

doi:10.15199/48.2017.08.35

Algorytmy rozpoznawania ruchów głowy w elektronicznym interfejsie człowiek-komputer

Streszczenie. Celem badań opisanych w niniejszym komunikacie jest wykorzystanie specjalnego interfejsu dla osoby niewidomej umożliwiającego sterowanie menu urządzenia nawigacyjnego za pomocą ruchów głowy. Zastosowano czujniki inercyjne wbudowane w układ stereowizyjny umieszczony na głowie użytkownika systemu. Dla 2-sekundowych okien czasowych rejestrowanych sygnałów wyliczono parametry statystyczne takie jak: średnie, minimum, maksimum, odchylenie standardowe, kurtosis oraz współczynniki korelacji dla sygnałów pochodzących z trój-osiowego akcelerometru i trój-osiowego żyroskopu. Dla tak przygotowanych danych treningowych zbadano skuteczność działania czterech różnych klasyfikatorów do rozpoznawania różnych ruchów głowy służących do sterowania interfejsem. Najlepsze wyniki uzyskano dla tzw. lasu losowego, którego skuteczność dla zbioru sygnałów testowych wyniosła 98,89%. Prawidłowe rozpoznawanie ruchów głowy umożliwi osobie niewidomej komunikowanie się z urządzeniem mobilnym i zarządzanie aplikacjami za pomocą ruchów głowy.

Abstract. The purpose of the study described in this communication is to apply a special interface for the blind person to control the menu of the navigation device using head movements. Inertial sensors built into the stereovision system positioned on the user's head are used. For two-second time windows of the recorded signals, statistical parameters such as: mean, minimum, maximum, standard deviation, kurtosis, and correlation coefficients for signals from a three-axis accelerometer and a three-axis gyroscope were calculated. For such training data, the effectiveness of the four different classifiers was assessed to distinguish different head movements. The best results were obtained for the so-called random forest whose efficiency for the test set signals achieved 98.89%. Proper recognition of head movements will allow the blind person to communicate with the mobile interface and control applications by means of head movements. (**Algorithms for head movements' recognition in an electronic human-computer interface**).

Słowa kluczowe: elektroniczny interfejs człowiek-maszyna, rozpoznawanie gestów, czujniki inercyjne, IMU, MEMS

Keywords: electronic human-machine interface, gesture recognition, inertial sensors, IMU, MEMS

Wprowadzenie

Systemy interakcji człowieka z systemami technicznymi (ang. *Human-System Interaction – HSI*) stanowią dzisiaj wyodrębnioną dziedzinę badawczą poświęconą rozwijaniu nowych technologii komunikacji człowieka z systemami. Zależnie od zastosowania systemy interakcji określa się też mianem interfejsów człowiek-komputer (ang. *Human-Computer Interface – HCI*) lub bardziej ogólnie interfejsów człowiek-maszyna (ang. *Human-Machine Interface*). Projektowanie i budowa takich systemów wymaga interdyscyplinarnego podejścia badawczego, obejmującego wiedzę z zakresu mechanizmów spostrzegania zmysłowego, procesów poznawczych i przetwarzania informacji przez człowieka oraz podstaw ergonomii. Dobrze zaprojektowany interfejs użytkownika często decyduje o użyteczności całego systemu [1].

Projektowanie interfejsów dostępnych dla osób z niepełnopraktycznymi zmysłowymi i ruchowymi jest szczególnie trudnym zagadnieniem badawczym [2, 3]. Dla niepełnosprawnych użytkowników konieczne staje się opracowanie innowacyjnych rozwiązań umożliwiających komunikację z komputerem. W interfejsach wykorzystuje się alternatywne, często wielomodalne sposoby komunikacji zastępujące utracone funkcje zmysłowe lub fizyczne użytkownika [4]. Osoby z utratą słuchu posługują się językiem migowym (należącym do tzw. języków wizualno-przestrzennych) i korzystają z aplikacji wyświetlających język migowy. Dla osób z niepełnosprawnościami fizycznymi budowane są specjalnej konstrukcji urządzenia wejścia-wyjścia (m.in. z wykorzystaniem czujników inercyjnych) oraz innowacyjne interfejsy umożliwiające komunikację „bezdotykową” z komputerem wykorzystujące, m.in. interfejsy mózg-komputer (ang. *Brain-Computer Interface – BCI*) [5], interfejsy wizyjne m.in. śledzące ruch gałek ocznych [6] lub wykrywające intencjonalne mrugnienia [7]. Osoby niewidome komunikują się z komputerami z wykorzystaniem syntezy mowy oraz tzw. monitorów brajlowskich umożliwiających pisanie i wyświetlanie alfabetu Braille'a. Bardziej złożonym problemem dla osób niewidomych jest manualna obsługa

urządzeń podczas marszu. Podstawową pomocą, z której korzystają osoby niewidome w czasie poruszania się jest biała laska. Angażuje ona jedną rękę i obsługa dodatkowego urządzenia, np. nawigacyjnego, podczas marszu jest utrudniona.

Celem badań opisanych w niniejszym komunikacie jest opracowanie specjalnego interfejsu dla osoby niewidomej umożliwiającego sterowanie menu urządzenia nawigacyjnego za pomocą ruchów głowy. Motywacją do podjęcia tych badań są prowadzone w Instytucie Elektroniki wieloletnie prace badawcze i rozwojowe (m.in. w ramach grantów krajowych i europejskich) ukierunkowane na opracowanie osobistych systemów nawigacji dla osób niewidomych [8].

Rozpoznawanie gestów z zastosowaniem czujników inercyjnych

Sterowanie aplikacjami komputerowymi odbywa się tradycyjnie przy pomocy klawiatury i/lub myszy komputerowej. Jednak coraz częściej można spotkać rozwiązania, umożliwiające pracę z komputerem osobom z różnymi chorobami kończyn, które mają ograniczoną możliwość korzystania z tych peryferyjnych urządzeń. W pracy [9] pokazano interfejs działający na podstawie rozpoznawania ruchów głowy i jednocześnie ruchów ust, i w zależności od kombinacji tych ruchów zarządzanie programem komputerowym. Nie tylko aplikacje komputerowe mogą być sterowane bez pomocy rąk. Ruch elektrycznego wózka osoby niepełnosprawnej może być sterowany za pomocą rozpoznanych ruchów głowy [10], a nawet sygnałów EEG, co zostało zaprezentowane w pracy [11].

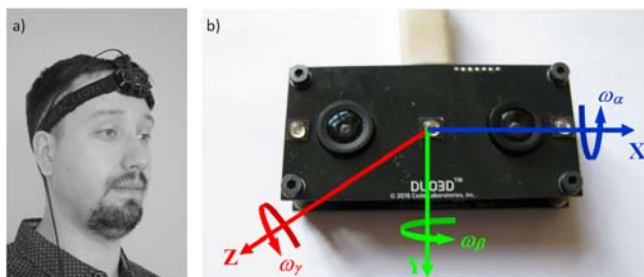
Współczesne technologie mikroukładów elektromechanicznych MEMS (ang. *Micro Electro Mechanical Systems*) umożliwiły budowanie miniaturowych i tanich czujników inercyjnych, a w konsekwencji ich coraz szersze zastosowania w interfejsach człowiek-komputer [12]. W pracy [13] w celu rozpoznania ośmiu różnych typów aktywności zastosowano trój-osiowy akcelerometr umieszczony na nadgarstku dominującej ręki. W pracy [14] wykorzystano

pięć czujników przyspieszenia do rozpoznania ruchu rąk oraz rozróżnienia chodu od bezruchu. Czujniki te były umieszczone na klatce piersiowej oraz na wszystkich kończynach. W pracy [15] opisano system, który może służyć do monitorowania ruchów ciała osób z chorobami neurodegeneracyjnymi. Analizowano sygnały pochodzące z trzech czujników, przy czym jeden zamocowany był na głowie. W pracy [16] zastosowano jednocześnie przetwarzanie obrazów oraz sygnałów z czujników inercyjnych do estymacji ruchu głowy.

W wymienionych pracach pierwszym krokiem analizy sygnałów rejestrowanych z czujników inercyjnych polega na parametryzacji sygnałów dla zadanych okien czasowych. Wyliczone są m.in. wartość średnia, odchylenie standardowe, energia sygnału i współczynniki korelacji pomiędzy sygnałami [13, 14]. Pozyskane parametry służą następnie do budowania odpowiednich klasyfikatorów rozpoznających zadane rodzaje ruchów.

Rejestracja sygnałów z czujników inercyjnych

W badaniu wykorzystano zestaw pomiarowy DUO MLX wyposażony w kamerę stereowizyjną, elektroniczny żyroskop i czujnik przyspieszenia, termometr oraz magnetometr [19]. DUO MLX ma małe rozmiary: 52x25x13 mm i waży 12,5 g. Dzięki tym właściwościom może być umieszczony na głowie użytkownika, jak pokazano na rysunku 1a. Układ ten umożliwia rejestrację sekwencji obrazów głębi otoczenia. Został wykorzystany w systemie wspomagającym osobę niewidomą w lokalizacji przeszkód i samodzielnym poruszaniu się.



Rys.1. Sposób zamocowania czujnika na głowie użytkownika a), układ stereowizyjny DUO MLX wyposażony w czujniki inercyjne wraz z zaznaczonymi osiami współrzędnych b)

W niniejszej pracy skupiono się na wykorzystaniu sygnałów z czujników inercyjnych układu DUO MLX, tj. trój-osiowego akcelerometru i trój-osiowego żyroskopu do sterowania menu interfejsu wspomagającego osobę niewidomą. Celem badania było automatyczne rozpoznawanie różnych ruchów głowy.

Przykładowe wykresy zarejestrowanych sygnałów przez czujnik przyspieszenia i żyroskop pokazano odpowiednio na rysunku 2a i rysunku 2b. Sygnały z czujników były rejestrowane z szybkością 100 próbek na sekundę.

W badaniu wzięło udział 8 osób – cztery kobiety (wiek: 34, 35, 37 i 37 lat) i czterech mężczyzn (wiek: 29, 36, 37 i 39 lat). Przy czym dla siedmiu osób wykonano rejestracje wchodzące w skład zbioru sygnałów treningowych oraz zbioru sygnałów testowych. Natomiast dla jednej losowo wybranej osoby wykonano tylko rejestracje przeznaczone dla zbioru sygnałów testowych.

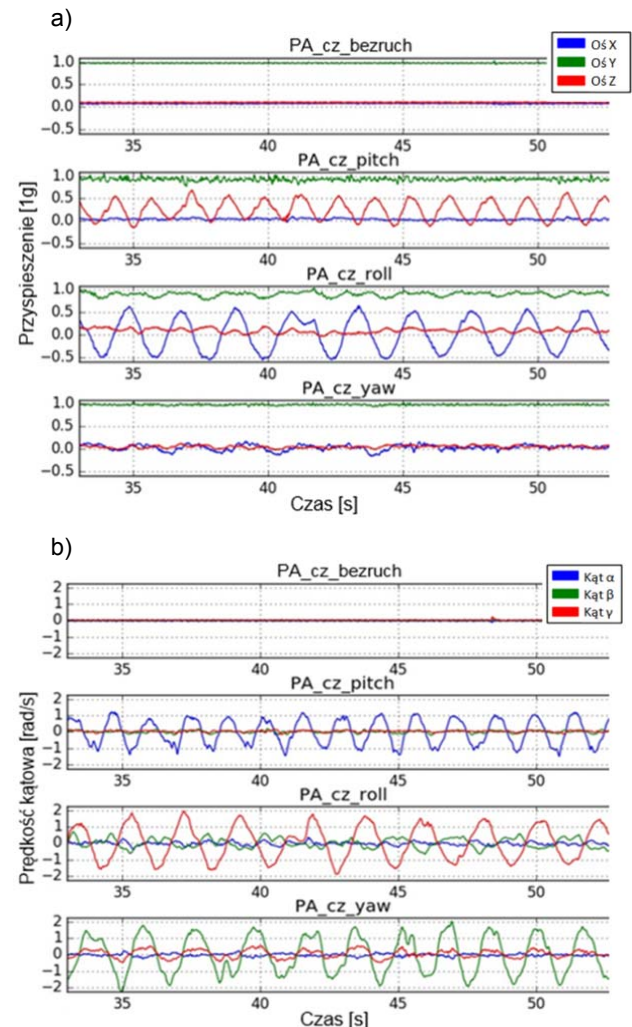
Nagrania ruchów głowy, które weszły w skład zbioru treningowego, zostały wykonane według następującego schematu:

10 sekund bezruch – 150 sekund (2,5 minuty) zadany ruch – 10 sekund bezruch

Zadany ruch to: *roll* (ruch głowy „na boki”), *pitch* (ruch głowy „górną-dół”) oraz *yaw* (ruch głowy „lewo-prawo”). Z nagrań tych wybrano 90-sekundowe odcinki sygnałów odpowiadających zadany ruchom, które weszły w skład zbioru treningowego.

Natomiast zapisy sygnałów stanowiących zbiór testowy wykonano wg następującego schematu:

20 sekund bezruch – 40 sekund yaw – 20 sekund bezruch – 40 sekund roll – 20 sekund bezruch – 40 sekund pitch – 20 sekund bezruch



Rys.2. Przebiegi sygnałów rejestrowanych przez: czujnik przyspieszeń a), żyroskop b)

Zbiór treningowy został dodatkowo podzielony na dwie części: uczącą oraz walidacyjną. Zbiór walidacyjny służył do oceny uzyskanego modelu rozróżniania ruchów głowy. Ponieważ w badaniu zbiór treningowy był duży, do podziału na część uczącą i walidacyjną wykorzystano metodę *holdout*, tj. podzielono zbiór na dwie części w stosunku 2:1 [17]. Zbiór testowy zaś, to zbiór rejestracji, które nie weszły w skład zbioru uczącego. Zbiór ten służył do weryfikacji poprawności działania klasyfikatora.

Rozpoznawanie ruchów głowy – wyniki

Sygnały, rejestrowane z czujników inercyjnych wg opisanych powyżej dwóch schematów, podzielono na odcinki dwusekundowe, z krokiem 0,1 sekundy. Dla każdego takiego odcinka i dla sześciu rejestrowanych sygnałów, tj. trzech sygnałów z akcelerometru (x, y, z)

i trzech sygnałów z żyroskopu ($\omega_x, \omega_y, \omega_z$) wyznaczono 36 parametrów: 1) średnie, 2) minima, 3) maksima, 4) odchylenia standardowe, 5) kurtozy, 6) współczynniki korelacji dla par sygnałów z akcelerometru i par sygnałów z żyroskopu. W ten sposób dla każdej rejestracji (tj. dla każdego rodzaju ruchu i bezruchu) wchodzącej w skład zbioru treningowego utworzono zbiór 900 wektorów z 36 parametrami opisującymi zadany ruch.

Do klasyfikacji ruchów głowy zastosowano tzw. drzewa decyzyjne, które są nieparametryczną metodą nadzorowanego uczenia maszynowego [18]. Drzewa decyzyjne pozwalają automatycznie budować reguły decyzyjne dzielące zbiór danych na rozłączne klasy. Drzewo decyzyjne można zdefiniować jako skierowany graf acykliczny i spójny. Składa się on z wierzchołków i łączących je krawędzi skierowanych. Drzewo decyzyjne posiada jeden wierzchołek początkowy, nazywany korzeniem. Węzły końcowe nazywamy liśćmi, zaś krawędzie gałęziami. W każdym węźle, oprócz węzłów końcowych (liści), ma miejsce podział elementów zbioru treningowego na podgrupy.

Istnieją różne algorytmy tworzenia drzew decyzyjnych. W niniejszej pracy zastosowano algorytm CART (ang. *Classification and Regression Trees*), w którym dla reguł podziału drzewa decyzyjnego zastosowano tzw. współczynnik Giniego jako miarę różnorodności klas w węźle [18].

Drzewa klasyfikacyjne można łączyć w grupy i w ten sposób tworzyć tzw. lasy losowe. W przypadku lasu losowego każde drzewo klasyfikacyjne uczone jest na zbiorze wektorów, który jest tworzony przez losowanie ze zwracaniem n -elementowego zbioru treningowego ze zbioru n wektorów. Następnie w każdym węźle takiego drzewa jest losowanych bez zwracania m spośród p cech wektora obserwacji ($m \leq p$) i na podstawie tych wybranych cech dokonywany jest podział. Końcowa klasyfikacja wektora odbywa się na zasadzie głosowania, tj. wektor poddawany jest klasyfikacji przez wszystkie drzewa lasu i zaklasyfikowany do klasy, która uzyskała najwięcej głosów. Dobre wyniki klasyfikacji otrzymuje się dla $m = \sqrt{p}$ [18]. Lasy losowe są skutecznymi klasyfikatorami dla bardzo licznych zbiorów danych i danych o bardzo dużych wymiarach.

Tabela 1. Sprawność klasyfikatorów dla danych testowych (liczba wektorów w zbiorze testowym 16074)

L.p.	Klasyfikator	Dane testowe – liczba poprawnie sklasyfikowanych wektorów	Sprawność klasyfikatora [%]
1	Drzewo decyzyjne	15516	96,53
2	Drzewo decyzyjne z min. 5 próbkami w liściu	15472	96,25
3	Las losowy	15835	98,51
4	Las losowy z min. 5 próbkami w liściu	15832	98,49

W przeprowadzonych badaniach zastosowano następujące klasyfikatory: 1) drzewo decyzyjne, 2) drzewo decyzyjne z minimalną liczbą próbek przypadających na liść równą 5, 3) las losowy złożony z 10 drzew decyzyjnych oraz 4) las losowy złożony z 10 drzew decyzyjnych, w których określono minimalną liczbę próbek przypadających na liść równą 5. Na podstawie uzyskanych wyników dla danych walidacyjnych obliczono tzw. sprawność klasyfikatora jako wyrażony w procentach stosunek liczby poprawnych klasyfikacji do liczby wszystkich wektorów w zbiorze walidacyjnym. Uzyskano sprawność wszystkich

czterech rozpatrywanych klasyfikatorów dla danych walidacyjnych nie mniejszą niż 99%.

Następnie sprawdzono poprawność działania klasyfikatorów dla nagrań testowych. Umożliwiło to weryfikację poprawności reguł każdego z klasyfikatorów oraz wyliczenie łącznej sprawności każdego z klasyfikatorów. Uzyskane wyniki zebrano w tabeli 1.

W zbiorze danych testowych znalazły się nagrania osoby, która nie nagrała sygnałów treningowych. Wyniki poprawnego rozpoznania ruchów głowy tej osoby zawiera tabela 2.

Tabela 2. Sprawność klasyfikatorów dla danych testowych osoby, która nie brała udziału w nagraniu danych treningowych (liczba wektorów 1989)

L.p.	Klasyfikator	Dane testowe – liczba poprawnie sklasyfikowanych wektorów	Sprawność klasyfikatora [%]
1	Drzewo decyzyjne	1930	97,03
2	Drzewo decyzyjne z min. 5 próbkami w liściu	1918	96,43
3	Las losowy	1967	98,89
4	Las losowy z min. 5 próbkami w liściu	1967	98,89

Wnioski i podsumowanie

Uzyskane wyniki rozpoznawania ruchów głowy pozwalają na sformułowanie następujących wniosków:

1. Drzewa decyzyjne oraz grupy drzew decyzyjnych (lasy losowe) są skutecznymi klasyfikatorami sygnałów z czujników inercyjnych umożliwiającymi rozpoznawanie ruchów głowy.
2. Klasyfikatory zbudowane dla zbiorów sygnałów rejestrowanych od siedmiu osób osiągnęły poprawność rozpoznawania ruchów głowy nie gorszą niż 96,25%.
3. Największą sprawność uzyskano dla lasów losowych: 98,89% poprawnie rozpoznanych ruchów.
4. Wyniki rozpoznawania rodzaju ruchów głowy dla danych testowych, rejestrowanych u osoby, która nie nagrała danych treningowych, są porównywalne do nagrań osób, które nagrały także nagrania treningowe.

Planowanym zastosowaniem opracowanej metody klasyfikacji ruchów głowy jest sterowanie interfejsem systemu mobilnego wspomagającego osoby niewidome w orientacji przestrzennej i samodzielnym poruszaniu się. W związku z tym planowane jest powiększenie zbiorów treningowych i testowych o nagrania pochodzące od większej liczby osób. Ponadto badanie zostanie rozszerzone o dodatkowe nagrania innych klas ruchów takich jak: marsz, siadanie / wstawanie, wchodzenie / schodzenie po schodach oraz ich klasyfikację.

Autorzy: mgr inż. Anna Borowska-Terka, Politechnika Łódzka, Instytut Elektroniki, ul. Wólczańska 211/215, 90-924 Łódź, E-mail: anna.borowska-terka@p.lodz.pl; prof. dr hab. inż. Paweł Strumiłło, Politechnika Łódzka, Instytut Elektroniki, ul. Wólczańska 211/215, 90-924 Łódź, E-mail: pawel.strumillo@p.lodz.pl.

LITERATURA

- [1] Nielsen J., Usability Engineering, Morgan Kaufman, Amsterdam, 2009.
- [2] Helal A., Mounir M., Abdulrazak B. (Eds.): The Engineering Handbook of Smart Technology for Aging, Disability, and Independence, John Wiley & Sons, Inc. (2008)
- [3] Strumiłło P., Materka A., Królak A., Systemy interakcji człowiek-komputer dla osób niepełnosprawnych, *Biuletyn Techniczno-Informacyjny Zarządu Oddziału Łódzkiego SEP*, nr.1 (2011), 2-9
- [4] Dumas B., Lalanne D., Oviatt S., Multimodal Interfaces: A Survey of Principles, Models and Frameworks, *Human*

- Machine Interaction*, Springer-Verlag, Berlin Heidelberg, (2009), 3-26
- [5] Poryzala P., Materka A., Cluster analysis of CCA coefficients for robust detection of the asynchronous SSVEPs in brain-computer interfaces, *Biomedical Signal Processing and Control*, Vol. 10, March 2014, 201-208
- [6] Kocejko T., Bujnowski A., Wtorek J., Eye mouse for disabled, *IEEE Conference on Human System Interaction (HIS)*, Kraków, czerwiec, 2008
- [7] Królak A., Strumiłło P., Eye-blink detection system for human-computer interaction, *International Journal on Universal Access in the Information Society*, vol. 11, no. 4 (2011), 409-419
- [8] Strumiłło P., Elektroniczne systemy nawigacji osobistej dla niewidomych i słabowidzących, Politechnika Łódzka, 2012 (<http://cybra.lodz.pl/dlibra/doccontent?id=3772>)
- [9] Song Y., Luo Y., Lin J., Detection of Movements of Head and Mouth to Provide Computer Access for Disabled, *2011 International Conference on Technologies and Applications of Artificial Intelligence*, 11–13 Nov. 2011.
- [10] Jia P., Hu H.H., Lu T., Yuan K., Head gesture recognition for hands-free control of an intelligent wheelchair, *Industrial Robot: An International Journal*, vol. 34 (2007), no.1, 60-68
- [11] Matsuzawa K., Ishii C., Control of an electric wheelchair with a brain-computer interface headset, *2016 International Conference on Advanced Mechatronic Systems (ICAMechS)*, Nov. 30-Dec. 3 2016
- [12] Mitra S., Acharya T., Gesture Recognition: A Survey, *IEEE Trans. Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37 (2007), no.3, 311-324
- [13] Yang J.-Y., Wang J.-S., Chen Y.-P., Using acceleration measurements for activity recognition: an effective learning algorithm for constructing neural classifiers, *Pattern Recognition Letters*, vol. 29, no. 16, 1st December 2008, 2213-2220
- [14] Kupryjanow A., Kaszuba K., Rozpoznawanie kategorii ruchu ludzkiego na podstawie analizy sygnałów pochodzących z trójosiowych czujników przyspieszenia, *Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej* Nr 26, XIX Seminarium Zastosowanie Komputerów w Nauce i Technice' 2009, Oddział Gdański PTETiS, Referat 17
- [15] Lorenzi P., Rao R., Romano G., Kita A., Irrera F., Mobile Devices for the Real-time detection of specific human motion disorders, *IEEE Sensors Journal*, vol, 16, no. 23, Dec.1, 2016, 8220-8227
- [16] He C., Kazanzides P., T.n Sen H., Kimand S., Liu Y., *An Inertial and Optical Sensor Fusion Approach for Six Degree-of-Freedom Pose Estimation*, *Sensors* 2015, 15(7), 16448-16465
- [17] Stapor K., *Automatyczna klasyfikacja obiektów*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2005
- [18] Koronacki J., Ćwik J., *Statystyczne systemy uczące się*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2015
- [19] Strona producenta DUO MLX: <https://duo3d.com/>