



AGH

**AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W
KRAKOWIE**

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

KATEDRA AUTOMATYKI I INŻYNIERII BIOMEDYCZNEJ

Praca dyplomowa magisterska

*Sprzętowo-programowy system wizyjny do detekcji
obiektów z wykorzystaniem termowizji*

*Hardware-software vision system for object detection with
the use of thermovision.*

Autor:

Tomasz Kańka

Kierunek studiów:

Automatyka i Robotyka

Opiekun pracy:

dr inż. Tomasz Kryjak

Kraków, 2017

Oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i że nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.

*Serdecznie dziękuję ...tu ciąg dalszych
podziękowań np. dla promotora, żony, są-
siada itp.*

Spis treści

1. Wstęp	7
1.1. Cel pracy	8
1.2. Struktura pracy	8
2. Cyfrowy system wizyjny	9
2.1. Metody akwizycja obrazu i kalibracja	9
2.2. Model geometryczny	10
2.3. Algorytmy detekcji pieszych	11
2.3.1. Ustalenie regionu zainteresowań	11
2.3.2. Wyodrębnienie cech	11
2.3.3. Klasyfikator	12
2.4. Wykorzystanie FPGA w analizie obrazu	12
2.5. Podejście sprzętowo - programowe	13
2.6. Sprzęt	14
2.6.1. Kamera termowizyjna Lepton	14
2.6.2. Zynq-7000	15
2.7. Interfejs AXI	15
2.7.1. Wykorzystanie AXI-Stream do transmisji sygnału video.	17
3. Implementacja	19
3.1. Opis modułów	20
3.1.1. Transformata projekcyjna	20
3.1.2. Kontroler kamery IR	20
3.1.3. Interpolacja bilinearna	20
3.1.4. Łączenie strumieni	21
3.1.5. Koloryzacja i nakładanie	21

1. Wstęp

Cyfrowa analiza obrazów znalazła szerokie zastosowanie w wielu dziedzinach życia. Umożliwia automatyczne uzyskanie istotnych dla podmiotu informacji na podstawie obrazu bez konieczności angażowania człowieka. Niektóre informacje zawarte w obrazie nie są dobrze dostrzegane przez ludzką percepcję np. kolor jest bardzo subiektywnym parametrem dla różnych jednostek. Przez ostatnie kilkadziesiąt lat opracowano tysiące różnych technik i algorytmów wyspecjalizowanych do określonych zadań np. kontrola jakości i przebiegu procesu przemysłowego, kontrola dostępu poprzez rozpoznawanie twarzy w iPhone, optymalizacja ruchu na skrzyżowaniach, bezobsługowe systemy bezpieczeństwa i monitoringu, autonomiczne pojazdy, leśne fotopułapki do badania zachowań i migracji zwierząt itp. Dodatkowym narzędziem w analizie obrazu jest możliwość analizy sceny poza zakresem światła widzialnego ludzkim okiem. Kamery na podczerwień stają się coraz tańsze i coraz bardziej popularne. Dostarczają nam informacje o temperaturze obserwowanych obiektów i jest coraz chętniej wykorzystywane w wielu różnych dziedzinach np. weterynarii do określenia miejsc urazów zwierząt, kontroli jakości artykułów spożywczych, analiza strat cieplnych w budynkach, detekcji gazów, systemy wspomagania kierowcy[1].

Większość systemów wizyjnych służących do rozpoznawania przechodniów są oparte o analizę obrazów z zakresu światła widzialnego, bądź podczerwieni. W przypadku światła widzialnego można uzyskać bardzo dobre wyniki pod warunkiem że wyszukiwane obiekty są dobrze oświetlone i wyróżniają się swoim kolorem od tła. Podczerwień, a szczególnie termowizja, umożliwia detekcję w warunkach nocnych i ograniczonej widoczności. Oba podejścia mają swoje wady i zalety które wzajemnie się uzupełniają np. duże nasłonecznienie powoduje że tło termiczne staje się dużo wyższe co utrudnia wyodrębnienie pieszego, natomiast daje idealne warunki do uzyskania dobrej jakości obrazu w zakresie widzialnym [2]. Połączenie tych dwóch obrazów daje możliwość uzyskania jeszcze lepszych metod rozpoznawania ludzi. W pracy [3] autorzy nazywają ten rozszerzony format jako RGBT (“Red-Green-Blue-Thermal”), natomiast inna praca jako analizę wielospektralną (Multispectral) [4], albo po prostu jako połączony obraz z kamery termowizyjnej i zwykłej[2].

Skuteczna detekcja obiektów jest często okupiona dużym zapotrzebowaniem na zasoby obliczeniowe. W wielu przypadkach nie da się uzyskać satysfakcjonującej wydajności by można było uznać że system za działający w czasie rzeczywistym wykorzystując jedynie komputer klasy PC. Dają to pole do popisu dla układów rekonfigurowalnych które mają możliwość dużego zrównoleglenia obliczeń. Układy FPGA (ang. field-programmable gate array) znalazły już zastosowanie w wielu systemach wizyjnych wykonując różnego rodzaju niskopoziomowe operacje kontekstowe, zamiany przestrzeni barw czy też binaryzacji nawet w czasie jednego cyklu zegara. Dodatkową zaletą układów FPGA jest mały pobór mocy co czyni je niezwykle atrakcyjną dla mobilnych aplikacji takich jak drony czy czujniki środowiskowe [5].

Niniejsza praca jest kontynuacją pracy inżynierskiej autora.

1.1. Cel pracy

Celem pracy jest realizacja wbudowanego systemu wizyjnego do detekcji wybranych obiektów (np. ludzi) na podstawie obrazu z kamery termowizyjnej. Zakłada się, że jako platforma obliczeniowa zostanie użyty układ heterogeniczny (np. Zynq firmy Xilinx), który umożliwia realizację sprzętowo-programową algorytmów.

1.2. Struktura pracy

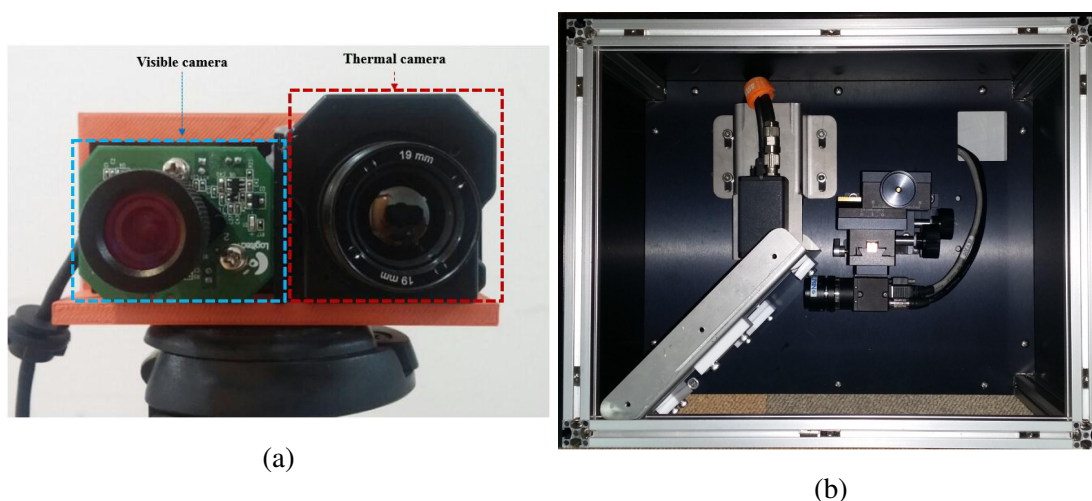
Struktura pracy

Jak wymyśle to będzie

2. Cyfrowy system wizyjny

2.1. Metody akwizycja obrazu i kalibracja

Większość implementacji wykorzystuje układ dwóch równoległych do siebie kamer. Do połączenia obrazów należy zastosować algorytm wyrównujący oba obrazy. Kalibrację wykonuje się specjalnymi planszami które pozwalają określić położenie punktów kalibracyjnych w obu rejestrowanych zakresach. Plansze mogą być aktywne (posiadają własne źródło ciepła) albo pasywne (przesłaniają obce źródło ciepła). W tym układzie występuję również zjawisko paralaksy które powiększa się wraz z wzrostem odległości obiektu od punktu kalibracji. W pracy [4] autorzy zastosowali zwierciadło półprzezroczyste wykonane z wafla krzemowego pokrytego cynkiem do rozdzielenia obrazu co wyeliminowało wady układu równoległego.



Rys. 2.1. Sposoby akwizycji obrazów: (a) dwie kamery równoległe [2], (b) z wykorzystaniem zwierciadła półprzezroczystego [4].

2.2. Model geometryczny

Do opisu matematycznego systemu wykorzystuje się model kamery otworowej. Dzięki niej można opisać relację między trójwymiarową przestrzenią a dwuwymiarowym obrazem za pomocą projekcji perspektywicznych. Nie stanowi on najdokładniejszego opisu matematycznego kamery, nie ma uwzględnionych w nim zakłóceń soczewkowych, ale jest wystarczające dobre dla niektórych zastosowań. Składa się ona z 2 zestawów parametrów: zewnętrznych oraz wewnętrznych. Parametry zewnętrzne definiują lokację kamery względem zewnętrznego układu współrzędnych. Są reprezentowane przez wektor translacji T między układem związanym z kamerą (X_c, Y_c, Z_c) a zewnętrznym (X, Y, Z) . Drugim parametrem jest macierz rotacji R (między osiami tych dwóch układów). Punkt $P = [X, Y, Z]^T$ będący w zewnętrznym układzie współrzędnym ma swój odpowiednik w układzie wewnętrznym który można określić zależnością

$$P_c = RP + T \quad (2.1)$$

Właściwości optyczne kamery można przedstawić w postaci macierzy kamery.

$$K = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$

gdzie:

f_x, f_y = ogniskowa kamery wyrażona w liczbie pikseli,

x_0, y_0 = współrzędne punktu głównego.

Macierz K określa związek między znormalizowanymi współrzędnymi w układzie odniesienia kamery, danych wzorem $x_n = \frac{X_c}{Z_c}, y_n = \frac{Y_c}{Z_c}$, a odpowiadającym im współrzędnymi punktów na obrazie u, v :

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} \quad (2.3)$$

W pracy [rangel20143d] autorzy sprawdzili kilka różnych plansz kalibracyjnych i najlepsza okazała się z kartonu... TO DO: dopisać.

2.3. Algorytmy detekcji pieszych

W cyfrowej analizie obrazu rozpoznawanie pieszych jest jedną z najbardziej aktywnych i rozwijanych dziedzin. W przeciągu kilkudziesięciu lat powstało ponad tysiąc artykułów poru-

szających to zagadnienie [6] i wiele różnych metod zostało już opracowanych. Większość metod opiera się o analizę obrazu tylko w jednym spektrum: widzialnym albo podczerwieni. Praca [4] pokazała że połączenie obu obrazów może dać lepsze wyniki. Podobnie w [7] ustalono że analiza multispektralna jest skuteczniejsza w dzień niż w nocy (o około 5% AMR (ang. average miss rate)). W artykule [8] autorzy podsumowują osiągnięcia w dziedzinie detekcji pieszych w latach 2004 – 2014 wyróżniono ponad 40 różnych podejść do problemu. Artykuł jest oparty o bazę danych Caltech-USA która oferuje obrazy w kolorze. Jednym z wniosków jest że przez ostatnie dziesięć lat największy postęp został osiągnięty głównie dzięki dopracowaniu cech jakie są wyodrębniane z obrazu niż ulepszanie klasyfikatora. Dodatkowo autorzy połączyli cechy dające najlepsze wyniki i stworzyli własną metodę która uzyska 12% zysk AMR względem najlepszej badanej wcześniej metody.

Dla typowego algorytmu detekcji pieszych można wyróżnić trzy podstawowe etapy:

2.3.1. Ustalenie regionu zainteresowań

Jest to obszar zwany ROI(ang. Region of interest) w którym potencjalnie mogą znajdować się przechodnie. Wiele podejść uznaje cały obraz jako ROI i stosuje okno przesuwne sprawdzając każdy możliwy fragment obrazu. Jeżeli obraz jest rejestrowany przez nieruchomą kamerę, ROI można określić poprzez różnicę między zapamiętanym tłem a aktualnym obrazem. Wyodrębnienie ROI jest bardzo istotne w przypadku pracy w czasie rzeczywistym ze względu na ograniczony czas analizy pojedynczego obrazu.

2.3.2. Wyodrębnienie cech

Do najbardziej popularnych cech można zaliczyć:

1. Histogramy zorientowanych gradientów (HOG) zaproponowany przez N.Dalala i B. Triggs w pracy [9] stała się jedną z najbardziej popularnych techniką w dziedzinie rozpoznawania ludzi. Jest cały czas rozwijana i modyfikowana w wielu pracach naukowych. Technika polega na zliczeniu kierunków gradientów, uzyskanych z 2 masek kierunkowych $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$ i $\begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T$, w komórkach o określonych wymiarach. Komórki te są organizowane w bloki w obrębie których następuje normalizacja. Wektorem cech jest połączony wszystkich histogramów z wszystkich bloków w jeden wektor.
2. Lokalne wzorce binarne LBP (ang. Local Binary Patterns).
3. Falki Haara.
4. Kolor. W analizie obrazu wykorzystuje różne przestrzenie barw np. RGB, HSV oraz LUV

5. Lokalna struktura. W odróżnieniu od pojedynczych pikseli można wyznaczyć lokalne struktury o podobnym kolorze. (np. głowa i ręce mają podobne kolory, jednolita koszula, spodnie)
6. i wiele innych

2.3.3. Klasyfikator

Otrzymany wektor cech jest poddany klasyfikacji której wynik decyduje czy obraz zawiera człowieka. W pracy [8] autorzy wyróżnili 3 dominujące rodziny:

1. Rodzina DPM (ang. Deformable Part Detectors) ??? wykrywacze deformowalnych elementów ???. Technika polega na klasyfikacji poszczególnych elementów człowieka (głowa, tułów, nogi). Następnie jest analizowany układ tych elementów na obrazie i podjęcie decyzji o obecności człowieka.
2. Deep networks – głębokie sieci neuronowe.
3. Decision forests – ?? lasy decyzyjne ?? zbiór nieskorelowanych drzew decyzyjnych.
4. inne: SVN (ang. support vector machine – maszyna wektorów nośnych), AdaBoost itp.

2.4. Wykorzystanie FPGA w analizie obrazu

Tradycyjne systemy wizyjne zwykle bazują na architekturze sekwencyjnej, po kolejnym przekształceniu obraz jest sukcesywnie poddawany następnym. W aplikacji procesorowej te operacje są wykonywane przez układ arytmetyczno-logiczny w który jest wyposażony. Kolejne kroki algorytmu są kompilowane w ciąg instrukcji dla procesora który oprócz operacji matematycznych dużą część pracy poświęca na pobieranie i dekodowanie rozkazów oraz na pobieranie i zapisywanie danych do pamięci. By taka aplikacja mogła pracować w czasie rzeczywistym cała procedura musi wykonać się szybciej przychodzące dane obrazu co wymusza wysoki taktowanie procesora sięgające GHz. W przypadku podejścia równoległego, implementacja poszczególnych kroków algorytmu odbywa się w osobnych procesach. Jeżeli kolejne kroki algorytmu wymagałyby danych otrzymanych z poprzednich to zysk takiego zabiegu byłby równy zero. By uzyskać znacznie przyspieszenie algorytm musi mieć możliwość podzielenia na wiele niezależnych części. Maksymalne do uzyskania przyspieszenie jest określone przez prawo Amdahla:

$$P_w = \frac{1}{s + \frac{1-s}{n_w}} \quad (2.4)$$

gdzie:

P_w = przyspieszenie algorytmu w systemie wieloprocesorowym,

s = część algorytmu niepodlegająca zrównolegleniu (wartość od zera do jeden),

n_w = liczba elementów obliczeniowych.

Teoretycznie jedynym ograniczaniem w możliwości zrównoleglenia obliczeń jest ilość zasobów dostępnych, jednak istotnym aspektem jest sposób dostarczania danych do zaimplementowanych w układzie procesorów. Czas i przepustowość jaka jest potrzebna do odczytania i zapisu obrazu po przetworzeniu z i do pamięci jest najczęściej wąskim gardłem systemu wizyjnego. Z tego powodu przetwarzanie obrazu bezpośrednio z sensora w czasie jego akwizycji jest chętnie wykorzystywane gdyż zmniejsza to ilość operacji odczytu i zapisu. [5]

2.5. Podejście sprzętowo - programowe

W pracy [10] autorzy wykorzystali układ FPGA oraz CPU małej mocy do skonstruowania systemu wizyjnego dla robotów. System analizował obraz stereoskopowy z dwóch kamer tworząc mapę głębi. Obie kamery są bezpośrednio podpięte do układu FPGA w którym obrazy są przetwarzane. Następnie dwa oryginalne obrazy oraz mapa głębi są przesyłane do CPU za pomocą specjalnej szyny danych. Moduł frame grabbera przechwytywał ten obraz i wykorzystując DMA (ang. Direct Memory Access) zapisywał do pamięci systemu. Ten zabieg gwarantował poprawną transmisję obrazu do CPU. Rozdzielczość oraz ilość klatek na sekundę są w pełni elastyczne dzięki czemu CPU dostawało obraz o szerokości trzy raz większej niż oryginalny obraz. Pozwalało to na przesłanie zsynchronizowanego lewego, prawego obrazu i mapy głębi. System pracował w rozdzielczości 752x480 piksele i 60 klatkach na sekundę. Całość systemu wizyjnego włącznie z kamerami, układem FPGA, CPU oraz konwerterami napięcia pobierał mniej niż 5W mocy. Całkowita latencja podana przez autorów rozwiązania wynosi około 2ms.

W pracy [11] autorzy wykorzystali układ SoC (ang. System on Chip) do detekcji pieszych dla zaawansowanego systemu wspomagania kierowcy (ADSA ang. advanced driver assistance system). Głównym wyzwaniem było opracowanie metody która działa w czasie rzeczywistym, ma mały pobór mocy oraz niski koszt wykonania. Większość topowych algorytmów wymaga znacznych zasobów obliczeniowych więc autorzy dokonali relaksacji problemu poprzez zastosowanie prostszego deskryptora jakim jest LBP oraz SVM jako klasyfikatora. Autorzy zamontowali po każdej stronie pojazdu inteligentną kamerę o 180° horyzontalnym kącie widzenia by jak najlepiej monitorować przestrzeń wokół niego. W kamerach została przeprowadzona wstępna obróbka obrazu (rektyfikacja i skalowanie). Przetworzony obraz z kamer był transmitowany do „Fusion-Box” gdzie odbywała się generacja kandydatów, klasyfikacja, weryfikacja oraz śledze-

nie. Wyniki były przesyłane do wbudowanego komputera PC. Rozwiązanie nie zostało jeszcze w pełni zaimplementowane ale pierwsze testy dawały obiecujące rezultaty.

2.6. Sprzęt

2.6.1. Kamera termowizyjna Lepton

Lepton jest zintegrowaną w pojedynczym układzie kamerą składającą się z soczewki, sensora podczerwieni fal długich (ang. LWIR – long wave infrared) oraz elektroniki sterującej i przetwarzającej sygnał. Układ ma możliwość domontowania dodatkowej przesłony która jest wykorzystywana do automatycznej optymalizacji procesu ujednolicania obrazu (kalibracji sensora). Prosty do integracji z dowolnym mikrokontrolerem dzięki zastosowaniu standardowych protokołów i interfejsów. Lepton po podłączeniu od razu pracuje w domyślnym trybie pracy, który może zostać zmieniony za pomocą CCI (ang. camera control interface – interfejs kontroli kamery).[lepton] Parametry:

- Wymiary: 11,8 x 12,7 x 7,2 mm,
- Sensor: niechłodzony mikrobolometr VOx (tlenek wanadu),
- Rejestrowany zakres: fale długie podczerwieni, $8\mu\text{m}$ do $14\mu\text{m}$,
- Wielkość piksela: $17\mu\text{m}$,
- Rozdzielczość: 80x60 pikseli,
- Ilość klatek na sekundę 8,6,
- Zakres rejestrowanych temperatur: -10°C do 140°C (Tryb wysokiego wzmocnienia),
- korekta niejednorodności matrycy: automatyczna na bazie przepływu optycznego
- kąt widzenia horyzontalny / diagonalny: 51° / 66° ,
- Głębina ostrości: od 10cm do nieskończoności
- Format wyjściowy: do wyboru: 14-bit, 8-bit (z AGC (ang. automatic gain control – automatyczna kontrola wzmacnienia)) 24-bit rgb (z ACG i koloryzacją).
- Interfejs video: VoSPI (Video over Serial Peripheral Interface)
- Interfejs sterujący: CCI (I2C podobny)

2.6.2. Zynq-7000

Rodzina układów Zynq-7000 bazuje na architekturze SoC (ang. System on Chip). Posiadają zintegrowany kompletny system składający podzielonego na dwie części: systemu procesorowego bazującego na procesorze ARM Cortex-A9 (PS ang. Porcessing System) oraz logikę programowalną (PL ang. programable logic) FPGA w jednym układzie scalonym. Na rysunku 2.2 przedstawiono schemat architektury. Prócz procesora część procesorowa posiada wbudowaną pamięć, kontroler pamięci zewnętrzne oraz szereg interfejsów dla układów peryferyjnych takich jak USB, GigEthernet, CAN, I2C, SPI. W części logiki programowalnej znajdują się bloki logiki konfigurowalnej (CLB ang. configurable logic block), 36Kb bloki pamięci RAM, procesory sygnałowe DSP48, układ JTAG, układy zarządzania zegarami oraz dwa 12-bitowe przetwornik analogowo-cyfrowy.

Komunikacji między częścią procesorową a logiką programowalną odbywa się za pośrednictwem Interfejsu AXI (ang. Advanced Extensible Interface), oraz bezpośrednio wykorzystując porty generalnego przeznaczenia, przerwania, oraz poprzez bezpośredni dostęp do pamięci (DMA ang. Direct Memory Access)

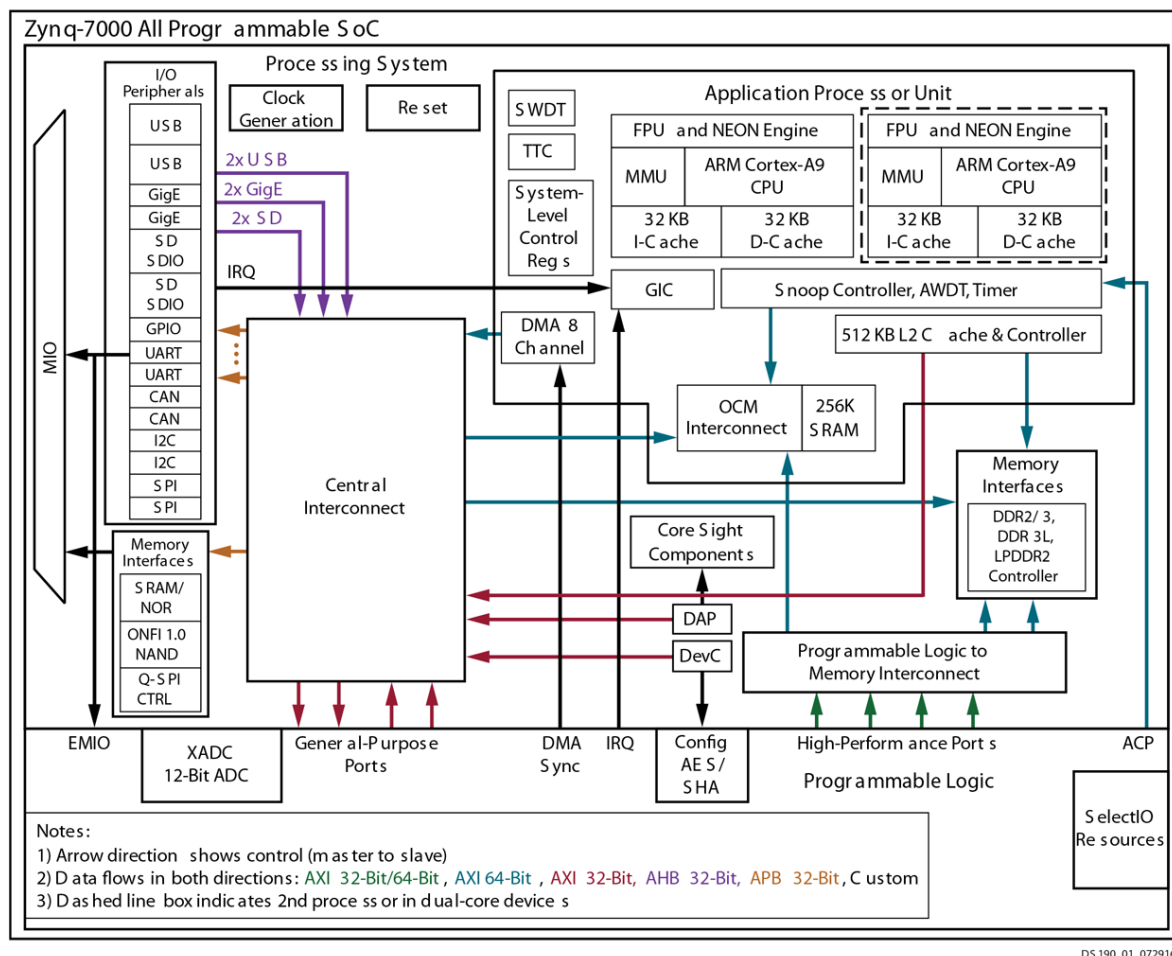
2.7. Interfejs AXI

AXI (ang. Advanced eXtensible Interface zawansowany rozszerzalny interfejs) jest częścią ARM AMBA (ang. Advanced Microcontroller Bus Architecture) – otwartego standardu, specyfikacją do zarządzania i połączeń między blokami funkcyjnymi w SoC. Aktualnie jest stosowana AMBA 4.0 która wprowadziła drugą wersję AXI, AXI4. Występują trzy typy interfejsów dla AXI4:

- AXI4 – stosowany w wysokowydajnych transferach w przestrzeni pamięci (ang. memory-mapped)
- AXI4-Ligt – stosowany dla prostszych operacji w przestrzeni pamięci (na przykład do komunikacji z rejestrami kontrolnymi i statusu)
- AXI4-Stream – stosowany do wysokiej prędkości transmisji strumieniowych

Specyfikacja interfejsu zakłada komunikację pomiędzy pojedynczym AXI master i pojedynczym AXI slave, która ma na celu wymianę informacji pomiędzy tymi dwoma blokami funkcyjnymi IP core. Kilkanaście interfejsów AXI master i slave mogą zostać połączone między sobą za pomocą specjalnej struktury zwanej interconnect block (blok międzypołączeniowy) w której odbywa się trasowanie połączeń do poszczególnych bloków.

AXI4 i AXI4-Lite składają się z 5 różnych kanałów:



DS 190_01_072916

Rys. 2.2. Schemat ogólny architektury układu Zynq-7000.

- Kanał adresu odczytu,
- Kanał adresu zapisu,
- Kanał danych odczytanych
- Kanał danych do zapisania
- Kanał potwierdzenia zapisu

Dane mogą płynąć w obie strony pomiędzy master a slave jednocześnie. Ilość danych które można przesłać w jednej transakcji w przypadku AXI4 wynosi 256 transferów, zaś AXI4-Lite pozwala na tylko 1 transmisję.

AXI4-Stream nie posiada pola adresowego, a dane mogą być przesyłane nieprzerwanie.

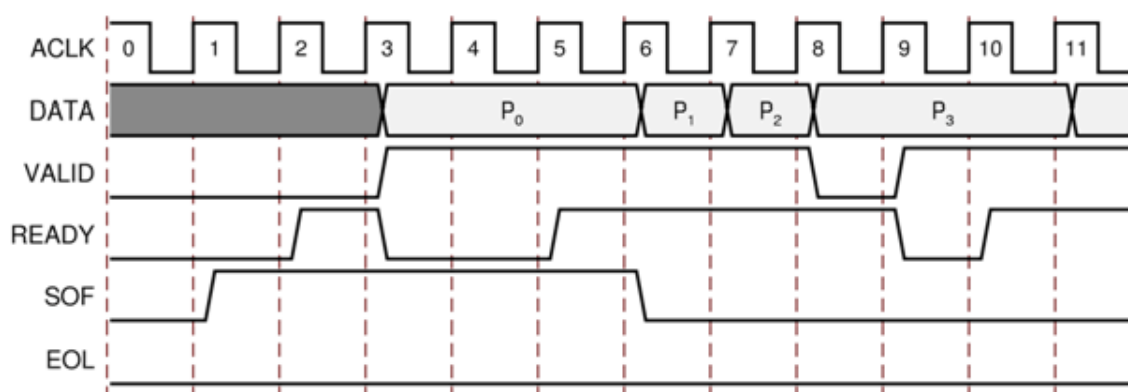
2.7.1. Wykorzystanie AXI-Stream do transmisji sygnału video.

W odróżnieniu od klasycznej implementacji przetwarzania strumieniowego video, w AXI-Stream przesyłane są jedynie aktywne piksele. Linie synchronizacji poziomej i pionowej są odrzucane albo są połączone do specjalnego bloku detekcji timingów który mierzy parametry wchodzącego strumienia wizyjnego (ilość pikseli na linii, czas ilość aktywnych linii, czas wyciemnienia itd.). Podobnie informacje o synchronizacji są dodawane przez blok generujący timingi.

Do transmisji wykorzystane jest 6 linii: jedna linia danych i pięć kontrolno-sterujących.

- Video Data – linia danych o szerokości jednego (albo dwóch) pikseli. Szerokość tej linii powinna być wielokrotnością liczby osim (16, 24, 48 itd.)
- Valid – Linia podająca czy dane piksela są poprawne,
- Ready – Linia kontrolna informująca urządzenie master że slave jest gotowy do transmisji danych,
- Start Of Frame – linia która wskazuje pierwszy piksel nowej ramki,
- End Of Line – linia wskazująca ostatni piksel w linii.

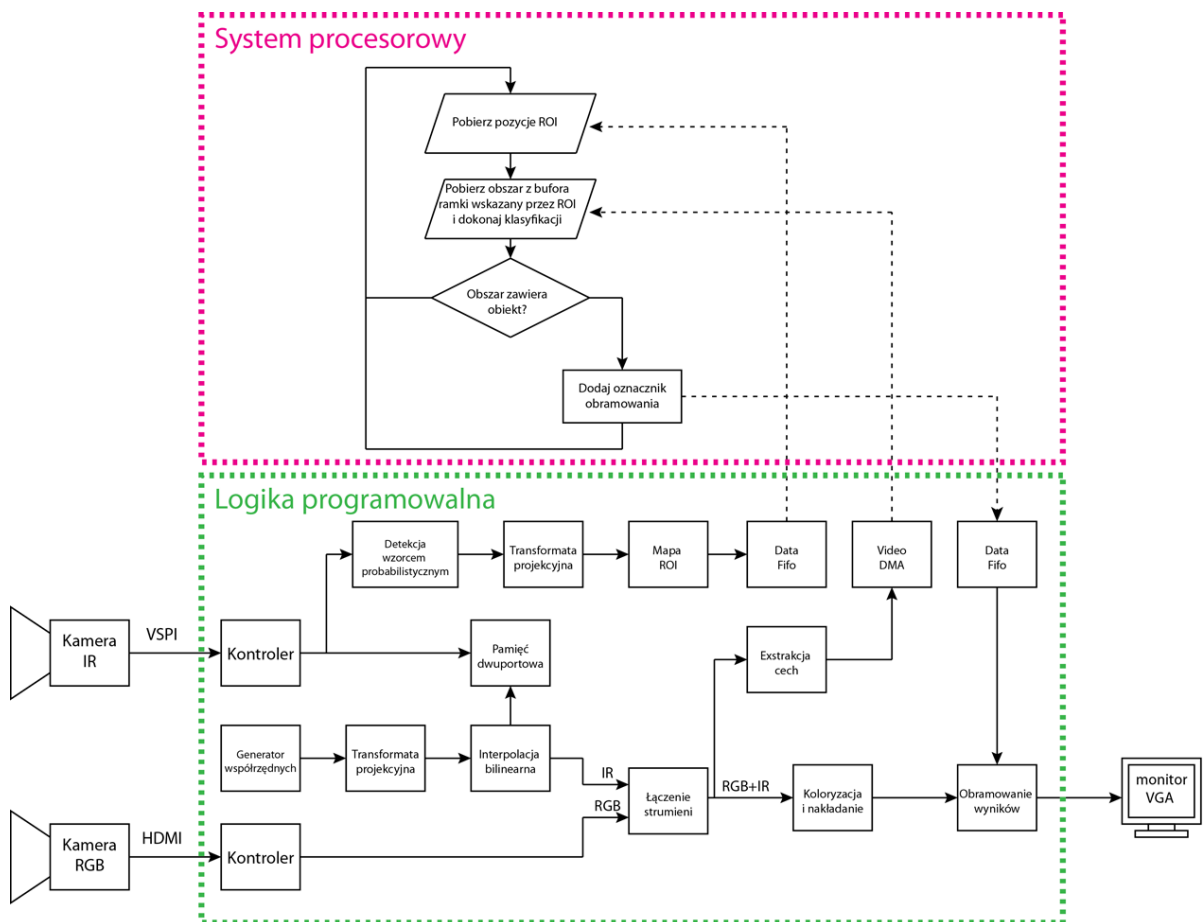
Aby mógł wystąpić poprawny transfer danych linie Valid i Ready muszą być w stanie wysokim podczas rosnącego zbocza zegara. Przykładowe nawiązanie transmisji przedstawia rysunek 2.3



Rys. 2.3. Przykład rozpoczęcia transmisji Reday/Valid.

3. Implementacja

Mając do dyspozycji układ heterogeniczny rodziny Zynq-7000 od firmy Xilinx operacje zostały podzielone między programowalną logiką a systemem procesorowym. Ogólny zarys systemu został przedstawiony na rysunku 3.1.



Rys. 3.1. Schemat blokowy systemu detekcji.

Programowalna logika:

- Akwizycja Obrazu poprzez HDMI (RGB) i VoSPI (IR),

- Transformata projekcyjna i interpolacja obrazu IR,
- Nałożenie i synchronizacja obrazu IR do obrazu RGB,
- Ekstrakcja cech z połączonych obrazów,
- Prezentacja wyników,
- Detekcja kandydatów za pomocą wzorca probabilistycznego.

System Procesorowy:

- konfiguracja parametrów systemu wizyjnego w logice programowalnej poprzez interfejs AXI-Lite,
- Klasyfikacja obszarów wytypowanych przez wzorzec probabilistyczny,
- Generowanie oznaczników.

3.1. Opis modułów

3.1.1. Transformata projekcyjna

Zadaniem modułu jest dostosowanie obrazu IR by pokrywał się z obrazem RGB. Moduł transformaty projekcyjnej zamienia wygenerowane współrzędne w zakresie wielkości przycho-
dzącego obrazu RGB na odpowiadające im punkt na obrazie IR (wraz z częścią ułamkową).
Moduł jest konfigurowalny poprzez interfejs AXI-Lite, za pomocą którego można ustawić war-
tość minimalną i maksymalną współrzędnych wyjściowych U i V oraz macierz transformaty.

3.1.2. Kontroler kamery IR

Pobiera obraz z kamery poprzez interfejs VoSPI który następnie zostaje zapisany do dwu-
portowej pamięci BRAM.

3.1.3. Interpolacja bilinearna

Prosty moduł przeznaczony głównie do powiększania obrazów. Pobiera wartość 4 otacza-
jących, podanych na wejściu punktu, pikseli z BRAM i na ich bazie jest wykonywana inter-
polacja. Moduł zapamiętuje 4 ostatnio użyte piksele które są na bieżąco aktualizowane wraz z
zmianą położenia punktu wejściowego na obrazie IR.

3.1.4. Łączenie strumieni

Moduł posiada dwa wejścia dla obrazu. Jeden strumień jest głównym i do niego jest dołączany drugi strumień. Do synchronizacji strumieni została wykorzystana możliwość AXI-Strem do wstrzymania transmisji. Piksele z dołączanego strumienia są odrzucane do momentu pojawienia się sygnału SOF. W momencie pojawienia się sygnału SOF w strumieniu głównym transmisja zostaje wznowiona pod kontrolą strumienia wyjściowego.

3.1.5. Koloryzacja i nakładanie

Połączone strumienie RGB+IR zostają połączone w jeden obraz. Obraz IR zostaje poddany koloryzacji na podstawie 12-bitowego LUT i nałożony w proporcjach 50 na 50 z obrazem RGB.

Bibliografia

- [1] Rikke Gade i Thomas B Moeslund. „Thermal cameras and applications: A survey”. W: *Machine vision and applications* 25.1 (2014), s. 245–262.
- [2] Ji Hoon Lee i in. „Robust pedestrian detection by combining visible and thermal infrared cameras”. W: *Sensors* 15.5 (2015), s. 10580–10615.
- [3] Louis St-Laurent, Xavier Maldague i Donald Prévost. „Combination of colour and thermal sensors for enhanced object detection”. W: *Information Fusion, 2007 10th International Conference on*. IEEE. 2007, s. 1–8.
- [4] Soonmin Hwang i in. „Multispectral pedestrian detection: Benchmark dataset and baseline”. W: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, s. 1037–1045.
- [5] Gabriel J Garcia i in. „A survey on FPGA-based sensor systems: towards intelligent and reconfigurable low-power sensors for computer vision, control and signal processing”. W: *Sensors* 14.4 (2014), s. 6247–6278.
- [6] Shanshan Zhang, Rodrigo Benenson i Bernt Schiele. „Filtered channel features for pedestrian detection”. W: *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE. 2015, s. 1751–1760.
- [7] Alejandro González i in. „Pedestrian detection at day/night time with visible and FIR cameras: A comparison”. W: *Sensors* 16.6 (2016), s. 820.
- [8] Rodrigo Benenson i in. „Ten years of pedestrian detection, what have we learned?” W: *arXiv preprint arXiv:1411.4304* (2014).
- [9] Navneet Dalal i Bill Triggs. „Histograms of oriented gradients for human detection”. W: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 1. IEEE. 2005, s. 886–893.

- [10] Dominik Honegger, Helen Oleynikova i Marc Pollefeys. „Real-time and low latency embedded computer vision hardware based on a combination of FPGA and mobile CPU”. W: *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE. 2014, s. 4930–4935.
- [11] Songlin Piao i in. „Real-time multi-platform pedestrian detection in a heavy duty driver assistance system”. W: *Proc. Int. Commercial Veh. Technol. Symp.* 2016, s. 61–70.