



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI,
INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

KATEDRA AUTOMATYKI I INŻYNIERII BIOMEDYCZNEJ

Praca dyplomowa magisterska

*Wykorzystanie systemu operacyjnego Linux we wbudowanych
systemach wizyjnych zrealizowanych na platformie Zynq.*

*The use of the Linux operating system in embedded vision systems
implemented on the Zynq platform.*

Autor:

Wojciech Gumuła

Kierunek studiów:

Automatyka i Robotyka

Opiekun pracy:

dr Tomasz Kryjak

Kraków, 2017

Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystycznego wykonania albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.): „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej «sądem koleżeńskim».”, oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i że nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Spis treści

1. Wstęp.....	5
1.1. Cel pracy.....	7
1.2. Zawartość pracy.....	7
2. Platforma Zynq-7000.....	9
2.1. Zastosowanie systemu operacyjnego.....	11
3. Przegląd wybranych funkcjonalności platformy Zynq i systemu operacyjnego PetaLinux...	17
3.1. Obsługa SSH	17
3.2. FPU i technologia NEON.....	18
3.3. Protokół AXI	20
3.4. Obliczenia równoległe.....	24
3.5. OpenCV	26
3.6. Integracja algorytmów w FPGA i CPU.....	27
3.7. Przerwania systemowe.....	28
4. System wizyjny zrealizowany na platformie Zynq z systemem operacyjnym PetaLinux.....	33
4.1. Moduł wyznaczania różnicy kolejnych ramek w sekwencji obrazów.....	34
4.2. Moduł generacji tła.....	37
4.3. Integracja z systemem PetaLinux	41
5. Konfiguracja projektu	45
5.1. Podstawowa konfiguracja projektu.....	45
5.2. Konfiguracja modułu AXI DMA.....	49
5.3. Konfiguracja modułu AXI VDMA.....	53
5.4. Obliczenia równoległe.....	55
5.5. Biblioteka OpenCV	55
5.6. Wykorzystanie mechanizmu przerwań systemowych	57
5.7. Konfiguracja projektu generacji tła	62
6. Podsumowanie	63
Dodatki.....	65

A.	Aplikacja w architekturze NEON.....	65
B.	Konwersja danych pomiędzy VDMA i OpenCV	67

1. Wstęp

Przetwarzanie obrazów i ich sekwencji stanowi pole rozległych badań naukowych i przemysłowych.

W ich ramach, projektowane są algorytmy umożliwiające akwizycję, modyfikowanie, analizę, rozpoznawanie treści i prezentację obrazów. Często motywacją badań jest próba naśladowania zjawisk związanych z narządem wzroku człowieka i dążenie do uzyskania takiego opisu sposobu działania, aby umożliwić wykonanie zbliżonego do nich algorytmu przy użyciu układów elektronicznych. Odmiernym zagadnieniem jest poszukiwanie możliwości realizacji przetwarzania obrazów w taki sposób, aby uzyskać informacje niewidoczne dla ludzkiego oka, w oparciu o parametry obrazu o niewielkiej zmienności. Temat ten obejmuje analizę obrazów w celu wykrycia możliwych modyfikacji obrazu oryginalnego czy algorytmy częstotliwościowe.

Techniki przetwarzania obrazów opierają się zwykle na analizie i redukcji informacji zawartej w sekwencji pikseli w taki sposób, aby uzyskać obraz wynikowy, na którym uwypuklone będą kluczowe z punktu widzenia algorytmu własności. Wynikiem działania procedury może być również zbiór cech opisujących badane zjawiska.

Zdefiniować można szereg operacji składających się na proces przetwarzania obrazu. [1]

- Akwizycja – przygotowanie cyfrowej reprezentacji obrazu, czytelnej dla układu obliczeniowego.
- Przetwarzanie – proces modyfikacji danych wejściowych w celu przystosowania do obróbki algorytmicznej. Wykorzystujący, między innymi, operacje skalowania, zmiany przestrzeni barw czy usuwania zakłóceń.
- Analiza – redukcja informacji wizyjnej w celu uzyskania opisu jakościowego lub ilościowego badanych cech i eliminacja zbędnych z perspektywy rozpatrywanego zadania informacji.
- Rozpoznawanie – proces uzyskiwania informacji wynikowych na podstawie wektora cech.

Techniki przetwarzania obrazów, a zwłaszcza ich sekwencji znajdują zastosowanie w coraz większej liczbie dziedzin.

Jedną z dziedzin wykorzystujących techniki wizyjne, która jest prężnie rozwijana w ostatnich latach jest budowa systemów ADAS (*ang.* Advanced driver-assistance systems). Ich działanie, poza sygnałami wizyjnymi, wymaga użycia sygnałów o innych charakterystykach, między innymi czujników optycznych

oraz systemów *LIDAR*. Celem projektowania zaawansowanych systemów wsparcia kierowcy jest stopniowe zwiększanie autonomii pojazdów i ograniczenie zaangażowania kierowcy. W szerszej perspektywie, rozwój systemów ADAS może pozwolić na zaprojektowanie pojazdów w pełni autonomicznych, pozwalających na transport osób i towarów bez udziału kierowców. Dane z czujników wizyjne mogą być przetwarzane w celu uzyskania informacji na temat jezdni, innych uczestników ruchu, oznakowania czy potencjalnych zagrożeń. Opracowanie współcześnie stosowanych technik znaleźć można w pracach [2, 3].

Inny zbiór technik wykorzystywany jest w celu detekcji i rozpoznawania twarzy oraz badania emocji. Zagadnienie to znajduje zastosowanie w ramach projektowania nie tylko systemów przemysłowych, ale jest również powszechnie stosowane w oprogramowaniu współcześnie dostępnych aparatów cyfrowych czy w ramach serwisów społecznościowych. Metody te mogą również pozwolić na budowę systemów weryfikacji użytkownika bez konieczności zdefiniowania hasła dostępu. Znajdują także zastosowanie w interfejsach przystosowanych do pracy z osobami niepełnosprawnymi. Analizę wykorzystywanych w tym celu algorytmów znaleźć można w pracy [4].

Współcześnie, coraz większe znaczenie mają również systemy śledzenia osób i analizy ich zachowań w celu wykrycia zachowań niepożądanych. Motywując to zwiększeniem bezpieczeństwa, badane są takie zagadnienia jak detekcja porzuconych bagaży, obecność osób nieuprawnionych w ustalonych strefach czy śledzenie ruchu i reidentyfikacja przy użyciu wielu kamer. Potrzeba automatyzacji wynika ze złożoności projektowanych systemów, które zasięgiem obejmować mogą całe aglomeracje i pozwalać na obserwację zachowań tysięcy osób. Z tego powodu, praktycznie niemożliwe jest zapewnienie liczby operatorów pozwalającej na wykorzystanie informacji wejściowych w pełni. Systemy te mogą działać niezależnie lub stanowić jeden z elementów zintegrowanego oprogramowania, wykorzystującego dane z wielu źródeł [5, 6, 7].

Równolegle do rozwoju algorytmów wizyjnych, badane są techniki implementacji pozwalających na wykorzystanie ich w systemach uruchamianych na układach elektronicznych różnego typu. Algorytmy wizyjne projektowane są z myślą o uruchamianiu na powszechnie stosowanych układach procesorowych w architekturach x86 lub ARM, mikrokontrolerach, układach ASIC (*ang.* Application Specific Integrated Circuit) i FPGA (*ang.* Field-Programmable Gate Array).

Pośród wymienionych platform wyróżnić można rodzinę Zynq [8], integrującą możliwości układów FPGA oraz procesorów ARM. Dzięki zastosowaniu logiki programowalnej, możliwe jest projektowanie algorytmów wizyjnych wykonywanych w sposób strumieniowy, zapewniając wysoką wydajność przy stosunkowo niskim zapotrzebowaniu na energię. Uzupełnieniem takiego układu jest procesor ARM, umożliwiający wykorzystanie algorytmów, które wymagają swobodnego dostępu do kontekstu obliczeniowego. Procesor sekwencyjny jest również, w porównaniu do układów logicznych, lepiej przystosowany do wykonywania algorytmów zdominowanych przez instrukcje lub takich, których sprzętowa implementacja jest trudna lub niemożliwa.

Układy Zynq pozwalają wykorzystać zalety algorytmów projektowanych z myślą o implementacji przy użyciu języków HDL (*ang.* Hardware Description Language) oraz powszechnie stosowanych

języków proceduralnych. Układ ten pozwala na uruchomienie systemu operacyjnego, ze szczególnym uwzględnieniem systemu PetaLinux [9], dzięki czemu możliwy jest dostęp do szerokiego zbioru narzędzi związanych z powszechnie stosowanymi systemami operacyjnymi.

1.1. Cel pracy

Celem niniejszej pracy było uruchomienie oraz konfiguracja systemu PetaLinux na platformie Zynq, ze szczególnym uwzględnieniem funkcjonalności, które mogą zostać wykorzystane we wbudowanych systemach wizyjnych.

Na pierwszym etapie analizę architektury układu oraz dostępnych systemów operacyjnych i systemów czasu rzeczywistego. Następnie, opracowano należy zagadnienia teoretyczne i praktyczne związane z funkcjonalnościami systemu, które mogą znaleźć zastosowanie we wbudowanych systemach wizyjnych.

Ostatecznie, działanie komponentów zaprezentowano na przykładzie wybranego systemu wizyjnego.

1.2. Zawartość pracy

Praca podzielona została na pięć rozdziałów.

Rozdział 2 zawiera opis i analizę platformy Zynq-7000. Omówiono krótko specyfikację układu. Poruszono zagadnienia związane z dostępnymi systemami operacyjnymi, z uwzględnieniem zalet i wad każdego z proponowanych rozwiązań. Zbadano również możliwość wykorzystania systemów czasu rzeczywistego.

Rozdział 3 zawiera analizę funkcjonalności układu, które mogą zostać wykorzystane w systemach wizyjnych. Zbadano możliwości wykorzystania systemu operacyjnego PetaLinux i jego integracji z układem reprogramowalnym. Opisano również protokół AXI, ze szczególnym uwzględnieniem modułów AXI DMA (*ang.* Direct Memory Access) oraz VDMA (*ang.* Video DMA).

W rozdziale 4 zaprezentowano system wizyjny wykorzystujący omawiane funkcjonalności, którego zadaniem jest generacja tła. Zaproponowano metody integracji rozwiązań budowanych z wykorzystaniem obu części układu, wskazano ograniczenia i potencjalne kierunki rozwoju.

Rozdział 5 zawiera zbiór instrukcji związanych z konfiguracją funkcjonalności omawianych w poprzednich rozdziałach, na przykładzie układu ZYBO. Zaprezentowano w nim kroki wymagane do poprawnej konfiguracji wykorzystywanych elementów systemu oraz wskazano metody umożliwiające weryfikację poprawności działania.

2. Platforma Zynq-7000

Karta uruchomieniowa ZYBO jest przedstawicielem rodziny układów SoC (*ang.* System-on-a-chip) Zynq-7000. SoC to układy scalone integrujące zbiór układów elektronicznych, takich jak mikroprocesor, układy koprocesujące, interfejsy wejścia i wyjścia, czy pamięci. Są one powszechnie wykorzystywane do projektowania systemów wbudowanych. Centralną część układu rodziny Zynq-7000 stanowi dwurdzeniowy procesor o architekturze ARM w wersji Cortex-A9, współpracujący z działającym równolegle układem FPGA, opartym na architekturze Artix-7 lub Kintex-7. [10] Są to układy heterogeniczne, łączące w sobie elementy klasycznego układu FPGA (*PL, ang.* Programmable Logic) oraz procesora ARM (*PS, ang.* Processing System).

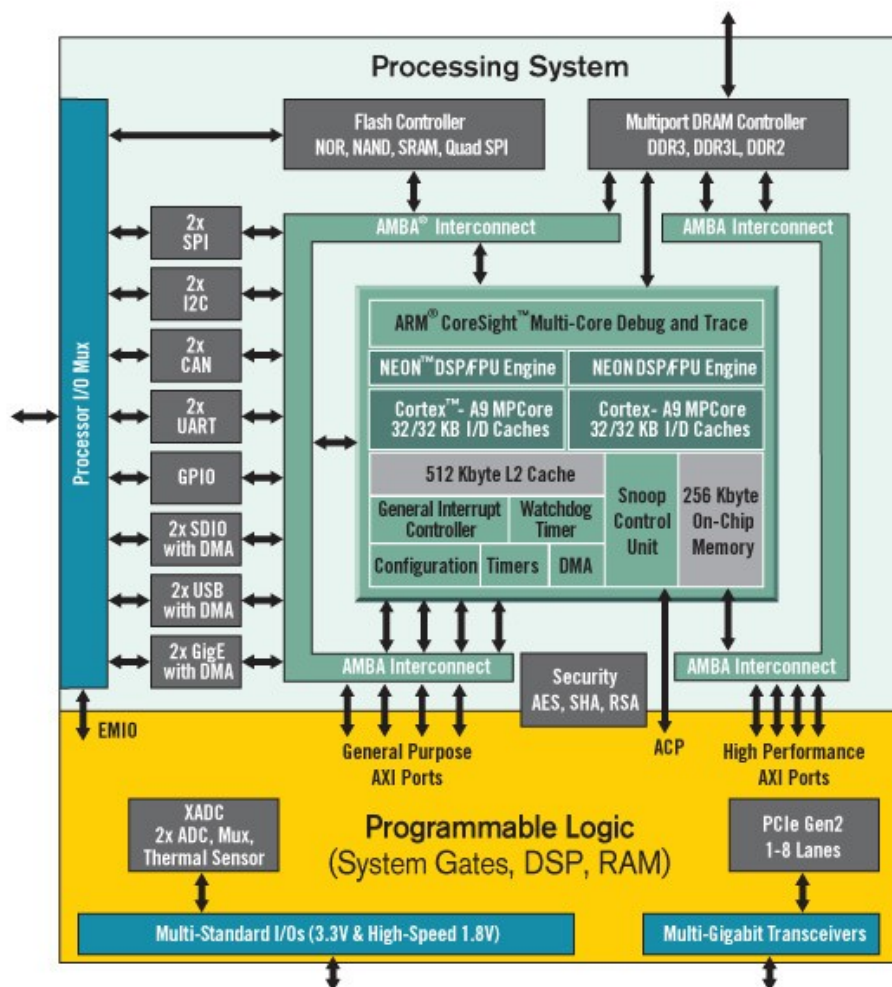
Karta ZYBO wyposażona jest ponadto w 512 MB pamięci RAM, złącza HDMI i VGA do transmisji obrazu, gniazda Jack do przesyłu sygnału dźwiękowego, gniazdo USB oraz slot pamięci MicroSD. Komunikacja sieciowa jest możliwa dzięki implementacji stosu TCP/IP i obecności gniazda RJ-45. [11]

Układy rodziny Zynq-7000 są stosowane w aplikacjach systemów wsparcia kierowców, systemach wizyjnych wysokich rozdzielczości, cyfrowego przetwarzania sygnałów czy kryptograficznych. Zaproponowano wykorzystanie zalet elementów FPGA i CPU do projektowania systemów wsparcia kierowców ADAS, co pozwoliło na redukcję czasu odpowiedzi systemu i zapotrzebowania na energię. [12] W innej pracy zbadano możliwość wykorzystania układu do transmisji sygnału o rozdzielczości 4K (3840×2160 pikseli) przy możliwie niewielkim zużyciu zasobów i energii. [13] Wśród zagadnień kryptograficznych, realizowanych przy użyciu omawianej rodziny układu wymienić można algorytmy generowania liczb pseudolosowych o wysokiej rozdzielczości. [14]

Na rysunku 2.1 przedstawiono schemat omawianej architektury.

Schemat przedstawia podział układu na części *PL* – oznaczonej kolorem żółtym, oraz *PS* – na zielono. Architektura części programowalnej zbliżona jest do powszechnie stosowanych układów FPGA. Wyposażono ją jednak w zbiór portów umożliwiających wydajną komunikację z procesorem. Ponadto, konfiguracja tej części wykonywana jest na starcie przez procesor lub przy użyciu interfejsu JTAG i układ nie zawiera elementów pozwalających na wykorzystanie logiki programowalnej niezależnie. Część procesorowa wyposażona jest w szereg interfejsów, w tym kontroler pamięci DDR3, interfejs komunikacji AMBA oraz zbiór interfejsów peryferyjnych.

Procesor wyposażony jest w koprocesor arytmetyczny (*FPU*), wspierający w obliczeniach na liczbach zmiennoprzecinkowych oraz wspiera obsługę architektury SIMD (*ang.* Single Instruction, Multiple



Rys. 2.1. Schemat architektury Zynq-7000. (Źródło: [8])

Data) – pozwalającej na przetwarzanie wielu strumieni danych przy użyciu jednego strumienia instrukcji. Zagadnienia te szerzej opisane zostały w sekcji 3.2.

Układ wyposażony jest w kontroler pamięci DDR, obsługujący żądania dostępu ze strony zarówno procesora, jak i logiki programowalnej. Pamięć jest współdzielona między obiema częściami. Pozwala to na wymianę danych, do czego wykorzystywany jest standard AXI. Zastosowanie znajduje również mechanizm DMA, pozwalający na przeprowadzanie operacji z użyciem pamięci bez udziału procesora. Interfejs pozwala na transmisję pojedynczych słów danych, umożliwiając konfigurację parametrów pracy modułów algorytmicznych, jak i na transmisję o wysokiej przepustowości. Pozwala to, dzięki zastosowaniu modułu VDMA, na przesyłanie obrazu o rozdzielczości HD z częstotliwością osiągającą wartości 680 klatek na sekundę [15]. Interfejs AXI opisano szerzej w sekcji 3.3.

Dzięki zastosowaniu kontrolera przerw (General Interrupt Controller) możliwe jest wykorzystanie modułów logiki programowalnej komunikujących się z procesorem z wykorzystaniem techniki zgłaszania żądań. Zagadnienie to opisano w sekcji 3.7.

Ponadto, dostępne są powszechnie spotykane układy zegarowe (*Timers*) i *Watchdog*, odpowiedzialny za przerwanie pracy procesora w przypadku wykrycia błędnego wykonania programu.

Kontekst pamięciowy synchronizowany jest pomiędzy oboma rdzeniami procesora dzięki modułowi *Snoop Control Unit*.

Układ logiki programowalnej należeć może do rodzin Artix-7 lub Kintex-7 i zbudowany jest z typowych elementów wykorzystywanych w FPGA o różnej liczebności, związanej z klasą układu:

- CLB (*ang.* Configurable Logic Blocks) – w tym tablice Look-up (*LUT*) – 14400 – 277400 elementów, przerzutniki (*FF*) – 28800 – 554800 elementów.
- Pamięci *Block RAM* – od 1,8Mb do 26,5Mb (50 – 755 elementów).
- Elementy *DSP* – wykorzystywane zwykle do implementacji operacji dodawania i mnożenia – 66 – 2020 elementów.
- Bloki *IOB* – umożliwiające budowę interfejsów wejściowych i wyjściowych.
- Inne – w tym interfejs JTAG, PCI Express czy konwertery analogowo-cyfrowe.

2.1. Zastosowanie systemu operacyjnego

Centralny element architektury stanowi dwurdzeniowy procesor ARM. Jest on odpowiedzialny za przeprowadzenie konfiguracji logiki reprogramowalnej. Ponadto pozwala na wykonanie dowolnego programu użytkownika. Powszechnie stosowana jest konfiguracja *bare-metal*, w której procesor wykonuje program zaprojektowany w pełni przez użytkownika. Pozwala to na uzyskanie możliwie największej kontroli nad pracą układu, ogranicza jednak możliwości wykorzystania pełni zasobów procesora oraz utrudnia projektowanie rozbudowanych aplikacji. Brak systemu operacyjnego ogranicza możliwość wykorzystania komunikacji sieciowej na etapie wykonania programu. Możliwości przechowywania wyników i logów aplikacji są niewielkie. Ponadto, użycie zewnętrznych bibliotek, w tym związanych z przetwarzaniem obrazów, takich jak OpenCV, jest niemożliwa. W efekcie, wykorzystanie konfiguracji pozbawionej systemu operacyjnego nie jest możliwe w przypadku aplikacji wymagających nadzoru bez fizycznego dostępu do układu czy przechowywania wyników.

W niniejszej pracy badano możliwość wykorzystania systemu operacyjnego Linux na przykładzie aplikacji systemów wizyjnych. Dzięki zastosowaniu Linuxa, możliwe staje się budowanie programów składających się z wielu modułów działających niezależnie. System ten wspiera obsługę sieci, co pozwala na wykorzystanie narzędzi komunikacji sieciowej, jak SSH [16], do konfiguracji i nadzorowania działania aplikacji, co opisano w rozdziale 3.1. Ponadto, możliwe jest wykorzystanie powszechnie dostępnych bibliotek, ułatwiających rozwój aplikacji w krótkim czasie. Zagadnienie to badano na przykładzie biblioteki OpenCV [17], udostępniającej narzędzia przetwarzania obrazów, co opisano w rozdziale 3.5.

Zbadano możliwość wykorzystania systemu PetaLinux, rozwijanego przez organizację Xilinx, jak i podstawowej wersji systemu, opartej wyłącznie na źródle jądra, oraz dystrybucji bazującej na Ubuntu Core. Ponadto, rozpatrzono możliwość użycia systemu czasu rzeczywistego do wykonania zadań obliczeniowych z zachowaniem reżimu czasowego.

2.1.1. PetaLinux

Firma Xilinx zapewnia dostęp do zbioru narzędzi *PetaLinux Tools* [9] umożliwiających przeprowadzenie procesu konfiguracji, budowania i uruchamiania systemu Linux na platformie Zynq. Dzięki zintegrowaniu koniecznych narzędzi w jednym pakiecie, proces ten jest w dużej części zautomatyzowany i ogranicza interakcję z programistą, zapewniając przy tym możliwość dowolnej konfiguracji systemu.

Celem wykorzystania omawianego pakietu narzędzi jest zbudowanie systemu operacyjnego gotowego do uruchomienia i umożliwiającego szybką rekonfigurację zarówno elementów logiki reprogramowalnej jak i samego systemu operacyjnego.

Pakiet wymaga dostarczenia zewnętrznych zależności, w tym narzędzi umożliwiających budowanie systemu – kompilatora, generatora parserów, systemu budowania – oraz zbioru narzędzi programistycznych i konfiguracyjnych. W przypadku dystrybucji Debian, zależności mogą być zainstalowane poleceniem:

```
apt-get install tofrodos iproute2 gawk gcc git make net-tools libncurses5-dev tftpd  
zlib1g-dev libssl-dev flex bison libselinux1
```

W przypadku pracy na systemie wspierającym architekturę 64-bitową, konieczne jest również zainstalowanie bibliotek programistycznych dla architektury 32-bitowej.

```
dpkg --add-architecture i386  
apt-get update  
apt-get install libc6:i386 libncurses5:i386 libstdc++6:i386  
apt-get install libgtk2.0-0:i386 libxtst6:i386 gtk2-engines-murrine:i386 lib32stdc  
++6 libxt6:i386 libdbus-glib-1-2:i386 libasound2:i386
```

Praca z pakietem wymaga ponadto wykorzystania oprogramowania *Vivado Design Suite* [18] do projektowania układu połączeń logiki reprogramowalnej oraz *Xilinx SDK* [19] do kompilacji programów uruchamianych w środowisku systemu operacyjnego układu.

Pierwszym krokiem jest wykonanie projektu w pakiecie *Vivado*. Szczegóły procesu opisano w sekcji 5.1.1. Wyeksportowany projekt Vivado jest konieczny do przeprowadzenia procesu budowania systemu, proces konfiguracji projektu PetaLinux opisano w sekcji 5.1.3.

Pakiet udostępnia możliwość dodania do budowanego systemu zbioru programów i bibliotek. Dostępne jest kilkaset pakietów oprogramowania oferowanych na zasadach wolnych licencji, w tym biblioteki do przetwarzania obrazów. Ponadto, pakiet umożliwia dodanie własnoręcznie zbudowanych aplikacji. Pozwala to na integrację etapu projektowania aplikacji oraz budowy i uruchamiania systemu operacyjnego w jednym procesie.

Projekt PetaLinux jest niezależny od projektu Vivado i może powstawać równolegle. Zmiany w strukturze modułów logiki reprogramowalnej wymagają ponownego zbudowania plików wynikowych systemu operacyjnego PetaLinux, jednak proces ten został wydzielony z oprogramowania Vivado. Pozwala to na wykorzystanie jednego projektu opisującego logikę współpracującego z aplikacjami bare metal i systemowymi.

Proces budowania systemu jest czasochłonny, na etapie prototypowania aplikacji praktyczne jest zastosowanie oprogramowania pracującego w trybie bare-metal. Pozwala to na przeprowadzanie procesu debugowania aplikacji bezpośrednio z poziomu oprogramowania Vivado/SDK. Po upewnieniu się, że sprzętowa część algorytmu działa poprawnie, zaprojektować można aplikację systemową, odpowiedzialną za komunikację, monitorowanie i wykorzystanie wyników działania algorytmu w kompletnym projekcie.

2.1.2. Inne dystrybucje systemu

Niezależnie do analizy zastosowania PetaLinux, zbadano również inne możliwości konfiguracji systemu operacyjnego do zastosowania na platformie ZYBO. Wśród dostępnych opcji, rozpatrzono dystrybucję Ubuntu Core oraz budowę systemu Linux ze źródeł.

2.1.2.1. Budowa ze źródeł

Wykorzystanie pakietu PetaLinux związane jest z ograniczeniem dostępności projektu do środowisk, dla których ten pakiet narzędzi jest dostępny. W przypadku konieczności uruchomienia projektu na systemie nie wspieranym przez twórców oprogramowania lub potrzeby wprowadzenia dużych zmian w kodzie źródłowym systemu i konfiguracji, konieczne może być przeprowadzenie pełnego procesu budowania samodzielnie. Takie podejście pozwala również na pełne zrozumienie znaczenia kolejnych kroków procesu konfiguracji.

Proces budowy systemu składa się z kilku kroków.

1. Plik binarny zawierający konfigurację części oprogramowania wykorzystującej logikę programowalną dołączany jest w trakcie procesu budowania systemu operacyjnego.
2. Konieczne jest zbudowanie dwóch programów rozruchowych (*ang.* bootloader). Pierwszy z nich – FSBL (*ang.* First Stage Boot Loader) – odpowiada za przeprowadzenie procesu wstępnej konfiguracji procesora, kontrolera pamięci i uruchomienie drugiego programu rozruchowego. Na drugim etapie rozruchu wykorzystywany jest program U-Boot. Jego zadaniem jest przygotowanie środowiska do uruchomienia właściwego systemu operacyjnego.
3. Kolejny krok wymaga zbudowania struktury drzewa urządzeń (*ang.* device tree). Pozwala ona na zdefiniowanie i konfigurację urządzeń połączonych z procesorem, dzięki czemu mogą być one obsługiwane przez system operacyjny. W przypadku układu Zynq, wykorzystanie tej struktury pozwala na konfigurację i komunikację z elementami układu FPGA.

4. Po przeprowadzeniu wstępnej konfiguracji elementów systemu, możliwe jest wykonanie procesu konfiguracji, budowania i przygotowania wynikowych plików binarnych.

Opisany proces jest skomplikowany i wymaga specjalistycznej wiedzy. Dostępne są obszerne opracowania dotyczące tematu, zawierające precyzyjny opis kolejnych wymaganych kroków [20, 21, 22]

2.1.2.2. Ubuntu Core

Ubuntu Core to dystrybucja systemu Linux dedykowana do zastosowań w urządzeniach tzw. internetu rzeczy (IoT – *ang.* Internet of Things). Dystrybucja ta oparta jest na podstawowej wersji systemu Ubuntu, przystosowana do uruchomienia na urządzeniach o ograniczonej mocy obliczeniowej.

Dzięki wykorzystaniu Ubuntu Core, możliwy jest dostęp do repozytorium oprogramowania udostępnianego przez dystrybucję. W przeciwieństwie do dystrybucji PetaLinux, instalowane oprogramowanie może być aktualizowane w trakcie pracy systemu. Cecha ta może być istotna w przypadku aplikacji działających przez długi czas, gdy aktualizacja oprogramowania jest korzystna ze względu na znalezione błędy lub poprawę bezpieczeństwa w kolejnej wersji.

System Ubuntu Core może być zbudowany i uruchomiony na karcie ZYBO przy użyciu dedykowanego narzędzia. [23]

2.1.3. RTOS

System operacyjny czasu rzeczywistego (*RTOS, ang.* Real Time Operating System) to system operacyjny, którego zadaniem jest obsługa aplikacji przy zachowaniu założeń o nieprzekroczeniu maksymalnego dopuszczalnego czasu odpowiedzi programu. Pozwala to na projektowanie aplikacji, w których czas odpowiedzi ma kluczowe znaczenie, w tym systemów sterowania lub krytycznych aplikacji wizyjnych. Dzięki zastosowaniu dwurdzeniowego procesora w układzie Zynq, rozważyć można zaprojektowanie rozwiązania, w których jeden z rdzeni odpowiada za wykonanie programu Linuxa, a drugi – aplikacji lub systemu czasu rzeczywistego.

Rozpatrzono możliwość uruchomienia systemu operacyjnego PetaLinux i jego współpracę z aplikacjami czasu rzeczywistego na przykładzie OpenAMP [24]. OpenAMP zapewnia interfejs umożliwiający komunikację pomiędzy programami działającymi w systemie Linux oraz aplikacjami czasu rzeczywistego, wykorzystując do tego narzędzia dostępne już w systemie. Z punktu widzenia klasycznego systemu operacyjnego, program działający na systemie czasu rzeczywistego jest zewnętrznym zasobem, który oczekuje na zlecenie wykonania konkretnego zadania i wysyła odpowiedź. Dzięki wykorzystaniu systemu czasu rzeczywistego FreeRTOS [25], aplikacje mogą mieć dostęp do zasobów systemowych, w tym pamięci i interfejsów komunikacji.

System czasu rzeczywistego może być wykorzystany do obliczeń o krytycznym znaczeniu. W przypadku wykorzystania klasycznego systemu operacyjnego, nie jest możliwe zagwarantowanie wykonania dowolnego zadania w określonym czasie. W trakcie działania aplikacji, system może zdecydować o jej czasowym zatrzymaniu i udostępnieniu zasobów innemu z oczekujących zadań. Aplikacja działająca w czasie rzeczywistym pozwala uniknąć tego zjawiska.

System PetaLinux oferuje dostęp do modułów `RPMsg`, `remoteproc`, `virtIO`, które są wymagane do zapewnienia komunikacji z systemem czasu rzeczywistego. Udostępnione zostały również aplikacje testowe, które pozwalają na sprawdzenie poprawności działania konfiguracji. Użycie systemu czasu rzeczywistego wymaga zmian projektowych, w tym konfiguracji dwóch instancji konsoli do komunikacji szeregowej i zdefiniowania struktury drzewa urządzeń określającej obszar pamięci dla obu systemów operacyjnych. Po zbudowaniu poprawnie skonfigurowanego systemu i jego uruchomieniu, przetestowanie działania aplikacji wymaga użycia poniższych poleceń.

```
modprobe zynq_remoteproc firmware=image_echo_test
modprobe rpmsg_user_dev_driver
echo_test
```

W rezultacie uruchomiono moduły odpowiedzialne za obsługę systemu czasu rzeczywistego i przeprowadzono test komunikacji. Konfiguracja i wykorzystanie systemów czasu rzeczywistego wykracza poza zakres niniejszej pracy, a zagadnienie jest obiektem obszernych opracowań [26, 27].

Podsumowanie

Zarówno wykorzystanie pakietu PetaLinux Tools, jak i obu pozostałych metod pozwala na zbudowanie w pełni funkcjonalnej dystrybucji systemu Linux i uruchomienie jej na układzie Zynq. Każda z metod wiąże się z innymi ograniczeniami i udostępnia inne możliwości. W przypadku użycia narzędzi PetaLinux, użytkownik uzyskuje dostęp do ograniczonego zbioru dodatkowych aplikacji, niewielkiej w porównaniu do repozytoriów udostępnianych w dystrybucji Ubuntu Core. Ponadto aktualizacja oprogramowania może wymagać ponownego zbudowania systemu lub nie być możliwa bez aktualizacji pakietu narzędzi.

Dystrybucja Ubuntu Core zapewnia dostęp do aktualizacji samego systemu, pozwalając na zachowanie bezpieczeństwa działania i dostęp do poprawek kodu oprogramowania. Może to być kluczowe w przypadku wykorzystania układu Zynq do działania przez długi czas z oczekiwaną niezawodnością.

W przypadku konieczności dostosowania kodu systemu operacyjnego do własnych potrzeb, praktyczne staje się natomiast wykorzystanie technik budowy systemu bezpośrednio ze źródeł. Ogranicza to jednak możliwości instalacji dodatkowego oprogramowania i wymaga dobrej znajomości zagadnień związanych z działaniem systemu Linux.

Pakiet PetaLinux pozwala jednak na największą integrację z oprogramowaniem Vivado, co ułatwia prototypowanie aplikacji. Dzięki udostępnieniu repozytorium oprogramowania oraz braku konieczności ingerencji użytkownika w proces budowania systemu, wykorzystanie go jest najlepszym rozwiązaniem w większości projektów. Z tego powodu, w niniejszej pracy zdecydowano się na użycie tego rozwiązania na dalszym etapie projektu.

Zastosowanie systemu czasu rzeczywistego współpracującego z innym systemem operacyjnym pozwala na wykonanie krytycznych sekcji kodu z zachowaniem ograniczeń czasowych. Pamiętać należy jednak, że wiąże się to z ograniczeniem maksymalnej wydajności operacji wykonywanych przez klasyczny system operacyjny.

Firma Xilinx zrezygnowała ze wsparcia dla systemu FreeRTOS i podobnych i zdecydowano się na oparcie na bibliotece OpenAMP do realizacji zadań wykonywanych w czasie rzeczywistym. W okresie powstawania pracy, literatura omawiająca integrację biblioteki z systemem operacyjnym dla kart innych producentów nie była powszechnie dostępna. Z tego powodu, realizacja omawianych zadań w przypadku karty ZYBO była poważnie utrudniona.

3. Przegląd wybranych funkcjonalności platformy Zynq i systemu operacyjnego PetaLinux

W ramach pracy przeprowadzono analizę wybranych funkcjonalności układów rodziny Zynq działających pod kontrolą systemu operacyjnego PetaLinux, które mogą zostać zastosowane podczas realizacji wbudowanych systemów przetwarzania obrazów. Wybrane zagadnienia przedstawiono w poniższych podrozdziałach.

3.1. Obsługa SSH

Po połączeniu karty z uruchomionym systemem PetaLinux z komputerem przez interfejs USB, możliwe jest otworenie konsoli komunikacji przy użyciu protokołu transmisji szeregowej. Komunikacja odbywa się z prędkością 115200 bodów, ośmioma bitami danych, jednym bitem stopu i bez bitu parzystości.

Komunikacja przy użyciu transmisji szeregowej jest jednak ograniczona do sytuacji, w których możliwy jest bezpośredni dostęp do układu. Ponadto, nie zapewnia wysokiej przepustowości transmisji czy możliwości przesyłu plików. Z tych powodów, korzystne staje się wykorzystanie protokołu SSH (*ang.* Secure Shell) do nawiązania komunikacji sieciowej. SSH jest najczęściej stosowanym protokołem bezpiecznej komunikacji, wspieranym przez zdecydowaną większość dystrybucji systemu Linux i nie wymagającym dodatkowej konfiguracji na etapie budowania systemu. Dzięki zastosowaniu technik szyfrowania połączenia, zapewnia mechanizm nawiązywania kryptograficznie bezpiecznej komunikacji między dwoma urządzeniami. Wśród najczęściej wykorzystywanych funkcji protokołu wymienić można udostępnienie zdalnej konsoli poleceń, przesyłanie plików czy tunelowanie połączeń.

Połączenie odbywa się przy użyciu poniższego polecenia.

```
ssh root@adres-ip-urzadzenia
```

Domyślne hasło administratora w przypadku dystrybucji PetaLinux to `root`. Może być ono zmienione na etapie konfiguracji systemu z wykorzystaniem poniższych poleceń.

```
petalinux-config -c rootfs  
Petalinux RootFS Settings -> Root password
```

Aby uprościć proces logowania, wykorzystać można mechanizm wymiany kluczy, zapewniany przez protokół. Weryfikacja obu stron połączenia przy użyciu kluczy wykorzystuje techniki kryptografii asymetrycznej. Użytkownik posiada parę związanych ze sobą kluczy kryptograficznych, umownie nazywanych kluczem publicznym i prywatnym. Wiadomość zaszyfrowana przy użyciu klucza publicznego może zostać odszyfrowana wyłącznie z użyciem klucza prywatnego. Użytkownik udowodnić może swoją tożsamość przez przesłanie oryginału otrzymanej wiadomości, zaszyfrowanej przy użyciu klucza publicznego. Korzystając z tej zależności, klucz publiczny może być powszechnie znany i wykorzystywany do budowania kryptograficznie bezpiecznych wiadomości, pod warunkiem zachowania klucza prywatnego w tajemnicy.

Algorytmy kryptografii asymetrycznej wykorzystywane są przez narzędzie SSH na etapie weryfikacji tożsamości obu stron komunikacji. Po nawiązaniu połączenia, komunikacja zabezpieczana jest wybranym algorytmem symetrycznym. Wykorzystanie mechanizmu kluczy wymaga użycia poniższego polecenia.

```
ssh-copy-id -i ~/.ssh/id_rsa.pub root@adres-ip-urzadzenia
```

Umożliwia to logowanie w przyszłości bez konieczności podania hasła użytkownika. Skonfigurowany w opisany sposób protokół daje dostęp do pełnego zbioru narzędzi, w tym zdalnej obsługi konsoli użytkownika, przesyłania plików, tunelowania portów czy zdalnego montowania systemów plików.

3.2. FPU i technologia NEON

Układ Zynq wyposażony jest w koprocessor arytmetyczny oraz wspiera polecenia wykorzystujące technologię NEON [28]. Elementy te pozwalają na zwiększenie wydajności projektowanych aplikacji w przypadku, gdy wykonywane operacje wymagają przeprowadzania obliczeń na liczbach zmiennoprzecinkowych lub działań wektorowych.

Koprocessor arytmetyczny, FPU (*ang.* Floating-point unit), to układ działający we współpracy z jednostką procesora, dedykowany do wykonywania obliczeń na liczbach zmiennoprzecinkowych. Wykorzystanie dedykowanego układu pozwala na zwiększenie szybkości wykonywania operacji arytmetycznych, pierwiastkowania i przesunień bitowych. W przypadku braku dedykowanego układu FPU, konieczne jest symulowanie jego działania dla jednej operacji na liczbach zmiennoprzecinkowych przez wykonywanie większej liczby operacji na typach całkowitych, co wiąże się ze spadkiem wydajności.

Technologia NEON pozwala na rozszerzenie puli rozkazów procesora ARM o polecenia wykorzystujące architekturę SIMD zdefiniowaną przez taksonomię Flynna[29]. SIMD (*ang.* Single Instruction stream, Multiple Data streams) to klasa systemów, które pozwalają na przetwarzanie wielu strumieni danych na podstawie jednego strumienia instrukcji. Zastosowania tej architektury obejmują zagadnienia, w których dla wielu wartości wejściowych konieczne jest wykonanie tej samej operacji. Cechę tę posiada wiele operacji związanych z przetwarzaniem sygnałów i obrazów, w tym wyznaczanie wartości szybkiej transformaty Fouriera, implementacje filtrów FIR i IIR czy operacje skalowania, rotacji i filtracji uśredniającej obrazu.

Rozpatrzono możliwość wykorzystania architektury NEON w zagadnieniach przetwarzania sygnałów. Działanie testowano na podstawie programu wyznaczającego wartość iloczynu skalarnego dwóch wektorów zadanej długości. Porównano trzy implementacje algorytmu, którego kod źródłowy zawarto w dodatku A. Wykorzystano implementację bazową oraz stosującą polecenia dostępne w architekturze NEON i porównano wyniki z implementacją zaprojektowaną w asemblerze.

Implementacja w architekturze NEON wykorzystuje dedykowane funkcje, udostępnione w bibliotece `arm_neon.h`, które mają na celu maksymalne zwiększenie wydajności aplikacji. W przypadku pozostałych implementacji, stosowane są polecenia wykonywane na koprocesorze VFP (*ang.* Vector Floating-Point). VFP to układ niezależny od FPU, pozwalający na wykonanie jednej instrukcji dla wektora danych wejściowych. Układ ten nie należy do rodziny SIMD i wykonuje instrukcje sekwencyjnie, w przeciwieństwie do architektury NEON.

Tabela 3.1. Wyniki testu wydajnościowego.

Bez optymalizacji			
Implementacja	min [s]	max [s]	średnio [s]
Bazowa	0,4266	0,4339	0,4296
NEON	0,1103	0,1108	0,1105
ASM	0,4082	0,4086	0,4083
Z optymalizacjami			
Bazowa	0,1080	0,1152	0,1092
NEON	0,1088	0,1147	0,1090
ASM	0,1087	0,1144	0,1089

Eksperyment przeprowadzono z wykorzystaniem karty ZYBO działającej pod kontrolą systemu PetaLinux. Rozpatrzono przeprowadzenie procesu kompilacji z wyłączonymi optymalizacjami kompilatora (flaga `-O0`) oraz z włączonymi wszystkimi optymalizacjami (`-O3`). Wykorzystane poleceń NEON wymaga użycia odpowiadających im parametrów kompilacji. Poniżej przedstawiono polecenie kompilacji testowej implementacji wykorzystującej NEON.

```
arm-linux-gnueabihf-gcc -Wall -O3 -mcpu=cortex-a9 -mfpu=neon -ftree-vectorize -
mvectorize-with-neon-quad -mfloat-abi=hard -ffast-math -funsafe-math-
optimizations -g -c -o "src/main.o" "../src/main.c"
```

Wyniki testów wydajności zebrano w tabeli 3.1. W sytuacji, gdy wyłączono optymalizację na etapie kompilacji, zauważalny jest znaczny wzrost wydajności w przypadku wykorzystania instrukcji udostępnianych przez architekturę NEON. Pozwala ona na niemal czterokrotne zwiększenie szybkości działania programu względem pozostałych implementacji. Różnica ta zanika w przypadku wykorzystania możliwości optymalizacji kodu programu na etapie kompilacji. Różnica w szybkości wykonania programu NEON jest niewielka, gdyż pozostałe implementacje zostały zoptymalizowane do stanu, w którym koszt ich wykonania porównywalny jest z implementacją w NEON.

Wyniki pozwalają wnioskować o słuszności wykorzystania instrukcji udostępnianych przez architekturę NEON ze względu na możliwy wzrost wydajności. Istotna jest jednak weryfikacja wyników i potwierdzenie poprawy działania aplikacji. W przypadku, gdy różnice między programami są niewielkie, użycie instrukcji NEON może być niekorzystne ze względu na zwiększoną latencję wykonania rozkazów.

3.3. Protokół AXI

Protokół AXI (*ang.* Advanced eXtensible Interface) zdefiniowany został w specyfikacji AMBA (*ang.* Advanced Microcontroller Bus Architecture) 3. W kolejnej wersji dokumentu sprecyzowano standard w najnowszej wersji – AXI4 [30]. Protokół wykorzystywany jest do komunikacji pomiędzy elementami układu lub modułami zbudowanymi wewnątrz logiki reprogramowalnej i jest dedykowany systemom o dużej wydajności i pracującym z wysoką częstotliwością.

Specyfikacja definiuje trzy typy interfejsu:

- AXI4 – wykorzystujące technikę MMIO (*ang.* Memory-Mapped Input/Output) do odwzorowania rejestrów w przestrzeni adresowej pamięci RAM i dedykowanej aplikacjom wymagającym dużej wydajności komunikacji.
- AXI4-Lite – uproszczona wersja protokołu, wykorzystująca MMIO i dedykowana aplikacjom o mniej rozbudowanych wymaganiach komunikacyjnych.
- AXI4-Stream – wersja przepływową protokołu, nie wykorzystująca technik MMIO.

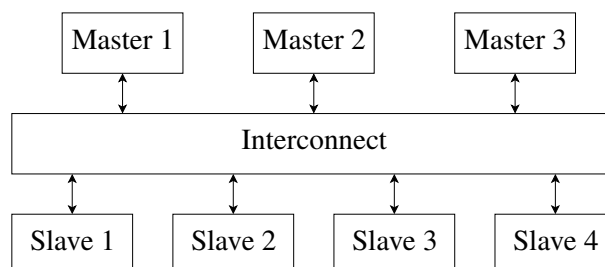
Interfejsy wykorzystujące technikę MMIO stosowane są powszechnie w zadaniach konfiguracji modułów aplikacji czy przesyłania informacji, takich jak ramka sygnału wizyjnego do pamięci procesora. Dzięki reprezentacji stanu elementów logiki reprogramowalnej w postaci komórek pamięci operacyjnej procesora, możliwa jest jednolita analiza działania całego systemu.

Interfejs w wersji *Stream* wykorzystywany jest natomiast do przesyłania sygnału pomiędzy kolejnymi elementami układu, na przykład transmisji kolejnych pikseli obrazu pomiędzy kolejnymi składowymi algorytmu przetwarzania obrazu. Proces przesyłania danych w takiej formie charakteryzuje się większą wydajnością, analiza działania aplikacji jest jednak utrudniona ze względu na brak reprezentacji przesyłanych danych w pamięci.

Możliwe jest również połączenie obu typów interfejsu wewnątrz jednego elementu. Technika ta wykorzystana została w przypadku elementu AXI VDMA, umożliwiając manipulowanie ramkami obrazu wizyjnego przesyłanymi przy użyciu interfejsu *Stream* dzięki buforowaniu w pamięci RAM. Zagadnienie to szerzej opisano w rozdziale 3.3.3. Podobne techniki wykorzystano również w przypadku interfejsu Ethernet DMA, umożliwiającego komunikację przy użyciu protokołu Ethernet.

3.3.1. Przebieg transakcji

Transakcja komunikacyjna odbywa się pomiędzy dwoma urządzeniami – *master* i *slave*, jednak dzięki zastosowaniu elementów AXI-Interconnect możliwe jest połączenie wielu urządzeń, co przedstawiono na schemacie 3.1.



Rys. 3.1. Schemat połączenia Interconnect w protokole AXI.

Komunikacja odbywa się przy użyciu pięciu niezależnych kanałów:

- Read Address
- Write Address
- Read Data
- Write Data
- Write Response

Każdy kanał zawiera zbiór sygnałów wykorzystywanych w trakcie wymiany danych.

Transmisja rozpoczyna się od wykorzystania sygnałów *valid* i *ready*. Urządzenie źródłowe wymusza stan wysoki sygnału *valid* i oczekuje na zmianę wartości sygnału *ready* urządzenia docelowego na stan wysoki. W chwili, gdy oba sygnały znajdują się w tym stanie, właściwe dane mogą zostać przesłane z urządzenia źródłowego do docelowego. Pozwala to na przekazanie takich danych jak adres odczytu/zapisu do pamięci, odczytywanych lub zapisywanych danych i potwierdzenia zapisu. Proces nawiązania transakcji odbywa się niezależnie dla każdego wykorzystywanego kanału.

Procedura odczytu danych składa się z dwóch etapów:

1. Zdefiniowanie przez urządzenie *master* adresu i parametrów transmisji danych na kanale *Read Address*.
2. Przesłanie przez urządzenie *slave* jednej lub więcej wartości danych na kanale *Read Data*.

Natomiast procedura zapisu wymaga trzech etapów:

1. Zdefiniowanie przez urządzenie *master* adresu i parametrów transmisji danych na kanale *Write Address*.

2. Przesłanie przez urządzenie *master* jednej lub więcej wartości danych na kanale *Write Data*.
3. Przesłanie przez urządzenie *slave* odpowiedzi na kanale *Write Response*.

Protokół pozwala ponadto na przesłanie do 256 wartości danych w trakcie jednej transmisji dzięki technice *burst*, a transakcje odczytu i zapisu danych mogą odbywać się równolegle.

Przepływ danych w interfejsie AXI4-Stream odbywa się wyłącznie w jednym kierunku i nie jest możliwy odczyt danych przesłanych wcześniej przez urządzenie *master* do *slave*. Procedura ta jest podobna do transakcji zapisu, jest jednak rozszerzona o możliwość dzielenia operacji na kilka mniejszych i łączenia wielu transakcji w jedną.

3.3.2. AXI DMA

DMA (*ang.* Direct Memory Access) to technika często stosowana w przypadku konieczności wykonywania operacji na pamięci RAM urządzenia z dużą szybkością. Wykorzystanie kontrolera DMA pozwala przeprowadzać operacje odczytu i zapisu do pamięci operacyjnej bez konieczności użycia głównej jednostki procesora. Dzięki temu, procesor odpowiada wyłącznie za skonfigurowanie kontrolera DMA i może wykonywać inne operacje w trakcie transmisji danych. Ponadto, stosowanie kontrolera DMA pozwala zwykle na uzyskanie wyższej przepustowości komunikacji z pamięcią i zmniejszenie zużycia energii. Kontroler DMA może również przeprowadzać podstawowe operacje konwersji sygnałów, na przykład, w przypadku sygnału wizyjnego, konwersję sygnałów synchronizacji obrazu – kontroler odpowiada za odpowiednie wyrównanie (*ang.* alignment) danych w pamięci, tak by zachować odstępy poprawnej długości pomiędzy kolejnymi liniami obrazu.

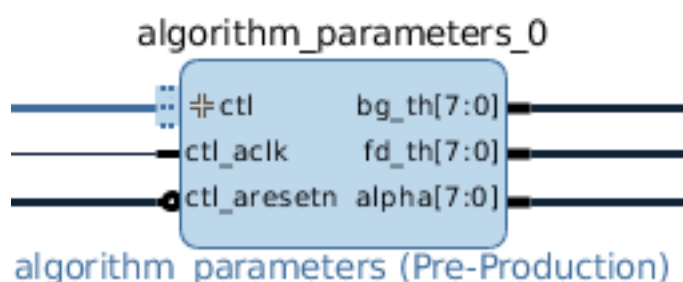
DMA pozwala na przesłanie wielu wartości danych w ramach jednej transakcji w trybie *burst*. *Master* przesyła wyłącznie adres pierwszego bajta danych, a kolejne adresy wyznaczane są w trakcie operacji przez urządzenie *slave*. Wyznaczany adres może być zwiększany, w przypadku, gdy operacja wykonywana jest w pamięci, bądź mieć stałą wartość, co ma miejsce w przypadku zapisu lub odczytu z kolejki FIFO (*ang.* First In, First Out). Interfejs pozwala również na ograniczenie dostępnej przestrzeni adresowej, w efekcie czego wartość adresu po przekroczeniu górnej granicy zakresu przyjmuje ponownie najniższą dopuszczalną wartość. Własność ta może być wykorzystana do projektowania linii buforujących.

Protokoły transmisji danych wykorzystywać mogą kolejność bajtów od najmniej znaczącego lub odwrotną. Należy o tym pamiętać na etapie projektowania modułów odpowiedzialnych za proces komunikacji. Protokół AXI DMA wykorzystuje kolejność bitów, w której najmniej znaczący bajt umieszczony jest jako pierwszy.

Dzięki zastosowaniu techniki DMA możliwa jest konfiguracja parametrów pracy algorytmu zaprojektowanego w układzie logiki reprogramowalnej oraz obserwacja jego działania na etapie wykonania z poziomu procesora ARM. Moduł algorytmiczny może udostępniać rejestry konfiguracyjne w przestrzeni adresowej pamięci procesora, a wyniki działania programu mogą być przesyłane do pamięci

operacyjnej. W szerszej perspektywie, pozwala to na udostępnienie interfejsu użytkownika, umożliwiającego nadzór nad pracą algorytmu, na przykład z poziomu konsoli dostępnej przez *ssh* lub w formie interfejsu strony internetowej. Możliwe jest również przesyłanie powiadomień z elementów logiki programowalnej do procesora ARM w celu wymuszenia reakcji na osiągnięty stan programu, na przykład przesłanie informacji o ukończeniu iteracji algorytmu dla aktualnej ramki obrazu. Można w tym celu wykorzystać mechanizm przerwań systemowych, co szerzej opisano w sekcji 3.7.

Mechanizm DMA zbadano na przykładzie projektu modułu umożliwiającego modyfikację parametrów oraz odczyt aktualnego stanu parametrów. Schemat strukturalny modułu przedstawiono na rysunku 3.2.



Rys. 3.2. Graficzna reprezentacja modułu AXI DMA w programie Vivado.

Moduł wyposażony jest w interfejs AXI, podpisany *ctl* oraz związane z nim sygnały, zegarowy – *ctl_aclk* oraz reset – *ctl_aresetn*. Sygnały wyjściowe pozwalają na odczyt zdefiniowanych parametrów z poziomu innych modułów logiki reprogramowalnej. Dzięki wydzieleniu modułu odpowiedzialnego za konfigurację algorytmu z części wykonującej obliczenia algorytmiczne, możliwe jest uproszczenie kodu języka opisu sprzętu związanego z każdym z modułów oraz zwiększenie czytelności schematu. Jeden moduł konfiguracyjny może być związany z kilkoma, działającymi niezależnie, modułami algorytmicznymi. Ponadto, zmiany w strukturze algorytmu są uproszczone. Proces projektowania oraz komunikacji z modułem przedstawiono w rozdziale 5.2.

3.3.3. AXI Video DMA

Interfejs AXI VDMA pozwala na wykorzystanie techniki DMA w przypadku aplikacji przetwarzających sygnał wizyjny. Mechanizm *Video DMA* oparty został na wykorzystaniu protokołu AXI w wersji Stream oraz Memory Mapped w połączeniu z techniką DMA do buforowania sygnału wizyjnego. Sygnał wizyjny przesyłany jest do modułu przy użyciu protokołu strumieniowego, gdzie następnie jest buforowany i zapisywany do komórek zewnętrznej pamięci RAM. Przechowywany obraz może być odczytany z poziomu procesora ARM. Moduł wspiera również komunikację w drugą stronę, pozwalając na odczyt obrazu z pamięci i przesłanie go dalej w postaci strumienia. Połączenie tych technik pozwala na wykorzystanie modułu do buforowania obrazu lub w celu rozdzielenia zadań algorytmicznych pomiędzy FPGA i CPU.

Moduł VDMA pozwala na zdefiniowanie do trzydziestu dwóch buforów ramek obrazu. Operacje mogą być wykonywane cyklicznie na każdym buforze lub stale na jednym z nich. Pozwala to na wielokrotną transmisję jednej klatki obrazu. Dane w buforze reprezentują kompletne ramki obrazu w ciągłym fragmencie pamięci, umożliwiając swobodny dostęp do poszczególnych pikseli. Struktura danych jest identyczna do tablicy zawierającej wartości kolejnych pikseli w obrazie.

Powszechnie wykorzystywanym zastosowaniem modułu jest mechanizm potrójnego buforowania, umożliwiający zmianę częstotliwości taktowania zegara sygnału wizyjnego. Zapis i odczyt danych może odbywać się niezależnie z tego samego lub różnych buforów. Dzięki zastosowaniu trzech buforów, zagwarantować można, że zapis i odczyt danych zawsze odbywa się z niezależnych obszarów pamięci, co pozwala uniknąć zjawiska nadpisania przechowywanych danych przed ich wyświetleniem.

W niniejszej pracy rozpatrzono możliwość wykorzystania modułu VDMA w celu obsługi algorytmów wymagających kontekstu w postaci dwóch kolejnych ramek obrazu. Proces konfiguracji modułu przedstawiono w sekcji 5.3.

3.4. Obliczenia równoległe

Ze względu na wykorzystanie w układzie Zynq procesora ARM o dwóch rdzeniach, możliwe jest rozważenie zagadnienia zwiększania szybkości wykonania algorytmu przez zrównoleglenie obliczeń w dwóch wątkach.

Stosując prawo Amdahla, wykazać można, że maksymalne przyspieszenie, jakie można uzyskać w systemie wieloprocessorowym jest proporcjonalne do liczby elementów obliczeniowych. Zależność ta zachodzi po warunkiem, że całe zadanie może być realizowane w sposób równoległy. W przypadku omawianego procesora, spodziewać się można korzyści nie przekraczających dwukrotnego zwiększenia szybkości wykonania algorytmu.

Zagadnienia związane z obliczeniami równoległymi stanowią obszar aktualnych badań, których efekty pozwoliły na zaprojektowanie zbioru bibliotek ułatwiających wykorzystanie własności systemów wieloprocessorowych w praktyce. W ramach pracy rozważono możliwości wykorzystania wątków natywnych oraz bibliotek Intel TBB(*ang.* Threading Building Blocks) i OpenMP do budowy aplikacji wielowątkowych.

3.4.1. Wątki natywne

Użycie wątków natywnych wymaga wykorzystania bibliotek systemowych – w przypadku aplikacji w języku C++ działającej w systemie PetaLinux zastosować można biblioteki `<thread>` lub `<pthread.h>` wchodzące w skład bibliotek standardowych [31].

Ich wykorzystanie pozwala na możliwie najbardziej efektywne użycie zasobów maszyny obliczeniowej. Wymaga to jednak dużych umiejętności programisty oraz dobrej znajomości architektury docelowej

oraz wykonywanego zadania. Ponadto zastosowanie biblioteki `<pthread.h>` wymaga zgodności systemu docelowego ze standardem POSIX, natomiast w przypadku `<thread>`, konieczne jest przeprowadzenie procesu kompilacji kompilatorem zgodnym ze standardem C++11. Założenia te mogą okazać się problematyczne w przypadku konieczności migracji aplikacji na system nie spełniający opisanych wymagań.

Stosowanie wątków natywnych pozwala na budowę wielowątkowych aplikacji działających heterogenicznie. Jest to najprostszy sposób na zbudowanie programu, w którym kilka wątków odpowiada za kilka różnych zadań. Na przykład, jeden wątek może być odpowiedzialny za przeprowadzenie obliczeń algorytmicznych, drugi za obsługę interfejsu użytkownika i przygotowanie danych wejściowych do właściwych obliczeń algorytmicznych, a kolejny – za niekrytyczne operacje po zakończeniu pracy algorytmu, takie jak przesłanie wyników do bazy danych.

3.4.2. Biblioteka Intel Threading Building Blocks

Biblioteka Intel Threading Building Blocks stanowi zbiór narzędzi rozszerzających standard języka C++ o elementy związane z obliczeniami równoległymi. Składają się na to implementacje algorytmów równoległych, struktury danych przeznaczone do wykorzystania w systemach wielowątkowych oraz implementacje operacji atomowych i algorytmów wzajemnego wykluczania [32].

Użycie biblioteki opiera się na zastosowaniu jej elementów na etapie powstawania aplikacji. Z tego względu, podobnie jak w przypadku wątków natywnych, konieczne jest zaprojektowanie aplikacji w sposób możliwie najlepiej wykorzystujący zalety biblioteki. Refaktoryzacja kodu istniejącego programu w taki sposób, by zastosować *TBB* może być utrudniona i ostatecznie nie pozwolić na uzyskanie zadowalających wyników.

Główną zaletą stosowania *TBB* jest większa skalowalność wynikowych rozwiązań. W przypadku zastosowania wątków natywnych, konieczne jest zaprojektowanie aplikacji w sposób umożliwiający wykorzystanie innej liczby wątków, gdy będzie to konieczne. W przypadku zastosowania dodatkowej biblioteki, stanowi ona warstwę abstrakcji pomiędzy programistą a warstwą obliczeniową, dzięki czemu proces możliwie najlepszej integracji aplikacji z platformą docelową odbywać się może przy niewielkiej interakcji ze strony projektanta. Biblioteka *TBB* zgodna jest z ideą programowania generycznego, paradygmatu powszechnie stosowanego w aplikacjach projektowanych w języku C++. Jej stosowanie stanowi naturalne rozszerzenie możliwości tego języka i nie wymaga szerokiej wiedzy na temat architektury systemu docelowego.

3.4.3. Biblioteka OpenMP

Biblioteka OpenMP to interfejs programowania aplikacji pozwalający na budowanie wieloplatformowych programów wykonywanych równoległe. Rozwiązanie to jest dedykowane aplikacjom powstającym w językach C i C++ [33].

OpenMP składa się z dyrektyw kompilatora i zbioru bibliotek, które pozwalają kształtować zachowanie programu na etapie wykonania. Ze względu na wykorzystanie dyrektyw kompilacji, możliwa jest integracja biblioteki z istniejącą aplikacją, nie wymagając przy tym modyfikowania właściwego kodu programu. Wymaga to wyłącznie znajomości aplikacji w stopniu umożliwiającym identyfikację obszarów, których równoległe wykonanie pozwoli na osiągnięcie największych zysków, obserwowanych w formie przyspieszenia działania programu.

Analogicznie jak w przypadku biblioteki *TBB*, zastosowanie OpenMP ma na celu zapewnienie skalowalności aplikacji i dodanie warstwy abstrakcji pomiędzy kod programu, a operacje wykonywane na wątkach obliczeniowych. Obie biblioteki wyposażone są również w algorytmy równoważenia obciążenia.

Wykorzystanie biblioteki OpenMP wymaga zastosowania kompilatora wspierającego dyrektywy wchodzące w jej skład. Ze względu na specyfikę stosowania części interfejsu biblioteki – w formie komentarzy do właściwego kodu aplikacji – możliwa jest kompilacja programu kompilatorem nie wspierającym jej. Wynikowa aplikacja nie będzie korzystać z zalet przetwarzania równoległego, jednak powinna pozwalać na uzyskanie poprawnych wyników algorytmu.

Podsumowanie

Trzy opisane rozwiązania zapewniają dostęp do różnych możliwości i obciążone są różnym kosztem stosowania. Z tego względu, nie jest możliwy jednoznaczny wybór najlepszego rozwiązania dla aplikacji wykonujących obliczenia równoległe. Często słuszne może okazać się wykorzystanie więcej niż jednej biblioteki w aplikacji, wykorzystując je do nadzoru nad zadaniami różnego typu. Część rozwiązań wymaga wsparcia kompilatora, część ograniczona jest do pewnej grupy systemów lub architektur. Wybór podejścia do obliczeń równoległych powinien stanowić etap projektowania aplikacji, a decyzja powinna uwzględniać szereg zagadnień.

W ramach pracy nie omówiono kilku innych popularnych rozwiązań związanych z obliczeniami wielowątkowymi, w tym *OpenCL* oraz *MPI*, ze względu na ograniczone możliwości ich wykorzystania w systemie PetaLinux działającym na platformie ZYBO. Badane zagadnienie stanowi temat obszernych dyskusji, których wyniki odnaleźć można w publikowanych pracach [34, 35]. W rozdziale 5.4 opisano kroki wymagane do zastosowania omawianych bibliotek w aplikacjach uruchamianych w systemie PetaLinux.

3.5. OpenCV

Algorytmy wizyjne znajdują zastosowanie w wielu aplikacjach realizowanych w ramach projektów związanych z szerokim spektrum dziedzin techniki. Ze względu na swoją popularność, część rozwiązań została zintegrowana w zbiorze bibliotek OpenCV [17]. Zbiór ten zawiera wydajne implementacje najczęściej stosowanych algorytmów, dedykowanych uruchamianiu na układach CPU oraz GPU (*ang.*

Graphics Processing Unit). Biblioteka jest szeroko stosowana ze względu na wysoką wydajność, stabilność działania i możliwość przenoszenia rozwiązań pomiędzy platformami o różnych architekturach.

Zastosowanie zewnętrznej biblioteki pozwala, w porównaniu do autorskiej implementacji algorytmu, na ograniczenie czasu wymaganego do zbudowania działającego prototypu algorytmu. Ponadto, gotowe rozwiązania zapewniają zwykle większą stabilność i wystarczającą wydajność w większości przypadków. Kluczowym ograniczeniem względem autorskiej implementacji jest brak możliwości dostosowania algorytmu do rozpatrywanego przypadku. Może to wiązać się z koniecznością sprowadzenia danych wejściowych do struktury wspieranej przez bibliotekę, co może prowadzić do ograniczenia wydajności działania aplikacji. Wśród algorytmów dostępnych w bibliotece znajdują się procedury detekcji i rozpoznawania twarzy, klasyfikacji zachowań, śledzenia obiektów, czy identyfikacji obrazów podobnych.

Wykorzystanie biblioteki OpenCV w aplikacjach *bare-metal* może być niemożliwe ze względu na liczbę zależności, które należy dostarczyć do poprawnego jej działania. Dostępność systemu operacyjnego, takiego jak PetaLinux, pozwala na dołączenie do systemu plików bibliotek, wybierając je z puli prekompilowanych zasobów, dostępnych w pakiecie PetaLinux lub dostarczając zewnętrzny zbiór bibliotek, przygotowany przez użytkownika. Możliwości te zbadano w pracy na przykładzie biblioteki OpenCV w wersjach 2.4 oraz 3.1.

Pakiet PetaLinux oferuje dostęp do prekompilowanej biblioteki w wersji 3.1. Wydanie to wciąż znajduje się w początkowej fazie rozwoju i nie zostało w pełni zaadaptowane w części środowisk. Z tego powodu, wersja 2.4 biblioteki wciąż znajduje zastosowanie. OpenCV w wersji 2.4 nie jest jednak oficjalnie dostępna w środowisku PetaLinux. Jego użycie wymaga zbudowania biblioteki na bazie kodu źródłowego wraz z zależnościami i dodanie plików wynikowych do zbioru bibliotek dostępnych w projekcie PetaLinux. Przebieg procesu dla obu przypadków opisano w rozdziale 5.5.

3.6. Integracja algorytmów w FPGA i CPU

Zagadnienie integracji rozwiązań algorytmicznych wewnątrz logiki programowalnej z obliczeniami prowadzonymi przez procesor ARM może pozwalać na zwiększenie wydajności działania aplikacji względem algorytmów realizowanych wyłącznie w jednym z tych układów. Podział algorytmu na sekwencję etapów, realizowanych w logice reprogramowalnej lub w systemie procesorowym pozwala wykorzystać zalety obu rozwiązań. Układy FPGA pozwalają na użycie zbioru algorytmów w technice strumieniowej, co umożliwia dokonać przetwarzania obrazów w czasie rzeczywistym, z zachowaniem przyjętych opóźnień.

Implementacja części algorytmów w sposób umożliwiający przetwarzanie strumieniowe może być niemożliwa. Procedury wymagające dużego kontekstu na etapie wykonania lub dostępu do danych historycznych projektowane są w celu uruchamiania na układach CPU o swobodnym dostępie do pamięci operacyjnej. Ponadto, realizacja algorytmów zdominowanych przez instrukcje lub wymagających zastosowanie wyrażeń warunkowych i pętli może być nieefektywna w porównaniu do implementacji na CPU.

Przykładami takich procedur mogą być śledzenie zmiennej liczby obiektów w kadrze czy stosunkowo proste operacje rotacji lub odbicia obrazu.

Podział algorytmu wizyjnego na etapy wykonywane naprzemiennie w logice programowalnej i procesorze ARM wymaga transmisji danych pomiędzy dwoma modułami. Realizacja tego zadania w obu kierunkach może być dokonana przy użyciu elementów AXI VDMA.

W ramach pracy, zbadano możliwość wykorzystania modułu VDMA do transmisji danych wizyjnych z poziomu FPGA do procesora ARM, w celu ich odczytu i prezentacji z wykorzystaniem komunikacji sieciowej. Zbadano również możliwość konwersji danych do struktur dostępnych w bibliotece OpenCV, co umożliwia wykorzystanie biblioteki do realizacji algorytmów wizyjnych. Procedury odczytu oraz zapisu danych przedstawiono w dodatku B. Sprawdzono szybkość działania procedur transmisji danych pomiędzy dwoma elementami obliczeniowymi dla sygnału wizyjnego o rozdzielczości HD (1280×720 pikseli), przyjmując rozmiar piksela równy trzydziestu dwóm bitom w cyklu dziesięciu tysięcy iteracji odczytu i zapisu.

Biblioteka OpenCV pozwala na użycie zewnętrznych źródeł danych do budowy struktur obrazu, a w konsekwencji pominąć etap kopiowania pamięci udostępnianej przez sterownik VDMA. W takim przypadku, czas odczytu i zapisu pełnej ramki obrazu nie przekracza kilkunastu mikrosekund. Należy jednak pamiętać, że stosując tę technikę, zagwarantować trzeba, że przetwarzanie obrazu zakończy się do chwili rozpoczęcia odczytu danych z wykorzystywanego bufora przez moduł VDMA. W przypadku zastosowania układu trzech buforów i sygnału wejściowego o częstotliwości sześćdziesięciu klatek na sekundę, daje to okno przetwarzania o długości szesnastu milisekund. Przekroczenie tej wartości może wpłynąć na błędną pracę algorytmu ze względu na nadpisanie danych przez moduł VDMA.

Z tego względu, rozważyć należy skopiowane danych wejściowych do nowego bufora na etapie odczytu oraz ponowne skopiowanie danych do bufora źródłowego w trakcie zapisu. Procedura odczytu wymaga średnio piętnastu milisekund, natomiast zapis zajmuje nie więcej niż dwanaście milisekund. Zastosowanie tej techniki nie pozwala na przetwarzanie sygnału wizyjnego w czasie rzeczywistym, gwarantuje jednak nienaruszalność danych. W połączeniu z zastosowaniem modułu VDMA o większej liczbie buforów, pozwala na integrację algorytmu o dużej złożoności obliczeniowej pomiędzy elementami obliczeniowymi obu typów. Jednakże, w celu zachowania limitów czasowych, konieczne może być ograniczenie przetwarzania danych do części klatek obrazu i pominięcie pozostałych. Nie powinno to jednak wpłynąć na zachowanie algorytmu wewnątrz logiki programowalnej.

3.7. Przerwania systemowe

Przerwanie to sygnał wysyłany przez urządzenie lub program, które ma na celu przekazanie do procesora informacji o zdarzeniu, które wymaga natychmiastowej obsługi. Przerwania podzielić można na maskowalne i niemaskowalne. Klasa przerwań maskowalnych może być zdefiniowana jako ignorowana przez właściwe ustawienie rejestrów kontrolera przerwań. Druga klasa przerwań związana jest zwykle

ze zdarzeniami o krytycznym znaczeniu, tak jak zdarzenia związane z działaniem zegarów czy układu *watchdog*¹, więc ich wystąpienie nie może zostać zignorowane przez układ procesora.

Wykorzystanie przerw pozwala na zaprojektowanie interfejsu aplikacji współpracującej ze zbiorem urządzeń peryferyjnych, takich jak czujniki, przyciski, czy klawiatury. W przypadku układu Zynq, rolę urządzeń wysyłających sygnał przerwania przyjąć mogą również elementy logiki programowalnej, takie jak układy VDMA, zegary, czy moduły zaprojektowane przez użytkownika.

Układ Zynq wyposażony jest w moduł GIC (*ang.* Generic Interrupt Controller), który pełni rolę kontrolera przerw, odpowiedzialnego za obsługę zdarzeń. *emphGIC* obsługuje przerwania z kilku źródeł:

- przerwania programowe – zbiór nie więcej niż szesnastu zdarzeń, które pozwalają na wywołanie procedury obsługi przerwania bezpośrednio z kodu aplikacji. Zachowanie to pozwala na komunikację z systemem operacyjnym i jest zwykle wykorzystywane do wywoływania operacji wejścia lub wyjścia.

Innym zastosowaniem przerw programowych jest wysłanie sygnału *yield*, który pozwala na dobrowolne wyłączenie obecnie aktywnego procesu przez układ planisty systemowego².

- przerwania systemowe – współdzielone i prywatne – zdarzenia zgłaszane przez sprzęt, na przykład klawiatury, moduły DMA czy układy pamięci. Dla każdego urządzenia lub zbioru zdefiniowana jest unikalna linia przerwania, a do kategorii zdarzenia przypisany jest identyfikator. Obsługa przerwania przez system operacyjny polega na wywołaniu agenta zdarzeń związanego z wyemitowanym identyfikatorem.

Przerwania współdzielone pozwalają na komunikację pomiędzy procesorem a urządzeniami peryferyjnymi oraz układem FPGA, mogą być definiowane przez użytkownika i być obsługiwane przez dowolny rdzeń procesora. Zdarzenia prywatne to przerwania definiowane niezależnie dla każdego rdzenia i pozwalają na obsługę zdarzeń zegarowych czy *watchdog*a.

W trakcie obsługi przerwania z dowolnego źródła, wykonanie kodu aplikacji jest wstrzymywane na czas wywołania kodu agenta obsługi zdarzenia. Stan rejestrów zapisywany jest na stosie i wykonywany jest kod odpowiedzialny za obsługę przerwania. Po zakończeniu wykonania, deklarowanego przez wywołanie instrukcji procesora *RETI*, przywracany jest zapamiętany stan aplikacji i wznowiane jest jej wykonanie.

Przerwania to zdarzenia wywoływane asynchronicznie do normalnego działania aplikacji i mogące pochodzić z wielu źródeł, co może prowadzić do sytuacji, w której kilka linii przerw sygnalizuje stan

¹Moduł odpowiedzialny za wykrycie błędnego działania systemu i wymuszający restart procesora w takiej sytuacji.

²Proces systemowy odpowiedzialny za wybór procesów do wykonania przez procesor w danej chwili. Wybrane zadania wykonywane są przez jednostkę przez ograniczony czas, po którym mogą zostać zatrzymane na rzecz innych oczekiwanych procesów.

alarmu jednocześnie, co może powodować problemy z właściwą obsługą zdarzeń. Procesor ARM pozwala na priorytetyzowanie przerwań. Przyznanie pierwszeństwa pewnemu zbiorowi krytycznych zdarzeń umożliwia rozwiązanie problemu kolejności wykonania w przypadku, gdy dwa zdarzenia o różnym priorytecie zostaną zgłoszone w tym samym czasie, a także na przerwanie obsługi przerwania o niskim priorytecie na rzecz wywołania agenta krytycznego zdarzenia. Ze względu na charakter przerwań, część zdarzeń może mieć krytyczne znaczenie dla poprawnego wykonania aplikacji, a brak ich obsługi może prowadzić do zatrzymania działania procesora lub przeprowadzenia procesu restartu.

Procedury obsługi przerwań mogą być definiowane zarówno w aplikacjach typu *bare-metal* jak i w systemach operacyjnych, takich jak PetaLinux. Obsługa przerwań w aplikacji *bare-metal* wymaga wykorzystania kontrolera do zarejestrowania agenta zdarzeń w formie funkcji aplikacji.

W przypadku systemu operacyjnego, konieczne jest wykorzystanie sterownika sprzętowego. W skład PetaLinux wchodzi sterowniki dedykowane zbiorowi modułów definiowanych w logice reprogramowalnej. Obsługa zdarzenia w niestandardowym module, zaprojektowanym przez użytkownika, wymagać może napisania dedykowanego sterownika urządzenia. Zagadnienie to wykracza poza zakres niniejszej pracy i stanowi rozbudowany proces, wymagający szerokiej wiedzy na temat działania systemów operacyjnych. Temat ten omawiany jest w pracach [36, 37].

W ramach pracy, zbadano możliwość wykorzystania przerwania emitowanego przez moduł AXI Timer, pozwalającego na wykonywanie operacji odliczania czasu oraz przerwań modułu AXI VDMA, pozwalających na przesłanie notyfikacji związanych ze zdarzeniami odczytu lub zapisu kolejnych ramek obrazu na właściwy im kanał.

Moduł AXI Timer wykorzystać można do odliczania czasu pomiędzy kolejnymi wykonaniami procedury, która powinna być wywoływana cyklicznie, w regularnych odstępach czasu. Działanie modułu AXI Timer może być warunkowane przez sygnały innych modułów logiki. Pozwala to na zaprojektowanie aplikacji, w której pewne zadanie wykonywane jest cyklicznie, pod warunkiem wystąpienia zdefiniowanego zdarzenia na poziomie logiki, a kontekst tego zdarzenia nie musi być znany z poziomu aplikacji wykonywanej przez procesor – na przykład, wykonanie przez aplikację operacji analizy danych uzyskiwanych przez algorytm wizyjny cyklicznie, pod warunkiem, że moduł algorytmu wykonywany przez układ logiczny ustawił sygnał spójności wyników. Podobne zachowanie uzyskać można korzystając z funkcji procesora i związanych z nim zegarów, co ogranicza jednak możliwości konfiguracji działania procedury do kontekstu znanego z poziomu CPU.

Przerwania definiowane przez moduł AXI VDMA pozwalają na zgłoszenie zdarzenia po wykonaniu procesu odczytu lub zapisu określonej liczby ramek obrazu lub po upływie określonego czasu od uzyskania sygnałów synchronizacji obrazu. Wykorzystanie tych mechanizmów pozwala na zaprojektowanie procedur aplikacji, które powinny być wykonane co określoną liczbę klatek sygnału wizyjnego. W szczególnym przypadku, mechanizm ten pozwala na wykonanie przez procesor operacji algorytmicznych dla każdej ramki obrazu, których implementacja sprzętowa byłaby trudna lub niemożliwa. Mechanizm ten

pozwała również na przeprowadzenie operacji końcowej analizy wyników algorytmu wizyjnego, zaimplementowanego w logice programowalnej. Proces konfiguracji projektu wykorzystującego mechanizm przerwań opisano w rozdziale 5.6.

4. System wizyjny zrealizowany na platformie Zynq z systemem operacyjnym PetaLinux

W ramach pracy zbadano możliwość realizacji systemów wizyjnych wykorzystujących możliwości obliczeniowe logiki programowalnej oraz procesora ARM i integrujących obie architektury w jednym procesie algorytmicznym. Zaproponowano moduł odpowiedzialny za generację tła w celu segmentacji obiektów pierwszoplanowych. Dane wejściowe pochodziły z kamery wideo połączonej z układem przy użyciu interfejsu HDMI. Sygnał wizyjny poddawany był przetwarzaniu i analizie przez elementy logiki programowalnej, a wyniki przesyłane były przy użyciu mechanizmu AXI VDMA do układu CPU. Proces rozpoznawania miał na celu indeksację obiektów pierwszoplanowych. Wykorzystano ponadto mechanizm AXI DMA do konfiguracji parametrów algorytmu logiki programowalnej oraz wykorzystano narzędzia systemu operacyjnego PetaLinux do obsługi procesu. Zaproponowano również mechanizm synchronizacji dwóch kolejnych ramek sygnału przy użyciu modułu AXI VDMA.

W przypadku algorytmów przetwarzania sekwencji obrazów, na etapie analizy jednej klatki wykorzystać można informacje uzyskane w trakcie obliczeń dla poprzednich ramek. Rozszerzenie kontekstu o parametry historyczne pozwala na projektowanie bardziej zaawansowanych systemów, zwykle wymaga jednak wykorzystania modułów zewnętrznej pamięci w celu przechowywania danych historycznych.

Wśród algorytmów wymagających kontekstu związanego z więcej niż jedną ramką obrazu wyróżnić można między innymi:

- segmentację obiektów pierwszoplanowych – umożliwia podział obrazu na elementy tła oraz znajdujące się na pierwszym planie, pozwalając zwykle ograniczyć obszar analizy do fragmentów obrazu, z którymi związane są obiekty pierwszoplanowe,
- śledzenie obiektów i analiza zachowań– indeksacja to proces przypisanie etykiet do obiektów i wyznaczenie zbioru niezależnych elementów obrazu. Pozwala to śledzić ruch każdego z obiektów oraz analizę zachowań. Badanie zachowań, na przykładzie przechodniów, może umożliwić detekcję obecności osób w strefach nieuprawnionych czy wykrycie osób o nieokreślonych motywacjach działań i w ten sposób przyczynić się do zwiększenia bezpieczeństwa osób,
- wyliczanie przepływu optycznego – pozwala na analizę ruchu obiektów znajdujących się w kadrze, umożliwiając estymację odległości czy parametrów ruchu.

Implementacja wymienionych typów algorytmów w architekturze potokowej może być utrudniona lub niemożliwa bez użycia zewnętrznego elementu pamięciowego. Jedna ramka obrazu kolorowego o rozdzielczości 1280×720 pikseli ma rozmiar 2,8 MB. Karta ZYBO wyposażona jest w bloki pamięci BRAM o łącznej pojemności 2,1 MB. Nawet w najprostszych systemach, wymagających buforowania wyłącznie jednej ramki obrazu nie jest więc możliwa realizacja tego zadania bez użycia pamięci zewnętrznej.

Ponadto, końcowa analiza wyników algorytmu w architekturze FPGA jest ogólnie ograniczona do przewidywanych parametrów działania systemu, a w efekcie mniej elastyczna – na przykład, zagadnienie śledzenia obiektów może być ograniczone do maksymalnej zadanej liczby niezależnych elementów. Proces analizy wymagać może budowania rozbudowanych maszyn stanu, których realizacja może być podatna na błędy, a dodanie nowych funkcjonalności utrudnione. Ze względu na te ograniczenia, korzystny może okazać się podział algorytmu na niezależne etapy, wykonywane przez moduły sprzętowe zrealizowane w logice programowalnej lub procedury uruchamiane na ARM.

Klasyczny sekwencyjny element obliczeniowy pozwala na adaptację algorytmu do zmieniających się w czasie parametrów obrazu – na przykład śledzenie zmieniającej się liczby obiektów pierwszoplanowych. Umożliwiają to mechanizmy dynamicznej alokacji pamięci czy obsługi wyrażeń warunkowych i pętli, właściwie niedostępne w przypadku implementacji sprzętowych.

Ponadto, użycie systemu operacyjnego pozwala wykorzystać zaawansowane możliwości prezentacji i przechowywania wyników działania algorytmu wizyjnego, na przykład prezentację danych wyjściowych przy użyciu interfejsu sieciowego lub zapis do bazy danych. W poniższym rozdziale zaproponowano metody wykorzystania platformy Zynq na przykładzie wybranych elementów systemów wizyjnych.

4.1. Moduł wyznaczania różnicy kolejnych ramek w sekwencji obrazów

Wyznaczenie różnicy pomiędzy dwoma kolejnymi ramkami strumienia wizyjnego wymaga obliczenia dla każdego piksela wartości różnicy, opisanej formułą (4.1).

$$d^i(x, y) = |p^i(x, y) - p^{i-1}(x, y)| \quad (4.1)$$

gdzie:

x, y – współrzędne piksela,

i – indeks ramki w sekwencji obrazów,

$p^i(x, y)$ – wartość w i -tej ramce dla piksela o współrzędnych (x, y) ,

$d^i(x, y)$ – wyznaczana wartość różnicy.

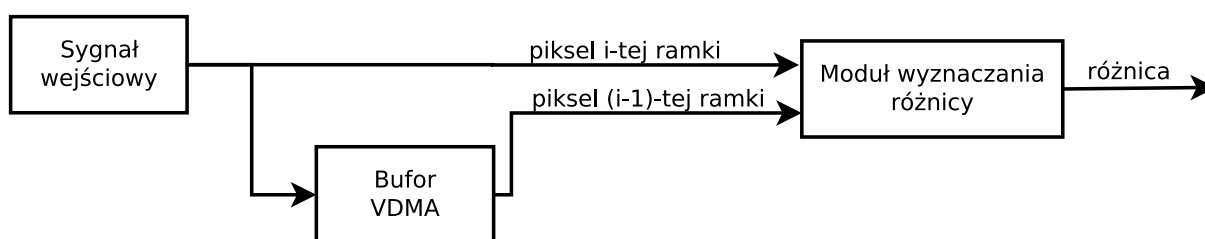
W omawianym przypadku, sygnał źródłowy i wynikowy mają postać obrazu przedstawionego w odcieniach szarości. Zastosować można również inne metryki odległości pomiędzy pikselami, na przykład

metrykę euklidesową, opisaną formułą (4.2), która w przypadku sygnałów o jednym kanale jest równoznaczna formule (4.1) ale może być również zastosowana dla obrazów kolorowych.

$$d_e^i(x, y) = \sqrt{(p^i(x, y) - p^{i-1}(x, y))^2} \quad (4.2)$$

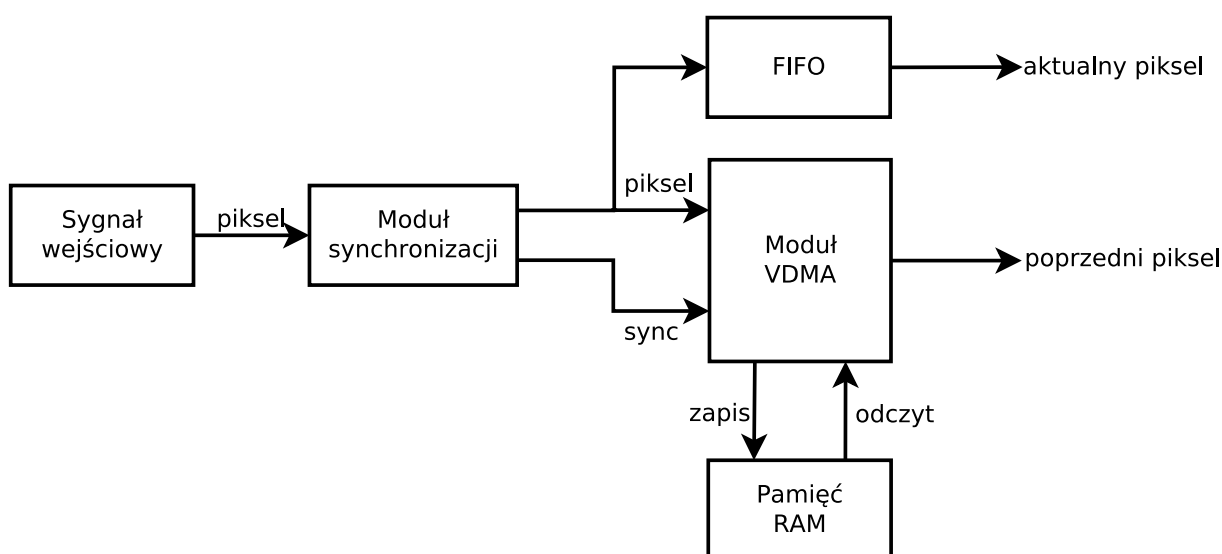
Wyznaczanie różnicy obrazów składających się z więcej niż jednego kanału jest możliwe, jednak interpretacja graficzna wyników może być nieczytelna. Obraz kolorowy poddać można redukcji do jednego kanału przed wykonaniem kroku wyznaczania różnicy. Zagadnienie obliczania różnicy dwóch kolejnych obrazów w sekwencji może stanowić przykład algorytmu, którego realizacja w systemach potokowych, pomimo niskiej złożoności obliczeniowej, może być utrudniona, ze względu na konieczność wykorzystania zewnętrznego modułu pamięci. W praktycznych realizacjach, konieczne jest wykorzystanie modułów pamięci operacyjnej w celu zapamiętania poprzedniej ramki obrazu.

Architekturę strumieniową realizującą omawiane zadanie przedstawiono na schemacie 4.1.



Rys. 4.1. Schemat architektury obliczającej różnicę sekwencji obrazów.

Wykorzystano moduł AXI VDMA w roli bufora sygnału, opóźniającego dane o pełen cykl strumieniowania ramki obrazu. Schemat elementu buforującego przedstawiono na schemacie 4.2.



Rys. 4.2. Schemat elementu buforującego ramkę obrazu.

Realizacja techniczna bufora wymagała zaprojektowania mechanizmu synchronizacji dwóch niezależnych klatek sygnału wizyjnego. W tym celu wykorzystano moduł kolejki FIFO dla protokołu AXI4-Stream oraz dedykowany element synchronizujący kanał odczytu z bufora VDMA z sygnałem rozpoczęcia nowej ramki obrazu strumienia wejściowego.

Założono, że algorytm będzie wykorzystywany w systemach wizyjnych czasu rzeczywistego, działających w architekturze potokowej. Aplikację przystosowano do działania z sygnałem wizyjnym o dowolnej rozdzielczości i próbkowaniu, składającym się z jednego lub wielu kanałów obrazu. Zastosowano kolejkę FIFO o długości 128 elementów oraz linie buforujące związane z modułem VDMA o tej samej długości.

W celu weryfikacji działania elementu wyznaczającego różnicę sekwencji obrazów zaprojektowano strukturę rozszerzoną o elementy umożliwiające komunikację przy użyciu protokołu AXI oraz przepływ sygnału wizyjnego. Zaprojektowano aplikację umożliwiającą konfigurację modułu w trybie *bare-metal* oraz przy współpracy systemu PetaLinux.

Sprawdzono działanie aplikacji dla sygnału wizyjnego o rozdzielczości 1280×720 pikseli i częstotliwości sześćdziesięciu ramek na sekundę. Szacowane zapotrzebowanie wynikowego systemu na energię elektryczną nie powinno przekroczyć $1,86W$. Właściwa energia wymagana do przeprowadzania operacji obliczeniowych nie przekracza wartości $1,723W$, w tym $1,559W$ (90%) to energia wymagana do obsługi układu ARM.

Tabela 4.1. Wykorzystanie zasobów przez aplikację wyznaczającą różnicę kolejnych ramek obrazu.

Rodzaj zasobu	Użycie	Dostępne	Procent użycia
FF	3059	17600	17,38%
LUT 6	5721	17600	32,51%
SLICE	2550	4400	57,95%
DSP 48	0	80	0%
BRAM	6	60	10%

W tabeli 4.1 przedstawiono zapotrzebowanie na zasoby FPGA układu Zynq. Moduł przeznaczony jest do pracy z częstotliwością $200MHz$, co pozwala na analizę sygnału wideo o rozdzielczości 1920×1080 pikseli i częstotliwości obrazu $60Hz$. Proces konfiguracji modułu VDMA wykorzystywanego do buforowania ramki obrazu przedstawiono w sekcji 5.3. Wynik działania aplikacji przedstawiono na rysunku 4.3. Projekty Vivado i PetaLinux związane z omawianą aplikacją dodano jako załączniki do pracy oraz udostępniono w repozytorium [38].



Rys. 4.3. Wynik działania aplikacji wyznaczającej różnicę dwóch kolejnych ramek obrazu.

4.2. Moduł generacji tła

Poza najprostszymi przypadkami analizy ruchu, informacja uzyskana w wyniku odejmowania ramek sekwencji wizyjnej nie jest wystarczająca do analizy strumienia obrazów. Zagadnienie to może być jednak elementem składowym bardziej rozbudowanych algorytmów, na przykład modułów realizujących algorytm generacji i modelowania tła, czy segmentacji obiektów pierwszoplanowych, gdzie pozwala określić zbiór obiektów będących w ruchu.

Generacja tła to zadanie ekstrakcji elementów *tła* badanego obrazu, a więc takich, które stanowią stały, niezmienny element sceny. Dzięki wydzieleniu obiektów tła, pozostałe elementy obrazu klasyfikowane są jako obiekty pierwszoplanowe. Zwykle, uważa się za nie elementy będące w ruchu. Uwzględnić jednak należy również obiekty, których ruch jest niejednostajny, na przykład zatrzymujący się piesi lub pojazdy na skrzyżowaniu.

Bardziej zaawansowane metody generacji tła uwzględniają ponadto dodatkowe warunki klasyfikacji obiektów do dwóch z omawianych grup:

- cienie – choć mogą być związane zarówno z elementami tła jak i pierwszoplanowymi, oczekiwane jest zwykle, by nie były uwzględniane w grupie obiektów wymagających analizy,
- ruchome elementy tła – występujące na przykład pod wpływem wiatru ruchy roślin czy deszcz nie powinny być traktowane jako obiekty pierwszoplanowe,

- obiekty o niejednostajnym ruchu – algorytm powinien klasyfikować poprawnie obiekty pierwszoplanowe, które pojawiają się na scenie a następnie zatrzymują, nie traktując ich jako elementy tła,
- obiekty wzajemnie przesłaniające się – elementy pierwszego planu mogą, w wyniku ruchu, zostać zasłonięte z perspektywy kamery przez elementy tła. Nie powinno to wpłynąć na zmianę klasyfikacji obiektów z obu grup.
- warunki oświetleniowe – możliwość zmiany warunków oświetleniowych może wymagać ciągłej korekty parametrów generowanego modelu tła. Uwzględnić należy zarówno zmiany długookresowe, wynikające na przykład z cyklu dobowego, jak i krótkookresowe, wynikające z nagłych zmian, takich jak włączenie lub wyłączenie sztucznego oświetlenia sceny.

Zagadnienie modelowania tła nie jest trywialne i wymaga metod uwzględniających część lub wszystkie z wymienionych powyżej ograniczeń. Opracowanie dostępnej literatury poruszającej ten temat znaleźć można w pracy [39]. W ramach niniejszego opracowania zdecydowano się na realizację generacji tła przy pomocy metody średniej bieżącej, opisanej zależnością (4.3). Wartość modelu tła wyznaczana jest niezależnie dla każdej składowej obrazu.

$$b^i(x, y) = \alpha p^i(x, y) + (1 - \alpha)b^{i-1}(x, y) \quad (4.3)$$

gdzie:

- x, y – współrzędne piksela,
- i – indeks ramki w sekwencji obrazów,
- $p^i(x, y)$ – wartość w i -tej ramce dla piksela o współrzędnych (x, y) ,
- $b^i(x, y)$ – wartość w i -tej ramce dla piksela modelu tła o współrzędnych (x, y) ,
- α – współczynnik bezwładności tła z przedziału $(0, 1]$.

Wadą przedstawionej metody jest jej wrażliwość na krótkookresowe zmiany oświetlenia. Jednym ze sposobów eliminacji zakłóceń pojawiających się cyklicznie – na przykład drgań liści pod wpływem wiatru – jest wykorzystanie wielu niezależnie wyznaczanych modeli tła. Stosując kilka modeli, budować można warianty dopasowane do najczęściej występujących przypadków, uporządkowanych według prawdopodobieństwa wystąpienia. Przy takim podejściu, obliczenia prowadzone są dla każdego modelu tła niezależnie, wartość nie jest jednak aktualizowana w sytuacji, gdy stan piksela nie jest zbliżony do oczekiwanego, związanego z wybranym modelem. Jednym z dostępnych rozwiązań jest algorytm GMM (*ang.* Gaussian Mixture Model), wykorzystujący kilka rozkładów prawdopodobieństwa Gaussa. W niniejszej pracy nie zdecydowano się na realizację opisanej powyżej metody eliminacji zakłóceń, uzasadniając to zachowaniem czytelności i prostoty implementacji.

Algorytm dostosowano do pracy z sygnałem opisanym w przestrzeni barw $YCbCr$. Procedura generacji tła odbywa się niezależnie dla każdej składowej sygnału. Przyjęto, że aktualizacja wartości modelu tła powinna mieć miejsce wyłącznie w przypadku, jeśli aktualnie badany piksel może być uznany za element tła. W tym celu wprowadzono dwa warunki wykonania obliczeń:

1. Warunek ruchu.

Aktualizacja powinna mieć miejsce wyłącznie w przypadku, jeśli wartość piksela nie uległa zmianie większej niż dopuszczalna względem poprzedniej ramki obrazu. W przeciwnym razie przyjąć można, że nastąpił ruch elementu i piksel nie należy do tła. Zależność opisano wzorem (4.4).

$$d_Y^i(x, y) > T_{fd} \quad (4.4)$$

gdzie:

$d_Y^i(x, y)$ – różnica ramek dla kanału Y obrazu, opisana wzorem (4.1),
 T_{fd} – próg ruchu z zakresu $[0, 255]$, zwykle nie przekraczający 30.

Większe wartości współczynnika T_{fd} pozwalają dokonać aktualizacji modelu tła dla elementów o coraz większej różnicy względem poprzedniej ramki obrazu.

2. Warunek tła.

Aktualizacja powinna mieć miejsce wyłącznie w przypadku, jeśli piksel został sklasyfikowany jako element tła na bazie aktualnego modelu. Zależność opisano równaniem (4.5).

$$w_Y m_Y^i(x, y) + w_{Cb} m_{Cb}^i(x, y) + w_{Cr} m_{Cr}^i(x, y) > T_{bg} \quad (4.5)$$

gdzie:

$m^i(x, y)$ – zmiana wartości piksela względem tła, opisana zależnością 4.6,
 w_k – współczynnik wagi związany z k -tym kanałem
 T_{bg} – współczynnik bezwładności przynależności do tła z zakresu $[0, 255]$.

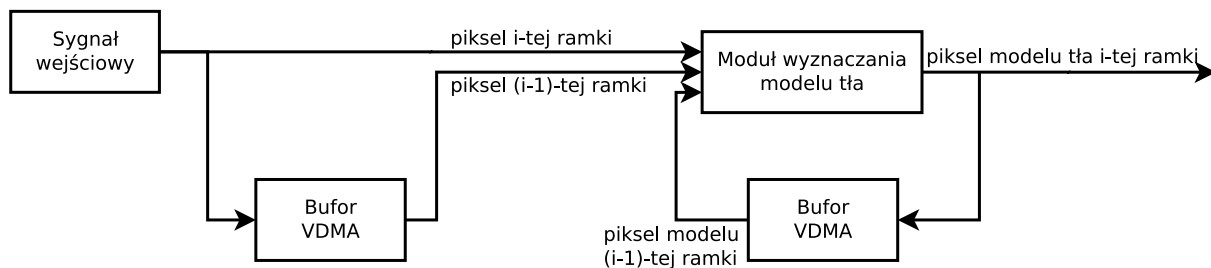
$$m_k^i(x, y) = |p_k^i(x, y) - b_k^{i-1}(x, y)| \quad (4.6)$$

gdzie:

k – indeks składowej barwnej sygnału.

Większe wartości parametru T_{bg} pozwalają na aktualizację modelu tła w sytuacji, gdy różnica piksela względem aktualnego modelu tła jest znaczna. Jego wartość nie przekracza jednak zwykle 30.

W trakcie eksperymentów przyjęto wartości współczynników wag przynależności do tła odpowiednio: $w_Y = 1$, $w_{Cb} = 2$, $w_{Cr} = 2$. Aktualizacja modelu tła powinna mieć miejsce wyłącznie w sytuacji, gdy spełnione są oba warunki przedstawione powyżej. Schemat blokowy algorytmu przedstawiono na rysunku 4.4.



Rys. 4.4. Schemat architektury wyliczającej model tła.

Algorytm wymaga wykorzystania dwóch buforów AXI VDMA. Jeden z nich przeznaczony jest do buforowania ramki obrazu wejściowego, natomiast drugi przechowuje aktualny model tła. Alternatywą jest zastosowanie wspólnego bufora i przechowywanie w nim dwóch scalonych sygnałów. Rozwiązanie to pozwala ograniczyć zapotrzebowanie na zasoby logiczne, może jednak wiązać się z trudnościami w synchronizacji wielu strumieni wizyjnych.

Algorytm zintegrowano z układem umożliwiającym komunikację z procesorem ARM, co pozwala na transmisję uzyskanego modelu tła i jego dalszą analizę. Wykorzystano w tym celu trzeci moduł AXI VDMA. W praktycznych zastosowaniach moduł ten może okazać się zbędny, ze względu na to, że w jednym z pozostałych modułów VDMA przechowywany jest model tła dla poprzedniej klatki obrazu. Opóźnienie jednego cyklu nie powinno wpłynąć negatywnie na jakość działania aplikacji. Niezależny moduł VDMA pozwala jednak na przesyłanie wyników również w przypadku, gdy algorytm generacji tła nie stanowi ostatniego etapu obliczeń.

Ze względu na duże zapotrzebowanie algorytmu na elementy obliczeniowe logiki reprogramowalnej, zdecydowano się ograniczyć rozmiar kolejek FIFO do 64 elementów. Sprawdzono działanie aplikacji dla sygnału wizyjnego o rozdzielczości 1280×720 pikseli i częstotliwości sześćdziesięciu ramek na sekundę. Szacowane zapotrzebowanie wynikowego systemu na energię elektryczną nie powinno przekroczyć $1,936W$. Właściwa energia wymagana do przeprowadzania operacji obliczeniowych nie przekracza wartości $1,797W$, w tym $1,565W$ (87%) to energia wymagana do obsługi układu ARM.

Tabela 4.2. Wykorzystanie zasobów przez aplikację.

Rodzaj zasobu	Użycie	Dostępne	Procent użycia
FF	6938	17600	39,42%
LUT 6	13670	17600	77,67%
SLICE	4400	4400	100%
DSP 48	15	80	18,75%
BRAM	12	60	20%

W tabeli 4.2 przedstawiono zapotrzebowanie na zasoby FPGA układu Zynq. Proporcjonalnie duże zużycie zasobów wynika z konieczności wykorzystania wielu modułów AXI VDMA oraz innych elementów wykorzystujących interfejs AXI. Może ono być ograniczone przez wykorzystanie jednego modułu do obsługi buforowania zarówno ramki obrazu, jak i modelu tła. W przypadku złożonych aplikacji, konieczne może okazać się użycie układu z większą liczbą dostępnych zasobów obliczeniowych. Moduł przeznaczony jest do pracy z częstotliwością $200MHz$, co pozwala na analizę sygnału wideo o rozdzielczości 1920×1080 pikseli i częstotliwości obrazu $60Hz$.

Wynik działania aplikacji przedstawiono na rysunku 4.5. Projekty Vivado i PetaLinux związane z omawianą aplikacją dodano jako załączniki do pracy oraz udostępniono w repozytorium [38]. Proces konfiguracji aplikacji przedstawiono w sekcji 5.7.



Rys. 4.5. Wynik działania aplikacji generacji tła obrazu.

4.3. Integracja z systemem PetaLinux

Na etapie prototypowania, elementy logiki reprogramowalnej kontrolowane były przez aplikację działającą w trybie *bare-metal*, bez wsparcia dla systemu operacyjnego. Po zakończeniu tego etapu, możliwe stało się zaprojektowanie aplikacji działającej pod kontrolą systemu PetaLinux, umożliwiającej wykorzystanie zaawansowanych funkcji systemu. Założono, że projektowana aplikacja powinna spełniać szereg wymagań:

- Konfiguracja modułów AXI i algorytmu.

Podstawowym zadaniem aplikacji powinno być przeprowadzenie wstępnej konfiguracji modułów, wykorzystując w tym celu interfejs AXI. Proces ten powinien mieć miejsce na etapie uruchamiania aplikacji. Aplikacja powinna być też odpowiedzialna za przeprowadzenie procesu konfiguracji parametrów wykonywanego algorytmu wizyjnego. Ponadto, działanie algorytmu nie powinno zostać przerwane w razie wyłączenia programu.

- Konfiguracja aplikacji przy użyciu argumentów wiersza poleceń.

Konfiguracja parametrów działania aplikacji, w tym rozmiar przetwarzanych obrazów i parametry algorytmu powinny być konfigurowane przy użyciu argumentów wiersza poleceń.

- Monitorowanie działania algorytmu.

Aplikacja powinna udostępniać opcję monitorowania stanu elementów algorytmu, ze szczególnym uwzględnieniem modułów AXI VDMA, odpowiedzialnych za buforowanie danych oraz komunikację z procesorem.

- Zapis wyników pracy algorytmu.

Program powinien być odpowiedzialny za zapis wyników działania algorytmu, na przykład w formie obrazów przechowywanych w pamięci.

- Wykorzystanie komunikacji sieciowej.

Uruchomienie aplikacji nie powinno wymagać fizycznego dostępu do układu Zynq. Docelowym narzędziem komunikacji jest protokół SSH. Ponadto, aplikacja powinna udostępniać interfejs wykorzystujący protokół HTTP, umożliwiający weryfikację stanu aplikacji przy użyciu przeglądarki internetowej.

- Kompatybilność z procedurami biblioteki OpenCV

Aplikacja powinna zapewniać zgodność z technikami programowania wykorzystywanymi przez bibliotekę OpenCV w stopniu umożliwiającym użycie algorytmów biblioteki ze strukturami danych wykorzystywanymi przez program.

Zaprojektowano aplikację w języku C, spełniającą przedstawione wymagania. Program odpowiedzialny jest za konfigurację elementów logiki reprogramowalnej na podstawie wartości przekazanych przy użyciu argumentów wiersza poleceń. Aplikacja wykonuje operacje monitorowania działania algorytmu i zapisu informacji logu do pliku. Ponadto, umożliwia cykliczny zapis obrazów będących wynikiem działania algorytmu do plików graficznych.

Proces obsługi aplikacji opiera się na wykorzystaniu protokołu SSH, program udostępnia również interfejs HTTP, umożliwiający uzyskanie aktualnych wyników działania algorytmu. Zbadano możliwość wykorzystania aplikacji w roli elementu obliczeniowego, odpowiedzialnego za przeprowadzenie części operacji algorytmicznych. Zaproponowano moduł indeksacji obiektów na bazie generowanego modelu tła. W tym celu wykorzystano procedurę `cv::connectedComponents` dostępną w bibliotece OpenCV.

Ze względu na ograniczenia sprzętowe, moduł indeksacji nie był w stanie spełnić wymagań pracy w czasie rzeczywistym dla sygnału wizyjnego o częstotliwości $60Hz$ i rozdzielczości 1280×720 pikseli – jego wydajność nie przekraczała piętnastu ramek na sekundę. Na rysunku 4.6 przedstawiono widok interfejsu www udostępnianego przez aplikację, przedstawiający aktualny stan i wyniki działania algorytmu. Związane z omawianą aplikacją dodano jako załączniki do pracy oraz udostępniono w repozytorium [38].



Rys. 4.6. Interfejs www aplikacji

Podsumowanie

Zaproponowane rozwiązania projektowe pozwalają na wykorzystanie części funkcji systemu operacyjnego, które badane były w ramach pracy. Szczególnie istotnym zagadnieniem jest komunikacja pomiędzy elementami zaprojektowanymi w dwóch architekturach. Dzięki wykorzystaniu transmisji danych, możliwe jest zaprojektowanie algorytmów podzielonych na moduły wykonywane naprzemiennie przez obie części układu, wykorzystując atuty obu architektur do możliwie maksymalnego zwiększenia wydajności pełnego algorytmu.

Ponadto, wykorzystanie systemu operacyjnego pozwala na realizację zadań zwykle niemożliwych w przypadku projektu aplikacji realizowanych wyłącznie przy użyciu elementów logiki reprogramowalnej lub sterowanych przez aplikację *bare-metal*. Program działający pod kontrolą systemu operacyjnego umożliwia prowadzenie zadań konfiguracji, kontroli i monitorowania działania aplikacji z wykorzystaniem komunikacji sieciowej.

Wykorzystanie systemu operacyjnego pozwala również na implementację aplikacji w dowolnym języku programowania. Dzięki temu, stosując dedykowane rozwiązania programistyczne, projektowanie aplikacji o rozbudowanych możliwościach zajmuje mniej czasu.

W trakcie realizacji projektu napotkano szereg ograniczeń.

- Liczba elementów obliczeniowych wykorzystywanego układu Zynq nie pozwala na realizację rozbudowanych rozwiązań algorytmicznych przy użyciu zaproponowanych technik. Ze względu na duże zapotrzebowanie na elementy logiki przez moduły AXI VDMA, buforowanie pełnych ramek obrazu jest kosztowne. W przypadku bardziej rozbudowanych algorytmów, konieczne może być wykorzystanie układu o większych możliwościach lub zastosowanie technik optymalizacji zużycia zasobów.
- Procesor ARM dostępny w układzie Zynq nie pozwala na realizację algorytmów wizyjnych o dużej złożoności obliczeniowej w czasie rzeczywistym. Próba wykorzystania rozwiązań biblioteki OpenCV do indeksacji obiektów pierwszoplanowych nie spełniała ograniczeń czasowych dla sygnału o częstotliwości $60Hz$.

W przypadku bardziej złożonych algorytmów, konieczne może być wykorzystanie układu o większej wydajności. Innym rozwiązaniem może być użycie protokołów sieciowych do transmisji danych do elementu obliczeniowego oferującego wydajność wystarczającą do realizacji zadań obliczeniowych. Dla zbioru algorytmów, których realizacja strumieniowa jest znana, możliwe jest również przeniesienie zadań obliczeniowych do elementów logiki reprogramowalnej. W przypadku, gdy żadne z zaproponowanych rozwiązań nie jest możliwe, konieczne jest ograniczenie częstotliwości działania algorytmu do poziomu, dla którego układ obliczeniowy będzie spełniać ograniczenia czasowe.

- Proces budowy systemu operacyjnego na bazie projektu sprzętowego jest złożony i wymaga dużych nakładów czasowych. Z tego powodu, na etapie projektowania połączeń logiki reprogramowalnej, wykorzystanie aplikacji typu *bare-metal* pozwala skrócić okres prototypowania.

W konsekwencji, konieczne może być zaprojektowanie dwóch aplikacji związanych z projektem: aplikacji *bare-metal*, wykorzystywanej na etapie prototypu oraz programu działającego pod obsługą systemu operacyjnego, projektowanego po ukończeniu implementacji sprzętowej.

Z tego powodu, aplikacje systemu operacyjnego nie pozwalają w pełni zastąpić programów *bare-metal* i powinny być traktowane jako metoda rozbudowy możliwości projektu.

5. Konfiguracja projektu

Użycie funkcjonalności opisywanych w niniejszej pracy wymaga przeprowadzenia konfiguracji wykorzystywanych modułów logicznych oraz systemu operacyjnego. W poniższych podrozdziałach zebrano informacje związane z poruszonymi zagadnieniami.

5.1. Podstawowa konfiguracja projektu

Wykorzystany w projekcie układ – Digilent ZYBO – nie jest bezpośrednio wspierany przez środowisko Vivado. Wynika z tego konieczność sprecyzowania parametrów układu na etapie tworzenia projektu. W trakcie modyfikowania projektu, w przypadku dodania modułów wykorzystujących interfejs wejścia/wyjścia, konieczna jest również konfiguracja parametrów interfejsu. Aby uprościć proces projektowania, zalecane jest skonfigurowanie obsługi układu przed utworzeniem projektu. Proces ten opisano w dokumentacji producenta [40].

5.1.1. Vivado

Utworzyć należy projektu typu „*RTL Project*”, z odznaczoną opcją „*Do not specify sources at this time*”.

W kroku „*Add Constraints*” dodać należy plik konfiguracyjny dla wybranego układu. W przypadku ZYBO, plik ten jest dostępny na stronie producenta. W kolejnym kroku możliwe jest skonfigurowanie parametrów układu. Wykorzystać do tego należy zakładkę „*Boards*” i wybrać wykorzystywany model.

Po utworzeniu projektu, skonfigurować należy właściwą przestrzeń roboczą dla projektu, wykorzystując do tego opcję „*IP Integrator → Create Block Design*”.

Do nowo utworzonej przestrzeni dodać należy moduł IP reprezentujący procesor ZYNQ – „*ZYNQ7 Processing System*”.

W kolejnych krokach należy dokonać konfiguracji modułu procesora, klikając dwukrotnie na moduł.

- Kanały interfejsu AXI mogą być konfigurowane przez zakładkę „*PS-PL Configuration*”. Możliwa jest aktywacja kanałów ogólnego przeznaczenia (*GP*) oraz wysokiej wydajności (*HP*).
- Interfejsy komunikacji konfigurowane mogą być z poziomu zakładki „*MIO Configuration*”. Zalecane jest aktywowanie interfejsów *ENET 0*, *SD 0* i *UART 1* ze względu na ich wykorzystanie na dalszym etapie pracy.

Przykład konfiguracji interfejsów wejścia/wyjścia przedstawiono na rysunku 5.1.

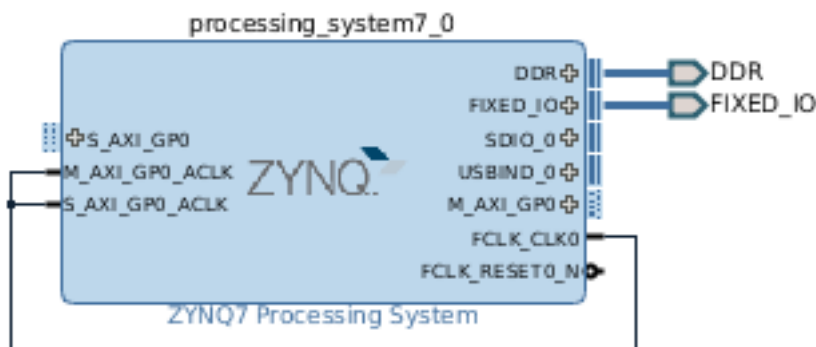
Peripheral	IO
I/O Peripherals	
<input checked="" type="checkbox"/> ENET 0	MIO 16 .. 27
<input type="checkbox"/> ENET 1	
<input checked="" type="checkbox"/> USB 0	MIO 28 .. 39
<input type="checkbox"/> USB 1	
<input checked="" type="checkbox"/> SD 0	MIO 40 .. 45
<input type="checkbox"/> SD 1	
<input type="checkbox"/> UART 0	
<input checked="" type="checkbox"/> UART 1	MIO 48 .. 49
<input checked="" type="checkbox"/> I2C 0	EMIO
<input type="checkbox"/> I2C 1	
<input type="checkbox"/> SPI 0	
<input type="checkbox"/> SPI 1	
<input type="checkbox"/> CAN 0	
<input type="checkbox"/> CAN 1	
GPIO	
<input checked="" type="checkbox"/> GPIO MIO	MIO
<input type="checkbox"/> EMIO GPIO (Width)	

Rys. 5.1. Okno konfiguracji interfejsów wejścia i wyjścia.

- Parametry sygnałów zegarowych dostępnych z poziomu układów logiki reprogramowalnej modyfikować można w zakładce „*Clock Configuration/PL Fabric Clocks*”.
- Częstotliwość pracy procesora oraz pamięci zmienić można w zakładce „*Clock Configuration/Processor/Memory Clocks*”.

Po ukończeniu etapu konfiguracji procesora i powrocie do głównego okna programu, należy użyć opcji „*Run Block Automation*”. Utworzone zostaną połączenia interfejsów pamięci DDR oraz FIXED_IO.

W przypadku zdefiniowania interfejsów AXI, połączyć należy właściwe sygnały zegarowe. Przykład wynikowej konfiguracji projektu przedstawiono na rysunku 5.2.



Rys. 5.2. Okno projektu.

Przedstawiona konfiguracja stanowi podstawę każdego projektu wykorzystującego moduł procesora Zynq.

Po zakończeniu konfiguracji, wygenerować należy warstwę HDL, korzystając z opcji „*Create HDL Wrapper*” dostępnej po kliknięciu prawym przyciskiem myszy na utworzony wcześniej plik źródłowy.

Skonfigurowany w ten sposób projekt może być budowany i uruchamiany na platformie Zybo.

5.1.2. SDK

W celu utworzenia projektu aplikacji w SDK, konieczne jest wyeksportowanie plików opisujących projekt z poziomu Vivado, wykorzystując do tego opcję „*File/Export/Export Hardware*” z zaznaczoną opcją „*Include bitstream*”.

W efekcie, dostępna powinna być aplikacja *nazwa_projektu_hw_platform_0*, zawierająca plik *nazwa_projektu.hdf*. Aplikacja ta stanowi podstawę każdego budowanego programu *bare-metal*.

W przypadku budowania aplikacji na platformę PetaLinux, na etapie tworzenia projektu, zmodyfikować należy pole „*OS Platform*” na wartość „*linux*”, „*Processor Type*” na „*ps7_cortexa9*” oraz wybrać właściwy język programowania.

W kontekście tak utworzonej aplikacji nie znajdują się biblioteki dostarczane przez firmę *Xilinx*, wykorzystywane w aplikacjach *bare-metal*, dostępna jest jednak pełna biblioteka języka *C* oraz *C++*.

W celu uruchomienia aplikacji systemowej na platformie ZYBO, przeprowadzić należy proces budowania i skopiować wynikowy plik z katalogu *Debug* lub *Release* do systemu plików systemu PetaLinux. Wykorzystać można do tego narzędzie SSH:

```
scp Debug/hello-world.elf root@adres-ip:~/
```

Aplikację uruchomić można przy użyciu konsoli użytkownika, również stosując narzędzie SSH.

5.1.3. PetaLinux

Utworzenie struktury katalogów projektu wykonywane jest przy użyciu poniższego polecenia.

```
petalinux-create -t project --template zynq --name nazwa-projektu  
cd nazwa-projektu
```

Powstała struktura zintegrowana jest z systemem kontroli wersji *git*, co pozwala utrzymać uporządkowanie danych wewnątrz projektu oraz wersjonowanie. Zarządzanie repozytorium wykorzystuje podstawowe komendy narzędzia.

Kolejnym krokiem jest zaimportowanie projektu *Vivado*.

```
petalinux-config --get-hw-description=/sciezka/do/projektu/projekt.sdk/
```

Jeśli polecenie wywołane zostało po raz pierwszy dla danego projektu, uruchomione zostanie narzędzie konfiguracyjne, domyślne ustawienia są jednak poprawne.

Konfiguracja projektu odbywa się przy użyciu polecenia `petalinux-config`. Skonfigurować należy metodę uruchamiania systemu – w omawianym przypadku, uruchomienie następuje na bazie plików znajdujących się na karcie SD.

```
petalinux-config
Image Packaging Configuration -> Root filesystem type -> SD card
```

Należy również zmodyfikować argumenty przekazywane systemowi na etapie uruchamiania, umożliwiając wykorzystanie sterowników do modułów logiki reprogramowalnej.

```
petalinux-config
Kernel Bootargs -> dezaktywować opcję Generate boot args automatically i zdefiniować
    własną wartość
console=ttyPS0,115200 earlyprintk uio_pdrv_genirq.of_id=generic-uio root=/dev/
    mmcblk0p2 rw rootwait
```

Następnie, przeprowadzić należy proces budowania systemu oraz wygenerować pliki wynikowe.

```
petalinux-build
petalinux-package --boot --fsbl images/linux/zynq_fsbl.elf --fpga images/linux/
    system_wrapper.bit --u-boot --force
petalinux-package --image -c rootfs --format initramfs
```

Uruchomienie systemu wymaga przygotowania karty SD – musi ona posiadać dwie partycje, pierwszą, z etykietą *boot* i systemem plików *fat32*, drugą – odpowiednio *sys* i *ext4*. Pierwsza z nich, zawierająca pliki wymagane na etapie inicjalizacji systemu, musi być poprzedzona 4 MB niezaalokowanej przestrzeni i mieć rozmiar co najmniej 40 MB. Druga partycja zawiera pliki systemowe, jej rozmiar powinien wynosić co najmniej kilkaset megabajtów. Proces formatowania przeprowadzić można przy użyciu narzędzia *gparted*.

Pliki wynikowe należy przenieść na kartę SD, korzystając z poleceń.

```
rm -rf /punkt-montowania/sys/*
cp images/linux/BOOT.BIN /punkt-montowania/boot/
cp images/linux/image.ub /punkt-montowania/boot/
cp images/linux/rootfs.cpio /punkt-montowania/sys/
cd /punkt-montowania/
pax -rvf rootfs.cpio
sync
cd -
```

Ze względu na mechanizm buforowania przez kontroler operacji zapisu danych, pamiętać należy o wywołaniu polecenia `sync`, zapewniającego zachowanie integralności danych.

Karta SD pozwala na uruchomienie systemu operacyjnego na układzie i przechowywanie danych użytkownika pomiędzy startami układu. Dalsza praca z systemem odbywać się może przez protokoły komunikacji *SSH* lub *UART*.

5.2. Konfiguracja modułu AXI DMA

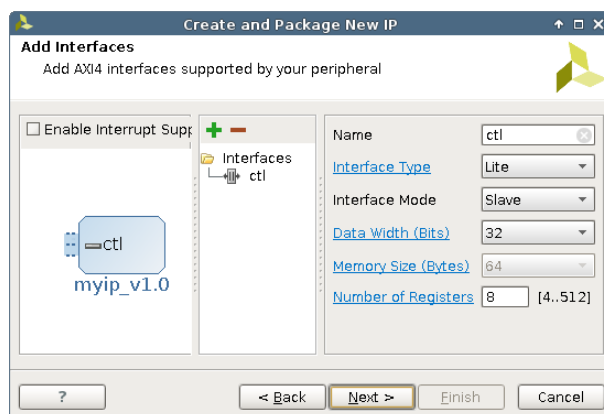
5.2.1. Vivado

Oprogramowanie Vivado umożliwia zbudowanie modułu wykorzystującego protokół AXI przez użycie opcji „*Create and package new IP...*”, zawartej w menu *Tools*.

Na ekranie wyboru zadania wybrać należy opcję „*Create a new AXI4 peripheral*”.

Po zdefiniowaniu podstawowych danych związanych z modulem, takich jak jego nazwa i nazwisko autora, w kolejnym kroku możliwe będzie zdefiniowanie interfejsu modułu. Na tym etapie konfiguracji skonfigurować należy wszystkie połączenia wykorzystujące interfejs AXI.

W przypadku modułu konfiguracyjnego o podstawowej strukturze, interfejs zawierać powinien jedno połączenie wykorzystujące protokół AXI w wersji *Lite*, działający w trybie *slave*, z oczekiwaną liczbą rejestrów. Każdy rejestr powinien być związany z jedną wartością, której konfiguracja ma być możliwa. Przykład konfiguracji przedstawiono na rysunku 5.3.



Rys. 5.3. Konfiguracja interfejsów modułu AXI DMA.

W omawianym przykładzie zdefiniowano interfejs AXI o nazwie *ctl*, związany z ośmioma rejestrami o długości trzydziestu dwóch bitów w pamięci.

Zdefiniowanie interfejsów kończy proces podstawowej konfiguracji modułu. W kolejnym kroku należy wybrać opcję „*Edit IP*” w celu dostosowania kodu źródłowego modułu.

Po wygenerowaniu, z modulem powinien być związany jeden plik źródłowy, zawierający instrukcje odpowiadające za obsługę komunikacji przy użyciu interfejsu AXI.

Do pliku dodać należy elementy odpowiedzialne za zdefiniowanie wyjść modułu oraz przypisanie im właściwych wartości.

W celu zdefiniowania wyjść modułu, odpowiadające im wpisy należy umieścić po komentarzu „// *Users to add ports here*”. Przykład przedstawiono na listingu 5.1.

Listing 5.1. Definicja interfejsów wyjściowych modułu.

```
// Users to add ports here
output wire parameter_a,
```

```
output wire [7:0] parameter_b,
output wire [15:0] parameter_c,
output wire [31:0] parameter_d,
// User ports ends
```

Zdefiniowano cztery sygnały wyjściowe, o różnej liczbie bitów.

Następnie, należy dokonać modyfikacji kodu odpowiedzialnego za powiązanie wartości parametrów z rejestrami modułu. Rejestry AXI zdefiniowane są poniżej linii „*//– Number of Slave Registers N*”, gdzie *N* to liczba dostępnych rejestrów. Rejestry te mają nazwy `slv_regn`, gdzie *n* to indeks rejestru – nie jest zalecana modyfikacja tych nazw.

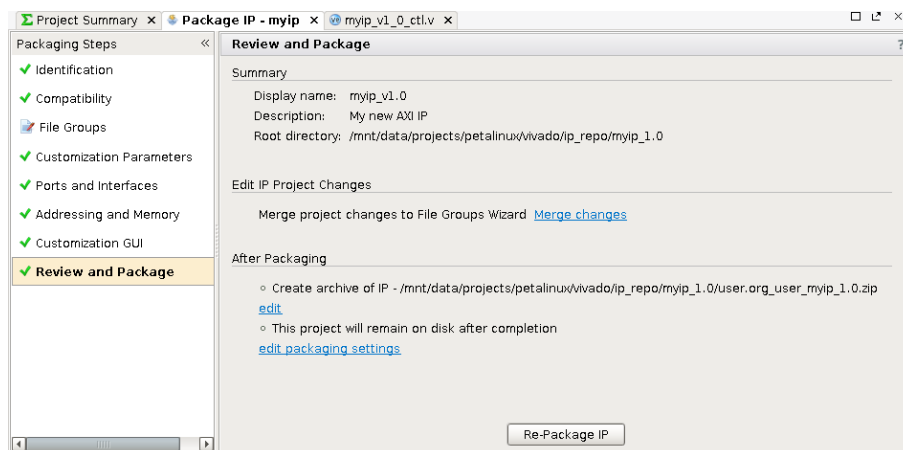
Modyfikacji kodu należy dokonać poniżej linii „*// Add user logic here*”. Przykład przedstawiono na listingu 5.2.

Listing 5.2. Powiązanie wyjść z rejestrami modułu.

```
// Add user logic here
assign parameter_a = slv_reg0[0];
assign parameter_b = slv_reg1[7:0];
assign parameter_c = slv_reg2[15:0];
assign parameter_d = slv_reg3[31:0];
// User logic ends
```

W przedstawionym przykładzie powiązano wartości parametrów bezpośrednio z danymi znajdującymi się w rejestrach. W rozbudowanych aplikacjach może być konieczne dodanie instrukcji modyfikujących wartości rejestrów przed przesłaniem ich na wyjście modułu.

Po ukończeniu modyfikacji modułu, konieczne jest zapisanie zmian i wygenerowanie plików wynikowych. W tym celu należy wykorzystać okno „*Package IP*”, sekcję „*Review and Package*”. Widok okna przedstawiono na rysunku 5.4.



Rys. 5.4. Okno finalizacji modyfikacji modułu.

Należy wybrać opcję „*Merge changes*”, umożliwiającą zintegrowanie wprowadzonych zmian z projektem bazowym. Następnie, można zakończyć edycję projektu przez wybór opcji „*Re-Package IP*”. Moduł będzie dostępny z poziomu interfejsu wyszukiwania modułów IP.

5.2.2. SDK

Konfiguracja wartości parametrów modułu opiera się na zapisie do właściwych sektorów pamięci. W przypadku pracy w trybie *bare-metal*, wykorzystać można instrukcję *Xil_Out32* z biblioteki *xil_io.h*. W przypadku pracy z systemem PetaLinux, wykorzystać należy biblioteki systemowe. Implementacje *bare-metal* przedstawiono na listingu 5.3, natomiast systemową na listingach 5.4, 5.5 i 5.6.

Listing 5.3. Obsługa modułu w trybie bare-metal.

```
#include "xparameters.h"
#include "platform.h"
#include "xil_io.h"

#define PARAMETER_A_REGISTER 0
#define PARAMETER_B_REGISTER 4
#define PARAMETER_C_REGISTER 8
#define PARAMETER_D_REGISTER 12

#define BASEADDR XPAR_ALGORITHM_PARAMETERS_0_CTL_BASEADDR

int main()
{
    init_platform();

    Xil_Out32(BASEADDR + PARAMETER_A_REGISTER, 1);
    Xil_Out32(BASEADDR + PARAMETER_B_REGISTER, 25);
    Xil_Out32(BASEADDR + PARAMETER_C_REGISTER, 1 << 10);
    Xil_Out32(BASEADDR + PARAMETER_D_REGISTER, 1 << 30);

    while(1);
}
```

Listing 5.4. Obsługa modułu w trybie systemowym - main.c.

```
#include <fcntl.h>
#include <stdio.h>
#include <stdlib.h>
#include <sys/mman.h>

#include "axi.h"

#define PARAMETER_A_REGISTER 0
#define PARAMETER_B_REGISTER 4
#define PARAMETER_C_REGISTER 8
#define PARAMETER_D_REGISTER 12

#define BASEADDR 0x43000000
```

```

typedef int memory_handle_t;

void setup_virtual_memory(struct axi_interface *interface, size_t length,
    memory_handle_t memory_handle, off_t base_addr) {
    interface->base_addr = base_addr;
    interface->virt_addr = (virt_address) mmap(NULL, length, PROT_READ |
        PROT_WRITE, MAP_SHARED, memory_handle, base_addr);
    if (interface->virt_addr == MAP_FAILED) {
        perror("Failed_to_map_virtual_memory.");
        exit(1);
    }
}

int main() {
    memory_handle_t memory_handle = open("/dev/mem", O_RDWR | O_SYNC);

    struct axi_interface* parameters = (struct axi_interface*) malloc(sizeof(
        struct axi_interface));
    if (parameters == NULL) {
        perror("Memory_allocation_failed.");
        exit(1);
    }
    setup_virtual_memory(parameters, 65535, memory_handle, BASEADDR);

    axi_write(parameters->virt_addr, PARAMETER_A_REGISTER, 1);
    axi_write(parameters->virt_addr, PARAMETER_B_REGISTER, 25);
    axi_write(parameters->virt_addr, PARAMETER_C_REGISTER, 1 << 10);
    axi_write(parameters->virt_addr, PARAMETER_D_REGISTER, 1 << 30);

    while(1);
}

```

Listing 5.5. Obsługa modułu w trybie systemowym - axi.h.

```

#include <stdbool.h>

typedef unsigned int* virt_address;

struct axi_interface {
    unsigned int base_addr;
    virt_address virt_addr;
};

void axi_write(virt_address virt_addr, int location, unsigned int value);

```

Listing 5.6. Obsługa modułu w trybie systemowym - axi.c.

```

#include "axi.h"

```



```
void axi_write(virt_address virt_addr, int location, unsigned int value) {  
    virt_addr[location >> 2] = value;  
}
```

5.2.3. PetaLinux

W celu wykorzystania techniki DMA w aplikacji działającej w systemie PetaLinux, konieczna jest aktywacja właściwych parametrów konfiguracji na etapie budowania systemu. W tym celu wykonać należy polecenie:

```
petalinux-config -c kernel
```

i aktywować funkcjonalność DMA:

```
Device Drivers -> DMA Engine Support
```

```
Device Drivers -> DMA Engine Support -> Xilinx AXI DMAS Engine
```

5.3. Konfiguracja modułu AXI VDMA

Proces konfiguracji modułu VDMA składa się z kroków podobnych do opisanych w sekcji poświęconej modułom AXI DMA.

5.3.1. Vivado

Do projektu dołączyć należy moduł *AXI Video Direct Memory Access*. Okno konfiguracji związane z nim pozwala na wybór obsługiwanych kanałów:

- write (*S2MM*) – kanał zapisu, pozwalający na transmisję danych z formatu strumieniowego do pamięci operacyjnej,
- read (*MM2S*) – kanał odczytu, umożliwiający konwersję danych przechowywanych w pamięci do strumienia.

Ustawienia pozwalają na wybór szerokości strumienia informacji dla jednego piksela, maksymalną liczbę buforów w pamięci oraz długość linii buforujących, związanych z oboma kanałami.

Wartości wielkości strumienia danych oraz liczby buforów związane są ściśle z projektowanym algorytmem, natomiast długość linii buforujących może wpłynąć na stabilność działania systemu. Zwiększenie tej wartości może poprawić działanie algorytmu w przypadku, gdy operacje związane z pamięcią operacyjną wykonywane są z opóźnieniem.

Zakładka ustawień zaawansowanych pozwala na zdefiniowanie parametrów związanych ze sterowaniem kanałami transmisji.

Wartość parametru „*Fsync Options*” w aplikacjach nie wymagających zewnętrznej sytuacji powinna być zdefiniowana jako `tuser` dla kanału zapisu oraz `none` dla kanału odczytu, dzięki czemu sygnał

synchronizacji modułu będzie związany z wejściowym strumieniem AXI. Część aplikacji może wymagać synchronizacji strumienia odczytu z drugim strumieniem danych, na przykład z inną ramką sygnału wizyjnego. W takiej sytuacji wykorzystać należy opcję synchronizacji `fsync`, a wejście układu `mm2s_fsync` połączyć z właściwym sygnałem synchronizacji.

W ramach pracy wykorzystywano również synchronizację pomiędzy kanałami przy użyciu parametru `GenLock`, o wartości `master` dla kanału zapisu i `slave` dla odczytu. Pozwalało to zachować przesunięcie o stałą, definiowaną z poziomu aplikacji, wartości pomiędzy buforami wykorzystywanymi przez oba kanały.

Ze względu na dużą wartość przepływu danych przez oba kanały, do komunikacji z procesorem wykorzystać należy połączenia o wysokiej wydajności. Kanały te można aktywować korzystając z opcji konfiguracyjnych modułu ZYNQ: „*PL-PS Configuration/HP Slave AXI Interface*”, i aktywując jeden lub wiele kanałów.

Z modułem AXI VDMA powiązać należy sygnał zegarowy o częstotliwości nie mniejszej od wartości zegara strumienia wejściowego. Sygnał ten powinien być generowany przez układ ZYNQ, a nie powiązany bezpośrednio z zegarem strumienia obrazu.

5.3.2. SDK

Konfiguracja modułu VDMA wymaga zastosowania technik opisanych w sekcji 5.2.2.

Proces uruchamiania transmisji dla modułu wymaga wykonania kroków zdefiniowanych przez producenta i opisanych w sekcji „*Programming Sequence*” dokumentacji [15].

5.3.3. Petalinux

Konfiguracja projektu zgodna jest z opisem dla modułów DMA, przedstawionym w sekcji 5.2.3.

W przypadku projektu aplikacji działającej pod kontrolą systemu operacyjnego, pamiętać należy, że konfiguracja buforów obrazu wymaga użycia adresów fizycznych, które mogą różnić się od adresów wirtualnych komórek pamięci.

Zdefiniować należy adresy buforów, odległe od siebie co najmniej o rozmiar jednej ramki sygnału wizyjnego. Zagwarantować należy nienaruszalność pamięci z perspektywy systemu operacyjnego. Efekt ten najprościej jest osiągnąć przez ograniczenie rozmiaru pamięci dostępnej dla systemu operacyjnego i zdefiniowanie adresów buforów poza tym zakresem. Wykorzystać do tego można argument `mem` przekazywany na etapie uruchamiania systemu, na przykład `mem=224M`. Proces dodawania argumentów uruchamiania systemu opisano w sekcji 5.1.3.

Adres fizyczny pamięci nie może być odczytany bezpośrednio, w tym celu musi zostać powiązany z adresem wirtualnym. Odpowiedzialną za to procedurę `setup_virtual_memory` przedstawiono na listingu 5.4, przy czym parametr `base_addr` to adres fizyczny pierwszej komórki bufora.

5.4. Obliczenia równoległe

Użycie rozwiązań omawianych w rozdziale 3.4 wymaga aktywacji właściwych funkcji kompilacji.

W przypadku zastosowania wątków natywnych lub biblioteki dostępnej w standardzie C++ wymagane jest zastosowanie przełączników:

```
g++ main.cpp -o main.out -pthread
g++ main.cpp -o main.out -std=c++11
```

Dla biblioteki TBB wymagane jest przeprowadzenie linkowania względem jej kodu źródłowego:

```
g++ main.cpp -o main.out -ltbb
```

Natomiast dla interfejsu OpenMP, konieczne jest użycie przełącznika:

```
g++ main.cpp -o main.out -fopenmp
```

5.5. Biblioteka OpenCV

5.5.1. OpenCV 2

Biblioteka OpenCV w wersji 2.4 nie jest oficjalnie dostępna w pakiecie PetaLinux, może jednak zostać dołączona do systemu operacyjnego dzięki mechanizmowi aplikacji użytkownika.

Twórcy biblioteki nie udostępniają prekompilowanych plików bibliotek na platformę ARM i konieczne jest procesu kompilacji kodu źródłowego wraz z zależnościami.

Poniżej przedstawiono proces instalacji zależności.

Listing 5.7. Definicja zmiennych środowiskowych.

```
export ARMPREFIX=ścieżka/instalacji
export CCPREFIX=arm-linux-gnueabihf-
```

Zmienna CCPREFIX wskazuje na prefiks kompilatora zawartego w pakiecie PetaLinux, a zmienna ARMPREFIX wskazuje na ścieżkę, gdzie zainstalowane zostaną pliki wynikowe.

Listing 5.8. Kompilacja biblioteki *xVideo*.

```
wget http://downloads.xvid.org/downloads/xvidcore-1.3.3.tar.gz
tar -zxvf xvidcore-1.3.3.tar.gz
cd xvidcore/build/generic/
./configure --prefix=${ARMPREFIX} --host=arm-linux-gnueabihf --disable-assembly
make
make install
```

Listing 5.9. Kompilacja biblioteki *x264*.

```
git clone git://git.videolan.org/x264
cd x264
```

```
./configure --enable-shared --host=arm-linux-gnueabi --disable-asm --prefix=${
    ARMPREFIX} --cross-prefix=${CCPREFIX}
make
make install
```

Listing 5.10. Kompilacja biblioteki *ffmpeg*.

```
git clone git://source.ffmpeg.org/ffmpeg.git
cd ffmpeg
git checkout release/2.6
./configure --enable-cross-compile --cross-prefix=${CCPREFIX} --target-os=linux \
    --arch=arm --enable-shared --disable-static --enable-gpl --enable-nonfree \
    --enable-ffmpeg --disable-ffplay --enable-ffserver --enable-swscale \
    --enable-pthreads --disable-yasm --disable-stripping --enable-libx264 \
    --disable-libxvid --prefix=${ARMPREFIX} --extra-cflags="-I${ARMPREFIX}"/
    include" \
    --extra-ldflags="-L${ARMPREFIX}"/lib"
make
make install
```

Po zainstalowaniu zależności, przystąpić można do pobrania i instalacji biblioteki OpenCV.

Listing 5.11. Pobieranie biblioteki OpenCV w wersji 2.4.10.

```
git clone https://github.com/Itseez/opencv.git
cd opencv
git checkout 2.4.10
```

Listing 5.12. Kompilacja biblioteki *OpenCV*.

```
mkdir build && cd build
cmake -DBUILD_DOCS=OFF -DBUILD_TESTS=OFF -DWITH_1394=OFF -DWITH_CUDA=OFF \
    -DWITH_CUFFT=OFF -DWITH_EIGEN=OFF -DWITH_GSTREAMER=OFF -DWITH_GTK=OFF \
    -DWITH_JASPER=OFF -DWITH_JPEG=OFF -DWITH_LIBV4L=OFF -DWITH_OPENEXR=OFF \
    -DWITH_PNG=OFF -DWITH_PVAPI=OFF -DWITH_TIFF=OFF -DWITH_V4L=OFF \
    -DENABLE_PRECOMPILED_HEADERS=OFF -DWITH_FFMPEG=ON \
    -DCMAKE_SYSTEM_NAME=Linux -DCMAKE_SYSTEM_PROCESSOR=arm \
    -DCMAKE_C_COMPILER=arm-linux-gnueabi-gcc \
    -DCMAKE_CXX_COMPILER=arm-linux-gnueabi-g++ \
    -DCMAKE_INSTALL_PREFIX=${ARMPREFIX} \
    -DCMAKE_FIND_ROOT_PATH=katalog/zawierający/narzędzia/kompilacji ../
make
make install
```

Aby zmniejszyć rozmiar biblioteki, a także skrócić proces instalacji, część modułów została dezaktywowana. Wartość zmiennej `CMAKE_FIND_ROOT_PATH` to ścieżka zawierająca strukturę katalogów wykorzystywanego kompilatora. W przypadku pakietu PetaLinux w wersji 2016.3, właściwa ścieżka względem punktu instalacji pakietu to `Xilinx/Petalinux/tools/linux-i386/gcc-arm-linux-gnueabi/arm-linux-gnueabi`.

Po zakończeniu procesu, pliki wynikowe znaleźć można w katalogu `$ARMPREFIX/lib`.

Pliki te mogą być dołączone do budowanego systemu operacyjnego jako dodatkowe pliki. W tym celu wykorzystać należy polecenie:

```
petalinux-create -t libs --template install --name opencv2
```

Utworzona zostanie struktura katalogów `components/libs/opencv2`, do której skopiować należy pliki wynikowe kompilacji biblioteki i jej zależności. Następnie, zmodyfikować należy plik `Makefile`, zgodnie z zawartymi w nim instrukcjami. W przypadku biblioteki OpenCV, wykorzystać można tekst generowany w wyniku wywołania polecenia:

```
for f in $(find . -type f -name "*.so*" -printf '%P\n'); \  
do echo -e '\t$(TARGETINST) -d' $f /lib/$f; done
```

Aktywacja biblioteki wewnątrz projektu wymaga wywołania polecenia przedstawionego na listingu 5.13 i wyboru biblioteki w zakładce „*Libs*”.

Listing 5.13. Dołączenie biblioteki do projektu PetaLinux.

```
petalinux-config -c rootfs
```

5.5.2. OpenCV 3

Biblioteka OpenCV w wersji 3.1 dołączona jest do pakietu PetaLinux. W celu jej aktywacji, wykorzystać należy polecenie przedstawione na listingu 5.13 i wybrać biblioteki w zakładce „*Filesystem Packages/libs/opencv*”.

5.5.3. SDK

Wykorzystanie bibliotek w projekcie SDK wymaga wskazania katalogu ze źródłami oraz bibliotekami w ustawieniach projektu.

W przypadku użycia biblioteki w wersji 3.1, wystarczające jest utworzenie aplikacji w języku C++ i typu *OpenCV Example Application*.

Dla wersji 2.4, konieczne jest ręczne zmodyfikowanie parametrów kompilacji projektu, w sposób analogiczny do konfiguracji aplikacji wykorzystującej OpenCV i działającej na platformie x86, wskazując jednak na skompilowane wcześniej pliki dla platformy ARM.

5.6. Wykorzystanie mechanizmu przerwań systemowych

Użycie mechanizmu przerwań systemowych wymaga zbudowania połączeń wewnątrz logiki programowalnej oraz konfiguracji agentów przerwań na poziomie aplikacji użytkownika. Poniżej opisano kroki wymagane do użycia omawianego mechanizmu w aplikacjach *bare-metal* oraz działających w systemie PetaLinux.

5.6.1. Vivado

Moduły wspierające mechanizm przerwań wyposażone są w dedykowane połączenia wyjściowe, wykorzystywane do transmisji sygnału przerwania. W przypadku modułu AXI Timer właściwe połączenie ma sygnaturę *interrupt*, natomiast w przypadku modułu AXI VDMA, sygnały przerwań dla kanałów odczytu i zapisu mają odpowiednio nazwy *mm2s_introut* oraz *s2mm_introut*.

Obsługa przerwań wymaga konfiguracji modułu procesora. Aktywować należy ścieżkę „*Fabric Interrupts* → *PL-PS Interrupt Ports* → *IRQ_F2P*” wewnątrz zakładki *Interrupts*. W rezultacie, dostępne będzie wejście procesora *IRQ_F2P* o szerokości do szesnastu linii. We wspomnianej zakładce ustawień aktywować można również inne połączenia przerwań, w tym szybkie przerywania w kierunku procesora oraz połączenia prowadzone od procesora do układów logiki, pozwalające na transmisję zdarzeń z interfejsów procesora, takich jak DMA, UART czy Ethernet.

Kanał *IRQ_F2P* pozwala na połączenie nie więcej niż szesnastu linii przerwań. W przypadku wykorzystania mechanizmu na platformie PetaLinux, pierwszym ośmiu liniom, zaczynając od najmłodszego bitu, przypisane będą identyfikatory przerwań w zakresie [61 – 68], natomiast pozostałym ośmiu – [84 – 91].

W przypadku konieczności zaprojektowania interfejsu wykorzystującego więcej niż szesnaście linii przerwań, konieczne jest zastosowanie układu dedykowanego obsłudze zdarzeń – „*AXI Interrupt Controller*”. Pozwala on na połączenie nie więcej niż trzydziestu dwóch linii przerwań do jednej linii na wejściu procesora i udostępnia interfejs umożliwiający identyfikację układu odpowiedzialnego za wysłanie sygnału przerwania. Zapewnia również mechanizmy priorytetyzacji i zagnieżdżania przerwań.

W sytuacji, gdy interfejs nie zawiera więcej niż szesnastu przerwań, wystarczające jest użycie modułu konkatenacji sygnałów zdarzeń do jednego wektora, którego wyjście połączone jest z wejściem *IRQ_F2P* procesora.

5.6.2. Aplikacja *bare-metal*

Wykorzystanie przerwań wymaga napisania procedury odpowiedzialnej za obsługę zdarzeń oraz zarejestrowanie jej jako agenta danego przerwania. Ponadto zwykle wymagane jest przeprowadzenie konfiguracji modułu w taki sposób, aby aktywować emisję przerwań.

Wymagane funkcje znaleźć można w plikach nagłówkowych *xparameters.h*, *xscugic.h*, *xil_exception.h*, oraz *xaxivdma.h* dla modułu AXI VDMA i *xtmrctr.h* dla AXI Timer.

Procedurę konfiguracji obsługi przerwań podzielić można na kilka etapów:

1. Zdefiniowanie agentów zdarzeń.

Konieczne jest zdefiniowane funkcji, które będą wywołane w przypadku wystąpienia przerwania. W najprostszym rozwiązaniu, ich celem jest akceptacja zdarzenia i przeprowadzenie konfiguracji modułu w taki sposób, aby umożliwić jego dalsze działanie – w przypadku modułu zegarowego jest to wykonanie restartu zegara. Moduł AXI VDMA nie wymaga żadnych kroków na etapie wywołania przerwania.

Ponadto, procedura jest odpowiedzialna za wykonanie obliczeń związanych z wystąpieniem przerwania.

Na listingu 5.14 przedstawiono funkcje agentów przerwań dla modułu zegara oraz obu kanałów AXI VDMA.

Listing 5.14. Procedury obsługi przerwań.

```
void Timer_InterruptHandler(void *data, u8 id) {
    // dodatkowe obliczenia

    // zerowanie przerwania
    XTmrCtr_Stop(data, id);
    XTmrCtr_Reset(data, id);
    XTmrCtr_Start(data, id);
}

void AxiRead_InterruptHandler(void *data, u32) {
    // dodatkowe obliczenia
}

void AxiWrite_InterruptHandler(void *data, u32) {
    // dodatkowe obliczenia
}
```

2. Konfiguracja modułów.

Oba omawiane moduły wymagają przeprowadzenia dodatkowych kroków konfiguracji. W przypadku modułu zegarowego, konieczne jest aktywacja obsługi przerwań w rejestrze kontrolnym – TCSR_n, natomiast w przypadku modułu VDMA, parametryzacja odbywa się przez rejestry MM2S_VDMACR dla kanału zapisu oraz S2MM_VDMACR dla kanału odczytu.

Ponadto, konieczna jest rejestracja agentów przerwań dla obu modułów. Proces ten przedstawiono na listingu 5.15, zmienne `TimerInstancePtr` i `AxiVdmaInstancePtr` są wskaźnikami do wykorzystywanych struktur typu `XTmrCtr` i `XAxiVdma`.

Listing 5.15. Rejestracja agentów przerwań.

```
XAxiVdma_SetCallBack(AxiVdmaInstancePtr, XAXIVDMA_HANDLER_GENERAL,
    &AxiWrite_InterruptHandler, AxiVdmaInstancePtr, XAXIVDMA_WRITE);

XAxiVdma_SetCallBack(AxiVdmaInstancePtr, XAXIVDMA_HANDLER_GENERAL,
    &AxiRead_InterruptHandler, AxiVdmaInstancePtr, XAXIVDMA_READ);

XTmrCtr_SetHandler(TimerInstancePtr, Timer_InterruptHandler, TimerInstancePtr);
```

3. Konfiguracja kontrolera przerwań.

W ostatnim kroku następuje konfiguracja kontrolera zdarzeń. Procedurę przedstawiono na listingu 5.16.

Listing 5.16. Konfiguracja kontrolera przerwań.

```
XScuGic InterruptController;
XScuGic_Config *GicConfig;
int ScuGicInterrupt_Init(ul6 DeviceId, XTmrCtr *TimerInstancePtr,
    XAxiVdma * AxiVdmaIntancePtr) {
    int Status;
    GicConfig = XScuGic_LookupConfig(DeviceId);
    if (NULL == GicConfig)
        return XST_FAILURE;

    Status = XScuGic_CfgInitialize(&InterruptController, GicConfig,
        GicConfig->CpuBaseAddress);
    if (Status != XST_SUCCESS)
        return XST_FAILURE;

    Xil_ExceptionRegisterHandler(XIL_EXCEPTION_ID_INT,
        (Xil_ExceptionHandler) XScuGic_InterruptHandler,
        &InterruptController);
    Xil_ExceptionEnable();

    Status = XScuGic_Connect(&InterruptController,
        XPAR_FABRIC_AXI_TIMER_0_INTERRUPT_INTR,
        (Xil_ExceptionHandler) XTmrCtr_InterruptHandler,
        TimerInstancePtr);
    if (Status != XST_SUCCESS)
        return XST_FAILURE;

    Status = XScuGic_Connect(&InterruptController,
        XPAR_FABRIC_AXI_VDMA_RESULT_S2MM_INTROUT_INTR,
        (Xil_ExceptionHandler) (XAxiVdma_WriteIntrHandler),
        AxiVdmaIntancePtr);
    if (Status != XST_SUCCESS)
        return XST_FAILURE;

    XScuGic_Enable(&InterruptController,
        XPAR_FABRIC_AXI_TIMER_0_INTERRUPT_INTR);

    XScuGic_Enable(&InterruptController,
        XPAR_FABRIC_AXI_VDMA_RESULT_S2MM_INTROUT_INTR);
    XScuGic_Enable(&InterruptController,
        XPAR_FABRIC_AXI_VDMA_RESULT_MM2S_INTROUT_INTR);
    return XST_SUCCESS;
}
```


Kolejne operacje wykonywane wewnątrz procedury odpowiadają za uruchomienie kontrolera przerwań, rejestrację w nim obu omawianych modułów AXI oraz uruchomienie trzech kanałów obsługi przerwań.

5.6.3. PetaLinux

Obsługa przerwań w systemie PetaLinux wymaga wykorzystania dedykowanych sterowników sprzętu i przeprowadzenia przy ich użyciu procesu konfiguracji. Pakiet PetaLinux udostępnia sterowniki do modułów AXI, które ich wymagają i w ramach niniejszej pracy ograniczono się do ich wykorzystania. W przypadku konieczności obsługi przerwania z niestandardowego modułu, konieczne może być dostarczenie dedykowanego mu sterownika, co wymaga szerokiej wiedzy z dziedziny działania systemów operacyjnych i komunikacji z urządzeniami peryferyjnymi.

Aby uzyskać dostęp do elementów logiki, konieczna jest aktywacja modułów systemowych.

Listing 5.17. Konfiguracja modułów systemowych.

```
petalinux-config -c kernel
```

```
Device Drivers -> Userspace I/O drivers
```

```
Device Drivers -> Userspace I/O drivers -> Userspace I/O platform driver with  
generic IRQ handling
```

```
Device Drivers -> Userspace I/O drivers -> Userspace I/O platform driver with  
generic iqr and dynamic memory
```

Konieczna jest również znajomość identyfikatorów linii przerwań. Można je odczytać z poziomu SDK, po utworzeniu projektu *Board Support Package* dla wykorzystywanej konfiguracji sprzętowej. Identyfikatory linii przerwań zdefiniowane są w pliku `xparameters.h`, na przykład:

```
/* Definitions for Fabric interrupts connected to ps7_scugic_0 */  
#define XPAR_FABRIC_AXI_VDMA_RESULT_MM2S_INTROUT_INTR 61  
#define XPAR_FABRIC_AXI_VDMA_RESULT_S2MM_INTROUT_INTR 62  
#define XPAR_FABRIC_AXI_TIMER_0_INTERRUPT_INTR 63
```

Wartości te mogą być również znalezione w strukturze *device tree*, generowanej przez pakiet PetaLinux na etapie parametryzacji, w której zdefiniowane są informacje o konfiguracji sprzętowej, wymagane do poprawnego działania systemu.

Wymagane informacje znajdują się w pliku `subsystems/linux/configs/device-tree/pl.dtsi`. Na listingu poniżej przedstawiono fragment konfiguracji związany z modułem AXI Timer.

```
axi_timer_0: timer@42800000 {  
    # ...  
    compatible = "xlnx,xps-timer-1.00.a";  
    interrupt-parent = <&intc>;  
    interrupts = <0 31 4>;  
    reg = <0x42800000 0x10000>;  
    # ...  
};
```

Kolejne wpisy w konfiguracji definiują informacje o sterowniku, który powinien być odpowiedzialny za obsługę modułu z poziomu procesora, module odpowiedzialnym za kontrolę przerw oraz definicję zdarzeń. Ostatni wpis zawiera informację o adresie urządzenia w pamięci oraz rozmiarze tego zasobu.

Definicja przerwania zawiera trzy elementy, z których kluczowa jest wartość 31. Ze względu na specyfikę formatu danych, w celu uzyskania właściwego identyfikatora przerwania, konieczne jest zwiększenie jej o 32. Uzyskany wynik – 63 – jest zgodny z definicją wewnątrz pliku `xparameters.h`.

W razie konieczności zaprojektowania dedykowanego sterownika sprzętu, wymagana jest wiedza na temat struktury *device tree* oraz zasad budowy oprogramowania tego typu. Informacje na ten temat znaleźć można we właściwych źródłach [37, 41].

Pakiet PetaLinux pozwala na dodanie do konfiguracji własnych modułów systemowych. W celu utworzenia struktury plików dla nowego modułu, wykorzystać można polecenie:

```
petalinux-create -t modules -n nazwa_modułu --enable
```

W wyniku działania polecenia utworzona zostanie struktura, którą następnie należy zmodyfikować dodając funkcjonalności sterownika.

Skompilowany na etapie budowania projektu moduł znajduje się w ścieżce `/lib/modules/identyfikator-kernela/extra` i może być uruchomiony poleceniem

```
insmod nazwa_modułu.ko
```

Logowane przez moduł wiadomości mogą być odczytane przy użyciu polecenia `dmesg`.

W celu weryfikacji poprawności konfiguracji przerw systemowych, wykorzystać można interfejs `/proc/interrupts`.

Wszystkie przerwy mogą być wypisane przy użyciu polecenia

```
cat /proc/interrupts
```

W przypadku modułu AXI Timer, spodziewany jest wpis o treści:

```
63:          1          0  axi-timer  40
```

Potwierdza on obecność linii przerwania o identyfikatorze 63, związanej ze sterownikiem `axi-timer`, która została wywołana jeden raz w przypadku pierwszego rdzenia procesora.

5.7. Konfiguracja projektu generacji tła

6. Podsumowanie

Celem niniejszej pracy była analiza możliwości systemu operacyjnego PetaLinux uruchomionego na układzie z rodziny Zynq i próba wykorzystania go w zagadnieniach związanych z wbudowanymi systemami wizyjnymi. Użycie systemu operacyjnego oferuje większy, względem aplikacji typu *bare-metal*, zasób możliwości – w tym komunikację sieciową, metody przechowywania danych, a także zaawansowane zadania algorytmiczne, dzięki zastosowaniu zewnętrznych bibliotek.

W rozdziale 2 autor przedstawił analizę układów rodziny Zynq. Zbadano możliwości wykorzystania istniejących systemów operacyjnych do działania na badanej platformie. Omówiono wady i zalety każdego ze znanych autorowi rozwiązań. Poruszono również zagadnienie realizacji obliczeń w czasie rzeczywistym w układzie Zynq.

W rozdziale 3 zebrano opis szeregu funkcjonalności, które, z perspektywy autora, znaleźć mogą zastosowanie w projektach realizujących zadania przetwarzania obrazów i sekwencji wizyjnych. Omówiono protokół AXI i wyjaśniono jego rolę w projektowaniu współczesnych układów obliczeniowych, realizowanych na platformie FPGA. W ramach rozdziału zebrano wiadomości teoretyczne, a także informacje praktyczne związane z implementacją badanych funkcjonalności.

Rozdział 4 zawiera wnioski z realizacji proponowanych projektów wizyjnych przy użyciu omawianych funkcjonalności. Autor zaproponował sposób realizacji podstawowego modułu wyznaczającego różnicę dwóch kolejnych obrazów należących do sekwencji wizyjnej. W dalszej części rozdziału, zaproponowano metodę wykorzystania modułu w systemach wizyjnych wymagających buforowania pełnych klatek obrazu na przykładzie modułu generacji tła. Ze względu na ograniczenia przedstawione w omawianym rozdziale, nie zrealizowano kompletnej aplikacji wizyjnej wykorzystującej proponowane techniki.

W rozdziale 5 zebrano informacje związane z praktycznym wykorzystaniem omawianych funkcjonalności w projektach wizyjnych. Przedstawiono metody konfiguracji kolejnych modułów w ujęciu ogólnym, nie związanym z konkretnym układem obliczeniowym.

Zdaniem autora, materiał zebrany i przedstawiony w ramach niniejszej pracy może posłużyć za podstawę do realizacji zaawansowanych algorytmów wizyjnych z wykorzystaniem platformy Zynq i systemu operacyjnego PetaLinux. Przedstawione techniki mogą zostać wykorzystane do rozwoju istniejących, jak i projektowania nowych aplikacji.

Zbiór przedstawionych przez autora w ramach pracy technik nie wyczerpuje możliwości badanego systemu operacyjnego. W ujęciu ogólnym, PetaLinux, lub inny system operacyjny działający na układzie z rodziny Zynq, pozwala na projektowanie aplikacji, które różnią się w znacznym stopniu od programów wykorzystywanych w systemach wbudowanych o ograniczonych możliwościach. Zastosowanie zaawansowanych technik projektowania aplikacji pozwala na uzyskanie efektu zbliżonego do programów stosowanych w życiu codziennym, na przykład dzięki wykorzystaniu protokołów komunikacji sieciowej.

Innym kierunkiem rozwoju może być zastosowanie systemów czasu rzeczywistego we współpracy z klasycznym systemem operacyjnym do realizacji obliczeń algorytmicznych z uwzględnieniem rygoru czasowego.

Dodatki

A. Aplikacja w architekturze NEON

Na listingach poniżej zaprezentowano implementacje procedury wyznaczającej wartość iloczynu skalarnego dwóch wektorów o zadanej długości. Listing 1 zawiera implementację bazową, wykorzystującą podstawowe operacje dostępne w języku C.

Listing 2 wykorzystuje funkcjonalności modułu NEON w celu potencjalnego zmaksymalizowania wydajności operacji. Obie implementacje porównano z procedurą napisaną w assemblerze, wykorzystującą instrukcje koprocesora VFP, przedstawioną na listingu 3.

Sposób wykorzystania procedur w programie przedstawiono na listingu 4. Kompilacja programu wymaga aktywacji przełączników odpowiedzialnych za obsługę instrukcji NEON. Wykorzystano również techniki optymalizacji udostępniane w kompilatorze *gcc*. Polecenie kompilacji przedstawiono poniżej:

```
arm-linux-gnueabi-gcc -Wall -O3 -mcpu=cortex-a9 -mfpu=neon -ftree-vectorize -
mvectorize-with-neon-quad -mfloat-abi=hard -ffast-math -funsafe-math-
optimizations -g -c -o "main.o" "main.c"
```

Listing 1. Implementacja bazowa.

```
float dot_product(float *first, float *second, unsigned int len) {
    float sum = 0.0;
    for (unsigned int i = 0; i < len; i++)
        sum += first[i] * second[i];
    return sum;
}
```

Listing 2. Implementacja w architekturze NEON. (Źródło: [42])

```
float dot_product_neon(float * restrict first, float * restrict second, unsigned int
len) {
    float32x4_t vec1_q, vec2_q;
    float32x4_t sum_q = {0.0, 0.0, 0.0, 0.0};
    float32x2_t tmp[2];
    float result;
    for( int i=0; i<( len & ~3); i+=4 )
    {
        vec1_q=vld1q_f32(&first[i]);
```

```

        vec2_q=vld1q_f32(&second[i]);
        sum_q = vmlaq_f32(sum_q, vec1_q, vec2_q );
    }
    tmp[0] = vget_high_f32(sum_q);
    tmp[1] = vget_low_f32 (sum_q);
    tmp[0] = vpadd_f32(tmp[0], tmp[1]);
    tmp[0] = vpadd_f32(tmp[0], tmp[0]);
    result = vget_lane_f32(tmp[0], 0);
return result;
}

```

Listing 3. Implementacja w asemblerze. (Źródło: [43])

```

float dot_product_asm(float * restrict first, float * restrict second, unsigned int
    len) {
    float net1D=0.0f;
    asm volatile (
        "vmov.f32_q8,_,#0.0"
        "1:"
        "subs_%,_,#,4"
        "vld1.f32_{d0,d1},_[%1]!"
        "vld1.f32_{d4,d5},_[%2]!"
        "vmla.f32_q8,_,q0,_,q2"
        "bgt_1b"
        "vpadd.f32_d0,_,d16,_,d17"
        "vadd.f32_%,_,s0,_,s1"
        : "=w"(net1D)
        : "r"(first), "r"(second), "r"(len)
        : "q0", "q2", "q8");
    return net1D;
}

```

Listing 4. Główna procedura programu.

```

int main() {
    float first[] = {1.1, 2.2, 3.3};
    float second[] = {-1.1, -2.2, -3.3};

    float result_base = dot_product(first, second, 3);
    float result_neon = dot_product_neon(first, second, 3);
    float result_asm = dot_product_asm(first, second, 3);

    return 0;
}

```

B. Konwersja danych pomiędzy VDMA i OpenCV

Transmisja sygnału wizyjnego pomiędzy elementami logiki programowalnej a procesorem CPU jest możliwa dzięki wykorzystaniu modułu AXI VDMA. W celu odczytu obrazu z bufora lub zapisu danych do niego wykorzystano bibliotekę OpenCV. W celu zwiększenia czytelności kodu, usunięto elementy związane z konfiguracją modułu VDMA. Założono, że pierwszy bajt bufora obrazu VDMA znajduje się pod adresem `framebuffer_ptr`. Procedura `from_vdma` jest odpowiedzialna za zbudowanie macierzy typu `cv::Mat` na bazie danych znajdujących się w buforze. Funkcja `to_vdma` pełni przeciwną rolę i umożliwia zapis danych z macierzy do bufora VDMA.

Domyślnie, procedura odczytu nie kopiuje danych z komórek pamięci związanych z modułem VDMA, co pozwala na uzyskanie możliwie największej wydajności. W przypadku, gdy macierz jest modyfikowana lub przechowywana przez okres dłuższy niż czas transmisji jednej ramki obrazu, konieczne jest uzyskanie kopii danych – odpowiedzialne jest za to polecenie `image = image.clone()`;

Na listingu 5 zaprezentowano metodę konwersji sygnału wizyjnego pomiędzy elementami obliczeniowymi wykonanymi w architekturach FPGA i ARM.

Listing 5. Konwersja sygnału wizyjnego pomiędzy AXI VDMA i `cv::Mat`.

```
#include "opencv2/core/core.hpp"

cv::Mat const from_vdma(unsigned char *ptr, std::size_t width, std::size_t height,
                        std::size_t bytes_per_pixel)
{
    return cv::Mat(height, width, CV_8UC(bytes_per_pixel), ptr);
}

void to_vdma(cv::Mat const &image, std::size_t bytes_per_pixel, unsigned char *ptr)
{
    assert(image.isContinuous());
    if (ptr != image.ptr())
        std::memcpy(ptr, image.ptr(),
                    image.rows * image.cols * bytes_per_pixel);
}

int main(int, char**)
{
    const std::size_t width = 1280, height = 720, bytes_per_pixel = 4;
    unsigned char framebuffer_ptr[width * height * bytes_per_pixel];

    for(int i = 0; i < 10000; i++)
    {
        cv::Mat image = from_vdma(framebuffer_ptr,
                                width, height, bytes_per_pixel);

        // opcjonalna kopia
```

```
        image = image.clone();

        algorithm(image);

        to_vdma(image, bytes_per_pixel, framebuffer_ptr);

        await_next_frame();
    }
    return 0;
}
```


Bibliografia

- [1] Ryszard Tadeusiewicz i Przemysław Korohoda. *Komputerowa analiza i przetwarzanie obrazów*. Wydawnictwo Fundacji Postępu Telekomunikacji, 1997.
- [2] Klaus Bengler i in. „Three Decades of Driver Assistance Systems: Review and Future Perspectives”. W: *IEEE Intelligent Transportation Systems Magazine* (2014).
- [3] Gorka Velez i Oihana Otaegui. „Embedding vision-based advanced driver assistance systems: a survey”. W: *IET Intelligent Transport Systems* (2017).
- [4] J Anil i L Padma Suresh. „Literature survey on face and face expression recognition”. W: *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on* (2016).
- [5] K Sriram i R Havaladar. „Human detection and tracking in video surveillance system”. W: *Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on* (2016).
- [6] Mudds Hussain i in. „Multi-target tracking identification system under multi-camera surveillance system”. W: *Progress in Informatics and Computing (PIC), 2016 International Conference on* (2016).
- [7] Huiwen Gou i in. „A novel approach for global abnormal event detection in multi-camera surveillance system”. W: *Information and Automation, 2015 IEEE International Conference on* (2015).
- [8] *Zynq AP SoC Architecture*. Dostęp: 2017.07.02. URL: https://reference.digilentinc.com/reference/programmable-logic/zybo/reference-manual#zynq_ap_soc_architecture.
- [9] *PetaLinux Tools*. Dostęp: 2017-07-04. URL: <https://www.xilinx.com/products/design-tools/embedded-software/petalinux-sdk.html>.
- [10] *Zynq-7000 All Programmable SoC*. Dostęp: 2017-07-02. URL: <https://www.xilinx.com/products/silicon-devices/soc/zynq-7000.html>.
- [11] *Zynq-7000 All Programmable SoC Data Sheet*. Dostęp: 2017-07-02. URL: https://www.xilinx.com/support/documentation/data_sheets/ds190-Zynq-7000-Overview.pdf.
- [12] Guanwen Zhong i in. „Design of Multiple-Target Tracking System on Heterogeneous System-on-Chip Devices”. W: *IEEE Transactions on Vehicular Technology* (2016).

- [13] Maleen Abeydeera i in. „4K Real-Time HEVC Decoder on an FPGA”. W: *IEEE Transactions on Circuits and Systems for Video Technology* (2016).
- [14] Paweł Dąbał i Ryszard Pełka. „Fast pipelined pseudo-random number generator in programmable SoC device”. W: *Signals and Electronic Systems (ICSES), 2014 International Conference on* (2014).
- [15] Xilinx. *AXI Video Direct Memory Access v6.2*. Dostęp: 2017-07-04. 2016. URL: https://www.xilinx.com/support/documentation/ip_documentation/axi_vdma/v6_2/pg020_axi_vdma.pdf.
- [16] *The SSH (Secure Shell) Remote Login Protocol*. Network Working Group.
- [17] *Open Source Computer Vision Library*. Dostęp: 2017-07-04. URL: <http://opencv.org/>.
- [18] *Vivado Design Suite*. Dostęp: 2017-07-04. URL: <https://www.xilinx.com/products/design-tools/vivado.html>.
- [19] *Xilinx Software Development Kit*. Dostęp: 2017-07-04. URL: <https://www.xilinx.com/products/design-tools/embedded-software/sdk.html>.
- [20] Jeremy Herbert. *Getting Started with the Linux Kernel and the Digilent Zybo/Xilinx Zynq*. Dostęp: 2017-08-11. URL: http://jeremyherbert.net/get/digilent_zybo_zynq_getting_started.
- [21] *Building stock (Xilinx) Linux For Zynq*. Dostęp: 2017-07-06. URL: <https://embeddedgreg.com/2014/04/08/step-4a-lets-build-stock-linux/>.
- [22] *Xilinx Wiki - Build Kernel*. Dostęp: 2017-07-06. URL: <http://www.wiki.xilinx.com/Build+Kernel>.
- [23] Peter Crosthwaite. <https://github.com/pcrost/ubuntu-core-zybo>. URL: <https://github.com/pcrost/ubuntu-core-zybo>.
- [24] *OPEN ASYMMETRIC MULTI PROCESSING (OpenAMP)*. Dostęp: 2017-07-06. URL: <http://www.multicore-association.org/workgroup/oamp.php>.
- [25] *FreeRTOS*. Dostęp: 2017-07-06. URL: <http://www.freertos.org/>.
- [26] *Adam Taylor's MicroZed Chronicles, Part 171: OpenAMP and PetaLinux Build*. Dostęp: 2017-07-06. URL: <https://forums.xilinx.com/t5/Xcell-Daily-Blog/Adam-Taylor-s-MicroZed-Chronicles-Part-171-OpenAMP-and-PetaLinux/ba-p/748583>.
- [27] *OpenAMP Framework for Zynq Devices*. Dostęp: 2017-07-06. URL: https://www.xilinx.com/support/documentation/sw_manuals/xilinx2017_1/ug1186-zynq-openamp-gsg.pdf.
- [28] *ARM NEON technology*. Dostęp: 2017-07-07. URL: <https://developer.arm.com/technologies/neon>.
- [29] Michael J. Flynn. „Some Computer Organizations and Their Effectiveness”. W: *IEEE TRANSACTIONS ON COMPUTERS* (1972).
- [30] *AXITM and ACETM Protocol Specification*. ARM.
- [31] Anthony Williams. *Język C++ i przetwarzanie współbieżne w akcji*. Helion, 2013.

- [32] James Reinders. *Intel Threading Building Blocks. Outfitting C++ for Multi-core Processor Parallelism*. O'Reilly Media, 2010.
- [33] *OpenMP Application Programming Interface*. Dostęp: 2017-07-20. OpenMP Architecture Review Board. 2016. URL: <http://www.openmp.org/wp-content/uploads/openmp-examples-4.5.0.pdf>.
- [34] *Choosing the right threading framework*. Dostęp: 2017-07-20. 2013. URL: <https://software.intel.com/en-us/articles/choosing-the-right-threading-framework>.
- [35] Philipp Kegel, Maraike Schellmann i Sergei Gorlatch. „Using OpenMP vs. Threading Building Blocks for Medical Imaging on Multi-cores”. W: *European Conference on Parallel Processing* (2009).
- [36] Robert Love. *Jądro Linuksa. Przewodnik programisty*. Helion, 2014.
- [37] Jonathan Corbet, Alessandro Rubini i Greg Kroah-Hartman. *Linux Device Drivers, Third Edition*. O'Reilly Media, 2005.
- [38] Wojciech Gumuła. *Repozytorium projektów Vivado i PetaLinux*. Dostęp: 2017-08-23. 2017. URL: <https://github.com/wgml/magister>.
- [39] Tomasz Kryjak. „Implementacja zaawansowanych algorytmów przetwarzania, analizy i szyfrowania obrazów w układach reprogramowalnych”. Prac. dokt. Akademia Górniczo-Hutnicza w Krakowie, 2012.
- [40] *Vivado Version 2015.1 and Later Board File Installation*. Dostęp: 2017-07-16. URL: <https://reference.digilentinc.com/reference/software/vivado/board-files>.
- [41] Xillibus. *A Tutorial on the Device Tree (Zynq)*. Dostęp: 2017-08-06. URL: <http://xillybus.com/tutorials/device-tree-zynq-1>.
- [42] Haoliang Qin. *Boost Software Performance on Zynq-7000 AP SoC with NEON*. Dostęp: 2017-07-07. Xilinx. URL: https://www.xilinx.com/support/documentation/application_notes/xapp1206-boost-sw-performance-zynq7soc-w-neon.pdf.
- [43] *Dot Product with Neon Intrinsics*. Dostęp: 2017-07-07. URL: <https://stackoverflow.com/a/17442498>.