

CHAPTER 1

Introduction

THE electromagnetic transmission and recording of image sequences requires a reduction of the multi-dimensional visual reality to the one-dimensional video signal. Scanning techniques for this purpose were already invented by F.C. Bakewell in 1848 and in a more practical form by P. Nipkov in 1884¹. These inventions had a great impact on our society, and *television*, i.e. the medium that allowed "viewing at a distance", became our main window on the world. The practical shape that resulted was that of a single source that *broadcasted* picture material to a large number of receivers. The costly equipment was concentrated at the transmitter side, that existed in small quantities, and the receivers were designed for low cost and high quantities, enabling the system to reach hundreds of million people.

1.1 Revolutions in a conservative world

An important consequence of this system architecture is the conservatism that handicaps technical progress even when it requires only modest adaptation of the very many receivers. This conservatism has evidently prevented rapid introduction of new scanning formats that enabled a higher spatial or temporal resolution. Also smaller modifications of the signal, e.g. the one required for the introduction of colour, usually were only acceptable when backward compatibility could be guaranteed. On the other hand this conservatism has stimulated the creativity of many researchers in the field of television to improve the perceived

¹An outline of these old methods and information on the history of television can be found in [39]

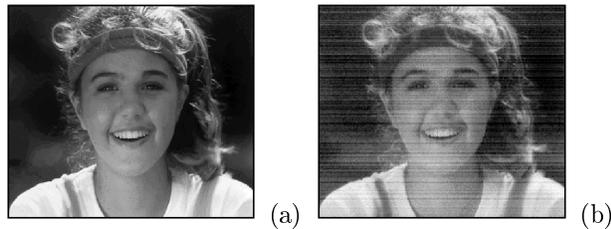


Figure 1.1: Screen photographs illustrating the specific form of noise sometimes encountered on television. The left hand image (a) shows the noise free original, and the right hand image (b) shows the same picture corrupted with clamp noise due to DC-restoration errors of some clamp circuit in the video chain.

image quality in a compatible way. Particularly the digital techniques that entered the television receiver around 1980, and parallel with it the option to store and delay image parts, pushed the use of *image enhancement* techniques.

The silicon technology enabled a complexity growth according to *Moore's Law* which helped the more robust, but less economical digital techniques to become the natural choice in areas where the earlier analogue solutions were more cost effective. More or less simultaneously, video, and by the way audio which, however, is out of our scope, entered the computer which became a consumer electronics product in the form of the Personal Computer, or PC.

By the end of the twentieth century, this synthesis led to the *multimedia* products in which the video typically is scalable in spatial as well as in the temporal domain. This caused an explosion of video formats, as in addition to the two main broadcast formats² PC monitors with picture rates between 60 Hz and 100Hz and spatial resolutions in a broad range (VGA, SVGA, XVGA, etc.), including high resolution, arrived on the market. Also television receivers profited from these techniques and decoupled their display format from the historically determined transmission format to eliminate artifacts as large area and line flicker, which resulted in new 100Hz (*flicker-free*), *non-interlaced*, and *widescreen* formats.

Apart from the broadcast video that that entered the multimedia PC and TV, also *videotelephony* and video from the *Internet* had to be merged with image signal generated internally in the PC, the *graphics*. Evidently, the conversion from one video format into the other, *Video Format Conversion*, became a key technology in a multimedia system that, therefore, in addition to the image enhancement techniques forms a very significant part of this book.

A last development, the launch of video broadcast satellites required a total digital approach for the channel format as well. The available transmission capacity could be used so much more efficiently, i.e. more economically, using digital bit rate reduction techniques, or *video coding*, that it prohibited the

²The 50 and 60 Hz television formats with 625 and 525 scanning lines, respectively, both interlaced.

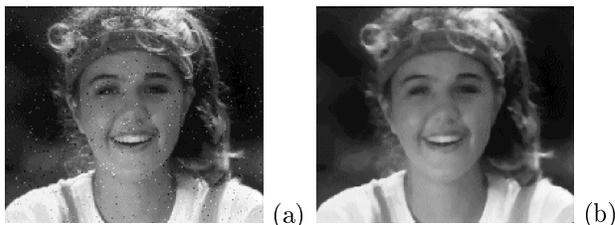


Figure 1.2: *Sometimes the advantage of an image processing algorithm depends on the taste of the viewer. Clearly, the speckle noise (a) has been removed by the processing, but a loss of resolution can be seen as a clear drawback of the method (b).*

use of the older analogue formats. Similar techniques became also dominant for storage, e.g. the *Digital Versatile Disk* (DVD), and eventually will replace the analogue formats for terrestrial video broadcast resulting in *Digital Video Broadcast* (DVB). The second part of this book shall, therefore, be dedicated to the video coding techniques that played a decisive role in this process.

1.2 Image enhancement: A matter of taste

The purpose of image enhancement is to improve the subjective picture quality, and *not* to estimate the true image from the observed one, which we shall call image restoration. An implication of this aim is that the success of image enhancement processing may be a matter of taste.

Sometimes there is not too much discussion about the improvement; An example of such type of processing is clamp error filtering shown in Figure 1.1.

In another case the advantage is clearly noticeable, but the drawback is visible as well, as illustrated for speckle noise removal, that often causes some blurring as a side effect, in Figure 1.2.

In some cases the advantage is rather specific for television images. For example the brightness of the television screen is a clear selling point, since many expect good and clear pictures even when viewed in broad daylight. Given this background, a trick as increasing the blue level slightly in near top white image parts, as in Figure 1.3 subjectively improves the brightness of the picture, although it is doubtful whether indeed the vision in daylight improves, while some may judge the white too cool.

More specifically, image enhancement is about removing, or reducing noise (*noise reduction*, or *image smoothing*), de-blurring edges (*peaking*, or *edge enhancement*), re-mapping the luminance values in order to obtain an improved contrast (*gray-level re-scaling*, or *histogram modification*), and modifying the colour hue (*skin tone correction*, *blue stretch*), or colour saturation (e.g. *green enhancement*). Many of these algorithms for video are quite similar to what can be found in publications on image processing. Specific differences occur in the

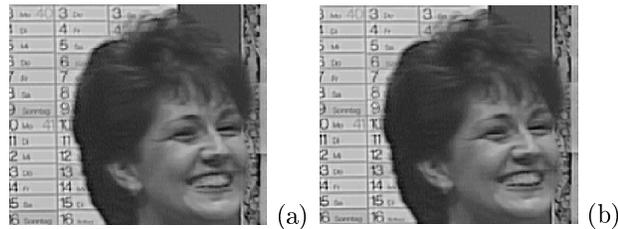


Figure 1.3: *With "blue-stretch" (b) the picture appears brighter, which is seen as a quality indicator for television sets. Whether it is a real improvement, is definitely a matter of taste.*

noise filtering methods, as in video the additional temporal dimension allows for different filtering techniques. At the same time, interlace complicates the use of this dimension, while the common DC-level restoration, or clamping, in video chains introduces a typical form of noise, shown in Figure 1.1, not dealt with in image processing literature. Because of these peculiarities, the focus in the part on image enhancement is on noise and noise reduction techniques.

It may come as a surprise that the processing of individual image enhancement algorithms sometimes have an opposite effect. A good example is found in the combination of noise reduction and peaking, where peaking generally increases the noise in the image and noise reduction may introduce some blurring of picture parts. In such cases, it is necessary to analyse the video data to conclude which processing algorithm is to be preferred. In our given example of noise reduction and peaking this could mean that we design a noise estimation circuit to decide upon the relative subjective effect of the noise and peaking algorithms. Such analysis tools shall also be discussed in the part on image enhancement.

The applications of image enhancement are typically found in medical image (sequence) processing, in remote imaging, and in consumer video, i.e. television, PCs, and video recorders. In the Chapters 2 and 3, which are dedicated to image enhancement and noise reduction, respectively, we shall focus explicitly on consumer vision applications of image enhancement, mainly television, and from the above categories in the first place on noise reduction.

1.3 Consumer video format conversion

The combination of new displays, new video broadcasting formats and the advent of multi-media equipment has rapidly increased the need for high quality consumer video format conversion. While the availability of digital signal processing has made the problem of spatial conversions almost trivial, the other elements of format conversion, i.e. de-interlacing and picture rate conversion have long been regarded highly complex. Therefore, picture rate converters simply repeated pictures until the next one arrived, although this results in artifacts



Figure 1.4: Photographs of a screen detail showing a comparison of up and down scaling using pixel dropping and repetition, (b), and polyphase-filtering, (c), respectively. The left hand picture, (a) shows the original. Although the result from polyphase filtering is clearly better than that of pixel dropping and repetition, it's resolution is less than that of the original image because of the intermediate scaling to a smaller size.

when motion occurs. Similarly, de-interlacing resulted from simple repetition, or averaging of neighbouring lines. In the somewhat more advanced de-interlacing concepts vertical-temporal processing was applied [40, 151, 12, 125], but even these degrade the image parts where motion occurs.

To appreciate the state of the art, as described in the Chapters 4, 5, and 6, we shall briefly introduce the capabilities, required for high quality video display format conversion, using mainly illustrations to elucidate the underlying concepts.

1.3.1 Spatial scaling

Spatial scaling is necessary whenever the number of pixels on a video line, or the number of lines in an input image, differs from the numbers required for the display. High quality spatial scaling of a time-discrete representation of a continuous signal results as a straightforward application of a long available theory [33], and perfection is achievable at a consumer price level.

An example of high quality scaling is found in wide screen television sets, where it is used for aspect ratio conversion of the image. The only prerequisite for application of this concept is that the continuous signal is sampled according to the demands of the sampling theorem, i.e. that its spectrum is bandwidth limited to half the sampling frequency. In the horizontal dimension of video signals this is usually the case, in the vertical and temporal domains usually not. The standard recipe for integer scaling then is to use decimating and interpolating low-pass filters for downscaling and up scaling respectively, as illustrated in Figure 1.5. Non-integer scaling results as a (virtual) cascade of up-scaling and down-scaling. Memories are used for buffering, as samples are written at a first (input) sampling frequency and read at a second (output) sampling frequency. A decimating low-pass filter reduces the bandwidth of the input signal to less than half of the output sampling frequency and throws away the redundant samples. The integer up-sampling low-pass filter acts on the (higher) output sampling

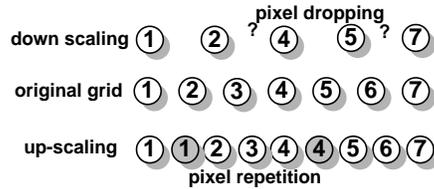


Figure 1.5: *Poly-phase filtering is a more sophisticated approach to sample rate conversion where output pixels result as a weighted average of input pixels. The up-sampling filter has the characteristic that it passes the baseband signal and suppresses the repeat spectrum resulting from the input sampling where possible. The down sampling filter has to suppress part of the baseband signal to prevent aliasing due to the lower output sampling frequency.*

frequency by adding zero valued samples to the input samples (zero-stuffing) and removing the first repeat(s) from the input spectrum.

Although this is a well established concept, it is useful to return to one of the outdated and inferior methods that have been used in the past, since this method is still quite popular for format conversions in the vertical and temporal domains. This earlier, or poor-mans, solution to spatial scaling consists in pixel dropping, or pixel repetition, illustrated in Figure 1.6. Though simple, which is its main advantage, this technique leads to clearly visible deterioration of the image, which made this technique for spatial scaling obsolete. Figure 1.4 shows screen photographs, illustrating the advantage of using polyphase filters for sample rate conversion compared with the simple pixel repetition and pixel dropping technique.

We repeat that the above-described procedure for high quality sample rate conversion, i.e. using polyphase filters, is only valid in case the demands of the sampling theorem are met. For vertical scaling of interlaced video material, this is not the case. A prior de-interlacing stage is then required, which is the topic of the next section.

1.3.2 De-interlacing

Interlace is the common video broadcast procedure to transmit alternately only the odd, or only the even numbered lines of a picture, respectively. De-interlacing attempts to restore the full vertical resolution, i.e. make odd and even lines available simultaneously, for every picture. This makes de-interlacing a basic requirement for video scanning format conversions, often necessary prior to conversions in the spatial and temporal domains.

As with spatial up scaling, the simplest method exists in pixel repetition. If the repeated pixel is a vertical neighbour we speak of line-repetition, if it is a temporal neighbour we speak of field-repetition. Field repetition preserves all detail in a stationary image part, but introduces severe artifacts in moving parts,

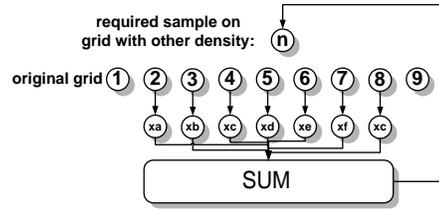


Figure 1.6: *Pixel repetition and pixel dropping are simple and straightforward techniques to adapt the density of sampling grids, using original input samples at the output only.*

as can be seen in Figure 1.7³. Line repetition, on the other hand, cannot eliminate the alias present in a single field and leads to jagged edges (also shown in Figure 1.7). Motion, however, does not further degrade the picture. All kinds of adaptive methods switching between field and line repetition have therefore been proposed to realize a reasonable compromise. Often line repetition is replaced by some linear interpolation on field basis.

Recognizing that de-interlacing is actually a 2-dimensional up-sampling, in the vertical-temporal domain, also 2-D, vertical-temporal, linear and non-linear filters have been proposed. More recently it has been concluded that only motion compensated methods (shown in the right hand image of Figure 1.7) can give satisfactory de-interlacing results. Although motion compensated methods are clearly best, they have for long been judged too expensive for consumer applications, mainly due to the high price of the motion vector estimator. Here, more recently breakthroughs have been reported [52, 55], which made it feasible that the motion compensated methods became available in consumer television scan conversion ICs [48, 49]. An overview of de-interlacing methods shall be given in Chapter 4.

1.3.3 Picture rate conversion

Picture rate conversion is the third and last aspect of video format conversion. Modern displays need to be able to display high-quality video from a range of sources. Common video cameras use picture rates of 50 or 60 Hz. Movie films are recorded at 24, 25 or 30 Hz, while the picture rate of TV and PC displays lies generally between 50 and 120 Hz. As before with spatial scaling and de-interlacing the poor-mans solution is pixel repetition, and again it results in degradation of the image quality. In the temporal domain, pixel repetition becomes picture repetition and it introduces artifacts which vary from "motion judder" (if the difference between input and output picture rate is below ap-

³An exception occurs when film material is broadcast. Film is progressively scanned, and although the odd and even lines of a film picture are transmitted in separate fields, they originate from the same film image and can be assembled to the original progressive picture without any disadvantage.

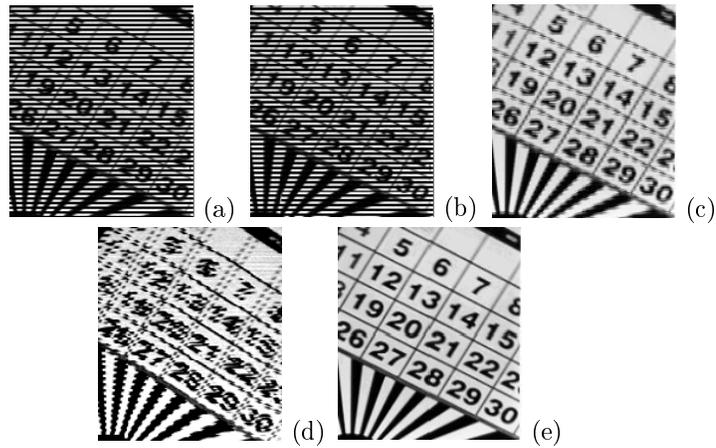


Figure 1.7: *The options in de-interlacing a video signal. In case of motion assembling the lines from the odd (a) and the even (b) field leads to strong artifacts as shown in (d). When interpolating the missing lines from one, e.g. the odd, field only, alias remains, clearly visible in (c). Motion adaptive processing can not prevent this alias in moving picture parts. Only if the motion between the fields is precisely compensated for, assembling leads to a perfect de-interlacing as shown in (e).*

proximately 30 Hz) to "motion blur" (for the higher difference frequencies). We shall clarify the effects using two examples.

The first example deals with a 24 Hz pictures per second film displayed on a 60 Hz display. As a result of picture repetition, some pictures will be shown two times and others three times (2-3 pull down). Figure 1.8b illustrates the situation with a moving ball registered by a 24 Hz film camera, and displayed on a 60 Hz display. As can be seen the input pictures are alternately repeated two or three times. In this case the viewer will observe an irregular, or jerky, motion, often indicated as "motion judder".

Figure 1.8c shows the second example; 50 Hz video input displayed on a 100 Hz display using picture repetition. In this case motion seems to be smooth, since the difference frequency is above 30 Hz, but instead the viewer tracks the ball on both possible positions simultaneously. Figure 1.9 shows the perceived image resulting from this second example. Indeed motion seems to be smooth, which obviously cannot be concluded from this stationary illustration, but instead the viewer will observe a double or blurred object ("motion blur"), as we shall clarify.

At the retina of the eye of an object-tracking viewer, the object is stationary; i.e. remains at the same position as long as the tracking is perfect. If the 'updating' of the image at the correct position occurs with a frequency that is high compared to the time constant of the human visual system, roughly above 30Hz, than the object is perceived continuously. This implies that at

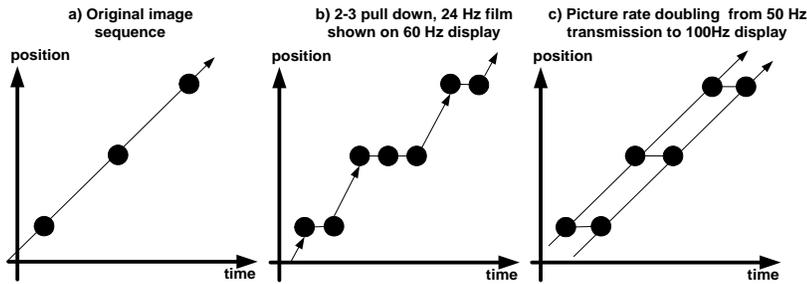


Figure 1.8: *The effect of picture repetition on the motion portrayal. In (a), the original picture sequence is shown. In (b) it is illustrated how film, with 24 pictures per second, is commonly displayed in the so-called 2-3 pull-down mode on a 60 Hz display device. Finally, in (c) the effect of picture repetition in case the difference between input and output frequency exceeds 30 Hz is illustrated.*

intermediate instances, the object is "seen" at its motion trajectory. If now, at the same time, the display shows the object at a different, repeated, position, the viewer incorrectly concludes that there must be two objects moving in parallel, as illustrated in Figure 1.8c.

The moving ball is perceived at both the "expected" and at the repeated position. The result is a double, or blurred, image. All this can be prevented using motion estimation and compensation techniques. The status of this technology shall be the topic of Chapters 5 and 6.

1.4 Goal, scope and outline of this book

High quality compatible picture improvement algorithms that can be economically implemented, applicable in future-generation consumer video electronics, are the topic of part one. This includes scan rate conversion, de-interlacing, feature enhancement, gray level rescaling, and noise reduction filters. It shall appear that often motion complicates the improvement algorithm. Therefore, motion compensation in many of these algorithms yields significant advantage. The enabling technology, here, is motion estimation from video data, which is dealt with separately.

After the general introduction in Chapter 1, Chapter 2 introduces the simpler image enhancement techniques designed to improve sharpness, contrast, or colour reproduction. This chapter also contains references to provide insight which products rely on what method. In Chapter 3, we shall discuss noise and noise reduction methods where, as an illustration of the principles, specific products are detailed in practical examples. The algorithms discussed in Chapters 2 and 3 are *not* related to scanning format conversion in contrast with the remaining chapters in part one.



Figure 1.9: *Picture repetition leads to motion blur if the difference in picture rate of input and output exceeds the 30 Hz.*

Chapter 4 provides an overview of de-interlacing techniques, partly available in products on the market, and for another part available in scientific literature. The complete spectrum of methods is discussed in this chapter, i.e starting from the simple linear methods up to the most advanced category of de-interlacing methods that relies on accurate knowledge of the movements of objects in the scene. The chapter also provides a comparative evaluation of a broad range of methods

Chapter 5 introduces an other dimension of scanning format conversion, picture rate conversion. The chapter too is focussed on algorithms feasible in consumer television sets. This focus, however, does not exclude advanced methods applying motion vectors, which are the only methods leading to high quality picture rate conversion.

Finally, Chapter 6 introduces several motion estimation methods, including pixel-recursive, block-matching, and object based algorithms, as well as post-processing methods proposed to improve the quality of the motion vectors generated with any of these methods. This chapter is highly relevant for video format conversion, as the better methods for de-interlacing video, described in Chapter 4, and the better methods for picture rate conversion of Chapter 5, all rely heavily on the availability of high quality motion estimation. The focus of this chapter shall therefore be on motion vector estimators applicable for these very demanding applications in scan rate conversion.