу досягти найкращого результату. Алгоритм тестувався на зображенні, що містить декілька облич (рис. 1).

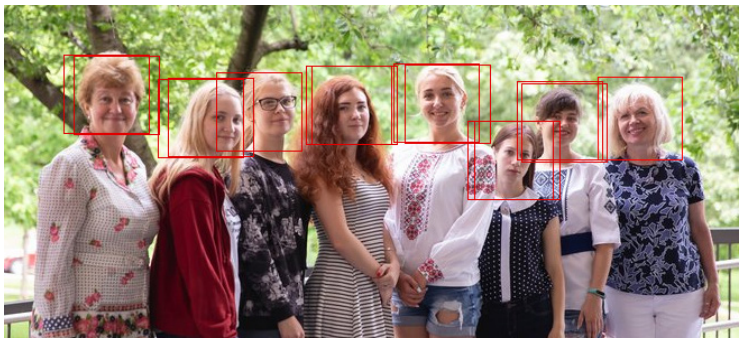

Рис. 1. Результат розпізнавання та ідентифікації на тестовому зображенні

Висновки. У роботі запропоновано мультибіометричну ідентифікацію студента за його голосовими та візуальними біометричними показниками. Система базується на витягненні акустичних характеристик із запису мови людини та подальшому віднесенню отриманих даних до одного з наперед заданих класів (дикторів). Класифікатори навчено на наборі даних з 43832 аудіофайлів від 108 дикторів. Багатошарова нейронна мережа на тестовій вибірці продемонструвала точність, що становить 91 %.

На етапі обробки кадрів відеопотоку здійснювалося виявлення обличчя в кадрі та розпізнавання виявленого обличчя. Під час розробки системи використано метод виявлення осіб Віоли-Джонса, а для розпізнавання – методи Eigenface і FisherFace. Розпізнавання облич у системі проводилося на основі пошуку найбільш відповідного шаблону базових зображень, що зберігається в базі даних.

Розроблено програмну систему для розпізнавання та індексації людей на відео одночасно з ідентифікацією особи за голосовими ознаками. Система впроваджується в освітній процес в університеті для реєстрації відвідування дистанційних занять. Зображення облич учнів створюватимуться за допомогою відеокамери, розпізнаватимуться та ідентифікуватимуться з даними, внесеними в журнал відвідуваності.

СПИСОК ПОСИЛАНЬ

Алимурадов, А.К. и Чураков, П.П., 2015. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи. *Измерение. Мониторинг. Управление. Контроль*, 2 (12), с.27-34.

Сорокин, В.Н., Вьюгин, В.В. и Тананыкин, А.А., 2012. Распознавание личности по голосу: аналитический обзор. *Информационные процессы*, 12 (1), с.1-30.

Юдін, О.К. та Зюбіна, Р.В., 2017. Аналіз сучасних систем та методів розпізнавання аудіосигналів у задачах ідентифікації та верифікації. *Проблеми інформатизації та управління*, 3 (59), с.75-79.

AT&T database of faces. [online] Avialable at: <https://www.kaggle.com/datasets/kasikrit/att-database-of-faces?resource=download> [Accessed 5 December 2021].

Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE transactions on pattern analysis and machine intelligence*, 19 (7), pp.711-720.

Das, T.K. and Nahar Khalid M.O., 2016. A Voice identification system using hidden Markov model. *Indian Journal of Science and Technology*, 9 (4), pp.1-6.

Distance Learning in 2021: How to make the most of this school year. *Lumin.* [online] Avialable at: <https://www.luminpdf.com/distance-learning-in-2021/> [Accessed 20 January 2022].

Ernawan, F., Abu, N. and Suryana, N., 2011. Spectrum analysis of speech recognition via discrete Tchebichef transform. *International Conference on Graphic and Image Processing (ICGIP 2011)*, 8285, pp.1619-1626.

Gupta, A., Raibagkar, P. and Palsokar, A., 2017. Speech Recognition Using Correlation Technique. *International Journal of Current Trends in Engineering & Research (IJCTER)*, 3 (6), pp.82-89.

Huang, X., Acero, A. and Hon, H.-W., 2001. Spoken language processing. Guide to algorithms and system development. United States: Prentice Hall.

Jadoul, Y., Thompson, B. and De Boer, B., 2018. Introducing Parselmouth: a Python interface to Praat. *Journal of Phonetics,* 71, pp.1-15.

Javed, A., 2013 Face Recognition Based on Principal Component Analysis. *International Journal of Image, Graphics and Signal Processing*, 2, pp.38-44.

Kobets, N. and Kovaliuk, T., 2020. Method of Recognition and Indexing of People's Faces in Videos Using Model of Machine Learning. *Advances in Intelligent Systems and Computing*, 1247, pp.534-544.

Kydyrbekova, A., Othman, M. Mamyrbayev, O., Akhmediyarova, A. and Bagashar, Z., 2020. Identification and authentication of user voice using DNN features and i-vector. *Cogent Engineering*, https://www.tandfonline.com/journals/oaen207 (1), pp.1-21.

Lavrynenko, O.Yu, Kocherhin, Y.A. and Konakhovych, G.F., 2018. Voice Control Command Recognition System of UAV Based on Steganographic-Cepstral Analysis. *Electronics and Control Systems,* 2 (56), pp.11-17.

Librosa: Audio and Music Processing in Python. [online] Avialable at: <https://librosa.org/> [Accessed 26 March 2022].

Lokesh, S. and Devi, M.R., 2019. Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method. *Cluster Computing,* 22, pp.11669-11679.

Mauch, M. and Dixon, S., 2014. PYIN: a fundamental frequency estimator using probabilistic threshold distributions. *International Conference on Acoustics, Speech, & Signal Processing,* pp.659-663.
Nair, S.R. and Shah, M.S., 2015. Applications of wavelet transform in speech processing: a review. *International Journal of Engineering Research & Technology*, 3 (1), pp.1-5.
Pandiaraj, S. and Kumar, K.R.S., 2015. Speaker identification using discrete wavelet transform. *Journal of Computer Science*, 11 (1), pp.53-56.
Pissarenko, D., 2003. *Eigenface-Based Facial Recognition.* [online] Avialable at: <https://www.researchgate.net/publication/2563672_Eigenface-Based_Facial_Recognition> [Accessed 20 January 2022].
Pramanik, A. and Raha, R., 2012. Automatic Speech Recognition using correlation analysis. *2012 World Congress on Information and Communication Technologies*, pp.670-674.
Sochman, J. and Matas, J., 2010. AdaBoost. *Prague: Center for Machine Perception, Czech Technical University*. [online] Avialable at: <https://cmp.felk.cvut.cz/~sochmj1/adaboost_talk.pdf> [Accessed 10 March 2022]
Tin, H. and Htake, H., 2012. Perceived gender classification from face images. *International Journal of Modern Education and Computer Science*, 1, pp.12-18.
Uchat, N.S., 2006. *Hidden Markov Model and Speech Recognition*. Indian Institute of Technology Mumbai.
Viola, P. and Jones, M.J., 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9
Viola, P. and Jones, M.J., 2004. Robust real-time face detection. *International Journal of Computer Vision,* 57 (2), pp.137-154.
VoxForge. [online] Avialable at: <http://www.voxforge.org/> [Accessed 5 December 2021].
Wang, Y-Q., 2014. An Analysis of the Viola-Jones Face Detection Algorithm. *Image Processing On Line,* 4, pp.128-148.
Wu, J.-D. and Lin, B.-F., 2009. Speaker identification based on the frame linear predictive coding spectrum technique. *Expert Systems with Applications,* 36 (4), pp.8056-8063.
Ye, F. and Yang, J., 2021. A Deep Neural Network Model for Speaker Identification. *Applied Sciences*, 11 (3603), pp.2-18.

REFERENCES

Alimuradov, A.K. and Churakov, P.P., 2015. Obzor i klassifikatciia metodov obrabotki rechevykh signalov v sistemakh raspoznavaniia rechi [Review and classification of methods for processing speech signals in speech recognition systems]. *Izmerenie. Monitoring. Upravlenie. Kontrol*, 2 (12), pp.27-34.
AT&T database of faces. [online] Avialable at: <https://www.kaggle.com/datasets/kasikrit/att-database-of-faces?resource=download> [Accessed 5 December 2021].
Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J., 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE transactions on pattern analysis and machine intelligence*, 19 (7), pp.711-720.
Das, T.K. and Nahar Khalid M.O., 2016. A Voice identification system using hidden Markov model. *Indian Journal of Science and Technology*, 9 (4), pp.1 6.

Distance Learning in 2021: How to make the most of this school year. *Lumin*. [online] Avialable at: <https://www.luminpdf.com/distance-learning-in-2021/> [Accessed 20 January 2022].
Ernawan, F., Abu, N. and Suryana, N., 2011. Spectrum analysis of speech recognition via discrete Tchebichef transform. *International Conference on Graphic and Image Processing (ICGIP 2011)*, 8285, pp.1619-1626.
Gupta, A., Raibagkar, P. and Palsokar, A., 2017. Speech Recognition Using Correlation Technique. *International Journal of Current Trends in Engineering & Research (IJCTER)*, 3 (6), pp.82-89.
Huang, X., Acero, A. and Hon, H.-W., 2001. *Spoken language processing. Guide to algorithms and system development*. United States: Prentice Hall.
Jadoul, Y., Thompson, B. and De Boer, B., 2018. Introducing Parselmouth: a Python interface to Praat. *Journal of Phonetics*, 71, pp.1-15.
Javed, A., 2013 Face Recognition Based on Principal Component Analysis. *International Journal of Image, Graphics and Signal Processing*, 2, pp.38-44.
Kobets, N. and Kovaliuk, T., 2020. Method of Recognition and Indexing of People's Faces in Videos Using Model of Machine Learning. *Advances in Intelligent Systems and Computing*, 1247, pp.534-544.
Kydyrbekova, A., Othman, M., Mamyrbayev, O., Akhmediyarova, A. and Bagashar, Z., 2020. Identification and authentication of user voice using DNN features and i-vector. *Cogent Engineering*, 7 (1), pp.1-21.
Lavrynenko, O.Yu, Kocherhin, Y.A. and Konakhovych, G.F., 2018. Voice Control Command Recognition System of UAV Based on Steganographic-Cepstral Analysis. *Electronics and Control Systems*, 2 (56), pp.11-17.
Librosa: Audio and Music Processing in Python. [online] Avialable at: <https://librosa.org/> [Accessed 26 March 2022].
Lokesh, S. and Devi, M.R., 2019. Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method. *Cluster Computing*, 22, pp.11669-11679.
Mauch, M. and Dixon, S., 2014. PYIN: a fundamental frequency estimator using probabilistic threshold distributions. *International Conference on Acoustics, Speech, & Signal Processing*, pp.659-663.
Nair, S.R. and Shah, M.S., 2015. Applications of wavelet transform in speech processing: a review. *International Journal of Engineering Research & Technology*, 3 (1), pp.1-5.
Pandiaraj, S. and Kumar, K.R.S., 2015. Speaker identification using discrete wavelet transform. *Journal of Computer Science*, 11 (1), pp.53-56.
Pissarenko, D., 2003. Eigenface-Based Facial Recognition. [online] Avialable at: <https://www.researchgate.net/publication/2563672_Eigenface-Based_Facial_Recognition> [Accessed 20 January 2022].
Pramanik, A. and Raha, R., 2012. Automatic Speech Recognition using correlation analysis. *2012 World Congress on Information and Communication Technologies*, pp.670-674.
Sochman, J. and Matas, J., 2010. *AdaBoost*. Prague: Center for Machine Perception, Czech Technical University. [online] Avialable at: <https://cmp.felk.cvut.cz/~sochmj1/adaboost_talk.pdf> [Accessed 10 March 2022]
Sorokin, V.N., Viugin, V.V. and Tananykin, A.A., 2012. Raspoznavanie lichnosti po golosu: analiticheskii obzor [Personality Recognition by Voice: An Analytical Review]. *Informatcionnye protcessy*, 12 (1), pp.1-30.
Tin, H. and Htake, H., 2012. Perceived gender classification from face images. *International Journal of Modern Education and Computer Science*, 1, pp.12-18.
Uchat, N.S., 2006. *Hidden Markov Model and Speech Recognition*. Indian Institute of Technology Mumbai.

Viola, P. and Jones, M.J., 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9
Viola, P. and Jones, M.J., 2004. Robust real-time face detection. *International Journal of Computer Vision*, 57 (2), pp.137-154.
VoxForge. [online] Avialable at: <http://www.voxforge.org/> [Accessed 5 December 2021].
Wang, Y-Q., 2014. An Analysis of the Viola-Jones Face Detection Algorithm. *Image Processing On Line*, 4, pp.128-148.
Wu, J.-D. and Lin, B.-F., 2009. Speaker identification based on the frame linear predictive coding spectrum technique. *Expert Systems with Applications*, 36 (4), pp.8056-8063.
Ye, F. and Yang, J., 2021. A Deep Neural Network Model for Speaker Identification. *Applied Sciences*, 11 (3603), pp.2-18.
Yudin, O.K. and Ziubina, R.V., 2017. Analiz suchasnykh system ta metodiv rozpiznavannia audiosyhnaliv u zadachakh identyfikatsii ta veryfikatsii [Analysis of modern systems and methods of recognition of audio signals in the problems of identification and verification]. *Problemy informatyzatsii ta upravlinnia*, 3 (59), pp.75-79.

UDC 004.032.26:004.357]:37.018.43

Kovaliuk Tetiana,
PhD, Associate Professor at the Software Systems and Technologies Department,
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine
tetyana.kovalyuk@gmail.com
https://orcid.org/0000-0002-1383-1589

Shevchenko Anastasiia,
Master at the Software Systems and Technologies Department,
Taras Shevchenko National University of Kyiv,
Kyiv, Ukraine
nastyashev99@gmail.com
https://orcid.org/0000-0001-5230-8339

Kobets Nataliia,
Engineer, UNITY-BARS LLC,
Kyiv, Ukraine
nmkobets@gmail.com
https://orcid.org/0000-0003-4266-9741

MULTIBIOMETRIC IDENTIFICATION OF THE STUDENT BY HIS VOICE AND VISUAL BIOMETRIC INDICATORS IN THE PROCESS OF DISTANCE EDUCATION

The purpose of the study is to reveal the essence of multibiometric identification of students and substantiate the feasibility of its use to improve quality and minimize errors in recognizing it using voice and visual biometric identifiers stored in audio files, video and photo images.

Research Methodology. A systematic approach to determining the software requirements for a multibiometric human identification system, sound processing methods, neural network models as classifiers that identify a person by the vector of voice characteristics and methods of visual identification of a person by video stream and photo images were applied.

Scientific Novelty. Methods for identifying the speaker's voice signs, methods for identifying and registering a person by his voice signs, algorithms for visual identification of a person from her images in a video stream and from photo images based on the Viola-Jones, Eigenface and FisherFace algorithms have been further developed, and the architecture of a multibiometric identification system has been designed.

Conclusions. Multibiometric identification of the student by voice and visual biometric indicators for the distance education system are offered. The system allows the extraction of acoustic characteristics from recording human language and further assignment of the obtained data to one of the predefined classes (speakers). A multilayer neural network (MNN) was used as a classifier. The classifier is trained on 43832 audio files from 108 speakers. MNN showed an accuracy of 91% in the test sample. The face in the frame is detected at the video stream frame processing stage, and the detected face is recognized. The system performed face recognition based on finding the most appropriate template of basic images stored in the database. A software system for recognising and indexing people on video simultaneously with the identification of a person by voice has been developed, to use in the educational process for recording attendance at distance learning classes.

Keywords: machine learning; artificial neural networks; speaker identification; biometrics; voice recognition; face recognition.

12.01.2022