ARCHITECTURAL STRATEGIESFOR HANDPOSTURE RECOGNITIONIN SYSTEMSONCHIP

Mahesh Chandra

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2017

ARCHITECTURAL STRATEGIESFOR HANDPOSTURE RECOGNITIONIN SYSTEMSON CHIP

by

Mahesh Chandra
DEPARTMENT OF ELECTRICAL ENGINEERING

Submitted
in partial fulfilment ofthe requirements for the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2017

गुरुर्ब्रह्मा गुरुर्विष्णुः, गुरुर्देवो महेश्वरः।

गुरुः साक्षात् परंब्रह्म, तस्मै श्री गुरुवे नमः।।

I bow to the ‘GURU’; the creator (Brahma), the preserver (Vishnu), the destroyer (Shiva) and the source of the Absolute.

गुरुजनों को समर्पित

To my GURUs

Certificate

This is to certify that the thesis entitled '**Architectural Strategies for Hand Posture Recognition in Systems on Chip'** being submitted by **Mahesh Chandra** for the award of the degree of **Doctor of Philosophy** to the Department of Electrical Engineering, Indian Institute of Technology Delhi, is a record of bonafide work done by him under our supervision and guidance. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree. The matter embodied in this thesis has not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. Brejesh Lall
Associate Professor
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi – 110016
India

Dr. Kaushik Saha
CTO,
Samsung Research Institute India - Delhi,
2A, Sector 126, Noida 201304
U.P., India

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisors Dr. Brejesh Lall and Dr. Kaushik Saha under whom I had the privilege to do this work. I am grateful for their expert guidance, kind suggestions and encouragement throughout the study and research work. Dr. Lall has been a perfect teacher and mentor during this period, even beyond the scope of the PhD, which has helped me immensely to carry out this work.

Besides my advisors, I would like to thank Prof. S. D. Joshi, Prof. G. S. Visweswaran and Prof. P. K. Kalra for agreeing to be part of my review committee and for their valuable suggestions and support. I would also like to extend my gratitude to all the faculty members and staff of IIT for making learning so much comfortable and fun.

I am thankful to STMicroelectronics management for giving me this opportunity. I would like to thank my managers at STMicroelectronics, Nikhil Kulshrestha, Gary McGillivray and Jerome Chossat for their continuous support, encouragement and patience.

I am thankful to my colleagues for their support and collaboration. I would like to thank Sandeep Chawla, Abhijeet Shrivastava, Jayesh Verma, Jwalant Joshipura, Sunil Kumar, Ashish Chawla, Kamal Mathpal, Deepak Chandola, Antoine Drouot, Mathieu Thivin, Alain Issard, Ludovic Chotard, Arnaud Bourge, Stephane Drouard, Giuseppe Montalbano, Arnaud Laflaquiere, David Sicard, Franck Bernard, Linda Kerrick, and Michael Markowitz of STMicroelectronics for their collaboration, help and support. I would also like to thank Rishi Kumar and Samer Hijazi of Cadence Design Systems for providing an opportunity to understand and work on the CNN.

I am thankful to my fellow students, particularly Pratyush, Prashanth, Anuj, Mansi, and Prerna for their help and support.

I am fortunate to have friends who are source of encouragement and support; I thank them for their friendship and support.

I would also like to thank the people who are selflessly contributing to the world of knowledge by online blogs, books, courses, sample codes and support. I have been immensely benefited from these unseen faces and am grateful to them.

Lastly, I would like to acknowledge the contribution of my mother, wife, children and other family members. As always, they have been very supportive and continuous source of encouragement. Without their sacrifices, it would have remained a distant dream.

New Delhi *Mahesh Chandra*

Abstract

The computer vision based hand gesture interfaces have many advantages. They are non-intrusive, do passive sensing, the hardware is commercially available at relatively lower cost and the camera can be used for other tasks. However, they require application specific algorithm development and deploying them is a challenge due to huge variation in operating conditions such as lighting condition, backgrounds, user differences, camera performance.

The hand gesture recognition has been an active area of research for more than two decades. The performance of the gesture recognition algorithms has improved significantly in the last few years. However, most of these algorithms require huge computing and memory resources and consume lot of power. These issues must be addressed for wider applications of the gesture based interfaces. The primary objective of this work is to provide simple and cost-effective solution for the hand posture recognition which is a sub-class of gesture recognition. This work is motivated by the efforts to move the vision processing closer to the sensor. So, the main focus is the implementation of the hand posture recognition algorithms.

Convolutional neural networks are the state of the art for the computer vision applications. These are evolving quite fast and the latest CNN implementations have significantly reduced the number of parameters and computational requirement while maintaining the detection performance. We have finetuned one such network, i.e. SqueezeNet, for static hand pose classification achieving 99.65% accuracy for ten classes. We have also explored the implementation of this CNN and proposed a novel hardware implementation which doesn't require a frame buffer and simplifies the integration and system level design. This hardware implementation has been synthesized for Xilinx vertex-7 FPGA.

To build the gesture recognition system, the CNN has to be integrated with the image acquisition subsystem. The images are captured using CMOS sensors in Bayer color format and an ISP is used to process these images for improving image quality. Using images in Bayer format for the hand pose classification will reduce the ISP complexity; so, we have explored CNN training using subsampled 'Bayer' color format. We have also proposed a novel method of implementing ISP for reducing power consumption. With this implementation, the power consumption can be reduced to 35-45%.

Typical applications, such as remote control for TV, have specific characteristics which can be used to save power at system level. We have proposed two techniques in this work, i.e. two stage detector and variable frame rate capture, for reducing the power consumption in such application. These simple techniques reduce the power consumption significantly.

सार

कंप्यूटर दृष्टि द्वारा हस्तमुद्रा की पहचान पर आधारित इंटरफेस के कई फायदे हैं। वे गैर-दखलंदाज हैं, निष्क्रिय सेंसिंग करते हैं, हार्डवेयर वाणिज्यिक रूप से अपेक्षाकृत कम लागत पर उपलब्ध है और कैमरे को अन्य कार्यों के लिए उपयोग किया जा सकता है। हालाँकि, उनके लिए अनुप्रयोग के अनुरूप विशिष्ट एल्गोरिथम विकास की आवश्यकता होती है और परिचालन की परिस्थितियों जैसे कि प्रकाश स्थिति, पृष्ठभूमि, उपयोगकर्ता अंतर, कैमरा प्रदर्शन में भारी भिन्नता के कारण उन्हें चुनना एक चुनौती है।

दो दशकों से अधिक समय के लिए यह अनुसंधान का एक सक्रिय क्षेत्र रहा है। इन एल्गोरिदम के प्रदर्शन में पिछले कुछ वर्षों में काफी सुधार हुआ है। हालांकि, इनमें से अधिकांश एल्गोरिदम को विशाल कंप्यूटिंग और मेमोरी संसाधनों की आवश्यकता होती है और बहुत सी विद्युत् शक्ति का उपभोग होता है। इन मुद्दों को इशारा आधारित इंटरफेस के व्यापक अनुप्रयोगों के लिए संबोधित किया जाना चाहिए। इस काम का प्राथमिक उद्देश्य हाथ मुद्रा की पहचान के लिए सरल और लागत प्रभावी समाधान प्रदान करना है जो कि इशारा मान्यता का एक उप-वर्ग है। यह काम संवेदक के करीब दृष्टि प्रसंस्करण को स्थानांतरित करने के प्रयासों से प्रेरित है। इसलिए, मुख्य विषय हाथ मुद्रा पहचान एल्गोरिदम के कार्यान्वयन है।

संकल्पनात्मक तंत्रिका (कोंवोलुशनल न्यूरल) नेटवर्क कंप्यूटर दृष्टि अनुप्रयोगों के लिए आधुनिकतम तकनीक हैं। यह काफी तेजी से विकसित हो रहे हैं और नवीनतम सीएनएन (CNN) लागूकरण ने पहचान के प्रदर्शन को बनाए रखते हुए काफी मापदंडों और संगड़क आवश्यकताओं की संख्या में कमी की है। हमने एक ऐसे नेटवर्क, स्क्वीज़नेटनेट, को इस काम के लिए प्रयोग किया है और हाथों की दस मुद्राओं के वर्गीकरण में 99.65% पहचान प्राप्त की है। हमने इस सीएनएन के कार्यान्वयन का भी पता लगाया है और एक नया हार्डवेयर कार्यान्वयन प्रस्तावित किया है जिसके लिए फ़्रेम बफर की आवश्यकता नहीं है और एकीकरण और सिस्टम स्तर डिजाइन को सरल करता है। यह हार्डवेयर कार्यान्वयन Xilinx vertex-7 FPGA के लिए संश्लेषित किया गया है।

इस प्रणाली पर आधारित सिस्टम का निर्माण करने के लिए, सीएनएन को कैमरा सबसिस्टम के साथ एकीकृत करना होगा। फोटो खींचने के लिए बेयर (Bayer) रंग स्वरूप में सीमोस (CMOS) सेंसर का उपयोग किया जाता है और फोटो की गुणवत्ता में सुधार के लिए आईएसपी (ISP) प्रयोग किया जाता है। हाथों वाले वर्गीकरण के लिए बेयर प्रारूप में फोटो का उपयोग आईएसपी जटिलता को कम करेगा; इसलिए, हमने सीएनएन प्रशिक्षण को 'बेयर' रंग प्रारूप का इस्तेमाल किया है। हमने बिजली की खपत को कम करने के लिए आईएसपी को कार्यान्वित करने की एक नयी पद्धति भी प्रस्तावित की है। इस क्रियान्वयन के साथ, बिजली की खपत को 35-45% तक घटाया जा सकता है।

विशिष्ट अनुप्रयोगों, जैसे टीवी के लिए रिमोट कंट्रोल, की विशिष्ट विशेषतायें होती हैं जिनका उपयोग सिस्टम स्तर पर बिजली बचाने के लिए इस्तेमाल किया जा सकता है। हमने इस काम में दो तकनीकों का प्रस्ताव रखा है, ऐसे आवेदन में बिजली की खपत को कम करने के लिए दो चरणों में पहचान करना और फोटो खींचने की गति को परिवर्तित करना। इन सरल तकनीकों के प्रयोग से बिजली की खपत काफी बचायी जा सकती है।

Table of Contents

List of Figures

List of Tables

Abbreviations

2D	Two-Dimensional Space
3D	Three-Dimensional Space
5DT	Fifth Dimension Technologies
ANN	Artificial Neural Network
AdaBoost	Adaptive Boosting
ADC	Analog to Digital Converter
AP	Associative Processor
ARM	Advanced RISC Machines
ASL	American Sign Language
ASIC	Application Specific Integrated Circuit
AWGN	Additive White Gaussian Noise
CamShift	Continuous Adaptive Mean Shift
CCD	Charge Coupled Device
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Network
Conv	Convolution
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
DB	Database
DCT	Discrete Cosine Transform
DOF	Degree of Freedom
DRAM	Dynamic Random Access Memory
DSP	Digital Signal Processor
DTW	Dynamic Time Warping
FAST	Feature from Accelerated Segment Test
FC	Fully Connected (Layer)
FFT	Fast Fourier Transform
FOV	Field of View
FPGA	Field Programmable Gate Array
FPS, fps	Frames per second
FSM	Finite State Machine
GB	Giga Byte
GMM	Gaussian Mixture Model
GDS	Gesture Data Score
GHz	Giga Hertz
GPU	Graphics Processing Unit
HCI	Human Computer Interface
HD	High Definition (image/video resolution)
HDR	High Dynamic Range
HMI	Human Machine Interface
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
HIS	Hue, Saturation, Intensity (color space)
HSV	Hue, Saturation, Value (color space)
HW	Hardware
I/F	Interface
I/P	Input
IC	Integrated Circuit

IP	Intellectual Property
IR	Infra-red
ISE	Integrated Synthesis Environment (Xilinx EDA tool)
ISO	International Organization for Standardization. In this document, ISO level represents the sensitivity of the image sensors in the context of capturing images.
ISP	Image Signal Processor
KB	Kilo Byte
KNN	K-Nearest Neighbors
LCD	Liquid Crystal Display
LED	Light Emitting Diode
LSC	Lens Shading Correction
LUT	Look Up Table
LVT	Low V_T (Threshold Voltage)
M	Million or Mega (10^6 or 2^{20})
MAC	Multiply Accumulate
MaxPool	Max Pooling (Layer)
MCU	Microcontroller Unit
Mem	Memory
MEMS	Micro Electro Mechanical Sensor
MIPI	Mobile Industry Processor Interface
MIPS	Million Instructions per second
MIT	Massachusetts Institute of Technology
NIOS	Soft core for Altera FPGAs
No.	Number
NRE	Non-Recurring Engineering (cost)
NUS	National University of Singapore
Op(s)	Operation(s)
O/P	Output
OS	Operating System
Param	Parameter
PC	Personal Computer
PCA	Principal Component Analysis
PE	Processing Elements
Perf	Performance
RAM	Random Access Memory
ReLU	Rectified Linear Unit
RGB	Red, Green, Blue (color space)
SAD	Sum of Absolute Differences
SDRAM	Synchronous DRAM
SIFT	Scale Invariant Feature Transform
SIMD	Single Instruction, Multiple Data
SoC	System on Chip
SOPC	System on a Programmable Chip (Altera EDA tool)
SNR	Signal to Noise Ratio
SRAM	Static Random Access Memory
SSD	Sum of Squared Differences
SVM	Support Vector Machine
SW	Software
TI	Texas Instruments
TV	Television

VGA	Video Graphics Array
VHDL	VHSIC (very high speed IC) Hardware Description Language
VLIW	Very Long Instruction Word (processor)
VPL	Visual Programming Language
YCbCr, YUV	Family of color spaces; Y is luma component and the other two are chroma components.