An Introduction to the Digital Still Camera Technology

Massimo Mancuso

Sebastiano Battiato

This paper gives an overview of the Digital Still Camera technology.

1. INTRODUCTION

The Digital Still Camera is a challenging application for Image Processing techniques. The basics of the algorithms applied to the data provided by the image sensor were first applied in camcorders. In fact, in the beginning, digital still cameras could be considered as a derivative of the camcorder technology, modified to capture still pictures.

But two factors contributed significantly in differentiating the two systems:

- the *Human Visual System* (HVS) characteristics;
- the quality of analog cameras.

It is well known that motion reduces the *Human Visual System* sensitivity to the high frequency contents - including artifacts - of a picture. Consequently, if the system is mainly intended for motion picture capture, the algorithms used can take advantage of that. Generally speaking, the complete system benefits from the reduced sensitivity of the HVS: simpler algorithms translate into less hardware, lower power consumption, and lower cost.

On the other hand, the image processing algorithms required in Digital Still Cameras are very demanding as far as quality is concerned due to the HVS characteristics for stills and the quality already provided by "Analog" or silver-halide still cameras.

Initially, Digital Still Cameras were strictly derived from Camcorders. Consequently, the demand in quality for DSC greatly differentiated the 2 systems. Today it is possible to say that camcorders benefit from the advances made to enhance the quality provided by DSC: for example, progressive sensors with Bayer CFA [1] developed for DSC now are used in Digital Camcorder.

Until recently, the DSC market focused in quality improvement and the reduction of the processing time, measured by the number of pictures which can be captured in a second. Quality improvement is obtained by increasing the resolution of the sensor and by using more sophisticated image processing algorithms. Now, with a sensor resolution reaching over 4 million pixels, and a picture acquisition speed of a fraction of a second, DSC competition among the different manufacturers is driven by new features (video clip acquisition, MP3 player, scanning of film negative etc).

The rest of the paper is organized as follows. The next section describes how digital cameras acquire images using CCD/CMOS sensors. Section 2 reports, in detail, the image generation pipeline from the input real scene to the final high quality picture. Section 3 is entirely devoted to the picture compression stage while section 4 reviews the various picture storage options, the display function and the various features provided by actual DSCs. A final section closes the paper showing possible evolution of the related scenario.

A digital still camera uses an electronic sensor to acquire the spatial variations in light intensity and then use image processing algorithms to reconstruct a color picture from the data provided by the sensor (see Fig. 1).

Two technologies exist to manufacture imaging sensors: CCD (Charge Coupled Device) or CMOS (Complementary Metal Oxide Semiconductor). While CMOS sensors will almost certainly improve and become more popular in the future, they probably won't replace CCD sensors in higher-end digital cameras. The CCD is a collection of tiny light-sensitive diodes, which convert photons (light) into electrons (electrical charge).

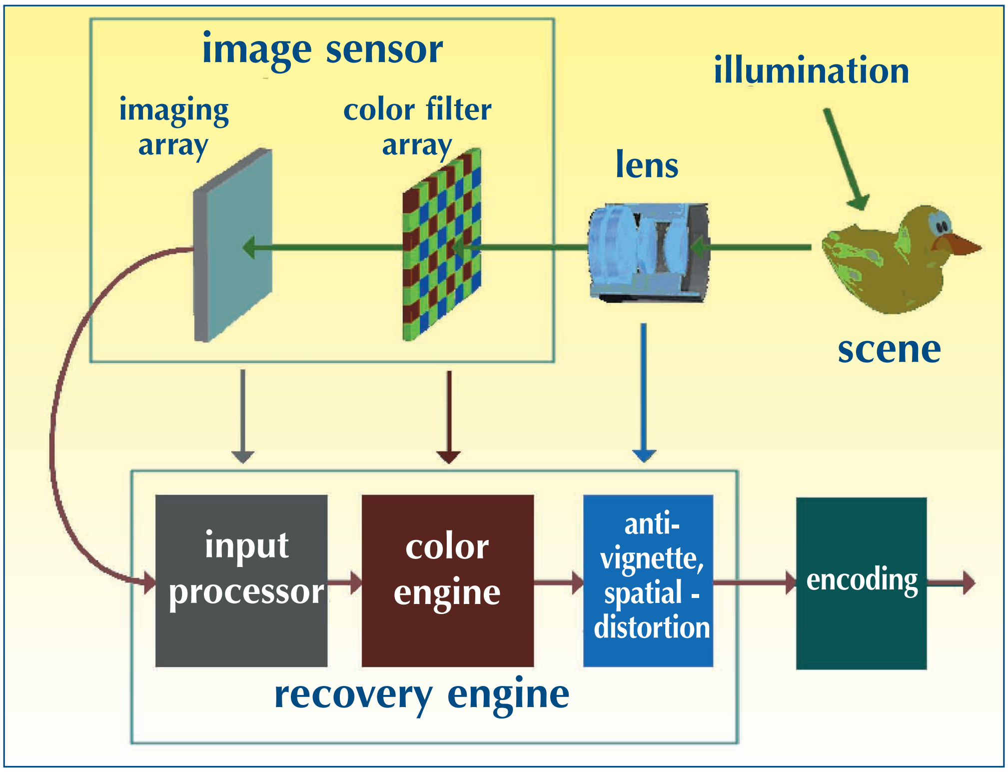

Figure 1:
DSC working principale

2. SENSORS AND DATA ACQUISITION

The working principle of the DSC is quite different from that of the conventional camera. Conventional cameras use a chemical reaction to capture the image, fixing it on film through an emulsion. The emulsion is composed by salt containing silver whose particles are sensitive to the quantum effect of light. Spatial variations of light intensity impacting the film appear as a picture.

These diodes are called *photosites*. Each *photosite* is sensitive to light: the brighter the light that hits a single *photosite*, the greater the electrical charge that will accumulate at that site. Both CCD and CMOS image sensors start at the same point: to convert light into electrons at the *photosites*. A simplified way to think about the sensor used in a digital camera (or camcorder) is to think of it as having a 2-D array of thousands or millions of tiny solar cells, each of which transforms the light from one small

portion of the image into electrons. Both CCD and CMOS devices perform this task using a variety of technologies.
The next step is to read the value (accumulated charge) of each cell in the image. In a CCD device, the charge is actually transported across the chip and read at one corner of the array. An analog-to-digital converter turns each pixel's value into a digital value. In most CMOS devices, there are several transistors at each pixel, which amplify and move the charge using more traditional wires. The CMOS approach is more flexible because each pixel can be read individually. CCDs use a special manufacturing process to create the ability to transport charge across the chip without distortion. This process leads to very high-quality sensors in terms of fidelity and light sensitivity. CMOS chips, on the other hand, use a normal manufacturing process to create the chip: the same silicon production line, so they tend to be extremely inexpensive compared to CCD sensors.
CCD sensors have been mass-produced for a longer period of time, so they are more mature. They tend to have higher quality pixels. Based on these differences, CCDs tend to be used in cameras that focus on high-quality images with lots of pixels and excellent light sensitivity. CMOS sensors usually have lower quality, lower resolution and lower sensitivity. However, CMOS cameras are much less expensive and have great battery life. Over time, CMOS sensors will improve to the point where they reach near parity with CCD devices in most applications, but they are not there yet. In any case, the sensor architecture and physical layout is essentially independent of the technology; for the purpose of this tutorial both CCD and CMOS can be thought as nearly identical devices.

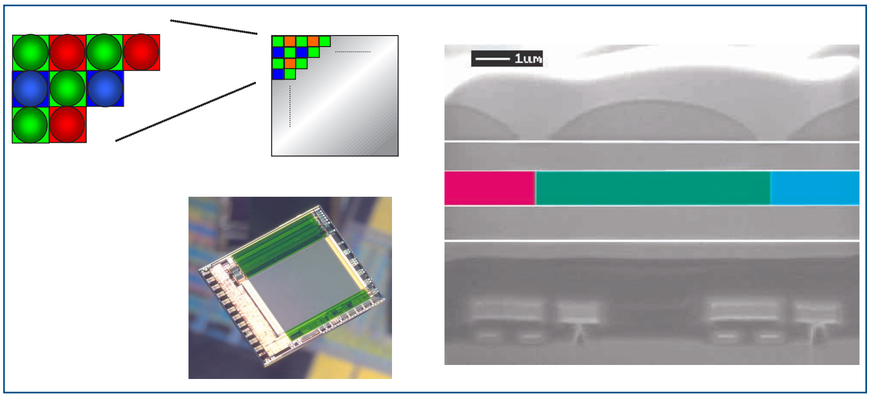

Figure 2:
A sensor color is obtained by a mono sensor plus suitable RGB color filters. Each pixel/color filter has a small lens a) placed on top, improving significantly light gathering at pixel site. On the right b) it is showed the typical layout of a single CMOS sensor.

process used to make most microprocessors. Because of the manufacturing differences, there are several noticeable differences between CCD and CMOS sensors.
CCD sensors create high-quality, low-noise images. CMOS sensors, traditionally, are more susceptible to noise.
Because each pixel on a CMOS sensor has several transistors located next to it, the light sensitivity of a CMOS chip is lower. Many of the photons hitting the chip hit the transistors instead of the photodiode.
CMOS sensors traditionally consume little power. Implementing a sensor in CMOS yields a low-power sensor.
CCDs, on the other hand, use a special process that consumes lots of power. CCDs consume as much as 100 times more power than an equivalent CMOS sensor.
CMOS chips can be fabricated on just about any standard

Each sensitive element of the sensor (known as pixel) is sensitive to one color component only. This is obtained through the deposition of color filters on top of a monochrome sensor. Filters are divided into primary and complementary colors. The first ones exhibit excellent color reproduction, but are less sensitive than the complementary. Primary filters are arranged into a pattern known as Bayer pattern. Picture quality is strictly related to the number of pixels composing the sensor: the higher the better. The amount of detail that the camera can capture is called the resolution, and it is measured in pixels. In general, the more pixels your camera has, the more detail it can capture. But increasing the sensor resolution without increasing its size reduces the area of individual pixels, and therefore their sensitivity. Micro-lenses can be used to

increase the sensor sensitivity by focusing more light on the pixel (see Fig.2).
Notice in Figure 2 the particular arrangement of the color filter array over the pixels, in what is called a Bayer pattern. Half of all pixels are green versus a quarter each for blue and red pixels. This particular arrangement relies on the higher sensitivity of our eyes to the green color. Therefore, what comes out of the sensor is one color component, red, green or blue, per sensor pixel. To construct a color image from the sensor's output, an RGB triplet must be computed for each pixel, in a filtering operation known as color interpolation.

The reconstruction process must guarantee the rendering of a high quality images avoiding typical artifacts that, due to the acquisition process, could be present. For this reason powerful and smart algorithms are applied to enhance quality in a sort of chain known as DSC pipeline (see Figures 3 and 5). They are detailed in the following section.

3. IMAGE PROCESSING AND DSC

3.1 Pre Capture Algorithms

During the Pre-Capture phase, just before the actual picture is captured, the sensor is read continuously and the output is analyzed in order to determine three parameters which will determine the quality of the final picture: white balancing, exposure and focus.

Auto-White-Balancing (AWB for short) compensates automatically for the dominant "color" of the scene. The human eye is able to compensate colors automatically through a characteristic known as *Color Constancy*, by which the color white, in particular, is always perceived as white independently of the spectral characteristics of the light source illuminating the scene. When a scene is captured on a picture, the illuminating context is lost, color constancy does not hold anymore, and white balancing is required to compensate colors. AWB relies on the analysis of the picture in order to match the white with a reference white point. White balance adjustment attempts to reproduce colors naturally so images are not affected by surrounding light.
To do that classical techniques use either simple global measure of energy of the scene analyzing the relative distribution of the various chromatic channels or try to adapt the white color to the particular light condition (sunset, ...). *Auto-white-balancing* is sufficient for most conditions, but if there is no near white color in the picture, colors that are not originally white may appear white in the image and the white balance of the image may not be correct. Also, *Auto-white-balancing* may not have the desired effect when shooting under white fluorescent or other fluorescent lights. In such cases, some cameras offer the possibility to use a white surface and quick reference white balance to achieve the correct white balance, or use preset white balance to select a color temperature for the incident light. Alternatively, it is possible to use preset white balancing to reproduce more red in a picture of a sunset, or capture a warmer artistic effect under artificial lighting.

AutoExposure determines the amount of light hitting the sensor and the sensor itself is used for light metering. The exposure - the amount of light that reaches the image sensor-determines how light or dark the resulting photograph will be. When the shutter opens, light strikes the image sensor inside the camera. If too much light strikes it, the photograph will be overexposed-washed out and faded. Too little light produces an underexposed photograph-dark and lacking in details, especially in shadow areas. To measure the light reflecting from the scene, a camera uses built in light meters. The part of the scene they measure makes all of the difference in the world. Most read the entire image area but give more emphasis to the bottom part of the scene because this reduces the possibility that the bright sky will cause the picture to be underexposed. They also emphasize the center of the image area based on the assumption the major subject is placed there. This is called a *center-weighted* system. Some system allows the user to select a small area of the scene and meter it directly using a spot meter. In this mode, only the part of the scene in the center of the viewfinder is metered.

Auto-Focus techniques are more proprietary and vary from one manufacturer to the other. The Auto-Focus algorithm directly affects picture sharpness. Essentially, it consists on extracting a measure of the high frequency content of the picture and changing the focus setting until this measure reaches a maximum. The article "In-focus image recovery through focus measure analysis" by Salvatore Curti examines in detail one such algorithm.

3.2 Post Capture Algorithms

Once the picture is taken a number of different techniques such as *Defect Correction, Noise Reduction* and *Color Correction* are applied to compensate/enhance the sensor output data. *Defect correction* manages pixel defects related to the sensor and/ or to the memory storing the picture. When systems on a chip solution for DSC are considered, both sensor and memory can be part of a more complex device. Exploiting the redundancy of image data, these defects can be corrected in a complete transparent way for the DSC manufacturer. *Noise Reduction* is performed to limit the visible effects of an electronic error (or interference) in the final image from a digital camera. Noise is a function of how well the sensor (CCD/CMOS) and digital signal processing systems inside the digital camera are prone to and can cope with or remove these errors (or interference). Visible noise in a digital image is often affected by temperature (high worse, low better) and ISO sensitivity (high worse, low better). *Color Correction* simply adjusts by mathematical operations the RGB components of a color separation and creates a new RGB output based on the relative values for the input components. Also called color matrixing or color mixing.

The key point of the image processing pipeline used in digital still cameras are the algorithms themselves but, also the kind of data they are applied to. Classical image processing algorithms are applied to color images simply working independently at the same manner in each color planes. As far as noise reduction is concerned, several references can be found on algorithms able to reduce noise without smoothing edges. Normally they are applied to gray level images. To use them on color pictures, it is common practice, to replicate the filter for the chromatic components (R, G, B). This will lead to a higher complexity and to possible false color introduction along the edges.
In a DSC, a better solution is working in the Bayer pattern domain so that all the subsequent algorithms will benefit of

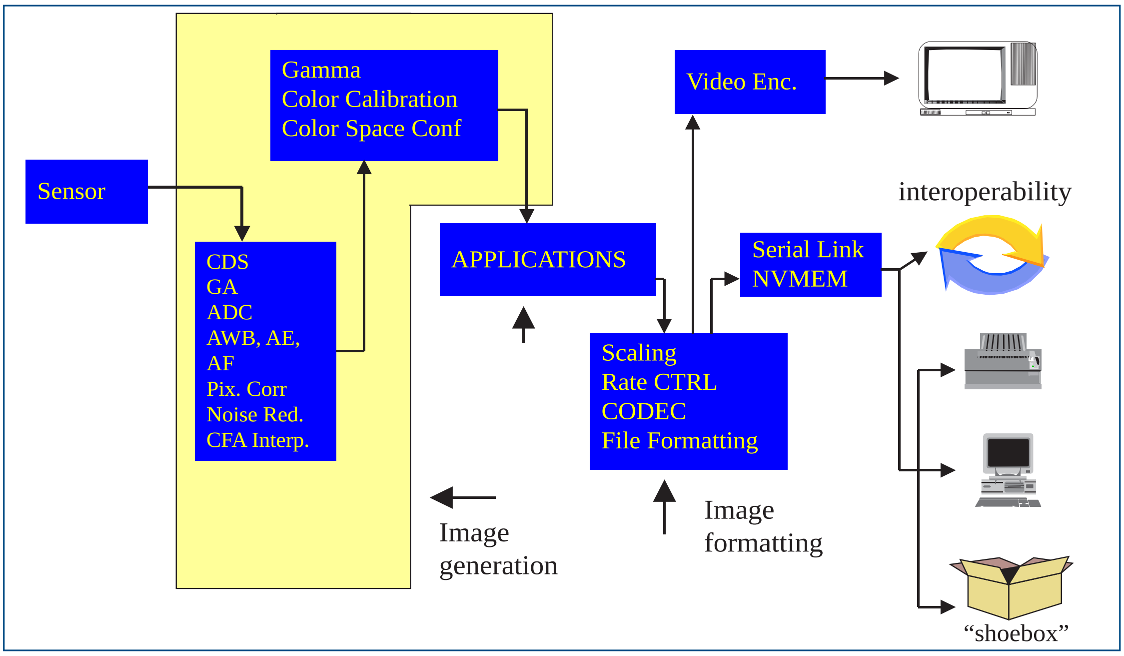

Figure 3: Typical DSC Imaging Pipeline

the "noisy-free" data, and the complexity will be lower than working on a RGB domain. Similar consideration can be developed for the scaling algorithm. Working in the Bayer domain requires a little effort to readapt ideas and techniques to the new particular environment but allows improving significantly quality of the final image reducing at the same time the associated computational complexity.

The *Color Interpolation* algorithm converts the sensor output to a true color image (e.g. RGB 24 bits/pixel), reconstructing for each pixel the three chromatic components. Using simple interpolation techniques (i.e. median or bicubic filtering) introduces aliasing and/or unpleasing artifacts; for this reasons each manufacturer applies proprietary solutions. Figure 4 illustrates the quality difference between a straight bicubic interpolation and an ST proprietary algorithm.

Typically, the camera includes additional image processing algorithms (proprietary or not, depending on the specific application) to sharpen, scale, etc.. A brief mention to the scaling algorithm used to downsample the input image in order to achieve the desired resolution (e.g an alternative way to achieve compression) saving computational resources. Before compression, usually, the data is transformed into a suitable color space splitting the luminance component from the chromatic components. Often such operations, together with chromatic subsampling, is one of the first steps performed before image/video compression. This is simply due to the fact that the human eye is more sensitive to luminance than chrominance.

All these algorithms contribute to the final quality of the image, and constitute one of the key features distinguishing one camera from another.

4. COMPRESSION

Managing high quality (e.g. 24 bit/pixel) digital images requires having enough on-board memory to do it. But in order to store them physically for later usage, an ad-hoc compression algorithm "must" be used in order to reduce the amount of physical bytes needed maintaining a very high quality. The subject of digital data compression divides neatly into two categories: lossless compression, in which exact recovery of the original data is ensured; and lossy compression, in which only an approximate reconstruction is available. Table 1 summarizes the main difference between the most classical uncompressed format, the TIFF (Tagged Image File Format) format, that requires 24 bit/pixel and the Jpeg format able to compress image without noticeable difference (visually lossless compression) using about 4 bit/pixel. Such kinds of performance are achieved using redundant and irrelevant data present in the original uncompressed image. The redundancy is related with repetitious, inherent in statistical

Figure 4: Images obtained using different color interpolation algorithms: classical bicubic (on the left), ST proprietary algorithm (on the right)

nature, having to do with similarities, correlation and predictability of the data. It can be either "spatial" (e.g. characterized by correlation of neighboring pixels) or "psycho-visual" (e.g. can be removed without complaint from a human observer). The irrelevant, instead, relates to an observer viewing an image: function of image resolution, noise, detail, and viewing conditions.
The degree of redundancy determines how much compression can be made.

A good quality image having, for example, dimension of 1024x768 pixels, requires 3 bytes for each pixel to represent all its colors (in an RGB color space, also called true-color). This implies that: 1024x768x3x8 = 18.874 Mbit are needed to represent the image.
To overcome this the lossy compression, the process of reducing the amount of information needed to represent and image without visibly deteriorating, is vastly used. Some cameras, also, allow exportation of data directly in "RAW" format. In this case the data is technically formatted in proprietary ways, describes the picture in the bayer check board pattern mentioned before. Such a feature can be used from a professional photograph working with original input data in order to apply your own enhancement techniques.

Format	bits/pixel	Quality	Comment
RAW	10-12	Lossless	Bayer Pattern
TIFF	24	Lossless	Very large files High quality
JPEG	2-4	Visually Lossless	Small/Medium size with Good/High quality

Table 1 - Main image formats.

The standard JPEG [3] allows managing properly the right trade-off between final quality and desired compression size. In the DSC environment, the JPEG compression stage has to obey two constraints:

- the compressed file size must not exceed a fixed size, determine from the camera settings such as picture quality and size, in other words, the compression factor is fixed
- the picture quality must be maximized for the given compression factor,

The first constraint stems from the fact that JPEG compression does not explicitly guarantee a compressed file size, nevertheless, the system must be predictable regarding the number of pictures that can be captured and stored. From a user perspective, it is not acceptable to tell the user the picture he just took cannot be saved. Commonly used algorithms are based on iterating the compression several times until the target compression factor is obtained. This, of course, will have impact on the processing time per picture.

The second constraint is quite obvious, since quality is a differentiating factor among DSCs. In JPEG, the primary means of controlling quality is the tuning of the quantization tables. Everything else being equal, selecting the right quantization tables can lead to a 10-20% improvement in compression factor or the corresponding increase in quality (2 DB on the average) for a fixed size.

Papers in this issue address these two issues in particular.

5. IMAGE DISPLAY, STORAGE AND NEW FEATURES

5.1 Image Display

Image output is the final stage of the image processing steps described so far. The basic forms of outputting the picture are, the LCD of the camera, the TV screen. Having the possibility of pre-viewing the picture immediately after it is taken is a feature that digital still camera users greatly appreciate. But the LCD, jointly with the flash-gun, is one of the power-hungry feature of the DSC. Some cameras have no LCD panel, and instead use a simple optical viewfinder. Other cameras have both an LCD panel and an optical viewfinder, in order to save the battery. The same LCD panel, can also act as the viewfinder. It also allows reviewing images in memory, deleting images and taking more pictures in their place.
Some newer cameras have advanced features for the

LCD (i.e. zooming in on parts of the image and see them in greater detail). Others offer an "MPEG movie" feature, allowing one to take short movies. It is important to mention that the display function requires the presence of a PAL/NTSC video encoder on chip.

5.2 Storage

Early generations of digital cameras had fixed storage inside the camera. To get the pictures out, they needed to be hooked up directly to a computer by cables so that the images could be transferred. Although most of today's cameras are capable of the connecting to a serial, parallel, SCSI, and/or USB ports, they usually are provided with some sort of removable storage device. Digital Still Cameras use flash memory to store pictures. The memory is available through removable flash-card of different capacity and type. The main are: Compact Flash, Smart Media and Memory Stick. These three are all small, removable, flash memory devices that have no moving parts. They are fast, inexpensive ways of store photos in order to transfer to a computer or printer later.

In order to transfer the files from a flash memory device to a personal computer without using cables, one needs a drive or other suitable reader-device. These devices behave much like floppy drives and are inexpensive to buy.

Micro hard disks are appearing as mass memory for the digital still cameras. Users will benefit of the augmented capacity by storing video clip.

5.3 New Features

Features allowing a number of different image effects are becoming increasingly common. This allows the selection of monochrome, negative and sepia modes. Apart from their use for artistic effect, the monochrome mode is useful for capturing images of documents for subsequent optical character recognition (OCR). Some digital cameras also provide a 'sports' mode - which adds sharpness to the captured images of moving objects - and a 'night shooting' mode that allows for long exposures.

Panoramic modes differ in their degree of complexity. At the simpler end of the spectrum is the option for a letterbox aspect image that simply trims off the top and the bottom

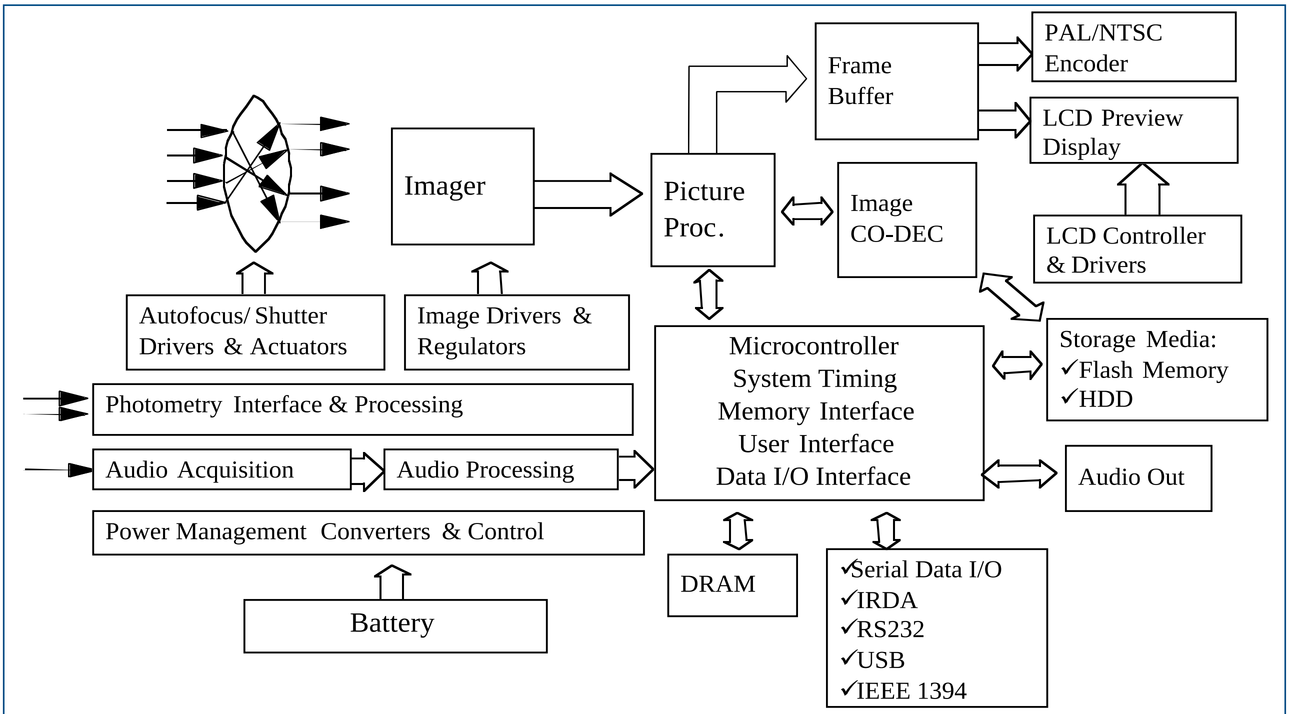

Figure 5: DSC: functional block diagram

edges of a standard image - taking up less storage space in the process. More esoteric is the ability to produce pseudo-panoramic shots by capturing a series of images and then combining them into a single panoramic landscape using special-purpose software.

A self-timer is a common feature, typically providing a 10 - second delay between the time the shutter is activated and when the picture is taken and all modern day digital cameras have a built-in automatic flash, with a manual override option. Typical settings allow to have a working range of up about 20', providing also a number of different modes, such as auto lowlight and backlight flash, fill flash for bright lighting shadow reduction, force-off for indoor and mood photography and red-eye reduction. Another feature commonly available with film cameras that is now available on their digital counterparts is the ability to watermark a picture with a date and time, or indeed some other chosen text. Yet, the recent innovation of built-in microphones provides for sound annotation, in standard "WAV" format. After recording, this sound can be sent to an external device for playback, or played back on headphones using an ear socket. A couple of other features which demonstrate the digital camera's close coupling with other aspects of PC technology are a function that allows thumbnail images to be emailed directly by camera-resident software and the ability to capture short video clips that can be stored in MPEG-like format.

6. Conclusions

The rapid growth of the digital still camera market segment is due to 3 factors mainly: the wider diffusion of personal computers, the increasing quality of the digital pictures and the introduction of new features enabled by the digital technology. To produce photo-realistic images, DSCs require significant amount of image processing. In addition, standard image processing features like image capture and playback, multimedia features such as video/speech capture and audio playbacks are becoming popular in consumers DSCs.

The high rate of innovation in the image processing technology used in DSCs includes changing the product features. The new features are not only related to functions implemented on board, but also to the availability of an infrastructure enabling the access to services such as, on line photo album, printing on demand, etc. The increased bandwidth and service capability of the 2G and 3G wireless communication is expected to contribute to this trend. Services such as Multimedia Messaging and Mobile Internet will contribute to make image acquisition more pervasive (imaging enabled mobile phones), but at the same time will make easier to access to remote imaging services through a link between DSC and wireless linked mobile devices.

The new JPEG2000 standard [4] is promising a lot as far as compression efficiency and features to be supported at a cost of a higher complexity with respect to JPEG. It provides a new image representation with a rich set of features, all supported within the same compressed bit-stream that can address a variety of existing and emerging compression applications. In particular, the Part 1 of the standard addresses some of the shortcomings of baseline JPEG by supporting the following set of features: improved compression efficiency, lossy to lossless compression, multiple resolution representation, embedded bit-stream (progressive decoding and SNR scalability), tiling, region-of-interest (ROI) coding, error resilience, random codestream access and processing, improved performance to multiple compression/ decompression cycles, a more flexible file format.

The complexity and the fact that in the range of visually loss-less compression factor the compression efficiency on JPEG2000 vs. JPEG is in the range 10~15% higher, both make JPEG2000 attractive for its flexibility and features that must be exploited by a killer application.

References

[1] B.E. Bayer, *Color imaging array,* U.S. Patent 3,971,065-1976;

[2] W.K. Pratt, *Digital Image Processing*,
New York: Wiley- Interscience, Ch. 10, 1978;

[3] W.B. Pennebaker, J.L Mitchell,
JPEG Still Image Data Compression Standard,
New York, NY, Van Nostrand Reinhold, 1993;

[4] Information Technology - JPEG2000 Image Coding System, ISO/IEC International Standard 15444-1,
ITU Recommendation T.800, 2000.