
Smart Camera Systems Technology Roadmap

Bruce Flinchbaugh

Texas Instruments
b-flinchbaugh@ti.com

This chapter outlines requirements for real-time vision, video and image processing in camera systems for consumer electronics, video surveillance, and automotive vision applications. Related technology trends as they affect the design and development of camera processors and systems are explained. Examples of smart camera prototypes and products are described, including the roles of the embedded digital signal processors (DSPs). Finally, the requirements and trends are extrapolated to speculatively project the future of smart cameras, as well as related implications and challenges for vision research and applications.

1 Camera System Requirements for Real-Time Vision

Let us begin by defining a “smart” camera as a software-programmable camera in which video data digitized from the image sensor is fully exposed to software for processing. We also refer to such cameras as “intelligent” cameras. This concept began to be seriously considered in the 1990s. In the early 2000s it led to the development of programmable processors that satisfy the many demanding requirements for consumer electronics and suggest the future of real-time vision in cameras.

While there are many diverse functions for programmable processors in cameras, in this chapter we focus on enabling vision functions: operations that process incoming video/images to estimate useful information about the environment. Our starting point will be consumer digital camera products available today that use fully-programmable processors and other embedded systems that can be programmed to incorporate vision functions. From there we will consider requirements for other examples of smart camera systems, emphasizing video surveillance, automotive vision, and future toys and games with embedded cameras.

Consumer digital cameras are an important starting point to understand, because any vision system for widespread consumer adoption will need to satisfy the same general requirements.

- **Very low cost.** First and foremost, the solution must be available at a low price. For the consumer to buy at a low price, the bill of materials for the electronics inside must be available at a much lower price. As a rule of thumb, expect the total budget for electronic components in a smart camera to be about 30% of the end-equipment retail price. Thus if a smart camera system concept requires \$1,000 worth of devices per unit to produce, it will probably need to sell for about \$3,000 to enable a successful business. While digital cameras that cost \$3,000+ have been available for many years now, consumers generally do not buy them of course. To reach most consumers, smart camera systems will likely need to be available for less than \$300, by analogy with what has happened in the digital camera market.
- **Very low power.** This is a critical requirement for battery-operated smart cameras. For reference, the peak image and video processing power budget of a typical consumer digital still camera is typically below 700 mW. Even for fix-mounted smart cameras with wall power, a processor that runs hotter than a light bulb can easily be unacceptable for many reasons. Thus, while a 150+ W, 4 GHz personal computer processor is useful in the lab for vision experiments, a 0.6 W, 0.6 GHz DSP is much more likely to be required for a smart camera system. Video processing in cellular camera phones faces even lower power requirements for long battery life.
- **Small size.** Clearly cellular phone cameras and other consumer digital cameras must be small enough to be hand carried and ideally to fit in a pocket. The other high-volume electronic camera market, traditional CCTV video surveillance cameras, also demands small size. Practically all future smart cameras will face the same requirements. In contrast to the low-power and small-size requirements for cameras, note that the heat sink alone needed to cool a 4 GHz PC processor is larger and heavier than a cellular camera phone with a high-performance DSP inside.
- **High-speed image and video processing.** The embedded processor(s) must apply the required functions to images and video fast enough to be useful and acceptable. For digital still cameras, the bar is generally regarded as being fixed at a one second shot-to-shot delay – regardless of how many megapixels are in the image. For digital video, 30 frames per second is the typical requirement, again, independent of the image size. While some vision applications may require more or less than this, those that need more will generally need to wait for affordable systems.
- **High-speed general-purpose processing.** Practically all digital cameras run an embedded operating system of some sort to manage system resources, user interaction and data communications. This generates a re-

quirement for a processor, often a reduced-instruction set processor, to re-use existing operating system and applications software.

- **Limited high-speed memory.** The computationally intense image and video processing of digital cameras requires enough high-speed memory to hold several frames of video, and several multi-megapixel uncompressed images for burst-mode capture. This is good for vision algorithms that generally share the same requirements. However, vision algorithms that require the relatively huge high-speed memory capacities of personal computers will need to wait longer for widespread smart camera applications.
- **Modular design.** Smart cameras will likely need to be loosely coupled with other systems, and in most cases will operate autonomously without relying on communications with other/remote systems. High-bandwidth communications for transmitting compressed digital video data will be a standard feature for many smart cameras. But the video communications will be infrequently used in applications where the primary purpose of the vision system is to “watch” the video, in contrast to the traditional purpose of cameras to provide images and video for people to watch.
- **Vision functions.** While an itemization of all algorithms required by various vision approaches could be perhaps as long as the list of all vision research publications, here are a few examples of applications and generic methods to illustrate the diversity of vision function requirements for smart cameras:
 - Video surveillance: motion analysis, object tracking, face detection, face classification, event recognition, ...
 - Automotive vision: range estimation using stereo video, lane detection, face/eye tracking and analysis, obstacle detection, ...
 - Toys and games: object detection and recognition, body tracking, gesture recognition, ...

Of course, for human-computer interaction in general, we regard the many methods described in other chapters of this book as candidate requirements for smart cameras.

2 Technology Trends Behind Smart Cameras

The digital camera and cellular phone industries are well along the way to making programmable DSPs commonplace in camera products. Here is an explanation of that trend and other underlying and contributing trends, providing insight to what has happened behind the scenes to influence current digital camera and cellular phone camera designs. To the extent that these trends are sustained, they also provide a basis for projecting the future of smart camera systems in Sect. 4.

2.1 DSP Crossover Trend from Fixed-Function Circuits to Programmable Devices

Since the first programmable DSPs were designed and produced in the early 1980s, the architectures and silicon technologies have progressed to provide very high processing performance with very low electrical power requirements. For example, the TMS320C64xTM family of DSPs includes processors that operate at various speeds, e.g., 720 MHz at about 1 W [7], with eight parallel functional units for multiple operations in a single instruction cycle, thus enabling over five billion operations per second. Whereas real-time video compression functions were generally beyond the reach of DSPs in the mid-1990s, they have crossed over from circuit-based designs to cost-effective DSP applications now because DSPs are fast enough. Television-quality MPEG-2 video encoding can be implemented entirely in software on a single DSP. And DSP video decoder software for the newest and more-complex video standard, H.264, is poised to receive digital television broadcasts (e.g., DVB-H) to handheld devices such as cellular phones. The accompanying advantages for smart cameras are compelling. These include the flexibility to add new vision applications to existing hardware systems via software without requiring development of new or custom electronics, the capability to upgrade embedded systems in the field via software downloads, and advantages of software re-use in development of next-generation camera systems [5]. Further, what we see is that once a video, image or vision function runs fast enough in DSP software to be useful, it remains in software. That function does not cross back over the line to become a fixed-function circuit implementation, because that would be more expensive all things considered.

2.2 Silicon Technology Trends

The semiconductor industry is undergoing two key changes that are affecting how and when new processors emerge. First, the term of Moore's Law has ended. The technology scaling rate is already slowing. While the industry will continue to develop higher density technology for several more generations, transistor performance is nearing physical limits, and on-chip interconnect is also running into performance limitations. Thus new approaches to architecture and design will be needed to continue to realize performance improvements and cost reductions that have been historically achieved. Clock speeds may not increase much beyond the levels already being produced, but alternate parallel implementations may still provide improvements in performance, power reductions, and cost reductions. At the same time, the industry is facing a form of economic limit: the cost of generating pattern masks to manufacture a new chip design with the most advanced semiconductor production processes already exceeds \$1M and is increasing. This nominally means that in order to justify manufacturing a chip to exploit a new circuit or processor architecture, the up-front fixed cost is so high that it significantly increases the risk for a

business to invest in the device. Only the very highest-volume markets can justify the expense of developing custom processors using the most advanced semiconductor technology. Unanticipated disruptive technology developments would be needed to avoid these trends.

2.3 From Closed-Circuit Video to Network Communications

Analog CCTV systems for video surveillance have begun to give way to digital network cameras. The transformation appears that it will take many years to complete, but it has begun. Campus-wide networks of video cables connecting dozens or hundreds of analog cameras to a centralized video monitoring room are starting to be displaced by digital systems. In the design and construction of new buildings, the additional expense of video cables is increasingly avoided altogether in favor of using one high-speed digital network for both data communications and video security functions. In Sect. 4, we will discuss some of the interesting opportunities and challenges this trend poses for future video surveillance systems.

2.4 From Wired to Wireless Communications

Perhaps the single trend with the most far-reaching implications yet to be comprehended is the embedding of smart cameras in wireless phones. This trend began almost instantaneously in 2003 when the number of cellular camera phones exceeded the number of digital still cameras sold. With programmable DSPs already in hundreds of millions of cellular phones at that time, many phones had the capacity for substantial digital image and video processing software functions before the image sensor modules were integrated in next-generation designs. The increasing adoption of wireless local area networking technology (e.g., 802.11) to replace wired digital communications networks is also changing the way people think about camera applications.

2.5 Toward Huge Non-Volatile Memory Capacities

The digital camera market drove the high-volume production of low-cost, non-volatile memory cards, which started with about 8 MB capacities around 2000 and exceeded 1 GB in 2004. At the same time, micro hard disk drives were developed in similarly small form factors and now provide tens of gigabytes of storage capacity for music players, digital cameras, and camera phones. While these memory technologies are too slow to meet the high-speed memory requirements for real-time vision, video and image processing, they serve well as storage for digital video recordings and information databases in smart cameras.

2.6 On the Integration of Image Sensors and Processors

The trend at the device level so far is one of status quo. A combination of economics and modular system constraints is keeping these devices from being integrated on a single chip. While CMOS imager technology enables digital processors to be integrated on-chip, and several such devices have been developed, practically all of the world's digital cameras and camera phones continue to keep these functions separate. At the system level, the trend is just the opposite. Whereas digital video processors, as in many machine vision applications for example, have traditionally been remote to the image sensors, the availability of high-performance DSPs has tipped the economic balance to favor co-locating the sensors and processors in the same end equipment in some cases, and sometimes on the same board.

3 Examples of DSP-Based Smart Cameras

This section provides some specific examples of how the technology trends are enabling smart cameras. The systems include research prototypes and consumer products developed by various companies, using DSPs to execute real-time vision, video and/or image processing functions implemented in software.

3.1 Network Camera Prototype

An early example of a DSP-based network camera was prototyped at Texas Instruments in 1998–99. This camera was motivated by vision research for autonomous video surveillance capabilities including object tracking, dynamic position mapping, and event recognition [2, 4]. The system was a network camera with an embedded hard disk drive, using a TMS320C6211TM DSP as the processor for all functions.

Image and video processing software demonstrated using this platform included tracking and position mapping of people and vehicles in 320×240-pixel frames at about 15 frames/second. JPEG software compressed video sampled at up to 15 fields/second. While tracking people and objects, event recognition software on the camera distinguished events such as when a person entered a room, placed an object on a table, or loitered in a specified area of the room. With a 6 GB hard disk drive designed into the camera, the camera could autonomously record video or selected snapshots of events as they were recognized. The system had an Ethernet interface to serve web pages and to be remotely configured. Remote web browsers could receive live or previously-recorded motion JPEG video, dynamic maps of where people and vehicles were moving in the field of view, and other information as it was produced by the vision algorithms in real-time, or stored in the database. Portions of the design of this prototype and its digital video recording software were used in

the Panasonic WJ-HD100 hard disk video recorder product for video security applications.

DSP software in systems such as this is typically written almost entirely in C, relying on compiler-optimizations to achieve high performance, and an embedded operating system to manage system resources. When higher performance is needed, usually only a few of the most computationally intensive image processing functions need to be optimized using a high-level assembly language. For example in this prototype, key “kernels” of the JPEG encoder (e.g., the DCT) were manually optimized, as well as image differencing and connected components labeling algorithms that provided inputs for tracking and event recognition. Other functions, e.g., face recognition, can also be implemented in C to achieve high-speed performance using the same kind of processor [1].

The DSP embedded in this early network camera prototype was much slower than the fastest available today. It operated at 166 MHz. In 2004, newer compatible DSP processors were available that operate at up to 1 GHz. Thus as smart cameras and software for autonomous video surveillance and monitoring are designed and developed as products, similar functions will run about six times faster than was possible with early prototypes, or process six times the amount of video data.

3.2 Consumer and Professional Digital Cameras

Keeping in mind that our definition of “smart” cameras means “programmable” cameras, here are some early examples of consumer digital cameras in which the image processing pipeline was implemented entirely in software: the 2 megapixel HP Photosmart 315 digital camera in 2000 and the Kodak DX3500 in 2001. In these systems the particular DSP was a multi-processor camera system-on-a-chip, the TMS320DSC21TM.

Since then several other system-on-a-chip camera processors have been developed to enable many cameras with more megapixels, improvements in algorithms, and video-rate processing. Among the latest and most advanced digital cameras based on DSPs are the 14 megapixel Kodak Professional DCS Pro SLR/n and SLR/c cameras announced in 2004. These cameras face a computational per-picture burden that is nominally seven times greater than the early 2 megapixel cameras.

Processing multi-megapixel images, starting with the raw Bayer pattern of sub-sampled colors from the image sensor and proceeding through JPEG compression, requires billions of operations per second to keep the photographer from waiting to take the next picture. The specific algorithms and parameters used are proprietary to camera companies. Generically, the operations include functions such as color filter array interpolation, color space conversion, white balance, faulty pixel correction, Gamma correction, false color suppression, edge enhancement, and lens distortion correction, as well as image compres-

sion at the end of the pipeline [8]. See also reference [5] for other insights to software camera systems and software designs.

3.3 Cellular Phone Cameras

Cellular phones with digital still camera, digital video recording, and interactive videoconferencing features are rapidly evolving. The early camera phones introduced VGA-sized picture snapshot capabilities. Now the image sizes are moving up to 3–5 MP in current and next-gen phones. Whereas the early video recording and streaming capabilities of various phones were limited to SQCIF, QCIF, or QVGA-sized video, they are moving up to VGA and are anticipated to reach television quality for digital camcorder-like video recording capabilities.

As in other smart cameras, the high-complexity video encode and decode computations of cellular phones can be implemented in DSP software. Video standards such as H.263 and MPEG-4 are used [3], as well as some proprietary formats. Various camera phone products are using programmable multimedia applications processors such as OMAP1510TM and OMAP-DM270TM for the image and video functions. These multi-processor systems-on-a-chip also enable many other software functions of camera phones.

3.4 Stereo Video Range Camera Prototype

David Hall, of Team Digital Auto Drive (Team DAD) that participated in the DARPA Grand Challenge of March 2004, designed and developed a real-time stereo video range camera prototype [6] to serve as the vision system for their autonomous vehicle entry in the desert race. A vehicle servo-control subsystem takes steering and acceleration commands from the vision system.

The vision system is a rather extraordinary example of a smart camera, comprising six CCD image sensors arranged in two 3-CCD prism modules with a 12" stereo baseline. Two TMS320C64xTM DSPs operate at 1.1 GHz to process the stereo video data. Software on the first DSP reads the digital video data directly from the sensor modules, calculates a range image of the scene, and produces a 3D terrain map in real-time. The second DSP receives the 3D terrain profile, estimates significant objects, and plans a path over the terrain to intersect a way point provided by a GPS and inertial navigation subsystem. Finally the vision system sends commands to the servo-controller to steer the vehicle.

4 Extrapolating the Trends for Future Smart Cameras

In this section, we take a stab at projecting the future design constraints and challenges for smart cameras and related vision algorithm applications. While

the speculations here are perhaps stated too boldly and could turn out to be wrong in various ways, this is an attempt to logically extrapolate the trends. To the extent that the trends outlined in Sect. 2 continue and no disruptive processing technology alternative emerges, perhaps much of this will turn out to be true.

4.1 Future Smart Camera Processors, Systems, and Software Products

Considering the trends and examples of DSP-based smart cameras and the already huge economic requirements to justify custom circuit designs for vision algorithms, it appears that smart camera processors will need to be designed and developed once, and then programmed many times in order to afford wide-ranging real-time vision system applications. Re-using system-on-a-chip programmable processors from high-volume consumer cameras will essentially become a requirement for implementing new kinds of low-cost end equipment and vision applications.

Multi-processor systems on a chip are becoming increasingly commonplace to achieve higher performance. The importance of power-efficient programmable architecture designs is increasing, and the amount of computation that is available at the lowest cost will eventually become relatively fixed. Engineers who design new smart cameras will increasingly select commercially available off-the-shelf system-on-a-chip processors that include many features that are overkill for the requirements – except the need for low cost.

A new trend seems likely to emerge: widespread availability of camera products that are designed to be programmed by the purchaser instead of the vendor. Then new smart camera systems will not need to be designed at all, for the most part, because cameras will be available in a variety of form factors and costs, ready to be programmed for custom vision, video and image processing applications. These cameras will enable the independent software vendor business model for vision applications software, to populate smart cameras and create new kinds of products that were previously cost-prohibitive.

Thus, for vision technology to be embedded in high-volume consumer products, the solutions will be provided in smart cameras. A strategy to develop such products is to look for where smart cameras are deployed now, or where they could be cost-effectively deployed to provide useful added value in the future, to see which vision applications can emerge next.

4.2 Generic Implications of Smart Cameras for Vision Research

Future smart cameras will provide powerful new tools for vision research. Analog cameras, frame grabbers and laboratory workstations will be displaced by smart cameras. Vision researchers will work from their office, from home, or

anywhere in the world for that matter, while conducting real-time data collection and algorithm development experiments with remote cameras directly via network. The smart cameras can be in a jungle on the other side of the world, in the depths of a mine or at the bottom of the sea, in controlled laboratory lighting conditions, in a car, or in a child's toy at home. Program the cameras for real-time data collection and in situ vision processing, and have the results emailed or streamed when they are ready. Or the remote camera can call a cellular phone to report results.

Vision research that aims to be useful to society someday, in the form of wearable, handheld or household appliances, must increasingly be computationally constrained. Whereas in the first forty years of vision systems research many computationally complex approaches could be justified by using the argument that processor technology may one day make applications affordable, that argument is going away. The premium will be on vision research that produces algorithms that can run on available smart cameras. We will not be able to afford vision research approaches that require custom algorithm-specific circuits for adequate performance unless the advantage is so compellingly valuable that it will be adopted by a mass market or command a very high price in a low-volume equipment market.

Traditionally, the viable locations of cameras have been extremely limited – and not many cameras in the world. As cellular phone cameras go into the hands of hundreds of millions of consumers, programmable cameras will be everywhere people are. The same technology will also enable cameras to be deployed in fixed positions that were cost-prohibitive to consider before. What new vision functions can these cameras be programmed to take on? While vision research has developed algorithms and prototypes that suggest many potential applications, there is a world of needs out there that vision research has only begun to address. New motivations will focus future vision research.

4.3 Future Human-Computer Interaction

As the field of human-computer interaction evolves to exploit smart cameras, new problems of human-system and human-environment interaction will arise. Some methods may find new applications as software add-ons for digital cameras, cellular phones, and personal data assistants. For example, multimodal human interaction methods could be adapted for smart camera systems. With a microphone and speaker already in camera phones, techniques that recognize speech and emotions using both audible and visual cues can be implemented using the same embedded processors. Thus interactive dialogue systems may emerge where smart cameras are deployed.

New ideas will lead to new kinds of handheld and wearable vision system tools. Among the possibilities: Gesture-based recognition algorithms can be implemented in personal smart cameras for rooms and vehicles to provide interactive remote controls. Mount a camera phone on the back of a bicycle

helmet to serve as a proactive “rear view mirror” monitoring system. And small smart cameras will enable new concepts for interactive toys and games.

Challenges for algorithms in this regard are to achieve sufficient robustness in wide-ranging imaging conditions. To be most useful, algorithms will need to operate reliably amid diverse ambient lighting conditions and backgrounds, indoors and out. Perhaps the biggest challenge facing the use of handheld vision systems for human interaction, or to devise new kinds of handheld vision tools, is that the cameras are moving during operation. In the classic human-computer interaction environment, the computer and connected camera(s) are in fixed positions, and the user is in a chair, providing key geometric constraints to help reduce the complexity of image and video analysis. Much more vision research is needed to develop reliable algorithms and applications for human interaction using smart cameras in motion.

4.4 Future Video Surveillance Systems

In the trend from closed-circuit video to network communications so far, most digital video security system products are essentially using the network as a replacement for analog video coax cable. For example, network video server equipment digitally compresses and transmits data from analog CCTV cameras to a remote network video storage system, or streams the data to a remote display for live observation.

While that approach provides several advantages and may be required for many security needs, smart cameras enable more. The traditional video surveillance security functions of centralized monitoring rooms will migrate to smart cameras, greatly reducing the overall cost of ownership and enabling new video security applications in homes, vehicles, schools, hospitals, etc. Using low-cost, high-capacity mass storage and system-on-a-chip processors embedded in smart network cameras to record video data and real-time observations from vision algorithms, centralized digital video storage systems will be avoided. Security personnel can obtain live or recorded video feeds direct from cameras via ordinary network communications when needed, without requiring separate network video server equipment. Traditional out-of-reach mounting positions of security cameras provide sufficient physical camera security for most applications, while real-time encryption algorithms and passwords protect smart camera data if the camera is stolen. In large campus installations, camera data can be backed up periodically on ordinary remote storage systems if needed, like computers are backed up, without requiring continuous streaming of video data to custom storage equipment.

But the big autonomous video surveillance and monitoring opportunities and challenges for vision research go far beyond the first-order cost-saving advantages of smart camera systems, and remain largely unrealized today. Security needs will continue to drive vision research for years to come, to help make the world a safer place. Smart camera systems will enable afford-

able deployment as research provides the useful, reliable, and computationally constrained algorithms.

4.5 Future Automotive Vision Systems

Automotive vision systems are starting to emerge. A prominent current example is the recent deployment of lane-departure warning camera systems in some car models in the industry. The economies of modular automotive designs, coupled with the expense of cabling, makes it preferable to co-locate the camera and the processor in such systems. As other automotive vision algorithms deploy, smart camera processors are likely to be adopted because automotive systems share the requirements outlined in Sect. 1.

A distinguishing challenge for automotive vision research is to assure a very high degree of reliability. Whereas limitations and mistakes of visual analysis may be acceptable or even exploited in smart camera applications such as interactive toys and games, the consequences of errors are clearly more serious for automotive safety. Vision research faces substantial challenges to collect sufficient image and video databases to measure reliability, in conjunction with human-system interaction research, to determine how reliable is reliable enough.

The large body of ongoing research and development for automotive vision systems is taking on this challenge to develop numerous new roles for smart cameras in cars. The longstanding quest for safe robotic driving continues, while research for other important automotive vision functions appears closer to improving safety for consumers. Stereo/multi-camera video analysis techniques may prove to be sufficient and cost-effective to meet increasing standards for air bag deployment safety. Prototype methods for visual analysis of drivers, to detect and alert if they start to fall asleep at the wheel, fit the smart camera approach as well. Examples of other smart camera applications in the works that seem likely to emerge for automotive safety include automatic monitoring of blind spots, obstacle detection, and adaptive cruise control.

What else will smart cameras do?

Acknowledgments

The observations in this chapter derive largely from lessons learned over the past twenty years in R&D projects for businesses of Texas Instruments, and involving the contributions of numerous others, but the speculative views expressed and any errors of fact that may appear are solely the author's.

References

1. A U Batur et al. A DSP-based approach for the implementation of face recognition algorithms. *Proc ICASSP*, 2003.
2. F Z Brill et al. Event recognition and reliability improvements for the Autonomous Video Surveillance System. *Proc Image Understanding Workshop*, 1998.
3. M Budagavi. Wireless MPEG-4 video communications. In: J G Proakis (Editor). *The Wiley Encyclopedia of Telecommunications*. Wiley, 2002.
4. T J Olson and F Z Brill. Moving object detection and event recognition algorithms for smart cameras. *Proc Image Understanding Workshop*, 1997.
5. B E Flinchbaugh. Advantages of software camera designs. *Electronic Products*, 2002. <http://www.electronicproducts.com>
6. D S Hall. Team Digital Auto Drive (DAD) White Paper. Personal communication, 2004.
7. T Hiers and M Webster. TMS320C6414T/15T/16TTM Power Consumption Summary. Application Report SPRAA45, Texas Instruments, 2004.
8. K Illgner et al. Programmable DSP platform for digital still camera. *Proc ICASSP*, 1999.