УДК 004.3 + 004.5 DOI 10.34131/MSF.21.3.82-99

ПОДХОДЫ К АВТОМАТИЧЕСКОМУ РАСПОЗНАВАНИЮ ЖЕСТОВОЙ ИНФОРМАЦИИ: АППАРАТНОЕ ОБЕСПЕЧЕНИЕ И МЕТОДЫ

Д.А. Рюмин, И.А. Кагиров

Канд. техн. наук, ст.н.с. Д.А. Рюмин; н.с. И.А. Кагиров
(СПб ФИЦ РАН)

В статье рассмотрены аппаратные и программные решения, предназначенные для автоматического распознавания жестовой информации. Проанализированы тенденции анализа изображения в современных подходах, основанных на методиках компьютерного зрения. Выполнено сравнение подходов на предмет выявления достоинств и недостатков. Проведен обзор исследований по юзабилити жестовых интерфейсов. Установлено, что системы, основанные на датчиках, не уступая системам, основанным на зрении, в точности и скорости распознавания, имеют ограниченное применение ввиду специфики устройств (перчатка, костюм) и их сравнительно узкого распространения. В то же время, подходы, основанные на компьютерном зрении, могут быть успешно применены только тогда, когда будут решены проблемы окклюзий и наборов данных. Полученные результаты могут быть внедрены при создании обучающих систем.
Ключевые слова: жестовые интерфейсы, распознавание жестов, компьютерное зрение, человеко-машинное взаимодействие, нейронные сети.

Approaches to Automatic Gesture Recognition: Hardware and Methods Overview. D.A. Ryumin, I.A. Kagirov

In this paper, hardware and software solutions addressed to automatic gesture recognition are considered. Trends in image analysis in the current computer vision-based approaches are analysed. Each of the considered approaches was addressed, in order to reveal their advantages and drawbacks. Research papers on the usability of gesture interfaces were reviewed. It was revealed that sensor-based systems, being quite accurate and demonstrating high speed of recognition, have limited application due to the specificity of devices (gloves, suit) and their relatively narrow distribution. At the same time, computer vision-based approaches can be successfully applied only when problems of occlusions and datasets are solved. The results obtained can be used for designing training systems.
Keywords: gesture interfaces, gesture recognition, computer vision, human-machine interaction, neural networks.

Развитие цифровых технологий в последние годы привело к развитию и проникновению во все сферы человеческой деятельности средств бесконтактного взаимодействия с разноуровневыми информационными системами посредством жестовых человеко-машинных интерфейсов. В первую очередь,

подобные технологии повышают качество жизни, расширяют пользовательские возможности людей с ограниченными возможностями по слуху и зрению [1], однако могут применяться и в других областях, что обусловлено следующими факторами:

1. Жестовая модальность может быть предпочтительной в условиях зашумленных и/или больших помещений/пространств.

2. Зачастую жестовое управление оказывается более быстрым с точки зрения задач пользователя, нежели традиционное сенсорное или голосовое; в первую очередь, это касается ситуаций, которым в определенной культуре четко соответствует определенный набор жестов (например, жест «стоп», передаваемый поднятыми вверх руками).

3. Человеческая коммуникация имеет многомодальный характер. Исследование и разработка интерфейсов, поддерживающих несколько модальностей, в первую очередь, жестовую, повышает качество распознавания переданной информации.

Применение жестовых интерфейсов в пилотируемой космонавтике является отражением общей тенденции, прослеживающейся и в других сферах повседневной и специальной человеческой деятельности. Наиболее сложной оказывается задача эргономичного внедрения подобных устройств и их адаптация к условиям деятельности экипажа.

В данной статье рассмотрены общие тенденции архитектуры устройств для распознавания жестовой информации, методы и подходы; перспективы и специфика их применения в отдельных видах операторской деятельности могут быть предметом отдельного развернутого исследования. Тем не менее, представленный материал позволяет решать вопросы построения моделирующих стендов, предназначаемых для сравнительного изучения применения разных типов интерфейсов в контексте профессионально-ориентированных целей проектирования эргатических систем.

Настоящая статья посвящена обзору подходов, методов и аппаратному обеспечению, применяемому в системах, предназначенных для автоматического распознавания жестовой информации при человеко-машинном взаимодействии.

Статья организована следующим образом.

В первом разделе дается характеристика систем, основанных на применении оснащенной сенсорами перчатки, маркерных систем и аппаратных систем.

Второй раздел посвящен собственно методам распознавания жестовой информации, в том числе, на базе систем компьютерного зрения. В следующем разделе рассматривается эргономика (оценки юзабилити) систем с жестовыми интерфейсами. Наконец, в заключении делаются выводы о перспективности конкретных подходов, обсуждаются их достоинства и недостатки.

Системы, основанные на применении датчиков

Существующие на сегодня подходы и методы, направленные на автоматическое распознавание жестовой информации, могут быть классифицированы в зависимости от типа используемых входных данных [2]:

– определение жеста при помощи оборудованной сенсорами перчатки;
– маркерная система захвата движения;
– аппаратные средства видеозахвата движений.

В некоторых обзорах и исследованиях предлагается более общая классификация: все подходы и методы делятся на основанные на применении датчиков (*англ.*: sensor-based) и основанные на зрении (*англ.*: vision-based).

Определение жеста при помощи сенсорной гарнитуры/перчатки

Методы, основанные на применении датчиков, подразумевают использование специализированных устройств и исторически являются самыми первыми в данной области. История развития систем автоматического распознавания жестов начинается с начала 2000-х годов, когда определение жестов базировалось на оборудованной сенсорами перчатке, которая надевалась непосредственно на руку диктора. Набор таких датчиков позволял регистрировать физический отклик в зависимости от движений руки или сгибания пальцев. Затем данные обрабатывались при помощи подключенного (по проводу) к перчатке компьютера.

В настоящее время данная технология эволюционировала в портативную за счет использования микроконтроллеров вместо настольного компьютера, как показано на рис. 1 [3]. Набор датчиков определения кривизны, углового смещения, изгиба позволяет с высокой точностью определять углы изгибов пальцев благодаря их разным физическим принципам функционирования.

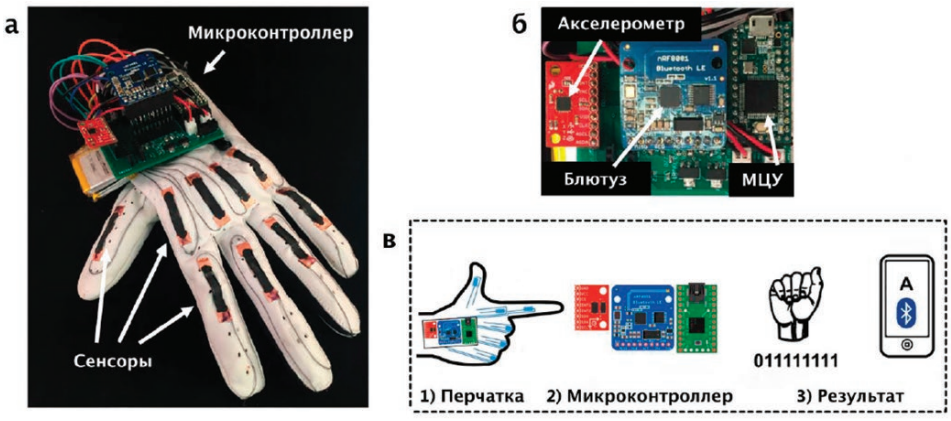


Рис. 1. Пример определения жеста
при помощи оборудованной сенсорами перчатки [3]

В качестве другого примера можно привести CyberGlove [http://www.cyberglovesystems.com/] – перчатку, предназначенную для отслеживания движений руки и пальцев пользователя. На поверхности перчатки закреплены датчики сгибания для пальцев, датчики абдукции (отведения пальцев в сторону), а также дополнительные датчики для более точного отслеживания формы кисти. Всего на перчатке закреплены 18 датчиков, таким образом, форма кисти отписывается 18 векторами признаков. Для передачи данных используется технология беспроводного USB.

В другом случае используются данные электромиографии. Так, прибор Polhemus FASTRAK [4] предназначен для отслеживания частей тела человека в режиме реального времени по 6 степеням свободы (*англ*. 6DOF) с минимальной задержкой. Он отслеживает положение (координаты X, Y и Z) и ориентацию (азимут, угол возвышения) небольшого датчика при его перемещении в пространстве.

Как было отмечено выше, с помощью оборудованной сенсорами перчатки можно достаточно точно определять координаты расположения, ориентации, конфигурации ладони и пальцев [5, 6]. Однако данный подход требует, чтобы перчатка была физически подключена к настольному компьютеру, ноутбуку или микроконтроллеру, что затрудняет человеко-машинное взаимодействие. К тому же цена на данное оборудование довольно высока.

Более новый подход заключается в распознавании жестов с помощью ультразвука. На руку крепится компактный ультразвуковой излучатель и приемник (рис. 2 [7]). Затем во время движения кистью руки или пальцами происходит анализ изменений акустических свойств ультразвука, по которым и определяется жест. Авторы исследования выделяют недостаток данного метода в необходимости обучать алгоритм для работы с конкретным информантом, что приводит к созданию дикторозависимой системы.

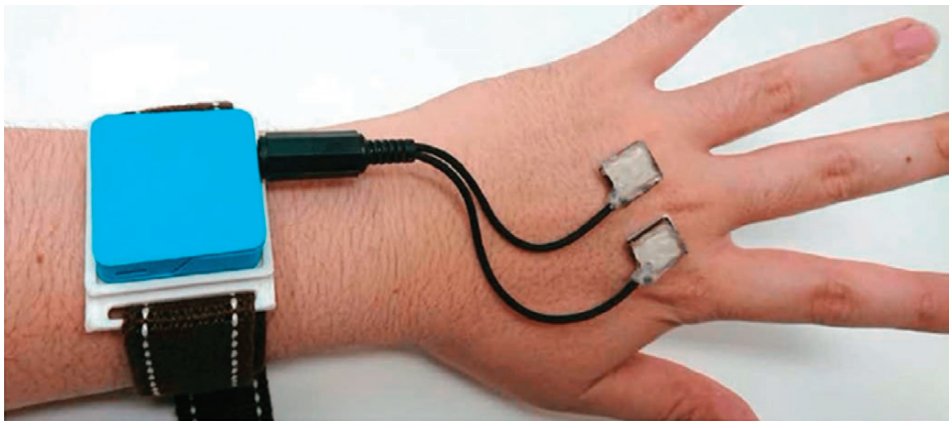


Рис. 2. Пример системы распознавания жестов
с помощью ультразвукового излучателя и приемника, цит. по работе [7]

Маркерная система захвата движения

В маркерных системах диктор или актер использует специальный костюм/перчатку, оснащенную набором так называемых «маркеров», траектории движения которых фиксируются камерой и обрабатываются компьютером. Обычно маркеры привязываются к характерным точкам на «анимационном скелете» и переносятся на создаваемую модель.

Существует несколько подходов к захвату движения, основанному на использовании маркеров. Так, можно выделить магнитные, акустические и оптические системы. В магнитных системах положение и ориентация маркера вычисляется относительно магнитного потока при помощи трехосевых магнитометров [8]. Координаты датчика вычисляются по изменениям напряжения и тока на катушках магнитометра.

Идея акустического принципа захвата движения заключается в использовании ультразвуковых передатчиков-маркеров. Ультразвуковые сигналы принимаются микрофонами, при этом расстояние до каждого маркера вычисляется по промежутку времени между принятым и переданным сигналами. Определение координат каждого отдельного маркера производится методом триангуляции. Примером современной системы акустического захвата движения может служить система MilliSonic [9]; в этой же статье дается обзор подобных систем, возникших за последние годы.

Наконец, оптические маркерные системы могут использовать как пассивные маркеры из светоотражающего материала, так и активные (чаще всего светодиодные). В случае с пассивной маркерной системой изображение фиксируется набором камер, производится поиск маркеров на изображении, а затем координаты каждого маркера рассчитываются посредством триангуляции. Как правило, пассивные системы захвата движения используют камеры, оснащенные инфракрасными светодиодами и инфракрасными фильтрами, что позволяет производить захват жеста даже в условиях недостаточной освещенности.

Активные системы оптического захвата движения основаны на снижении мощности свечения светодиода при удалении от камеры, что и используется для вычисления координат маркера. Активные системы обладают значительно большей точностью, если сравнивать их с пассивными системами. Тем не менее, на сегодняшний день наибольшую популярность снискали системы с пассивными оптическими маркерами. Существуют готовые решения от различных производителей (например, VICON) с различными техническими характеристиками.

Впервые принцип маркерного захвата движения для распознавания ручного жеста был предложен в работе [10] (рис. 3).

Подход к распознаванию жестов, основанный на маркерной системе захвата движения, хоть и исключает обязательное наличие микроконтроллеров на носимой перчатке, но все еще обладает недостатком в виде

Рис. 3. Пример перчатки
для маркерной системы захвата движения, цит. по работе [10]

непосредственного наличия самой перчатки или костюма. Кроме того, работа осложняется наличием большого массива камер, каждая из которых должны быть точно позиционирована и откалибрована.

Аппаратные средства видеозахвата движений

В последнее десятилетие благодаря развитию цифровых технологий появилась возможность разработки аппаратных средств видеозахвата жестов (оптические, инфракрасные, тепловизорные и другие камеры) и визуальных методов компьютерного зрения и машинного обучения для их обработки, которые исключают недостатки предыдущих описанных подходов. Такие автоматизированные системы распознавания жестов могут иметь широкий диапазон применения [11, 12], начиная от управления роботами [13] и заканчивая помощью врачам при клинических операциях [14]. Так, в работе [15] рассматривается роль визуальной интерпретации жестов рук в контексте человеко-машинного взаимодействия. Другая работа [16] посвящена выделению признаков для классификации определенных жестов рук. Кроме того, в работе [17] рассматривается разработка мобильного приложения для распознавания языка жестов Южно-Африканской Республики (ЮАР). В работе [18] авторы предложили систему распознавания жестов на основе компьютерного зрения, которая может быть использована в условиях со сложной фоновой обстановкой. Они разработали метод адаптивного обновления цветовой модели кожи для разных людей и при различных условиях освещения. Для описания контуров и основных моментов изображений с жестами рук авторы исследования объединили три вида функций (анализ главных

компонент, линейный дискриминантный анализ и метод опорных векторов) для создания нового метода иерархической классификации жестов. Оценка метода производилась на собранном наборе данных, в котором изображения одного и того же жеста были получены при различных условиях освещения.

Системы, основанные на компьютерном зрении

Подходы, подразумевающие применение датчиков или специального оборудования для диктора, демонстрируют довольно хорошие результаты распознавания, однако имеются и такие серьезные недостатки, как внушительные габариты костюма/перчатки, сложности подключения и настройки, что делает его не всегда удобным и применимым. Кроме того, люди, страдающие хроническими заболеваниями, которые приводят к потере мышечной функции, могут быть не в состоянии носить и снимать перчатки. Также датчики могут вызывать повреждения кожи, инфекцию или побочные реакции у людей с чувствительной кожей или тех, у кого имеются ожоги.

Подходы, основанные на компьютерном зрении и машинном обучении для задач обнаружения и распознавания жестов рук, на текущий момент являются наиболее применяемыми, поскольку они обеспечивают бесконтактное человеко-машинное взаимодействие [10]. Тем не менее, имеется также множество проблем в виду использования различных аппаратных средств видеозахвата жестов (оптические, инфракрасные, тепловизорные и другие камеры) [19].

К таковым можно отнести:

1) постоянное изменение освещенности;
2) эффекты окклюзии;
3) динамический фон;
4) время обработки, которое зависит от разрешения и частоты кадров;
5) дополнительные объекты переднего и заднего плана, представляющие тот же оттенок кожи или иначе похожие на руки человека [8, 20].

Из достоинств данных подходов можно выделить простоту использования и относительно невысокую цену по сравнению с оборудованной сенсорами перчатки.

Можно назвать следующие подходы к распознаванию жестовой информации, характерные для систем, основанных на компьютерном зрении:

1. Методы, основанные на выявлении цветовых закономерностей.
2. Подходы, основанные на внешнем виде.
3. N-точечные модели.
4. 3D-модели.

Методы, основанные на выявлении цветовых закономерностей, к которым относятся и различные подходы определения цвета кожи, до сих пор являются одними из самых популярных методов сегментации рук и используются в широком спектре приложений, таких, как: реконструкция

и классификация конфигураций рук, сегментация и идентификация жестов. Определение цветовой закономерности может быть выполнено несколькими способами. Первый способ заключается в определении цветовых сходств на основе пикселей, где каждый отдельно стоящий пиксель классифицируется без расположенных рядом пикселей. В свою очередь, второй способ учитывает набор пикселей в пространстве на основе такой информации, как интенсивность и текстура, что позволяет использовать цветовое пространство в качестве математической модели для представления информации о цвете кожи рук человека. В работе [20] представлен подход к распознаванию жестов рук, в котором цветное изображение из формата RGB (*англ*. Red Green Blue) преобразуется в цветовое пространство HSV (*англ*. Hue Saturation Value). Затем используются фильтры Габора для извлечения признаков, которые масштабируют и вращают изображение в 5 и 8 различных вариациях. Выходом такого фильтра является свертка исходного изображения с отфильтрованными изображениями. Результаты фильтрации отражают взаимосвязь между локальными пикселями (например, градиент или корреляция текстуры), но так как векторы признаков, полученные с использованием фильтра Габора, имеют высокую размерность, то для ее уменьшения используется метод нелинейного уменьшения размерности (*англ*. Kernel Principal Component Analysis, сокращенно Kernel PCA). На следующем шаге с помощью метода опорных векторов (*англ*. Support Vector Machine, сокращенно SVM) выполняется классификация жестов рук.

Для сегментации кожи на основе ее цветовой зависимости в основном используются такие форматы цветового пространства, как RGB, HSV, яркость. Более подробное описание цветовых зависимостей на основе модели RGB можно найти в работе [21]. Но стоит заметить, что данный способ на основе цветовых закономерностей для задач детектирования рук не является предпочтительным, так как набор цветовых каналов, которые затем образуются в модели, нельзя отнести к статическим характеристикам той или иной руки человека. Более того, в связи с тем, что цвет кожи определяется пороговым значением трех каналов (в случае с RGB), а в случае с нормализованными значениями информация о цвете кожи просто отделяется от яркостных значений, то при изменении освещенности сегментация и дальнейшая классификация жестов будет невозможна [22]. Стоит учитывать и тот факт, что характеристики цветового пространства, такие, как набор оттенков, насыщенность или яркость, следует применять в случаях с большой вариативностью освещения.

Подходы, *основанные на внешнем виде*, предполагают извлечение элементов изображения для моделирования внешнего вида объекта, например, руки. В таких подходах признаки вычисляются напрямую по яркости пикселей без предварительного процесса сегментации. В работе [23] представлен подход, основанный на примитивах Хаара. Такие примитивы могут эффективно анализировать контраст между темными и яркими объектами на

изображении. Кроме того, такой подход невосприимчив к окклюзиям и изменению освещения, поскольку он вычисляет разницу оттенков в градациях серого цветового пространства. Для задачи распознавания конфигураций рук использовался каскадный алгоритм машинного обучения (*англ*. Adaptive Boosting, сокращенно AdaBoost).

Подход извлечения гистограмм направленных градиентов (*англ*. Histogram of Oriented Gradients, сокращенно HOG) также основан на внешнем виде. Так признаки рук представляются в виде градиентов с резким изменением интенсивности по краям и углам объекта, то есть содержат контурную информацию о руке. HOG признаки устойчивы к освещению, однако не устойчивы к ориентации объекта. Для решения этой проблемы в работе [24] предложена статическая система оценки жестов рук на основе сравнения извлеченных HOG признаков с соответствующими сгенерированными эталонными жестами рук. Генерация эталонов жестов рук происходит с помощью специального программного обеспечения, которое позволяет визуализировать реалистичную трехмерную модель руки. Таким образом, для каждого жеста руки формируется набор изображений с разной ориентацией. С увеличением числа вариаций для каждого жеста возрастает точность распознавания жестов рук.

В другой работе [25] авторы предлагают улучшения для работы [24]. В качестве признаков жестов рук выступает комбинация HOG и локальных бинарных шаблонов (Local Binary Pattern, сокращенно LBP). Метод LBP заключается в том, что каждому пикселю изображения присваивается значение, характеризующее локальный узор вокруг этого пикселя. Эти значения вычисляются путем сравнения уровня градаций серого от центрального пикселя со значениями соседних пикселей. Таким образом, признаки LBP представляют собой текстурную информацию об объекте. Для сравнения эффективности системы использовался алгоритм машинного обучения AdaBoost.

Стоит отметить использование сверточных нейронных сетей, которые принимают на вход необработанные изображения, самостоятельно извлекают отличительные визуальные признаки и выполняют классификацию жестов рук [26, 27].

N-точечные модели представляют собой данные скелета руки, которые описывают ее геометрические и статические характеристики. Наиболее часто используемые характеристики – ориентация суставов, их расстояние и расположение относительного друг друга. В работе [28] представлен 3D-подход для сегментации рук с применением карты глубины сенсора Kinect v2, который определяет местоположения пальцев рук с использованием трехмерных соединений, евклидова и геодезического расстояний (*англ*. geodesic distance) по пикселям скелета руки. Другой 3D-подход к распознаванию жестов рук, основанный на модели машинного обучения с использованием двунаправленных сверточных нейронных сетей, представлен в работе [29].

Недостаток всех описанных ранее подходов заключается в анализе только статических жестов рук. Так, для анализа динамических жестов рук применяются другие методы. В работе [30] признаки из изображений жестов рук извлекались с помощью алгоритма ускоренных надежных функции (*англ*. Speeded Up Robust Feature, сокращенно SURF). Алгоритм SURF извлекает признаки в два этапа: 1) обнаружение ключевых точек на изображении и 2) создание их векторного представления, которое будет инвариантно к вращению и изменению масштаба. Векторное представление состоит из 64 или 128 значений для каждой ключевой точки интереса (пиксели). Далее извлеченные признаки подаются на вход скрытой марковской модели (*англ*. Hidden Markov model, сокращенно HMM). HMM применяются для предсказания последовательности изменений состояний на основе наблюдаемых последовательностей. Для процесса обучения в HMM традиционно используется алгоритм Витерби, который позволяет обнаружить наилучшее предположение о последовательности состояний, основываясь на последовательности наблюдений. Авторы предлагают использовать подход анализа последовательности состояний (*англ*. State Sequence Analysis, сокращенно SSA). В отличие от алгоритма Витерби, алгоритм SSA напрямую выполняет нахождение наиболее вероятной последовательности состояний. Полученные результаты показывают превосходство алгоритма SSA над алгоритмом Витерби.

В работе [31] авторы предлагают отслеживание жестов рук на основе траекторий центра их масс. В работе распознавались 16 букв английского алфавита, которые рисовались рукой информанта в воздухе. Алгоритмом классификации жестов выступал HMM.

Для анализа динамических жестов рук используются также сети с долгой кратковременной памятью (Long Short-Term Memory, сокращенно LSTM), в которых на вход принимаются несколько последовательных кадров. Так, в работе [32] предложен гибридный подход к распознаванию жестов рук, в котором на вход сверточной нейронной сети подается необработанное изображение, затем сеть LSTM используется для задачи классификации жестов рук.

Подходы по распознаванию жестов рук на основе 3D-моделей руки используют информацию о дальности визуальных элементов, тем самым позволяют формировать объемную модель руки. В работе [33] предложена модель распознавания действия рукой с использованием одного изображения RGB. В исследовании [34] предложен новый алгоритм на основе сверхточной нейронной сети (*англ*. 3D Convolution Neural Networks, сокращенно 3D CNN), который обучается детектировать руку из 3D-изображения. В работе [35] также применяется подход для обнаружения и распознавания трехмерных жестов одной руки с использованием CNN для обнаружения конфигураций рук. К недостаткам подходов, основанных на 3D-моделях, можно отнести потребность в больших наборах данных и вычислительные затраты.

Юзабилити жестовых интерфейсов

Представляется интересным оценить эргономичность и удобство применения жестовых интерфейсов с точки зрения пользователя. Одним из предметов эргономики является привлекательность и удобство человеко-машинного интерфейса для пользователя. Обыкновенно совокупность факторов, отвечающих за эргономичность интерфейса, обозначается термином «юзабилити», заимствованным из англоязычной литературы по эргономике [36]. К сожалению, четкой дефиниции этого собирательного термина не существует. Скорее, исследователям в каждом конкретном случае приходится оперировать конкретными наборами признаков, сформулированных по аналогии с теми, которые даны в ряде основных исследований по эргономике. Так, в основополагающей работе [37: 26] дается список из пяти признаков, из которых складывается юзабилити. Впоследствии этот список был расширен и переработан, в том числе, другими исследователями. Тем не менее, можно выделить перечень критериев оценки юзабилити, которые достаточно часто упоминаются в специальной литературе: а) эффективность (точность и полнота работы); б) результативность (затрачиваемые усилия, скорость); в) удовлетворенность (позитивное отношение к изделию).

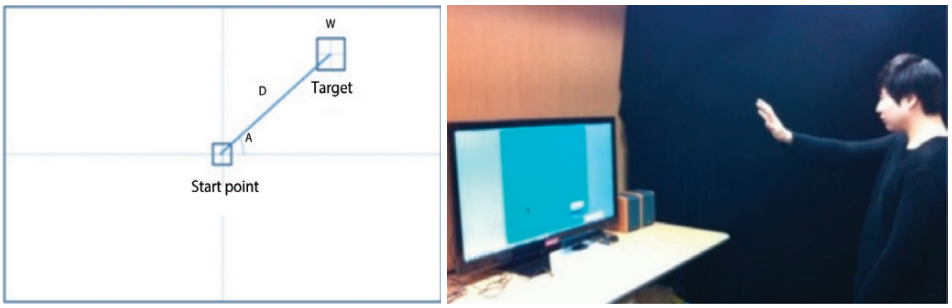
Для оценивания юзабилити применяются специальные методы тестирования. Обыкновенно подобные тестирования строятся на интервьюировании пользователей по определенному сценарию с использованием опросников. В процессе опроса пользователям предлагается выполнять задачи с помощью тестируемого продукта.

Смысл оценивания юзабилити состоит в поиске трудностей, с которыми сталкиваются пользователи. В результате тестирования появляется возможность улучшения интерфейса, соответствующая запросам пользователя. Отличие юзабилити-тестирования от экспертной оценки состоит в том, что в тестировании принимают участие потенциальные пользователи продукта, а не эксперты.

В ряде статей приводятся данные по экспериментам с юзабилити жестовых интерфейсов. Например, в исследовании [38] проведено сравнение сенсора MS Kinect и компьютерной мыши с точки зрения юзабилити. В исследовании приняли участие 50 пользователей. В ходе экспериментов пользователям предлагалось навести курсор на объект, представленный на экране компьютера и совершить действие (щелчок мышью) (рис. 4, а).

Эксперимент оценивался по закону Фиттса [39], связывающему время движения с точностью движения и с расстоянием перемещения: чем дальше или точнее выполняется движение, тем больше коррекции необходимо для его выполнения и, соответственно, больше времени требуется для внесения этой коррекции.

Результаты эксперимента показали, что в терминах индекса сложности [36] жестовый интерфейс проигрывает компьютерной мыши: 0.5312 и 1.2680 *бит/с*, соответственно.



(а) (б)

Рис. 4. (а) Пример задания для оценки управления курсором;
(б) общий вид системы, цит. по работе [38]

В статье [40] описывается дизайн жестового интерфейса на основе MS Kinect. Интерфейс распознает 5 жестов, которые использовались в компьютерной игре.

Тестирование приводилось при помощи опросника, и основными параметрами, которые интересовали исследователей, были узнаваемость жестов (*англ*. Memory Test) и уровень стресса при использовании жестового интерфейса. Как оказалось, из 38 опрошенных только 3 ошиблись в тесте на узнаваемость и только 5 раз жесты классифицировались как «раздражающие» в тесте на уровень стресса.

В другом исследовании [41] было проанализировано использование жестов в смоделированной среде рабочего стола персонального компьютера с использованием MS Kinect. Для исследования были выбраны максимально естественные жесты, которые использовались для выполнения различных задач в окружающей среде. Было создано два набора жестов, один с использованием кисти руки, а другой с использованием исключительно пальцев. Выполняемые задачи были разработаны с разными уровнями сложности.

Комплексное исследование юзабилити показало, что жесты, в которых были задействованы только пальцы, показались участникам эксперимента более естественными и менее утомительными. Однако для выполнения небольших задач, не требующих много времени, жесты кистью руки оказались предпочтительней, поскольку пользователи не успевали устать за короткий промежуток времени. Основная проблема заключается в том, что жесты руками требуют больших усилий и скорее утомляют, чем жесты только пальцами/манипуляции с мышью. Авторы делают вывод о том, что жесты как инструмент ввода информации значительно медленнее и утомительнее, чем при использовании мыши, однако выяснилось, что использование жестов руками на большом экране значительно естественнее и приятнее, чем использование мыши как на рабочем столе, так и на большом экране.

В статье [42] представлена мультимодальная система, которая использует жестовый и речевой интерфейсы для рисования трехмерного объекта в AutoCAD при помощи Leap Motion и микрофона. Оценка результатов показала, что выполнение задачи с использованием речевого управления воспринимается как утомительное по сравнению с использованием традиционных устройств ввода. Лишь небольшая часть пользователей (менее 7 %), смогла выполнить все задания с достаточной точностью. Жестовое управление показалось более естественным и менее утомительным в тех случаях, когда пользователи могли использовать обе руки, а не одну. Кроме того, пользователям хотелось иметь несколько жестов для одного и того же действия. Анализ результатов опросника показал, что пользователи готовы пользоваться жестами и речью и что эти модальности кажутся им удобными и естественными, однако основным условием для этого является скорость и стабильность работы системы. Однако интеграция Leap Motion, речевого ввода и AutoCAD не соответствовала этому стандарту, как отмечают сами авторы, что вызвало у пользователей некомфортные ощущения от работы с приложением.

На основании полученных в процитированных статьях выводов складывается впечатление, что жестовые интерфейсы, воспринимаясь пользователями как интуитивно понятные и естественные, все же проигрывают традиционным средствам человеко-машинного взаимодействия с точки зрения юзабилити. Это можно объяснить следующими факторами:

1) техническое несовершенство программно-аппаратной части: самым ярким примером будет процитированный в [42], когда опрошенные пользователи признали, что, несмотря на удобство жестов, интерфейс в целом вызывал неприятные ощущения (ошибки трекинга, долгий отклик системы);

2) неправильное применение: для ряда задач жестовый интерфейс является совершенно излишним и утомляющим; в то же время, как было отмечено в [41], в некоторых случаях именно жестовое управление оказывается приятным и предпочтительным.

Заключение

Проведенный обзор существующих моделей, методов, способов, принципов, а также интеллектуальных решений для автоматического распознавания жестовой информации показывает, что оба основных подхода к распознаванию жестовой информации обладают своими достоинствами и недостатками, которые необходимо оценивать в контексте стоящих исследовательских или проектных задач.

Краткое резюме по вышепредставленному сравнительному описанию состоит в следующем. Системы, оборудованные датчиками для работы с жестовой информацией, в том числе, сенсорные перчатки, оказываются хорошим средством для точного моделирования ручных жестов, обычно имеют компактную конструкцию и не препятствуют движениям рук. Кроме того,

системы с датчиками демонстрируют высокую точность при работе. Представляется, что, во-первых, к основным недостаткам подобных систем можно отнести их «лабораторность»: иными словами, большинство сенсорных перчаток являются уникальными, лабораторными (хоть и рабочими) системами, и не существует промышленных стандартов на проектирование и изготовление подобных устройств; это обстоятельство приводит и к высокой стоимости доступных коммерческих продуктов, что затрудняет их использование. Во-вторых, за исключением перчаток, основанных на датчиках растяжения, большинство перчаток имеет фиксированный размер, и их трудно подобрать для рук разных пользователей. Наконец, перчатки непригодны для использования в тех случаях, когда речь идет о дикторах, страдающих от определенных заболеваний, ограничивающих подвижность рук и суставов.

Методы, основанные на компьютерном зрении, также сопряжены с рядом проблемам.

В первую очередь, серьезной проблемой являются окклюзии. Поскольку руки часть используются для манипулирования объектами, высока вероятность, что они будут частично или полностью перекрыты объектами во время взаимодействия. Уже некоторое время как эта проблема находится в центре внимания исследователей. Так, авторы [43] предложили сквозную архитектуру для совместной оценки трехмерных поз рук и объектов на основе эгоцентрических изображений RGB.

В [44] предложен набор данных, содержащий сцены взаимодействия рук и 148 объектов, что в целом позволяет разрабатывать алгоритмы решения задачи распознавания взаимодействия рук и объектов.

Во-вторых, поскольку многие методы компьютерного зрения основаны на данных, большое значение имеют качество и охват обучающих наборов данных. Однако замечено, что все методы базируются на анализе исключительно 2D и 3D внешнего вида графического объекта (формы и позиции рук).

В этом случае не используется информация о физических свойствах рассматриваемого объекта [45]. К таковым методам относятся:

1) распознавание позиции и ориентации (кисти) руки с помощью моментов изображения, которое осуществляется только в том случае, если получаемое изображение включает в себя однородный фон, а также наличие одной руки человека, при условии, что рука является преобладающим объектом;

2) распознавание движения рук на основе анализа разности изображений через нахождение центра масс рук при движении;

3) распознавание конфигурации рук на основе анализа гистограмм направленных градиентов;

4) распознавание конфигурации и позиции рук на основе анализа контура изображения рук;

5) распознавание с применением алгоритма «случайного леса»;

6) распознавание жестов рук с применением скрытых марковских моделей и искусственных нейронных сетей, включая глубокие нейронные сети и методы глубокого обучения.

Результаты современных исследований дают основания считать, что методы машинного обучения, основанные на глубоких нейронных сетях, по сравнению с традиционными классическими подходами [46], которые базируются на линейных классификаторах (например, метод опорных векторов) имеют определенную специфику. Они показывают хорошие результаты в решении задач сегментации, классификации, а также распознавании как статических, так и динамических жестов.

При этом большинство методов, основанных на глубоком обучении, требуют больших объемов вычислительных ресурсов на этапах обучения и вывода. Многие алгоритмы должны запускаться на графическом процессоре (GPU) для достижения частоты кадров в реальном времени, что затрудняет развертывание на портативных устройствах, таких, как мобильные телефоны и планшеты. Таким образом, важно найти эффективные и действенные решения на мобильных платформах для повсеместных приложений.

Наконец, результаты юзабилити-тестирований показывают, что в целом жестовые интерфейсы обладают определенным потенциалом, однако их реальная имплементация должна отличаться от лабораторных систем, созданных и исследованных в последнее время.

ЛИТЕРАТУРА/REFERENCES

[1] Mahmud S., Lin X., Kim J.H. Interface for Human Machine Interaction for Assistant Devices: a Review // In Annual Computing and Communication Workshop and Conference (CCWC). – IEEE. – 2020. – pp. 0768–0773.

[2] Гриф М.Г., Козлов А.Н. Сравнительный анализ программно-аппаратных средств в задачах распознавания жестовой речи // Сборник научных трудов НГТУ. – 2014. – № 3. – Т. 77. – С. 63–72.
Grif M.G., Kozlov A.N. Comparative Analysis of Soft Hardware Regarding Gesture Recognition // Collection of NSTU Scientific Papers. – 2014. – No 3. – V. 77. – pp. 63–72.

[3] O'Connor T.F., Fach M.E., Miller R., Root S.E., Mercier P.P., Lipomi D.J. The Language of Glove: Wireless Gesture Decoder with Low-power and Stretchable Hybrid Electronics // PLoS ONE. – 2017. – Vol. 12. – No 7. P. e0179766. doi: 10.1371/journal.pone.0179766

[4] Maebatake M., Suzuki I., Nishida M., Horiuchi Y. and Kuroiwa S. 2008. Sign Language Recognition Based on Position and Movement Using Multi-Stream HMM. 2nd International Symposium on Universal Communication. – pp. 478–481.

[5] Rautaray S.S., Agrawal A. Vision Based Hand Gesture Recognition for Human Computer Interaction: a Survey // Artificial intelligence review. – 2015. – Vol. 43. – No 1. – pp. 1–54.

[6] Ibraheem N.A., Khan R.Z. Survey on Various Gesture Recognition Technologies and Techniques // International Journal of Computer Applications. – 2012. – Vol. 50. – No 7. – pp. 38–44.

[7] Kubo Y., Koguchi Y., Shizuki B., Takahashi S., Hilliges O. AudioTouch: Minimally Invasive Sensing of Micro-Gestures via Active Bio-Acoustic Sensing // In Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI). – 2019. – pp. 1–13.
[8] Yabukami Sh., Kikuchi H., Yamaguchi M., Arai K.I., Takahashi K., Itagaki A., Wako N. Motion Capture System of Magnetic Markers Using Three-axial Magnetic Field Sensor // IEEE Transactions on Magnetics 36. – pp. 3646–3648.
[9] Wang A., Gollakota Sh. MilliSonic: Pushing the Limits of Acoustic Motion Tracking // Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. May 2019. Paper No 18. – pp. 1–11.
[10] Kaur H., Rani J. A review: Study of Various Techniques of Hand Gesture Recognition // In International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES). – 2016. – IEEE. – pp. 1–5.
[11] Rajesh R.J., Nagarjunan D., Arunachalam R.M., Aarthi R. Distance Transform Based Hand Gestures Recognition for PowerPoint Presentation Navigation // Advanced Computing. – 2012. – Vol. 3. – No 3. – p. 41.
[12] Desai S., Desai A. Human Computer Interaction Through Hand Gestures for Home Automation Using Microsoft Kinect // In Proceedings of International Conference on Communication and Networks. – 2017. – pp. 19–29.
[13] Van den Bergh M., Carton D., De Nijs R., Mitsou N., Landsiedel C., Kuehnlenz K., Wollherr D., Van Gool L., Buss M. Real-time 3D Hand Gesture Interaction with a Robot for Understanding Directions From Humans // In Proceedings of the 2011 RoMan. – 2011. – IEEE. – pp. 357–362.
[14] Wachs J.P., Kölsch M., Stern H., Edan Y. Vision-based Hand-gesture Applications // Communications of the ACM. – 2011. – Vol. 54. – No 2. – pp. 60–71.
[15] Murthy G.R.S., Jadon R.S. A Review of Vision Based Hand Gestures Recognition // International Journal of Information Technology and Knowledge Management. – 2009. – Vol. 2. – No 2. – pp. 405–410.
[16] Khan R.Z., Ibraheem N.A. Hand Gesture Recognition: a Literature Review // International Journal of Artificial Intelligence and Applications. – 2012. – Vol. 3. – No 4. – p. 161.
[17] Seymour M., Tsoeu M. A Mobile Application for South African Sign Language (SASL) Recognition // In Proceedings of the AFRICON. – 2015. – IEEE. – pp. 1–5.
[18] Pan T.Y., Lo L.Y., Yeh C.W., Li J.W., Liu H.T., Hu M.C. Real-Time Sign Language Recognition in Complex Background Scene Based on a Hierarchical Clustering Classification Method // In Proceedings of the Second International Conference on Multimedia Big Data (BigMM). – 2016. – IEEE. – pp. 64–67.
[19] Sonkusare J.S., Chopade N.B., Sor R., Tade S.L. A Review on Hand Gesture Recognition System // In International Conference on Computing Communication Control and Automation. – 2015. – IEEE. – pp. 790–794.
[20] Uddin M.A., Chowdhury S.A. Hand Sign Language Recognition for Bangla Alphabet Using Support Vector Machine // In Proceedings International Conference on Innovations in Science, Engineering and Technology. – 2016. – IEEE. – pp. 1–4.
[21] Jones M.J., Rehg J.M. Statistical Color Models with Application to Skin Detection // International Journal of Computer Vision. – 2002. – Vol. 46. – No 1. – pp. 81–96.
[22] Brown D.A., Craw I., Lewthwaite J. A Som Based Approach to Skin Detection with Application in Real Time Systems // In BMVC. – 2001. – Vol. 1. – No 2001. – pp. 491–500.

[23] Chen Q., Georganas N.D., Petriu E.M. Real-time Vision-based Hand Gesture Recognition Using Haar-like Features // In the Instrumentation and Measurement Technology Conference (IMTC). – 2007. – IEEE. – pp. 1–6.
[24] Prasuhn L., Oyamada Y., Mochizuki Y., Ishikawa H. A HOG-Based Hand Gesture Recognition System On A Mobile Device // In International Conference on Image Processing (ICIP). – 2014. – IEEE. – pp. 3973–3977.
[25] Lahiani H., Neji M. Hand Gesture Recognition Method Based on HOG-LBP Features for Mobile Devices // Procedia Computer Science. – 2018. – Vol. 126. – pp. 254–263.
[26] Alnaim N., Abbod M., Albar A. Hand Gesture Recognition Using Convolutional Neural Network for People Who Have Experienced A Stroke // In the International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). – 2019. – IEEE. – pp. 1–6.
[27] Chung H., Chung Y., Tsai W. An Efficient Hand Gesture Recognition System Based on Deep CNN // In International Conference on Industrial Technology (ICIT). – 2019. – IEEE – pp. 853–858.
[28] Xi C., Chen J., Zhao C., Pei Q., Liu L. Real-time Hand Tracking Using Kinect // In the International Conference on Digital Signal Processing. – 2018. – pp. 37–42.
[29] Devineau G., Moutarde F., Xi W., Yang J. Deep Learning for Hand Gesture Recognition on Skeletal Data // In the International Conference on Automatic Face and Gesture Recognition (FG). – 2018. – IEEE. – pp. 106–113.
[30] Sagayam K.M, Hemanth D.J. A Probabilistic Model for State Sequence Analysis in Hidden Markov Model for Hand Gesture Recognition // Computational Intelligence. – 2019. – Vol. 35. – No 1. – pp. 59–81.
[31] Premaratne P., Yang S., Vial P., Ifthikar Z. Centroid Tracking Based Dynamic Hand Gesture Recognition Using Discrete Hidden Markov Models // Neurocomputing. – 2017. – Vol. 228. – pp. 79–83.
[32] John V., Boyali A., Mita S., Imanishi M., Sanma N. Deep Learning-based Fast Hand Gesture Recognition Using Representative Frames // In the International Conference on Digital Image Computing: Techniques and Applications (DICTA). – 2016. – pp. 1–8.
[33] Tekin B., Bogo F., Pollefeys M. H+ O: Unified Egocentric Recognition of 3D Hand-object Poses and Interactions // In the Conference on Computer Vision and Pattern Recognition. – 2019. – IEEE. – pp. 4511–4520.
[34] Malik J., Elhayek A., Stricker D. Structure-aware 3D Hand Pose Regression From a Single Depth Image // In the International Conference on Virtual Reality and Augmented Reality. – 2018. – pp. 3–17.
[35] Ryumin D., Kagirov I., Ivanko D., Axyonov A., Karpov A.A. Automatic Detection and Recognition of 3D Manual Gestures for Human–machine Interaction // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2019. – Vol. XLII-2/W12. – pp. 179–183.
[36] Сергеев С.Ф. Методы тестирования и оптимизации интерфейсов информационных систем. Учебное пособие. – Спб.: СПбГУ, 2015.
Sergeev S.F. Testing and Optimization Methods For Information Systems Interfaces. Text edition. – St. Petersburg. SPbU, 2015.
[37] Nielsen J. Usability Engineering. Boston: AP Professional, 1993.
[38] Lim C.J. & Jung Y.G. (2013). A Study on the Usability Testing of Gesture Tracking-Based Natural User Interface. Communications in Computer and Information Science. 373. – pp. 139–143.

[39] Fitts P.M. The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology 47(6). – pp. 381–391 (1954).
[40] Medeiros A.C.S., Tavares T.A., da Fonseca I.E. (2015) How to Design an User Interface Based on Gestures? // Marcus A. (eds) Design, User Experience, and Usability: Design Discourse. Lecture Notes in Computer Science, vol 9186. Springer, Cham.
[41] Farhadi-Niaki F., Etemad S.A., & Arya A. (2013). Design and Usability Analysis of Gesture-Based Control for Common Desktop Tasks. Lecture Notes in Computer Science. – pp. 215–224.
[42] Farzana Alibay, Manolya Kavakli, Jean-Rémy Chardonnet, Muhammad Zeeshan Baig. The Usability of Speech and/or Gestures in Multi-Modal Interface Systems. International Conference on Computer and Automation Engineering (ICCAE 2017), Feb. 2017, Sydney, Australia. – pp. 1–5.
[43] Tekin B., Bogo F., Pollefeys M.H. Unified Egocentric Recognition of 3D Hand-Object Poses and Interactions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019.
[44] Myanganbayar B., Mata C., Dekel G., Katz B., Ben-Yosef G., Barbu A. Partially Occluded Hands: A Challenging New Dataset for Single-Image Hand Pose Estimation. In Proceedings of the 14th Asian Conference on Computer Vision (ACCV 2018), Perth, Australia, 2–6 December 2018.
[45] Oyedotun O., Khashman A. Deep Learning in Vision-based Static Hand Gesture Recognition // Neural Computing and Applications. – 2017. – Vol. 28. – No 12. – pp. 3941–3951.
[46] Kagirov I., Ryumin D., Axyonov A. Method for Multimodal Recognition of One–Handed Sign Language Gestures Through 3D Convolution and LSTM Neural Networks // In: Salah A., Karpov A., Potapova R. (eds) Speech and Computer (SPECOM). Lecture Notes in Computer Science. – 2019. – Vol. 11658. – pp. 191–200.