

Інерційна система розпізнавання жестів

Ярошенко^f М. О., ORCID [0000-0002-3092-3856](https://orcid.org/0000-0002-3092-3856)

Варфоломєєв^s А. Ю., к.т.н., ORCID [0000-0002-6990-7140](https://orcid.org/0000-0002-6990-7140)

Яганов^s П. О., к.т.н., доц., ORCID [0000-0001-7358-9846](https://orcid.org/0000-0001-7358-9846)

Кафедра конструювання електронно-обчислювальної апаратури

Факультет електроніки

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Київ, Україна

Анотація—В роботі розглянуто інерційну систему розпізнавання жестів. Дана система побудована на основі апаратного модуля, що формує образи-жести за допомогою інерційних МЕМС давачів (акселерометра і гіроскопа) та розпізнає отримані образи за допомогою нейронної мережі з BLSTM шаром. Реалізовано прототип запропонованої системи, в якому апаратна частина побудована на мікроконтролері STM32F401RE та мікросхемі МЕМС давачів MPU9250, а розпізнавання жестів виконується на персональному комп'ютері засобами пакету прикладних програм МАТЛАВ. За допомогою розробленого прототипу сформовано навчальну та тестову вибірки для трьох типів жестів та показано спроможність підсистеми розпізнавання забезпечити похибку ідентифікації даних жестів не більше 3,5% (коефіцієнт правильного розпізнавання 96,7%).

Ключові слова — розпізнавання жестів; інерційна система; нейронні мережі; BLSTM

I. ВСТУП

Розвиток інформаційно-обчислювальних засобів електронних систем створює нові можливості для розв'язку складних ресурсномістких задач, які знаходять своє місце у сучасних системах штучного інтелекту. Серед них чільне місце займають системи розпізнавання образів. При цьому термін «образ» трактується не тільки у розумінні візуального образу, але і як деякий вектор, що однозначно характеризує стан об'єкта.

До розпізнавання образів належать також системи розпізнавання жестів людини. Ці системи знаходять своє застосування для комунікативного спілкування у різних сферах людської діяльності: від спорту і медицини до промисловості і бізнесу. Особливо це актуально для людей з обмеженими можливостями. Згідно даних Всесвітньої організації охорони здоров'я майже 5% населення планети мають проблеми зі слухом, а за даними Всесвітньої федерації глухих у світі налічують понад 70 млн. людей, для яких формою повсякденного спілкування є мова жестів [1].

Системи розпізнавання жестів активно впроваджуються у ігрові консолі відеоігор з віртуальною реальністю, персональні комп'ютери і смартфони [2]. Наприклад, смартфони серії Google Pixel 4 на додаток до системи розпізнавання обличчя Face Unlock мають здатність розпізнавати прості жести завдяки системі Motion Sense.

Розробка ефективних недорогих швидкодіючих систем з розпізнавання жестів триває, а їх інтеграція до складу систем зі штучним інтелектом є актуальною задачею.

Більшість сучасних систем розпізнавання жестів аналізують їх візуальний образ. Проблеми, які при цьому доводиться вирішувати, пов'язані з необхідністю виділити на загальному зображенні об'єкт, який ідентифікують як «жест». Потім цей образ потрібно класифікувати як інформативну складову. При цьому рівень освітленості, наявні світлотіні, колір світла, антропометричні особливості руки людини, що передає жест, положення руки тощо належать до факторів інформаційного шуму, який потрібно фільтрувати. Крім того, в системах реального часу слід визначати початок і кінець жесту у потоці відеокадрів. Для вирішення всіх перелічених вище складностей виникає потреба використання спеціальних алгоритмів обробки даних.

Переважає більшість цих алгоритмів базуються на методах і технологіях нейронних мереж. Оскільки нейронні мережі здатні ефективно здійснювати класифікацію та ідентифікацію образів, вони використовуються як інструмент вирішення різноманітних задач: від відносно простих одношарових пересептронних класифікаторів [3] до складних ідентифікаторів на глибоких нейромережах. В роботах [4, 5] повідомляється про використання згорткової нейронної мережі (CNN, англ. «Convolutional Neural Networks»), за допомогою якої здійснюють ідентифікацію жесту на загальному фоні зображення. Створюють також синтез алгоритмів обробки даних CNN, рекурентної нейронної мережі (RNN, англ. «Recurrent Neural Networks») та нейронної мережі з довгою короткостроковою пам'яттю (LSTM, англ. «Long Short-Term Memory») [6]. Точність ідентифікації жестів в таких системах зазвичай знаходиться в межах від 77% до 97% [2, 6]. Структури нейронних мереж при цьому можуть бути доволі складними, а кількість глибоких



шарів досягати десятків [7, 8]. Все це вимагає значних обчислювальних ресурсів, що можуть бути забезпечені лише сучасними надпотужними мікропроцесорами та графічними прискорювачами.

Водночас, для зменшення обчислювального навантаження на систему розпізнавання жестів, замість обробки візуальних образів, використовують спеціальні електронно-механічні пристрої, що реєструють рух за сигналами з давачів, створюючи попередній вектор-образ жесту, який потім ідентифікує нейронна мережа. Конструктивно такі пристрої виконані у формі рукавичок, що надівають на руку. Давачами сигналів зазвичай слугують або мікромеханічні інерційні пристрої (гіроскопи, акселерометри), або гнучкі тензорезистори, які змінюють свій опір під час деформації суглобів пальців (виробник Spectra Symbol Flex Sensor [9]). Такі апаратно-програмні рішення є менш вартісними у порівнянні з оптичними системами, не містять складних блоків попередньої обробки даних та потребують значно менших обчислювальних ресурсів. З огляду на це, в даній роботі вирішено зосередитись на дослідженні систем саме такого типу.

Таким чином, метою роботи є встановлення особливостей функціонування системи розпізнавання жестів на рекурентній BLSTM нейронній мережі з інерційними MEMS давачами. Поставлена мета досягається за рахунок вирішення наступних задач:

- розроблення апаратного пристрою для формування образу-жесту на основі інерційних MEMS давачів;
- створення процедури розпізнавання потокових даних на основі алгоритму функціонування нейронної мережі з довгою короткостроковою пам'яттю (LSTM).

Дослідження системи розпізнавання виконуватиметься на прикладі трьох жестів: рух по колу (circle), постукування (tap) та невідомий жест (other).

II. АПАРАТНИЙ БЛОК ЗАХОПЛЕННЯ РУХУ

Дані для розпізнавання жестів отримані за допомогою апаратного пристрою для захоплення руху, структурна схема якого показана на рис. 1, а.

Макет пристрою реалізовано засобами налагоджувальних плат NUCLEO-F401RE на базі мікроконтролера STM32F401RE, модуля GY-9250 з мікросхемою MPU9250, що включає трьохосові акселерометр і гіроскоп, а також модуля-перехідника Тройка Shield для забезпечення зв'язку між ними (рис. 1, б). Макет пристрою працює наступним чином. Мікроконтролер здійснює налаштування режимів роботи блоку давачів та отримує від них дані шиною I²C. Далі він пересилає отримані дані через віртуальний СОМ-порт на ПК з частотою 25 Гц, де в середовищі MATLAB виконується розпізнавання жестів.

Для зменшення впливу шумів використовуються внутрішні цифрові фільтри сенсора: 92 Гц частота зрізу для гіроскопа та 99 Гц – для акселерометра. Діапазон вимірювань акселерометра $\pm 2g$, гіроскопа – ± 250 °/с.

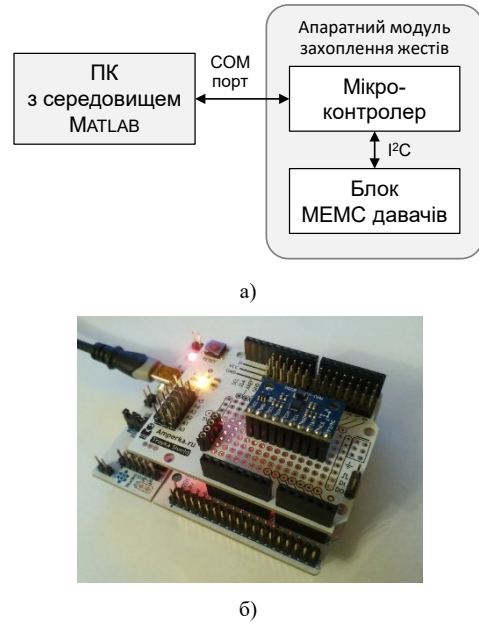


Рис. 1 Пристрій для захоплення руху: а) структурна схема пристрою; б) макет пристрою

III. ЗАГАЛЬНА СТРУКТУРА НЕЙРОННОЇ МЕРЕЖІ ДЛЯ РОЗПІЗНАВАННЯ ЖЕСТИВ

Для розпізнавання жестів використовується нейронна мережа з BLSTM шаром (цей шар також інколи називають BiLSTM) [10], яка належить до класу рекурентних. Ця нейромережа є модифікованою версією більш простої мережі LSTM, тому наведемо спрощену коротку відомість саме про неї.

Нейромережі LSTM мають властивість «запам'ятовувати» вектори-образи за рахунок існування у їх структурі зворотних зв'язків. У багатопшарових нейронних рекурентних нейромережах з'являється можливість охопити такими зв'язками не лише вихідний шар із вхідним, але і внутрішні шари у різних комбінаціях. Внаслідок цього у таких нейромережах з'являється можливість не тільки запам'ятовувати образ, але і накопичувати інформацію, здійснюючи на цій основі прогностичну функцію.

Особливість функціонування LSTM нейронної мережі полягає у тому, що синаптичні коефіцієнти петлі зворотнього зв'язку у ній не постійні, а програмно керовані. Цим досягають посилення ефекту самонавчання, оскільки процес накопичення інформації для прийняття рішень стає динамічним і залежить від вхідної послідовності векторів-образів [11]. Крім того, LSTM нейромережа має здатність «забувати» непотрібні складові інформації під час обробки послідовності даних. Це забезпечують спеціальні блоки-вентилі, що входять до її складу. Особливість їх функціонування ілюструє рис. 2.

Стан c_t комірки LSTM у момент часу t визначається наступним чином:

$$c_t = f_t \otimes c_{t-1} + g_t \otimes i_t,$$

а її прихований стан h_t як:

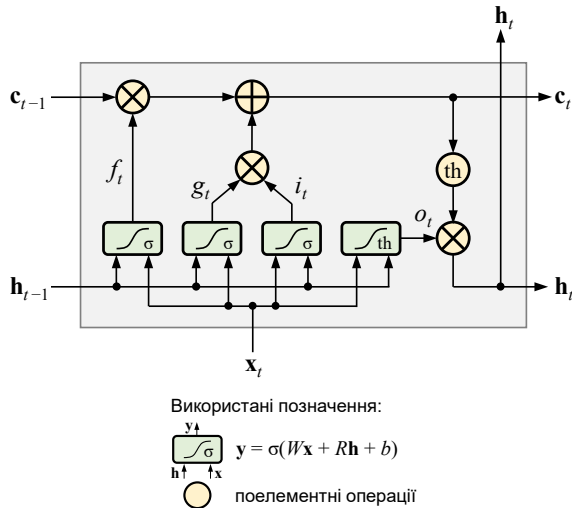


Рис. 2 Комірка LSTM

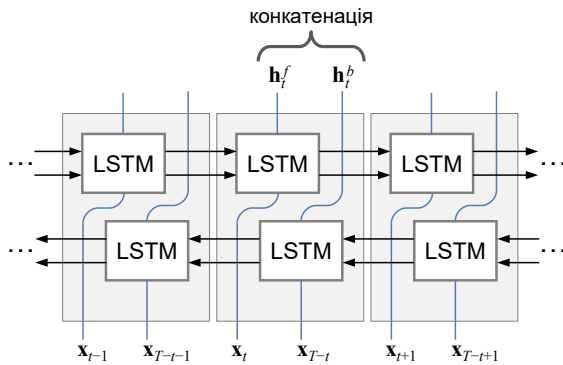


Рис. 3 Розгортка шару BLSTM нейронної мережі

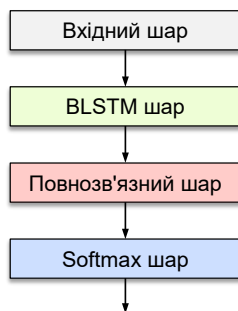


Рис. 4 Структура нейронної мережі для розпізнавання жестів із BLSTM шаром

$$\mathbf{h}_t = o_t \otimes \text{th}(\mathbf{c}_t),$$

де \otimes – позначає поелементне множення; $\text{th}(\cdot)$ також застосовується поелементно.

Значення f_t у комірку формується вентилем *забування* і контролює ступінь видалення старої інформації зі стану комірки. Значення g_t формується вентилем *кандидата* і визначає інформацію, що має додаватися до стану комірки. Значення i_t формується вентилем *входу* та задає рівень оновлення стану комірки. Значення o_t обчислюється вентилем *виходу* та контролює рівень інформації, що потрапляє зі стану комірки в її

прихований стан. Наведені значення обчислюються наступним чином:

$$\begin{aligned} f_t &= \sigma(W_f \mathbf{x}_t + R_f \mathbf{h}_{t-1} + b_f), \\ i_t &= \sigma(W_i \mathbf{x}_t + R_i \mathbf{h}_{t-1} + b_i), \\ g_t &= \sigma(W_g \mathbf{x}_t + R_g \mathbf{h}_{t-1} + b_g), \\ o_t &= \text{th}(W_o \mathbf{x}_t + R_o \mathbf{h}_{t-1} + b_o) \end{aligned} \quad (1)$$

де W – вагові коефіцієнти входу; R – рекурентні вагові коефіцієнти та b – зміщення; \mathbf{x}_t – вектор вхідного сигналу в поточний момент часу; \mathbf{h}_{t-1} – прихований стан комірки в попередній момент часу; $\sigma(x) = (1 + e^{-x})^{-1}$ – сигмоїдна функція активації. Добутки у виразах (1) є матричними або скалярними – залежно від розмірів \mathbf{x}_t , \mathbf{h}_{t-1} , W , R та b .

Розміри вхідного вектора і матриць W , R та b визначають кількість елементів векторів стану \mathbf{c}_t та прихованого стану \mathbf{h}_t комірки. Кількість елементів у векторі \mathbf{h}_t дорівнює кількості прихованих елементів LSTM шару.

Шар BLSTM мережі будується на основі LSTM комірок. На рис. 3 показано розгортку даного шару. По суті він включає дві LSTM комірки, що дозволяє краще виявляти взаємозв'язки між вимірюваннями у послідовності в двох напрямках – від початку до кінця послідовності та навпаки. Цей тип шару показав свою ефективність при розв'язанні таких задач аналізу послідовностей, як класифікація фону, розпізнавання тексту, розпізнавання ритму та інших. Також, BLSTM шар у випадку розпізнавання жестів є менш чутливим до довжини послідовності жесту за часом та зсуву жесту в часі [11].

Навчання LSTM та BLSTM шарів виконується тими ж методами, що і звичайних рекурентних мереж. Зокрема, одним з найбільш поширених способів є алгоритм *зворотного розповсюдження помилки в часі* (BPTT) [10, 11], який передбачає розгортання рекурентного шару у послідовність шарів без зворотного зв'язку, до яких далі застосовується звичайний алгоритм зворотного розповсюдження помилки.

Повна нейронна мережа, що виконує розпізнавання вхідної послідовності, має вигляд, показаний на рис. 4.

У нейронній мережі, показаній на рис. 4, вхідний шар виконує функцію поелементного передавання послідовності вхідного сигналу до наступного шару. Вхідний сигнал при цьому складається зі 100 нормованих до діапазону $[-1, +1]$ вимірювань акселерометра та гіроскопа по трьом осям, тобто по 6 значень на кожне вимірювання. Вимірювання здійснюються з частотою вибірки 25 Гц. Зазначений розмір послідовності та частота вимірювань обрані за рекомендаціями з [12]. Із вхідного шару дані послідовно передаються до шару BLSTM, який формує вихідний вектор з вдвічі більшою кількістю елементів, ніж число його прихованих вузлів, що обумовлено наявністю двох LSTM комірок у даному шарі. Після обробки всіх вимірювань вхідної послідовності вихід BLSTM шару передається на повнозв'язний шар з трьома нейронами, які класифікують отриману послідовність відповідно на 3 класи. Остаточна класифікація вхідної

послідовності виконується шаром з м'яким максимумом (Softmax шаром), на який передаються сигнали з попереднього повнозв'язного шару. Для повноти викладення наведемо формулу, за якою обчислюється значення функції Softmax:

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^N e^{x_j}},$$

де x_i та y_i – це відповідно i -й вхід та вихід шару; j – пробігає індекси всіх N виходів шару (в мережі, що розглядається $N = 3$).

Під час навчання мережі використовується функція втрат на основі взаємної ентропії, що мінімізує похибку мережі з вихідним Softmax-шаром:

$$L = \sum_{i=1}^N \sum_{j=1}^K t_{ij} \ln(y_{ij}),$$

де N – кількість послідовностей у навчальній вибірці; K – кількість класів; t_{ij} – індикатор належності i -ї послідовності навчальної вибірки j -му класу; y_{ij} – значення j -го виходу Softmax-шару для i -ї послідовності навчальної вибірки.

IV. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

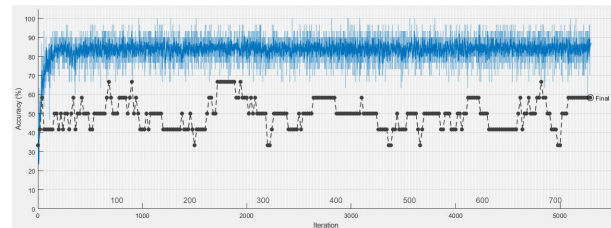
Для того, щоб виконувати розпізнавання жестів, застосовуючи описану в попередньому розділі нейронну мережу, її необхідно навчити. Крім того, необхідно також впевнитися, що навчена нейронна мережа здатна розпізнавати жести належним чином. Для цього за допомогою пристрою захоплення руху було отримано навчальну та тестову вибірки. Зокрема, навчальна вибірка складається з 228 послідовностей (по 76 послідовностей на кожний клас жестів) розміром 6×100 (100 вимірювань гіроскопа та акселерометра по кожній з 3-х осей). Тестова вибірка має аналогічну структуру, проте складається з 30 послідовностей (по 10 прикладів на кожен жест).

Для налаштування параметрів нейронної мережі було випробувано декілька методів навчання: стохастичний градієнтний спуск з інерційністю (SGDM), RMSProp та Adam [11]. Проте за досягнутим значенням коефіцієнта розпізнавання найкращим серед них виявився метод з адаптивною інерційністю Adam [13], в якому використовувались наступні параметри:

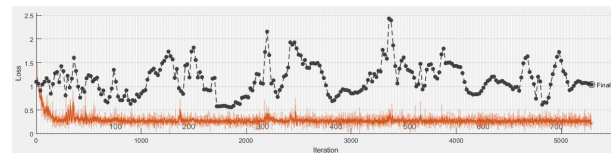
- швидкість навчання: $\alpha = 0,01$;
- коефіцієнт затухання градієнта: $\beta_1 = 0,99$;
- коефіцієнт затухання квадрата градієнта: $\beta_2 = 0,99$;
- параметр для запобігання ділення на 0: $\epsilon = 10^{-8}$;
- максимальна кількість епох: $N_E = 2000$;
- розмір навчальної підмножини (mini-batch) $B = 30$.

ТАБЛИЦЯ 1 Залежність коефіцієнту розпізнавання від кількості прихованих елементів у BLSTM шарі

Кількість прихованих елементів у BLSTM шарі	Точність розпізнавання, %
25	83,3
50	86,7
75	86,7
100	96,7
125	86,7



а)



б)

Рис. 5 Динаміка навчання нейронної мережі: а) зміна точності розпізнавання (синій колір, суцільна лінія); б) зміна значення функції втрат (червоний колір, суцільна лінія)

Під час експериментів було досліджено нейронні мережі з різною кількістю прихованих елементів у BLSTM шарі (див. табл. 1). Виявлено, що найкращий результат розпізнавання досягається при 100 прихованих елементах в BLSTM шарі. Цікаво відмітити, що при використанні 50, 75 та 125 прихованих елементів у BLSTM шарі, точність розпізнавання є однаковою і меншою та складає 86,7%. Причиною цього можуть бути особливості початкової ініціалізації ваг нейронів під час навчання, що призводить до потрапляння алгоритму оптимізації у локальний мінімум функції втрат.

Динаміка зміни точності розпізнавання під час навчання (співпадіння результатів роботи нейронної мережі із навчальною вибіркою) та зміни значення функції втрат наведена на рис 5. Зауважимо, що задовільна якість розпізнавання досягається вже при 750 епохах навчання.

Випробування нейронної мережі на тестовій вибірці, приклади з якої не приймали участь під час навчання, показало вірність розпізнавання жестів на рівні 96,7%, що може вважатися цілком прийнятним результатом.

ВИСНОВКИ

В роботі запропоновано та досліджено апаратно-програмний комплекс розпізнавання жестів. Встановлено, що апаратна реалізація блоку формування сигналів жестів можлива на доступній елементній базі – мікроконтролері STM32 та модуля MEMS давачів GY-9250, а класифікацію жестів може здійснювати рекурентна BLSTM нейромережа, синтезована та навчена засобами пакету прикладних програм

MATLAB. Процедура розпізнавання потокових даних на основі алгоритму функціонування нейронної мережі BLSTM при цьому здатна забезпечити похибку ідентифікації жестів в межах до 3,5%, що є прийнятним для подібного класу систем. Це обґрунтовує ефективність процесу розпізнавання жестів апаратно-програмними засобами, представленими у даній роботі, оскільки корисний результат досягається незначними затратами, які зумовили його одержання.

Експериментально визначено оптимальну кількість прихованих елементів у BLSTM шарі нейронної мережі заданої конфігурації, що опосередковано вказує на особливості функціонування системи, для якої характерна нелінійна залежності між кількістю прихованих елементів і якістю розпізнавання жестів.

Подальші дослідження можуть бути спрямовані на підвищення ідентифікаційних можливостей системи з розпізнавання жестів за рахунок підвищення роздільної здатності давачів апаратної частини та якості навчання BLSTM нейромережі. Крім того, може бути підвищена й автономність пристрою шляхом перенесення на мікроконтролер обчислень виходів нейронної мережі та передавання розпізнаного жесту через бездротовий канал.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] "World Federation of the Deaf," [Online]. Available: <http://wfdeaf.org/faq>. [Accessed 22 08 2019].
- [2] S. Alshabbab, A.-J. Gallego and M. Lozano, "Hand gesture detection with convolutional neural networks," *Advances in Intelligent Systems and Computing*, p. 45–52, 2018. DOI: [10.1007/978-3-319-94649-8_6](https://doi.org/10.1007/978-3-319-94649-8_6).
- [3] P. O. Yahanov and I. V. Redko, "Perseptronnyi klasyfikator teplovoho komfortu," *Visnyk KNUTD*, no. 6 (128), p. 29–38, 2018. DOI: [10.30857/1813-6796.2018.6.3](https://doi.org/10.30857/1813-6796.2018.6.3).
- [4] J. Tompson, M. Stein, Y. LeCun and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," *ACM Transactions on Graphics (ToG)*, vol. 33, no. 5, p. 169–173, 2014. DOI: [10.1145/2629500](https://doi.org/10.1145/2629500).
- [5] S. Ji, W. Xu, M. Yang and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*, vol. 35, no. 1, p. 221–231, 2013. DOI: [10.1109/TPAMI.2012.59](https://doi.org/10.1109/TPAMI.2012.59).
- [6] N. Nishida and H. Nakayama, "Multimodal gesture recognition using multi-stream recurrent neural network," *Image and Video Technology PSIVT 2015. Lecture Notes on Computer Science*, vol. 9431, p. 682–694, 2015. DOI: [10.1007/978-3-319-29451-3_54](https://doi.org/10.1007/978-3-319-29451-3_54).
- [7] N. Neverova, C. Wolf, G. W. Taylor and F. Nebout, "Multi-scale deep learning for gesture detection and localization," *Computer Vision – ECCV 2014 Workshops. ECCV. Lecture Notes in Computer Science*, vol. 8925, p. 474–490, 2014. DOI: [10.1007/978-3-319-16178-5_33](https://doi.org/10.1007/978-3-319-16178-5_33).
- [8] R. V. Siriak and I. S. Skarha-Bandurova, "Model obrobky potokovykh danykh dlia rozpoznavannia okremykh odynyts zhestovoi movy," *Visnyk Natsionalnoho tekhnichnoho universytetu "KhPI"*, vol. 42, no. 1318, p. 73–81, 2018.
- [9] "Spectra Symbol Flex Sensor," [Online]. Available: <https://www.spectrasymbol.com/product/flex-sensors/>. [Accessed 22 08 2019].
- [10] A. Graves and J. Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures," *Neural Networks*, vol. 18, no. 5–6, p. 602–610, 2005. DOI: [10.1016/j.neunet.2005.06.042](https://doi.org/10.1016/j.neunet.2005.06.042).
- [11] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, Massachusetts: MIT Press, 2016. ISBN: [978-0262035613](https://doi.org/978-0262035613).
- [12] G. Lefebvre, S. Berlemont, F. Mamalet and C. Garcia, "Inertial Gesture Recognition with BLSTM-RNN," in *Artificial Neural Networks Methods and Applications in Bio-/Neuroinformatics*, P. Koprinkova-Hristova, V. Mladenov and K. N. K., Eds., Springer International Publishing Switzerland, 2015, pp. 393–410. DOI: [10.1007/978-3-319-09903-3_19](https://doi.org/10.1007/978-3-319-09903-3_19).
- [13] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations (ICLR)*, San Diego, 2015. URL: <https://arxiv.org/pdf/1412.6980.pdf>.

Надійшла до редакції 20 вересня 2019 р.

УДК 004.358

Инерционная система распознавания жестов

Ярошенко^f М. А., ORCID [0000-0002-3092-3856](https://orcid.org/0000-0002-3092-3856)

Варфоломеев^s А. Ю., к.т.н., ORCID [0000-0002-6990-7140](https://orcid.org/0000-0002-6990-7140)

Яганов^s П. А., к.т.н., доц., ORCID [0000-0001-7358-9846](https://orcid.org/0000-0001-7358-9846)

Кафедра конструирования электронно-вычислительной аппаратуры
Факультет электроники
Национальный технический университет Украины
"Киевский политехнический институт имени Игоря Сикорского"
Киев, Украина

Аннотация— В работе рассмотрены инерционную систему распознавания жестов. Данная система построена на основе аппаратного модуля, формирующего образы-жесты с помощью инерционных МЭМС датчиков



(акселерометра и гироскопа) и распознает полученные образы с помощью нейронной сети с BLSTM слоем. Реализовано прототип предложенной системы, в котором аппаратная часть построена на микроконтроллере STM32F401RE и микросхеме МЭМС давлений MPU9250, а распознавание жестов выполняется на персональном компьютере средствами пакета прикладных программ MATLAB. С помощью разработанного прототипа сформирован учебную и тестовую выборки для трех типов жестов и показано способность подсистемы распознавания обеспечить погрешность идентификации данных жестов не более 3,5% (коэффициент правильного распознавания 96,7%).

Ключевые слова — распознавание жестов; инерционная система; нейронные сети; BLSTM

UDC 004.358

The Inertial Gesture Recognition System

M. A. Yaroshenko^f, ORCID [0000-0002-3092-3856](https://orcid.org/0000-0002-3092-3856)

A. Yu. Varfolomeiev^s, PhD, ORCID [0000-0002-6990-7140](https://orcid.org/0000-0002-6990-7140)

P. A. Yaganov^s, PhD Assoc.Prof., ORCID [0000-0001-7358-9846](https://orcid.org/0000-0001-7358-9846)

Department of Design of Electronic Computing Equipment

Faculty of Electronics

National technical university of Ukraine “Igor Sikorsky Kyiv polytechnic institute”

Kyiv, Ukraine

Abstract—The paper is devoted to the implementation of the gestures recognition system. Existing gesture recognition approaches are considered. Among them the system of inertial type is chosen for research, since the systems of this type are less difficult to implement and require relatively low computational complexity. Thus, the proposed system is based on two components: the hardware module for gesture capturing and the subsystem for the recognition of captured gestures. The hardware component of the system employs MEMS accelerometer and gyroscope for motion acquisition. The prototype of this component is implemented using the STM32F401RE microcontroller and the MEMS sensor MPU9250. The communication between the microcontroller and the MEMS sensor is carried out via the I2C bus. The microcontroller initializes the sensor and acquires the motion data. Then it sends the collected data via the virtual COM port to the PC, where the captured motions are recognized. The system component for the recognition of gestures is implemented as the neural network in MATLAB environment. This network consists of four layers: the input layer, the BLSTM layer, the fully-connected layer, and the Softmax layer. The key component of the selected neural network architecture is the BLSTM layer, which (due to its properties) provides some invariance to the length and the shift of the gesture signal in time. The suggested system was tested on three gesture classes: circular motion, tapping and other/unknown gesture. Using the hardware module the training and the testing sets were collected. In course of experiments, it was established that the best training results for the selected network architecture and the collected training set are achieved using the Adam optimizer. The experiment with different number of hidden units in the BLSTM layer showed that the highest recognition rate may be achieved when 100 units are used. Particularly, the suggested system with the specified number of hidden units in BLSTM layer was able to attain the recognition error less than 3.5% (recognition rate of 96.7%) on the test set.

Keywords — gesture recognition; inertial system; neural networks; BLSTM

