SHAXSNI IDENTIFIKATSIYALASHDA YUZ TASVIRI VA NUTQ BELGILARINI MULTIMODAL BIRLASHTIRISH MODELI

Ergasheva Durdona Sherali qizi
Namangan Viloyati Uchqorg'on tumani
20-sonli umumiy o'rta ta'lim maktabi
matematika o'qituvchisi
E-mail: dergasheva209@gmail.com
https://doi.org/10.5281/zenodo.23230393

Annotatsiya. Mazkur maqolada shaxsni avtomatik identifikatsiyalashda yuz tasviri va nutq belgilarini birgalikda qayta ishlashga asoslangan multimodal biometrik model taklif etiladi. An’anaviy bir modal biometrik tizimlarda yuz tasvirining yoritilish, pozitsiya, niqob, kamera sifati kabi omillarga, nutq signalining esa shovqin, kanal xususiyatlari, talaffuz, emotsional holat va mikrofon sifatiga sezgirligi identifikatsiya ishonchliligini pasaytirishi mumkin. Shu sababli maqolada yuz va nutqdan mustaqil biometrik belgilarni ajratib olish, ularni umumiy latent fazoga proyeksiyalash, sifat ko‘rsatkichlari asosida vaznlash va yakuniy qarorni score-level hamda feature-level fusion tamoyillari asosida shakllantirish modeli ishlab chiqiladi. Modelning konseptual arxitekturasi yuzni aniqlash va embedding olish moduli, nutqdan speaker embedding olish moduli, sifatni baholash mexanizmi, multimodal attention/fusion qatlami va identifikatsiya klassifikatoridan tashkil topadi. Taklif qilinayotgan modelning asosiy afzalligi bir modalning sifati pasaygan vaziyatda ikkinchi modal orqali identifikatsiya qarorini barqarorlashtirish imkoniyatidir. ISO/IEC TR 24722:2024 multimodal biometrik birlashtirishda feature, score va decision-level kabi yondashuvlarni ajratib ko‘rsatadi; 2026-yilda ushbu yo‘nalish bo‘yicha yangi qo‘mita loyihasi ham ishlab chiqilmoqda.

Kalit so‘zlar: biometrik identifikatsiya, yuz tasviri, speaker recognition, nutq signali, multimodal biometrika, feature fusion, score fusion, deep learning, embedding, attention, shaxsni aniqlash.

Abstract. This paper proposes a multimodal biometric model for person identification based on the joint analysis of facial images and speech characteristics. Unimodal biometric systems may suffer from reduced reliability due to illumination, pose, occlusion, camera quality, acoustic noise, channel variability, pronunciation, and other environmental factors. The proposed approach extracts independent facial and speech embeddings, projects them into a common latent space, estimates modality quality, and combines the resulting representations using adaptive fusion mechanisms. The conceptual architecture includes a face feature extraction module, a speaker embedding module, quality assessment, multimodal attention-based fusion, and a final identification classifier. The proposed model is designed to increase robustness when one biometric modality is degraded. The study also discusses feature-level and score-level fusion, normalization, decision thresholds, evaluation metrics, privacy, and security considerations.

Keywords: biometric identification, face recognition, speaker recognition, speech features, multimodal biometrics, feature fusion, score fusion, deep learning, embedding, attention.

Аннотация. В статье предлагается мультимодальная биометрическая модель для автоматической идентификации личности, основанная на совместной обработке изображений лица и речевых характеристик. В традиционных унимодальных биометрических системах надежность идентификации может снижаться из-за чувствительности изображений лица к таким факторам, как освещение, ракурс, наличие масок и качество камеры, а также из-за чувствительности речевых сигналов к шуму, характеристикам канала связи, особенностям произношения, эмоциональному состоянию и качеству микрофона. В связи с этим в статье разработана модель, которая извлекает независимые биометрические признаки из изображений лица и речи, проецирует их в общее латентное пространство, взвешивает на основе метрик качества и формирует итоговое решение с использованием принципов объединения данных на уровнях признаков и оценок (score-level fusion). Концептуальная архитектура модели включает модуль обнаружения лица и извлечения векторного представления (эмбеддинга), модуль извлечения эмбеддинга диктора, механизм оценки качества, слой мультимодального внимания/объединения данных и классификатор для идентификации. Основным преимуществом предложенной модели является возможность стабилизации решения об идентификации за счет использования второй модальности в случае снижения качества первой. Стандарт ISO/IEC TR 24722:2024 разграничивает подходы к объединению данных в мультимодальных биометрических системах на уровнях признаков, оценок и решений; в настоящее время также ведется разработка нового проекта стандарта в этой области, запланированного к выпуску в 2026 году.

Ключевые слова: биометрическая идентификация, изображение лица, распознавание диктора, речевой сигнал, мультимодальная биометрия, объединение признаков, объединение оценок, глубокое обучение, эмбеддинг, механизм внимания, идентификация личности.

Kirish. Biometrik texnologiyalar shaxsni uning fiziologik yoki xulq-atvor xususiyatlari asosida aniqlashga imkon beradi. Yuz tasviri va ovoz insonni identifikatsiyalashda keng qo‘llaniladigan ikki muhim biometrik modal hisoblanadi. Yuzni tanish tizimlari vizual belgilarni, speaker recognition tizimlari esa nutq signalidagi shaxsga xos akustik va fonetik xususiyatlarni tahlil qiladi. NIST speaker recognition yo‘nalishida nutq asosidagi identifikatsiya texnologiyalarini muntazam baholash va standartlashtirish bo‘yicha uzoq yillik tadqiqot dasturlarini olib boradi.

Bitta biometrik modalga tayanish tizimni tashqi muhitga nisbatan zaiflashtirishi mumkin. Masalan, yuz tasvirida past yoritilish, yuzning burilishi, qisman yopilishi yoki kamera sifati muammolar keltirib chiqaradi. Nutq signalida esa fon shovqini, reverberatsiya, mikrofonning akustik xususiyatlari, signalning qisqaligi va nutqdagi o‘zgaruvchanlik identifikatsiya aniqligiga ta’sir qilishi mumkin. Shu nuqtayi nazardan, yuz tasviri va nutq belgilarini yagona

modelda birlashtirish shaxsni identifikatsiyalashning ishonchliligi va moslashuvchanligini oshirish uchun istiqbolli yo‘nalish hisoblanadi.

Tadqiqot maqsadi va vazifalari. Tadqiqotning asosiy maqsadi — yuz tasviri va nutq belgilarini birgalikda qayta ishlaydigan, biometrik ma’lumotlarning sifatiga mos ravishda ularning hissasini adaptiv ravishda o‘zgartira oladigan multimodal shaxs identifikatsiyasi modelini ishlab chiqish.

Tadqiqot vazifalari quyidagilardan iborat:

1. yuz tasviridan shaxsga xos embedding vektorini olish;
2. nutq signalidan speaker embedding hosil qilish;
3. har ikkala modalning sifat ko‘rsatkichlarini aniqlash;
4. turli modal belgilarini umumiy latent fazoga keltirish;
5. feature-level va score-level fusion usullarini taqqoslash;
6. attention mexanizmi asosida modal vaznlarini adaptiv aniqlash;
7. yakuniy identifikatsiya qarorini shakllantirish;
8. model samaradorligini Accuracy, FAR, FRR, EER va F1 kabi ko‘rsatkichlar yordamida baholash.

Nutq moduli audio signalni dastlabki qayta ishlash, akustik xususiyatlarni ajratish va speaker embedding olish bosqichlaridan tashkil topadi. An’anaviy tizimlarda MFCC, log-Mel spectrogram, pitch, formant va boshqa akustik belgilar ishlatilishi mumkin. Zamonaviy chuqur o‘rganish tizimlarida esa nutqdan bevosita speaker embedding olishga mo‘ljallangan neyron arxitekturalardan foydalanish mumkin.

Modelning asosiy yangiligi — yuz va nutq belgilarini oddiy qo‘shish o‘rniga ularning sifatini hisobga olgan holda adaptiv birlashtirishdir. Ushbu mexanizmning muhim jihati shundaki, tizim har doim yuz va nutqqa bir xil vazn bermaydi. Qaror modal ma’lumotlarning joriy sifatiga mos ravishda shakllantiriladi.

Identifikatsiya jarayonining algoritmik ketma-ketligi. Taklif etilayotgan algoritm quyidagi bosqichlarda ishlaydi:

1. kamera va mikrofon orqali foydalanuvchidan ma’lumot olish;
2. yuzni aniqlash;
3. audio signalni ajratish;
4. yuz tasvirining sifatini baholash;
5. nutq signalining sifatini baholash;
6. yuz embeddingini hosil qilish;
7. speaker embeddingini hosil qilish;
8. embeddinglarni umumiy fazoga proyeksiyalash;
9. modal sifat koeffitsiyentlarini hisoblash;
10. attention yoki sifatga asoslangan fusionni amalga oshirish;
11. ma’lumotlar bazasidagi shaxs embeddinglari bilan taqqoslash;
12. yakuniy identifikatsiya skorini hisoblash;
13. belgilangan threshold asosida “mos” yoki “mos emas” qarorini chiqarish.

Modelni tadqiq qilish uchun multimodal datasetda har bir shaxsga kamida ikki turdagi biometrik ma’lumot biriktiriladi:

Shaxs ID	Yuz tasviri	Nutq namunasi	Sifat ko‘rsatkichi
P01	mavjud	mavjud	yuqori
P02	mavjud	mavjud	o‘rta
P03	mavjud	mavjud	yuqori
P04	mavjud	mavjud	past

Ma’lumotlar bazasi quyidagi qismlarga ajratilishi maqsadga muvofiq:

- training set — modelni o‘qitish;
- validation set — giperparametrlarni tanlash;
- test set — yakuniy baholash.

Bir shaxsning o‘qitish va test namunalarining bevosita takrorlanib qolishi model samaradorligini sun’iy ravishda oshirib yuborishi mumkin. Shu sababli identifikatsiya tajribasida ma’lumotlarni shaxslar yoki sessiyalar bo‘yicha to‘g‘ri ajratish muhim.

Multimodal tizimlar barcha muammolarni avtomatik ravishda hal qilmaydi. Asosiy cheklovlar quyidagilardan iborat:

- yuz tasviri va nutqni bir vaqtning o‘zida olish zarurati;
- hisoblash resurslariga talabning oshishi;
- ikki xil biometrik ma’lumotni saqlash bilan bog‘liq maxfiylik muammolari;
- audio shovqini;
- yuzning qisman yopilishi;
- deepfake yoki presentation attack xavfi;
- turli modal ma’lumotlarning vaqt bo‘yicha mos kelmasligi;
- modelning ma’lum demografik yoki akustik guruhlarda turlicha ishlashi ehtimoli.

Biometrik ma’lumotlar oddiy paroldan farqli ravishda shaxsga doimiy bog‘langan bo‘lishi sababli ularni himoyalash alohida ahamiyatga ega.

Tizimda quyidagi tamoyillarni qo‘llash maqsadga muvofiq:

1. xom yuz tasvirlarini imkon qadar markaziy serverda saqlamaslik;
2. xom audio yozuvlarini cheklangan muddat saqlash;
3. embeddinglarni shifrlangan ko‘rinishda saqlash;
4. kirish huquqlarini rollar asosida boshqarish;
5. biometrik ma’lumotdan foydalanish maqsadini aniq belgilash;
6. foydalanuvchining roziligi va qonuniy asosni ta’minlash;
7. presentation attack detection mexanizmlaridan foydalanish.

Taklif qilinayotgan yondashuvning ilmiy yangiligi quyidagi konseptual yechimlarda namoyon bo‘ladi:

- yuz va nutq embeddinglarini umumiy latent fazoga keltirish;
- modal sifatini alohida baholash;
- sifatga bog‘liq adaptiv fusion koeffitsiyentlarini hisoblash;
- attention asosida modal hissasini dinamik o‘zgartirish;
- feature-level va score-level fusionni yagona identifikatsiya tizimida taqqoslash;
- multimodal identifikatsiyani xavfsizlik va maxfiylik mexanizmlari bilan birgalikda ko‘rib chiqish.

Tadqiqot muhokamasi. Yuz tasviri va nutq belgilarining kombinatsiyasi shaxsni identifikatsiyalashda o‘zaro to‘ldiruvchi axborotdan foydalanish imkonini beradi. Yuz tasviri vizual morfologik belgilarni, nutq esa ovoz apparati va nutq ishlab chiqarish jarayoniga bog‘liq akustik belgilarni ifodalaydi. Shu sababli ularning xatoliklari qisman mustaqil bo‘lishi mumkin.

Masalan, qorong‘i muhitda yuz tasvirining sifati pasaysa, model nutq signaliga ko‘proq vazn berishi mumkin. Aksincha, kuchli fon shovqinida yuz tasviri asosiy axborot manbasiga aylanishi mumkin. Demak, multimodal tizimning asosiy afzalligi shunchaki ikki xil ma’lumotni qo‘shish emas, balki ularning ishonchliligini dinamik boshqarish bilan bog‘liq.

Kelgusidagi tadqiqotlarda modelni real multimodal ma’lumotlar bazasida sinash, turli yoritilish va shovqin darajalarida baholash, deepfake va presentation attacklarga chidamliligini tekshirish, demografik guruhlar bo‘yicha fairness tahlilini o‘tkazish hamda yuz va nutq encoderlarini multimodal kontrastiv o‘qitish asosida optimallashtirish maqsadga muvofiq.

Xulosa. Shaxsni identifikatsiyalashda yuz tasviri va nutq belgilarini multimodal birlashtirish modeli biometrik tizimlarning barqarorligi va moslashuvchanligini oshirish uchun istiqbolli yo‘nalish hisoblanadi. Tadqiqotda yuz va nutqdan mustaqil embeddinglar olish, ularni umumiy latent fazoga proyeksiyalash, modal sifatini baholash va adaptiv fusion yordamida yakuniy identifikatsiya qarorini shakllantirishga asoslangan model taklif qilindi.

Modelning muhim jihati — yuz yoki nutq signalining sifati pasaygan sharoitda ikkinchi modalning nisbiy ahamiyatini oshirish imkoniyatidir. Feature-level, score-level va attention-based fusion usullarining kombinatsiyasi tizimni turli amaliy sharoitlarga moslashtirish imkonini beradi.

Foydalanilgan adabiyotlar

1. Sadjadi O., Greenberg C., Singer E., Mason L., Reynolds D. NIST 2021 Speaker Recognition Evaluation Plan. NIST, 2021.
2. Watson C. I., Ulery B., Hicklin R. A., Fellner W., Halinan P. Studies of Biometric Fusion: Executive Summary. NIST, 2004.
3. Snelick R. D., Indovina M. D., Yen J. H., Mink A. Multimodal Biometrics: Issues in Design and Testing. NIST, 2003.
4. O'Toole A. J., Phillips P. J., Weimer S. et al. Recognizing people from dynamic and static faces and bodies: Dissecting identity with a fusion approach. Vision Research, 2010.
5. ISO/IEC JTC 1/SC 37. Biometrics — Standards catalogue and current work programme. ISO, 2026.