УДК 004.93’11

МУЛЬТИМОДАЛЬНАЯ СИСТЕМА ДЛЯ АУДИОВИЗУАЛЬНОЙ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПОЛЬЗОВАТЕЛЯМИ С НАРУШЕНИЕМ ЗРЕНИЯ

И. П. Бурукина, к.т.н., доцент, заведующий кафедрой САПР ПГУ, Пенза, Россия;
orcid.org/0009-0006-1953-2914, e-mail: burukinairina@gmail.com
Г. О. Фельдман, бакалавр ПГУ, Пенза, Россия;
e-mail: gl.feldman2018@yandex.ru
Д. А. Гришаев, бакалавр ПГУ, Пенза, Россия;
e-mail: dima_grishaev28@mail.ru

Работа посвящена решению проблемы пространственной ориентации лиц с нарушением зрения. Цель исследования заключается в разработке мультимодальной системы, объединяющей алгоритмы сегментации изображений и методы нейронных сетей для преобразования визуальных сигналов в детальное звуковое представление окружающего пространства. Система успешно прошла комплексное техническое тестирование, показав высокие функциональные и эксплуатационные характеристики. Основными достоинствами являются высокая точность распознавания объектов, определение достоверной дистанции, быстрота реакции и низкое энергопотребление. Исследование играет важную роль в развитии отечественных технологий реабилитации, нацеленных на обеспечение доступности инфраструктуры и повышение самостоятельности людей с нарушением зрения.

***Ключевые слова**: система, нарушение зрения, сегментация, нейронная сеть, пространственная ориентация, тестирование, эффективность.*

DOI: 10.21667/1995-4565-2026-95-226-235

Введение

Люди с нарушением зрения ежедневно сталкиваются с множеством трудностей, значительно осложняющих свободное передвижение и способность полноценно ориентироваться в окружающей среде. Основной проблемой является невозможность точно оценить расположение предметов и дистанции до них [1], что создает риск случайных столкновений с препятствиями.

За последнее десятилетие предложено несколько интересных разработок, направленных на улучшение условий передвижения людей с нарушением зрения [2]. Среди наиболее известных решений можно выделить следующие проекты:

– система *Lookout*, представленная корпорацией *Google*, интегрируется с камерой мобильного устройства и автоматически описывает обстановку вокруг пользователя, выделяя важные элементы, такие как ступеньки, дорожные знаки, предметы мебели, и проговаривая описание вслух. Существенным минусом данной системы является отсутствие детального позиционирования объектов относительно пользователя, что затрудняет точное определение дистанции и угла расположения предмета;

– приложение *Seeing Al* от корпорации *Microsoft* также основано на анализе видеокадра и озвучивании происходящего. Оно способно считывать текст, распознавать лица знакомых людей, определять денежные купюры и многое другое. Недостатком программы являются узкая специализация функций и сложность быстрого освоения для пожилых людей или тех, кому непросто пользоваться смартфоном;

– головные очки *eSight* используют специальные линзы и камеру, транслируя увеличенное и четкое изображение на дисплее внутри очков. Хотя продукт помогает пользователям с нарушением зрения, он имеет серьезные ограничения по стоимости.

Каждый из перечисленных продуктов, наряду с высоким потенциалом применяемых цифровых технологий, обладает определенными недостатками [3, 4]. Следовательно, возникает потребность в разработке новой системы, учитывающей специфику целевой аудитории и российского рынка. Система должна обладать возможностью локализации всех компонент интерфейса, поддерживать взаимодействие с пользователями на русском языке и иметь умеренную цену.

Постановка задачи

Целью исследования является разработка и апробация системы аудиовизуальной поддержки принятия решений людьми с нарушением зрения, основанной на принципах интеграции мультимодальных технологий обработки изображений и синтетического звучания и направленной на существенное повышение автономности, мобильности и общего качества жизни данной категории пользователей. Важнейшими характеристиками системы выступают простота использования, надежность функционирования и гибкость настройки под индивидуальные предпочтения каждого потребителя.

Для достижения обозначенной цели сформулированы следующие задачи:

– анализ состояния вопроса и определение основных направлений исследования;

– теоретическое обоснование метода сегментирования изображений;

– внедрение технологий нейронной сети для распознавания образов и генерации звуковых подсказок;

– проектирование и разработка мультимодальной системы;

– экспериментальная проверка работоспособности и оценка эффективности системы.

Теоретические исследования

Процесс сегментации изображения является важным моментом обработки визуальной информации в разработанной системе аудиовизуальной поддержки [5].

Изображения, полученные с камер телефонов, часто содержат значительные шумы, вызванные ограничениями аппаратуры и условиями съемок. Поэтому первым этапом является применение гауссового фильтра для сглаживания высокочастотных помех и уменьшения влияния случайных шумовых артефактов [6]. Операция свертки с ядром Гаусса исходного изображения *I(x, y)* определяется выражением:

$$I_{smooth}(x,y) = I(x,y) * G_{\sigma}(x,y),$$

где I_{smooth} – итоговое изображение; $G_{\sigma}(x,y)$ – двумерное гауссово распределение с дисперсией σ.

Причем само ядро задается формулой:

$$G_{\sigma}(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{x^2+y^2}{2\sigma^2}\right).$$

Следующим шагом является выделение границ объектов на изображении алгоритмом Канни (*Canny Edge Detector*). Несмотря на то, что данный алгоритм включает собственную процедуру гауссовой фильтрации, предварительное сглаживание снижает риск формирования ложных пограничных областей, обусловленных изменениями яркости, и одновременно усиливает выраженность реальных контрастных переходов, что значительно повышает точность выделения границ объектов [6].

Классический оператор Канни определяет «сильные» границы путем анализа изменения интенсивностей пикселей по контору изображений:

$$\left|\nabla I\right| = \sqrt{I_x^2 + I_y^2}\ ,\ \theta = arctan\frac{I_y}{I_x},$$

где $\left|\nabla I\right|$ – модуль градиента; θ – угол наклона края.

Для сохранения «сильных» границ применяется пороговая бинаризация – устанавливаются значения верхнего t_h и нижнего t_l порога, и пиксели с величиной выше верхнего порога признаются частью границы:

$$E(x,y)=\begin{cases}1 & \left|\nabla I(x,y)>t_h\right| \\ 0 & \left|\nabla I(x,y)>t_l\right| \\ \text{не определено} & t_l \le \left|\nabla I(x,y)>t_h\right|\end{cases}.$$

Для разделения выявленных объектов используется метод кластеризации *Mean Shift*, основанный на сходстве атрибутов пикселей:

$$m(x)=\frac{\sum_y K(\left|x-y\right|)\cdot y}{\sum_y K(\left|x-y\right|)},$$

где $K(\,)$ – ядро распределения.

Полученные сегменты подвергаются дальнейшему анализу, включая измерение геометрических параметров (размеры, форма) [7], после чего определяется пространственное положение объекта относительно его центра тяжести:

$$P_{obj}=\frac{1}{\left|S\right|}\sum_{(x,y)\in S}(x,y),$$

где $\left|S\right|$ – количество пикселей в сегменте; P_{obj} – центр тяжести сегмента.

Однако алгоритм сегментации, эффективно выделяя границы объекта на изображениях [8], не позволяет устанавливать, какой именно объект обнаружен [9]. Для идентификации объектов используется сверточная нейронная сеть (*Convolutional neural network, CNN*), предварительно обученная на больших выборках изображений и находящаяся в открытом доступе [10]. Между процедурой сегментации и функционированием нейронной сети установлена тесная взаимосвязь. Во-первых, сегментация позволяет избавиться от информационной избыточности, неизбежно присутствующей в полнокадровых изображениях. Представление объекта отдельно от фона существенно уменьшает нагрузку на нейронную сеть, позволяя концентрироваться исключительно на релевантных характеристиках объекта, что заметно повышает точность и быстроту распознавания. Во-вторых, сегментация дает возможность нормализовать и стандартизировать выделенные объекты перед подачей на вход нейронной сети. Приведение всех объектов к одному формату позволяет минимизировать влияние условий съемки и индивидуальных особенностей отдельных кадров. В-третьих, нейронные сети, работающие с отдельными объектами, требуют меньших ресурсов и могут эффективно работать даже на маломощных смартфонах. После завершения этапа сегментации подготовленный фрагмент изображения подается на вход *CNN*, который осуществляет внутренний анализ признаков и присваивает каждому объекту соответствующий класс согласно результатам вероятностной оценки.

Информация об идентифицированных объектах и расстоянии используется для формирования звуковых сигналов. Озвучивание реализуется программно на стороне смартфона встроенными интерфейсами для синтеза речи (*application programming interface, API*), которые обеспечивают стабильную и качественную работу вне зависимости от условий эксплуатации.

Таким образом, применение комбинации алгоритма сегментации и нейросетевых технологий обеспечивает качественное преобразование визуальной информации в доступные звуковые сигналы, позволяя человеку с нарушением зрения «услышать» расположение окружающих предметов.

Экспериментальные исследования

В интегрированной среде *Delphi* с использованием языка программирования *Object Pascal* реализована главная форма мультимодальной системы со специализированными компо-

нентами, предназначенными для захвата видеопотока с камеры смартфона [11]. Для управления функционалом программы разработан обработчик событий (*event handler*) [12], выполняющий последовательные действия:

– захват текущего кадра с камеры;
– выполнение процедур сегментации и идентификации объектов на изображении;
– генерирование звуковых сигналов;
– передачу голосовых подсказок через интерфейс вывода смартфона.

Алгоритм сегментации полученных изображений программно реализуется посредством классификации каждого пикселя в зависимости от уровня его яркости: темные тона автоматически преобразуются в черный цвет, а светлые – в белый. Процедура выполняется последовательно для всех пикселей кадра, создавая бинарное представление исходного изображения (рисунок 1). Данный подход основан на концепции обработки графической информации и является эффективным инструментом для извлечения доминантных цветов и оценки расстояния до объектов.

```
void blacknwhite(int x, int y, int sizex, int sizey, HDC hdc, int porog) {
    std::vector<COLORREF> colors;     // Вектор для хранения цветов пикселей

    for (int i = 0; i < sizex; i++) {
        for (int j = 0; j < sizey; j++) {
            COLORREF color = GetPixel(hdc, x + i, y + j);
            if (color != 0x00ffffff) {
                colors.push_back(color);   // Сохраняем цвет в вектор

                int sumcolor = 0;
                for (int k = 0; k < 3; k++) {
                    COLORREF onecolor = color & 0x0000ff;
                    sumcolor += onecolor;
                    color = color >> 8;
                }

                if (sumcolor / 3 < porog) {
                    SetPixel(hdc, x + i, y + j, 0x00000000);
                }
                else {
                    SetPixel(hdc, x + i, y + j, 0x00ffffff);
                }
            }
        }
    }

}
```

Рисунок 1 – Фрагмент программного кода реализации алгоритма сегментации
Figure 1 – Fragment of program code for segmentation algorithm implementation

Для распознавания конкретных объектов используются библиотека *PyTorch* на языке *Python* и архитектура *Faster R-CNN* [13]. Полученные результаты обработаны и систематизированы в специализированной коллаборационной среде разработки *Google Colaboratory* (рисунок 2).

Для преобразования информации о выявленных объектах в аудиосигналы используется библиотека *pyttsx* [14]. Данная технология позволяет озвучивать предварительно обнаруженные элементы сцены, преобразуя текстовые описания объектов в голосовую форму (рисунок 3).

Пользователь может воспользоваться программой, запустив ее на своем мобильном устройстве и направив встроенную камеру смартфона на интересующий объект (рисунок 4).

После короткого периода обработки (2 – 3 секунды) приложение выдаст подробное описание окружения в форме звукового сообщения, доступного через встроенные динамики телефона или подключенные наушники. Важной особенностью является вариативность тона голоса: близкие объекты озвучиваются высоким голосом, а отдаленные – низким, что создает дополнительное восприятие пространства расстояния.

При участии 48 взрослых пользователей с различными степенями нарушения зрения проведена оценка эффективности разработанной мультимодальной системы с точки зрения повышения уровня автономности и самостоятельности респондентов. Эксперимент проводился на городских улицах в дневное и вечернее время, а также внутри помещений с искусственным освещением. Первый этап – передвижение без вспомогательных устройств, второй

этап – использование смартфона с установленной программой. Каждый этап длился один час и состоял из серии практических заданий: ориентирование в пространстве, определение местоположения предметов и распознавание объектов.

```
import torch
import torchvision
from torchvision import transforms as T  # Импорт трансформаций для преобразования изображений

from PIL import Image
import cv2  # Библиотека для обработки
from google.colab.patches import cv2_imshow  # Специальная функция отображения изображений в Colab

model.eval() # Переключение модели в режим оценки (выключение обучения и градиентов)

!wget 'test.jpg' # Скачиваем тестовое изображение

img = Image.open("/content/test.jpg") # Загружаем изображение

transform = T.ToTensor() # Преобразуем изображение в формат тензора
img = transform(img)

with torch.no_grad(): # Оборачиваем в список и передаем в модель через no_grad (для экономии памяти)
    pred = model([img])

print(pred[0].keys()) # Посмотрим ключи результата

boxes, labels, scores = pred[0]["boxes"], pred[0]["labels"], pred[0]["scores"] # Выделяем предсказанные боксы, метки и вероятности

num = torch.sum(scores > 0.9).item() # Считаем количество предсказаний с уверенностью выше 0.9
print(f"Количество надёжных объектов: {num}")

# Список классов COCO dataset для подписей
coco_names = ["person", "bicycle", "car", "motorcycle", "airplane", "bus", "train", "truck",
              "boat", "traffic light", "fire hydrant", "street sign", "stop sign", "parking meter",
              "bench", "bird", "cat", "dog", "horse", "sheep", "cow", "elephant", "bear", "zebra",
              "giraffe", "hat", "backpack", "umbrella", "shoe", "eye glasses", "handbag", "tie",
              "suitcase", "frisbee", "skis", "snowboard", "sports ball", "kite", "baseball bat",
              "baseball glove", "skateboard", "surfboard", "tennis racket", "bottle", "plate",
              "wine glass", "cup", "fork", "knife", "spoon", "bowl", "banana", "apple", "sandwich",
              "orange", "broccoli", "carrot", "hot dog", "pizza", "donut", "cake", "chair", "couch",
              "potted plant", "bed", "mirror", "dining table", "window", "desk", "toilet", "door",
              "tv", "laptop", "mouse", "remote", "keyboard", "cell phone", "microwave", "oven",
              "toaster", "sink", "refrigerator", "blender", "book", "clock", "vase";, "scissors",
              "teddy bear", "hair drier", "toothbrush", "hair brush"]

font = cv2.FONT_HERSHEY_SIMPLEX  # Шрифт для подписей

img_cv = cv2.imread("/content/test.jpg") # Загружаем исходное изображение в формате OpenCV (для рисования)

for i in range(num): # Проходимся по каждому доверительному объекту и рисуем на изображении bbox и название класса
    x1, y1, x2, y2 = boxes[i].cpu().numpy().astype("int")  # Координаты бокса
    class_name = coco_names[labels[i].item() - 1]  # Имя класса, индексы COCO начинаются с 1
    cv2.rectangle(img_cv, (x1, y1), (x2, y2), (0, 255, 0), 1) # Рисуем зеленый прямоугольник bbox
    cv2.putText(img_cv, class_name, (x1, y1 - 10), font, 0.5, (255, 0, 0), 1, cv2.LINE_AA)  # Подписываем объект синим цветом над bbox

cv2_imshow(img_cv) # Показываем конечное изображение с отрисованными объектами в Colab

print(labels) # Выводим списки меток и боксов
print(boxes)
```

Рисунок 2 – Фрагмент программного кода реализации алгоритма Faster R-CNN
Figure 2 – Fragment of program code for Faster R-CNN algorithm implementation

```
engine = pyttsx3.init() # Инициализация синтезатора речи
engine.setProperty('rate', 150)  # скорость речи
engine.setProperty('volume', 1)  # громкость
```

Рисунок 3 – Инициализация синтезатора речи для озвучивания объектов
Figure 3 – Initialization of the speech synthesizer for voicing objects

Целью эксперимента было сравнение результатов выполнения заданий участниками на разных этапах. В качестве объективных критериев выбраны скорость перемещения, количество допущенных ошибок и общее время выполнения заданий [15]. Обработка собранных данных проводилась методами статистического анализа. Субъективные критерии, отражающие восприятие комфорта и удобства использования программного продукта, определялись анкетированием с вопросами по пятибалльной шкале оценивания Лайкерта (от полного несогласия до полного согласия).

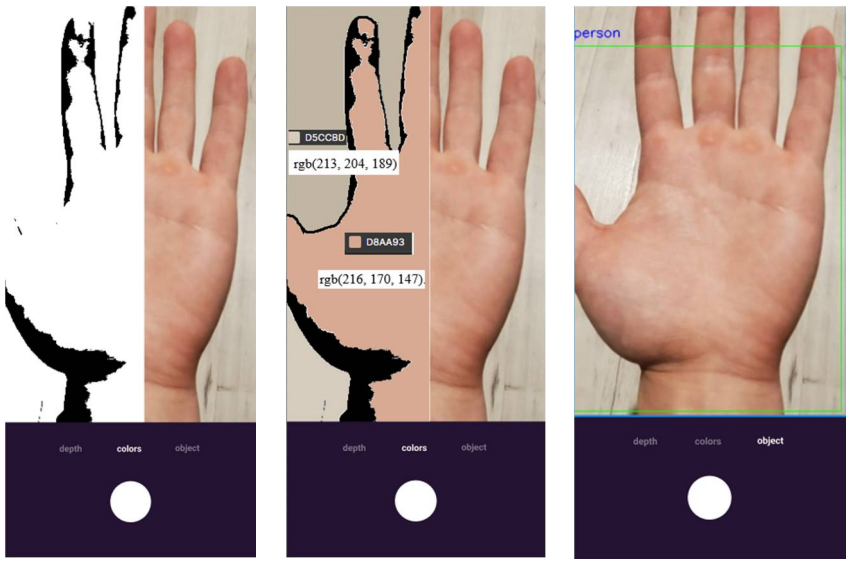

Рисунок 4 – Демонстрация работы приложения на экране смартфона
Figure 4 – Demonstration of the application on smartphone screen

Полученные результаты (таблица 1) позволили сделать вывод об улучшении качества взаимодействия пользователей с окружающей средой при использовании мультимодальной системы.

Таблица 1 – Результаты эксперимента
Table 1 – Experimental results

Критерий оценки	Результат без использования системы	Результат с использованием системы	Изменение	P-value
Скорость перемещения (см/сек)	28,5 ± 3,2	32,8 ± 2,9	+14 %	p = 0,004
Количество допущенных ошибок (%)	16,2 ± 4,1	12,4 ± 3,7	–24 %	p = 0,012
Время выполнения заданий (сек)	123,4 ± 18,61	108,7 ± 15,21	–12 %	p = 0,008
Комфорт работы с системой (баллы)	–	4,5 ± 0,7	–	–
Удобство использования (баллы)	–	4,2 ± 0,6	–	–

Повышение скорости передвижения и снижение количества совершаемых ошибок при идентификации объектов подтверждены статистически значимым уровнем ниже порогового значения ($p < 0,05$), положительное восприятие системы среди пользователей доказано высокими оценками комфорта и удобства использования.

Для проверки работоспособности и устойчивости, а также сбора количественных и качественных показателей оценки производительности организовано три этапа технического тестирования мультимодальной системы.

Первый этап предполагает модульное тестирование отдельных компонентов системы.

Камера проверяется на уровень съемки и устойчивость к изменениям освещения и погодным условиям. В качестве основных критериев оценки уровня съемки выбраны показатель пикового значения сигнал-шум (*Peak Signal To Noise Ratio, PSNR*) и индекс подобия структур (*Structural Similarity Index Measure, SSIM*). Показатель *PSNR* вычисляет отношение среднего уровня сигнала к уровню шума, причем высокие значения указывают на меньшие искажения. Показатель *SSIM* измеряется в диапазоне [0, 1], где единица соответствует лучшему пониманию структур на изображении. Устойчивость к изменению уровня освещения

протестирована на диапазоне яркости от 10 люкс до 5000 люкс. Чувствительность к природным осадкам рассмотрена с точки зрения снижения контрастности и резкости изображения.

Алгоритм сегментации оценивается по точности границ объектов через коэффициент Джаккара (*Jaccard index*) и показатель точности (*Boundary recall*). Коэффициент Джаккара определяет степень перекрытия между двумя областями и варьируется в пределах от 0 до 1, где значение 1 указывает на идеальное совпадение выделенных областей. Показатель *Boundary recall,* определяемый как доля верно обнаруженных границ среди общего числа реальных границ, также лежит в интервале от 0 до 1, стремясь к максимальному значению при полном успешном обнаружении границ.

При тестировании *CNN* основным показателем эффективности служит точность классификации объектов (*Accuracy*), представляющая процент правильно классифицированных объектов относительно общего количества тестовых образцов.

Эффективность речевого синтеза тестировалась относительно естественности речи *(Naturalness Mean Opinion Score, MOS-NAT)* и ее разборчивости *(Intelligibility Mean Opinion Score, MOS-INTEL)*. Показатель *MOS-NAT* определяется как среднее значение оценок пользователей, выражающих степень сходства синтезированной речи с естественной человеческой речью по пятибалльной шкале, где 1 соответствует искусственному звучанию, а 5 – максимально естественному. Показатель *MOS-INTEL* фокусируется на понятности и ясности произносимых слов. Его оценка аналогично проводится по пятибалльной шкале, где наивысший балл указывает на полную разборчивость речи.

На втором этапе проводится интеграционное тестирование системы путем объединения всех модулей и проверки их совместной работоспособности в реальных эксплуатационных условиях. Испытания охватили разнообразные сценарии движения пользователей, в разных условиях, с наличием препятствий.

Третий этап – нагрузочное тестирование системы, направленное на исследование поведения системы при длительных рабочих сессиях, высоких уровнях нагрузки и большом числе одновременно поступающих запросов.

Результаты технического тестирования представлены в таблице 2.

Таблица 2 – Результаты технического тестирования

Table 2 – Technical testing results

Тестирование	Описание процесса	Показатели	Результат
Модульное тестирование	Проверка камеры	Пиковый сигнал-шум	*PSNR* = 38 дБ
		Индекс структурного подобия	*SSIM* = 0,92
		Устойчивость	Потеря детализации = 1 %
	Проверка алгоритма сегментации	Коэффициент Джаккара	*Jaccard index* = 0,89
		Точность границы объекта	*Boundary recall* = 0,91
	Проверка CNN	Точность классификации	*Accuracy* = 94,5 %
		Задержка обработки кадра	15 мс
	Проверка речевого синтеза	Естественность речи	*MOS-NAT* = 4,2
		Разборчивость речи	*MOS-INTEL* = 4,4
Итерационное тестирование	Совместная работа компонентов	Разнообразие маршрутов	Городская среда, в закрытых помещениях
	Реальные условия испытаний	Время реакции системы	< 100 мс
Нагрузочное тестирование	Длительная эксплуатация	Производительность	Снижение на 5 % после 8 часов непрерывной работы
	Высокая нагрузка	Энергопотребление	3 Вт/ч

Результаты технического тестирования подтверждают высокую функциональность и надежность мультимодальной системы. Вместе с тем дальнейшее поддержание оптимального режима работы возможно при соблюдении определенных условий эксплуатации:

– рекомендуемое расстояние до объектов от 5 до 25 метров, что обеспечивает наиболее точные результаты измерения дистанций;

– предпочтительными для эффективной работы системы являются дневные часы и солнечная погода в условиях открытого пространства;

– в пасмурную или дождливую погоду желательно применять дополнительную подсветку, так как наблюдается снижение точности распознавания объектов на 2 %;

– закрытые помещения рекомендуется выбирать с мощными лампами искусственного освещения, создающими равномерное распределение света с минимальным уровнем теней;

– рекомендуется воздержаться от эксплуатации системы в темное время суток, поскольку уровень распознавания объектов снижается до 46 %.

Для повышения производительности работы мультимодальной системы в различных условиях, включая темное время суток, предполагается проведение дальнейших усовершенствований.

Заключение

Таким образом, предложенная мультимодальная система доказала свою ценность и перспективность в качестве инструмента реабилитации для лиц с нарушенным зрением. Обоснованность выводов подтверждается результатами проведенных экспериментов, показавших положительную динамику развития навыков ориентации в пространстве среди участников исследования. Система позволила существенно увеличить независимость и уверенность пользователей в повседневной жизнедеятельности, облегчая передвижение и повышая общий уровень комфорта и социального участия.

Дальнейшие исследования направлены на устранение существующих недостатков и расширение функциональных возможностей системы. Среди приоритетных задач выделяются разработка рекомендаций по оптимизации дизайна интерфейса, созданию специализированных методик обучения пользователей и исследование возможных негативных последствий длительной эксплуатации аудиовизуальных интерфейсов.

В целом, работа открывает широкие перспективы для внедрения инновационных аудиовизуальных решений в практику социальной реабилитации и медицинской техники, создавая основу для разработки новых стандартов оказания помощи лицам с инвалидностью по зрению.

Библиографический список

1. **Мясникова Л.В**. Использование высокотехнологичных средств ориентирования в обучении мобильности лиц с нарушениями зрения // Специальное образование и социокультурная интеграция-2018. 2018. № 4. С. 46-53.

2. **Mountapmbeme A., Okafor O., Ludi S.** Addressing accessibility barriers in programming for people with visual impairments: A literature review // ACM Transactions on Accessible Computing (TACCESS). 2022. № 1. С. 1-26.

3. **Hoggan E.** Multimodal Interaction // Interaction Techniques and Technologies in Human-Computer Interaction. 2024. № 2. С. 45-63.

4. **Baskar A., Kishan D., Lingesh B.V.** A vision system to guide visually challenged to pick up ingredients from rack // Journal of Advanced Research in Dynamical and Control Systems. 2017. № 11. С. 27-35.

5. **Mittal H.** A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets // Multimedia Tools and Applications. 2022. С. 1-26.

6. **Белокуров В.А.** Система угловой ориентации на основе гауссовского парциального фильтра // Вестник рязанского государственного радиотехнического университета. 2016. № 56. С. 11-16.

7. McGuinness K., O'connor N.E. A comparative evaluation of interactive segmentation algorithms // Pattern Recognition. 2010. Т. 43. № 2. С. 434-444. DOI: 10.1016/j.patcog.2009.03.008

8. **Ershov D., Phan MS., Pylvänäinen J.W**. et al. TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines // Nat Methods. 2022. Т. 19. С. 829-832. DOI: 10.1038/s41592-022-01507-1

9. **Kohler R.A** segmentation system based on thresholding // Computer Graphics and Image Processing. 1981. Т. 15. № 4. С. 319-338. DOI: 10.1016/S0146-664X(81)80015-9

10. **Wu J.** Introduction to convolutional neural networks // National Key Lab for Novel Software Technology. 2017. Т. 5. № 23. С. 495.

11. **Fish L.S., Busby D.M.** The delphi method // Research methods in family therapy. 1996. Т. 469. С. 482.

12. **Zhang J.** A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging // Scientific reports. 2022. № 1. С. 11905.

13. **Yuan J.** Performance analysis of deep learning algorithms implemented using PyTorch in image recognition // Procedia Computer Science. 2024. Т. 247. С. 61-69. DOI: 10.1016/j.procs.2024.10.008

14. **Anikin A., Herbst C.T.** How to analyse and manipulate nonlinear phenomena in voice recordings // Philosophical Transactions B. 2025. № 1923. С. 20-24.

15. **Брукина И.П., Фельдман Г.О.** Система пространственной ориентации людей с нарушением зрения на основе аудиовизуальных технологий // Труды международного симпозиума «Надежность и качество». 2025. Т. 1. С. 388-391.

UDC 004.93’11

A MULTIMODAL SYSTEM FOR AUDIOVISUAL DECISION SUPPORT FOR VISUALLY IMPAIRED USERS

I. P. Burukina, Ph.D., Associate Professor, Head of CAD Department, PSU, Penza, Russia;
orcid.org/0009-0006-1953-2914, e-mail: burukinairina@gmail.com
G. O. Feldman, Bachelor, PSU, Penza, Russia;
e-mail: gl.feldman2018@yandex.ru
D. A. Grishaev, Bachelor, PSU, Penza, Russia;
e-mail: dima_grishaev28@mail.ru

The work is devoted to solving the problem of spatial orientation of persons with visual impairment. The aim of the research is to develop a multimodal system that combines image segmentation algorithms and neural network methods to transform visual signals into a detailed audio representation of surrounding space. The system has successfully passed comprehensive technical testing, showing high functional and operational characteristics. The main advantages are high accuracy of object recognition, reliable distance detection, quick response and low power consumption. The research plays an important role in the development of domestic rehabilitation technologies aimed at ensuring accessibility of infrastructure and increasing the independence of people with visual impairment.

Keywords*: system, visual impairment, segmentation, neural network, spatial orientation, testing, efficiency.*

DOI: 10.21667/1995-4565-2026-95-226-235

References

1. **Myasnikova L.V.** The use of high-tech orientation tools in teaching mobility to people with visual impairments. *Special education and socio-cultural integration-2018*. 2018, no. 4, pp. 46-53. (in Russia).

2. **Mountapmbeme A., Okafor O., Ludi S.** Addressing accessibility barriers in programming for people with visual impairments: A literature review. *ACM Transactions on Accessible Computing (TACCESS*). 2022, no. 1, pp. 1-26.

3. **Hoggan E.** Multimodal Interaction. *Interaction Techniques and Technologies in Human-Computer Interaction.* 2024, no. 2, pp. 45-63.

4. **Baskar A., Kishan D., Lingesh B.V.** A vision system to guide visually challenged to pick up ingredients from rack. *Journal of Advanced Research in Dynamical and Control Systems.* 2017, no. 11, pp. 27-35.

5. **Mittal H.** A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*. 2022, pp. 1-26.

6. **Belokurov V.A.** Angular orientation system based on a Gaussian partial filter. *Bulletin of the Ryazan State Radio Engineering University*. 2016, no. 56, pp. 11-16. (in Russia).

7. **McGuinness K., O'connor N.E.** A comparative evaluation of interactive segmentation algorithms. *Pattern Recognition*. 2010, vol. 43, no. 2, pp. 434-444. DOI: 10.1016/j.patcog.2009.03.008

8. **Ershov D., Phan MS., Pylvänäinen J.W.** et al. TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat Methods*. 2022, vol. 19, pp. 829-832. DOI: 10.1038/s41592-022-01507-1

9. **Kohler R.** A segmentation system based on thresholding. *Computer Graphics and Image Processing*. 1981, vol. 15, no. 4, pp. 319-338. DOI: 10.1016/S0146-664X(81)80015-9

10. **Wu J.** Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology*. 2017, vol. 5, no. 23, p. 495.

11. **Fish L.S., Busby D.M.** The delphi method. *Research methods in family therapy*. 1996, vol. 469, p. 482.

12. **Zhang J.** A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging. *Scientific reports*. 2022, no. 1, p. 11905.

13. **Yuan J.** Performance analysis of deep learning algorithms implemented using PyTorch in image recognition. *Procedia Computer Science*. 2024, vol. 247, pp. 61-69. DOI: 10.1016/j.procs.2024.10.008

14. **Anikin A., Herbst C.T.** How to analyse and manipulate nonlinear phenomena in voice recordings. *Philosophical Transactions B*. 2025, no. 1923, pp. 20-24.

15. **Burukina I.P., Feldman G.O.** A system of spatial orientation of people with visual impairment based on audiovisual technologies. *Proceedings of the international symposium «Reliability and quality».* 2025, vol. 1, pp. 388-391. (in Russia).