Real-time Face Detection and Recognition on the Vision-In-Package System

E. Türetken, E. Franzi, P.-A. Beuchat

Automated facial analytics has a broad range of applications in many industries such as biometrics and market research. Existing solutions require significant hardware resources that consume substantial amounts of power and are costly. To address these limitations, we developed a complete real-time face detection and recognition system that is low-power, compact and inexpensive.

Facial analytics and recognition is a rapidly growing market expected to reach $ 20 billion by 2025 [1]. From surveillance and access control systems in smart buildings to retail stores that automatically collect viewership and demographics, there is an ever growing need for compact and reliable solutions.

One of the major barriers to rapid adaptation of this technology, especially in applications with low-power, small-size and low-cost requirements, is the computationally demanding nature of the algorithms involved. As a result, existing solutions on the market rely on GHz-clocked multi-core processors that are neither power-efficient nor cost-effective.

The Vision-In-Package (VIP) system is designed to address these challenges. It is a compact camera system with a low-power processor (ARM Cortex M4/M7 with 8 MB RAM), a high-dynamic range imager, optics, and a communication interface. The system occupies only around 4 cm^3 and weighs less than 20 g including a battery cell. It features a complete facial analysis pipeline running in real time and fully embedded within the VIP system.

The software is compact and stand-alone with no external dependencies. It is comprised of a minimal version of the uKOS operating system [2] and a face analysis package running on it. Unlike existing systems that run on powerful hardware architectures, our system requires several orders of magnitude less CPU time and memory. The analysis pipeline runs at around 4-5 frames per second at QVGA resolution on the VIP and consists of the following steps, also depicted in Figure 1.

- Face detection: All the faces in an acquired frame are detected, which typically takes less than a hundred ms to run and requires only a few hundred KB of RAM memory;
- Facial landmark localization: Facial attributes, such as corners of the eyes and nose, are located within each detected face region;
- Normalization: Involves a rough geometric transformation that aligns the eyes horizontally and scales the face to a standard size, and a photometric normalization that re-moves non-linear intensity variations caused by shadows and non-uniform illumination;
- Face recognition: Descriptive features are extracted at landmark locations and used for uniquely identifying people in a database of registered faces. New individuals can be registered to this database instantly at any time with just a single click and without requiring any re-training.

These steps are made possible by using efficient machine learning algorithms including the Adaboost, ensemble of regression trees and LBP algorithms, which we trained on millions of examples with ground truth annotations. The resulting classifiers typically take a few hundred kilobytes of space and are fast to run even on low-end mobile processors.

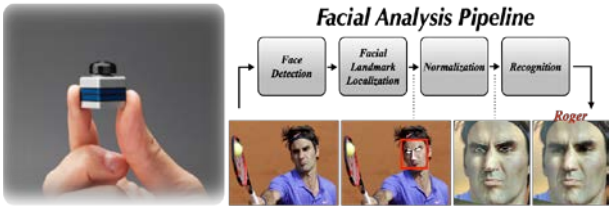

Figure 1: The VIP system (left). The facial analysis pipeline (right).

The combination of compact design offered by the VIP system and the efficiency of the face processing pipeline allows for applications with low-power consumption, small-size & weight, or low-cost requirements. We believe the following industries offer numerous opportunities for a widespread use of our technology:

- Wearables: Fast and compact facial analytics on smartwatches or eyewear can give users a secure and more personalized experience, and possibly provide new health screening tools for children and elderly;
- Marketing & advertisement industry: Collecting viewership and demographics data, such as age and gender that require prior facial analysis, allows advertisers to objectively assess the effectiveness of their campaigns. The data can be collected from cameras installed in digital advertising displays in retail outlets, or billboards on streets;
- Robotic/toy industry: Facial recognition and analysis can provide more personalized interactions for robotic pets allowing them to recognize their owners, play with them, read their moods and emotions;
- TV manufacturers: Combined with hand gesture recognition, facial analytics can provide a unique and personalized interaction experience with TV sets;
- Automotive industry: Facial expression analysis can be used to help improve driver safety by monitoring driver drowsiness and distraction. Furthermore, recognition of the driver allows driver customization such as automatically adjusting vehicle settings to his/her personal preferences;
- Security: Using low-power or solar-powered facial recognition systems opens up new possibilities in the security domain such as monitoring remote locations without electricity, or easing the deployment of camera networks in populated public areas such as airports.

[1] Research and Markets, "Global facial recognition market analysis & trends - Industry forecast to 2025", (2016).

[2] E. Franzi, "The µKernel project", www.ukos.ch (2016).