

Design and implementation of Pedestrian detection based on Zynq APSoC

Wei Chen

Southwest Minzu University, Chengdu 610225, China

Abstract: Pedestrian recognition is an important content of intelligent security surveillance video processing system. In order to meet the application requirements of miniaturization and real-time security system, a moving target recognition system based on APSoC platform is designed. The system uses ov5640 camera as the video collector, zynq7020 as the development platform, and uses the soft and hard co-design to realize each function module. Finally, the pedestrian image acquisition and real-time display system is successfully completed. The test results show that the system can effectively and stably track pedestrians and moving objects within a certain range of distance, and can be displayed in real time, miniaturized and low power consumption, which can be further applied in the field of security.

Keywords: APSoC; Zynq; Pedestrian recognition; Soft and hard co-design.

1. Introduction

With the enhancement of security industry and public security awareness, security market scale is expanding rapidly, and its application fields can be seen everywhere. Security supervision and control domain has also entered the period of digitalization, high-definition and networking at the technical level, among which pedestrian recognition is an important content in the intelligent security surveillance video processing system. Although the previous ARM-based digital image processing system can meet the complex image processing needs, it is difficult to obtain the bandwidth and real-time guarantee because of the low sampling rate of the processor. ARM+FPGA heterogeneous processing framework can take into account both of them well. This paper designs a framework for real-time video data acquisition, processing, control and output based on Xilinx company's ZYNQ series ARM+FPGA heterogeneous structure as the processing core. Combining the powerful driving capability, high flexibility and configurability of the dual-core ARM-Cortex-A9 processor with the parallel, high speed and reprogramming features of FPGA, the structure of both soft and hard reprogramming is realized.

2. Methodology

2.1. System principle

The system uses the ZYNQ series XC7Z020 as the main control, the OV5640 camera as the video data source, and includes a display with HDMI HD interface support. The input and output channels and hardware algorithm modules of video data are set up. The overall block diagram is shown in Figure 1. PL hardware part is responsible for video input and hardware calculation

Method design, video output and other parts. PS processor part to realize the control of the whole system, including the hardware initialization process, and through AXI4 bus to control the image processing algorithm module. PL communicates with PS through the Advanced eXtensible Interface (AXI) protocol to enable software and hardware to work together.

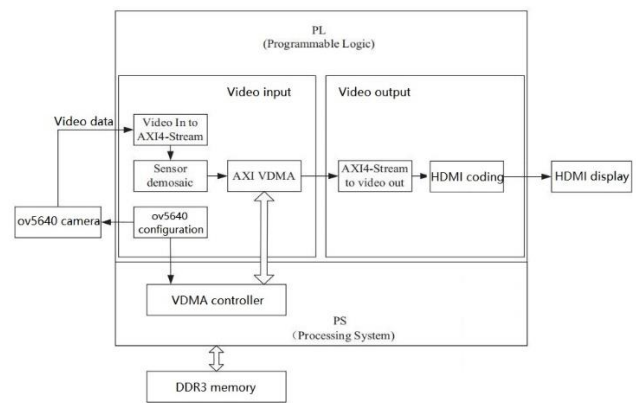


Figure 1. System block diagram

From video data acquisition, transmission and processing in the system to the display Angle, the system can be divided into the following main modules according to functions:

(1) Video data acquisition module: The system collects external video data through a 5-megapixel CMOS camera and inputs it into the PL hardware part to complete data acquisition. The camera initialization is completed by the PL part of the IIC timing controller. The initialization can be completed within 2 seconds after power-on, and the video stream data can be output normally.

(2) Video input module: After the RAW data of the camera is imported, the data format is converted into AXI-stream format through Video in to AXI4StreamIP and into the demosaic module. The video color is interpolated and restored. At the same time, the single pixel data is converted from RAW to RGB888 format. Finally, the Video Direct Memory Access (VDMA) is cached to DDR3 to complete the video input path construction.

(3) visual frequency output module: first through AXI4Stream to Video out image data by line read from the DDR memory, and then

After the hardware algorithm processing module, the video data is pipelined, and the processed video data is output in real time. After the HDMI encoder, the standard VGA sequential protocol video is encoded into HDMI differential signal for output.

2.2. Hardware module implementation

2.2.1. Introduction to OV5640

Image sensor is the core component of camera. OV5640 camera is a kind of CMOS digital image sensor. The sensor supports the output of images up to 5 million pixels (2592x1944 resolution), supports the use of VGA sequential output image data, Support DVP (DC), MIPI interface output image data format support YUV (422/420), YCbCr422, RGB565, RAW and JPEG format, if directly output JPEG format image can greatly reduce the amount of data, convenient network transmission. It can also be acquired image compensation, support curve, white balance, saturation, chroma and other basic processing. According to different resolution configurations, the frame rate of the sensor output image data can be adjusted from 15-60 frames, and the working power is between 150mW and 200mW. OV5640 uses SCCB bus configuration, which is compatible with I2C bus, and is configured as RGB565 output in the development board. Because the data bus is 8Bit, the data of a pixel is transmitted in 2 clock cycles, and then spliced into complete pixel data at the FPGA receiving end.

2.2.2. The use of VDMA

VDMA is a key IP in xilinx video processing, and VDMA is a special DMA with special design for video processing. As shown in the figure below, we see that VDMA has an AXI4 Memory Map interface for reading and writing video data to memory, an AXI4-Lite interface for reading VDMA status and configuring VDMA parameters, and an AXI4-Stream interface, For video input and output.

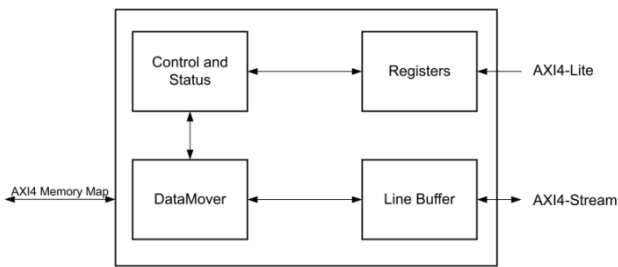


Figure 2. VDMA interface diagram

2.2.3. Block Designer project drawing

The following is a Block Designer project drawing created using Vivado IP integrator, in which the core modules are alinx ov5640, video timing controller, axi_vdma, axi4-stream to video out, they correspond to acquisition, processing, caching and display.

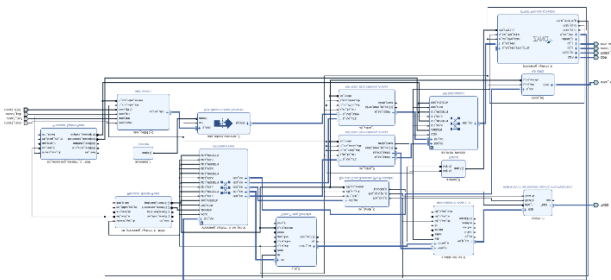


Figure 3. BD engineering drawing

2.3. SDK software part design

The SDK needs to complete the following tasks: (1) Initialize EMIO; (2) Initialize the OV5640 camera; (3) Configure VDMA; (4) Initialize the display timing controller;

(5) Set the display mode. Figure 5 shows the SDK application design flow.

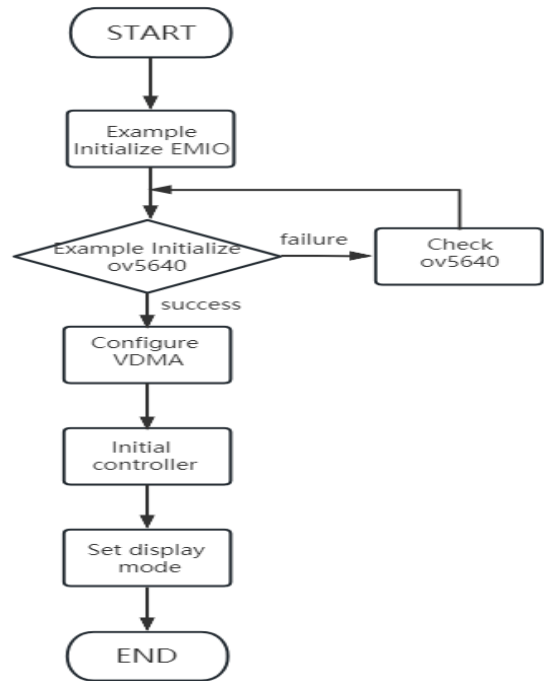


Figure 4. SDK programming flow chart

All of the work in the SDK is done by calling library functions in the BSP (Board Support Package), which is provided according to the previous Vivado hardware design section. First, the two EMIO pins on Zynq are initialized to be used as SCCB communication pins. Since only register write to OV5640 is required here, both the clock line and data line are set as output pins. This encloses an emio_init function for initializing the EMIO pin; Then initialize the OV564 camera by calling a wrapped ov5640_init function, first check the ov5640 connection is correct, if the check passes then call sccb_write_reg16 function, write different register values to the ov5640 register address. Here the OV5640 is configured as a DVP (Digital Video Port) output interface, 1 280x720 resolution and 30FPS (Frame Per Second). The VDMA is then configured by calling a wrapped run_vdma_frame_buffer function, which includes the configuration of the frame cache address, the write channel configuration, the read channel configuration, the open write channel configuration and the open read channel configuration. VDMA is configured with 3 frame image cache capability. Initialize the Display Initialize function of the visual Timing Controller, including VTC (Video Timing Controller) and dynamic clock configuration. Set Display mode and Start Display image by Display SetMode function, start VTC by display start function to start work output image to HDMI interface.

3. Results and discussion

After the system design and simulation was completed on vivado platform and modelsim simulator, the project was integrated, realized and generated bit stream files, and the hardware was exported to SDK environment for ps terminal program development. Then create the FSBL (First Stage Boot Loader) project and solidify the C program in the SDK. After starting the system and waiting for the completion of the initialization process, the display receives 30 frames of video data at 1280x720 resolution from the system through

the HDMI cable. The following figure shows the pedestrian display results collected by the ov5640 camera.



Figure 5. Hardware connection diagram

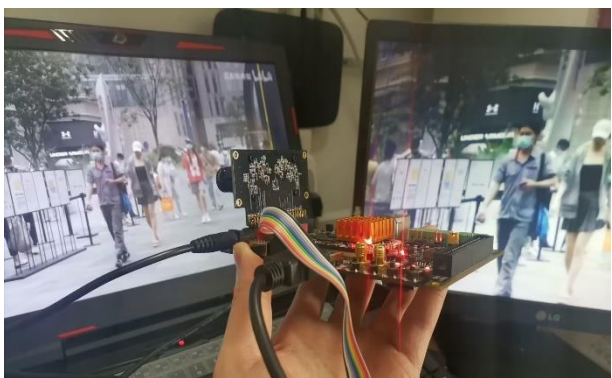


Figure 6. Pedestrian detection diagram

4. Conclusion

This paper proposes and implements a real-time video acquisition, processing and display platform based on ZYNQ APSoC chip, which fully combines the flexibility of ARM with the high real-time performance of FPGA in image processing field. The platform has the advantages of low delay, easy replacement and extension, which provides a good basic platform for the realization of complex image

processing algorithms. The system has high real-time performance and stable performance. It can be applied to dynamic pedestrian detection, video front-end processing and other fields.

References

- [1] Bobick A F, Andy Wilson. Using configuration states for the representation and recognition of gestures. MIT Media Lab Perceptual Computing Section Technical Report, No. 308, 1995.
- [2] Bobick A F, Wilson A D. A state-based approach to the representation and recognition of gesture [J] .IEEE Trans PAMI, 1997, 19(12): 1325-1337.
- [3] Maher M W, Marais M L. A Field Study on the Limitations of Activity-Based Costing When Resources are Provided on a Joint and Indivisible Basis[J]. Journal of Accounting Research, 1998, 36(1):129-142.
- [4] Wei S E, Ramakrishna V, Kanade T, et al. Convolutional Pose Machines[J]. 2016:4724-4732.
- [5] He, K., Gkioxari, G., Doll'ar, P., Girshick, R.: Mask R-CNN[C]. International Conference on Computer Vision, 2017: 2961-2969.
- [6] Fang, H., Xie, S., Tai, Y., Lu, C. RMPE: Regional Multi-Person Pose Estimation[C]. International Conference on Computer Vision,2017:2334-2343.
- [7] Yilun C., Zhicheng W., Yuxiang P., Zhiqiang Z. Cascaded Pyramid Network for Multi-Person Pose Estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018:7103-7111.
- [8] Ren S, Girshick R, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [9] Cao, Z., Simon, T., Wei, S.E., Sheikh, Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields[C]. IEEE Conference on Computer Vision and Pattern Recognition,2017:7291-7299.
- [10] Newell, A., Huang, Z., Deng, J. Associative Embedding: End-to-End Learning for Joint Detection and Grouping[C]. Neural Information Processing Systems, 2017.