

Mateusz KOMORKIEWICZ, Marek GORGOŃAGH AKADEMIA GÓRNICZO-HUTNICZA, WYDZIAŁ EAIIB, KATEDRA AUTOMATYKI I INŻYNIERII BIOMEDYCZNEJ,
Al. Mickiewicza 30, 30-059 Kraków**Rekonfigurowalny moduł akwizycji i transmisji obrazów dla systemów monitoringu wizyjnego****Mgr inż. Mateusz KOMORKIEWICZ**

Autor jest absolwentem kierunku Automatyka i Robotyka na Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie (2010). Obecnie jest słuchaczem studiów doktoranckich na wydziale EAIIB tej samej uczelni. Interesuje się przetwarzaniem i analizą obrazów oraz sprzętową akceleracją algorytmów wizyjnych z wykorzystaniem układów FPGA.



e-mail: komorkie@agh.edu.pl

Dr hab. inż. Marek GORGOŃ

Autor jest absolwentem Akademii Górniczo-Hutniczej w Krakowie, gdzie uzyskał kolejno dyplom magistra inżyniera elektroniki (1988) oraz stopnie doktora (1995) i doktora habilitowanego (2007), oba w dyscyplinie Automatyka i Robotyka. Pracuje w Katedrze Automatyki AGH i zajmuje się metodami i systemami do przetwarzania i analizy obrazu. Od 1992 roku jego głównym obszarem zainteresowania są systemy obrazowe oparte na układach FPGA.



e-mail: mago@agh.edu.pl

Streszczenie

W artykule opisano działający w czasie rzeczywistym sprzętowy moduł do odbioru strumienia wizyjnego z kamery HDMI, zmiany rozdzielczości obrazu oraz dalszej jego transmisji przy wykorzystaniu sieci Ethernet (1 Gbps). Aby możliwe było ograniczenie kosztów oraz wykorzystanie zaprezentowanego modułu w urządzeniach wbudowanych, na żadnym etapie przetwarzania nie jest wykorzystywane buforowanie danych w zewnętrznej pamięci. W ramach prac zostało przebadane, jak parametry transmisji (rozmiar obrazu, maksymalny rozmiar pakietów) wpływają na jej przepustowość. Omówiono budowę każdego z modułów, zużycie zasobów FPGA całego systemu, rozpraszanie mocy, a także przykładowe rezultaty działania na płycie ewaluacyjnej SP605 firmy Xilinx.

Słowa kluczowe: transmisja obrazów, zmiana rozdzielczości, układy rekonfigurowane, systemy monitoringu wizyjnego.

A reconfigurable image acquisition and transmission module for video surveillance systems**Abstract**

Automated video surveillance systems are an important means of providing security. In projects such as SIMPOZ, INDECT or VIRAT the main tendency was to replace the human operator in a tedious task of video analysis. Because computer vision algorithms demand a lot of computational power, reconfigurable devices are often used for this type of applications. In the paper a module for video acquisition and transmission for a reconfigurable device is presented. It is the basic component of a reconfigurable based video surveillance system. An FMC card is used to allow FPGA to receive a video from the HDMI source (other FMC cards can be used if needed). In the next step, the image is streamed to module which scale it down. This operation is necessary to meet the bandwidth of transmission media and other modules processing capabilities. A hardware module provides Ethernet communication with 1 Gbps speed. Packet forming, checksum computation, ARP requests, IP and UDP protocols are realized in hardware using several finite state machines. The images or data obtained from analysis are transferred in UDP packets. The proposed system can process both grayscale and color images. The idea was verified using the Xilinx SP605 board with a low power Spartan 6 device.

Keywords: image transmission, image resizing, reconfigurable devices, video surveillance systems.

1. Wstęp

Automatyczne systemy monitoringu wizyjnego wykorzystywane są w celu zapewnienia bezpieczeństwa w chronionych obiektach. W projektach badawczych takich jak SIMPOZ [6], VIRAT [2] lub INDECT [4] głównym celem było stworzenie algorytmów, które pozwoliłyby na zastąpienie człowieka w żmudnym procesie analizy obrazu dostarczanego przez kamery. W rezultacie powstały systemy, które w sposób automatyczny pozwalają monitorować chronione strefy i wykrywać potencjalnie niebezpieczne sytuacje,

między innymi takie jak pozostawienie przedmiotu albo ruch osób w nieprawidłowym kierunku.

Okazało się jednak, że algorytmy wykorzystywane w systemach monitoringu wizyjnego wymagają dużej mocy obliczeniowej, której nie zawsze są w stanie dostarczyć procesory oraz karty graficzne. W związku z tym pojawił się pomysł, aby wykorzystywać układy rekonfigurowalne do realizacji tych zadań. Jednak, aby zastosować je do tego celu, oprócz implementacji algorytmów przetwarzania i rozumienia obrazów, wymagane są także moduły, które umożliwią komunikację z resztą systemu (najczęściej poprzez sieć Ethernet). Muszą one umożliwić odbiór strumienia wizyjnego z kamery, odpowiednie jego przeskalowanie oraz możliwość transmisji zarówno odebranego obrazu jak i danych obliczonych przez system (np. obszary w których stwierdzone zostały alarmy itd.).

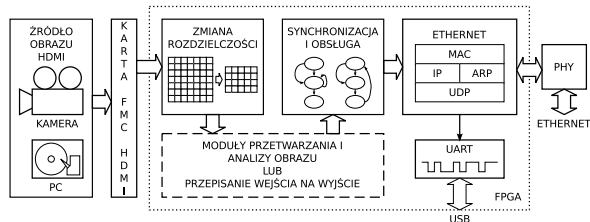
W dostępnych na rynku inteligentnych kamerach, obraz jest przesyłany głównie w postaci mocno skompresowanej (kompresja MJPEG lub H.264). Nie zawsze mogą one być jednak stosowane w omawianych wcześniej automatycznych systemach monitoringu. Jeśli kamera inteligentna nie realizuje wszystkich operacji, a jest jedynie elementem w większym rozproszonym systemie obliczeniowym, nie wszystkie dane mogą być w niej kompresowane. Bardzo niekorzystna byłaby np. kompresja przepływu optycznego, który najczęściej jest zapisany tylko na dwóch kanałach, za to z 32 bitową precyzją zmiennoprzecinkową. Jedno z nielicznych rozwiązań, które spełniają te wymagania, jest oparte o standard GigE Vision [1]. Posiada ono jednak dwie zasadnicze wady. Po pierwsze jest rozwiązaniem zamkniętym (opłata za korzystanie), po drugie narzuca ono dość skomplikowany format transmisji. Nie jest więc proste w implementacji na układach reprogramowalnych.

Naukowcy zaproponowali więc inne rozwiązania. W pracy [10] został omówiony system umożliwiający transmisję obrazów w sieci Ethernet (tylko 100Mbps) przy wykorzystaniu układu reprogramowalnego z układem Virtex 4 oraz procesorem PowerPC. Całkowicie sprzętowy stos został zaprezentowany w pracach [7] i [9]. W roku 2012 zostało zaprezentowane rozwiązanie [12] wykorzystujące strukturę typu SoC do odbioru sygnałów telewizji cyfrowej (DVB-T) i ich dalszej transmisji po sieci Ethernet. Zaprezentowane rozwiązanie umożliwiło transmisję skompresowanych obrazów o rozmiarach 352x288 pikseli z prędkością 30 klatek/s (bitrate na poziomie 458kbps).

W niniejszym artykule opisano całkowicie sprzętowy system, umożliwiający odbiór i transmisję obrazów. Zaproponowane rozwiązanie wyróżnia się wysoką przepustowością (na poziomie 60 MB/s). Dzięki mechanizmowi synchronizacji nadajnika z odbiornikiem nie wymaga używania zewnętrznej pamięci do buforowania obrazów. Modułowemu rozwiązaniu pozwala na dodawanie kolejnych bloków do systemu (analiza obrazów, kompresja itd.). Całość została również przetestowana na karcie ewaluacyjnej z układem FPGA Spartan 6.

2. Zaproponowany system i jego implementacja sprzętowa

Schemat blokowy zaproponowanego rozwiązania został przedstawiony na rys. 1. Jako źródło obrazu może zostać wykorzystany dowolny nadajnik zgodny ze standardem HDMI [3]. Do jego odbioru przez kartę ewaluacyjną Xilinx SP605 [5] jest wykorzystywana karta rozszerzeń standardu FMC. Dzięki temu, poprzez wymianę karty rozszerzeń możliwe jest wprowadzenie sygnału wideo w innym standardzie (np. analogowym). Dalsze etapy przetwarzania zostały zrealizowane w układzie FPGA serii Spartan 6 XC6SLX45T. W pierwszej kolejności odbiera on i buforuje obraz z karty FMC. W kolejnym etapie dostosowywana jest jego rozdzielczość od wydajności modułów przetwarzania oraz przepustowości sieci. Dalej obraz może być dostarczony do bloków przetwarzania i analizy obrazu (w rzeczywistym systemie) lub przekazany bezpośrednio na wejście kolejnego bloku (stosowane do testów systemu). Blokiem tym jest blok synchronizacji i obsługi. Jego zadaniem jest odbieranie pakietów UDP z informacją o żądanej liczbie ramek do przesłania, synchronizacja strumienia wizyjnego do kanału transmisji (rozpoczęcie transmisji dopiero gdy zacznie się nowa ramka) oraz odpowiednie podzielenie obrazu na pakiety.



Rys. 1. Schemat systemu akwizycji i transmisji obrazów
Fig. 1. Schematic of image acquisition and transmission system

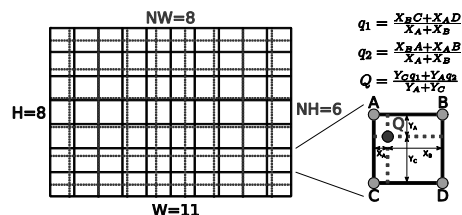
Dodatkowo wykorzystywany jest moduł UART, który umożliwia podgląd takich informacji jak numery portu czy adres IP klienta do którego przesyłane są dane (mogą się zmieniać w zależności od przesłanego pakietu). Moduł ten może również służyć do podglądu wszystkich odebranych pakietów IP. Ostatni z bloków implementuje wewnątrz układu reprogramowalnego funkcjonalności wymagane do obsługi transmisji Ethernet.

3. Zmiana rozdzielczości obrazu

Zmiana rozdzielczości obrazu, o skalę różną niż potęga liczby dwa, w systemie potokowym, gdzie nie ma możliwości jego buforowania w pamięci, jest wyzwaniem. Rozwiązania wykorzystujące interpolację dwuliniową były prezentowane w pracach [11] i [13]. W artykule [14] został zaprezentowany system, który umożliwiał tworzenie tzw. piramidy obrazów w celu umożliwienia wykrywania obiektów na obrazie w różnych skalach. W pracy [8] został zaprezentowany system typu SoC (magistrala AXI) do obsługi wyświetlaczy wysokiej rozdzielczości z interpolacją bikubiczną.

Bazując na tych pracach, w opisywanym rozwiązaniu zdecydowano się na wykorzystanie dwuliniowej metody interpolacji. Dodatkowo podczas projektowania założono, że rozdzielczość obrazu odbieranego z kamery może być jedynie zmniejszana. Obserwacja ta wynika z faktu, że interfejs HDMI ma znacznie większą przepustowość niż Ethernet (1 Gbps). W związku z tym, kamera zawsze będzie w stanie dostarczyć obraz w większej rozdzielczości niż wydajność medium transmisyjnego.

Zasada interpolacji dwuliniowej została przedstawiona na rys. 2. Jeśli przedstawimy obraz jako siatkę pikseli, których środek leży na przecięciu się linii, to na obraz o oryginalnej szerokości ($w=11$) i wysokości ($h=8$) może być nałożona nowa siatka (linie przerywane), która odpowiada rozmiarowi nowego obrazu. Wartości pikseli dla nowego obrazu są wyznaczane jako odpowiednio zważone wartości czterech najbliższych mu pikseli na oryginalnym obrazie.



Rys. 2. Zmiana rozdzielczości obrazu
Fig. 2. Image rescaling

Aby umożliwić potokowe wykonywanie tej operacji w układzie FPGA, wykorzystane zostało buforowanie kontekstu 2x2 aktualnego obrazu. Po analizie przykładu na rys. 2, można zauważyć, że:

- jeśli obraz jest zmniejszany, nie jest możliwe, aby środki nowych pikseli wystąpiły dwukrotnie pomiędzy środkami oryginalnych pikseli,
- jest możliwe, że pomiędzy dwoma pikselami oryginalnego obrazu nie wystąpi żaden piksel nowego obrazu.

W związku z tym została zaprojektowana logika, która ustala, czy pomiędzy pikselami z aktualnego kontekstu oryginalnego obrazu znajduje się środek pikseli z nowego obrazu. Jeśli ten warunek jest spełniony, obliczana jest jego wartość na podstawie wzorów z rys 2. Aby nie wykonywać trzech dzieleń (dla q_1, q_2 i Q), najpierw obliczana jest wartość liczników, a na samym końcu następuje dzielenie przez mianownik:

$$Q = \frac{Y_C(X_B C + X_A D) + Y_A(X_B A + X_A C)}{(X_A + X_B)(Y_A + Y_C)} \quad (1)$$

Dzięki temu wykorzystywana jest tylko jedna dzielnarka. Przy czym można zauważyć, że wartość mianownika jest stała (odległość pomiędzy punktami siatki nie ulega zmianie).

4. Moduł transmisji Ethernet

Wstępne badania pokazały, że z uwagi na dużą ilość danych, która musi być przesyłana przy transmisji wideo, wykorzystanie rozwiązań typu SoC (np. programowy stos IP na procesorze Microblaze) powoduje obniżenie maksymalnej prędkości transmisji. Jest to spowodowane tym, że procesor wykonuje wiele operacji w celu sformowania transmitowanych pakietów danych.

W opisywanym rozwiązaniu zdecydowano się więc na rozwiązanie całkowicie sprzętowe. W celu przesyłania danych protokołem Ethernet zaprojektowano moduł, który realizuje część warstw modelu OSI. Zaimplementowane zostały warstwa łącza danych, formowanie i odbieranie pakietów, dodawanie i sprawdzanie sumy kontrolnej oraz adresów MAC. Obsługiwany jest także protokół ARP (adres resolution protocol). Realizowana jest również warstwa sieciowa (protokół IPv4).

W warstwie transportowej wykorzystany został protokół UDP (user datagram protocol). Pomimo tego, iż nie zapewnia on kontroli przepływu i pewne pakiety mogą zostać zgubione, z uwagi na jego prostotę oraz szybkość jest powszechnie wykorzystywany do transmisji dużej ilości danych szczególnie multimedialnych. Został zaplanowany następujący model komunikacji. Urządzenie chcące odebrać dane z układu FPGA, wysła do systemu zdefiniowany pakiet UDP, w którym zawarte są informacje takie jak typ ramki, rozmiar oraz liczba ramek które mają być przesłane. Dane te są odbierane przez moduł synchronizacji i obsługi, który transmituje żadaną liczbę ramek pod odpowiedni adres. Jeśli urządzenie odebrało mniej danych niż powinno, oznacza to że nastąpił błąd transmisji i aktualny obraz powinien być odrzucony, a transmisja powtórzona. Z uwagi na uproszczony protokół, karta musi być bezpośrednio połączona z komputerem (bez węzłów pośredniczących).

5. Rezultaty i wnioski

Cały system został zrealizowany w formie modułów sprzętowych opisane w języku Verilog przy wykorzystaniu oprogramo-

wania Xilinx ISE 14.4 Design Suite. Zużycie mocy wyznaczone za pomocą narzędzia Xilinx XPower Analyzer wyniosło 0,354 W. Wykonano również pomiary zużycia mocy dla płyty SP 605: jeden bez skonfigurowanego układu FPGA (9,516 W), drugi z uruchomionym systemem (9,756 W). Zatem faktyczne zużycie mocy przez układ FPGA wyniosło ok. 0,24 W. Zużycie zasobów dla wersji obsługującej obrazy kolorowe zaprezentowano w tab. 1.

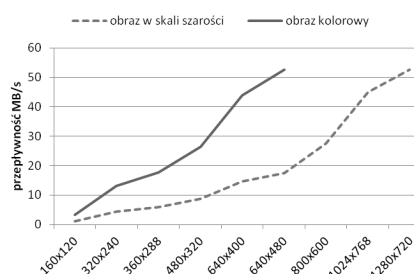
Tab. 1. Zużycie zasobów FPGA
Tab. 1. Utilisation of FPGA resources

Zasób	Wykorzystane	Dostępne	Procent
FF	4614	54576	8%
LUT6	4740	27288	17%
SLICE	1652	6822	24%
BRAM_18	14	116	12%
DSP_48	33	58	56%

Maksymalna przepustowość sieci Ethernet 1 Gbps wynosi 125 MB/s. Osiągnięcie takiej przepustowości przy przesyłaniu danych jest jednak niemożliwe, ponieważ redukują ją narzuty kolejnych protokołów oraz transmisji (minimalne odstępy pomiędzy pakietami itd). W zależności od parametrów strumienia wideo, przesłanie obrazu wymaga następującej przepustowości sieci:

$$Q = c \cdot p \cdot f \cdot I_w \cdot I_h \quad (2)$$

gdzie: c jest liczbą kanałów (1-skala szarości, 3-obraz kolorowy RGB), p – liczbą bajtów potrzebnych do zapisania wartości kanału, f – liczbą klatek przesyłanych na sekundę a I_w i I_h wysokością i szerokością obrazu.



Rys. 3. Przepływność systemu w zależności od rozmiaru obrazu (60 klatek/s)
Fig. 3. Bitrate for different image size (@ 60 fps)

Z uwagi na przyjętą architekturę, która nie wykorzystuje pamięci RAM do buforowania ramek obrazu, każda ramka musi być przesłana w dostępnym dla niej oknie czasowym (dla 60 klatek na sekundę jest to 16,6 ms).

Przy wykorzystaniu prostego oprogramowania wykorzystującego funkcje Socket oraz bibliotekę OpenCV, przebadano działanie systemu. Na rys. 3 został przedstawiony wykres, który obrazuje osiągniętą przepływność dla rozważanego systemu w przypadku obrazów kolorowych i w skali szarości o różnych rozdzielczościach przesyłanych 60 klatek na sekundę.

Tab. 2. Maksymalna przepływność systemu (obrazy kolorowe) 60 klatek/sek.
Tab. 2. Maximum system bitrate for color images @ 60 fps

Rozmiar pola danych (bajty)	Maksymalna rozdzielczość obrazu	Przepływność
600	720 x 300	41 MB/s
900	720 x 420	52 MB/s
1200	720 x 490	61 MB/s

Maksymalne uzyskane przepływności oraz odpowiadające im rozdzielczości, w zależności od zastosowanego rozmiaru pakietu zostały przedstawione w tab. 2. Przykładowy rezultat działania systemu zaprezentowano na rys. 4.



Rys. 4. Działający system: kamera, karta oraz obraz odbierany przez komputer
Fig. 4. The working system: camera, board and received video stream

6. Podsumowanie

Omówione w pracy moduły sprzętowe mogą zostać wykorzystane do budowy tzw. kamer inteligentnych. Maksymalna uzyskana przepływność wyniosła ponad 60 MB/s. System może przesyłać wyniki np. generacji tła lub obliczenia stereowizji bez utraty jakości spowodowanej kompresją.

Prace były finansowane ze środków na działalność statutową Katedry Automatyki i Inżynierii Biomedycznej (11.11.120.612).

7. Literatura

- [1] GigE Vision video streaming and device control over Ethernet standard version 2.0, AIA, Ann Arbor, Michigan, November 2011.
- [2] Broad Agency Announcement Video and Image Retrieval and Analysis Tool BAA 08-20 03, DARPA, Arlington Virginia, March 2008.
- [3] High-definition multimedia interface specification version 1.3, HDMI Licensing LLC, Sunnyvale, California, June 2006.
- [4] INDECT description of system architecture 2.2, The INDECT Consortium, October 2010.
- [5] SP 605 Hardware User Guide UG526 v1.8, Xilinx Inc., Sunnyvale, California, September 2012.
- [6] Chmiel W., Kwiecień J., Mikrut Z.: Realization of scenarios for video surveillance, *Automatyka* (3), 2012.
- [7] Herrmann F. L., Perin G., de Freitas J.P.J., Bertagnolli R. and dos Santos Martins J.B.: A Gigabit UDP/IP Network Stack in FPGA, 16th IEEE International Conference on Electronics, Circuits, and Systems ICECS 2009, pp.836-839, December 2009.
- [8] Hongbin Sun, Longjun Liu, Qiubo Chen, Baolu Zhai, Nanning Zheng: Design and implementation of a video display processing SoC for full HD LCD TV, International SoC Design Conference (ISOCC), pp.297-300, 4-7 November 2012.
- [9] Khalilzad N.M., Yekeh F., Asplund L., Pordel M.: FPGA implementation of real-time Ethernet communication using RMI interface, IEEE 3rd International Conference on Communication Software and Networks (ICCSN), pp.35-39, May 2011.
- [10] Llamocca D., Pattichis, M., Vera G.A.: A dynamic computing platform for image and video processing applications, Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp.412-416, November 2009.
- [11] Perez J.M., Sanchez P., Martinez M.: Low-cost Bayer to RGB bilinear interpolation with hardware-aware median filter, 16th IEEE International Conference on Electronics, Circuits, and Systems ICECS 2009, pp.916-919, December 2009.
- [12] Salman A.H., Adiono T., Cahyadi W.A., Kurniawan Y.: SOC design and FPGA implementation of Digital TV receiver, 7th International Conference on Telecommunication Systems, Services, and Applications (TSSA), pp.125-129, October 2012.
- [13] Villalpando C.Y., Morfopolous A., Matthies L., Goldberg S.: FPGA implementation of stereo disparity with high throughput for mobility applications, IEEE Aerospace Conference, pp.1-10, March 2011.
- [14] Zemčík P., Přibyl B., Žádník M., Korček P.: Fast and Energy Efficient Image Processing Algorithms using FPGA, Workshop of 21th Conference on Field Programmable Logic and Applications, Chania, 2011.