Systemy wizyjne w zastosowaniach przemysłowych

Marek Gorgoń, Maciej Aleksandrowicz, Krzysztof Błachut, Andrzej Brodzicki, Zbigniew Bubliński, Artur Cyba, Michał Daniłowicz, Adam Głowacz, Mirosław Jabłoński, Joanna Jaworek-Korjakowska, Aleksander Kostuch, Marcin Kowalczyk, Tomasz Kryjak, Dariusz Kucharski, Konrad Lis, Michał Machura, Zbigniew Mikrut, Piotr Pawlik, Michał Piekarski, Dominika Przewłocka-Rus, Joanna Stanisz, Hubert Szolc, Mateusz Wąsala, Anna Wójcicka, Piotr Wzorek

AGH Akademia Górniczo-Hutnicza, Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej, Kraków

Streszczenie: Systemy wizyjne to dynamicznie rozwijająca się dziedzina nauki i techniki. W niniejszym artykule przedstawiono przegląd zagadnień, nad którymi pracuje zespół Laboratorium Systemów Wizyjnych. Obejmuje on systemy wizyjne dla pojazdów autonomicznych, aplikacje analizy obrazów medycznych, systemy wspomagania szkolenia pracowników, analizę obrazów termowizyjnych, implementacje i akceleracje algorytmów wizyjnych w układach SoC FPGA i GPU, przetwarzanie strumienia wizyjnego w rozdzielczości 4K, a także analizę sygnałów z wykorzystaniem głębokich sieci neuronowych i współpracę przy eksperymencie ALICE w CERN-ie.

Słowa kluczowe: systemy wizyjne, SoC FPGA, GPU, głębokie sieci neuronowe, termowizja

COMPUTER VISION SYSTEMS IN INDUSTRIAL APPLICATIONS

Abstract: Computer vision systems are a dynamically developing field of science and technology. This article presents an overview of issues on which the Computer Vision Laboratory team is working. It covers vision systems for autonomous vehicles, medical image analysis applications, employee training support systems, analysis of thermal images, implementation and acceleration of vision algorithms in SoC FPGAs and GPUs, processing of 4K resolution video streams, as well as signal analysis using deep neural networks and a cooperation with the ALICE experiment at CERN.

Keywords: computer vision systems, SoC FPGA, GPU, deep neural networks, thermal imaging

https://doi.org/10.7494/978-83-67427-00-5_1

1. Wstęp

System wizyjny składa się z oprogramowania (zestawu algorytmów), które analizują obrazy lub sekwencje wideo pozyskane z różnego rodzaju kamer (wizyjnych, podczerwonych, termowizyjnych, hiperspektralnych czy zdarzeniowych) oraz sprzętu, na którym realizowane są obliczenia. W Laboratorium Systemów Wizyjnych, które działa w ramach Katedry Automatyki i Robotyki na Wydziale Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej (EAIiIB) Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, prowadzone są prace badawcze nad oboma wymienionymi komponentami – rozwijane są różnego rodzaju algorytmy, w tym także bazujące na głębokich sieciach neuronowych, które są następnie implementowane i akcelerowane z wykorzystaniem kart graficznych (GPU – *graphic processing unit*) lub układów SoC FPGA (*system on chip field programmable gate array*). Od strony aplikacyjnej prace wpisują się w szeroko rozumianą robotykę, w której istotną rolę pełni tzw. wizyjne sprzężenie zwrotne. Rozwijane są zaawansowane systemy monitoringu wizyjnego (AVSSs – *advanced video surveillance systems*), zaawansowane systemy wspomagania kierowcy (ADASs – *advanced driver assistance systems*), systemy percepcji pojazdów autonomicznych (samochodów, dronów), systemy automatycznej analizy obrazów medycznych, systemy wspomagające wykonywanie prac konserwacyjnych czy systemy diagnostyki silników elektrycznych z użyciem termowizji. Istotnym wątkiem są również prace nad implementacją algorytmów wizyjnych w systemach wbudowanych. Ponadto członkowie zespołu Laboratorium Systemów Wizyjnych zaangażowani są w projekt detekcji anomalii podczas działania synchrotronu czy też projektowanie i rozwijanie konfiguracji układów FPGA, które są wykorzystywane m.in. w układach elektronicznych tworzących podsystem tzw. triggera dla detektora FV0 eksperymentu ALICE w CERN-ie. W kolejnych rozdziałach wymienione wątki badawcze zostały skrótowo opisane i scharakteryzowane.

2. Wbudowane systemy wizyjne czasu rzeczywistego

W wielu aplikacjach systemów wizyjnych w robotyce kluczowe jest jak najszybsze opracowanie sposobu sterowania na podstawie percepcji i analizy otoczenia, w którym działa robot. Dobrym tego przykładem są systemy wspomagania kierowcy, np. automatyczne hamowanie awaryjne, które ma zapobiec kolizji z pieszym. W takim ujęciu praca w reżimie czasu rzeczywistego jest rozumiana jako przetwarzanie strumienia wideo (bez pomijania ramek), w pełnej rozdzielczości, z szybkością akwizycji (np. 60 ramek na sekundę) i przy jak najmniejszej latencji. Dodatkowo w części aplikacji dostępny budżet energetyczny jest ograniczony, a obliczenia muszą być prowadzone w urządze-

niu (*embedded, edge*) – przykładowo kamerze inteligentnej (*smart camera*). Spełnienie wymienionych wyżej wymagań nie jest zadaniem łatwym i najczęściej nie jest możliwe przy użyciu jednostek obliczeniowych ogólnego przeznaczenia (GPPs – *general purpose processors*). Stąd jednym z aktywnie rozwijanych w Laboratorium wątków badawczych są tzw. wbudowane systemy wizyjne czasu rzeczywistego, implementowane głównie w układach SoC FPGA, ale też częściowo eGPU (*embedded GPU*). Ponadto nowym wątkiem badawczym są obliczenia neuromorficzne, tj. z wykorzystaniem sieci impulsowych (SNNs – *spiking neural networks*) oraz specjalizowanych platform asynchronicznych.

2.1. Przetwarzanie obrazu w rozdzielczości 4K

W ostatnich latach obserwuje się dynamiczny rozwój rozdzielczości urządzeń do akwizycji (kamer) oraz wyświetlania (monitorów, telewizorów). Obecnie typowa rozdzielczość to 4K, tj. 3840 × 2160 pikseli (zwana też UHD – *ultra high definition*), dostępne są już urządzenia 8K, ale także 10K i 16K. Przetwarzanie wideo 4K, szczególnie w czasie rzeczywistym, stanowi duże wyzwanie dla projektantów wbudowanych systemów wizyjnych. Dla typowej częstotliwości 60 ramek na sekundę strumień danych to 1424 MB/s przy tzw. zegarze piksela około 600 MHz, co jest wartością za wysoką dla obecnie dostępnych układów SoC FPGA. Stąd konieczność stosowania formatów wektorowych, tj. 2, 4 lub 8 ppc (*pixel per clock cycle*), co w przypadku implementacji niektórych algorytmów stanowi spore wyzwanie.

W Laboratorium w latach 2016–2022 prowadzony był projekt finansowany przez Narodowe Centrum Nauki (konkurs „Sonata") pt. *Opracowanie sposobu organizacji obliczeń w heterogenicznych układach rekonfigurowalnych najnowszej generacji umożliwiającego przetwarzanie strumienia wizyjnego UHD/4K w czasie rzeczywistym*, w ramach którego zaimplementowano szereg algorytmów wizyjnych, udowadniając, że układy SoC FPGA są bardzo dobrą platformą do prototypowania wizyjnych systemów czasu rzeczywistego.

W pierwszym etapie projektu skupiono się na algorytmach podstawowych – operacjach punktowych i kontekstowych. Implementacja metod z pierwszej grupy jest prosta, ale wymaga zwielokrotnienia zastosowanych modułów obliczeniowych. Przykładowo konwersja z przestrzeni RGB do odcieni szarości dla strumienia 1 ppc wymaga jednego modułu, a dla 4 ppc – czterech identycznych modułów działających równolegle. W przypadku operacji kontekstowych, takich jak np. filtracja Gaussa lub Sobela, realizacja jest również względnie prosta, choć wymagana jest uważna kontrola synchronizacji obliczeń (dla 4 ppc równoległa analiza czterech kontekstów). Zagadnienie to zostało opisane w publikacji Kowalczyka i in. (2018).

Bazując na powyższych podstawowych modułach, wykonano implementację szeregu bardziej zaawansowanych algorytmów, takich jak: poprawa kontrastu metodą CLAHE (*contrast limited adaptive histogram equalization*), wyliczanie przepływu optycznego metodami Horna–Schuncka oraz Lucasa–Kanade (również w wersji wieloskalowej), detekcji obiektów z wykorzystaniem podejścia HOG+SVM (*histogram of oriented gradients, support vector machine*), detekcji i deskrypcji punktów charakterystycznych metodą ORB (*oriented FAST and rotated BRIEF*), segmentacji obiektów pierwszoplanowych, wyznaczania map głębi metodą *semi-global matching* czy indeksacji jednoprzebiegowej. Szczególnym wyzwaniem okazały się ostatnie dwa z wymienionych algorytmów, gdyż zachodzi w nich bezpośrednia zależność pomiędzy danymi w wektorze (2 lub 4 ppc) – przykładowo wyliczenie wartości dla piksela 4 wymaga wyznaczenia wszystkich poprzednich (dla 1, 2 i 3), i to wszystko w jednym takcie zegara, aby zachować paradygmat obliczeń w czasie rzeczywistym. Szczegóły dotyczące algorytmu indeksacji zostały przedstawione w artykule Kowalczyka i in. (2021).

Ostatnim wątkiem były prace nad akceleracją głębokich konwolucyjnych sieci neuronowych. Z uwagi na bardzo dużą złożoność obliczeniową tego typu modeli przy wykorzystaniu obecnie dostępnej technologii nie jest możliwe przetwarzanie obrazu w czasie rzeczywistym – nie tylko w układach SoC FPGA, ale też w wysokowydajnych GPU. W ramach projektu podjęto jednak dwa wątki: detekcję i klasyfikację znaków drogowych z wykorzystaniem sieci binarnych (XNOR) (Przewłocka-Rus i in. 2021) oraz śledzenie obiektów z wykorzystaniem sieci syjamskiej.

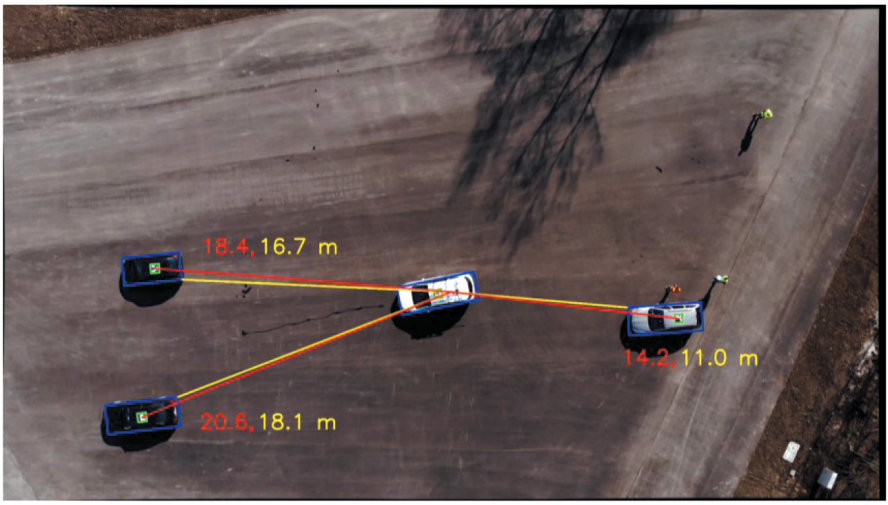
Podsumowując uzyskane wyniki, należy uznać, że przetwarzanie strumienia wideo w rozdzielczości 4K w układach SoC FPGA jest w pełni możliwe, choć niekiedy parametry systemu ograniczone są przez dostępne zasoby logiczne (np. przetwarzanie w wielu skalach) lub transfer do zewnętrznej pamięci RAM (np. segmentacja obiektów pierwszoplanowych). Szczególne wyzwanie stanowią algorytmy, w których występują zależności pomiędzy pikselami w pojedynczym wektorze. Z uwagi na dużą złożoność obliczeniową implementacja sieci głębokich dla strumienia 4K (przykładowo segmentacji semantycznej) nie wydaje się obecnie możliwa.

2.2. Systemy percepcji dronów i pojazdów autonomicznych

Kolejnym wątkiem rozwijanym przez zespół Laboratorium Systemów Wizyjnych są szeroko rozumiane systemy percepcji pojazdów autonomicznych. Rozumie się przez to pozyskiwanie informacji o otoczeniu – przykładowo detekcję obiektów (przeszkód) czy wyznaczenie przestrzeni, po której może poruszać się samochód (dron). W przypadku samochodów prace koncentrują się obecnie na dwóch wątkach – wykorzystaniu czujnika LiDAR (*light detection and ranging*) oraz nowoczesnej kamery zdarzeniowej (*event camera*, *dynamic vision sensor*). W pierwszym przypadku celem jest przetwarza-

nie w układzie SoC FPGA chmury punktów z czujnika LiDAR w czasie rzeczywistym i pozyskanie informacji o obiektach (innych samochodach, pieszych, rowerzystach). Wykorzystuje się sieci głębokie, gdyż tylko one cechują się odpowiednio wysoką skutecznością. Niestety z uwagi na specyficzny format danych i związaną z tym nietypową architekturę sieci, uzyskanie działania w czasie rzeczywistym stanowi istotne wyzwanie (Stanisz i in. 2021). Równolegle prowadzone są prace nad określeniem potencjału kamer zdarzeniowych do detekcji różnych obiektów związanych z ruchem drogowym – obecnie pionowych znaków drogowych, ale też pieszych i pojazdów.

W przypadku dronów badania prowadzone są w dwóch obszarach. Pierwszym jest ich wykorzystanie do szeroko rozumianej inspekcji. Przykładem może być użycie informacji wizyjnej z kamery drona do oceny działania systemów percepcji pojazdów autonomicznych (Błachut i in. 2022). W pracy przeanalizowano wykorzystanie marketów ArUCo oraz sieci typu Mask-RCNN (rys. 1). Uzyskano dokładność pomiaru około 0,5 m z wysokości do 80 m, co uznano za wynik satysfakcjonujący.



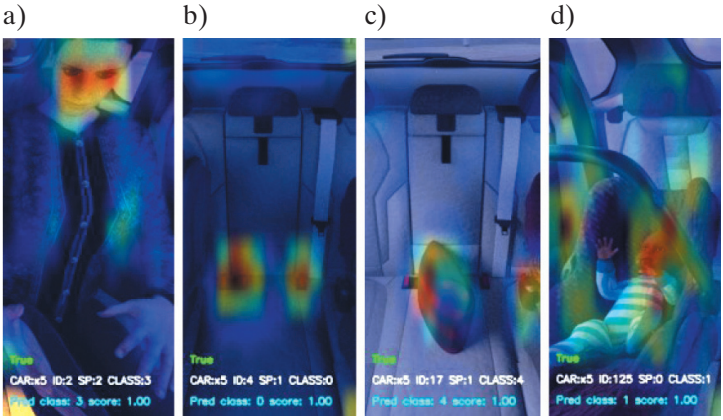
Rys. 1. Przykład działania wizyjnego algorytmu do oceny systemów percepcji pojazdów autonomicznych. Na niebiesko zaznaczono prostokąty otaczające pojazdy, a na zielono znaczniki ArUco. Czerwona linia wizualizuje odległość między znacznikami, żółta minimalny dystans do prostokątów otaczających samochody

Drugim obszarem badawczym są drony autonomiczne, w szczególności wyścigowe (przelot przez serię bramek), oraz eksplorujące nieznane obszary (problem jednoczesnej lokalizacji i mapowania, SLAM – *simultaneous localization and mapping*). Dodatkowo rozważane są różne metody sterowania – klasyczne i wykorzystujące uczenie przez wzmacnianie (*reinforcement learning*). W pracach stosowane są typowe kamery wizyjne oraz kamery zdarzeniowe.

3. Głębokie sieci neuronowe w problemach widzenia maszynowego

W ostatnich latach kluczową rolę w rozwiązywaniu problemów widzenia maszynowego zaczęły odgrywać architektury bazujące na głębokich sieciach neuronowych. Architektury oparte na warstwach konwolucyjnych są wykorzystywane do przetwarzania obrazów w problemach klasyfikacji (CNN), detekcji (RCNN, YOLO, SSD), segmentacji (UNet), korekcji i augmentacji czy generowania nowych obrazów (VAE, GAN). W ostatnich latach coraz większą popularnością cieszą się sieci typu *transformer*, które, w przeciwieństwie do sieci konwolucyjnych, wykorzystują mechanizm uwagi. Poniżej przedstawiono zastosowanie głębokiego uczenia w takich obszarach, jak pojazdy autonomiczne, analiza zdjęć medycznych czy detekcja anomalii w dużych infrastrukturach badawczych.

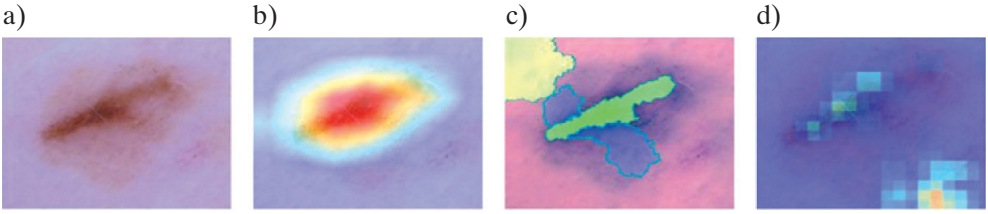
Klasyfikacja zajętości siedzeń we wnętrzu pojazdu to obiecująca funkcjonalność samochodów nowej generacji, która wciąż pozostaje poważnym wyzwaniem. Informacja ta może być wykorzystana do przypominania kierowcy o pasażerach na tylnym siedzeniu. Ze względu na wysoką śmiertelność dzieci pozostawianych w pojazdach wykrywanie obecności dziecka to jedna z funkcji na mapie drogowej standardu Euro NCAP. Ponieważ większość wypadków wiąże się z błędami kierowcy, konsekwencje niezapinania lub niewłaściwego zapinania pasów bezpieczeństwa są oczywiste. NHTSA donosi, że 47% z 22 215 pasażerów samochodów osobowych, którzy zginęli w wypadkach w 2019 roku, nie miało zapiętych pasów bezpieczeństwa. Aby rozwiązać powyższe problemy, w pracy Jaworek-Korjakowskiej i in. (2021b) zaproponowano podejście bazujące na głębokich sieciach neuronowych, które klasyfikują zajętość miejsc w pojeździe w czterech najważniejszych kategoriach: niemowląt, dzieci i dorosłych oraz obiektów codziennego użytku (rys. 2).



Rys. 2. Wejściowy obraz z wnętrza pojazdu z informacjami w postaci mapy cieplnej dla czterech klas zajętości siedzenia: a) dorosły; b) pusty fotel; c) obiekt codziennego użytku; d) niemowlę w foteliku

W badaniach przedstawiono interpretowalne podejście AI, które wykorzystuje wstępnie wytrenowane sieci, w tym ResNet152V2, DenseNet121 i EfficientNetB0-B5-B7, do obliczania wektorów cech, a następnie korzysta z dostosowanego do konkretnego problemu klasyfikatora. Model ten zapewnia interpretację wyników w trybie nienadzorowanym. W pracy uzyskano skuteczność 79% i wykazano również, że obliczone metryki statystyczne umożliwiają pracownie interpretowalnego modelu, który koreluje z dokładnością architektury.

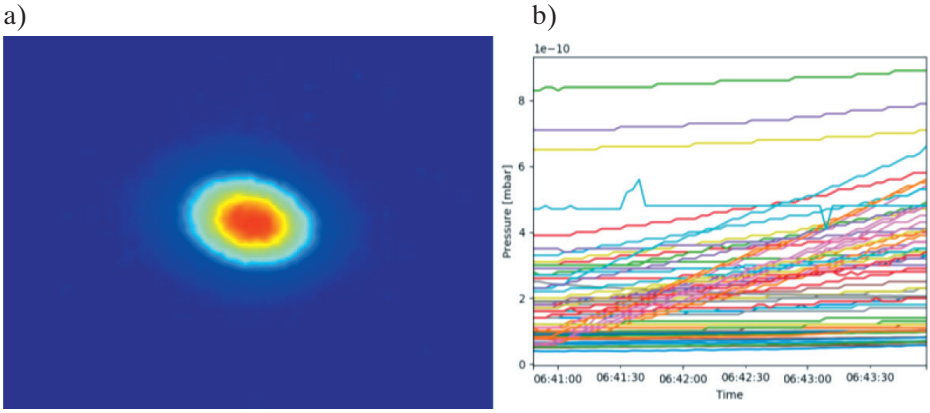
Głębokie sieci neuronowe znalazły również szerokie zastosowanie w komputerowym wspomaganiu diagnostyki medycznej. W ramach prowadzonych badań wykorzystuje się architektury głębokie do rozpoznawania czerniaka złośliwego oraz raka płuc czy wątroby. W pracach Jaworek-Korjakowskiej i in. (2021a) oraz Cassidy'ego i in. (2022) zaobserwowano, że specyficzne kryteria czerniaka *in situ* i wczesnego inwazyjnego czerniaka w dużym stopniu zależą od anatomicznej lokalizacji na ciele oraz od dostępnego zbioru danych służącego do treningu sieci. Zaproponowano strukturę bazującą na architekturze głębokiego uczenia, aby klasyfikować zmiany skórne w trzech najważniejszych miejscach na ciele – na twarzy, tułowiu i kończynach oraz w okolicy dłoni i stóp. W badaniu wykorzystano wstępnie wytrenowane sieci, m.in. ResNet50, Xception, DenseNet121 i EfficientNetB0. Eksperymenty potwierdziły zdolność opracowanych algorytmów do rozpoznawania najważniejszych miejsc anatomicznych z ogólną dokładnością 91,45% w przypadku architektury EfficientNetB0. Wykonano także wizualną analizę interpretowalności rozwiązania (rys. 3) przy użyciu map ciepła, algorytmu LIME czy map okluzji oraz zaproponowano metrykę bazującą na pokrywaniu map ciepła z maskami segmentacji.



Rys. 3. Zdjęcia dermoskopowe analizowane przy użyciu algorytmów do oceny interpretowalności: a) obraz wejściowy; b) GradCAM; c) LIME; d) mapa okluzji

Najistotniejszym celem wszystkich instytutów badawczych, w tym Narodowego Centrum Promieniowania Synchrotronowego SOLARIS, jest zapewnienie naukowcom prowadzącym badania w ośrodku dostępu do niezawodnej infrastruktury. Ze względu na złożoność systemu sterowania synchrotronem często trudno jest określić przyczynę niestabilności wiązki lub – w skrajnych przypadkach – jej utraty. Celem prowadzonych badań jest wykorzystanie architektury głębokiego uczenia do badania

stabilności wiązki na podstawie analizy różnych sygnałów diagnostycznych, takich jak wartości ciśnienia z podsystemu ultrawysokiej próżni czy obrazy przekroju wiązki elektronów z linii diagnostycznej Pinhole – przykłady zostały przedstawione na rysunku 4. Badania członków zespołu opisane w publikacji Piekarskiego i in. (2020) zawierają bardzo obiecujące wyniki dotyczące wykorzystania metod głębokiego uczenia do rozwiązania problemu detekcji anomalii. Proponowane rozwiązanie wykrywa anomalie w sygnałach diagnostycznych z 92-procentową dokładnością i 85,5-procentową precyzją. W badaniach wykorzystywane są zaawansowane rekurencyjne sieci neuronowe, autoenkodery i architektury wielowejściowe, w szczególności sieci z wieloma widokami, które są zdolne do wykonywania wielu zadań eksploracji sygnałów diagnostycznych. Pozwala to na analizę informacji zawartych zarówno w sygnałach 2D (obrazach), jak i skalarnych sygnałach diagnostycznych (jako dwa wejścia do proponowanej architektury).



Rys. 4. Dane diagnostyczne analizowane w zaproponowanym systemie głębokiego uczenia:
a) przekrój wiązki z linii diagnostycznej Pinhole;
b) sygnały z podsystemu ultrawysokiej próżni

4. Wspomaganie procesu montażu zaworu systemem wizyjnym[1]

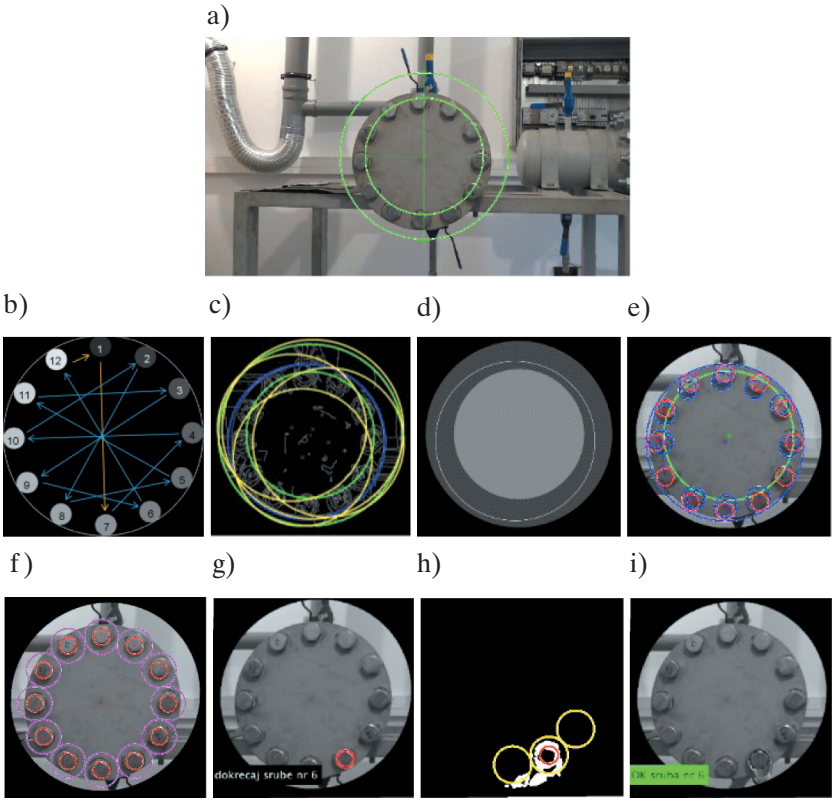
W przedsiębiorstwach remontowych coraz większym problemem staje się rotacja pracowników. Nowo zatrudnieni muszą być w miarę szybko, bezpiecznie i efektywnie przeszkoleni. W dobrych firmach standardem staje się szkolenie praktyczne, przeprowadzane na specjalnie w tym celu konstruowanych stanowiskach.

Stanowisko dydaktyczne, wykorzystywane przez jedną z firm współpracujących z AGH (rys. 5a), służy do przeszkolenia pracownika pod kątem poprawnej konserwacji

[1] Praca wykonana w ramach Programu Operacyjnego Inteligentny Rozwój (NCBiR) POIR.01.01.01-00-0170/17.

połączenia kołnierzowego zaworu. Prezentowany system wizyjny wspomaga ostatni etap konserwacji, czyli proces skręcania zaworu, podczas którego istotne jest zachowanie odpowiedniej kolejności dokręcania śrub (rys. 5b) zadanym momentem siły.

W tym celu niezbędne jest wygenerowanie modelu 2D pokrywy zaworu i skorelowanie go z obrazem uzyskanym z kamery stacjonarnej. W algorytmie wykorzystano transformatę Hougha (Davies 2005), która wykrywała okręgi duże (rys. 5c) oraz małe (rys. 5e). Dopasowanie modelu polega na wyborze okręgu odpowiadającego krawędzi pokrywy, zawartego pomiędzy okręgami kalibracyjnymi (rys. 5d), a następnie na lokalizacji obszarów obrazu zawierających śruby i przyporządkowaniu ich odpowiednio ponumerowanym obszarom (okręgom) modelu.



Rys. 5. Stanowisko badawczo-dydaktyczne (a), przykładowa kolejność dokręcania śrub (b), kolejne kroki algorytmu (c–i) (opis w tekście)

Na rysunku 5e czerwone okręgi to model, który ma zostać dopasowany do reprezentacji śrub (małe niebieskie okręgi). Czerwone linie obrazują pary punktów, na podstawie których obliczono transformatę afiniczną. Niebieskie linie symbolizują wyeliminowane pary punktów. Powodem eliminacji było wielokrotne odwoływanie się

punktów modelu (środków czerwonych okręgów) do reprezentacji śrub (środków niebieskich okręgów) wykrytych transformatą Hougha. Rysunek 5f przedstawia wynik dopasowania po transformacji afinicznej: obraz rzeczywisty został nieco powiększony i przekręcony. Oprócz czerwonych okręgów oznaczających model pokazano na nim „obszary zainteresowania" otaczające poszczególne śruby (kolor różowy), w których prowadzona będzie detekcja klucza.

Na obrazie po transformacji afinicznej (rys. 5g) pokazano wygenerowany komunikat i zaznaczono kolejną śrubę kolorem czerwonym. Pracownik powinien rozpocząć dokręcanie wskazanej śruby. Na rysunku 5h pokazano wynik detekcji klucza. Analizę prowadzono dla śruby nr 6 oraz dwóch śrub sąsiednich. Z dwóch przetestowanych metod wykrywania klucza – odejmowania od tła oraz pola ruchu (OF – *optical flow*) – do detekcji wybrano tę ostatnią.

Zbinaryzowane pole ruchu oznaczono kolorem białym (rys. 5h). Warunkiem wykrycia klucza jest znaczna przewaga sumy wykrytych białych pikseli w obszarze odpowiadającym wskazanej śrubie nad sumami wykrytymi w obszarach sąsiednich. Po spełnieniu tego warunku pracownik jest informowany o prawidłowości wyboru śruby (rys. 5i).

Opisany wyżej algorytm został przetestowany – z wynikiem pozytywnym – na kilkunastu filmach, na których dokręcano śruby rozmieszczone w różnych miejscach pokrywy zaworu.

5. Realizacja algorytmów wizyjnych z wykorzystaniem kart graficznych GPU

Podstawowym, choć nie jedynym, sposobem komunikacji człowieka z komputerem jest monitor. Prezentacja informacji wizyjnej wymaga stosowania specjalizowanych komponentów sprzętowych określanych jako GPU, które – uzupełnione o biblioteki programowe (np. OpenGL) – pozwalają generować realistyczną animację obiektów trójwymiarowych na płaskim ekranie, również z zastosowaniem emulacji widoku 3D przez obrazy stereo. Rozwój w tej dziedzinie był powodowany głównie zapotrzebowaniem rynku rozrywkowego oraz branży CAD/CAM (*computer aided design*/*computer aided manufacturing*). Wysoka elastyczność architektur GPU sprawiła jednak, że około dwóch dekad temu stały się one atrakcyjnym narzędziem do obliczeń wykraczających poza generowanie syntetycznych obrazów. Początkowo struktura potoku OpenGL odwzorowana w architekturze sprzętowej GPU pozwalała jedynie na odwrócenie kierunku przepływu danych, co umożliwiało przetwarzanie obrazów rastrowych pozyskanych z kamer wizyjnych. W ten sposób udawało się realizować większość algorytmów przetwarzania i analizy obrazów bez obciążania jednostki CPU.

Zespół Laboratorium Systemów Wizyjnych prowadził szereg prac mających na celu realizację typowych algorytmów wizyjnych: bezkontekstowych, kontekstowych, a także tych bardziej złożonych, takich jak na przykład segmentacja obiektów pierwszoplanowych metodą GMM (*Gaussian mixture model*), etykietowanie obiektów, operacje wykorzystujące histogram czy analiza obrazów w wielu skalach (Jabłoński 2012, 2013, Jabłoński i Przybyło 2016). Przemysłowe, komercyjne stacje wizyjne (*machine vision*) niejednokrotnie zawierają w sobie takie specjalizowane biblioteki dla kart graficznych.

Wysoka użyteczność obliczeń w GPU sprawiła, że od ponad dekady powszechnie dostępne są środowiska rozwojowe określane jako GPGPU (*general-purpose computing on graphics processing units*) OpenCL (*open computing language*) oraz CUDA (*compute unified device architecture*). Ich cechą wspólną jest możliwość realizacji równoległych obliczeń z wykorzystaniem wydajnych jednostek SIMD (*single instruction multiple data*) dostępnych we współczesnych układach GPU, a obszar ich zastosowań jest w zasadzie nieograniczony, bo obejmujący centra obliczeniowe ośrodków badawczych (np. CERN, Cyfronet) (Jabłoński i Przybyło 2016, Rohr i Lindenstruth 2017), a także platformy wbudowane i urządzenia mobilne.

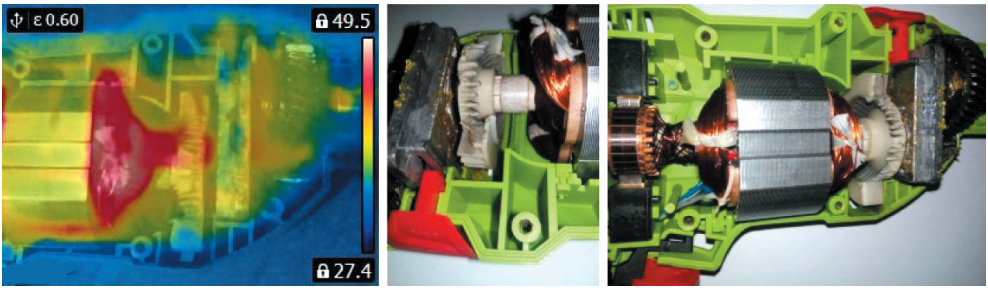
W trakcie prowadzonych prac powstało autorskie środowisko pozwalające na graficzne projektowanie złożonych algorytmów wizyjnych wykorzystujących biblioteki operacji przetwarzania i analizy obrazów (Świerk 2013). Służy ono do przetwarzania w czasie rzeczywistym obrazu z przemysłowych kamer wysokiej rozdzielczości. Poszczególne etapy przetwarzania mogą być realizowane w sposób równoległy na GPU lub CPU za pomocą biblioteki OpenCL albo też sekwencyjnie z zastosowaniem funkcji bibliotecznych OpenCV.

Duża moc obliczeniowa układów GPU stała się przed dekadą katalizatorem rozwoju bardzo popularnych dzisiaj technik sztucznej inteligencji i uczenia głębokiego bazujących na konwolucyjnych sieciach neuronowych CNN (*convolutional neural networks*). Dzięki nim uczenie i wykorzystanie sztucznych sieci neuronowych zawierających niekiedy setki milionów wag jest obecnie możliwe również na powszechnie dostępnym sprzęcie (Przybyło i Jabłoński 2019).

6. Diagnostyka silników elektrycznych z zastosowaniem termowizji

Diagnostyka silników elektrycznych z zastosowaniem termowizji jest kolejnym z tematów podejmowanych w Laboratorium i została przedstawiona w pracach Głowacza (2021a, 2021b, 2021c). Pierwsza z wymienionych prac opisuje diagnostykę wiertarek elektrycznych z zastosowaniem analizy obrazów termowizyjnych. Do tego celu została zaproponowana oryginalna metoda ekstrakcji cech z obrazów termowizyjnych BCAoID (*binarized common areas of image differences*). Przeanalizowano trzy różne

stany wiertarek elektrycznych: sprawna wiertarka, wiertarka z dziesięcioma uszkodzonymi łopatkami wentylatora (rys. 6) oraz wiertarka z uszkodzoną przekładnią. Cechy obrazów, obliczone z zastosowaniem BCAoID, zostały przeanalizowane przy użyciu klasyfikatora najbliższego sąsiada i sieci neuronowej. Skuteczność rozpoznawania obrazów mieściła się w zakresie 97,91–100%.



Rys. 6. Wiertarka elektryczna z uszkodzonym wentylatorem

W kolejnym artykule wspomnianego autora przedstawiono metodę analizy stanu pracy szlifierek kątowych z zastosowaniem obrazów termowizyjnych (Głowacz 2021b). Zaproponowano do tego celu innowacyjny algorytm BCAoMID-F (*binarized common areas of maximum image differences-fusion*). Przeanalizowano trzy różne stany szlifierki kątowej: sprawna szlifierka kątowa, szlifierka kątowa z jednym zablokowanym wlotem powietrza oraz szlifierka kątowa z dwoma zablokowanymi wlotami powietrza. Klasyfikację wektorów cech przeprowadzono z zastosowaniem klasyfikatora najbliższego sąsiada i SVM. Całkowita skuteczność rozpoznawania dla trzech klas mieściła się w przedziale 98,5–100%.

W ostatnim z wymienionych artykułów opisano termograficzną diagnostykę uszkodzeń wentylacji silników BLDC (Głowacz 2021c). Przeanalizowano ich następujące stany:

- sprawny silnik BLDC pracujący z prędkością 1450 obr/min,
- sprawny silnik BLDC pracujący z prędkością 2100 obr/min,
- silnik BLDC z zablokowaną wentylacją pracujący z prędkością 1450 obr/min,
- silnik BLDC z zablokowaną wentylacją pracujący z prędkością 2100 obr/min,
- sprawna maszynka do strzyżenia,
- maszynka z zablokowaną wentylacją.

Do ekstrakcji cech obrazów termowizyjnych zaproponowano metodę CPoAMoTI (*common part of arithmetic mean of thermographic images*). Testowe obrazy termowizyjne zostały pomyślnie rozpoznane ze skutecznością wynoszącą 100%.

Należy podkreślić, że wszystkie zaproponowane metody dostarczają wielu korzyści dla przemysłu i społeczeństwa. Są z powodzeniem wykorzystywane zarówno do samochodów elektrycznych oraz pociągów, jak i wentylatorów, komputerów, elektronarzędzi, a także różnych rodzajów silników elektrycznych.

7. Podsumowanie

Prace podejmowane w Laboratorium Systemów Wizyjnych przy Katedrze Automatyki i Robotyki na Wydziale EAIiIB Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie obejmują bardzo szeroki zakres zastosowania systemów wizyjnych w przemyśle. Dotyczą one bowiem m.in. problemów nadzoru instalacji technicznych (wykrywanie uszkodzeń przyrządów, detekcja anomalii wiązki generowanej w synchrotronie) czy też wsparcia procesów montażowych (np. zaworów). Działalność naukowa obejmuje również takie branże jak diagnostyka medyczna (np. czerniaka złośliwego) czy automatyka pojazdowa (np. sterowanie pojazdami autonomicznymi, wykrywanie zajętości tylnych siedzeń). Dzięki wykorzystywaniu najnowszych rozwiązań algorytmicznych (np. głębokich konwolucyjnych sieci neuronowych) oraz wydajnych platform obliczeniowych (takich jak np. procesory graficzne GPU, układy SoC FPGA) prowadzone prace zyskują uznanie na arenie krajowej oraz międzynarodowej, o czym świadczą liczne publikacje na renomowanych konferencjach i w czasopismach naukowych.

Zespół aktywnie podejmuje też nowe zagadnienia naukowe i prowadzi badania nad zastosowaniem kamer zdarzeniowych w systemach percepcji pojazdów autonomicznych, platformami neuromorficznymi, sieciami kwantowymi, uczeniem przez wzmacnianie w zastosowaniach robotycznych czy najnowszymi architekturami głębokich sieci neuronowych. Ponadto współpracuje z przemysłem w różnych projektach związanych z szeroko rozumianą sztuczną inteligencją.

Literatura

Błachut K., Daniłowicz M., Szolc H., Wąsala M., Kryjak T., Komorkiewicz M., 2022, *Automotive Perception System Evaluation with Reference Data from a UAV's Camera Using ArUco Markers and DCNN*, Journal of Signal Processing Systems. https://doi.org/10.1007/s11265-021-01734-3.

Cassidy B., Kendrick C., Brodzicki A., Jaworek-Korjakowska J., Hoon Yap M., 2022, *Analysis of the ISIC image datasets: usage, benchmarks and recommendations*, Medical Image Analysis, vol. 75, 102305. https://doi.org/10.1016/j.media.2021.102305.

Davies E.R., 2005, *Circle Detection*, [w:] Davies E.R., *Machine Vision: Theory, Algorithms, Practicalities*, 3rd ed., Elsevier, s. 283–313.

Głowacz A., 2021a, *Fault diagnosis of electric impact drills using thermal imaging*, Measurement, vol. 171, 108815. https://doi.org/10.1016/j.measurement.2020.108815.

Głowacz A., 2021b, *Ventilation Diagnosis of Angle Grinder Using Thermal Imaging*, Sensors, vol. 21, no. 8, 2853. https://doi.org/10.3390/s21082853.

Głowacz A., 2021c, *Thermographic Fault Diagnosis of Ventilation in BLDC Motors*, Sensors, vol. 21, no. 21, 7245. https://doi.org/10.3390/s21217245.

Jabłoński M., 2012, *Silhouette processing via mathematical morphology with pose-aware structuring elements based on 3D model*, Image Processing & Communications, vol. 17, no. 4, s. 71–78. https://doi.org/10.2478/v10248-012-0031-1.

Jabłoński M., 2013, *Data fusion in Web-based visualization of 3D scene semantic model*, Image Processing & Communications, vol. 18, no. 2–3, s. 61–70. https://doi.org/10.2478/v10248-012-0080-5.

Jabłoński M., Przybyło J., 2016, *Evaluation of MoG video segmentation on GPU-based HPC system*, Computing and Informatics, vol. 35, no. 5, s. 1141–1159.

Jaworek-Korjakowska J., Brodzicki A., Cassidy B., Kendrick C., Hoon Yap M., 2021a, *Interpretability of a deep learning based approach for the classification of skin lesions into main anatomic body sites*, Cancers, vol. 13, no. 23. https://doi.org/10.3390/cancers13236048.

Jaworek-Korjakowska J., Kostuch A., Skruch P., 2021b, *SafeSO: interpretable and explainable deep learning approach for seat occupancy classification in vehicle interior*, [w:] *CVPRW 2021: 2021 IEEE/CVF conference on Computer Vision and Pattern Recognition Workshops: virtual, 19–25 June 2021: proceedings*, IEEE, Piscataway, s. 103–112.

Kowalczyk M., Przewłocka D., Kryjak T., 2018, *Real-time implementation of contextual image processing operations for 4K video stream in Zynq UltraScale+ MPSoC*, [w:] *DASIP: the 2018 conference on Design & Architectures for Signal & Image Processing: Porto, Portugal, October 10–12, 2018*, IEEE, Piscataway, s. 37–42.

Kowalczyk M., Ciarach P., Przewłocka-Rus D., Szolc H., Kryjak T., 2021, *Real-time FPGA implementation of parallel connected component labelling for a 4K video stream*, Journal of Signal Processing Systems for Signal, Image, and Video Technology, vol. 93(5), s. 481–498. https://doi.org/10.1007/s11265-021-01636-4.

Piekarski M., Jaworek-Korjakowska J., Wawrzyniak A., Gorgoń M., 2020, *Convolutional neural network architecture for beam instabilities identification in Synchrotron Radiation Systems as an anomaly detection problem*, Measurement, vol. 165, 108116. https://doi.org/10.1016/j.measurement.2020.108116.

Przewłocka-Rus D., Kowalczyk M., Kryjak T., 2021, *Exploration of hardware acceleration methods for an XNOR traffic signs classifier*, [w:] Choraś R.S., Kurzyński M., Trajdos P., Pejaś J., Hyla T. (eds.), *Progress in image processing, pattern recognition and communication systems: proceedings of the conference (CORES, IP&C, ACS) – June 28–30 2021*, Lecture Notes in Networks and Systems, vol. 255, Springer, Cham, s. 34–45. https://doi.org/10.1007/978-3-030-81523-3_4.

Przybyło J., Jabłoński M., 2019, *Using deep Convolutional Neural Network for oak acorn viability recognition based on color images of their sections*, Computers and Electronics in Agriculture, vol. 156, s. 490–499. https://doi.org/10.1016/j.compag.2018.12.001.

Rohr D., Lindenstruth V., 2017, *Portable and Vendor-Independent Low-Level Programming and Performance Benchmarking for Graphics Cards and Processors*, [w:] *2017 IEEE 19th International Conference on High Performance Computing and Communications Workshops (HPCCWS)*, IEEE, Piscataway, s. 1–8. https://doi.org/10.1109/HPCCWS.2017.00007.

Stanisz J., Lis K., Gorgoń M., 2021, *Implementation of the PointPillars Network for 3D Object Detection in Reprogrammable Heterogeneous Devices Using FINN*, Journal of Signal Processing Systems. https://doi.org/10.1007/s11265-021-01733-4.

Świerk K., 2013, *Równoległa realizacja złożonych algorytmów wizyjnych przy pomocy procesorów graficznych*, Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie [praca dyplomowa magisterska].