

SYSTEMY WIZYJNE SWOBODNEGO PUNKTU WIDZENIA O WYSOKIEJ POSTRZEGANEJ JAKOŚCI USŁUG FREE-VIEWPOINT VIDEO SYSTEMS WITH HIGH QUALITY OF EXPERIENCE

Adam Grzelka

Politechnika Poznańska, Poznań,
adam.grzelka@put.poznan.pl

Streszczenie: Artykuł dotyczy systemów swobodnego punktu widzenia, które są systemami wizyjnymi umożliwiającymi użytkownikowi wybór kierunku i miejsca obserwowania prezentowanej sceny. Autor podejmuje za cel zapewnienie wysokiej postrzeganej jakości usług w takich systemach, który realizuje poprzez szereg propozycji oryginalnych rozwiązań. Rozwiązania te dotyczą między innymi optymalizacji ustawienia kamer systemu wielokamerowego, zbadania wpływu stratnej kompresji, opóźnień transmisji do terminala użytkownika oraz realizacji nowego systemu wielokamerowego i serwera swobodnej nawigacji.

Abstract: The work deals with free-viewpoint video systems, which allow a user to choose the location and viewing direction for observing the presented scene. The author proposes a set of original solutions to obtain a high-quality experience in such systems. These proposals relate to optimizing the positioning of the cameras in a multi-camera system, investigating the impact of lossy compression, the transmission delay to the user terminal, and implementing a new multi-camera system and a free-viewpoint server.

Słowa kluczowe: estymacja map głębi, postrzegana jakość usług, swobodny punkt widzenia, synteza widoku wirtualnego

Keywords: depth map estimation, free-viewpoint, quality of experience, virtual view synthesis

1. WSTĘP

Systemy swobodnego punktu widzenia (ang. Free-viewpoint Video – FVV; Free-viewpoint Television – FTV)[20][5] umożliwiają użytkownikowi samodzielny wybór miejsca i kierunku, z którego obserwuje on zarejestrowaną scenę. Widok prezentowany użytkownikowi nie musi pochodzić bezpośrednio z rzeczywistej kamery, lecz być wirtualnie wygenerowany na podstawie obrazów z wielu kamer, rozmieszczonych wokół sceny. Dzięki temu użytkownicy mogą swobodnie „poruszać się” w przestrzeni wizualnej – niezależnie od siebie i w czasie rzeczywistym.

Systemy tego typu znajdują zastosowanie m.in. w transmisjach wydarzeń sportowych czy koncertów, a także w rozwiązaniach immersyjnych, takich jak wyświetlacze nagłowne (HMD)[17]. Kluczowym wyzwaniem w ich

projektowaniu jest zapewnienie wysokiej jakości doświadczenia użytkownika (Quality of Experience – QoE), która zależy nie tylko od parametrów technicznych, ale również od subiektywnego komfortu odbiorcy.

Celem pracy, jest znalezienie rozwiązań, które umożliwiają uzyskanie wysokiej jakości postrzeganej (QoE) w systemach swobodnego punktu widzenia. Autor skoncentrował się na analizie, jak różne elementy systemu – w tym ustawienie kamer, kompresja danych i opóźnienia transmisji – wpływają na odbiór jakości przez użytkownika. Autor przeprowadził badania zarówno teoretyczne, jak i eksperymentalne w czterech obszarach: konfiguracja kamer, kompresja stratna, opóźnienia w wyświetlaczach nagłownych oraz budowa prototypowego systemu swobodnej nawigacji. Pełen opis przeprowadzonych badań można znaleźć w [10].

2. SYSTEM SWOBODNEJ NAWIGACJI

Projektowanie systemów swobodnej nawigacji wymaga uwzględnienia szeregu aspektów technicznych: od rozmieszczenia i synchronizacji kamer, przez przetwarzanie obrazu (w tym estymację głębi[13]), aż po syntezę widoków[8] i transmisję danych do terminala użytkownika[5]. W literaturze opisano wiele różnych metod reprezentacji sceny trójwymiarowej. Należą do nich m.in. chmury punktów[21], siatki trójkątów, reprezentacje objętościowe[24] czy opisy geometryczne z teksturowaniem[19]. Dostępne są również rozwiązania oparte na wielowidokowej wizji z dodatkowymi mapami głębi (MVD – Multiview Video plus Depth)[16][12][3], które pozwalają na rekonstrukcję widoku z dowolnego punktu przy pomocy tzw. DIBR (Depth-Image-Based Rendering)[9] czyli syntezy widoku wirtualnego na podstawie obrazów i odpowiadających im map głębi[13][8].

W niniejszej pracy skoncentrowano się na podejściu DIBR – polegającym na przesyłaniu rzeczywistych widoków (np. z dwóch wybranych kamer), kompresowanych dowolną techniką stratną (np. AV1[2], HEVC[11], VVC[22]). Następnie, po stronie odbiornika, generowane są mapy głębi dla przesyłanych widoków oraz widok wirtualny z punktu wybranego przez użytkownika. Taki schemat (referencyjne widoki i głębia) zapewnia rozsądny kompromis między jakością a złożonością systemu[1].

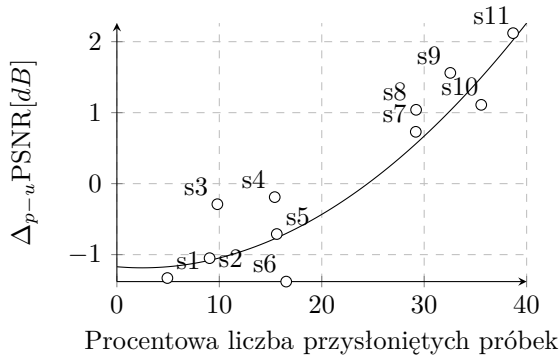
Do oceny jakości systemu wykorzystano obiektywne

miary jakości takie jak PSNR oraz IV-PSNR[6]. Wpływ opóźnień transmisji na postrzeganą jakość swobodnej nawigacji oceniano na podstawie testów subiektywnych.

3. OPTIMALIZACJA USTAWIENIA KAMER W SYSTEMACH SWOBODNEJ NAWIGACJI

W celu określenia optymalnego rozmieszczenia kamer w systemie FTV, w pracy przyjęto pewien model systemu wielokamerowego. Składa się on z czterech kamer. Kamery skrajne nie zmieniają swojej pozycji, natomiast kamery wewnętrzne mogą zostać ustawione tak, aby odległości między sąsiednimi kamerami w systemie były równe lub przesuwać się w kierunku kamer skrajnych, tworząc pary kamerowe. Model ten pozwala na teoretyczne wyprowadzenie wpływu zmiany położenia kamer na najważniejsze zjawiska wpływające na jakość widoku wirtualnego. Na jego podstawie wyznaczono liczbę przysłoniętych punktów obrazu oraz wyprowadzono dwa uproszczone modele teoretyczne opisujące wpływ wybranych parametrów systemu na jakość wirtualnego obrazu.

Pierwszy model uwzględnia zmianę odległości pomiędzy kamerami (odległości bazowej), analizując wpływ tej zmiany na teoretyczną dokładność mapy głębi i jakość wirtualnego widoku. Drugi skupiał się na analizie powierzchni obszarów przysłoniętych w scenie, których nie można poprawnie zsintezować. Oba modele pozwoliły na uzyskanie wyników teoretycznych, które są sumą omówionych wyżej modeli. Wynik, w funkcji procentowej liczby przysłoniętych próbek dla kamer ustawionych równomiernie, jako $\Delta_{p-u}PSNR$ czyli zmiana wartości metryki PSNR pomiędzy systemem z kamerami ustawionymi w pary (ang. pair) oraz równomiernie (ang. uniform) przedstawiono w postaci linii na rysunku 1.



Rysunek 1: Porównanie wyników eksperymentalnych i modelu teoretycznego

Dla weryfikacji modeli teoretycznych przeprowadzono eksperymenty na wielokamerowych sekwencjach testowych, analizując wpływ ustawienia kamer na jakość syntezy widoku. Na wykresie 1 wyniki eksperymentalne są przedstawione w postaci kropek, każda oznacza inną sekwencję testową, do której odniesienie można znaleźć w [10]. Eksperyment wykazał zgodność pomiędzy modelami teoretycznymi a jakością ocenianą przez PSNR.

Wyniki teoretyczne, jak i eksperymentalne pokazują, że stosowanie par kamer o małej odległości bazowej daje zysk, kiedy w scenie występuje wiele przysłonień. W przypadku małej liczby przysłonień lepiej stosować układy kamerowe z równomiernym ich rozłożeniem. Liczba

przysłoniętych punktów, dla kamer ustawionych równomiernie, dla których warto stosować pary kamer to około 25%. Jest to istotna wskazówka dla projektowania rzeczywistych systemów FTV[3][20][5].

4. WPŁYW KOMPRESJI NA JAKOŚĆ WIRTUALNEGO WIDOKU

W tym rozdziale przebadano, w jaki sposób jakość przesyłanych widoków rzeczywistych wpływa na jakość widoku syntetycznego generowanego techniką DIBR[9]. We wszystkich przypadkach przysyłano jedynie zakodowaną stratnie wizję, natomiast mapy głębi oraz widoki wirtualne były estymowane lokalnie po stronie odbiornika. W ramach tej części badań przeprowadzono trzy eksperymenty: wstępny, ogólny oraz wykorzystujący technikę MIV[14].

Eksperyment wstępny miał charakter eksploracyjny i został przeprowadzony z wykorzystaniem starszych technik kompresji, estymacji głębi oraz syntezy widoku wirtualnego[4]. Dla każdej sekwencji testowej zakodowano widoki przy użyciu trzech kodeków: MPEG-2[18], AVC (H.264)[23] i HEVC (H.265)[11]. Widok wirtualny generowano ze zdekodowanych widoków, i porównywano z widokami syntezywanymi z oryginalnych, niezakodowanych widoków. Wyniki pokazały, że kompresja w systemach swobodnego widzenia jest możliwa i nie powoduje dużej straty jakości widoku wirtualnego.

W eksperymencie ogólnym przeanalizowano wiele sekwencji testowych i cztery techniki kompresji: MPEG-2[18], AVC(H.264)[23], HEVC(H.265)[11] oraz VVC[22]. Dla różnych konfiguracji kodera mierzono PSNR widoków przesyłanych oraz IV-PSNR[6] widoku syntezywanego. Wyniki potwierdziły obserwację z eksperymentu wstępnego oraz pozwoliły na zebranie danych dla dalszych badań, z wykorzystaniem aktualnych metod estymacji map głębi i syntezy wirtualnego widoku.

W ostatnim eksperymencie wykorzystano technikę MPEG Immersive Video (MIV)[14][7]. W tym przypadku kodowano reprezentację sceny przy użyciu czterech kodeków: HEVC[11], VVC[22], AV1[2] i VP9[15]. Wykorzystano profil normy MIV GA (ang. Geometry Absent)[14], w którym mapy głębi nie są wysyłane a wyznaczane w dekodерze. Także w tym przypadku jakość widoku syntetycznego była silnie zależna od jakości zdekodowanych widoków.

Wyniki przedstawione w eksperymentach ogólnym oraz z użyciem MIV pozwoliły na dokładne przebadanie zależności pomiędzy jakością widoków przesyłanych i syntezywanym. W tym celu obliczono współczynnik korelacji Pearsona[25], który był bardzo wysoki dla prawie wszystkich przeprowadzonych eksperymentów. Następnie zaproponowano prosty model liniowy, który nie jest zależny od techniki kompresji. Pozwala on na oszacowanie jakości widoku syntezywanego na podstawie jakości przesyłanych widoków, dla danej sekwencji, zgodnie ze wzorem:

$$IV-PSNR_p \approx f_{AG}(PSNR_p) = a \cdot PSNR_p + b, \quad (1)$$

gdzie IV-PSNR_p oraz PSNR_p to zmierzone wartości metryk oceny jakości odpowiednio widoku wirtualnego oraz przesyłanego, a f_{AG} jest zaproponowanym liniowym modelem tej zależności[10].

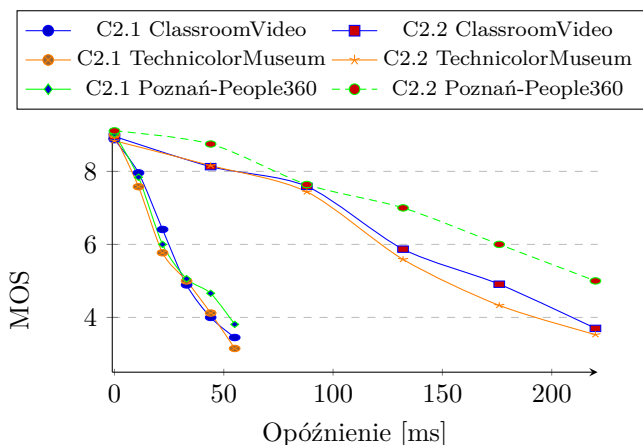
Podsumowując przeprowadzone badania, kluczowy wpływ na jakość widoku syntetycznego ma jakość przesyłanych widoków a metoda kompresji (np. VVC, AV1,

MPEG-2) ma drugorzędne znaczenie. Prosty model jakości może służyć jako skuteczne narzędzie predykcyjne, wspierające projektowanie i optymalizację systemów FTV.

5. ANALIZA WPŁYWU OPÓŹNIEŃ NA POSTRZEGANĄ JAKOŚĆ USŁUGI W WYŚWIETLACZACH NAGŁOWNYCH

W kolejnej części pracy przeanalizowano wpływ opóźnień na jakość swobodnej nawigacji użytkownika korzystającego z wyświetlacza nagłownego[17]. W ramach tych prac przeanalizowano różne scenariusze wymiany danych pomiędzy terminalem użytkownika i serwerem. Zidentyfikowano dwa rodzaje opóźnień, które mogą mieć różne wartości dla badanych scenariuszy. Są to opóźnienia wynikające ze zmiany pozycji użytkownika w scenie oraz zmiany kierunku patrzenia.

Następnie przeprowadzono eksperyment, który miał zbadać subiektywną jakość dla różnych scenariuszy transmisji, a co za tym idzie wartości opóźnień. Został on przeprowadzony za pomocą urządzenia Oculus Rift CV1. W trakcie eksperymentu uczestnicy oceniali w 11-punktowej skali MOS jakość swobodnej nawigacji w trzech sekwencjach testowych, dla dwóch scenariuszy transmisji oraz kilkunastu wariantów opóźnienia. Scenariusz C2.1 to taki, w którym opóźniona jest pozycja oraz kierunek patrzenia użytkownika o tę samą wartość. Wariant C2.2 to taki, w którym występuje jedynie opóźnienie pozycji użytkownika, a kierunek patrzenia syntezowany jest bez opóźnienia. Wyniki zostały przedstawione na rysunku 2.



Rysunek 2: Uśrednione wyniki testów subiektywnych dla wszystkich sekwencji i scenariusza C2.1 oraz C2.2

Pierwsza grupa krzywych to wyniki uzyskane dla przypadku z opóźnioną zarówno pozycją użytkownika w scenie, jak i kierunkiem patrzenia. Druga grupa krzywych obrazuje przypadek, gdzie jedynie opóźnienie wiąże się z pozycją użytkownika w scenie.

Wyniki wykazały, że opóźnienia rotacyjne (scenariusz C2.1) są znacznie bardziej odczuwalne i dokuczliwe dla użytkowników niż opóźnienie jedynie pozycji użytkownika w scenie (scenariusz 2.2). Wykorzystując scenariusz transmisji, który zapewnia niską wartość tego opóźnienia, wysoka jakość usługi swobodnej nawigacji może być zapewniona nawet przy czterokrotnie większym czasie komunikacji terminala użytkownika z serwerem.

6. EKSPERYMENTALNY SYSTEM SWOBODNEJ NAWIGACJI

Celem tej części badań było opracowanie i uruchomienie eksperymentalnego systemu FTV umożliwiającego swobodną nawigację widza w zarejestrowanej rzeczywistej scenie. System miał być możliwie prosty i tani w realizacji.

System składał się z dwóch głównych elementów, pierwszy to system wielokamerowy, drugi to serwer brzegowy, który dostarcza dane do terminala użytkownika, którym może być dowolne urządzenie z przeglądarką internetową.

System wielokamerowy został zbudowany z popularnych, konsumenckich kamer Go Pro Hero 4. Jest to kamera sportowa, która posiada złącze w tylnej części służące do podłączania akcesoriów, oraz umożliwiające złączenie dwóch kamer w stereopary[27]. Ta funkcjonalność została wykorzystana do budowy systemu wielokamerowego. Zaprojektowano dedykowany moduł „plecaków” kamer, czyli płytki elektronicznej dołączonej do kamery. Płytki ta odpowiadała za zarządzanie kamerą oraz odbiór informacji sterujących i synchronizacyjnych od zarządcy systemu. Z uwagi na duże odległości, zbudowano również moduły rozdzielające i regenerujące przesyłane sygnały. Zarządca systemu jest urządzeniem elektronicznym, które umożliwia sterowanie wszystkimi kamerami, a w szczególności zapewnia synchronizację rejestrowanych ramek.

Drugi element systemu to serwer swobodnej nawigacji. Korzysta on z przygotowanych widoków oraz map głębi. Użytkownicy systemu podłączają się do serwera przez stronę internetową. Zawiera ona kod skryptu JavaScript, który realizuje wyświetlanie nowych przesyłanych widoków wirtualnych, jak i umożliwia ich zmianę, czyli swobodną nawigację w scenie. Sam serwer jest wielowątkową implementacją zrealizowaną w języku C++. Serwer może być uruchomiony nawet na konsumenckim komputerze i obsługiwać jednocześnie kilku użytkowników.

Przeprowadzone badania pokazały, że budowa systemu swobodnego punktu widzenia jest możliwa, nawet z wykorzystaniem prostych konsumenckich komponentów. Serwer ten, mimo prostej wielowątkowej realizacji, może obsługiwać nawet kilkunastu użytkowników. Zarejestrowana scena musi jednak zostać odpowiednio przygotowana, w szczególności muszą zostać wyznaczone parametry systemu oraz mapy głębi.

7. PODSUMOWANIE

Przedstawione badania dotyczą systemów swobodnego punktu widzenia (FTV), ze szczególnym uwzględnieniem czynników wpływających na jakość odbioru usługi (QoE). Obejmują one zarówno prace teoretyczne, jak i praktyczne, prowadzone w czterech głównych obszarach: konfiguracji kamer, kompresji, wpływu opóźnień oraz implementacji systemu eksperymentalnego.

W pracy przeanalizowano wpływ rozmieszczenia kamer na jakość syntezowanego widoku. W szczególności wykazano, że układ kamer w postaci par o małej odległości bazowej jest korzystniejszy w scenach zawierających przysłonięcia. Poza tym zaproponowano model jakości widoku wirtualnego, który pozwala oszacować jakość widoku syntezowanego na podstawie jakości przesyłanych danych. Model ten, pomimo prostej liniowej definicji, wykazał dużą zgodność z wynikami eksperymentalnymi oraz co najważniejsze, nie zależy od zastosowanej techniki kompresji. Również eksperymentalnie zbadano wpływ transmisji

w systemach z wyświetlaczem nagłownym z uwzględnieniem opóźnień kierunku patrzenia oraz pozycji użytkownika. Wyniki subiektywne pokazały około czterokrotnie większą wrażliwość użytkowników na opóźnienia kierunku patrzenia – co pozwala na dobranie takiego scenariusza transmisji, który poprzez minimalizację tego czasu może zapewnić wyższą postrzeganą jakość. W ramach pracy również zaprojektowano i uruchomiono prosty kompletny system FTV. System umożliwia synchroniczne nagrywanie rzeczywistych wydarzeń oraz ich dostarczenie do terminali użytkowników końcowych, z usługą swobodnej nawigacji w scenie. Stanowi on w pełni funkcjonalną platformę badawczo-testową.

Pełne wyniki przedstawionych badań można znaleźć w pracy doktorskiej [10].

LITERATURA

- [1] Boyce J. M. i in. 2021. „MPEG Immersive Video Coding Standard”. *Proceedings of the IEEE* 109.9. s. 1521–1536.
- [2] Chen Y. i in. 2018. „An Overview of Core Coding Tools in the AV1 Video Codec”. *2018 Picture Coding Symposium (PCS)*. s. 41–45.
- [3] Domański M. i in. 2014. „Experiments on Acquisition and Processing of Video for Free-Viewpoint Television”. *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. s. 1–4.
- [4] Domański M. i in. 2016. „Optimization of Camera Positions for Free-Navigation Applications”. *2016 International Conference on Signals and Electronic Systems (ICSES)*. s. 118–123.
- [5] Domański M. 2015. „Widz wybiera punkt widzenia”. *Academia, Magazyn Polskiej Akademii Nauk*, No. 2 (42)., s. 48–49.
- [6] Dziembowski A. i in. 2022. „IVPSNR — The Objective Quality Metric for Immersive Video Applications”. *IEEE Transactions on Circuits and Systems for Video Technology* 32.11. s. 7575–7591.
- [7] Dziembowski A. i in. 2022. „Nowe techniki kompresji wizji dla rzeczywistości wirtualnej – MPEG Immersive Video”. *Przegląd telekomunikacyjny*. s. 116–123.
- [8] Dziembowski A. i in. 2018. „Szybka synteza widoków wirtualnych dla systemów telewizji swobodnego punktu widzenia”. *Przegląd Telekomunikacyjny*, Vol. 91, No. 6. s. 330–333.
- [9] Fehn C. 2003. „A 3D-TV Approach Using Depth-Image-Based Rendering DIBR”. in *Proceedings of 3rd IASTED Conference on Visualization, Imaging, and Image Processing*, Benalmadena. s. 482–487.
- [10] Grzelka A. 2024. „Systemy wizyjne swobodnego punktu widzenia o wysokiej postrzeganej jakości usług”. *Praca doktorska*. Politechnika Poznańska.
- [11] Hu Q. i in. 2014. „Analysis and optimization of x265 encoder”. *2014 IEEE Visual Communications and Image Processing Conference*. s. 502–505.
- [12] Merkle P. i in. 2007. „Multi-View Video Plus Depth Representation and Coding”. *2007 IEEE International Conference on Image Processing*, San Antonio, TX, USA, s. I - 201-I - 204.
- [13] Mieloch D. i in. 2017. „Estymacja głębi dla systemów wielowidokowych”. *Przegląd Telekomunikacyjny*, Vol. 86, No. 6. s. 479–482.
- [14] Mieloch D. i in. 2022. „Overview and Efficiency of Decoder-Side Depth Estimation in MPEG Immersive Video”. *IEEE Transactions on Circuits and Systems for Video Technology* 32.9. s. 6360–6374.
- [15] Mukherjee D. i in. 2013. „The latest open-source video codec VP9 - An overview and preliminary results”. *2013 Picture Coding Symposium (PCS)*. s. 390–393.
- [16] Muller K. i in. 2011. „3-D Video Representation Using Depth Maps”. *Proceedings of the IEEE* 99.4, s. 643–656.
- [17] Shibata T. 2002. „Head Mounted Display”. *Displays* 23.1, s. 57 –64. issn: 0141-9382. <http://www.sciencedirect.com/science/article/pii/S0141938202000100>
- [18] Sikora T. 1997. „MPEG digital video-coding standards”. *IEEE Signal Processing Magazine* 14.5. s. 82–100.
- [19] Smolic A. i in. 2005. „3D Video Objects for Interactive Applications”. *2005 13th European Signal Processing Conference*. s. 1–4.
- [20] Tanimoto M. 2010. „FTV (Free-viewpoint TV)”. *2010 IEEE International Conference on Image Processing*. s. 2393–2396.
- [21] Wei K. i in. 2013. „Point-Based Model Construction for Free-Viewpoint TV”. *2013 IEEE Third International Conference on Consumer Electronics Berlin (ICCE-Berlin)*. s. 220–221.
- [22] Wieckowski A. i in. 2021. „Vvenc: An Open And Optimized Vvc Encoder Implementation”. *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. s. 1–2.
- [23] Wiegand T. i in. 2003. „Overview of the H.264/AVC video coding standard”. *IEEE Transactions on Circuits and Systems for Video Technology* 13.7, s. 560–576.
- [24] Wu G. i in. 2023. „4D Gaussian Splatting for Real-Time Dynamic Scene Rendering”. *arXiv preprint arXiv:2310.08528*.
- [25] N.A. Weiss. 2012. *Introductory Statistics*. Pearson Education, isbn:9780321691224. url: https://books.google.pl/books?id=_r5ucgAACAAJ
- [26] C. Fehn i in. 2006. „Interactive 3-DTV-Concepts and Key Technologies”. W: *Proceedings of the IEEE* 94.3, s. 524–538.
- [27] Dual HERO System. Dostęp: 2020-05-06. https://gopro.com/content/dam/help/dual-hero-system/manuals/DualHeroSystem_UG_SPA_REVD_WEB.pdf.