

УДК: 004.89

ТЕХНОЛОГИИ ИДЕНТИФИКАЦИИ И РАСПОЗНАВАНИЯ ЖЕСТОВ**Сіряк Р.В.****GESTURE IDENTIFICATION AND RECOGNITION TECHNIQUES****Siryak R.V.**

В статье рассмотрены технологии распознавания и методы идентификации жестов. В статье приведены отрасли применения технологии распознавания жестов, выделены нерешенные проблемы в этой области. Определены этапы работы системы распознавания жестов, извлечения и предварительной обработки данных, подходы к получению данных. Выделены основные признаки распознавания жестов руки. Приведены подходы к распознаванию жестов на основе низкоуровневых признаков и на основании поиска кинематических параметров. Определено направление дальнейших исследований направленных на разработку метода улучшения существующих алгоритмов распознавания жестов руки в режиме реального времени в условиях меняющегося освещения.

Ключевые слова: распознавание жестов, нейронные сети, глубокое обучение, обработка динамических данных

Введение. Распознавание жестов – процесс, в котором жесты, совершаемые пользователем, используются для ввода информации в электронное устройство с целью управления, идентификации или коммуникации. Жесты могут быть статичными – одна конкретная позиция, или динамичными – последовательность позиций, извлекаемых из потокового видео. Основным средством передачи информации в большинстве приложений, использующих бесконтактный интерфейс, являются руки, хотя могут использоваться и лицо, и положение тела в целом. Распознавание жестов является комплексной задачей, включающей в себя моделирование и анализ движений, шаблоны распознавания и машинное обучение.

За последние несколько десятилетий разработаны различные инструменты и алгоритмы для применения в человеко-компьютерном взаимодействии, в роботоправлении, в видеонаблюдении, коммуникации, играх и прочих областях, где совершаемые человеком движения интерпретируются как определенные команды или коммуникационное взаимодействие.

Анализ исследований, области применения.

Удаленное управление техникой с помощью жестов достаточно широко используется в военной и космической сферах [1], но имеет реальные возможности для применения во многих других областях деятельности – там, где человеку затруднительно или опасно находиться. Жесты используются для взаимодействия с роботом и управления им в режиме реального времени

Распознавание жестов является актуальным для распознавания языка жестов. На сегодняшний день разработано немало систем распознавания языка жестов для различных языков. Подобная система может использоваться не только в качестве переводчика, но и служить средством ввода текста в компьютер, что для глухонемых людей может быть проще и естественней, чем клавиатура. Кроме того, учитывая, что языки жестов являются высокоструктурированными, они в целом хорошо подходят для проверки того или иного визуального алгоритма распознавания динамических образов.

Некоторые исследования [2] предлагают жестовое управление компьютером как альтернативу клавиатуры и мыши, и для решения этой задачи также используют распознавание жестов. В работе [3] управление компьютером для решения задач манипулирования графикой или редактирования документов включает в себя использование жестов, отражающих движения ручки.

В виртуальной и дополненной реальности используются жесты, позволяющие производить реалистично выглядящие манипуляции с виртуальными объектами через трехмерный интерфейс [4] или двухмерный интерфейс, имитирующий трехмерное взаимодействие [5].

Фриман и соавторы [6] изучили отслеживание руки игрока или его тела для контроля движения и положения игровых интерактивных объектов, таких, как автомобиль. Конрад и соавторы [77] использовали жесты для движения аватаров в виртуальном мире. PlayStation 2 представили

камеру Eye To You для отслеживания движений руки в интерактивных играх.

Постановка задачи. Несмотря на широкую область применения данных технологий, в области распознавания жестов рук существует целый ряд нерешенных задач. Существующие техники показывают достаточно высокие показатели, но преимущественно в специально созданных для этого условиях. В реальной жизни производительность значительно снижается вследствие различных условий освещения, шумов изображения, неоднородности фона, окклюзий и движения камеры. Помимо этого, в большинстве систем не решена проблема положения руки по отношению к камере: рука должна находиться в предопределенном месте и не уклоняться далее некоторого угла. Проблемой остается определение начала и конца жеста в последовательности кадров.

Все это обуславливает необходимость проведения дальнейших исследований, направленных на разработку комплексных систем, использующих сильные стороны различных подходов, взаимно перекрывающих недостатки.

Этапы работы системы распознавания жестов. Работу системы распознавания жестов можно разделить на определенные этапы. Их количество может различаться, в зависимости от используемых алгоритмов, назначения приложения и целей, которые ставят перед собой разработчики. Обобщенная схема, признаваемая большинством исследователей [8], представлена на рис. 1.

Извлечение и предварительная обработка данных. После получения информации с входного источника запускается процесс моделирования жеста, на котором определяется относятся ли данные на изображении к классу искомым объектам. Первым шагом моделирования является сегментация, которая представляет собой процесс разбиения изображения на области по сходству их точек. Таким образом, поле зрения D разбивается на области объектов D_1, \dots, D_s и область фона D_f .

Затем изображение обрабатывается с более четким делением на объект и фон. Снижаются помехи и подавляются внешние шумы - по сути это операции усреднения и выравнивания гистограмм. Проводится также цветовая нормализация в соответствие с условиями освещения и световой температурой.

Если жест статичен, то на этом шаге достаточно только выделить его контур или силуэт; если динамичен – тогда следует также определить его направленность.

На этапе определения контура объекта выявляются его значимые свойства, которые можно обнаружить по фотометрическим, геометрическим и физическим разрывам. Для этого могут быть использованы градиентные или лапласовы техники определения края.

Выделение характерных признаков. Выделение признаков является ключевым признаком для распознавания жестов руки. Они включают в себя: положение руки, ладони, пальцев, направление и углы, под которым расположены фаланги пальцев. Признаки могут быть геометрическими – контур руки, кончики пальцев, сами пальцы, и негеометрическими – цвет, силуэт, текстура.

При выделении признаков может использоваться кинематический подход, основанный на модели, при которой положение ладони и углы, под которыми расположены суставы, принимаются в качестве признаков [9]. Суть модели состоит в поиске кинематических параметров, переводящих двухмерную проекцию трехмерной модели руки в соответствие с обрезанным по краю изображением руки (рис. 2). Данные поступают с двух или более камер. Были показаны хорошие результаты при модели руки с семью степенями свободы.

При другом подходе, основанном на внешнем виде руки [11], используются двумерные параметры изображения, которые сравниваются с такими же параметрами, выделенными из входного изображения. В то же время жесты моделируются как последовательность представлений. Используемое в данном подходе собственное пространство (eigenspace) обеспечивает эффективное представление большого набора точек высокой размерности с использованием малого набора ортогональных базисных векторов. Эти векторы охватывают подпространство обучающего набора, называемого собственным пространством, и линейная комбинация этих изображений может быть использована для реконструкции изображений из обучающего набора.



Рис. 1. Схема системы распознавания жестов

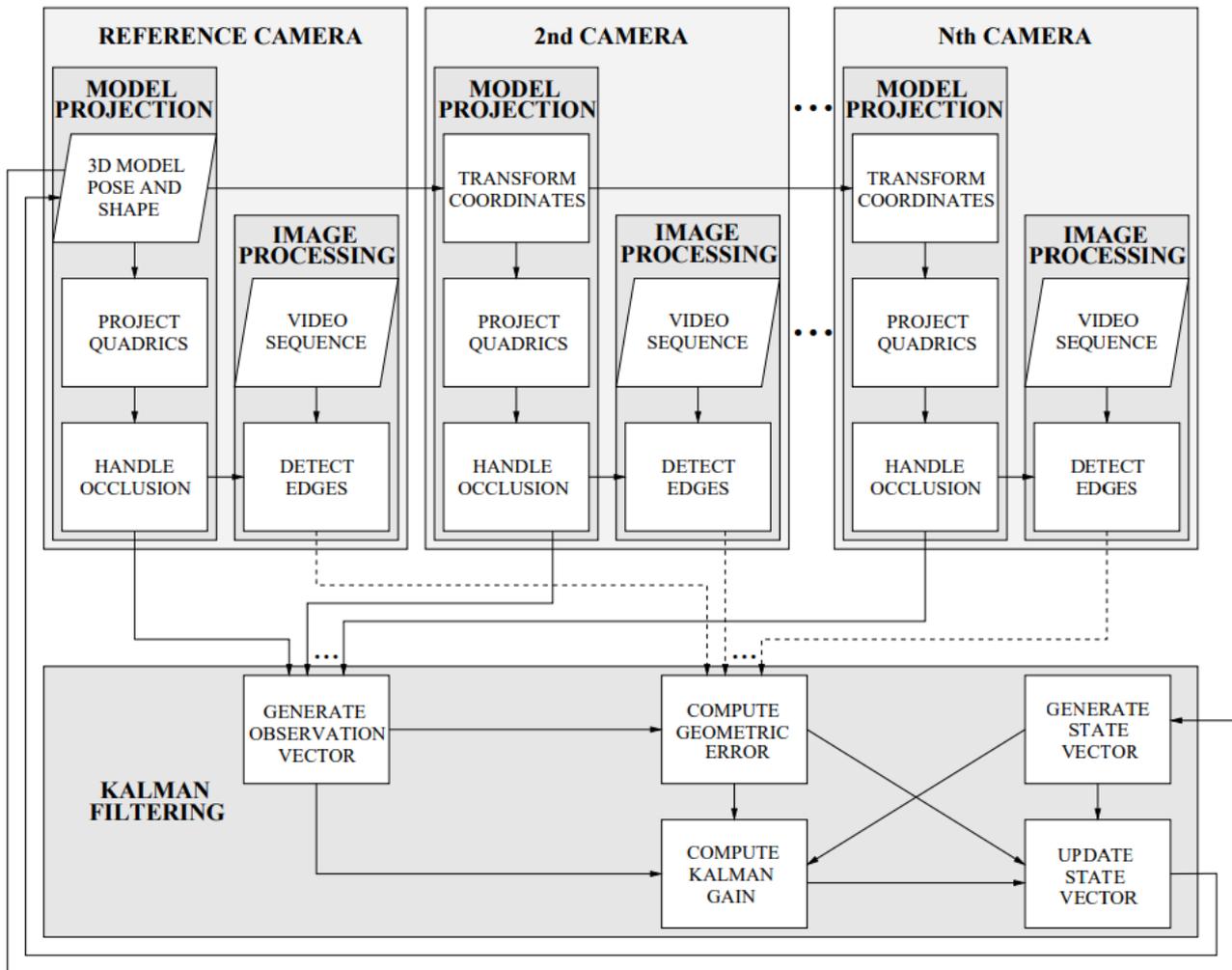


Рис. 2. Схема кинематической модели [10]

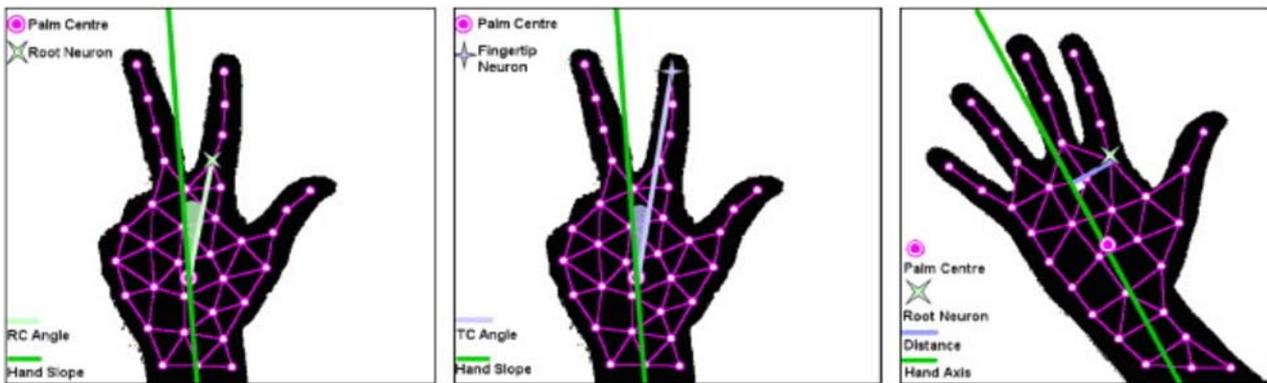


Рис. 3. Использование низкоуровневых признаков

Идея еще одного подхода – основанного на низкоуровневых признаках – исходит из того, что вовсе необязательно реконструировать всю руку. Вместо этого было предложено извлекать из изображения низкоуровневые параметры, которые устойчивы к шумам и могут быть быстро извлечены. К таким параметрам могут относиться: центр ладони [12]; оси, определяющие эллиптическую область руки [13] и оптический аффинный поток руки [14]. Одни из первых

исследователей, добившихся успеха при данном подходе, Старнер и соавторы [13] использовали в качестве низкоуровневых признаков оттенок кожи, положение руки по x и y, угол оси наименьшей инерции и ограничивающий эллипс. Комбинируя набор низкоуровневых признаков через сеть скрытые марковские модели, при распознавании американского языка жестов они добились точности в 97%. Пример использования низкоуровневых признаков представлен на рис. 3.

Выделенные признаки могут быть сохранены на этапе обучения в системе в виде шаблонов. Или же могут быть включены в уже имеющуюся нейронную сеть, скрытые марковские модели или деревья решений, имеющие ограниченную память.

Этап распознавания жеста. На заключительном этапе проводится идентификация значения жеста. На этой стадии обычно работает классификатор, который относит каждый входящий проверяемый жест к определенному классу с некоторой вероятностью его соответствия этому классу. Далее каждый этап должен быть рассмотрен более подробно с точки зрения применяемых формальных методов и подходов.

Подходы к получению данных. Используемые в распознавании жестов методы можно разделить на визуальные методы распознавания жестов; отслеживание рук с помощью оборудованной перчатки; захват движения на основе маркеров.

Данные методы основаны на восприятии человеческим зрением информации о видимых объектах. Обычно, информация на компьютер поступает через одну или две видеокamеры. Характеристиками, используемыми при идентификации жеста, обычно служат контуры и силуэты. При создании базы данных распознаваемых жестов каждый жест может содержать несколько образцов для повышения точности системы.

Один из распространенных визуальных методов основывается на внешнем подобии. При этом используются двумерные признаки изображения для моделирования визуальной внешности руки. Самым простым ориентиром при этом в процессе сегментации может быть цвет кожи, поскольку он не изменяется при масштабировании, сдвигах и поворотах [15]. Однако, подобный параметр сильно зависит от условий освещения и различия в цвете кожи у разных людей. Исследователи решают эту проблему, либо используя датчики [16], либо фреймворки глубокого обучения [17]. В свою очередь, в работах Неверовой и соавторов [18] или Он-Бара и соавторов [19] работу по распознаванию выполняют сенсоры цвета и глубины (RGB-D). В исследовании последних извлеченные из сенсоров RGB-D признаки HOG использовались для обучения SVM-классификатора. Молчанов и соавторы [16] для повышения точности, помимо RGB-D, использовали в работе глубину пространства, цвет и радарные сенсоры. Визуальные методы способны распознавать как статические, так и динамические жесты руки, порой с точностью более 90 процентов [20].

Другим способом визуального распознавания жеста является скелетный метод. В нем используется детализированная модель скелета руки, которая описывается с указанием длин сегментов и углов между ними. Моделирование

руки с учетом всех ее степеней свободы является сложной задачей, поэтому обычно используется модель с 27-ю степенями свободы. Система отслеживания положения руки в пространстве использует заранее созданную базу данных всевозможных конфигураций рук. Томаси и соавторам [21] удалось создать систему отслеживания руки в пространстве на основе трехмерной модели руки, которая может описать форму и сочлененную структуру руки.

Другим визуальным средством сегментации руки являются метрические методы распознавания жестов [20], в которых на основе множества входных сигналов строится метрика и выполняется классификация путем сравнения входного изображения с набором эталонов. К примеру, предлагается метрика, характеризующая степень сходства скелетов силуэтов ладони, и выполняется классификация жестов с помощью ближайшего соседа.

Достаточно большим интересом среди исследователей [22] пользуется Microsoft Kinect, оборудованный камерой глубины пространства. Благодаря лазерному проектору и сенсору CMOS он способен работать в любых условиях освещения.

Часто используется сочетание сразу нескольких способов получения данных: через цвет, глубину пространства, каркас скелета и т.д.

Главной проблемой при визуальных методах распознавания жестов является изменение фона и освещения, отчего снижается качество переднего плана и сегментированного изображения жеста. Для решения этой проблемы Zhang и соавт. [23] предлагается метод с использованием 3σ -принципа нормального распределения, исходя из разницы смежных видеокadров. Для выбора оптимального порога предлагается адаптивный метод автоматического выбора, основанный на методе максимальной дисперсии между классами.

Отслеживание рук с помощью оборудованной перчатки. Специально оборудованные перчатки позволяют максимально точно с высокой скоростью отобразить естественную механику человеческой руки и могут быть оснащены рядом датчиков: тактильными, вибрационными, акселерометром и др. Недостатком данного подхода является некоторая скованность движений, поскольку подобные перчатки довольно громоздки, и, помимо этого, человеку несвойственно использовать устройства-посредники для жестового общения. На настоящее время создано несколько достаточно удачных образцов систем, оборудованных перчатками. Бесконтактный интерфейс Fingual [24], созданный японскими исследователями из университетов Осаки и Шиншю представляет собой систему, направленную на перевод языка глухонемых в письменную речь.

Захват движения на основе маркеров. При маркерной системе [26] на руку человека наносятся цветные метки, с помощью которых передаются

данные о движении. На основе полученных данных строится трехмерная модель, в точности отображающая жесты. Маркерные метки наносятся на руку или же надеваются разноцветные перчатки. Подобные системы достаточно точны в передаче информации, просты в использовании и обладают низкой стоимостью.

Методы динамического распознавания жестов на основе глубокого обучения. В последние годы в научной литературе оформилось два основных подхода к алгоритмам распознавания на основе глубокого обучения.

Для первой категории характерна структура с обучением сети по входным данным без явного моделирования межкадровой информации. Как правило, множественная сверточная нейронная сеть адаптируются и обучаются с наборами масштабируемых данных, полученными от нескольких датчиков. Карпати и соавт. [26] предлагают мультиразмерную архитектуру сверточной нейронной сети, которая бы обучалась с исходными и усеченными кадрами, чтобы распознавать действие. О схожей множественной нейронной сети писал Молчанов и соавторы [16], в которой цвет и глубина комбинировались и использовались для обучения двух подсетей. В работе Неверовой и соавторов [1818] использовался мультимасштабный и мультимодальный фреймворк глубокого обучения. Мультимасштабная информация о жесте с системы захвата движения с видео подается как входные данные на несколько нейронных сетей. Выходные сигналы различных сверточных нейронных сетей сливаются многослойным перцептроном в выходной слой и классифицируются. Мультимасштабная нейронная сеть была адаптирована Пиго и соавторами [27] чтобы классифицировать американский язык жестов. В их работе несколько кадров RGB и данных глубины извлекались из видеопотока, содержащего жесты рук и тела, которые подавались на вход двух сверточных нейронных сетей для классификации. Также в работе Симонян и соавт. [28] архитектура двух сверточных нейронных сетей обучалась на видеопотоке и получаемом из него оптическом потоке.

Во второй категории распознавания движения посредством глубокого обучения исследователями моделируется межкадровая временная информация с использованием рекуррентных нейронных сетей (возвратных нейронных сетей). Муруками и соавторы [14] создали рекуррентную нейронную сеть для распознавания японского языка жестов. Ордез и соавторы [29] разработали рекуррентную нейронную сеть с долгой краткосрочной памятью (LSTM), чтобы классифицировать жесты через данные, поступающие с датчиков, размещенных на теле. Донахью и соавторы [30] интегрировали сверточную нейронную сеть в рекуррентную сеть с LSTM, создав таким образом рекуррентную сверточную сеть с долгосрочной памятью (LRCN).

В подобной сети несколько кадров выбираются из более длинной видеопоследовательности и передаются в качестве входных сигналов на несколько сверточных нейронных сетей, причем каждая из них получает отдельный кадр. Их выходные сигналы в свою очередь подаются на вход LSTM для предсказания типа класса. Классификаторы действий в данной сети сообщают с большой точностью о проблемах, связанных с распознаванием. Джон и соавторы [17] предложили использовать точность классификации LRCN для извлечения характерных кадров из видеопотока. Кроме того, для репрезентативного извлечения кадров используется новый фреймворк глубокого обучения, основанный на семантической сегментации.

Выводы. Распознавание жестов в последние два десятилетия в связи с развитием технологий нейронных сетей и повышением вычислительных возможностей пользуется среди исследователей большим интересом. Данная область является перспективной в плане развития человеко-машинных интерфейсов, межчеловеческой коммуникации, области виртуальной реальности и игровых приложениях. Перед разработчиками приложений для распознавания жестов, как и перед разработчиками распознавания образов вообще, стоят задачи совершенствования существующих алгоритмов и поиска новых. Дальнейшие исследования направлены на разработку метода, который бы позволил улучшить существующие алгоритмы распознавания жестов руки в режиме реального времени в условиях меняющегося освещения. Те, из существующих на сегодняшний день подходов, которые используют простые видеокamеры для получения информации, в значительной степени зависят от освещения и неоднородности фона изображения.

Л і т е р а т у р а

1. Goza, S. M., Ambrose, R. O., Diftler, M. A. & Spain, I. M. "Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform". In: Proceedings of the 2004 Conference on Human Factors in Computing Systems. ACM Press, (2004) 623–629.
2. Stotts, D., Smith, J. M. & Gyllstrom, K. "Facespace: Endo- and Exo-Spatial Hypermedia in the Transparent Video Facetop". In: Proc. of the Fifteenth ACM Conf. on Hypertext & Hypermedia. ACM Press, (2004) 48–57.
3. Smith, G. M. & Schraefel, M. C. "The Radial Scroll Tool: Scrolling Support for Stylus-or Touch-Based Document Navigation". In Proc. 17th ACM Symposium on User Interface Software and Technology. ACM Press, (2004) 53–56.
4. Sharma, R., Huang, T. S., Pavovic, V. I., Zhao, Y., Lo, Z., Chu, S., Schulten, K., Dalke, A., Phillips, J., Zeller, M. & Humphrey, W. "Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists". In: Proc. of ICPR'96 II (1996), 964-968.
5. Gandy, M., Stamer, T., Auxier, J. & Ashbrook, D. "The Gesture Pendant: A Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control

- and Medical Monitoring”. Proc. of IEEE Int. Symposium on Wearable Computers. (2000), 87-94.
6. Freeman, W., Tanaka, K., Ohta, J. & Kyuma, K. “Computer Vision for Computer Games”. Tech. Rep. and International Conference on Automatic Face and Gesture Recognition, (1996)
 7. Konrad, T., Demirdjian, D. & Darrell, T. “Gesture + Play: Full-Body Interaction for Virtual Environments”. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. ACM Press, (2003) 620–621.
 8. Sarkar, Arpita Ray, G. Sanyal, and S. Majumder. "Hand gesture recognition systems: a survey." International Journal of Computer Applications 71.15 (2013).
 9. Kristensson, P. O., Nicholson, T. F. W. and Quigley, A. 2012. Continuous Recognition of One-handed and Two-handed Gestures using 3-D Full-body motion tracking sensors. In Proc. of IUI 12.
 10. Stenger, Bjoern, Paulo RS Mendonça, and Roberto Cipolla. "Model-based 3D tracking of an articulated hand." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 2. IEEE, 2001.
 11. Gupta, N., Mittal, P., Dutta Roy, S., Chaudhury, S. & Banerjee, S. “Developing a Gesture-Based Interface”. IETE Journal of Research, 48(3) (2002) 237–244.
 12. New, J. R., Hasanbelliu, E. and Aguilar, M. “Facilitating User Interaction with Complex Systems via Hand Gesture Recognition.” In Proc. of Southeastern ACM Conf., Savannah, (2003).
 13. Starner, T., Weaver, J. & Pentland, A. “Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video”. PAMI, 20(12) (1998) 1371–1375.
 14. M. H. Yang, N. Ahuja, and M. Tabb, “Extraction of 2-D Motion Trajectories and its Application to Hand Gesture Recognition,” in PAMI., 29(8) (2002) 1062–1074.
 15. E. Stergiopoulou, N. Papamarkos. (2009). “Hand gesture recognition using a neural network shape fitting technique,” Elsevier Engineering Applications of Artificial Intelligence
 16. P. Molchanov, S. Gupta, K. Kim, and K. Pulli, “Multi-sensor system for driver’s hand-gesture recognition,” in FGR, 2015.
 17. V. John, A. Boyali, S. Mita, M. Imanishi, and N. Sanma, “Deep learning-based fast hand gesture recognition using representative frames,” in DICTA, 2016.
 18. N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout, Multi-scale Deep Learning for Gesture Detection and Localization, 2015.
 19. E. Ohn-Bar and M. M. Trivedi, “Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations,” IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 6, pp. 2368–2377, Dec 2014.
 20. Suk H., Sin B.K., Lee S.W. Hand gesture recognition based on dynamic Bayesian network framework // Journal Pattern Recognition, volume 43, number 9, Elsevier Science Inc.: 2010. — p. 3059-3072
 21. Tomasi C., Petrov S., and Sastry A. 3D tracking = classification + interpolation // Proc. Ninth IEEE International Conference on Computer Vision (Nice, France, 2003). – IEEE Computer Society, 2003. –P. 1441–1448
 22. Li, Y. 2012. Hand Gesture Recognition using KINECT. MS thesis. University of Louisville
 23. Q. Y. Zhang, F. Chen and X. W. Liu, “Hand Gesture Detection and Segmentation Based on Difference Background Image with Complex Background,” Proceedings of the 2008 International Conference on Embedded Software and Systems, Sichuan, 29-31 July 2008, pp. 338-343.
 24. Fukushima, Taishi, Fumio Miyazaki, and Atsushi Nishikawa. "The Development of a Noncontact Letter Input Interface “Fingual” Using Magnetic Dataset." Transactions of the Society of Instrument and Control Engineers 48.3 (2012): 159-166.
 25. Dadgostar, Farhad, Andre LC Barczak, and Abdolhossein Sarrafzadeh. "A color hand gesture database for evaluating and improving algorithms on hand gesture and posture recognition." (2005).
 26. Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei, “Large-scale video classification with convolutional neural networks,” in CVPR, 2014
 27. L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, Sign Language Recognition Using Convolutional Neural Networks, 2015.
 28. K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” 2014.
 29. F. J. Ordez and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” Sensors, vol. 16, no. 1, p. 115, 2016
 30. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, “Long-term recurrent convolutional networks for visual recognition and description,” in CVPR, 2015.

References

1. Goza, S. M., Ambrose, R. O., Diftler, M. A. & Spain, I. M. “Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform”. In: Proceedings of the 2004 Conference on Human Factors in Computing Systems. ACM Press, (2004) 623–629.
2. Stotts, D., Smith, J. M. & Gyllstrom, K. “Facespace: Endo- and Exo-Spatial Hypermedia in the Transparent Video Facetop”. In: Proc. of the Fifteenth ACM Conf. on Hypertext & Hypermedia. ACM Press, (2004) 48–57.
3. Smith, G. M. & Schraefel, M. C. “The Radial Scroll Tool: Scrolling Support for Stylus-or Touch-Based Document Navigation”. In Proc. 17th ACM Symposium on User Interface Software and Technology. ACM Press, (2004) 53–56.
4. Sharma, R., Huang, T. S., Pavovic, V. I., Zhao, Y., Lo, Z., Chu, S., Schulten, K., Dalke, A., Phillips, J., Zeller, M. & Humphrey, W. “Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists”. In: Proc. of ICPR'96 II (1996), 964-968.
5. Gandy, M., Starner, T., Auxier, J. & Ashbrook, D. “The Gesture Pendant: A Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring”. Proc. of IEEE Int. Symposium on Wearable Computers. (2000), 87-94.
6. Freeman, W., Tanaka, K., Ohta, J. & Kyuma, K. “Computer Vision for Computer Games”. Tech. Rep. and International Conference on Automatic Face and Gesture Recognition, (1996)
7. Konrad, T., Demirdjian, D. & Darrell, T. “Gesture + Play: Full-Body Interaction for Virtual Environments”. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. ACM Press, (2003) 620–621.

8. Sarkar, Arpita Ray, G. Sanyal, and S. Majumder. "Hand gesture recognition systems: a survey." *International Journal of Computer Applications* 71.15 (2013).
9. Kristensson, P. O., Nicholson, T. F. W. and Quigley, A. 2012. Continuous Recognition of One-handed and Two-handed Gestures using 3-D Full-body motion tracking sensors. In Proc. of IUI 12.
10. Stenger, Bjoern, Paulo RS Mendonça, and Roberto Cipolla. "Model-based 3D tracking of an articulated hand." *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 2. IEEE, 2001.*
11. Gupta, N., Mittal, P., Dutta Roy, S., Chaudhury, S. & Banerjee, S. "Developing a Gesture-Based Interface". *IETE Journal of Research*, 48(3) (2002) 237–244.
12. New, J. R., Hasanbelliu, E. and Aguilar, M. "Facilitating User Interaction with Complex Systems via Hand Gesture Recognition." In Proc. of Southeastern ACM Conf., Savannah, (2003).
13. Starmer, T., Weaver, J. & Pentland, A. "Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video". *PAMI*, 20(12) (1998) 1371–1375.
14. M. H. Yang, N. Ahuja, and M. Tabb, "Extraction of 2-D Motion Trajectories and its Application to Hand Gesture Recognition," in *PAMI.*, 29(8) (2002) 1062–1074.
15. E. Stergiopoulou, N. Papamarkos. (2009). "Hand gesture recognition using a neural network shape fitting technique," Elsevier Engineering Applications of Artificial Intelligence
16. P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *FGR*, 2015.
17. V. John, A. Boyali, S. Mita, M. Imanishi, and N. Sanma, "Deep learning-based fast hand gesture recognition using representative frames," in *DICTA*, 2016.
18. N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout, Multi-scale Deep Learning for Gesture Detection and Localization, 2015.
19. E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2368–2377, Dec 2014.
20. Suk H., Sin B.K., Lee S.W. Hand gesture recognition based on dynamic Bayesian network framework // *Journal Pattern Recognition*, volume 43, number 9, Elsevier Science Inc.: 2010. — p. 3059-3072
21. Tomasi C., Petrov S., and Sastry A. 3D tracking = classification + interpolation // Proc. Ninth IEEE International Conference on Computer Vision (Nice, France, 2003). – IEEE Computer Society, 2003. –P. 1441–1448
22. Li, Y. 2012. Hand Gesture Recognition using KINECT. MS thesis. University of Louisville
23. Q. Y. Zhang, F. Chen and X. W. Liu, "Hand Gesture Detection and Segmentation Based on Difference Background Image with Complex Background," *Proceedings of the 2008 International Conference on Embedded Software and Systems, Sichuan, 29-31 July 2008*, pp. 338-343.
24. Fukushima, Taishi, Fumio Miyazaki, and Atsushi Nishikawa. "The Development of a Noncontact Letter Input Interface "Fingual" Using Magnetic Dataset." *Transactions of the Society of Instrument and Control Engineers* 48.3 (2012): 159-166.
25. Dadgostar, Farhad, Andre LC Barczak, and Abdolhossein Sarrafzadeh. "A color hand gesture database for evaluating and improving algorithms on hand gesture and posture recognition." (2005).
26. Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei, "Large-scale video classification with convolutional neural networks," in *CVPR*, 2014
27. L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, *Sign Language Recognition Using Convolutional Neural Networks*, 2015.
28. K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," 2014.
29. F. J. Ordez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016
30. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *CVPR*, 2015.

Сіряк Р.В. Технології ідентифікації та розпізнавання жестів

У статті наведено огляд технологій розпізнавання і методів ідентифікації жестів. Наведені приклади застосування технології, виділені невирішені проблеми в цій галузі. Визначено етапи роботи системи розпізнавання жестів, вилучення та попередньої обробки даних, підходи до отримання даних. Наведено підходи до розпізнавання жестів на основі низькорівневих ознак і на підставі пошуку кінематичних параметрів. Визначено напрямки подальших досліджень спрямованих на розробку методу поліпшення існуючих алгоритмів розпізнавання жестів руки в режимі реального часу в умовах мінливого освітлення.

Ключові слова: розпізнавання жестів, нейронні мережі, глибоке навчання, обробка динамічних даних

R.V. Siryak Gesture Identification and Recognition Techniques

The examples of application of technology are given, outstanding unsolved problems in this area are highlighted. Stages of the system of recognition of gestures, extraction and preliminary processing of data, approaches to data acquisition are determined. The approaches to recognition of gestures on the basis of low-level signs and on the basis of search of kinematic parameters are given. The direction of further research aimed at the development of a method for improving the existing algorithms for recognizing hand gestures in real time in conditions of changing light is determined.

Keywords: gesture recognition, neural networks, deep learning, dynamic data processing

Сіряк Р.В. – пошукач кафедри комп'ютерної інженерії Східноукраїнського національного університету імені Володимира Даля, e-mail: hashem.r@gmail.com

Рецензент: д.т.н., проф. **Рязанцев О.І.**