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 zwykorzystaniem 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

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.

Serdecznie dziękuję . . . tu ciąg dalszych podziękowań np. dla promotora, żony, sąsiada itp.

Spis treści

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 ludzi. Przez ostanie 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. Dzisiejsza technologia nie ogranicza nas tylko do stosowania spektrum wiatła 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 nasłonecznienia 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 system za działający w czasie rzeczywistym wykorzystując jedynie komputer. Daję to pole do popisu dla układów rekonfigurowalnych które mają możliwość dużego zrównolegnienia obliczeń. Układy FPGA (ang. field-programmable gate array) znalazły już zastosowani w wielu systemach wizyjnych wykonując różnego rodzaju niskopoziomowe operacje kontekstowe, zamiany przestrzeni barw czy też binearyzacji nawet w czasie jednego cyklu zegara. Dodatkową zaletą układów FPGA jest mały pobór mocy co czyni je niezwykle atrakcyjna 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

W pierwszej części została opisana budowa cyfrowego systemu wizyjnego z wykorzystaniem połączonych obrazów RGB oraz IR. Zawiera teorię tworzącą podstawę dla następnych rozdziałów oraz kilka przykładów już zrealizowanych systemów. W następnym rozdziale została podana specyfikacja techniczna zastosowanych urządzeń oraz technologii. W rozdziale czwartym opisano realizacja autorskiego systemu detekcji ludzi. Prace zakończono podaniem osiągniętych wyników i wnioskami.

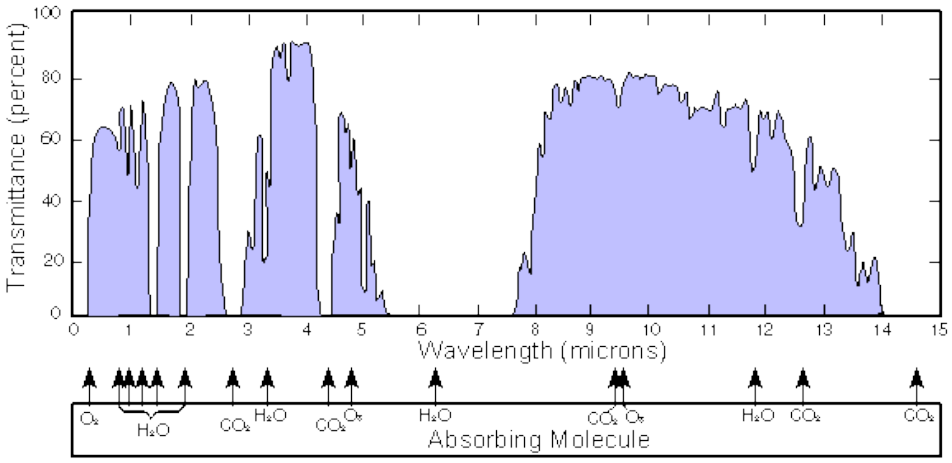
2. Cyfrowy system wizyjny

2.1. Podczerwień

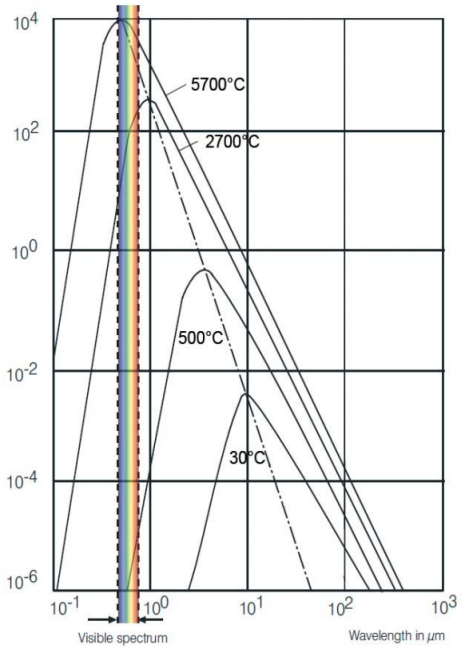
Jako podczerwień określa się promieniowanie elektromagnetyczne w zakresie długości fali od 0,75 μm do 1000 μm. Ciało które ma temperaturę powyżej zera absolutnego emituje swoją powierzchnią promieniowanie. Im większa jest temperatura ciała tym większa jest jego emisja. Dla każdej temperatury danego ciała istnieje charakterystyczna długość fali o najwyższej wartości mocy promieniowania. Im wyższa temperatura tym ta częstotliwość przesuwa się w zakres fal widzialnych. Można to zaobserwować gdy stal osiąga wysoką temperaturę powodując tym emisję światła. Ciało doskonale czarne całkowicie pochłania padające na nie promieniowanie, oraz emituję promieniowanie ściśle związane z jego temperaturą. Wykres na rysunku 2.1 przedstawia tą charakterystykę. Promieniowanie podczerwone jest częściowo pochłaniane przez atmosferę ziemską . Na rysunku 2.2 przedstawiono transmisyjność atmosfery. W aparaturze obrazującej w podczerwieni wykorzystuję się dwa zakresy przy których transmisyjność jest największa: 3 – 5 μm (MIWR, ang. *mid wave infrared* - podczerwień fal średnich) oraz 8 – 14 μm (LWIR , ang. *long wave infrared* - podczeriwń fal długich)[6].

2.2. Metody akwizycja obrazu

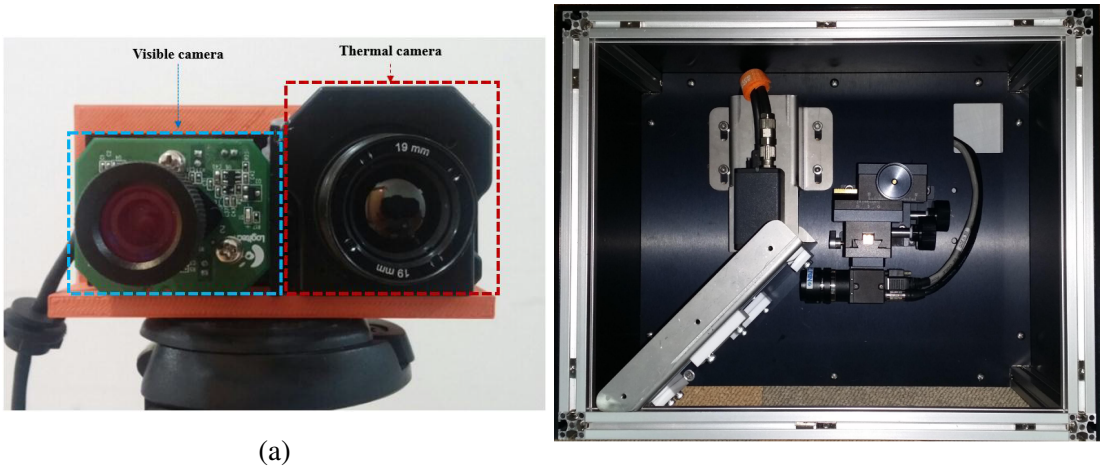
Większość implementacji wykorzystuje układ dwóch równoległych do siebie kamer. Do połączenia obrazów należy zastosować algorytm wyrównujący oba obrazy. Kalibrację wykonuję się specjalnymi planszami które pozwalają określić położenie punktów kalibracyjnych w obu rejestrowanych zakresach. Plansze mogą 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. Wykres transmisyjności atmosfery dla promieniowania podczerwonego [7].



Rys. 2.2. Emisyjność ciała idealnie czarnego.



(a) (b)

Rys. 2.3. Sposoby akwizycji obrazów: (a) dwie kamery równolegle [2], (b) z wykorzystaniem zwierciadła półprzezroczystego [4].

2.3. 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

$$P_c = RP + T \tag{2.1}$$

Właściwości optyczne kamery można przedstawić w postaci macierzy kamery.

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

gdzie:

f_x, f_y = ogniskowa kamery wyrażona w liczbie pikseli,
x_0, y_0 = współrzęne punktu głównego.

Macierz K określa związek między 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:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} \tag{2.3}$$

2.4. 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 poruszających to zagadnienie [8] 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 [9] ustalono że analiza multispektralna jest skuteczniejsza w dzień niż w nocy (o około 5% AMR (ang. avrange miss rate). W artykule [10] 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 ostanie 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:

2.4.1. 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.

2.4.2. Wyodrębnienie cech

Do najbardziej popularnych cech można zaliczyć:

1. Histogramy zorientowanych gradientów (HOG) zaproponowany przez N.Dalala i B. Triggs w pracy [11] 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 następuję normalizacja. Wektorem cech jest połączony wszystkich histogramów z wszystkich bloków w jeden wektor.

2. Lokalne wzorce binarne LBP (ang. Local Binary Paterns). Oryginalnie przeznaczone do opisu tekstur. Obraz zostaje podzielony na bloki. Następnie każdego piksela w bloku zostaje przypisany wzorzec binarny na podstawie wartości pikseli w jego sąsiedztwie. Jeżeli wartość sąsiadującego piksela jest większa od centralnego to przyjmuje on wartość 1. Następnie zostaje obliczony histogram dla każdego bloku. Histogramy z wszystkich bloków wchodzących w skład obrazu tworzą wektor cech [12].

3. Falki Haara. Określają różnicę w kontraście między dwoma przylegającymi prostokątnymi obszarami. Są łatwe do skalowania i nie wymagają dużych nakładów obliczeniowych.

4. Kolor. W analizie obrazo wykorzystuje różne przestrzenie barw np. RGB, HSV oraz LUV.

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

2.4.3. Klasyfikator

Otrzymany wektor cech jest poddany klasyfikacji której wynik decyduje czy obraz zawiera człowieka. W pracy [10] autorzy wyróżnili 3 dominujące rodziny:

1. Rodzina DPM (ang. Deformable Part Detectors) ??? wykrywacze deformowlnych 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.

2.5. 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:

$$P_w = \frac{1}{s + \frac{1-s}{n_w}} \tag{2.4}$$

gdzie:

P_w = przyspieszenie algorytmu w systemie wieloprocesorowym,
s = cześć algorytmu niepodlegająca zrównolegleniu (wartość od zera do jeden),
n_w = liczba elementów obliczeniowych.

Teoretycznie jedynym ograniczaniem w możliwości zrównolegnia 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]

2.6. Przegląd literatury

2.6.1. Podobne rozwiązania

W pracy [13] autorzy opracowali algorytm pozwalający na szybką i efektywną detekcję przechodniów w czasie rzeczywistym . Termowizja pozwala na uzyskanie dobrego kontrastu między poszukiwanym przechodniem a otoczeniem. System dedykowany jest do pracy w nocy

kiedy kontrast między człowiekiem pozwala na jednoznaczne ich rozróżnienie. Rozwiązanie bazuję na ulepszonym algorytmie progowania i segmentacji obrazu. Pierwszym etapem jest wyodrębnienie obszarów zainteresowań (ang. ROI). Pozwala to na znaczne ograniczenie obszaru obrazu do analizy. Obraz w odcieniach szarości zostaję poddany binaryzacji z użyciem dwóch progów: mniejszym i większym. Dodatkowo każdy wykryty obszar tworzy dodatkowy ROI przylegający do pierwotnego. Progowanie z pojedynczym progiem jest niewystarczające w wielu wypadach dlatego autorzy zastosowali podwójne progowanie. Pozwala to na detekcję przechodniów w różnych rejonach obrazu o różnym kontraście. Progi zmieniają się wraz z dynamiką obrazu wejściowego. W obrazie termicznym człowieka często występuję obszar o niższej temperaturze w okolicach bioder. Skutkuje to przerwą w zbinearyzowanym obiekcie i błędną klasyfikację np.: samych nóg. Autorzy opracowali technikę polegającą na powiększeniu obszaru. Łączy ona dwie połówki człowieka jeżeli posiadaj wspólne współrzędne wzdłuż pionowej osi tworząc nowy obszar. Ostatecznie obie grupy obszarów zainteresowania uzyskanych z obu progowań zostają połączone.

Następnym krokiem jest filtracja wyników. Ma na celu zredukowanie ilości obszarów do końcowej analizy. Autorzy zastosowali filtrację opierającą się na proporcji obszaru zainteresowań. Pozytywnie zakwalifikowane zostały tylko obszary o odpowiednich proporcja wysokości do szerokości (1:1.3 do 1:4). Z racji że badany obraz pochodzi z kamery zamontowanej na stałe na samochodzie autorzy wykorzystali filtrację perspektywiczną. W większej odległości na horyzoncie obiekty są mniejsze. Zakłada ona że w określonych obszarach obrazu istnieje maksymalna możliwa wysokość kandydata. Filtracja jednorodnych regionów pozwoliła na odrzucenie obszarów które często występują jako część szerszych obiektów nie mających nic wspólnego z przechodnimi. Autorzy zaproponowali by obliczenie odchylenia standardowego tych obszarów w odcieniach szarości i odrzucenie części która jest poniżej pewnego progu.

Ostatnim krokiem algorytmu jest klasyfikacja wytypowanych kandydatów. Autorzy wykorzystują Histogram zorientowanych gradientów jako cechę tworząc wektor 3780 cech które są przetwarzane przez maszynę wektorów nośnych.

W celu zbadania dokładności algorytmu został przeprowadzony test na zbiorze CVC-14 zawierający obrazy nagrane kamerą FIR podczas nocnego przejazdu samochodem. Testy wykazały że metoda podwójnego progowania daje trzy razy lepsze rezultaty niż przy wykorzystaniu pojedynczego progu. Wraz z zaproponowanymi technikami filtracji zaowocowało bardzo efektywnym mechanizmem segmentacji. Cała procedura detekcji przechodniów osiągnęła wysoki poziom wydajności na poziomie 33 klatek na sekundę przy wykorzystaniu pojedynczego rdzenia CPU.

„Pedestrian detection using infrared images and histograms of oriented gradients”

W pracy [] autorzy zaproponowali wykorzystanie dwóch kamer termowizyjnych tworząc system stereowizyjny. By wyodrębnić obszary zainteresowania, potencjalnie zawierający w sobie przechodnie, zgrupowano piksele o wartościach powyżej kilku różnych progów. Porównując te dwa obrazy można określić pozycję i odległość źródła ciepła od kamery. W obrazie termowizyjnym człowieka można zauważyć że najbardziej ciepłym i odsłoniętym obszarem ciała jest głowa. Wykorzystując ten fakt, oraz informację o odległości od kamery, zostają wytyczone obszary wokół tych pikseli o wielkości zależnej od tej odległości. Następnie wszystkie wyodrębnione tak obszary zostają przeskalowane do wymiaru 128x64 piksele i poddane klasyfikacji za pomącą kombinacji HOG+SVM. W tej pracy autorzy skupili się na optymalnym dobraniem parametrów HOG. Badanie zostały przeprowadzone na bazie obrazów termowizyjnych o wymiarach 128x64. Zestaw zawierał 4400 obrazów: 2200 zawierających przechodnia oraz 2200 nie zawierających. Został wykorzystany następujący zestaw parametrów HOG:

1. Wielkość komórki: 4x4, 8x8, 16x16,
2. wielkość bloku: 1x1, 2x2, 4x4,
3. nakładanie się bloków: 1, 2,
4. ilość przedziałów histogramu: 4, 8, 16,
5. - metoda dopasowania: ważony lub nie
6. metoda normalizacja bloku: L1, L2, brak

Parametry dla klasyfikatora SVM :

1. wielkość zestawu do nauki: 10, 100, 1000 obiektów na klasę,
2. waga źle sklasyfikowanych punktów C: 0.01, 1, 100.

Autorzy przeprowadzili po 10 nauczań klasyfikatora dla każdej kombinacji wykorzystując różne kombinacje danych do nauki i testów. Po przeprowadzonych badaniach został wytypowany optymalny zestaw parametrów:

1. Wielkość komórki: 8x8,
2. wielkość bloku: 2x2
3. nakładanie się bloków: 1,
4. ilość przedziałów histogramu: 4, 8, 16,

5. metoda dopasowania histogramu: ważona

6. metoda normalizacja bloku: L2.

Badanie parametrów dla nauczania SVM wynikło że im większy zestaw uczący tym lepszą można uzyskać skuteczność detekcji. Parametr C miał marginalne znaczenie na wyniki.

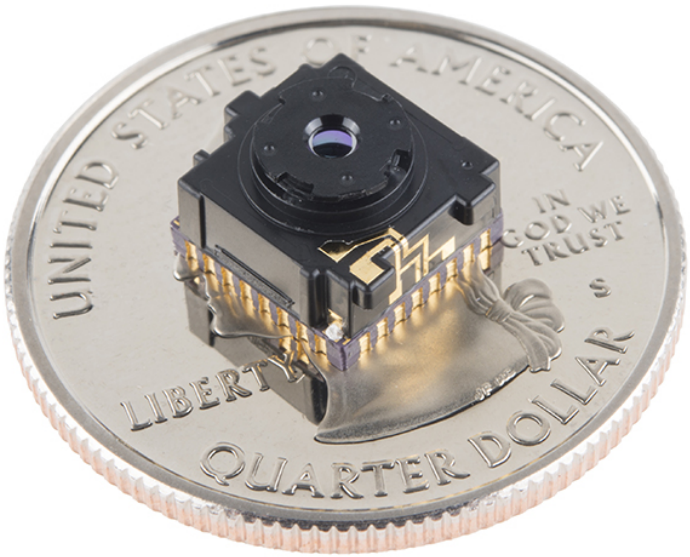
2.6.2. Podejście sprzętowo - programowe

W pracy [14] autorzy wykorzystali układ FPGA oraz CPU małej mocy do skonstruowania systemu wizyjnego dla robotów. System analizował obraz steroskopowy 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 Memeory Acces) 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 piksle i 60 klatkach na sekundę. Całość systemu wizyjne 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 [15] 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 śledzenie. Wyniki były przesyłane do wbudowanego komputera PC. Rozwiązanie nie zostało jeszcze w pełni zaimplementowane ale pierwsze testy dawały obiecujące rezultaty.

3. Wykorzystane zasoby sprzętowe i technologie

3.1. Kamera termowizyjna Lepton



Rys. 3.1. Widok poglądowy na kamere Flir 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. Checuje siębardzo małymi wymiarami co czyni go idealnym do zastosowań mobilnych. 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 pracuję w domyślnym trybie pracy,

który może zostać zmieniony za pomocą CCI (ang. camera control interface – interfejs kontroli kamery).[**lepton**] Parametry:

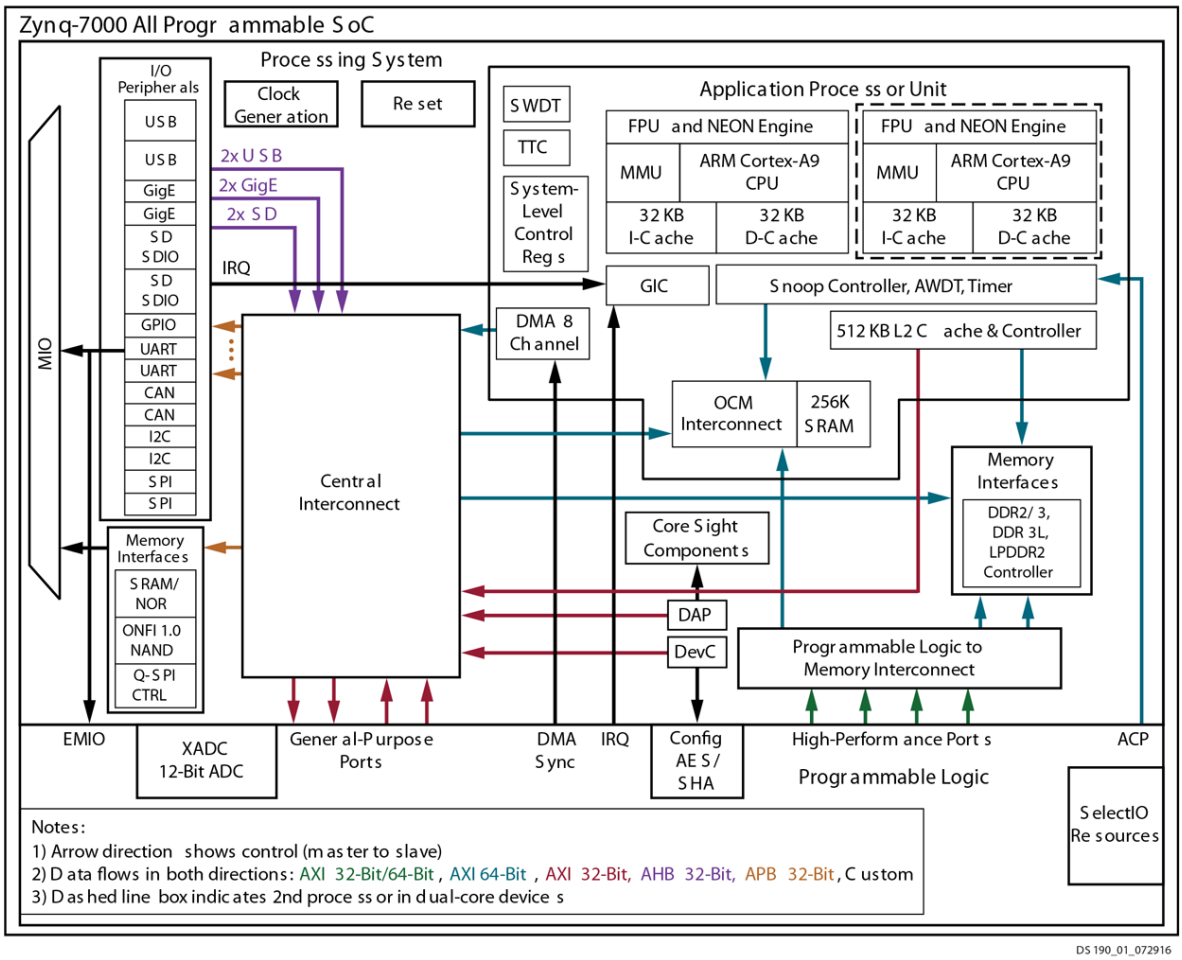
- Wymiary: 11,8 x 12,7 x 7,2 mm,
- Sensor: niechłodzony mikrobolometr VOx (tlenek wanadu),
- Rejestrowany zakres: fale długie podczerwieni, $8\mu m$ do $14\mu m$,
- Wielkość piksela: 17 μm,
- Rozdzielczosć: 80x60 pikseli,
- Ilość klatek na sekundę 8,6,
- Zakres rejestrowanych temperatur: -10 ° C 140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

3.2. 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 3.2 przedstawiono schemat architektury. Prócz procesora cześć 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)



Rys. 3.2. Schemat ogólny architektury układu Zynq-7000.

3.3. 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-Lite – 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:

– 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.

3.4. 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 połączane do specjalnego bloku detekcji timingów który mierzy parametry wchodzącego strumienia wizyjnego (ilość pikseli na linie, 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.)

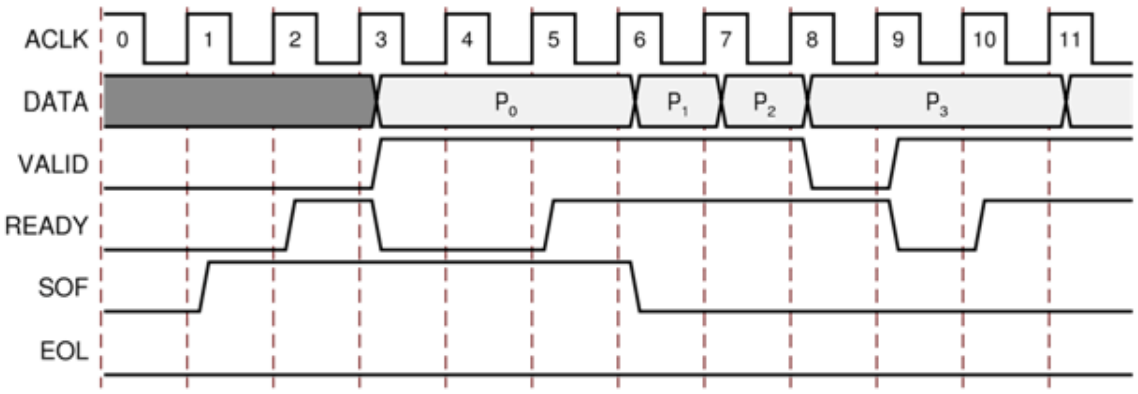
– Valid – Linia podająca czy dane piksela są poprawne,

– 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 3.3



Rys. 3.3. Przykład rozpoczęcia transmisji Reday/Valid.

4. Realizacja

W celu rozpoznania przechodnia został użyty połączony obraz termowizyjny i kolorowy nazywany dalej RGBIR. Następnie ten obraz zostaje poddany analizie HOG oraz klasyfikacji za pomocą SVM. W celu ustalenia obszaru zainteresowania na obrazie termowizyjnym za pomocą wzorca probabilistycznego zostają wytypowani kandydaci.

4.1. Akwizycja obrazu

Obraz kolorowy służy jako obraz bazowy. Rozdzielczości 640 x 480 pikseli, prędkością 30 klatek na sekundę i głębi 8 bitów na kanał. Źródłem tego obrazu jest kamera podłączona do układu za pomocą interfejsu HDMI. Na obraz bazowy zostaje nałożony obraz termowizyjny z kamery Lepton, który różni się znacząco parametrami. Abu je zsynchronizować zastosowano bufor ramki, do którego jest zapisywany obraz z prędkością 9 klatek na sekundę a odczytywany z prędkością 30. Kolejnym przekształceniem jest transformacja projekcyjna. Ma na celu powiększenie i dopasowanie obrazu by poprawnie pokrywał się z obraz wizyjnym. W tym celu został zaimplementowany moduł który oblicza na podstawie parametrów macierzy transformaty i koordynatami piksela obrazu źródłowego odpowiadającą mu pozycję na obrazie termowizyjnym zapisanym w buforze ramki. Następny moduł dokonuje interpolacji dwuliniowej. Do poprawnej interpolacji wymagane są 4 piksele otaczające obliczony z projekcji punkt. W celu zredukowania liczby dostępów do pamięci i zwiększenie szybkość działania moduł zapamiętuje 4 ostatnio użyte wartości pikseli. Rozwiązanie to pozwala na pracę w czasie rzeczywistym małym kosztem zasobów układu. Strumień wizyjny jak i termowizyjny działają w AXI-Stream. Umożliwia to łatwą synchronizację obu obrazów na podstawie sygnału SOF (ang. Start o frame). Moduł synchronizacji czeka na pojawienie się tego sygnału w strumieniu termowizyjnym. Do tego momentu wszystkie napływające piksele są odrzucane. Gdy pojawi się sygnał strumień IR zostaje zatrzymany i czeka na pojawienie się sygnału SOF w bazowym strumieniu wizyjnym. Po jego wykryciu strumień IR rusza. Oba strumienie zostają zsynchronizowane tworząc strumień wizyjna obrazu RGBIR. Następnie ten strumień zostaje przesłany do

pamięci za pośrednictwem VDMA oraz (po koloryzacji i nałożeniu) wyświetlony na monitorze przez port VGA.

4.2. Wyznaczanie ROI

Strumień IR z kamery zostaje zbinearyzowany i zbadany w detektorze DPM. Moduł DPM przesyła do pamięci listę koordynatów kandydatów wraz z mocą dopasowania. Moduł DPM został zaczerpnięty z pracy inżynierskiej. Moduł wykorzystuję strumień bezpośrednio z kamery. Wielkość okna detekcji wynosi 16 x 40 pikseli. Jeżeli badany obraz binarny wykazał odpowiedni poziom dopasowania do wzorca zostaje wysłana o tym informacja poprzez AXI-Stream do pamięci. Informacja zawiera koordynaty okna w układzie odniesienia kamery IR oraz wartość mocy dopasowania. Gdy zostanie zbadane ostatnie okno w obrazie zostaje wysłany sygnał LAST co wygeneruje przerwanie dla systemu procesorowego.

4.3. Klasyfikacja za pomcą SVM

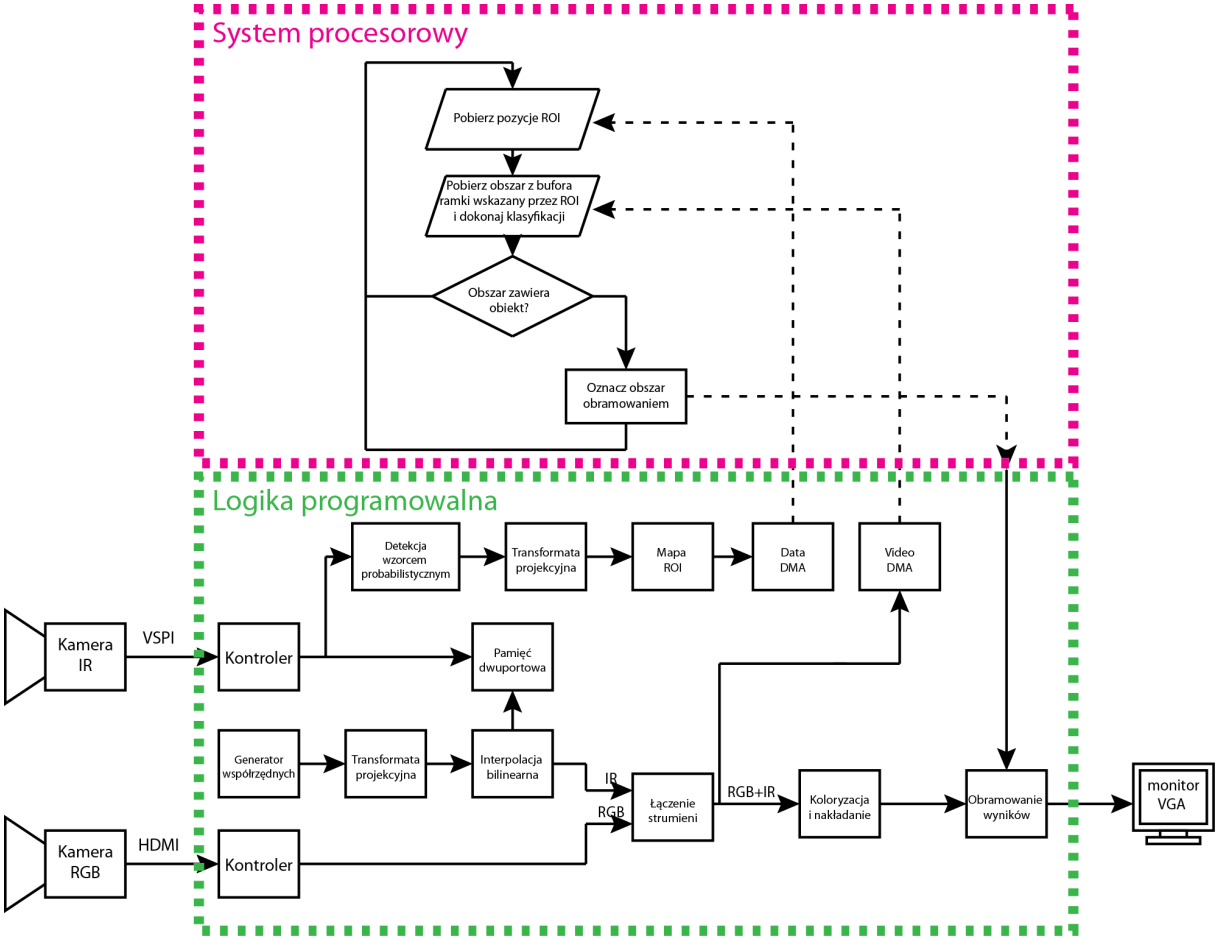
Z lisy kandydatów wygenerowanej przez moduł DPM wybierany jest wynik o najwyższej mocy dopasowania. Koordynaty z układu odniesienia kamery zostają poddane transformacie projekcyjnej do układu odniesienia kamery RGB. Z obszaru na obrazie RGBIR zawierającym potencjalnie człowieka zostają wyodrębnione cechy HOG które następnie służą jako wektor dla SVM.

Klasyfikator został opracowany i nauczony na podstawie 60 wyselekcjonowanych obrazów. 30 z nich stanowiło próbką pozytywną zawierającą osobę a 30 nie. Nauczanie zostało zrealizowane przy użyciu oprogramowania Matlab. Próbki pozytywne zostały wygenerowane poprzez zapis ROI wyznaczonych przez wzorzec probabilistyczny.

4.4. Prezentacja wyników

Na wyjściu konsoli zostają podane współrzędne oraz moc dopasowania i klasyfikacja obiektu. Na obrazie wyjściowym VGA obszar ten zostaje zaznaczony zieloną ramką. Jeżeli potencjalny obszar nie został zakwalifikowany jako człowiek ale miał największą moc dopasowania DPM to obszar zostaję zaznaczony czerwoną ramką. Czarna ramka oznacza że nie został wykryty żaden obiekt.

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 4.1.



Rys. 4.1. 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,
- 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.

4.5. Opis modułów

4.5.1. Kontroler kamery IR

Pobiera obraz z kamery poprzez interfejs VoSPI który następnie zostaje zapisany do dwuportowej pamięci BRAM.

4.5.2. Transformata projekcyjna

Zadaniem modułu jest dostosowanie obrazu IR by pokrywał się z obrazem RGB. Moduł transformaty projekcyjnej zamienia wygenerowane współrzędne w zakresie wielkości przychodzącego 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 wartość minimalną i maksymalną współrzędnych wyjściowych U i V oraz macierz transformaty.

4.5.3. Interpolacja bilinearna

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

4.5.4. Łączenie strumieni

Moduł posiada dwa wejścia dla obrazu. Jeden strumień jest głównym i do niego jest dołączany drugi strumień. Do synchronizacja 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.

4.5.5. 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.

5. Wyniki i wnioski

Aby sprawdzić działanie i dokładność systemu została zaimplementowana możliwość zapisu użytego obliczonego wektora cech na karcie SD. Następnie został obliczony przykładowy błąd względny między wektorem cech wyliczonym w implementacji programowej a uzyskanym z sytemu wizyjnego. Błąd oscyluje w granicy 10^{-6} co czyni go marginalnym i najprawdopodobniej wynika z różnić użytych bibliotek numerycznych.

<TU WSTAW WYKRES>

Na przebadanie jednego okna zaproponowany system procesorowy potrzebuje 75ms (dla porównania te same obliczenia w pakiecie Matlab zajmują około 23 ms). Dzięki zastosowaniu sprzętowego wyszukiwania ROI zadanie sytemu procesorowego zostało ograniczone do obliczenia jednego okna z największym prawdopodobieństwem zawierania w sobie przechodnia. Kamera termowizyjna będąca źródłem sygnału dla wzorca probabilistycznego pracuję z prędkością 9 klatek na sekundę dając w przybliżeniu 111 ms na zbadanie danego okna więc system procesorowy mieści się w tych ramach czasowych z dużym zapasem.

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 Garcıa 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] Frank Niklaus, Christian Vieider i Henrik Jakobsen. „MEMS-based uncooled infrared bolometer arrays: a review". W: *Photonics Asia 2007*. International Society for Optics i Photonics. 2007, s. 68360D–68360D.

[7] Wikipedia. *Infrared, Wikipedia, The Free Encyclopedia*. [Dostęp: 9 stycznia 2016]. 2016.

[8] 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.

[9] Alejandro González i in. „Pedestrian detection at day/night time with visible and FIR cameras: A comparison". W: *Sensors* 16.6 (2016), s. 820.

[10] Rodrigo Benenson i in. „Ten years of pedestrian detection, what have we learned?" W: *arXiv preprint arXiv:1411.4304* (2014).

[11] 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.

[12] Timo Ojala, Matti Pietikainen i Topi Maenpaa. „Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". W: *IEEE Transactions on pattern analysis and machine intelligence* 24.7 (2002), s. 971–987.

[13] Karol Piniarski i Pawel Pawlowski. „Efficient pedestrian detection with enhanced object segmentation in far IR night vision". W: wrz. 2017, s. 160–165.

[14] 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.

[15] 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.