

УДОСКОНАЛЕНА МОДЕЛЬ РОЗПІЗНАВАННЯ ЕМОЦІЙ УЧАСНИКІВ ВІДЕОКОНФЕРЕНЦІЙ НА ОСНОВІ МУЛЬТИМОДАЛЬНОГО ПІДХОДУ

Вінницький національний технічний університет

Анотація

Запропоновано удосконалену модель розпізнавання емоцій учасників відеоконференції на основі мультимодального підходу, а саме комбінації голосу та виразу обличчя. Наведено етапи розпізнавання емоцій на основі голосу та описано ознаки, які подаються на вхід згорткової нейронної мережі. Описано етапи розпізнавання емоцій на основі виразу обличчя. Наведено узагальнену схему розпізнавання емоцій учасників відеоконференції на основі мультимодального підходу, що комбінує процеси розпізнавання на основі голосу та виразу обличчя.

Ключові слова: розпізнавання емоцій, згорткові нейронні мережі, розпізнавання голосу, обробка зображень.

Abstract

An improved model for recognizing the emotions of video conference participants based on a multimodal approach, specifically, a combination of voice and facial expression, is proposed. The stages of emotion recognition based on voice are presented and the features that are used as an input of the convolutional neural network are described. The stages of emotion recognition based on facial expressions are described. A generalized scheme for recognizing the emotions of video conference participants based on a multimodal approach that combines voice and facial expression recognition processes is presented.

Keywords: emotion recognition, convolutional neural networks, voice recognition, image processing.

Вступ

Останні роки аналіз емоцій учасників відео комунікації набуває все більшої значущості, оскільки відео зв'язок став невід'ємною частиною бізнесу, освіти та повсякденного спілкування. Розуміння емоційних станів учасників сприяє покращенню якості взаємодії [1]. Дослідження емоційних реакцій у процесі комунікації між людьми відкриває можливості для машинного розпізнавання невербальних сигналів і відповідного реагування на них. Аналіз емоцій у відеоконференціях дозволяє визначати настрої учасників, адаптуватися до них та впливати на сприйняття взаємодії, що сприяє попередженню конфліктів і зниженню напруги на ранніх етапах [2].

Результати дослідження

Для підвищення точності розпізнавання емоцій учасників відеоконференцій пропонується комбінування процесів розпізнавання на основі голосу та виразу обличчя. Розглянемо детальніше основні етапи кожного процесу окремо та їх комбінацію у межах мультимодального підходу.

Розпізнавання голосових емоцій (Speech Emotion Recognition – SER) визначає емоційний зміст голосу незалежно від слів. Процес включає навчання моделі за допомогою машинного або глибокого навчання та тестування у реальному часі. Також, важливими етапами процесу є підготовка даних (фільтрація шуму), вилучення ключових ознак та їх класифікація.

Голосові ознаки поділяються на локальні (короткочасні) та глобальні (довготривалі). SER аналізує їх у чотирьох категоріях: просодичні (ритм, інтонація), спектральні (MFCC, LPCC), ознаки якості

голосу (джиттер, шиммер) та характеристики на основі енергетичного оператора Тігера (ТЕО). Просодичні ознаки особливо важливі для розпізнавання гніву та радості, а спектральні – для визначення голосового тракту [3].

Методи класифікації включають SVM, RF, GMM, KNN та глибокі нейронні мережі (CNN, DNN, LSTM). Традиційні підходи мають обмеження через залежність від створених вручну ознак. Глибоке навчання покращує SER завдяки кращому аналізу складних структур мовлення.

Більшість SER-систем використовує спектрограму MFCC, але цей підхід не завжди точно передає емоційний стан. Проблеми реального застосування включають змінність записів і різні мовні середовища. Покращення можливо шляхом поєднання локальних і глобальних ознак для точнішого розпізнавання емоцій.

Загалом, розпізнавання емоцій учасників відеоконференцій за фізичними характеристиками голосу буде містити такі основні етапи:

1. Попередня обробка (Pre-processing).
 - a. Отримання сигналу. Запис або завантаження аудіо файлу.
 - b. Нормалізація амплітуди. Приведення сигналу до стандартного діапазону.
 - c. Фільтрація шуму: Видалення небажаних частотних компонентів.
 - d. Видалення тиші: Використання Voice Activity Detection (VAD).
 - e. Фреймування: Розділення сигналу на короткі вікна (10–30 мс) з 50% перекриттям.
2. Виділення ознак (Feature Extraction).
 - a. Спектральні ознаки (Spectral Features): MFCC Coefficients, WPT Features, LPCC Coefficients, Spectral Rolloff, Spectral Kurtosis, Spectral Centroid, Formants, Std. Deviation of Formants, Mean of Formants.
 - b. Ознаки часової області (Time Domain Features): Pitch Frequency, Zero-Crossing Rate (ZCR).
 - c. Ознаки якості голосу (Voice Quality Features): Jitter, Shimmer, RMS.
3. Формування вектора ознак.
 - a. Зібрати всі обчислені ознаки в єдиний набір параметрів.
 - b. Використовувати статистичні характеристики (середнє, стандартне відхилення, медіана, мін/макс).
 - c. Передача ознак у модель класифікації.
 - d. Нормалізувати дані.
 - e. Використовувати традиційні (SVM, RF, KNN) або глибокі (CNN, LSTM) моделі для класифікації емоцій.
4. Обробка вектора ознак згортковою нейронною мережею.
5. Вибір найбільш ймовірної емоції.

Загалом, розпізнавання емоцій учасників відеоконференції за виразом обличчя буде містити такі основні етапи [4]:

1. Отримання зображення учасника відеоконференції.
2. Нормалізації зображення.
3. Розпізнавання обличчя за допомогою ознак Хаара.
4. Виділення обличчя із зображення.
5. Обробка зображення обличчя згортковою нейронною мережею.
6. Вибір найбільш ймовірної емоції.

З метою підвищення точності розпізнавання емоцій було об'єднано процес розпізнавання емоцій за характеристиками голосу та процес розпізнавання емоцій на основі виразу обличчя в межах удосконаленої моделі розпізнавання емоцій учасників відеоконференції. Узагальнену схему розпізнавання емоцій учасників відеоконференцій на основі мультимодального підходу, що комбінує процеси розпізнавання на основі голосу та виразу обличчя представлено на рис. 1. Схема містить графічний опис комбінації процесів розпізнавання від моменту отримання медіа даних від учасника відеоконференції, до остаточного отримання кінцевої найбільш ймовірної його емоції у кожен окремий момент часу базуючись на результатах, отриманих з розпізнавання емоцій у голосі та розпізнавання емоцій на обличчі.

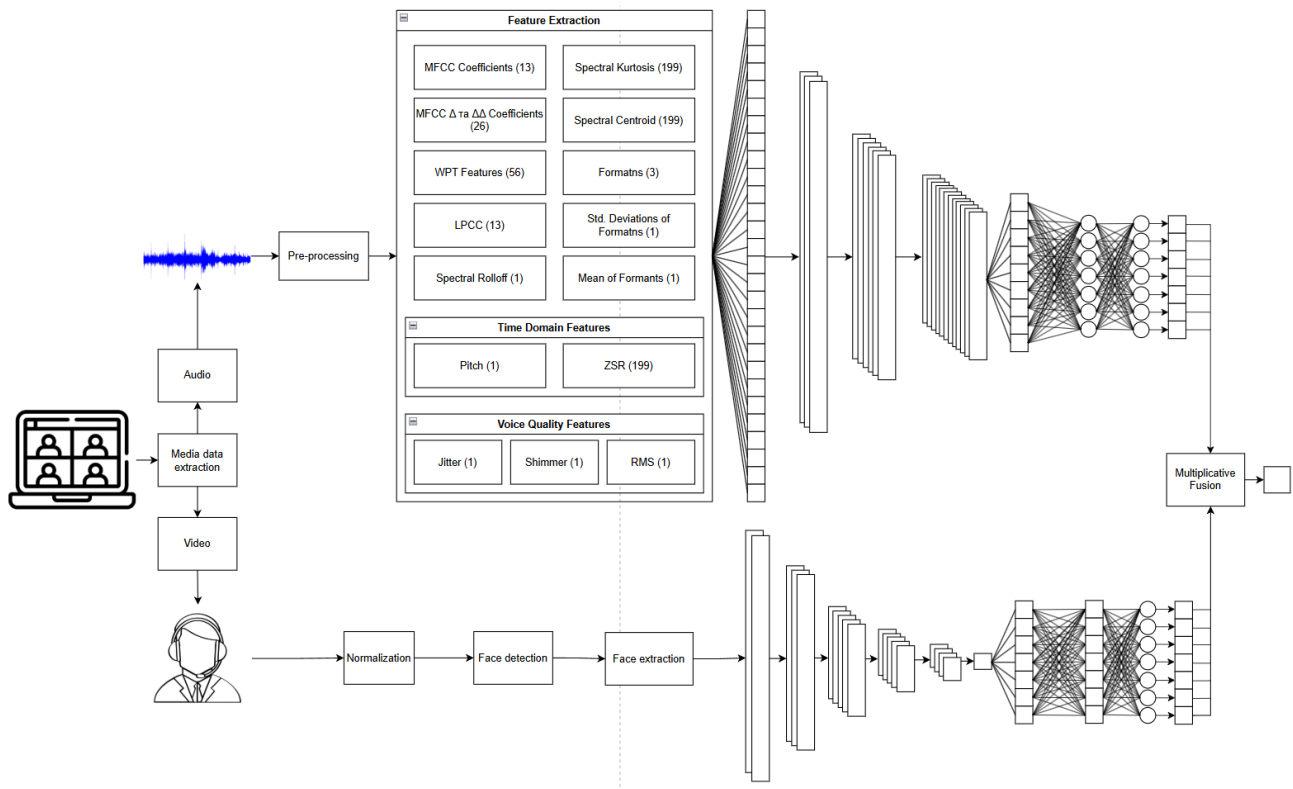


Рисунок 1 – Схема розпізнавання емоцій учасників відеоконференцій на основі мультимодального підходу

Висновки

Таким чином, у даному дослідженні запропоновано удосконалену модель розпізнавання емоцій учасників відеоконференцій за виразом обличчя та характеристиками голосу. Покроково розглянуто основні етапи процесу розпізнавання емоцій на основі голосу, описано список ознак, які виділяються з голосу та передаються у згорткову нейронну мережу у вигляді вектора ознак. Також, покроково розглянуто основні етапи процесу розпізнавання емоцій на основі виразу обличчя. Особливості комбінування вказаних процесів представлено в узагальненій схемі розпізнавання емоцій учасників відеоконференцій на основі мультимодального підходу.

Список використаної літератури

1. Ekman, P. (1999). Basic emotions. Handbook of cognition and emotion, 45-60.
2. Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. American psychologist, 56(3), 218-226.
3. Kishor Bhangale, Mohanaprasad Kothandaraman. Speech Emotion Recognition Based on Multiple Acoustic Features and Deep Convolutional Neural Network. Electronics 2023, 12(4), 839.
4. Савчук Т. О., Пастух І. П. Розпізнавання емоцій учасників відеоконференцій в Microsoft Teams. Таврійський науковий вісник. Серія: Технічні науки. Херсон: Видавничий дім «Гельветика», 2023. Вип. 6. С. 18-24

Пастух Ігор Петрович – аспірант кафедри комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця.

Яровий Андрій Анатолійович – д.т.н., проф., завідувач кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Козловський Андрій Володимирович – к.т.н., доц., доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Ihor P. Pastukh – Post-Graduate Student of the Department for Computer Science, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia.

Andrii A. Yarovy – Dr. Sc. (Eng), Professor, Head of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.

Andrii V. Kozlovskyi – Ph. D. (Eng), Associate Professor of the Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.