

Image Signal Processor for Artificial Intelligence

Karthick Kumaran Ayyalluseshagiri Viswanathan¹, Karthik Poduval²

Member, IEEE

Abstract: *Cameras have become very common in most consumer electronics and industrial products like mobile phones, augmented/virtual reality headsets, tablets, robotics, drones, automobiles, door cameras. There are two types of cameras in most consumer electronic products, one used for human consumption which requires high quality visually appealing images and another used for computer vision applications powered by various deep learning models. High quality visually appealing images for human consumption have to undergo rigorous image signal processing for improving the quality of the captured image or video whereas the images produced for the computer vision tasks doesn't have to be enhanced in quality as the algorithms can be trained to operate without any image enhancement. In this paper, we will introduce the lightweight ISP for the various computer vision tasks and the benefits of that.*

Keywords: Cameras, ISP, Computer vision, Deep Learning, DMA, MIPI, CSI, Artificial Intelligence

1. Introduction

Many of the consumer electronics products that are selling in the market now have cameras for some use cases. Cameras have become an important peripheral in the various consumer electronics and industrial products. Mobile phones now have more than two cameras in general, one front camera and two or more rear cameras for main, Ultra-wide, telephoto captures. Augmented/virtual reality, robotics, automotive products includes multiple cameras for environment tracking for SLAM, object detection, object recognition for computer vision applications and typical color image capture for human vision consumption. Camera frames produced by the Image Signal Processor (ISP) for human consumption have to be visually appealing to differentiate the product from competitors. Image Signal Processor has various hardware blocks for enhancing the image quality as shown in Fig.2 and the most notable ones are dead pixel correction, black level compensation, gamma correction, color space conversion, edge enhancement, noise filter etc., These hardware blocks take significant amount of die size in the SoC (System-on-a-chip) and it also consumes power when the cameras are streaming.

2. ISP

ISP stands for Image Signal Processor. Cameras are connected to one of the camera interfaces like MIPI, USB,

Parallel bus. Cameras in mobile phones and other latest consumer electronic products like drones, AR/VR headsets use MIPI CSI interface to transfer pixels. Color cameras output Bayer pattern data to the SoC over the MIPI CSI interface. SoC receives the camera data in the CSI PHY interface and decodes the MIPI packets with the MIPI CSI Decoder as shown in Fig.1. This decoded data can either be dumped into the DRAM or processed by the ISP. Major blocks in the ISP hardware blocks are shown in Fig.2. Bayer pattern data goes through the image processing pipeline as shown in Fig. 2. Sensor linear output converted into non-linear form by the gamma correction block for better visual looking. Then the bayer pattern output is converted into the RGB color space domain, a process known as demosaicing. Then the pixels are passed to the noise filtering hardware blocks to enhance the visual quality of the captured images. Then finally the image will be converted from RGB to YUV color space as the other hardware blocks like the video encoder, display work in the YUV color domain. YUV data will then be dumped to the image buffers in the DRAM by the DMA (Direct Memory Access). This buffer will then be shared with the video encoders for compression, computer vision algorithms for object detection & recognition, graphics engine for rendering on the display.

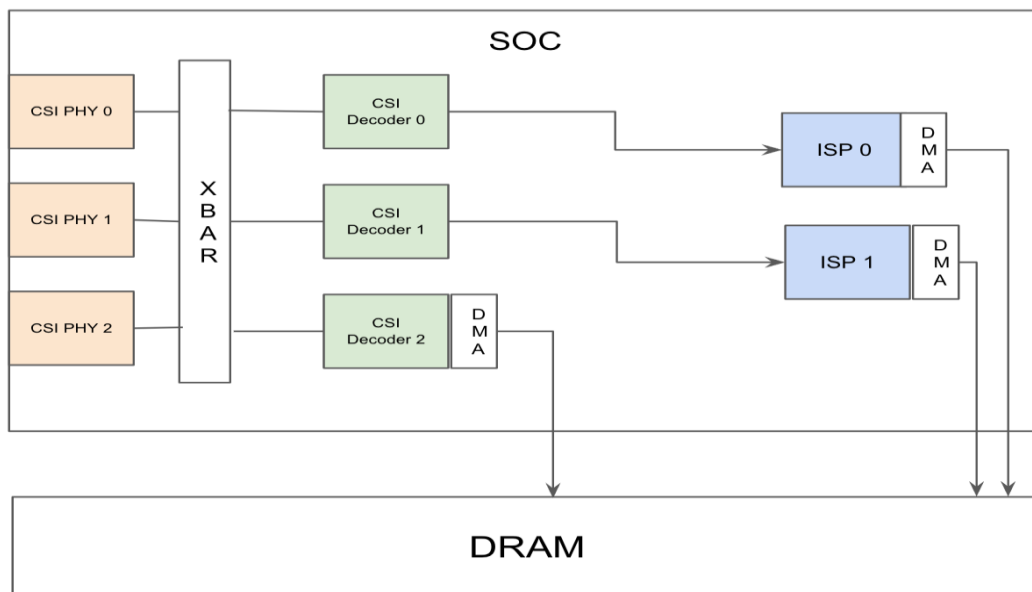


Figure 1: Typical Mobile phone camera interface

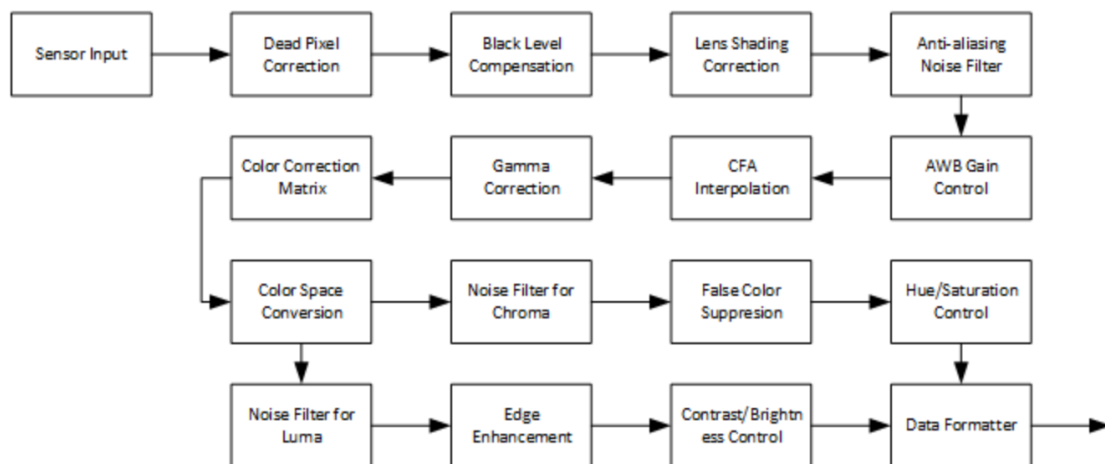


Figure 2: Image Signal Processor hardware blocks (ISP Pipeline)

3. Light Weight ISP

Typically monochrome cameras are used for most of the computer vision algorithms like object detection, tracking, object counting etc., These monochrome cameras are connected to the SoC over the same MIPI interface just like the Bayer color sensor. However, the monochrome cameras output grayscale data in the RAW8 or RAW10 format. This data will be consumed by the AI/ML based neural network model for various computer vision tasks said above. The neural network models are trained with the data collected from the same sensor or similar sensor. The output of the monochrome sensors doesn't need to be enhanced for two different reasons. Firstly they are not going to be visually presented to humans. Secondly they are going to be processed by the neural network models which can be trained with the raw (unprocessed by the ISP) output from the camera sensor. So for the computer vision activities with monochrome cameras we can have a light weight ISP (CSI PHY 2 -> CSI Decoder 2 -> DMA -> DRAM) pipeline in Fig.1. This reduces the die size of the SoC as the ISP hardware blocks consume a significant amount of space in the SoC as there are usually

more monochrome cameras in the modern CV based systems. This also saves power because the data is unprocessed by the hardware blocks which will extend the lifetime of the battery operated devices. Finally this also saves overall cost of the SoC and the product.

4. Conclusion

In this paper we discussed about the need for the ISP for generating visually appealing images from the color cameras and a lightweight ISP for deep neural network models that are used by computer vision applications like object detection, tracking etc., The lightweight ISP allows us to save die size of the SoC, cost of the product and reduces overall power consumption of the device.

References

- [1] <https://github.com/cruxopen/openISP>
- [2] MIPI CSI-2 Specification, MIPI Alliance
- [3] Park, Hyun. (2016). Architectural Analysis of a Baseline ISP Pipeline. 10.1007/978-94-017-9987-4_2.