

Abstract

Estymacja spojrzenia, będąca kluczowym aspektem interakcji człowiek–komputer (HCI), znajduje zastosowanie w takich obszarach jak rzeczywistość rozszerzona, rzeczywistość wirtualna, gry komputerowe, analiza zachowań, opieka zdrowotna oraz technologie wspomagające. Tradycyjnie do tego celu wykorzystywano eye trackery w kontrolowanych warunkach laboratoryjnych, co wymagało udziału wykwalifikowanych specjalistów oraz skomplikowanej konfiguracji. Jednak w ostatnich latach badania naukowe skupiły się na tworzeniu prostszych w użyciu systemów śledzenia spojrzenia, które nie wymagają dodatkowego sprzętu — poza standardową kamerą wbudowaną w komputer, tablet czy urządzenie mobilne — w celu upowszechnienia tej technologii wśród szerszego grona odbiorców. Takie podejście wiąże się jednak z dodatkowymi trudnościami, takimi jak niska rozdzielczość obrazu z kamery internetowej, zmienne warunki oświetleniowe w tle, różnice w wyglądzie poszczególnych użytkowników, zmiany pozycji głowy i wiele innych. Współczesne badania w tej dziedzinie koncentrują się na analizie tych wyzwań, aby poprawić dokładność estymacji spojrzenia w warunkach rzeczywistych. Niniejsza rozprawa doktorska podejmuje temat estymacji spojrzenia z wykorzystaniem standardowych kamer, dążąc do zwiększenia dostępności i przystępności cenowej systemów śledzenia wzroku. Głównym celem pracy jest zastosowanie standardowych kamer internetowych w celu udostępnienia technologii estymacji spojrzenia szerszemu gronu użytkowników.

Celem badań było rozwiązanie problemu śledzenia spojrzenia przy użyciu standardowej kamery internetowej. Na początku przeprowadzono szczegółową analizę istniejących badań w tej dziedzinie, dokładnie oceniając zalety i wady poszczególnych podejść. Szczególną uwagę poświęcono wyzwaniom związanym z wykorzystaniem standardowych kamer, takim jak niska rozdzielczość, zmienna pozycja głowy czy zmienne warunki oświetleniowe. W niniejszej pracy zastosowano metodę estymacji spojrzenia opartą na wyglądzie (appearance-based), która polega na bezpośrednim odwzorowaniu obrazu na przewidywany punkt spojrzenia (POG) lub kierunek patrzenia. W pracy zastosowano nowoczesne metody uczenia głębokiego. Przeprowadzono trening i testy z użyciem różnych architektur sieci neuronowych, w szczególności konwolucyjnych sieci neuronowych (CNN). Dokonano optymalizacji różnych hiperparametrów tych sieci w celu uzyskania najbardziej efektywnego modelu do przewidywania kierunku spojrzenia. Dodatkowo zastosowano techniki przetwarzania obrazu w celu poprawy wyników. Ponadto zbudowano model dostosow-

any do konkretnej osoby (person-specific), który wykazał, że odpowiednio dostrojona sieć może przewyższyć model trenowany na danych pochodzących od wielu osób. W dalszej części pracy zastosowano podejście transfer learning w środowisku z ograniczoną ilością danych i porównano je z modelem trenowanym bez wstępnie wyuczonych wag. Wyniki pokazują, że modele z transfer learningiem osiągają lepsze wyniki przy ograniczonych danych, a także szybciej się uczą i stabilizują w porównaniu z modelami bez pretrenowanych wag.