Advanced Imaging Applications on Smart-phones

– Convergence of General-purpose computing, Graphics acceleration, and Sensors

Sriram Sethuraman
Technologist & DMTS, Ittiam

Ittiam

Presented at Silicon India Global Embedded Conference, 21 May 2011

Overview

- Imaging on Smart-phones
- GPP, Graphics acceleration & Sensors
- Advanced Imaging
  - Image registration
  - Image warping
  - Camera movement & sensor readings
- Advanced imaging Application needs
- Samples
  - Advanced HDR
  - 2D to 3D
- Conclusions

Imaging on Smart-phones

- Standard on most Smart-phones today
- Aim to be replacements for DSCs
- ~8 Mpixel CMOS sensors
- Most use a hardware imaging pipeline to get a reasonably good looking picture
  - Of late, include features such as automatic face detection in h/w
- Some models have flashes
- Many app. store applications

Sensors in Smart-phones

- Other than CMOS sensor
  - MEMS based Accelerometers
    - Originally added to detect orientation (portrait/landscape)
    - Sometimes used to take shake free pictures
  - MEMS based 3-axis accelerometer + 3-axis magnetometer
  - MEMS based 3-axis gyroscopes
  - 3D-TOF sensors
  - IR sensors

General-purpose Compute Power in Smart-phones

- ARMv7 architecture
  - Super-scalar
  - A8, A9, A15
- Vector Floating Point + NEON SIMD
  - Suitable for geometry calculations or pixel processing
- 1-2 cores is standard (with options of up to 4 cores)
- Each core is clocked in the 600MHz - 1.5GHz range with talks of up to 2GHz

Graphics Acceleration in Smart-phones

- Heavily driven by user-interfaces and gaming
- Multiple options in h/w
  - 40-50 Million triangles per second
  - IMG (MBX/SGX), ARM (Mali), nVidia (GeForce), Qualcomm (Adreno), etc.
- OpenGL ES 2.0 compliant

Advanced Imaging

- Multiple possibilities
  - Bridging quality gaps with DSCs
  - Providing new imaging modes obtained by moving the sensor in a particular manner
    - Panorama
    - 3D stereoscopic
    - 3D Panorama
  - Augmented reality
  - Gestural interfaces
  - Gaming
- Most of them involve computer vision (scene understanding) algorithms

Building Blocks for Scene Understanding

- Features or invariants detection
  - Low level image processing such as corner detection
- Correspondence establishment
  - Identify corresponding feature points
    - Used in a multitude of applications
- Model parameter estimation
  - Learn underlying scene motion
    - Fundamental matrix
  - Outliers – object motion
- Object Segmentation
  - Morphological processing on spatio-temporal data
- Object motion tracking
  - E.g. Hand motion
- Object motion trajectory classification
  - Gesture for UI
- Co-ordinate system mapping
- OpenCV has a good collection of reference algorithms for these

Advanced Imaging Application Needs

- Typical needs
  - Driven by ease of use
  - Human factors
  - All the advanced processing needs to be completed within a short time due to the following needs
    - Shot to shot delay is reasonably low for still image capture
    - Real-time processing for augmented reality, gestural interfaces, and gaming
    - Preferable to include most processing at capture time to allow use of standard viewing options (to allow sharing with others who will have entirely different setups)

Speeding up the algorithms - 1

- Use of GPUs
  - General-purpose computing on GPUs
    - Stream processing (massively parallel simple operations)
    - Huge body of implementations exist
    - Some of the feature detection steps can be done
    - Morphological filtering can be done
  - Warping
    - Some motion model estimation methods require iterative registration requiring warping of images
    - Advanced imaging methods require temporal alignment before stitching or combining
    - Rectification methods in 3D require applying transformations such as rotation correction

Speeding up the algorithms - 2

- Use of available sensors
  - Conventional method of global motion analysis
    - Compute features, perform correspondence, and then solve using least squares method for the rotation matrix part of the Fundamental matrix
      - Computationally intensive
      - Fails in the presence of noisy data
      - De-generate cases are possible based on scene content
  - With smart sensors (such as 3-axis Acccelerometer + 3-axis magnetometer or 3-axis Gyro)
    - Rotation matrix can be obtained
    - Most smartphone OS frameworks (such as iOS, Android) provide APIs to get the rotation matrix
      - Depending on the sensitivity of the sensors used, accuracy can be different
      - Local magnetic fields can cause big errors
      - Frequency of update may be limited

Speeding up the algorithms - 3

- After off-loading suitable algorithms to GPU or obviating the need for some algorithms using sensor data,
  - Optimize any core pixel processing modules using NEON SIMD
  - Optimize any floating point operations using VFP instructions

Advanced HDR Imaging

- HDR – High Dynamic Range
- Involves taking pictures at multiple exposures and combining them to get the impression of HDR
  - Normally under-exposed parts get resolved well
  - Areas that get saturated at higher exposure retain their details from the lower exposure
  - However, conventional HDR works well only for cameras on tripod and scenes that are highly static to facilitate such combining
- Advanced HDR involves getting HDR even when slight camera or object movement is there
- Involves registering the multiple exposure pictures and then combining them robustly

Advanced HDR Imaging acceleration

- Sensor data can be used to compensate for rotation
- Graphics accelerator h/w can be used to warp the images after computing the global transformation
- Robust combining can then be done using General-purpose computing on the application processor and is highly SIMD friendly
- All the above processing can be done in the time that it takes to capture the multiple exposure pictures + a few milliseconds by using the sensor + GPU + CPU

3D using 2D Sensor

- Simple iPhone applications exist that require a user to move the camera by a specified amount to take a stereo image pair
  - Cumbersome
  - Improperly taken pairs cannot be fused very well
- Automated 3D from 2D
  - Found on some Sony DSCs as 3D sweep and 3D sweep panorama
  - Requires moving the camera parallel to the scene and capturing multiple pictures, analyze them and create one or more stereo pairs

3D from 2D - Acceleration

- Sensor can be used to obtain rotation matrix
- Resizer, if any, available in h/w can be used to decimate pictures for faster analysis
- Graphics h/w can be used to perform rotation adjustment
- General purpose processing can be used to select optimum stereo pairs based on disparity analysis
- Final chosen pairs at the high resolution can be rectified using graphics h/w
- Fast 3D capture that is intuitive and simple for users to capture can be realized – speeds can be similar or better than on DSCs
  - Demo!

Conclusions

- Advanced Imaging Applications require a lot of image processing and scene understanding algorithms
- Smart-phones are increasingly equipped with multiple sensors, graphics acceleration, and high general-purpose compute power on the application processor
- Good user experience can be realized for Advanced Imaging Apps by leveraging the above to accelerate the processing