Информационные технологии

УДК 0045

Распознавание жестов ручной азбуки ASL

В. Э. Нагапетян

Кафедра информационных технологий
Российский университет дружбы народов
ул. Миклухо-Маклая, д. 6, Москва, Россия, 117198

Предложен метод и разработана программная система автоматического распознавания жестов ручной азбуки глухонемых ASL (American Sign Language). В качестве устройства ввода информации о статических жестах, отображающих цифры и латинские буквы, выступает трёхмерный сенсор нового поколения Asus Xtion Pro Live. Распознавание жестов осуществляется посредством извлечения, предварительной обработки и последующего сравнения нормализованных геометрических скелетов руки на основе анализа дальностных изображений, формируемых сенсором. Сравнение скелетов осуществляется на основе алгоритма динамической трансформации шкалы времени (Dynamic Time Warping, DTW), имеющего полиномиальную сложность.

Ключевые слова: распознавание жестов, ASL, DTW, дальностное изображение.

1. Введение

Жесты являются неотъемлемой частью человеческого общения, причём для глухонемых и слабослышащих они служат единственным методом оперативного обмена информацией и обучения. На сегодняшний день не существует системы, полностью автоматизирующей процесс распознавания символов ручной азбуки. Для решения данной задачи предлагаются разные методы и подходы. Например, в работе [1] для распознавания жестов ручной азбуки ASL используется трёхмерная камера Microsoft Kinect, а само распознавание осуществляется применением алгоритма машинного обучения случайных лесов (random forest). Для повышения эффективности распознавания используется словарь английского языка, а при неоднозначном распознавании оператору предоставляется возможность самому выбрать правильную букву. В статье [2] для распознавания 24 жестов ручной азбуки в видеоряде используется двухслойная нейронная сеть типа feed-forward с алгоритмом обучения Левенберга–Маркара. В ней достигнута точность распознавания на уровне 99%, но авторы, к сожалению, не раскрывают информацию о количестве жестов в наборе данных и о других важных деталях эксперимента. В настоящей работе рассматривается решение задачи автоматизации распознавания жестов азбуки ASL на основе предложенного автором метода анализа геометрических характеристик формы дальностного изображения руки.

2. Ручная азбука ASL

Ручная азбука (дактильный алфавит, дактилемы) — это азбука, используемая в дактилологии, воспроизводящая посредством пальцев рук орфографическую форму слова речи. Согласно данным [3], при построении дактильных алфавитов упор делается на сходство с буквенными обозначениями. В то же время в состав некоторых алфавитов включаются дактилемы, обозначающие не буквы, а фонемы. Так же как чтение и произношение, дактилирование требует соблюдения

Статья поступила в редакцию 12 января 2013 г.

Работа выполнена при частичной поддержке проекта № 2.10 по программе фундаментальных исследований ОНИТ РАН «Интеллектуальные информационные технологии, системный анализ и автоматизация» и проекта РФФИ № 13-01-90602 Арм_а «Исследование методов человеко-машинного взаимодействия на основе отслеживания и распознавания жестов руки в режиме реального времени».

особых правил словесной речи. Согласно этим правилам, жесты должны показываться плавно и слитно, слова должны разделяться паузой, фразы — остановкой. Общепринято показывать жесты правой рукой, хотя существуют системы, например в Великобритании, где используются обе руки. При дактилировании рука должна находиться на уровне плеча, слегка вынесена вперёд и обращена ладонью от себя к собеседнику. При показе одной и той же буквы кисть руки сдвигается вправо. На рис. 1 приведена последовательность жестов, означающая слово "settlement" (поселение) на языке ASL (примечание: пример доступен по ссылке: `http://asl.ms`).

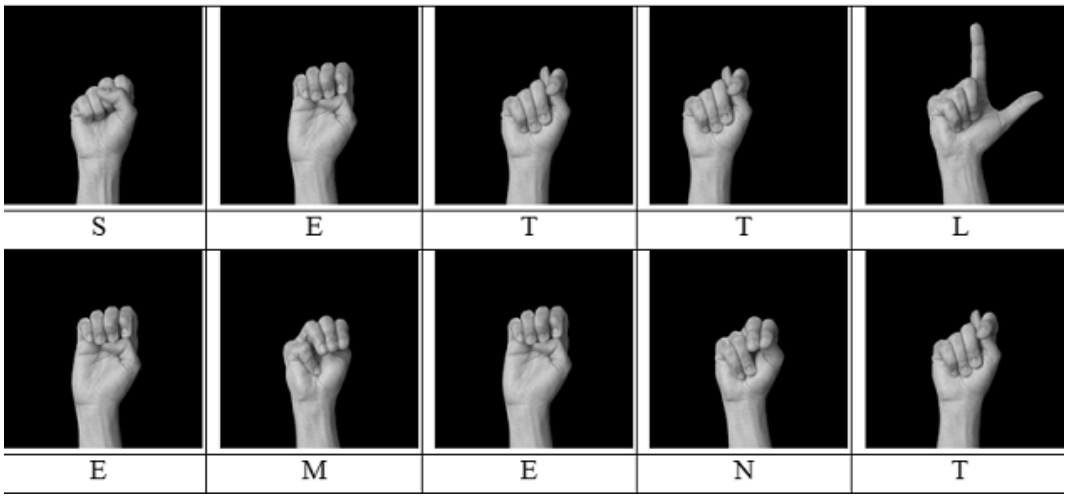


Рис. 1. Слово «settlement», показанное с помощью ручной азбуки ASL

3. Задачи распознавания жестов ручной азбуки

Основными задачами процесса распознавания жестов являются их ввод в автоматизированную систему, предварительная обработка и классификация.

3.1. Автоматический ввод жеста руки

Ввод изображения жеста руки может осуществляться с применением различных аппаратных средств и алгоритмов. Например, в работах [4,5] для ввода жеста используются веб-камеры низкого разрешения. В работах [6,7] для захвата жеста используется времяпролетная камера, в работах [8,9] ввод жеста осуществляется с применением перчатки со встроенными датчиками, которая в системах виртуальной реальности известна как «управляющая перчатка» (англ. dataglove). В основе работы [10] лежит трёхмерный сенсор нового поколения, работающий на принципах триангуляции и структурированного света. Захват жеста можно также осуществить с применением нескольких видеокамер, алгоритмов стереозрения или новых типов сенсоров, которые содержат встроенные алгоритмы стереозрения и распознавания кончиков пальцев руки. К такому классу принадлежат, например, сенсоры Leap Motion [11] и Edge3 [12]. Характеристики перечисленных устройств ввода приведены в табл. 1.

В настоящей работе для решения задачи распознавания жестов в качестве устройства ввода был применён сенсор Asus Xtion Pro Live, работающий на принципах структурированного света [13]. При выборе сенсора учитывались такие особенности как скорость выполнения операций, устойчивость работы при слабом освещении, цена и расширяемость функций для распознавания динамических жестов руки человека. Выбранный сенсор уступает сенсорам, работающим

Таблица 1

Характеристики устройств ввода жестов ручной азбуки

Название	Разрешение	Скорость (кадры/сек.)	Вид информации на выходе устройства:	Цена ($)
Цветная веб-камера	2560×1920	30	Цветное	10
Времяпролетная камера (Swiss Ranger SR4000, PMD CamBoard nano)	160×120	90	Дальностное изображение	10000
Сенсоры, работающие на принципах структурированного света (MS Kinect, Asus Xtion, PrimeSense depth sensor)	640×480	30/60	Дальностное изображение	100
Сенсоры, работающие на принципах стереозрения (Leap Motion, Edge3)	1.3/4.5 мегапиксель	160	Дальностное изображение и трёхмерные координаты кончиков пальца руки	70
Перчатки, оснащённые сенсорами (CyberGlove)	-	150	Трёхмерные координаты точек руки	10000

на принципах стереозрения по скорости передачи кадров и качеству изображения, но даёт возможность отследить не только пальцы руки, но и всю руку в целом. Это важно для дальнейших исследований, которые предполагают расширение перечня распознаваемых жестов ASL.

3.2. Задача классификации жеста руки

Задачей классификации жеста руки является присвоение входному жесту одного из 36 классов (26 букв и 10 цифр) ручной азбуки. В компьютерном зрении существует широкий класс алгоритмов, основанных на разных принципах, которые могут быть использованы для идентификации объекта в изображениях. Например, алгоритмы, основанные на геометрическом хешировании, позволяют распознать двумерные и трёхмерные объекты даже в том случае, когда объект был подвержен аффинной трансформации [14,15]. Исходными для распознавания являются основные характеристики объекта (дескрипторы), которые должны быть извлечены из изображения, например, количество вершин и рёбер контура объекта, объем, площадь поверхности объекта и другие свойства объекта. Результаты вычислений дескрипторов хранятся в хэш-таблице и используются в дальнейшем для классификации жестов. Другой метод распознавания объектов в изображениях основан на применении многомасштабных графов Риба (Multi-resolution Reeb graphs, MRGs), где в качестве меры используется геодезическое расстояние [16]. Метод инвариантен к таким трансформациям как вращение и сдвиг.

В настоящей работе для решения задачи классификации жестов используется метод, основанный на анализе геометрического скелета объекта, который определяется как множество точек, равноудалённых от двух ближайших точек контура

данного объекта. Для идентификации объекта сначала вычисляется геометрический скелет с применением методов итеративного истончения, диаграмм Вороного или же на основе вычисления расстояний точек до границ фигуры. Затем скелеты сравниваются друг с другом для выявления степени схожести объектов. Краткий анализ методов определения расстояний между скелетами дан в работах [17, 18].

Сложность задачи классификации жестов заключается в различии форм изображений руки при показе одного и того же жеста. При сравнении изображений жестов, должны учитываться такие факторы как особенности показа жеста разными людьми, наличие шума в изображениях и возможные отклонения от эталонов. В работе [19] отмечено, что при показе одного и того же жеста формы руки разных людей в изображении отличаются друг от друга настолько, что по ним можно идентифицировать личность человека. Однако этот факт снижает качество вышеупомянутых алгоритмов при использовании в системах автоматического распознавания жестов.

4. Алгоритм распознавания жестов руки

Укрупнённый алгоритм автоматизированного распознавания жестов руки состоит из нескольких шагов (рис. 2):

1) выделение ладони и пальцев руки в дальностном изображении;
2) вычисление геометрического скелета руки;
3) нормализация геометрического скелета;
4) сравнение полученного скелета со скелетами эталонных жестов.

Сенсор Asus Xtion Pro Live [13], используемый в качестве устройства ввода, снабжён одной RGB камерой, излучателем структурированного инфракрасного света и приёмником, который принимает отражённый свет от поверхностей объектов. В результате сенсор возвращает цветное изображение и дальностное изображение с разрешением 640×480 со скоростью 30 кадров в секунду, что вполне приемлемо для создания приложений, работающих в реальном времени.

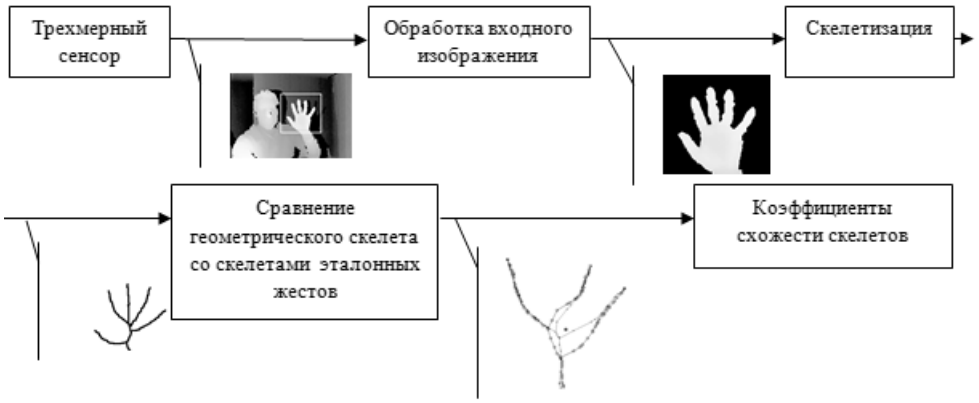


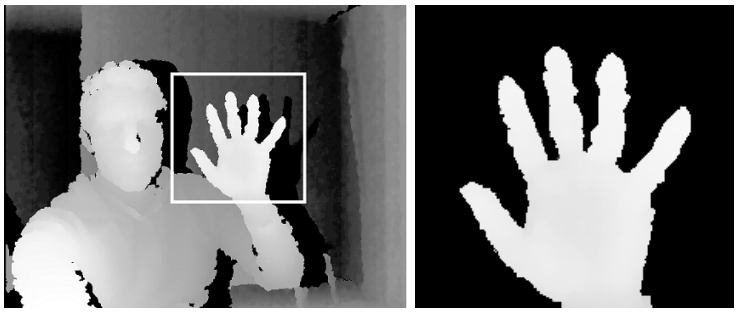
Рис. 2. Архитектура системы автоматизированного распознавания жестов

Далее рассмотрим более подробно основные этапы разработанного алгоритма.

4.1. Выделение ладони и пальцев руки в дальностном изображении

Дальностное изображение (или карта глубины) — это изображение, каждый пиксель которого характеризуется расстоянием до камеры наблюдения. Выделить ладонь руки в дальностном изображении можно разными способами. Согласно правилам дактилирования? во время показа жестов рука должна находиться на уровне плеча, слегка вынесена вперёд. Учитывая этот фактор, можно ставить естественное ограничение для жестикулирующего человека: рука должна быть

самым близким объектом к сенсору. В этом случае пиксель дальностного изображения, имеющий минимальное расстояние, будет точкой руки. По ней можно выделить целую ладонь и пальцы руки с помощью алгоритмов сегментации дальностных изображений [20]. В настоящей работе используется иной метод — перед началом жестикуляции человек делает определённый жест рукой (трясёт руку, держит руку в определённой позиции и пр.). Точка руки распознаётся и отслеживается в последующих кадрах. Ладонь и пальцы руки выделяются алгоритмом сегментации [21]. Результаты выделения ладони и пальцев руки предложенным способом показаны на рис. 3. Можно заметить, что в извлечённом изображении отсутствует задний фон, удаление которого является трудной проблемой при работе с цветными изображениями.



а б

Рис. 3. Дальностное изображение человека (а), извлечённое дальностное изображение руки (б)

4.2. Построение и сравнение геометрических скелетов руки

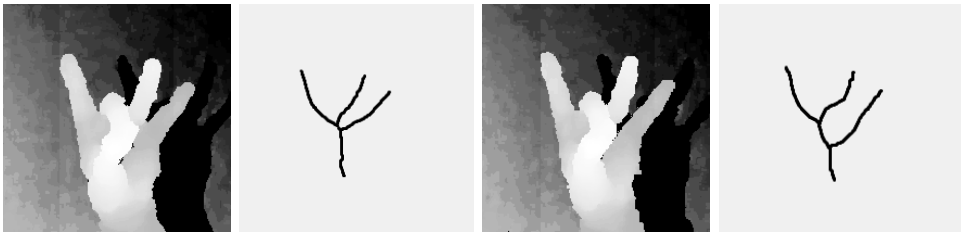
Для построения геометрического скелета руки (скелетизации) используется алгоритм, основанный на применении диаграмм Вороного [22]. На последующем этапе лишние дуги скелета удаляются посредством алгоритма «стрижки» (рис. 4).

а б в

Рис. 4. Дальностное изображение руки (а), монохромное изображение извлечённой руки (б), геометрический скелет руки после «стрижки» (в)

Геометрический скелет является одним из дескрипторов объекта, который широко используется в компьютерном зрении для распознавания двумерных и трёхмерных объектов [18, 22]. Несмотря на большое количество алгоритмов распознавания объектов, непосредственное использование классических алгоритмов

сравнения скелетов для распознавания жестов ручной азбуки не даст удовлетворительных результатов. Это связано с тем, что геометрические скелеты одного и того же жеста могут отличаться друг от друга как количеством дуг, так и длиной и углами отдельных дуг скелетов (рис. 5). Для сравнения скелетов руки была



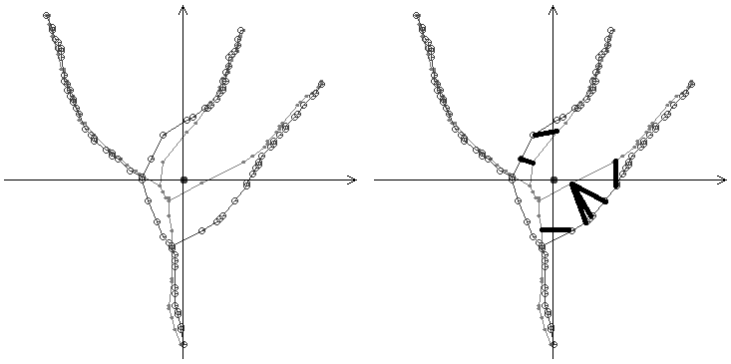
а б в г

Рис. 5. Различие скелетов руки, при показе одного и того же жеста: (а) Жест «8» из ручной азбуки ASL (б) Количество дуг — 83 (в) Жест «8» из ручной азбуки ASL (г) Количество дуг — 104

разработана и программно реализована система, основанная на алгоритме DTW. В качестве меры расстояния была выбрана метрика Манхеттена. Сравнение скелетов руки осуществляется в два этапа:

1. Нормализация скелетов
2. Вычисление коэффициента схожести скелетов

Нормализация включает в себя изменение масштаба и местоположения скелета руки. Масштаб нормализуется на основе расстояния от точки центра руки до сенсора. Скелет сдвигается алгоритмом до совпадения его геометрического центра с началом координатной оси (рис. 6(а)).



а б

Рис. 6. Сравнение скелетов руки: (а)Нормализированные скелеты руки (б) Сопоставленные вершины скелетов

В качестве коэффициента сходства двух скелетов предложено использовать суммарное расстояние путей, которые должны пройти вершины одного скелета при его трансформации до точного попадания на вершины другого скелета. Сопоставленные вершины определяются алгоритмом DTW. На рис. 6(б) жирной линией показаны некоторые сопоставленные вершины для двух скелетов. Обозначим через U и V множества вершин первого и второго скелета соответственно, где $|U| = m$, $|V| = n$. $A \in R^{m \times n}$ — матрица расстояний между вершинами U и V на метрике Манхеттена: $a_{i,j} = d(u_i(x,y), v_j(x,y)) = |u_{ix} - v_{jx}| + |u_{iy} - v_{jy}|$. Следующим шагом является поиск пути в матрице $A \in R^{m \times n}$, начинающегося

с элемента $a_{1\times1}$ и достигающего элемента $a_{m\times n}$, для которого сумма значений элементов данного пути минимальна. Решить задачу за полиномиальное время можно посредством алгоритма динамического программирования. Для матрицы $A \in R^{m\times n}$ создаётся новая матрица $B \in R^{(m+1)\times(n+1)}$. Элементу $b_{1,1}$ присваивается значение 0, а прочим элементам первой строки и первого столбца матрицы B — значение ∞. Остальные элементы матрицы вычисляют следующим образом: $b_{i,j} = a_{i,j} + \min\{b_{i-1,j}, b_{i,j-1}, b_{i-1,j-1}\}$. В качестве коэффициента схожести двух скелетов выбирается значение элемента $b_{m+1,n+1}$. Заметим, что порядок выбора вершин скелета для построения матрицы влияет на коэффициент схожести. По этой причине вершины скелета изначально сортируются.

5. Результаты тестирования алгоритма распознавания жестов

Предложенный алгоритм был протестирован на базе изображений жестов цифр и букв ручной азбуки ASL. Тестовая база включала в себе 800 изображений жестов цифр и 1920 изображений жестов букв двух разных людей. В табл. 2 и 3 приведены характеристики качества распознавания алгоритма, где точность распознавания определяется как доля жестов действительно принадлежащих данному классу относительно всех жестов, которые система отнесла к этому классу. Полнота распознавания определяется как доля найденных классификатором жестов, принадлежащих классу относительно всех жестов этого класса в тестовой выборке. Заметим, что в табл. 2 отсутствуют буквы J и Z, жесты которых не являются статическими.

Таблица 2

Характеристики качества распознавания для жестов букв

Характеристики качества распознавания	Тестовая выборка											
	A	B	C	D	E	F	G	H	I	K	L	M
Точность	0.81	0.79	0.87	0.95	0.86	0.97	0.85	1	0.81	1	0.6	0.4
Полнота	0.6	0.6	1	0.9	0.9	0.8	0.8	1	0.9	0.9	1	0.3
	N	O	P	Q	R	S	T	U	V	W	X	Z
Точность	1	1	1	1	0.9	0.5	0.5	0.73	0.87	0.77	1	1
Полнота	0.5	0.45	1	1	0.85	1	0.38	0.75	0.85	1	1	1

Таблица 3

Характеристики качества распознавания для жестов цифр

Характеристики качества распознавания	Тестовая выборка									
	0	1	2	3	4	5	6	7	8	9
Точность	1	0.83	0.83	0.97	0.7	1	0.5	1	0.97	0.97
Полнота	0.55	1	0.88	0.73	0.98	0.98	0.98	0.33	0.9	0.93

Из проведённых экспериментов видно, что средняя точность распознавания цифр составила $87,7\%$, а полнота — $82,6\%$. Эти же показатели для распознавания букв составили соответственно $84,08\%$ и $81,2\%$. Из табл. 3 видно, что жест цифры 0 в 45% случаев был отнесён к неправильному классу, что объясняется существенным различием форм и размеров руки людей, показывающих этот жест. Решение данной проблемы требует дальнейшего совершенствования метода распознавания. При тестировании алгоритма на базе жестов одного человека средняя полнота распознавания цифр достигла $87,5\%$, а букв – $84 - 38\%$.

6. Заключение

Выполненные эксперименты показали, что разработанный алгоритм с большой точностью и полнотой распознает большинство статических жестов ручной азбуки ASL. Эффективность алгоритма при распознавании некоторых букв падает по причине большой схожести фигур соответствующих жестов. Такими являются буквы G и H, M, N, A и S.

Для повышения качества распознавания жестов в последующих исследованиях предполагается существенное расширение набора дескрипторов, описывающих скелет. Например, можно учитывать такую характеристику, как расстояние дуг скелета до ближайших точек контура руки, что позволит различить, жесты букв G и H. Дальностное изображение руки может быть дополнено анализом цветных изображений, что также может повлиять на улучшение качества распознавания.

Литература

1. *Pugeault N., Bowden R.* Spelling It Out: Real–Time ASL Fingerspelling Recognition // Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. — Barcelona, Spain: IEEE, 2011. — P. 1114 – 1119.
2. *Isaacs J., Foo S.* Hand Pose Estimation for American Sign Language Recognition // System Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on. — IEEE, 2004. — Pp. 132–136.
3. *Зайцева Г. Л.* Жестовая речь. Дактилология: Учеб. для студ. высш. учеб. заведений. — М.: ВЛАДОС, 2000. — 192 с. [Zaitseva G. L., Gestural language. Cheirology: Textbook for high school students. — M.: VLADOS, 2000. — 192 p.]
4. Hand Tracking and Gesture Recognition for Human-Computer Interaction / C. Manresa, J. Varona, R. Mas, F. Perales // ELCVIA. — 2005. — No 5(3). — Pp. 96–104.
5. Flutter - Play and Pause Your Music and Movies with a Gesture. — `https://flutterapp.com/`. — Accessed: 10/01/2013. Accessed: 10/01/2013.
6. Gesture Recognition with a Time-of-Flight Camera / E. Kollorz, J. Penne, J. Hornegger, A. Barke // IJISTA. — 2008. — Vol. 5, issue 3/4. — Pp. 334–343.
7. *Breuer P., Eckes C., Müller S.* Hand Gesture Recognition with a Novel IR Time-of-Flight Range Camera: a Pilot Study // Proceedings of the 3rd International Conference on Computer Vision / Computer Graphics Collaboration Techniques. — MIRAGE'07. — Berlin, Heidelberg: Springer-Verlag, 2007. — Pp. 247–260.
8. *Kevin N. Y. Y., Ranganath S., Ghosh D.* Trajectory Modeling in Gesture Recognition using Cybergloves Reg; and Magnetic Trackers // TENCON 2004. 2004 IEEE Region 10 Conference. — Vol. A. — 2004. — Pp. 571–574.
9. *Ji-Hwan K., Nguyen D. T., Tae-Seong K.* 3-D Hand Motion Tracking and Gesture Recognition using a Data Glove // Industrial Electronics, 2009. ISIE 2009. IEEE International Symposium on. — 2009. — Pp. 1013–1018.
10. *Li Y.* Hand gesture recognition using Kinect // Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on. — 2012. — Pp. 196–199.

11. Leap Motion. — https://leapmotion.com. — Accessed: 10/01/2013. Accessed: 10/01/2013.
12. Edge3. — http://edge3technologies.com. — Accessed: 10/01/2013. Accessed: 10/01/2013.
13. Asus Xtion Pro Live. — http://www.asus.com/Multimedia/Motion_Sensor/Xtion_PRO_LIVE/. — Accessed: 10/01/2013. Accessed: 10/01/2013.
14. *Hecker Y., Bolle R.* On Geometric Hashing and the Generalized Hough Transform // Systems, Man and Cybernetics, IEEE Transactions on. — 1994. — Vol. 24, No 9. — Pp. 1328–1338.
15. *Lamdan Y., Schwartz J., Wolfson H.* Affine Invariant Model-Based Object Recognition // Robotics and Automation, IEEE Transactions on. — 1990. — Vol. 6, No 5. — Pp. 578–589.
16. Topology Matching for Fully Automatic Similarity Estimation of 3D shapes / M. Hilaga, Y. Shinagawa, T. Kohmura, T. L. Kunii // Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. — SIGGRAPH '01. — New York, NY, USA: ACM, 2001. — Pp. 203–212.
17. *Bunke H., Shearer K.* A Graph Distance Metric Based on the Maximal Common Subgraph // Pattern Recogn. Lett. — 1998. — Vol. 19, No 3-4. — Pp. 255–259.
18. *Brennecke A., Isenberg T.* 3d shape matching using skeleton graphs // In Simulation and Visualization. — 2004. — Pp. 299–310.
19. Shape-Based Hand Recognition / E. Yoruk, E. Konukoglu, B. Sankur, J. Darbon // Image Processing, IEEE Transactions on. — 2006. — Vol. 15, No 7. — Pp. 1803–1815.
20. Depth-Supported Real-Time Video Segmentation with the Kinect / A. Abramov, K. Pauwels, J. Papon et al. // Applications of Computer Vision (WACV), 2012 IEEE Workshop on. — 2012. — Pp. 457–464.
21. *Нагапетян В. Э.* Обнаружение пальцев руки в дальностных изображениях // Искусственный интеллект и принятие решений. — 2012. — № 1. — С. 90–95. [Nahapetyan V. E. Fingertip detection in depth images // ISA RAS. № 7. — 2012. — P. 90–95]
22. *Местецкий Л. М.* Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры. — М.: ФИЗМАТЛИТ, 2009. — 288 с. [Mestetskiy L. M., Continuous morphology of binary images: figures, skeletons, circulars. — M.: Fizmatlit, 2009. — 288 p.]

UDC 0045

ASL Fingerspelling Recognition

V. E. Nahapetyan

Information Technology Department
Peoples' Friendship University of Russia
Miklukho-Maklaya str., 6, Moscow, Russia, 117198

A method is proposed and software is developed for automatic recognition of gestures used in ASL fingerspelling. Static gestures are captured using the new generation 3D sensor Asus Xtion Pro Live. Gesture recognition is achieved by extracting and further comparing the normalized geometric skeletons of the hand. Hand skeletons are compared using Dynamic Time Warping algorithm, which has polynomial complexity.

Key words and phrases: gesture recognition, ASL, DTW, depth image.