

ЭТАПЫ И МЕТОДЫ РАСПОЗНАВАНИЯ ЖЕСТОВ БЕЛОРУССКОГО ЖЕСТОВОГО ЯЗЫКА

Ковалева-Евсюченя А.В.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Бранцевич П.Ю. – канд. техн. наук, доцент

В данной работе рассмотрены этапы и современные методы распознавания жестов. Выбраны технологии для реализации программного средства для распознавания жестов белорусского жестового языка.

Речь – это неотъемлемая часть жизни любого человека. При помощи речи мы каждый день общаемся со своими близкими, выражаем свое мнение и взаимодействуем с окружающим миром. Однако, более 450 миллионов людей в мире страдают от нарушений слуха и полной глухоты и большинство из них не может использовать речь как основное средство коммуникации [1]. Для этой цели они используют жестовые языки. Так как слышащие люди редко владеют жестовыми языками, для коммуникации с ними глухие люди вынуждены прибегать к услугам сурдопереводчиков.

В Беларуси на законодательном уровне закрепили понятие «белорусский жестовый язык» в 2022 году [2]. Государство способствует подготовке сурдопереводчиков и повышению их квалификации, но сейчас существует нехватка специалистов – в реестре переводчиков жестового языка Республики Беларусь зарегистрирован 21 человек [3]. Поэтому, сейчас есть потребность в технологии распознавания белорусского жестового языка, которая сможет помочь в обучении новых специалистов и в будущем дополнить живой сурдоперевод.

Жестовое общение использует верхнюю часть туловища, включая жесты руками, мимику, чтение по губам и положение тела в пространстве. Распознавание жестового языка, основанного на образах, включает в себя пять этапов: получение изображений для составления обучающей выборки; предварительную обработку полученных изображений; сегментацию изображений; извлечение признаков; классификацию образа.

Для получения изображений для распознавания используются различные устройства. Самое широко используемое устройство - веб-камера. Веб-камеру легко использовать, и она встроена в многие устройства. Самым точным устройством считается специальная перчатка, но она обладает высокой стоимостью и может быть некомфортна в использовании. Контроллер Kinect позволяет получать изображение и одновременно траектории движения рук, но он также обладает высокой стоимостью [4].

Предварительная обработка изображений осуществляется с целью убрать посторонние шумы и улучшить качество изображения. Применяемые подходы для улучшения качества изображения включают в себя выравнивание гистограммы (HE), адаптивное выравнивание гистограммы (AHE), контрастно-ограниченное выравнивание гистограммы (CLAHE) и логарифмическое преобразование. Сравнительный анализ методов улучшения качества изображений приведен в таблице 1. Используемые алгоритмы шумоподавления включают в себя медианный фильтр, фильтр Гаусса, фильтр Винера, адаптивный фильтр и фильтр скользящего среднего. Правильный выбор методов обработки изображений является ключевым для корректного распознавания жеста [5].

Таблица 1 – Сравнительный анализ методов улучшения качества изображений.

Метод улучшения качества изображений	Преимущества	Недостатки
Выравнивание гистограммы	Простота реализации. Эффективность при использовании на черно-белых изображениях	Изменение яркости изображений. Повышение контраста посторонних шумов
Адаптивное выравнивание гистограммы	Усиление локального контраста во всех областях изображений	Потеря исходного уровня яркости изображений. Усиление посторонних шумов
Контрастно-ограниченное выравнивание гистограммы	Низкий уровень шумов в изображениях. Сохранение исходной яркости изображений	Потеря информации при использовании изображений с повышенной яркостью
Логарифмическое преобразование	Сжатие динамического диапазона изображений, имеющих большие вариации в значениях пикселей	Потеря информации при использовании изображений, на которых большая часть

		деталей сосредоточена в светлых областях
--	--	--

Сегментация изображений – это разделение изображений на значимые сегменты. Для сегментации изображений используются два подхода: контекстуальный и неконтекстуальный [6]. Контекстуальная сегментация использует отношения между признаками изображения, такими как расстояние между признаками на изображении, контрастность сегментов изображения, границы признаков и другие. Неконтекстуальная сегментация игнорирует признаки отдельных областей изображения и осуществляет разделение, используя глобальные атрибуты изображения. Подробная классификация методов сегментации изображения показана на рисунке 1.



Рисунок 1 – Классификация методов сегментации изображения

На этапе извлечения признаков, для увеличения точности распознавания, выделяются самые релевантные признаки изображения, которые представляют собой отличия между классами распознавания программного средства [7]. Результатом выполнения данного этапа является вектор признаков, который в дальнейшем используется этапом классификации. Важные методы извлечения признаков включают в себя метод главных компонент (PCA), Фурье-дескрипторы (FD), гистограмму направленных градиентов (HOG), масштабно-инвариантную трансформацию признаков (SIFT), ускоренные устойчивые функции (SURF).

При классификации образа, изображение, не входящее в обучающую выборку, поступает в алгоритм классификации. Задачей алгоритма классификации является указание наименования класса на основе имеющегося вектора признаков. Среди существующих классификаторов, алгоритм k-ближайших соседей (KNN), искусственная нейронная сеть (ANN), метод опорных векторов (SVM), скрытая марковская модель (HMM) и сверточная нейронная сеть (CNN) являются основными [8].

Согласно рассмотренным этапам распознавания, оптимальным выбором классификатора для создания программного средства для распознавания жестов белорусского жестового языка является сверточная нейронная сеть в сочетании с масштабно-инвариантной трансформацией признаков. SIFT-CNN автоматически распознает релевантные признаки образа без вмешательства программиста и выполняет классификацию жеста с высокой точностью. Сегментация на основе искусственной нейронной сети позволяет получить качественное разделение на сегменты с низким уровнем шума, несмотря на большую вычислительную сложность. Для получения изображения для распознавания используется веб-камера как самое доступное устройство, с применением фильтра Гаусса для шумоподавления.

Список использованных источников:

1. World report on hearing process / Department of Noncommunicable Diseases ; ed.: Jackie Clark [et al.]. – Geneva : WHO, 2021. – 252 p.
2. О правах инвалидов и их социальной интеграции [Электронный ресурс] : Закон Республики Беларусь, 30 июня 2022 г., № 183-З // Национальный правовой Интернет-портал Республики Беларусь. – Режим доступа: <https://pravo.by/document/?guid=12551&p0=H12200183&p1=1>. – Дата доступа: 10.04.2023.
3. Реестр переводчиков жестового языка Республики Беларусь [Электронный ресурс]. – Режим доступа: <http://belog.org/reestr-perevodchikov-belorusskogo-zhestovogo-yazyka/reestr-perevodchikov-bzhya>. – Дата доступа: 10.04.2023.
4. Kamal, S. M. Technical Approaches to Chinese Sign Language Processing: A Review / Kamal, S. M. [et al.] // IEEE Access. – 2019. – Vol. 7. – P. 96926–96935.
5. Majeed, S. H. Adaptive Entropy Index Histogram Equalization for Poor Contrast Images / Majeed, S. H., Isa, N. A. M. // IEEE Access. – 2021. – Vol. 9. – P. 2169–3536.

59-я научная конференция аспирантов, магистрантов и студентов БГУИР

6. Xu, W. *A novel star image thresholding method for effective segmentation and centroid statistics* / Xu, W. [et al.] // *Optik*. – 2013. – Vol. 124. – Iss. 20. – P. 4673–4677.
7. Patel, J. M. *A review on feature extraction techniques in Content Based Image Retrieval* / Patel, J. M., Gamit, N. C. // *Proc. IEEE Intern. Conf. on Wireless Communications, Signal Processing and Networking*. – Chennai, India, 2016. – P. 2259–2263.
8. Gopika, P. *Single-layer convolution neural network for cardiac disease classification using electrocardiogram signals* / Gopika, P. [et al.] // *Deep Learning for Data Analytics* / Gopika, P. [et al.]. – Cambridge : Academic Press, 2020. – Chap. 2. – P. 21–35.