embeddedworld2020
Exhibition&Conference
... it's a smarter world

Touchless HMI with Low Power Gesture Detection

Dirk Seidel
Senior Marketing Manager
Lattice Semiconductor
San Jose, California, U.S.A.
Dirk.Seidel@latticesemi.com

Abstract— Early uses of gesture detection as an HMI began with gaming and entertainment applications on platforms like Microsoft's Xbox or Nintendo's Wii, but they required significant computational performance and power. Designers have been less successful using gesture detection technology in broad market applications. Analysts at MarketWatch forecast consumer electronics using gesture recognition technology will grow at a 30.37 percent rate over the next five years. A reason for this is the broad availability of devices and tools using Artificial Intelligence (AI)/Machine learning (ML), which makes the rapid adoption of hand gesture detection as a way to control the human machine interface (HMI) less elusive. But there are challenges to address before developers can implement gesture detection more broadly: they need to run costly, power-hungry sensor hardware in always-on applications, they require high signal processing capabilities, and enable fast response times that demand significant computational resources. As a result, gesture detection systems often consume significant power and computational resources that limit how they may be used in resource and power-constrained devices.

This paper will explore an FPGA-based approach for use in developing low power gesture detection applications in products targeting the broad market.

I. Introduction

Nine years after Microsoft introduced its Kinect motion sensor as an accessory to the Xbox video game system, the rapid adoption of hand gesture detection as a way to control the human machine interface (HMI) remains elusive. Early uses of gesture detection as an HMI have largely revolved around gaming and entertainment applications on platforms like Microsoft's Xbox or Nintendo's Wii. Designers have been less successful using gesture detection technology in broad market applications except for a small number of IoT devices.

Despite the slow adoption, market researchers remain optimistic. Analysts at MarketWatch forecast consumer electronics using gesture recognition technology will grow at a 30.37 percent rate over the next five years. And industry analysts agree that the use of simple gestures to control or interact with a device without physically touching it offers multiple advantages in applications like the smart home and smart factory.

So, what has held up the adoption of gesture detection technology in mainstream consumer applications? Over the last few years developers have brought to market multiple innovative gesture detection approaches such as those based on infrared, electric field sensing and wireless signals. But existing approaches still face challenges: they need to run costly, power-hungry sensor hardware in always-on applications, they require high signal processing capabilities, and enable fast response times that demand significant computational resources. As a result, gesture detection systems often consume significant power and computational resources that limit how they may be used in resource and power-constrained devices.

As an example, Microsoft's Kinect features an RGB camera and a depth sensor that uses an IR light source to output 3-D positional data in real time. Kinect data is organized as a stream of two 640 x 480 images acquired at a rate of 30 frames per second (fps). One stream serves as an ordinary 24-bit RGB video image, while the other provides an 11-bit depth image that is used to calculate 3-D positional data.

II. DIFFERENT APPROACH

But there are ways to implement gesture detection technology without the use of expensive hardware, high resolution image processors or a high-performance CPU. Indeed, there are many applications that don't require support for complex gesture detection capabilities, like waving a hand to turn a smart light switch or water faucet on and off. This whitepaper looks at a new approach to reducing the cost of gesture recognition solutions on IoT devices. By combining recent advances in Artificial Intelligence (AI)/Machine learning (ML) with improvements in a new generation of very low-power Field Programmable Gate Arrays (FPGAs), gesture detection can be easier to implement and more cost effective. Designers can now use convolutional neural networks (CNN) in Lattice FPGAs with non-invasive sensor technologies like IR and passive IR (PIR). The CNN can be trained for specific gesture detection and then deployed on the FPGA for inferencing the gesture.

There are several potential advantages to this approach. The designer does not need to be an expert in image processing algorithms. Also there is no need for a powerful processor to make sense of the input data. Machine learning techniques simplify this process and the Lattice sensAI™ solution stack makes it simple to implement specific gesture detection solutions. Finally, by using Lattice's ultra-low-power iCE40 UltraPlus™ FPGAs, designers can take advantage of significant power savings, a crucial consideration in battery-powered applications.

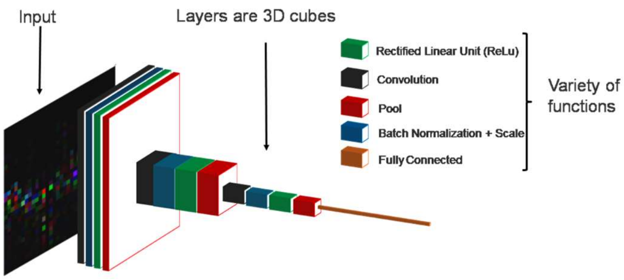

Figure 1: Overview of a BNN structure

Figure 1 illustrates the CNN model that runs on the Lattice FPGA. This is a ready-to-go NN model specifically optimized for FPGA resources and supporting multiple layer types as shown above. Convolution is an 8-bit multiplication of the input data and the 8-bit weights. Rectified Linear Unit (ReLu) sets data below a specific threshold to 0 and higher than the same threshold to 1. Pool is performed on each for adjacent pixels of the image and chooses the highest probability meaningful pixel. This function reduces the amount of computation needed in subsequent steps. The fully-connected layer is usually the last layer and it takes every neuron from the previous layer. This function is generally computationally expensive, so it is performed as a last operation where there are significantly fewer neurons.

The CNN model is trained using a GPU and running standard training tools such as Café and TensorFlow. The training dataset used is collected using actual hardware and includes about 1,000 images of each of the patterns of interest. This phase is known as the training phase. The output of the training tool is then passed through the NN compiler tool to format it for use by the FPGA design as a set of instructions for the CNN IP and weights. The weights should be thought of as a template for the gesture to be used for comparison to the input data the system receives.

The system would run a Lattice FPGA with the CNN IP instantiated along with the image sensor interface block, downscaling block and output post processing block. The post processing block can connect the FPGA to the rest of the system through standard interfaces such as SPI/I2C/UART or GPIO.

III. NEW REFERENCE DESIGN

To help accelerate the use of this technology across a wider array of applications, Lattice now offers a reference design for the development of hand gesture solutions. The reference design detects hand gestures using the CNN IP implemented in a Lattice 5K iCE40 UltraPlus FPGA. Engineers can develop their design on the HM01B0 UPduino Shield, a complete hardware development platform which uses the UPduino 2.0 board and a daughter boards with HM01B0 image sensor. Developers can retrain the network with an updated model using deep learning frameworks like Caffe, Keras or Tensorflow. The same application can also be developed using an IR sensor, which is less sensitive to lighting conditions.

By taking advantage of the FPGA's inherent parallel data processing capability, solutions based on this reference design can dramatically reduce power consumption in always-on applications. The end to end reference design provides the ability to modify the input sensor, the gesture the design is trained for and the post-processing interface.

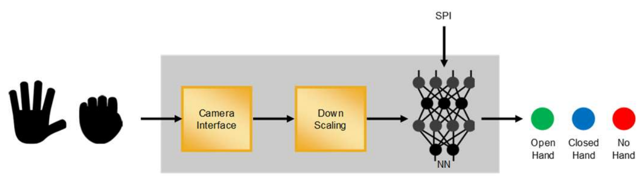

Figure 2: A system diagram for a low power gesture detection solution for use in Edge devices

The Lattice iCE40 UltraPlus FPGA used for this application features 5,280 LUTs, 120 Kb of Block RAM, eight DSP blocks and 1 Mb of RAM. With 128 Kbytes of integrated memory, the iCE40 UltraPlus allows the storage of weights/activations directly inside the FPGA. This architecture is well suited for AI/ML applications because of the distribution of the resources around the FPGA fabric enable efficient data movement. At five frames per second, power consumption is only 3.3 mW. The frame rate can be modified by trading off more power consumption for speed. Input resolution can also be adjusted up to 128x128x3, trading accuracy for power consumption. For this demo a hand must be located no less than six and no more than 12 inches from the image sensor. An on-board LED indicates when a gesture is detected.

This reference design can be modified to perform a wide range of HMI functions using gesture detection and can be used in a range of applications, from using a hand gesture to turn office equipment and displays on or off, or to adjust thermostats and other IoT-connected control appliances in a smart home.

IV. COMPREHENSIVE ECOSYSTEM

To help developers extend gesture detection to a wider array of applications, Lattice developed sensAI, a full-featured ML inferencing technology stack. Optimized for low power operation from 1 mW to 1 W, in package sizes as small as 5.5 mm^2, and priced for high volume production, this stack accelerates development and deployment of always-on, on-device AI in a wide range of Edge applications.

sensAI is supported by a comprehensive ecosystem that brings together hardware kits, neural network IP cores, software tools, reference designs and custom design services. sensAI can be implemented on the HM01B0 UPduino Shield, a low cost modular development platform featuring a Lattice iCE40 UltraPlus FPGA and an image sensor, which can be changed to the sensor of choice for a specific application. Gesture detection designs can take advantage of Lattice's compact CNN IP to support designs with a power budget under a few mW. On the software tools side, Lattice has added a neural network compiler tool for networks developed in Keras, Caffe or TensorFlow to its Radiant design software. With the compiler developers can implement their designs into Lattice FPGAs without any prior knowledge of RTL.

The sensAI ecosystem also offers a wide range of reference designs beyond hand gesture detection, including face tracking, human face detection, human presence detection, key phrase detection, object counting and speed sign detection. Lattice has also built a community of certified design service partners that offer custom solutions and services for a wide range of end applications including the smart home, smart factory, smart city and smart cars.

V. CONCLUSION

Interest is growing in the use of hand gestures to interact with devices in mass market applications for the smart home and smart factory. But current approaches require hardware that is simply too costly and not flexible enough for mainstream consumer applications. By taking advantage of new technology stacks and reference designs based on small, low power FPGAs, developers can now build solutions that meet the stringent power, cost and footprint limitations those markets demand.