Digital Image Processing in CCTV

By Nikolai Uvarov, a Russian CCTV expert from Bezopasnost company, Moscow.

Digital technologies in CCTV systems are primarily identified as video data transmission via communication links and recording (back up) on hard magnetic disks or integrated digital memory. Such application of digital technologies is aimed at improving the immunity and decreasing the redundancy. Digital technologies are also applied as image (video data) processing for specific purposes in various applications of science and engineering, space research and defense, i.e. where TV systems are used for receiving video information and displaying spatial distribution of features of invisible physical fields. For instance, digital technologies may be used for improving the quality of biological and medical images, including X-ray images and image thermograms in isotopic and ultrasonic diagnostics as well as for improving the contrast and definition of images created by electronic microscopes and industrial X-ray equipment, correcting distorted images received from space, performing automatic analysis of terrain characteristics, exploring natural resources by spectra-zonal TV systems, etc. [1].

Generally speaking, image processing is a type of signal transformation. As for the salient features of an image represented as a signal, they are as follows:

- First, this signal is bi- (monochrome image) or multi-dimensional (spectra-zonal or color image) – the essential thing here is random links, not the display format (e.g. multi-dimensional display of white noise does not create an image);

- Second, an image, as a rule, is displayed in such a way that dimensional characteristics of its display form do not match the dimensional characteristics of the image itself (e.g. TV images are displayed by means of mono-dimensional video signals);

- Third, an image is always displayed in a finite interval: it is confined within a frame – hence, care should be taken in the usage of true spectral estimators in infinite intervals, in issues related to ensuring isotropic processing in edge zones, etc.

Having this interpretation of images as signals in mind, the return to the original space for the sake of exploring and changing the correlations (they are the only things that give meaning to an image, as the white noise image is deprived of intelligible information) is, perhaps, possible only with the help of digital technologies. Digital technologies permit to interpret temporal distribution of a feature as spatial distribution of digital readouts in memory cells with further fragmental access to the readouts for the sake of processing.

Two questions always emerge in image processing and in signal processing, in general: why to process and by what means. In TV - the processing is traditionally done to get a better view. Historically, during the TV evolution image transmission & display equipment were always improving. Since the early times of TV image processing has been considered as an instrument for correcting hardware errors. In the early days of television image processing was in the form of, for example, aperture processing or shading correction – the former type of processing has been still flourishing on up to now, but the latter has already been long forgotten. The basic mission of processing then was to ensure high precision of display, i.e. image restoration. Such processing operation was aimed at transforming the image parameter in such a way as to make up for expected hardware-originated distortions, ensuring the principle of video data transmission accuracy.

Nevertheless, it was becoming clear even in these early development stages that display accuracy was not always an advantage. Is there any need in accurate display of illumination, if in nature it varies from 100 000 lx on a bright sunny day to 100 lx in twilight, or even to 0.005 lx on a dark night? Who needs the image displayed at a luminance equivalent to 0.005 lx illumination level on a monitor? So display accuracy is not needed, at least, as far as illumination is concerned. Here one must adapt the equivalent luminance of a monitor to the conditions under which the image is watched, i.e. *to normalized luminance*. Hence, there is another mission of processing – bringing an image parameter to its optimal value, i.e. normalization.

In TV applications a TV system, as a rule, must help actively to ensure a TV message continuously maintains the optimal (normalised) conditions. This is done by means of image pre-processing.

The third mission of image processing – noise suppression – requires no special explanation. The need for suppression of fluctuation, background and impulse noises is quite evident in all types of TV.

The fourth mission of image processing that stands apart is related to a semantic analysis, i.e. semantic decoding of images. An image is analyzed when you want to learn something new from it. It is not known for certain, how people (human brains) manage to do it. In known engineering practice the image analysis process is normally a combination of measurements and solutions, which is often rather complicated and recurrent.

The issue of selecting the processing means it cannot be considered separately from the processing content. For instance, it is known that a rapid form of an elementary integral Walsh-Adamar transform for a mono-dimensional sequence from N readouts requires N $\log_2 N$ operations of addition and subtraction [2], and a bi-dimensional transformation needs N^2-fold recurrence of this procedure. Thus, $N^3 \log_2 N$ operations are needed to process

www.cctv-focus.com

a square picture segment with a side of N readouts, i.e. about 30 operations per picture element will be needed to make an elementary high-frequency processing of an image segment 3 x 3. Therefore, at least 10 million readouts per second will be needed for a standard TV camera with a video-frequency bandwidth of 5 MHz to get a non-distorted display of the signal. The processor speed required for processing an image segment of 3 x 3 at the TV system operation rate will be 300 million operations per second. Increase of an image segment size will sharply increase the requirements processing speed. There are a few ways to decrease the requirements in hardware computing power. One is, for example, by means of parallel multi-processors. Another would be by extending the processing procedure for the time required for generation of all elements of an image segment. So far, most of these methods are too exotic to be applied in surveillance and security CCTV systems, which normally consist of tens or sometimes even hundreds of TV cameras.

One possible way to perform image processing consists in creating customized computing machine, at the cost of abandoning the advantages of a general-purpose computer. This seems to be the way security CCTV equipment developers and manufacturers have chosen to follow. We have to be aware though that there are essential limitations here. For example, to isolate low-frequency spatial components, the whole picture measuring 600 x 800 elements has to be processed as a whole, not by individual segments. This is, of course, unachievable at the TV system rate even at the cost of programmed control abandonment.

That is why "outdated" methods of analogue engineering are still used for image processing in security systems. Here, the pragmatic approach to selection a method proved to be justified, i.e. selecting the method that minimizes the cost under the circumstances when pre-set constraints have been applied to processing quality as well as to weight and size of the equipment. With this approach in mind, it turns out that the selection of a method depends on the physical details of the image transformed by this function. Then image parameters may be grouped into:

1. Spatial distribution parameters (e.g. spectral content, geometrical similarity, etc.);

2. Level distribution parameters (e.g. luminance, contrast, saturation);

3. Parameters of reciprocal statistical relationship between multi-dimensional signal components (e.g. segment wise inter frame correlation, reciprocal correlation of chrominance signals);

4. Parameters of image variation in time (e.g. shift rate, display lag, etc.).

If a processing operation transforms a parameter in such a way as to make up for hardware-originated distortions, the image restoration method will be applied. In this case known hardware-originated errors are considered (e.g. photoelectric transformation errors). If a parameter is brought to its optimal value, the normalization will be done OK. In this case the most essential thing is the criteria for image quality as seen by the TV recipient. If in this parameter the image distribution differs from noise distribution, this difference may be used for signal selection in the noise background, i.e. for achieving noise suppression and proper Signal/Noise ratio. If the distribution differences are result from the picture elements, processing of distribution in this parameter will permit to select picture elements for semantic analysis. During the semantic analysis the properties of the image itself are considered (e.g. sizes of the objects, relationship of contours, parameters of motion, etc.).

CCTV systems in security and video monitoring are characterized, as a rule, by the fact that the number of cameras is considerably larger than the number of monitors. Therefore, video signal processing hardware is traditionally embedded directly in the cameras. This allows: (1) to apply customized computing devices adapted to a specific photoelectric converter and (2) to perform the most important mission – normalizing the video signals as it comes from the camera. Indeed, it is practically impossible (and inappropriate) to arrange video data processing "on the fly" in a single customized digital device. It is much simpler and more reliable for the video switching, recording and displaying equipment to operate with normalized video signals in CCTV network.

Maintaining the stable video signal swing at the camera output under the conditions of varying illumination of a scene is the most important function of video signal normalization. To achieve this goal, special mechanisms are applied in the camera for controlling its sensitivity. It is important here that electronic devices of the CCTV camera and the CCTV operator together evaluate (measure) and set the video signal value, which ensures optimal conditions for displaying visual scenes on the monitor at the protected facility. This measurement is not easy, because a CCTV video signal is uni-polar by its nature and its form is quite different from the form of harmonic functions, whereupon the traditional notion of an amplitude loses its sense. It is also wrong to equate the value of a video signal with its average value or with the mean square (standard) deviation, as deviations of real video signals from the normal distribution law are very considerable.

Therefore, in actual practice we have to either limit the abnormal readouts prior to measurement, or to refer to stable (robust) response of the signal values. Any constrains of the signal and making it stick to a frame of specific parameters will always be faced with the risk of neglecting something that afterwards may turn out to be extremely important. In cases like these non-parametric evaluation of signals by plotting level distribution histograms is preferred. This can be done correctly only during the digital processing of video data of *the whole frame* – and as such is implemented in modern digital CCTV cameras. Besides, to ensure stable operation of automatic control loops used for controlling camera sensitivity, a sequence of video signal level evaluations made per frame is subject to statistical processing by means of a digital filter with the parameters adapted to varying input illumination. Thus, the traditional evaluation of video signal level in analogue cameras by means of peak- or quasi-peak detection, which was faced with inadequate selection of peak value for luminance distribution in the scene, is past. It is known from practice that use of analogue cameras and evaluating video signal close to the average value leads to bleeding whites on luminous segments of the picture. On the other hand, evaluating signals close to the peak video leads to "blinding" when the whole picture fades to black, like when some small shining or bright object appears in the camera view. The above-mentioned example cannot be applied when using PTZ cameras with unequal illumination distribution pattern in a frame, which is sometimes subject to rather sharp changes in time.

Digital algorithms for measuring video signal levels and processing of sensitivity control allow for a more precise implementation of the functions for background illumination compensation, automatic black level compensation and electronic exposure control.

Fig. 1 – Picture of a low video signal prior to (top) and after (bottom) compensation of black level

Digital algorithms allow to program scheduled changes in the operation modes or when external conditions change. Contrary to analogue cameras (whose functions are uniquely determined by hardware) the functions of digital cameras depend on the software and thus are highly flexible. Their abilities for adaptation to specific conditions are practically unlimited.

At the same time digital technologies permit to implement service functions for installation and maintenance of video camera lenses. For example, the "Lens Wizard" service menu mode on some cameras enables automatic detection of the lens iris drive as "Manual", "DC Driven AI", or "Video Driven AI." It is also possible to see a display of video signal level scale or digital readout during its adjustment and set full opening of the automatic lens iris in order to adjust the back-focus optimally.

The advantages of digital technologies are evident, and there are a few camera manufacturers in CCTV taking a full advantage of the video information processing. One such manufacturer that was amongst the first to apply such technologies is Bosch, with their "Dinion-XF" cameras. In addition to the traditional DSP camera processing (contour imaging, gamma correction, background light compensation, black level compensation and automatic white balance) which are superbly implemented in "Dinion-XF", for the first time a function of contrast normalization on the basis of luminance signal distribution level histogram equalization is implemented. This function allows for contrast increase of important subject details on the image, especially important when there is a considerable difference in luminance, be that a light-and-shade, or non-uniform illumination of objects.

Fig. 2 – Picture prior to (top) and after (bottom) level distribution equalization

Beside the traditional interpolation technique applied to values of RGB-signals typical for each pixel of CCD sensor, a quite radical form of reciprocals normalization of the colour separation signals (de-correlation) has been introduced into the "Dinion-XF." This de-correlation has been implemented like a standard colour restoration, but matrix coefficients are selected in an adaptive way, so as to delete components common for all signals and to leave the distinctive components, which permit to distinguish separate signals. Of course, the de-correlation distorts color tints, but it also permits to increase saturation sharply, thus assisting in detection of extremely weak color contrasts, which is very important in detection and identification.

Fig. 3 – Original picture (top) and picture after normalizing colour-separation signals (bottom)

Fig. 4 – Colour picture under low illumination conditions (top) and change-over to monochrome mode with noise reduction system (bottom)

The 15-bit digital video signal processing technology applied in "Dinion-XF" cameras combined with the powerful DSP video processor chip – "Hercules" - it ensures high-quality video data processing allowing operators to perform detection and identification operations successfully, even under very difficult visual conditions. The dynamic noise reduction system (DNR) applied in this camera and combined with super-high sensitivity and high signal/noise ratio allows for high-definition images with thorough half-tone processing and good colour saturation even under low illumination conditions.

A four-zone video motion detector, with individual configuration of parameters for each zone, has been implemented in the "Dinion-XF" cameras. Also, it is possible to compensate for the losses due to the coaxial cable length and type. A small and special hardware/software device is provided for remote adjustment via the coaxial cable, as well as firmware updates (by Bilinx Technology). The setup menu control and configuring makes operations related to its installation and maintenance considerably easier [3].

Beside the above-mentioned methods of video signal processing, in most modern PTZ dome cameras the original video information is processed by spatial frequency parameters so as to provide automatic control functions, such as, for example, focusing. In this case, non-parametric evaluation is also applied for image analysis by plotting spectral distributions and auto correlation functions, which, if taken alone, cannot serve efficiently as direct control signals. Nevertheless, these results of measurements are brought to individual control parameters by actuating mechanisms with the help of bi-dimensional spatial processing algorithms. Frequency parameters expressing average spectral density of frequency range are used successfully, for example, for optimizing peak-holding control in auto-focusing systems. But when an image is analyzed, not in itself, but as part of some assembly of images for example, video sequence of frames, the degrees of their mutual relationship are assessed. So, the frame-to-frame difference module averaging provides for an estimation of general speed of action, while the averaging of readout products of optional correlated images provides for the factor of the co-variances showing the statistical degree of their similarity. Based on similar procedures implemented in the DSP, the correlation algorithms allows for calculation of the value of reciprocal coordinate overlay error of correlated images; in automatic positioning systems this value serves as the control action, while in tracking systems it assures tracking of mobile objects.

All parameters calculated need not be applied to the whole picture – digital technologies permit to calculate them by fragments, thus obtaining the spatial distribution of relevant feature within the original picture. This fact is widely used in identification of images, in motion detectors as well as in plotting and display of motion traces on monitors. Yet, it must be kept in mind that, as the size of the fragments decreases, the reliability of correlation evaluations also decreases and the spatial frequency analysis gradually loses spectral selectivity.

Without going any further in describing the use of digital technologies in CCTV cameras - we shall only mark their development tendency in regards to the individual processing means to achieve the goal: “to have a better view of what needs to be seen”. Here, the prevailing criteria is the parameter normalization of the image displayed on a monitor directly from the video cameras. If this conclusion is true, we can assert that other CCTV equipment is related to functions of video data distribution, switching and recording. In this regard - the digital technologies, as opposed to analogue, performs practically without any losses in quality, which is very important.

It must be noted that the tendency of equipping CCTV cameras with the simplest video motion detectors can be understood, as motion detectors are typical of CCTV. If however, a more detailed semantic analysis of images is required (like for example, detecting inappropriate items left in a protected area, or motion detection under poor visibility) then use of more sophisticated video processing is required.

Semantic decoding of images is one of the most complex types of video processing. In surveillance such decoding is basically reduced to comparative analysis of current images and the pre-stored reference images. Detection of similarity between correlated pictures, or their fragments, may be assessed by the difference between the current and the reference video signals. The storage space required for recording of a reference image may be decreased by recording only its individual parameters, provided, of course, that assessed correlation of images may be expressed through correlation of their parameters.

The parameters fit for this purpose, among others, are integral transformation factors (transformants). Two-fold Walsh and Fourier transforms are most often used in this case. Sometimes contour image descriptions are used. Although they are complex and less compact than the description through individual transformants, they assure considerably more precise and stable evaluation. The procedure of description based on transformants or contours is similar to a special filter that selects only the features, which a sufficient to solve the task. As for the differences expressed by discarded features, they cannot be detected by comparison only. Correlation algorithms related to detection of maximum likelihood function for correlation of current and reference images do not permit any essential losses in description. Nevertheless, they are rather intolerant of scale-and-angle disparities between analyzed images and reference images. This is why, for example, in facial recognition applications - fixed positioning of face relative to the camera is required, while, for retinal/iris recognition the analysis must be concentrated on one point.

Extensive efforts have been taken lately to create survey-and-retrieval systems capable of remote survey and visual identification of people and vehicles. Here a current image is transformed by means of topological transformations with the help of digital technologies from real space into the space of reference. If accuracy of the transformation is not sufficient, further analysis becomes considerably more difficult and correct identification of an object becomes less probable. It is with the aid of these intelligent survey-and-retrieval systems designed for protected space scanning, that an operator is spared from routine and weary viewing many monitors and entrusted only with the decision-making function in contingency cases. It should be noted that the processing of these duties are not so intensive, as is the digital recording and distribution of video information over the digital networks.

Remote monitoring and control of systems are always associated with visual control of a considerable number of places at the protected facility from a central security post. The system operation efficiency depends on the number of cameras, their positions, as well as their image quality. In the past, video recording was always considered less important then live viewing. Today however, based on the video products supplied to market, there is a considerable switch towards digital video recording. This tendency still remains.

The main idea behind the concept of security with image processing consists in prediction and suppression of dangerous events, and not recording them for future analysis and diminishing the probability of finding something. It is within this framework that the development of digital technologies and image processing must proceed.

Digital technologies must contribute to the successful and efficient work of the security personnel, and not to review them after the tragic events occurred. The CCTV practice proves that the human factor should remain in charge of the security systems. In any CCTV system surveyed by cameras a considerable part of the security assessment work is still performed by the security operators. In this context, the tendency of digital technology implementation in CCTV cameras "to have a better view of what needs to be seen" is the most decisive.

References:

1. W. Pratt, "Digital Image Processing", 2nd edition, John Wiley & Sons, 1991.
2. N.Ahmed, K.R. Rao, “Orthogonal Transforms for Signal Processing”, Springer-Verlag Berlin – Heidelberg – New York, 1975.
3. “CCTV Focus” (Russian edition) #5, 2005, pp. 16-34.