AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ

KATEDRA AUTOMATYKI I ROBOTYKI

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, 2018

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.

Streszczenie

Słowa kluczowe:

Abstract

Keywords:

Spis treści

1. Wprowadzenie

Cyfrowa analiza obrazów znalazła szerokie zastosowanie w wielu dziedzinach życia. Dzięki niej możliwe jest automatyczne uzyskanie istotnych dla użytkownika informacji. Przez ostanie kilkadziesiąt lat opracowano tysiące różnych technik i algorytmów wyspecjalizowanych do określonych zadań np. leśne fotopułapki do badania zachowań i migracji zwierząt, systemu kontroli jakości i przebiegu procesu przemysłowego, metody kontroli dostępu poprzez rozpoznawanie twarzy m.in. w smartfonach, algorytmu analizy zdjęć satelitarnych ziemi umożliwiające prognozowanie pogody, sterowanie ruchem drogowym na podstawie obrazu z kamer zamontowanych nad skrzyżowaniami, a także systemy do badania przekroju żołędzia pozwalające określić czy jest dobrym kandydatem na sadzonkę.

Wzrok ludzki operuje w pewnym zakresie promieniowania elektromagnetycznego zwanego światłem widzialnym. Dzisiejsza technologia daje możliwość rejestracji obrazów wykraczających poza to widmo. Kamery termowizyjne stają się coraz tańsze i przez to bardziej popularne. Dostarczają one informację o temperaturze obserwowanych obiektów. Jest to coraz chętniej wykorzystywane np. w weterynarii do określenia miejsc urazów zwierząt, w przemyśle do kontroli jakości artykułów spożywczych, w budownictwie do analizy strat cieplnych w budynkach, w systemach wspomagania kierowcy, przez ratowników do odnajdywania zasypanych ludzi w gruzowiskach, straż graniczną do monitorowania granic, przez wojsko do odnajdywania celów i zagrożeń podczas misji m.in. z wykorzystaniem dronów. [1].

Większość systemów wizyjnych służących do rozpoznawania przechodniów jest oparta 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 lepszej skuteczności rozpoznawania ludzi. W pracy

[3] autorzy nazywają ten rozszerzony format jako RGBT ("Red-Green-Blue-Thermal"), natomiast inna praca jako analizę wielospektralną (ang. *Multispectral*) [4], albo po prostu jako połączony obraz z kamery termowizyjnej i wizyjnej [2].

Skuteczna detekcja obiektów często wymaga dużego zapotrzebowania na zasoby obliczeniowe. W wielu przypadkach nie da się uzyskać satysfakcjonującej wydajności – tak by można było uznać system za działający w czasie rzeczywistym – wykorzystując jedynie typowy komputer wyposażony w procesor ogólnego przeznaczenia. Stosuje się zatem różne metody akceleracji obliczeń. Karty graficzne (GPU ang. *graphics processing unit*) pozwalają na duże zrównoleglenie obliczeń, jednak charakteryzują się znacznym zużyciem energii. Tworzenie specjalizowanych układów scalonych (ASIC ang. *application-specific integrated circuit*) daje najlepsze rezultaty w implementacji systemu wizyjnego, ale ich opracowanie i produkcja wymaga bardzo dużych nakładów finansowych. Dobre rozwiązanie stanowią układy rekonfigurowalne, które charakteryzują się podobnymi możliwościami w realizacji wyspecjalizowanych zadań co układy ASIC, ale nie wymagają dużych nakładów finansowych w ich tworzeniu.

Układy FPGA (ang. *Field-Programmable Gate Array*) umożliwiają zrównoleglenie obliczeń i są szeroko stosowane w systemach wizyjnych. Szczególnie chętnie są wykorzystywane do realizacji operacji niskiego poziomu, przygotowując wstępnie obraz do dalszej analizy na wysokim poziomie. Przykłady takich operacji to: filtry konwolucyjne, filtry 2D, podpróbkowanie, wykrywanie krawędzi, obliczanie SAD (ang. *sum of absolute differences*) z regionu zainteresowania, obliczanie orientacji krawędzi i histogramów, obliczanie strumieniowo statystyk (wartość maksymalna, minimalna, średnia), zmiana przestrzeni barw [**kisacanin2008embedded**]. Dodatkową zaletą układów FPGA jest mały pobór mocy, co czyni je niezwykle atrakcyjne dla aplikacji mobilnych – takich jak drony czy czujniki środowiskowe [5]. Układy heterogeniczne łączą w jednej obudowie dwa układy o różanej architekturze i funkcjonalności. Przykładem takiego połączenia jest Zynq-7000 firmy Xilinx, który integruje w sobie układ FPGA oraz procesor ARM. Największą zaletą takiego rozwiązania jest wysoka przepustowość transferu danych między procesorem a logiką programowalną.

Niniejsza praca stanowi kontynuację i rozwinięcie pracy inżynierskiej autora.

1.1. Cel pracy

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

1.2. Struktura pracy

W pierwszej części pracy została opisana budowa cyfrowego systemu wizyjnego z wykorzystaniem połączonych obrazów RGB oraz IR. Rozdział 2zawiera teorię stanowiącą podstawę dla realizowanych prac oraz kilka przykładów już zrealizowanych systemów. W rozdziale 4 zostały przedstawione wykorzystane zasoby sprzętowe oraz technologie użyte w opracowaniu systemu wizyjnego. W rozdziale 5 zawiera realizację autorskiego systemu detekcji ludzi. Prace zakończono omówieniem uzyskanych wyników, wnioskami oraz wskazaniem dalszych kierunków rozwoju stworzonego systemu.

2. System multispektralny

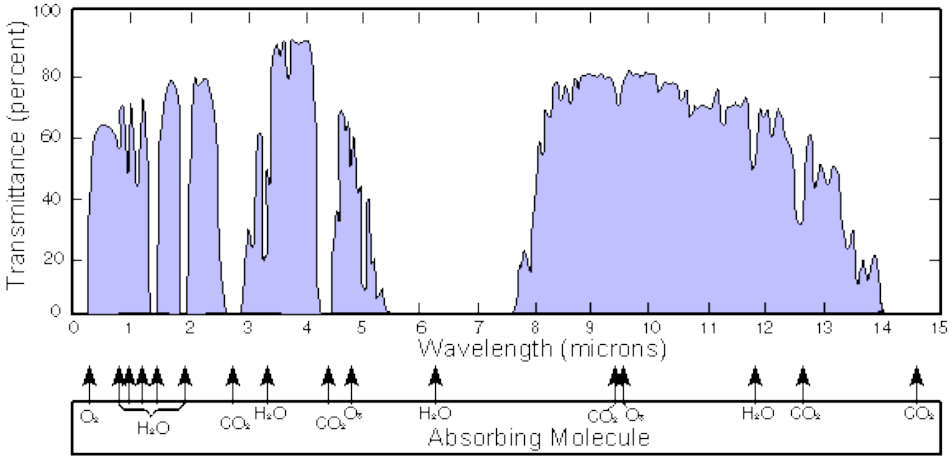
2.1. Podczerwień

Każde ciało które ma temperaturę wyższą niż zero absolutne emituje swoją powierzchnią promieniowanie którego natężenie rośnie wraz z jej wzrostem. Dla każdej temperatury danego ciała istnieje charakterystyczna długość fali o najwyższej wartości mocy promieniowania. Wraz z wzrostem temperatury ta częstotliwość przesuwa się w zakres fal widzialnych. Można to zaobserwować, gdy stal osiąga wysoką temperaturę co skutkuję emisją światła. Zależność ta jest opisana prawem Plancka, które opisuję emisję promieniowania elektromagnetycznego przez ciało doskonale czarne. Ciało doskonale czarne to wyidealizowane ciało fizyczne, które całkowicie pochłania padające na nie promieniowanie oraz emituję promieniowanie ściśle związane z jego temperaturą. Wykres na rysunku 2.1 przedstawia tą zależność.

Mianem podczerwieni określa się promieniowanie elektromagnetyczne w zakresie fali o długości od 0,75 μm do 1000 μm. Wyróżnia się następujące pasma podczerwieni:

- Bliska podczerwień (NIR ang. *near infrared*) w zakresie 0,75 μm do 1,4 μ.
- Podczerwień fal krótkich (SWIR ang. *short-wawelength infrared*) w zakresie 1,4μm do 3μm.
- Podczerwień fal średnich (SWIR ang. *mid-wavelength infrared*) w zakresie 3μm do 8μm.
- Podczerwień fal długich (LWIR ang. *long-wavelength infrared*) w zakresie 8μm do 15μm.
- Daleka podczerwień (FIR ang. *long-wavelength infrared*) w zakresie 15μm do 1000μm.

Bliska podczerwień znajduje się tuż za zakresem światła widzialnego ludzkim wzrokiem i jest możliwa do rejestracji przez typowe dla kamer sensory CCD czy CMOS (często z zastosowaniem oświetlaczy oświetlaczy IR). Wraz z wzmacniaczem światła jest również stosowana w noktowizji.



Rys. 2.1. Wykres transmisyjności atmosfery dla promieniowania podczerwonego [7].

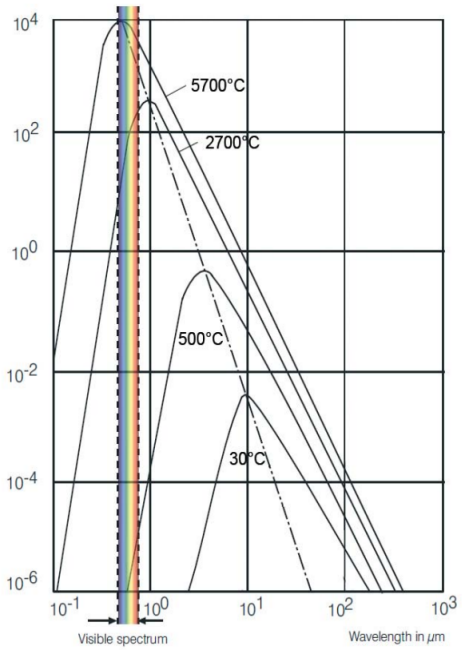
SWIR i LWIR występują również pod nazwą termowizji. Promieniowanie podczerwone jest częściowo pochłaniane przez atmosferę ziemską. Na rysunku 2.2 przedstawiono tzw. transmisyjność atmosfery. W aparaturze rejestrującej w podczerwieni wykorzystuję się dwa zakresy przy których transmisyjność jest największa: 3 – 5 μm oraz 8 – 14 μm [6].

2.2. Kamera termowizyjna

2.2.1. Sensor podczerwieni

W kamerach do rejestracji obrazu w termowizji są wykorzystywane sensory FPA (ang. *Focal Plane Array* - płaskie zespoły ogniskujące). Najbardziej popularne typy to : InSb, InGaAs, HgCdTe (w postaci fotodiod; wymagają kriogenicznych warunków pracy) and QWIP (ang. *Quantum well infrared photodetector*). Najnowsze technologie wykorzystują niskobudżetowe, niewymagające chłodzenia mikrobolometry.

Firma Flir wykorzystuje tlenek wanadu do budowy mikrobolometrów m. in. w kamerach Lepton. Tlenek walaniu cechuje się dużym temperaturowym współczynnikiem rezystancji (TWR) oraz małym szumem 1/f co zapewnia doskonałą wrażliwością oraz stabilną jednolitość. Do uzyskania obrazu zespół soczewek skupia promieniowanie z rejestrowanej sceny na macierz detektorów. W każdym z detektorów w odpowiedzi na padającą na niego wiązkę promieniowania, zmienia się temperaturę zawartego w nim tlenku wanadu. Zmiana temperatury wiąże się proporcjonalnie z zmianą rezystancji. Rejestracja sceny polega na odczycie rezystancji każdego detektora poprzez przyłożenia napięcia i odczyt przepływającego przez nie prądu. [8]



Rys. 2.2. Emisyjność ciała idealnie czarnego.

2.2.2. Kamera termowizyjna FLIR Lepton

Lepton jest miniaturową kamerą termowizyjną. W pojedynczym układzie został zintegrowany kompletny system składający się soczewki, sensora podczerwieni fal długich (ang. LWIR – *long wave infrared*) oraz elektroniki sterującej i przetwarzającej sygnał. Cechuje się bardzo małymi wymiarami, co czyni ją idealnym do zastosowań mobilnych. Układ ma możliwość domontowania dodatkowej przesłony, która jest wykorzystywana do automatycznej optymalizacji procesu ujednolicania obrazu (kalibracji sensora). Układ jest prosty do integracji z dowolnym mikrokontrolerem dzięki zastosowaniu standardowych protokołów i interfejsów. Lepton po podłączeniu zasilania od razu uruchamia się w domyślnym trybie pracy. Kamera jest konfigurowalna poprzez CCI (ang. *camera control interface* – interfejs kontroli kamery poprzez który jest dostęp do rejestrów zawierających konfigurację[**lepton**]

Parametry kamery:

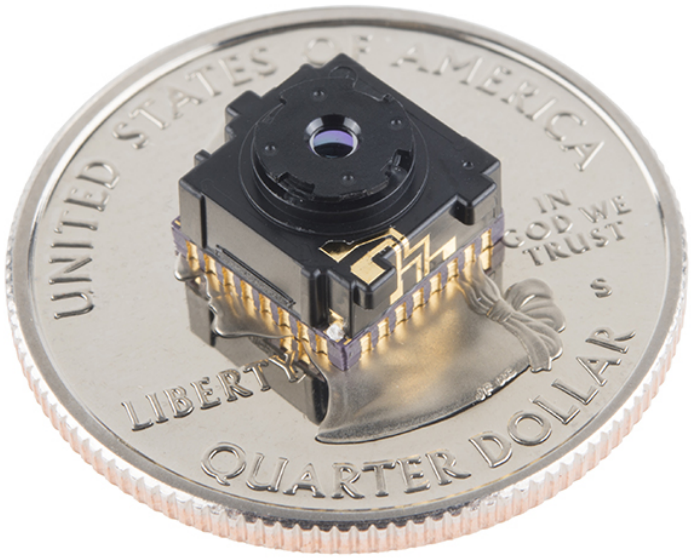
– Wymiary: 11,8 x 12,7 x 7,2 mm,

– Sensor: niechłodzony mikrobolometr VOx (tlenek wanadu),

– Rejestrowany zakres: fale długie podczerwieni, $8\mu m$ do $14\mu m$,

– Wielkość piksela: 17 μm,

– Rozdzielczość: 80x60 pikseli,



Rys. 2.3. Widok poglądowy na kamerę FLIR Lepton.

- Liczba klatek na sekundę: 8,6,
- Zakres rejestrowanych temperatur: -10 ° C 140 ° C (tryb wysokiego wzmocnienia),
- 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 wzmocnienia)) 24-bit RGB (z ACG i koloryzacją),
- Interfejs wideo: VoSPI (Video over Serial Peripherial Interface),
- Interfejs sterujący: CCI (zbliżony do I2C).

2.3. Rejestracja obrazu multispektralnego

Widmo elektromagnetyczne docierające do kamery składa się fal o różnych długościach. Sensory w kamerach rejestrują obraz tylko w pewnym zakresie tego widma, więc aby uzyskać

obraz w wymaganym paśmie należy odfiltrować niepożądane elementy widma np. kolorowy obraz z kamery wizyjnej jest otrzymywany poprzez zastosowanie trzech filtrów: czerwonego, zielonego i niebieskiego. Ponieważ wszystkie trzy kolory mogą być zarejestrowane przez pojedynczą matrycę, filtry są nałożone bezpośrednio na sensor a wartość koloru w danym punkcie jest interpolowana z sąsiadujących ze sobą pikseli. W przypadku gdy nie jest możliwe zastosowanie jednego sensora do wszystkich pożądanych zakresów należy rozdzielić wiązkę pomiędzy różne aparaty, albo wykorzystać równoległy układ kamer.

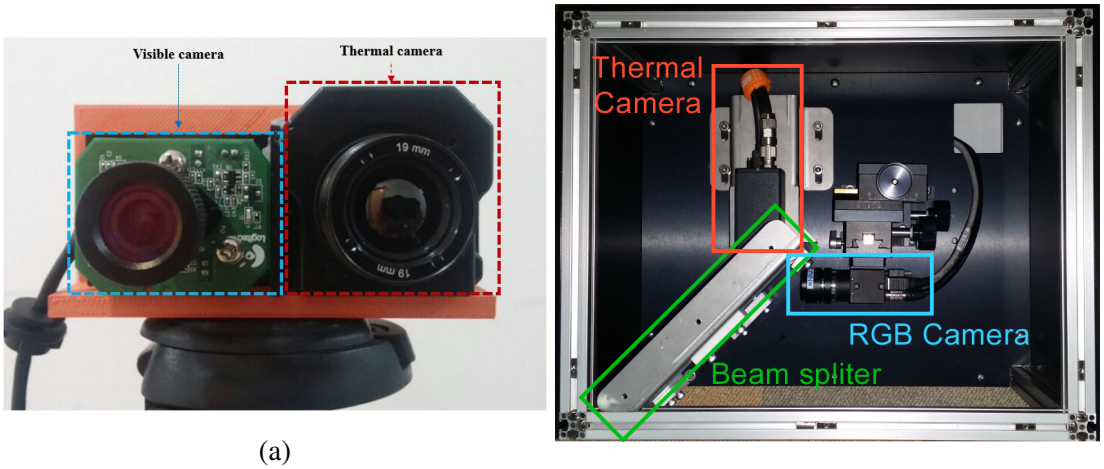
W przypadku rejestracji obrazu wizyjnego i termicznego większość implementacji wykorzystuje układ dwóch równoległych do siebie kamer, której przykład przedstawia rysunek 2.4a. W tym przypadku została zastosowana kamera termowizyjna Flir Tao 2 oraz kamery wizyjnej logitech webcam c600. Zazwyczaj obrazy z kamer różnią się, wynika to z ich budowy, różnej rozdzielczości, kąta widzenia jak oraz zniekształceń soczewkowych. Do poprawnego odwzorowania tej samej sceny w obu widmach należy zastosować algorytm mający na celu dopasowanie obu obrazów. Tworzony jest w ten sposób nowy obraz na którym wszystkie piksele łączą informację o kolorze i temperaturze.

Pierwszym z etapów poprawnego dopasowania obrazów jest kalibracja. Wykonuje się ją z wykorzystaniem specjalnych plansz, które pozwalają określić położenie pewnych punktów w przestrzeni w obu rejestrowanych zakresach promieniowania. Punkty te pozwalają na obliczenie relacji między obrazami. Plansze mogą być aktywne (posiadają własne źródło ciepła) albo pasywne (przesłaniają obce źródło ciepła). W równoległym układzie kamer występuje 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 wizyjnego od termicznego(rysunek 2.4b). Wykorzystując trójosiowy uchwyt, kamery zostały ustawione tak by ich osie optyczne pokrywały się. Następnie obrazy z obu kamer zostały zrektyfikowane tak by miały tą samą wirtualną ogniskową.

2.3.1. Model geometryczny

Do opisu matematycznego systemu wykorzystuje się model kamery otworkowej. Dzięki niemu można opisać relację między trójwymiarową przestrzenią, a dwuwymiarowym obrazem za pomocą projekcji perspektywicznej. Nie stanowi on najdokładniejszego opisu matematycznego kamery, nie ma w nim uwzględnionych zakłóceń soczewkowych, jednakże zapewnia dobre rezultaty w wielu aplikacjach. Model składa się z 2 zestawów parametrów: zewnętrznych oraz wewnętrznych. Parametry zewnętrzne definiują lokację kamery względem zewnętrznego



(a) (b)

Rys. 2.4. Sposoby akwizycji obrazów: (a) dwie kamery równolegle [2], (b) z wykorzystaniem zwierciadła półprzezroczystego [4].

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.

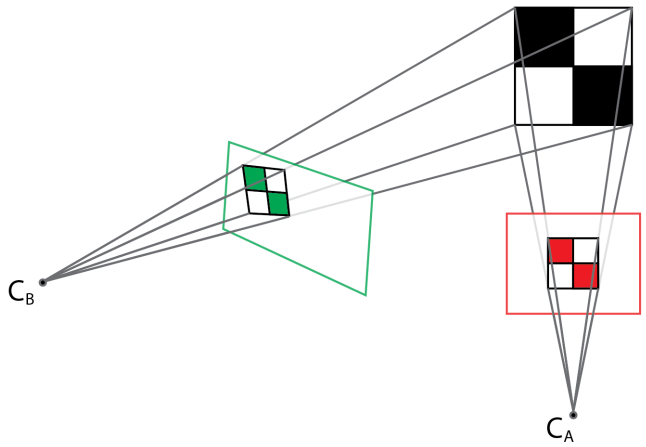
$$K = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.2}$$

gdzie:

f_x, f_y = ogniskowa kamery wyrażona w liczbie pikseli,
x_0, y_0 = 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:

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

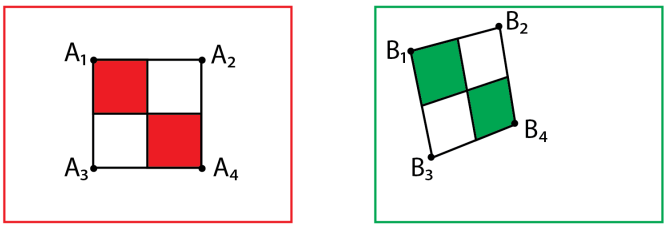


Rys. 2.5. Dwie kamery rejestrujące jeden obiekt.

2.3.2. Kalibracja

Obrazy które przedstawiają tę samą scenę ale zostały wykonane dwoma różnymi kamerami w innych położeniach, różnią się. Na rysunku 2.5 czarna szachownica jest uchwycona przez dwie kamery ustawione w punktach C_A)(na wprost obiektu), C_B (po skosie i lekko obrócona).

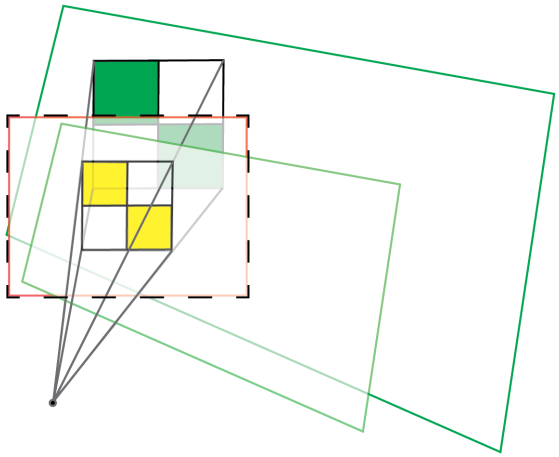
Aby dopasować te dwa obrazy tak by szachownice były ujęte tak samo należy na jednym z nich przeprowadzić transformację projekcyjną. Jest to przekształcenie pomiędzy dwoma płaszczyznami, które wykorzystuje model geometryczny kamery. Wymaga to obliczenia macierzy transformacji A na podstawie co najmniej 4 punktów kalibracyjnych.



(a) (b)

Rys. 2.6. Zarejestrowane obrazy: (a) przez kamerę C_A, (b) przez kamerę C_B. Punkty A i B są punktami kalibracyjnymi.

Na rysunkach 2.6a i **??** punkty A_1 do A_4, będące czterema rogami zarejestrowanej szachownicy przez kamerę C_A, odpowiadają punktom B_1 do B_4 będącymi tymi samymi czterema rogami zarejestrowanymi kamerą C_B. Macierz transformacji A można obliczyć rozwiązując



Rys. 2.7. Interpretacja transformacji projekcyjnej: rzutowanie płaszczyzny.

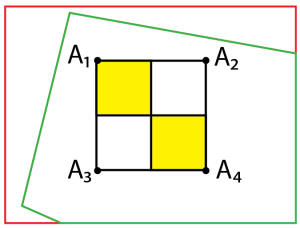
równanie (2.4).

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \tag{2.4}$$

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ \dots \\ u_n \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ \dots \\ v_n \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1x_1 & -u_1y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -u_2x_2 & -u_2y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -u_3x_3 & -u_3y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -u_4x_4 & -u_4y_4 \\ \dots & & & & & & & \\ x_n & x_n & 1 & 0 & 0 & 0 & -u_nx_n & -u_ny_n \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1x_1 & -v_1y_1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -v_2x_2 & -v_2y_2 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -v_3x_3 & -v_3y_3 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -v_4x_4 & -v_4y_4 \\ \dots & & & & & & & \\ 0 & 0 & 0 & x_n & y_n & 1 & -v_nx_n & -v_ny_n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{bmatrix} \tag{2.5}$$

$u_n, v_n =$ współrzędne punktu kalibracji n, na obrazie bazowym
$x_n, y_n =$ współrzędne punktu kalibracji n, na obrazie dopasowywanym

Transformacje projekcyjną można zinterpretować jako rzutowanie płaszczyzny, co obrazuję rysunek 2.7. Wynikiem transformacji (a zarazem rzutowania) jest obraz dopasowany do obrazu bazowego (2.8)



Rys. 2.8. Wynik transformacji.

3. Algorytmy detekcji pieszych

W cyfrowej analizie obrazu rozpoznawanie pieszych jest jedną z najbardziej aktywnie rozwijanych dziedzin. W przeciągu kilkudziesięciu lat powstało ponad tysiąc artykułów poruszających to zagadnienie [9], w których zaproponowano wiele różnych metod. 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 artykule [10] wykazano, że analiza multispektralna jest skuteczniejsza w dzień niż w nocy (o około 5% AMR (ang. *avrange miss rate*). W artykule [11] autorzy podsumowują osiągnięcia w dziedzinie detekcji pieszych w latach 2004 – 2014. Wyróżniono ponad 40 różnych podejść do problemu. Eksperymenty w artykule są oparte o bazę danych Caltech-USA, która zawiera obrazy w kolorze. Jednym z wniosków jest, że przez ostanie dziesięć lat największy postęp został osiągnięty głównie dzięki dopracowaniu cech, które 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:

3.1. Ustalenie regionu zainteresowań

Jest to obszar zwany ROI (ang. *region of interest*), w którym potencjalnie mogą znajdować się piesi. Wiele podejść uznaje cały obraz jako ROI i stosuje okno przesuwne sprawdzając każdy możliwy fragment obrazu. Jeżeli scena jest rejestrowana przez nieruchomą kamerę, ROI można określić poprzez różnicę między zapamiętanym tłem, a aktualnym obrazem (tzn. modelowanie i odejmowanie tła). Analiza przepływu optycznego również pozwala na wyodrębnienie obszaru który swoim ruchem różni się od reszty. Inną metodą jest zastosowanie słabszego, bardziej ogólnego klasyfikatora ale mniej wymagającego obliczeniowo. Wyodrębnienie ROI jest bardzo istotne w przypadku pracy w czasie rzeczywistym, ze względu na ograniczony czas analizy pojedynczego obrazu.

3.2. Wyodrębnienie cech

Do najbardziej popularnych cech można zaliczyć:

1. Histogramy zorientowanych gradientów (HOG *Histogram of Oriented Gradients*). Algorytm został zaproponowany przez N.Dalala i B. Triggs w pracy [12] i stał się jednym z najbardziej popularnych technik w dziedzinie detekcji ludzi. Jest cały czas rozwijany i modyfikowany 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 połączenie wszystkich histogramów z wszystkich bloków w jeden wektor.

2. Lokalne wzorce binarne LBP (ang. *Local Binary Paterns*). Oryginalnie deskryptory te zaproponowane zostały do opisu tekstur. Analizowany obraz zostaje podzielony na bloki. Następnie, do 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. W ten sposób do każdego piksela przypisywany jest wzorzec binarny (np. 100110). Następnie zostaje obliczony histogram dla każdego bloku. Histogramy z wszystkich bloków wchodzących w skład obrazu tworzą wektor cech [13].

3. Falki Haara. Określają różnicę w kontraście między dwoma przylegającymi prostokątnymi obszarami. W oryginalnej pracy P.Viola i M.Jones z 2001 **??** autorzy rozważali 3 rodzaje cech: Dwa obszary mające ten sam rozmiar i kształt oraz przylegają do siebie horyzontalnie bądź wertykalnie, gdzie cechę stanowi różnica sumy pikseli zawartych w każdym z regionów. Obszar składający się z 3 prostokątów przylegających do siebie gdzie od sumy środkowego elementu jest odejmowana suma dwóch zewnętrznych oraz układ 4 prostokątów, gdzie suma jest różnicą między obszarami po przekątnej. Cechy są łatwe do skalowania i nie wymagają dużych nakładów obliczeniowych.

4. Kolor. W analizie obrazów wykorzystuje różne przestrzenie barw np. RGB, HSV oraz LUV. Wykorzystywane głównie gdy kolor wykrywanego obiektu jest kluczowy(np. znaki drogowe, światła na skrzyżowniu). Jako cecha można go wykorzystać w kilku formach. Momenty koloru (ang. *Color Moments*) jest to średnia, wariancja i odchylenie standardowe występowania danego koloru w obrazie. Histogramu określa częstość występowania danego koloru na obrazie. Wektor koherencji koloru (CCV ang. *Color Coherence Vectors*) określa w jakim stopniu piksele danego koloru są częścią obszaru o podobnym

kolorze (np. Weźmy obraz zielonej łąki na którym pasła by się jedna fioletowa krowa. Kolor zielony na obrazie byłby rozłożony równomiernie natomiast fioletowy byłby skupiony w pojedynczym rejonie koherencji - krowy) **??**.

3.3. Klasyfikator

Otrzymany wektor cech jest poddany klasyfikacji, której wynik decyduje czy obraz zawiera człowieka. W pracy [11] autorzy wyróżnili 3 dominujące rodziny metod:

1. Rodzina DPM (ang. *Deformable Part Model*)

   Technika zakłada że obiekty mogą być zamodelowane poprzez części ułożone w deformowanych konfiguracjach. Model składa się z głównego, globalnego filtra, który stanowi punkt odniesienia dla pozostałych części. Każda część zawiera swój własny filtr wraz z zestawem dozwolonych pozycji względem okna detekcyjnego, oraz koszt deformacji dla każdej z tych pozycji. Suma wyniku uzyskanego z filtra głównego wraz z jego częściami stanowi o wyniku detekcji **??**.

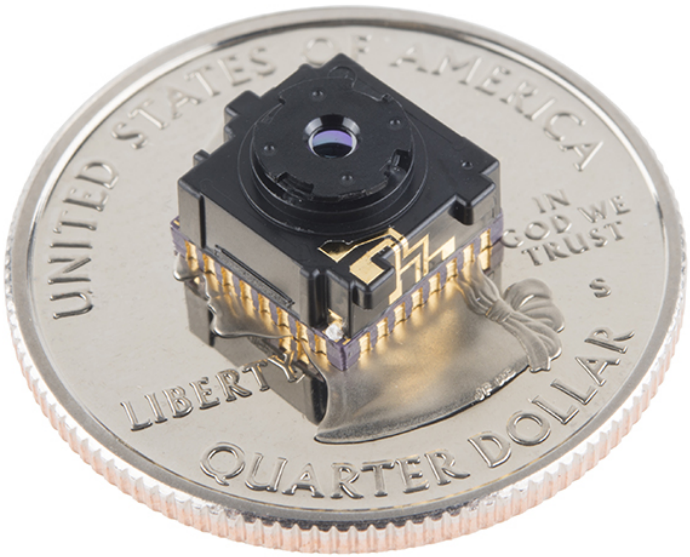
2. Deep networks. Głębokie sieci neuronowe posiadają kilkanaście warstw ukrytych między warstwą wejściową i wyjściową. Jej działanie polega na tym że po podaniu wektora cech na warstwę wejściową wytrenowanej sieci, w warstwie wyjściowej aktywuje się neutron odpowiedzialny za daną klasę. W analizie obrazu szczególnie chętnie są wykorzystywane sieci konwolucyjne. Neurony pierwszej warstwy ukrytej są podłączone jedynie do wybranego fragmentu warstwy wejściowej (np. okna 5x5 obrazu albo pojedynczego histogramu w komórce). Jest to tzw. warstwa konwolucyjna. Neurony w tej warstwie dzielą wspólne wagi dla swoich wejść i bias. Sieć posiada zazwyczaj kilkanaście takich warstw każda wykrywająca pojedynczą cehę. Pozwala to na redukcję ilości potrzebnych neutronów i mniejszą ilość parametrów potrzebnych do uzyskania w procesie uczenia. Do warstw konwlucyjnych dochodzą warstwy sumujące (ang. *pooling layers*). Zadaniem warstwy jest generalizacja informacji z poprzedniej warstwy. Sieć zamyka w pełni połączona z poprzednimi, warstwa wyjściowa.

3. Decision forests – lasy decyzyjne zbiór nieskorelowanych drzew decyzyjnych. Drzewo jest graficznym odwzorowaniem procesu decyzyjnego. Algorytm uczenia drzew wykorzystuje przykłady (wektor cech) i związane z nimi konsekwencje (klasyfikacja obiektu).[**wikiedia**].

4. inne: np. SVM (ang. support vector machine – maszyna wektorów nośnych), AdaBoost itp.

4. Wykorzystane zasoby sprzętowe i technologie

4.1. Kamera termowizyjna Lepton



Rys. 4.1. Widok poglądowy na kamerę 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. Cechuje się bardzo małymi wymiarami, co czyni ją 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 uruchamia się w domyślnym trybie pracy,

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

Parametry kamery:

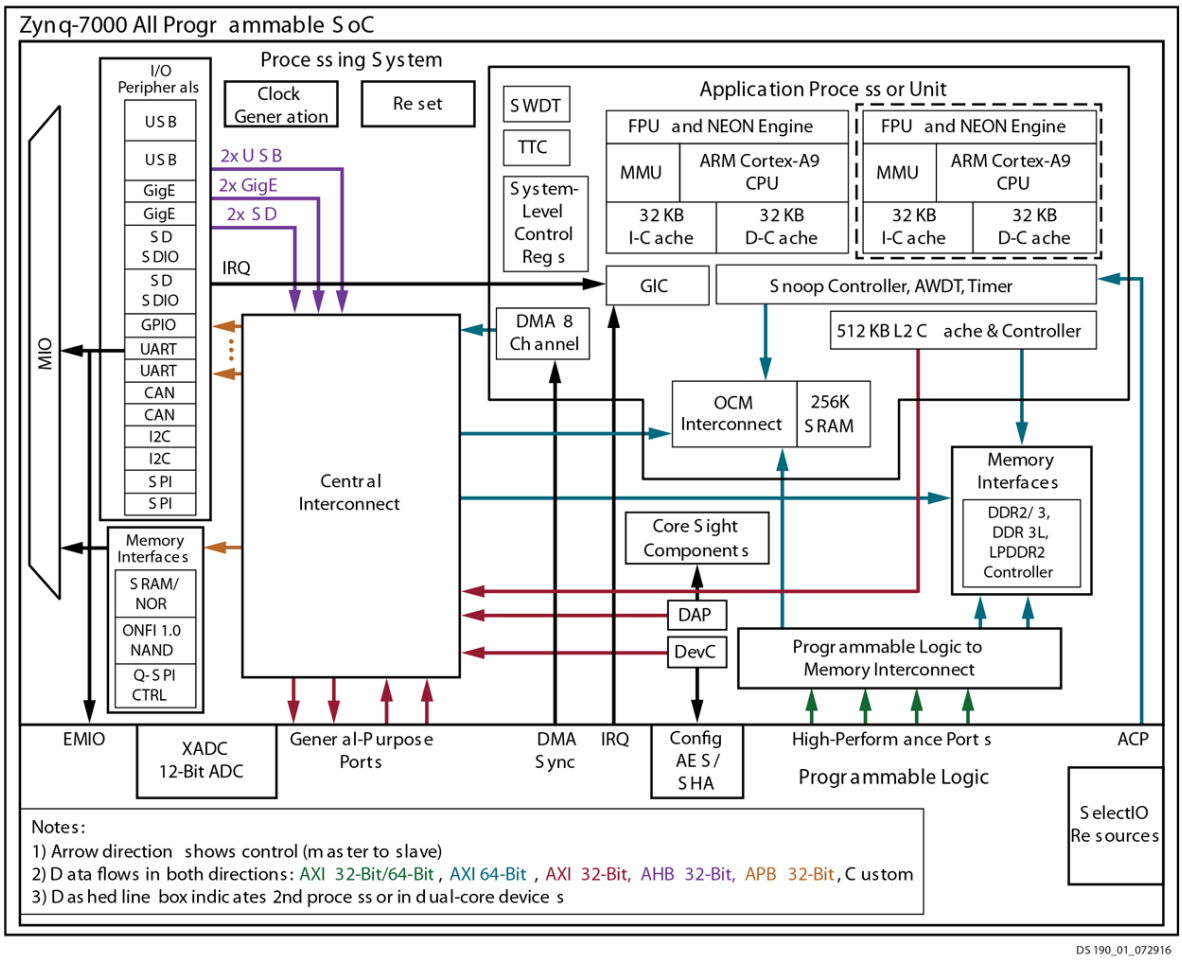
- Wymiary: 11,8 x 12,7 x 7,2 mm,
- Sensor: niechłodzony mikrobolometr VOx (tlenek wanadu),
- Rejestrowany zakres: fale długie podczerwieni, $8\mu m$ do $14\mu m$,
- Wielkość piksela: 17 μm,
- Rozdzielczość: 80x60 pikseli,
- Liczba klatek na sekundę: 8,6,
- Zakres rejestrowanych temperatur: -10 ° C 140 ° C (tryb wysokiego wzmocnienia),
- 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 wzmocnienia)) 24-bit RGB (z ACG i koloryzacją),
- Interfejs wideo: VoSPI (Video over Serial Peripherial Interface),
- Interfejs sterujący: CCI (zbliżony do I2C).

4.2. 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 4.2 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 przetworniki analogowo-cyfrowe.

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



Rys. 4.2. Schemat ogólny architektury układu Zynq-7000.

4.3. Interfejs AXI

AXI (ang. *Advanced eXtensible Interface* – zaawansowany 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 transmisji strumieniowych (wysokiej prędkości).

Specyfikacja interfejsu zakłada komunikację pomiędzy pojedynczym AXI master i pojedynczym AXI slave, która ma na celu wymianę informacji. 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.

4.4. Wykorzystanie AXI-Stream do transmisji sygnału wideo.

W odróżnieniu od klasycznej implementacji przetwarzania strumieniowego wideo, 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 (liczba pikseli w linii, liczba 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 osiem (16, 24, 48 itd.)

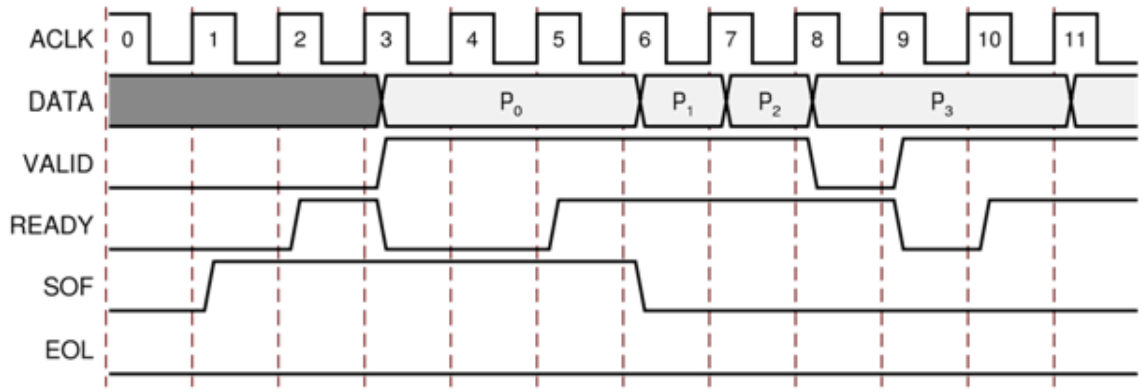
– Valid – linia określająca czy dane piksela są poprawne,

– 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 4.3



Rys. 4.3. Przykład rozpoczęcia transmisji Reday/Valid.

4.5. AXI VDMA

Wiele aplikacji wizyjnych wymaga przechowania całej ramki obrazu w celu jej dalszej obróbki np. podczas skalowania, przycinania bądź dopasowania liczby klatek na sekundę. Część programowalna układu Zynq zazwyczaj nie posiada wystarczającej liczby zasobów pamięciowych do przechowanie klatki obrazu w swojej strukturze. W tym celu jest wykorzystywany mechanizm bezpośredniego dostępu do pamięci, który pozwala na przesłanie i wczytanie danych z logiki programowalnej do pamięci RAM bez konieczności angażowania procesora. Realizuje się to poprzez IP-Core AXI VDMA. Zapewnia on przejście między interfejsem AXI4-Stream, a AXI4 Memory Map w obu kierunkach. Przed rozpoczęciem przesyłania IP-Core jest konfigurowany poprzez interfejs AXI4-Lite. Konfiguracja zawiera adres w pamięci RAM do którego ma być zapisana bądź wczytana ramka obrazu. Po wgraniu do pamięci ramki kontroler może wywołać przerwanie dla systemu procesorowego.

5. Realizacja

W celu detekcji pieszych, wykorzystany został połączony obraz termowizyjny (IR) i kolorowy (RGB) 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.

5.1. 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 ona na celu powiększenie i dopasowanie obrazu termowizyjnego, tak by poprawnie pokrywał się z obraz wizyjnym. W tym celu został zaimplementowany moduł, który oblicza na podstawie parametrów macierzy transformaty i współrzędnych 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 oczekuje na pojawienie się tego sygnału w strumieniu termowizyjnym. Do tego momentu wszystkie napływające piksele są odrzucane. Gdy pojawi się sygnał, strumień IR zostaje za-

trzymany i czeka na pojawienie się sygnał SOF w bazowym strumieniu wizyjnym. Po jego wykryciu strumień IR rusza. Oba strumienie zostają zsynchronizowane tworząc strumień wizyjny obrazu RGBIR. Następnie ten strumień zostaje przesłany do pamięci za pośrednictwem VDMA oraz (po koloryzacji i nałożeniu) wyświetlony na monitorze przez port VGA.

5.2. Kalibracja

Aby obraz termowizyjny poprawnie pokrywał się z obrazem RGB należy wykonać procedurę kalibracji. Kalibracja przeprowadzana jest ręcznie. Oprogramowanie kamery pozwala na zapisanie na karcie SD specjalnego obrazu kalibracyjnego, na którym jest zawarty zrzut aktualnie wyświetlanego obrazu wraz z nieprzetworzonym projekcyjnie obrazem IR. Następnie w pakiecie Matlab zostaje obliczona macierz transformaty projekcyjnej za pomocą wbudowanej funkcji. Wymaga ona wskazania 4 par odpowiadających sobie punktów na obrazie RGB oraz IR. Nową macierz można wgrać podając jej parametry w konsoli.

5.3. Wyznaczanie ROI

Strumień IR z kamery zostaje zbinaryzowany i poddany analizie w detektorze DPM korzystającym z wzorca probabilistycznego. 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. Zawiera ona koordynaty okna w układzie odniesienia kamery IR oraz wartość mocy dopasowania. Gdy zostanie zbadane ostatnie okno w obrazie, zostaje wysłany sygnał TLAST co wygeneruje przerwanie dla systemu procesorowego.

5.4. Klasyfikacja za pomocą HOG+SVM

Z lisy kandydatów wygenerowanej przez moduł DPM wybierany jest wynik o najwyższej mocy dopasowania. Koordynaty z układu odniesienia kamery zostają poddane transformacji 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.

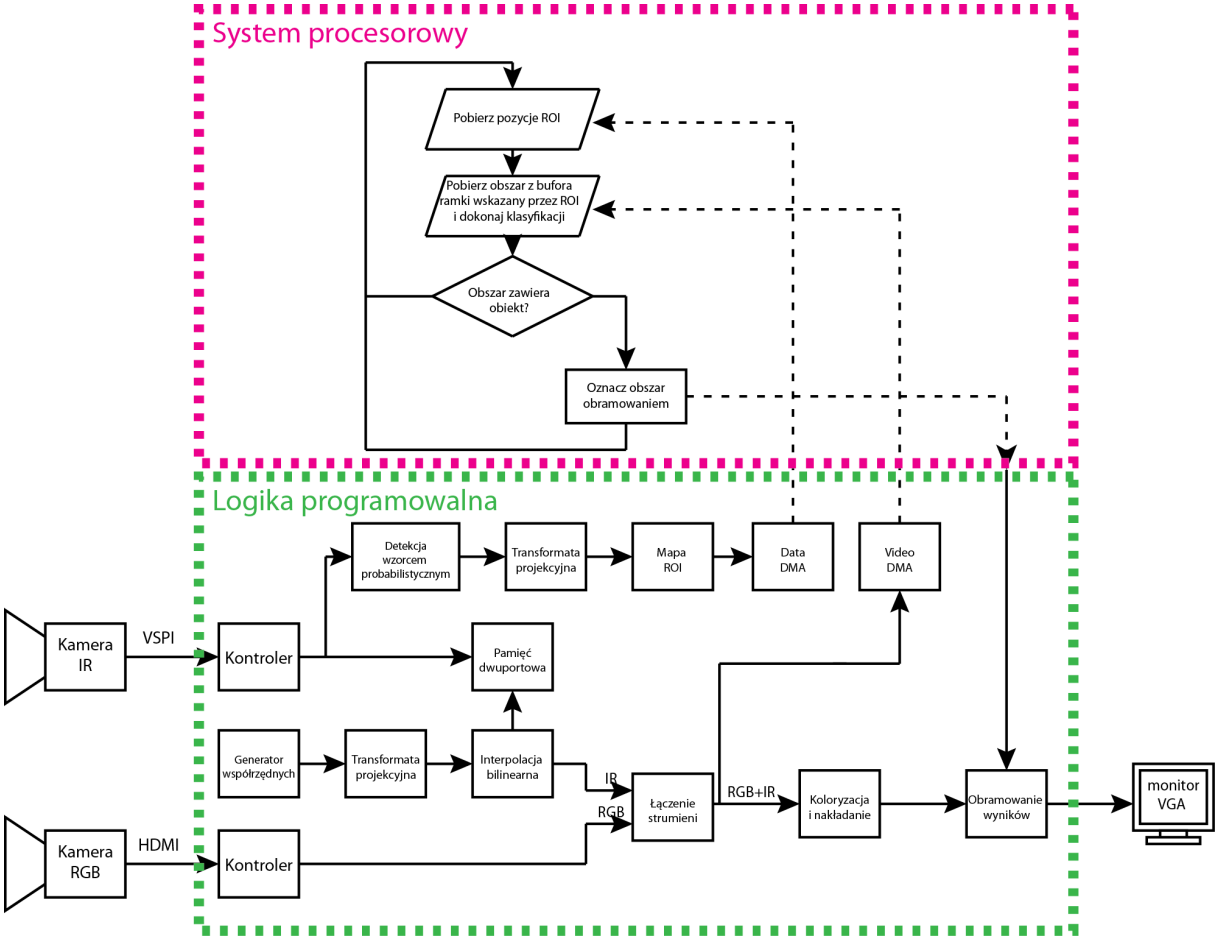
Klasyfikator został opracowany i nauczony na podstawie 60 wyselekcjonowanych obrazów. 30 z nich stanowiło próbką pozytywną zawierającą osobę, a 30 negatywną. Nauczanie zostało

zrealizowane przy użyciu oprogramowania Matlab. Próbki pozytywne zostały wygenerowane poprzez zapis ROI wyznaczonych przez wzorzec probabilistyczny.

5.5. 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 z rodziny Zynq-7000 firmy Xilinx, operacje zostały podzielone między logikę programowalną, a system procesorowy. Ogólny zarys rozwiązania został przedstawiony na rysunku 5.1.



Rys. 5.1. Schemat blokowy systemu detekcji.

Logika programowalna :

– 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.

5.6. Opis modułów

5.6.1. Kontroler kamery IR

Kontroler odpowiada za pobieranie obrazu z kamery IR poprzez interfejs VoSPI, który następnie zostaje zapisany do dwuportowej pamięci BRAM. Na początku pracy w stan niski ustawiany jest pin CS (ang. *Chip Select*), a po chwili rozpoczyna transmisję poprzez taktowanie zegarem SCK. Kamera reaguje na opadające zbocze zegara i wystawia kolejny bit danych na swoim porcie MISO. Strumień VoSPI składa się z 63 pakietów na ramkę obrazu. Pakiet rozpoczyna identyfikator składający się z numeru linii oraz sumy CRC pakietu (2 bajty na numer linii i 2 na sumę). Dane pakietu stanowi 160 bajtów – po dwa bajty na piksel w linii. Dane są przesyłane w 14-bitową wartość piksela oraz 2 zera wypełnienia. W przypadku niepoprawnej ramki numer identyfikator przyjmuje wartość xFxx. Ostatnie trzy pakiety stanowią telemetrię i są ignorowane.

5.6.2. Transformata projekcyjna

Moduł zamienia współrzędne z układu odniesienia kamery RGB odpowiadającym im na obrazie IR. Na wejściu podawany jest strumień AXI4-Stream zawierający timingi oraz 12 bitowe współrzędne X i Y. Moduł realizuję operację:

$$\begin{bmatrix} u_n & v_n & n \end{bmatrix} = \begin{bmatrix} x & y & 1 \end{bmatrix} T \tag{5.1}$$

$$u = \frac{u_n}{n} \tag{5.2}$$

$$v = \frac{v_n}{n} \tag{5.3}$$

Moduł wystawia na wyjściu strumień timingów, 12 bitowe wartości U i V oraz ich części ułamkowe w U_fraction i V_fraction (14 bitów). W module zostały wykorzystane 34 z 80 dostępnych w układzie Zynq procesorów DSP48 do wykonania operacji arytmetycznych. Najwięcej zasobów jest pochłonięte przez moduł dzielarki dostarczony od producenta układu. Do implementacji jednej dzielarki zostało wykorzystane 14 modułów DSP. Dzielenie nie odbywa się w pełni potokowo. Użyty w dzielarce algorytm High_Radix wymaga zatrzymania strumienia na czas obliczeń. Jednak dzięki zastosowaniu wyższej częstotliwości niż zegar pikseli obrazu RGB oraz bufora (250 MHz) nie stanowi to wąskiego gardła systemu. Macierz T jest zapisana w dziewięciu 32 bitowych rejestrach i konfigurowalna poprzez interfejs AXI4-Lite. Elementy macierzy są 25 liczbami w notacji stałoprzecinkowej: 1 bit znaku 10 – część całkowita, 14 – część ułamkowa.

5.6.3. Interpolacja bilinearna

Prosty moduł przeznaczony głównie do powiększania obrazów. Ma za zadanie pobrać z pamięci dwuportowej obrazu IR wartość piksela wskazaną na wejściu układu i wystawić na wyjście. Podobnie jak reszta systemu używa AXI4-Stream do przekazywania danych między poszczególnymi modułami. Dane na wejściu to współrzędne U i V oraz ich części ułamkowe U_fraction i V_fraction. Moduł został wyposażony w 4 rejestry, w których przechowywane są współrzędne oraz wartości 4 ostatnio użytych pikseli. Zabieg ten znacznie redukuje liczbę potrzebnych zapytań do pamięci. Podczas powiększania obrazów jest duża szansa, że kolejne koordynaty na wejściu UV odwołują się do tych samych czterech otaczających ich pikseli. W module jest sprawdzane, czy w pamięci są już wartości z koordynatów [U,V], [U+1,V] [U,V+1], [U+1,V+1]. Jeżeli któregoś piksela brakuje, jest on pobierany z pamięci i zapisywany w rejestrze przechowującym niepotrzebny piksel. Jeżeli wszystkie koordynaty się zgadzają , obliczana jest wartość piksela wyjściowego zgodnie ze wzorem (5.4).

$$Ir = A(1-U_f)(1-V_f) + BU_f(1-V_f) + C(1-U_f)V_f + DU_fV_f \tag{5.4}$$

gdzie: A, B, C, D odpowiadają wartościom pikseli w [U,V], [U+1,V] [U,V+1], [U+1,V+1], a Ir to wartość wyjściowa piksela wyjściowego. U_f i V_f stanowią U_fraction i V_fraction.

Moduł działa strumieniowo. W przypadku gdy jest wymagana aktualizacja rejestrów strumień jest wstrzymywany. biera 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.

5.6.4. Łączenie strumieni

Moduł posiada dwa wejścia dla obrazu. Jeden strumień jest głównym i do niego jest dołączany drugi strumień. Do synchronizacji została wykorzystana możliwość wstrzymania transmisji poprzez AXI4-Stream. 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. Po przejściu całej ramki strumienie są ponownie synchronizowane.

5.6.5. Koloryzacja i nakładanie

Strumień RGBIR zostaje połączony w jeden obraz. Obraz IR zostaje poddany koloryzacji na podstawie 12-bitowego LUT i nałożony w proporcjach 50 na 50 z obrazem RGB. Na wyjściu jest podany 24 bitowy strumień RGB.

5.6.6. Obramowanie wyników

Moduł dodaje do obrazu podanego na strumień wejściowy ramkę, która następnie jest podawana dalej strumieniem wyjściowym. Parametry ramki są ustawiane przez dwa 32 bitowe rejestry. Pierwszy (`position_reg`) zawiera pozycję, gdzie ma się znajdować ramka na obrazie (lewy górny róg ramki), drugi (`parameters_reg`) odpowiada za kolor i wielkość ramki. Rejestry są konfigurowane poprzez AXI4-Lite.

5.7. System procesorowy

System procesorowy spełnia dwa podstawowe zadania: konfiguracja modułów zawartych w logice programowalnej za pomocą interfejsu AXI4-Lite, takich jak macierz projekcji, wartość progu binaryzacji i wartość progu mocy dopasowania dla modułu DPM, wzmocnienie oraz offset modułu normalizacji sygnału IR. Pozwala on również na zapisanie na karcie SD aktualnej ramki bądź pozytywnie sklasyfikowanego obrazu okna detekcji, jak i obrazu do przeprowadzenia kalibracji.

Drugim zadaniem jest przeszukanie listy kandydatów w celu znalezienia tego z największą mocą dopasowania, wyliczenie cech HOG i klasyfikacji SVM. Oryginalny rozmiar okna detekcji w układzie kamery IR wynosi 16x40 zaś na obrazie RGBIR analizowane jest okno 80x192

piksele. Jest ono podzielone na 60 komórek o wielkości 16x16 pikseli. Następnie obliczane są gradienty oraz histogram dla każdej komórki. Wykorzystany jest histogram ważony o 9 przedziałach. Do każdego histogramu jest przypisana dodatkowo suma kwadratów wszystkich wartości przedziałów. Następnie komórki są łączone w bloki 2 na 2, w obrębie których dokonuje się normalizacji wykorzystując wcześniej obliczone sumy kwadratów. Bloki nakładają się na siebie dając w sumie 44 bloki. Suma histogramów z wszystkich bloków tworzy 1584 elementowy wektor cech. Wektor jest przemnożony przez wektor beta uzyskany w procesie nauczania SVM i dodany bias. Jeżeli uzyskany wynik jest większy od 0, badane okno zostaje sklasyfikowane z wynikiem pozytywnym.

6. Wyniki i wnioski

Aby sprawdzić działanie i dokładność systemu została zaimplementowana możliwość zapisu 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, pracuje 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.

Tabela 6.1. Wykorzystane zasoby logiki programowalnej.

Resource	Utilization	Available	Utilization %
LUT	12583	17600	71,49
LUTRAM	617	6000	10,28
FF	19924	35200	56,60
BRAM	25,50	60	42,50
DSP	36	80	45,00
IO	43	100	43,00
BUFG	7	32	21,88
MMCM	1	2	50,00
PLL	1	2	50,00

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] *LEPTON® Long Wave Infrared (LWIR) Datasheet*. Lepton. FLIR Commercial Systems. 2015.

[9] 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.

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

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

[12] 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.

[13] 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.