



## Special issue on real-time image and video processing in mobile embedded systems

Awais Ahmad<sup>1</sup> · Marco Anisetti<sup>2</sup> · Ernesto Damiani<sup>3</sup> · Gwanggil Jeon<sup>4</sup>

Published online: 8 December 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Recent advancements in the field of mobile devices, such as smartphones, smart wearable gadgets, and smart watches have changed the way we connect with the world around us. Users rely on smart devices wish to maintain an always-on access to information, personal and social networks, etc. With the rapid advancements in embedded mobile sensors, such as cameras, GPS, digital compass, and proximity sensors, a variety of data is recorded, thus enabling new sensing applications across diverse research domains, including mobile information retrieval, mobile media analysis, mobile computer vision, mobile social networks, mobile human–computer interaction, mobile gaming, mobile entertainment, mobile healthcare, and mobile learning. Irrespective of the application fields, the majority of the challenges and issues brought by emerging mobile imaging and video still need to be studied and many research queries remain to be answered. For instance, seamless user experience has been recognized as one of the major factors in designing mobile image and video processing applications for multiple form factors with different device size (small, medium, and large). However, its provision is a challenging task and requires effective integration of mobile sensors and multidisciplinary research such as visual content adaptation, and user behavior analysis. Fetching the information of high-quality

video and images through mobile devices increases the battery consumption and processing power. Therefore, most of the intense battery required applications are uploaded to the cloud servers to save energy and enhance battery lifetimes for mobile users. Keeping the fact of the immense popularity of the mobile phones' high resolution data throughput applications such as video streaming and high-quality gadgets, a notable improvement and enhancement in video and image processing in mobile embedded systems has been seen over the last decade. For instance, MPEG has designed as standard for mobile visual search (MPEG Compact Description for Visual Search). Also, a mobile application Instagram has attracted 200 million monthly active users till this date. Some of the other wearable devices, such as Apple Watch or Google Glass, are also a type of personal digital assistants that have shown their potential to be the next-generation mobile media and embedded services. Keeping the face of the mobile embedded devices for high-quality streaming and the technology development, this special issue is designed to examine an open study and investigation of the current development and future trends and directions of the real-time aspects in mobile video and image video processing in mobile embedded systems.

This special issue is intended to provide a highly recognized international forum to present recent advances in Real-Time Image and Video Processing in Mobile Embedded Systems. The ultimate objective is to bring together well-focused, top quality research contributions, providing to the general real time image processing community an opportunity to get an overall view of research results, projects, surveying works and industrial experiences that are dealing with theory and applications within the theme of Real-Time Image and Video Processing in Mobile Embedded Systems.

In the last few years, CNN (Convolutional Neural Network), especially Faster Region-Convolutional Neural Networks (Faster R-CNN) has shown great advantages in image classification and object detection. It has superiority to traditional machine learning methods by a large margin.

---

✉ Gwanggil Jeon  
gjeon@inu.ac.kr

Awais Ahmad  
aahmad.marwat@gmail.com

Marco Anisetti  
marco.anisetti@unimi.it

Ernesto Damiani  
ernesto.damiani@kustar.ac.ae

- <sup>1</sup> Bahria University, Islamabad, Pakistan
- <sup>2</sup> Università degli Studi di Milano, Crema, Italy
- <sup>3</sup> Khalifa University, Abu Dhabi, United Arab Emirates
- <sup>4</sup> Incheon National University, Incheon, South Korea

The contribution by Zhang et al. “Real-time vehicle type classification with deep convolutional neural networks” proposes a deep learning based vehicle type classification system. The proposed system uses Faster R-CNN to solve the task. The authors test the system on an NVIDIA Jetson TK1 board with 192 CUDA cores that is envisioned to be forerunner computational brain for computer vision, robotics and self-driving cars.

Face recognition, expression identification, age determination, racial binding and gender classification are common examples of computational image processing. The conventional sequence for recent real-time facial image processing consists of five steps: face detection, noise removal, face alignment, feature representation, classification. The contribution by Malik et al. “Deepgender: real-time gender classification using deep learning for smartphones” proposes feature representation which is presented in multi-layer deep neural network, hence the authors name it Deepgender. It is reported that the proposed Deepgender system has registered 98% accuracy by combined use of both databases with the specific preprocessing procedure.

The single-image upsampling with denoising influences the quality of the resulting images. Image upsampling is known as super-resolution, which refers to restoration of a higher-resolution image from a given low-resolution image. The contribution by Kim et al. “Kernel design for real-time denoising implementation in low-resolution images” proposes a filter-based image upsampling and denoising method for low-resolution images. The proposed method involves two stages: (1) designing least squares method-based filters, (2) implementation of image upsampling and denoising process. In addition, image quality effect on various-sized filters are tested on low-resolution noisy images.

The infrared (IR) radiance of an aerial target owing to the reflection of the external sources including the sun, atmosphere and the earth’s surface is a key factor to consider in the modeling and simulation of the IR image in the studies of target detection and tracking, guidance and camouflage. The inherent parallelism that the reflection of radiations incident from different directions in each spectral wavelength can be calculated in parallel in this problem encourages us to accelerate it on multi-core platforms, which are common nowadays. The contribution by Wu et al. “Parallel BRDF-based infrared radiation simulation of aerial targets implemented on Intel Xeon processor and Xeon Phi coprocessor” firstly uses a dual socket Intel Xeon E5-2620 nodes running at 2.00 GHz. Subsequently, implementations using native and offload modes on the Intel Xeon Phi 5110p coprocessor are described in detail. By increasing the scalability and vectorization, speedups obtained in the native and offload mode implementations were 13.8× and 13.0×, respectively.

The demosaicking is a digital image process used to reconstruct a full color image from the incomplete color

samples output from an image sensor overlaid with a color filter array. The contribution by Lee et al. “Real-time demosaicking method based on mixed color channel correlation” proposes a new demosaicking method based on mixed color channel correlation. Different from conventional interpolation methods based on only two or four directions, the proposed method exploits the mixed color channel correlation within the local sliding window to improve the interpolation performance. In addition, the proposed method uses both of correlation of spatial closeness and spectral similarity between the high and low-resolution of the raw color filter array image. By using geometric duality of Bayer CFA pattern, a robust interpolation model is proposed with optimal interpolation coefficients.

The very-large-scale integration (VLSI) is the process of creating an integrated circuit by combining hundreds of thousands of transistors or devices into a single chip. The contribution by Chang et al. “VLSI implementation of anisotropic probabilistic neural network for real-time image scaling” proposes an VLSI implementation of anisotropic probabilistic neural network (APNN) for real-time video processing applications. The APNN interpolation method achieves good sharpness enhancement at edge regions, and reveals the noise reduction at smooth regions. For real-time applications, the APNN interpolation is further implemented with efficient pipelined VLSI architecture. The VLSI architecture of APNN has a five-layer structure which is comprised of Euclidian layer, Gaussian layer, weighting layer, summation layer and division layer. The presented VLSI implementation of APNN interpolation method can reach  $1920 \times 1080$  at 30 frames per second with a reasonable hardware cost.

Image super-resolution (SR) plays an important role in many areas as it promises to generate high resolution (HR) images without upgrading image sensors. Many existing SR methods require a large external training set, which would consume a lot of memory. In addition, these methods are usually time-consuming when training models. Moreover, these methods need to retrain models once the magnification factor changes. The contribution by Wu et al. “A fast single-image super-resolution method implemented with CUDA” proposes a method, which does not need external training set by using self-similarity. Firstly, authors rotate original low resolution (LR) image with different angles to expand the training set. Secondly, multi-scale Difference of Gaussian (DoG) filters are exploited to obtain multi-view feature maps. Multi-view feature maps could provide an accurate representation of images. Thirdly, feature maps are divided into patches in parallel to build an internal training set. Finally, nonlocal-means is applied to each LR patch from original LR image to infer high resolution patches. The authors implement the proposed method with CUDA.

Interest on anomaly detection for hyperspectral images has increasingly grown during the last decades due to the diversity of applications that benefit from this technique. However, the high computational cost inherent to this detection procedure seriously limits its processing efficiency, especially for onboard application scenarios. The contribution by Gao et al. “Approximate computing for onboard anomaly detection from hyperspectral images” proposes a novel spectral and spatial approximate computing approach, named SSAC, for onboard anomaly detection from hyperspectral images. To efficiently design the proposed approach, authors deeply analyze two preliminary aspects: (1) data correlation in hyperspectral images in both spectral and spatial dimensions, (2) Error resilience of a popular hyperspectral anomaly detection algorithm in both data-level and algorithm-level. The results obtained with a nonlinear anomaly detector for hyperspectral imagery, such as the well-known kernel RX algorithm, show that the proposed SSAC approach greatly improves anomaly detection efficiency compared to the traditional method with negligible degeneration in accuracy.

RCF-Retinex is a novel Retinex-based image enhancement method which can improve contrast, eliminate noise, and enhance details simultaneously. It utilizes region covariance filter (RCF) to estimate the illumination. However, RCF-Retinex encounters a time-consuming problem, since the region covariance filter is computationally intensive, which restricts the practical application in real-time systems. The contribution by Peng et al. “Implementing real-time RCF-Retinex image enhancement method using CUDA” proposes a GPU-based RCF-Retinex, which can accelerate the computation of the region covariance filters by using CUDA. To decrease the computational complexity by parallelization, authors propose a GPU-based RCF-Retinex, which can accelerate region covariance filter by using CUDA. Experiments have proved the improvement of run time and the enhancement results are similar with those by using the unaccelerated RCF-Retinex method.

The contribution by Martin Fleury “Prospects for live higher resolution video streaming to mobile devices: achievable quality across wireless links” demonstrates that live video streaming to mobile devices with pixel resolutions from standard definition up to 4 k ultra high definition (UHD) is now becoming feasible by means of high-throughput IEEE 802.11ad at 60 GHz or 802.11ac at 5 GHz, and 4kUHD streaming is even possible with 802.11n operating at 5 GHz. The author also shows that real-time compression assisted by GPUs at 4kUHD is also becoming feasible. The author further considers the impact of packet loss on H.264/AVC and HEVC compressed video streams in terms of Structural Similarity (SSIM) index video quality. Findings suggest that, for medium-range transmission, the video quality may be acceptable at low packet loss rates.

For hardware-accelerated 4kUHD encoding, standard frame rates may be possible but appropriate higher frame rates are only just being reached in hardware implementations.

The implementation of a video reconstruction pipeline is required to improve the quality of images delivered by highly constrained devices. These algorithms require high computing capacities—several dozens of Giga Operations per second (GOPs) for real time HD 1080 p video streams. Today’s embedded designs constraints impose limitations both in terms of silicon budget and power consumption—usually 2 mm<sup>2</sup> for half a Watt. The contribution by Thevenin et al. “A templated programmable architecture for highly constrained embedded HD video processing” presents the eISP architecture that is able to reach 188 MOPs/mW with 94 GOPs/mm<sup>2</sup> and 378 GOPs/mW using TSMC 65 nm integration technology. This fully programmable and modular architecture is based on an analysis of video processing algorithms. Synthesizable VHDL is generated taking into account different parameters, which simplifies the architecture sizing and characterization.

A large number of battery-powered systems will then integrate an HEVC codec, implementing the latest video encoding standard from MPEG, and these systems will need to be energy efficient. Constraining the energy consumption of HEVC encoders is a challenging task, especially for embedded applications based on software encoders. The most efficient approach to manage the energy consumption of an HEVC encoder consists in optimizing the quad-tree partitioning and balance compression efficiency and energy consumption. For the purpose of budgeting the energy consumption of a real-time HEVC encoder, the contribution by Mercat et al. “On predicting the HEVC intra quad-tree partitioning with tunable energy and rate-distortion” proposes a variance-aware quad-tree prediction which limits the energetic cost of the RDO process. The predictor is moreover adjustable by two parameters, offering a trade-off between energetic gains and compression efficiency.

Real-time operation and low-power dissipation in video coding systems have become important research challenges, especially in mobile devices with limited battery and computational resources. Given the variety of applications able to manipulate videos and the growing number of video coding standards, current devices are expected to provide native support to multiple coding standards. Although state-of-the-art coding requires a wide set of tools focusing on coding efficiency, major tools are usually present in different standards with limited differences. The contribution by Penny et al. “High-throughput and power-efficient hardware design for a multiple video coding standard sample interpolator” presents a multi-standard sample interpolator hardware design for the MC and FME with full support to MPEG-2, MPEG-4, H.264/AVC, HEVC, AVS, and AVS2. The proposed design is capable of UHD 8K (Ultra High

Definition—4320 p@60 fps) real-time interpolation when synthesized using a 45 nm standard-cell library. The circuit footprint occupies 65,508  $\mu\text{m}^2$  and the power dissipation ranges from 14.58 to 65.316 mW for MPEG-2 and AVS2 operation modes, respectively.

For high compression ratios (smaller output size), execution time is dominated by the transformation algorithm, which plays a progressively smaller role as the compression ratio gets smaller (larger output size). The contribution by Alvares et al. “Real-time rate-distortion optimized image compression with region of interest on the ARM architecture for underwater robotics applications” proposes the use of a real-time progressive image compression and Region Of Interest (ROI) algorithm for the design of an underwater image sensor, to be installed in an Autonomous Underwater Vehicle for Intervention (I-AUV), with constraints on the available bandwidth for the ARM architecture, allowing a more agile data exchange between the vehicle and a human operator supervising the underwater intervention. The authors focus on a novel efficient in-place, multithreaded, and cache-friendly parallel 2-D wavelet transform algorithm, based on the lifting transform using the ARM Architecture.

With the advent of touch screens in mobile devices, sketch-based image search is becoming the most intuitive method to query multimedia contents. The absence of such critical information from sketches increased the ambiguity between natural images and their sketches. Although it was previously considered too cumbersome for users to add colors to hand-drawn sketches in image retrieval systems, the modern day touch input devices make it convenient to add shades or colors to query sketches. The contribution by Baik et al. “Partially shaded sketch-based image search in real mobile device environments via sketch-oriented compact neural codes” proposes deep neural codes extracted from partially colored sketches by an efficient convolutional neural network (CNN) fine-tuned on sketch-oriented augmented dataset. The authors also studied the effects of shading and partial coloring on retrieval performance and show that the proposed method provides superior performance in

sketch-based large-scale image retrieval on mobile devices as compared to other state-of-the-art methods.

Vision-based Advanced Driver Assistance Systems (ADAS), appeared in the 2000s, are increasingly integrated on-board mass-produced vehicles, as off-the-shelf low-cost cameras are now available. Integrating accurate localization and mapping functionalities meeting the constraints of ADAS would pave the way towards obstacle detection, identification and tracking on-board vehicles at potential high speed. While the SLAM problem has been widely addressed by the robotics community, very few embedded operational implementations can be found, and they do not meet the ADAS-related constraints. The contribution by Piat et al. “HW/SW co-design of a visual SLAM application” implements the first 3D monocular EKF SLAM chain on a heterogeneous architecture, on a single SoC, meeting these constraints. In order to do so, the authors picked up a standard co-design method and adapted it to the implementation of such a complex processing chain. The authors also have designed original hardware accelerators for all the image processing functions involved, and for some algebraic operations involved in the filtering process.

We hope that this special issue would shed light on major developments in the area of real time image processing for mobile embedded systems and attract attention by the scientific community to pursue further investigations leading to the rapid implementation of these technologies.

**Acknowledgements** We would like to express our appreciation to all the authors for their informative contributions and the reviewers for their support and constructive critiques in making this special issue possible.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.