Engineering & Ergonomics

Postbus 267
7500 AG Enschede
tel: 053 - 428 0500

# Human Factors of CCTV

Reference: S386-R1.CCT
Date: September 2012

drs. K.B.J. Schreibers Eur. Erg.
ir. R.B. Landman
ir. R.N. Pikaar Eur. Erg.

## Part 1
## Technology and Literature review

## Content

**Project partners**

Human Factors of CCTV in practice is a pooled funded research. Project partners are:

- H.I.T.T. Traffic
- IHC Dredgers / IHC Beaver Dredgers
- Nedap Security Management
- NS Concernveiligheid, dep. Security
- ProRail ICT Services
- Royal Haskoning
- RWS - Dienst Verkeer en Scheepsvaart
- Total E&P Netherlands
- Vopak Management Netherlands B.V.
- Waterschap Hollandse Delta
- DG-Organisatie Bedrijfsvoering Rijk
- Provincie Noord-Holland
- Rijksbelastingdienst.

**Research partners**

- ErgoS Engineering & Ergonomics
- Informatica Communicatie Academie - Hogeschool Arnhem Nijmegen
- Intergo
- Vhp-ergonomie.

## Management Summary

The aim of the research project *Human Factors of CCTV in practice* is to deliver practical Human Factors guidelines for the design of IP-based CCTV systems. The project consists of 5 phases:

1. Orientation and definition of the research, including a literature review.
2. Situation analyses, to gain insight in relevant CCTV tasks.
3. Field experiments and/or laboratory experiments, regarding workplace variables (such as image quality, number of images to be observed).
4. Development and testing of preliminary Guidelines.
5. Final Guidelines.

This report represents the deliverable of phase 1, i.e. an overview of CCTV-system elements and the results of a literature search.

**Technology**
Literature and Human Factors data on CCTV systems, i.e. camera, lens, sensor chip, coding, transmission, decoding and displays is entirely based on analogue systems. There is little state-of-the-art HF research available related to IP-network technology and digital systems in CCTV control centres.
There is a difference in topics discussed in literature between surveillance systems and other applications. For surveillance systems, research focusses on image reduction in view of a limited transmission capacity relative to a (very) large number of camera views one would like to transmit. For all other applications, imaged reduction or quality degradation because of technology is not an issue.

**Human Factors of CCTV - general**
Approximately 40 recent publications on Human Factors and CCTV have been found. In the UK and the USA, several research projects have been published. Nevertheless, the HF literature is limited, when privacy aspects, legal matters and public surveillance (how does the public respond to this) are excluded from a search. There are no references to process control tasks supported by CCTV, such as lock and bridge control, loading & unloading supervision, and process unit control.

Generally speaking, the level of available design guidance is disappointing. There are listings of "things you should think of" in designing CCTV control centres, however no design requirements. An exception concerns the well established guidelines available from other areas, such as control centre workplace design (measurements), text labelling, controls, and principles for a logical mapping of cameras on displays.

**Human Factors of CCTV - detailed guidelines**

There is some evidence for the following guidelines:

- The number of images (live streams) an operator can handle depends on the task complexity. The majority of literature mentions a maximum of 16. Some research mentions that the probability of detecting an incident is substantially reduced when there are more than 100 cameras per operator.
- Useful field experiments have been carried out in the UK, concerning traffic control (checking hard shoulders for objects), rail safety (level crossing), and incident detection in shopping areas. However, there is (yet) too little evidence to be able to generalise the outcomes.
- Image quality and quality reduction due to coding, transmission, decoding and displaying has not been researched so far.

**Tasks and workload**

- Operator tasks vary from proactive monitoring to reactive screening based on alarm messages. For surveillance purposes the following topology of tasks is frequently mentioned in literature: detect, monitor or observe, recognise, and identify.
  Unfortunately, the topology is based on the perception of persons (on a display). It is expected that this particular topology will not hold for functions in a (process) control setting. In addition, clear definitions for each of these tasks are missing.
- Regarding workload, there is some evidence that the perceived image complexity correlates to task complexity (or workload). It has been suggested to develop a metric for image complexity (a task driven metric). In the researchers view, there is also the additional need to develop a metric for image quality (technology driven metric or measurement).
- The limited evidence indicates that CCTV surveillance tasks have a high workload, i.e. frequent breaks are recommended, however in line with general visual display terminal guidelines (VDT-guidelines).
- An interesting topic mentioned in literature is the effect of becoming overly dependent on the CCTV displays, while the same task(s) may be performed by other data (displayed in a suitable GUI). An example given concerns traffic detection loops and related software.

# 1 Project Human Factors of CCTV in practice

## 1.1 Project objectives

CCTV systems are used for traffic supervision, tunnel safety, lock and bridge control, surveillance, remote production control, etc. Centralised CCTV image observation and supervision is carried out by operators in control centres. CCTV systems consist of cameras, monitors, recorders, interconnecting IP network hardware and support infrastructure. New installations are usually digital. The quality of image presentation of CCTV-systems is determined by the weakest link of all elements of a CCTV chain. The chain starts at the scenery and ends after information processing by the human operator.

At face value, many CCTV-images seem to lack quality. Related issues are:
- camera position
  - appropriate vertical or horizontal viewing angle (avoid distorted images)
  - camera to object distance (related to relevant object details).
- visual noise may occur as a result of
  - sub optimal lighting situations in the object’s environment
  - insufficient camera resolution
  - limited transmission capacity.
- operator workload and optimal performance will be related to
  - number of cameras and images to be monitored
  - workstation layout.

A pilot study by Landman and Pikaar [1] in 2010 revealed that there are many questions, but few answers. Therefore, a research project *Human Factors of CCTV in practice* was specified, aiming to deliver practical Human Factors guidelines for the design of CCTV systems. The research project consists of 5 phases:
1. Orientation and definition of the research, including a literature review.
2. Situation analyses, to gain insight in relevant CCTV tasks.
3. Field experiments and/or laboratory experiments, regarding workplace variables (such as image quality, number of images to be observed).
4. Development and testing of preliminary Guidelines.
5. Final Guidelines.

The research format is based on the active participation of 13 project partners with an interest in CCTV-system design. The project is carried out by four research partners with an extensive Human Factors (HF)/ergonomics knowledge on control centres.
During the project, four workshops will be organised for all participants. Topics to be discussed during the workshops will be respectively, the problem definition phase including literature overview, results of situation analyses, preliminary guidelines, and the final guidelines.

This report presents the results of a literature search on human factors of CCTV-systems, including an overview of current technology (Chapter 2). In Chapters 3 the literature regarding Human Factors (HF) aspects of CCTV-systems will be reviewed. Chapter 4 elaborates literature on CCTV operator tasks. Refer for definitions, abbreviations and acronyms to Appendix A.

## 1.2 Scope

Phase 1, orientation and literature review, covers HF aspects of camera selection, transmission, and presentation of CCTV-images, in order to list actual knowledge and to identify gaps to be researched. The project scope will be limited to digital (and IP-based) CCTV-systems for monitoring traffic, facilities and processes, and surveillance.

The project does not include:
- analogue technology of CCTV system components, because they are expected to be replaced by digital systems within a few years.
- intelligent software, such as advanced digital image enhancement, automated video content analytics (VCA), decision support, and post-event analysis of CCTV imagery.
- legal and privacy related topics of CCTV use.

## 1.3 Literature review - method

Researchers sought out guidelines, standards and relevant literature using internet (Google Scholar, Association for Computing Machinery) and snowball method. The review was carried out from April to July 2012. Keywords and acronyms in English and Dutch, used for internet search are:
- cctv
- ergonomics and CCTV
- human factors and CCTV
- guidelines CCTV
- human factors "traffic management centre"
- guidelines camera supervision
- guidelines supervision tasks
- guidelines video surveillance.

The draft of this report has been presented during the first workshop (June 2012). Feed back by project partners has been included in this final version.

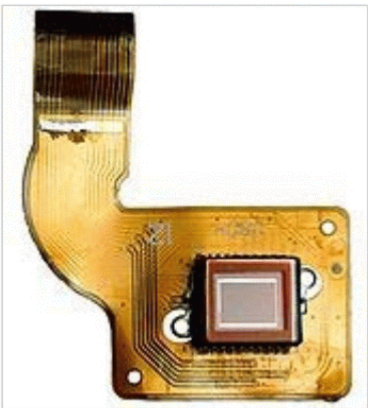
Literature will be indicated in the main text by a number in square brackets "[xx]" and listed in the Reference section in order of first appearance.

# 2 Remote Supervision - CCTV systems

This Chapter summarizes the technology of CCTV systems, as far as its components are expected to influence image quality. The information presented here comes from many sources including manufacturers data sheets, and has been verified in the English Wikipedia, the free encyclopedia, and CCTV - Networking and Digital Technology (Damjanovski, [2]).

## 2.1 Scene and camera lens

In order to describe CCTV-camera systems an analogy to the traditional 35 mm cameras will be used. The 35 mm refers to the width of a traditional film with a 24 x 36 mm picture format. A standard 35 mm camera has a 50 mm lens (=focal length). The image projection on the film compares well with the visual properties of the eye. I.e., the image represents the area we can see in one glance and without apparent distortions. A lens with a short focal length (for example 25 mm) is called a wide-angle lens. The picture has a wider angle (shows more of the scenery), however some distortions may occur near the edges of the picture, depending on the quality of the lens.



At the heart of a digital camera is a CCD or a CMOS image sensor.

**Digital camera**
Instead of a film, the digital camera has a CCD or CMOS image sensor chip. The chip has light sensitive areas (pixels). Sensor chips are smaller than the traditional 24 x 36 mm film. The factor by which the sensor is smaller is indicated by the "Crop-factor".
A Crop-factor 2 means that the sensor chip is a factor 2 smaller than the reference surface of 24 x 36 mm (in both directions), in this case 12 x 18 mm (see figure 1). As a consequence, the image on the sensor compares to the image of a 100 mm (traditional) lens.

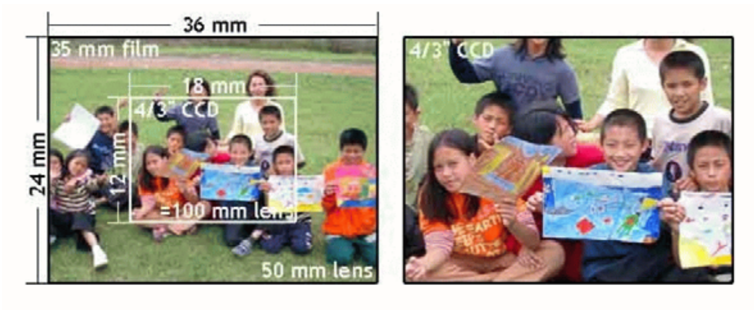



Figure 1. Illustration of crop factor 2 picture.

The other way around: taking a 25 mm lens and a sensor of 12 x 18 mm, yields the same result as a traditional 35 mm camera equipped with a 50 mm lens. However, because of the wide angle (25 mm lens), some distortion of the scenery will occur.

**Image on the sensor chip**

- The image quality on the sensor chip depends on:
  - lens quality (mechanical)
  - lens aperture (f-number, denoted by f/..)
  - sensitivity of the sensor-chip;
  - exposure time per frame
  - the number and density of pixels on the sensor-chip;
  - Depth of Field (DoF).

- The lens aperture determines the amount of light on the sensor. The greater the f-number, the darker the image projected by the lens. The amount of light transmitted from each object in the lens's field of view to each unit of area of sensor decreases with the square of the f-number. Doubling the f-number decreases the exposure by a factor of four. To maintain the same exposure when doubling the f-number, the exposure time would need to be four times as long.

- A digital camera with a small sensor-chip may still produce a high resolution image. A common 1/1.8" lens has a crop-factor 5 and a sensor size of 5.32 x 7.18 mm. The image projected by a 5 mm lens resembles the traditional 25 mm lens (of a 35 mm camera). However, a 5 mm lens is an extreme wide-angle lens, resulting in large distortions of the image. Advanced software corrects the distortion, however at the cost of resolution or sharpness of the image. An example of a CCTV camera specification:
  *Surveillance CCTV camera: 1/3" CMOS Sensor (1.3 Mpixels)*
  *Lens focal length: 2.8 - 10 mm; F/1.4*
  *Visual angle: 81° (at 2.8 mm) - 27° (at 10 mm).*
  *Minimum illumination: 0.003 lux at F/1.4 (at 50% videosignal)*

- Field of View (FoV)
  The area of a scene, observed by a camera and lens combination and measured both horizontally and vertically, that can be seen through the camera. An on-line calculation sheet is provided by www.jvsg.com/online/#. (see example below).

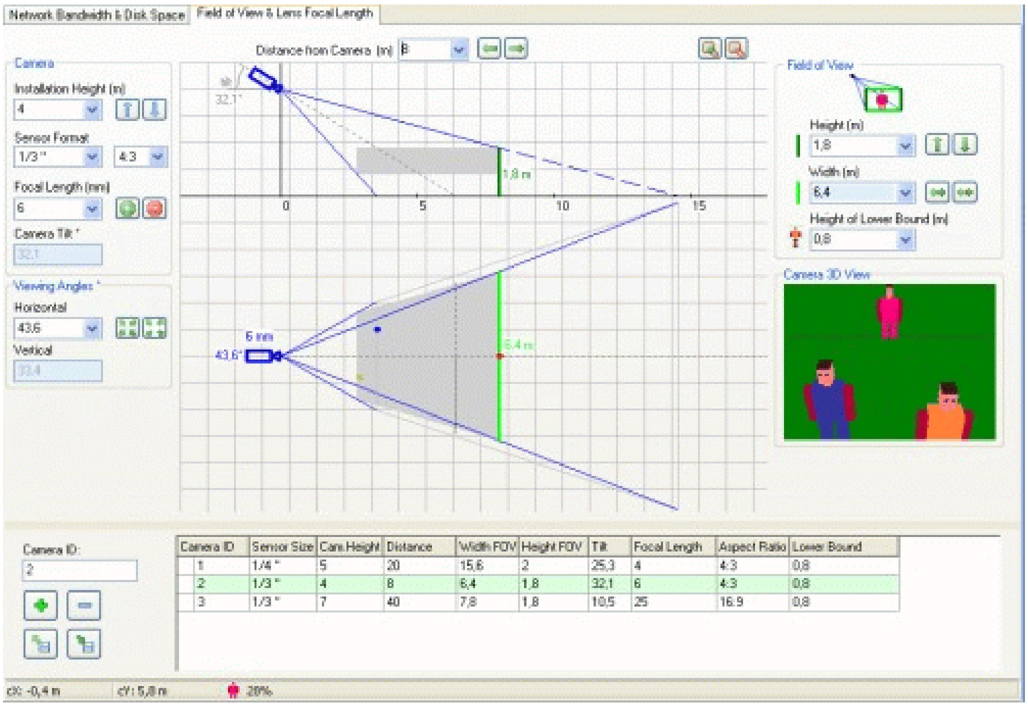



Figure 2. Example of a field of view calculation tool.

- Depth of Field (DoF)
  DoF is defined as the range of distance within the image that is acceptably sharp. When you focus precisely on a subject, a certain distance in front of and behind the subject also will be in focus, although not necessarily as sharp (illustrated in figure 3).
  - DoF increases or decreases based on the focal length of the lens, the distance to the subject, and the aperture.
  - In light (sunny) environment, a high f-number is possible, resulting in a large DoF. In a dark environment, a low f-number is necessary to get sufficient light on the sensor chip (leading to a small DoF).
  - Exposure time (shutter time) also determines the illumination of the sensor chip. Longer exposure time = more light = high f-number = large DoF.

  The internet offers several DOF-calculators (search for: depth of field calculator, or cctv dof).

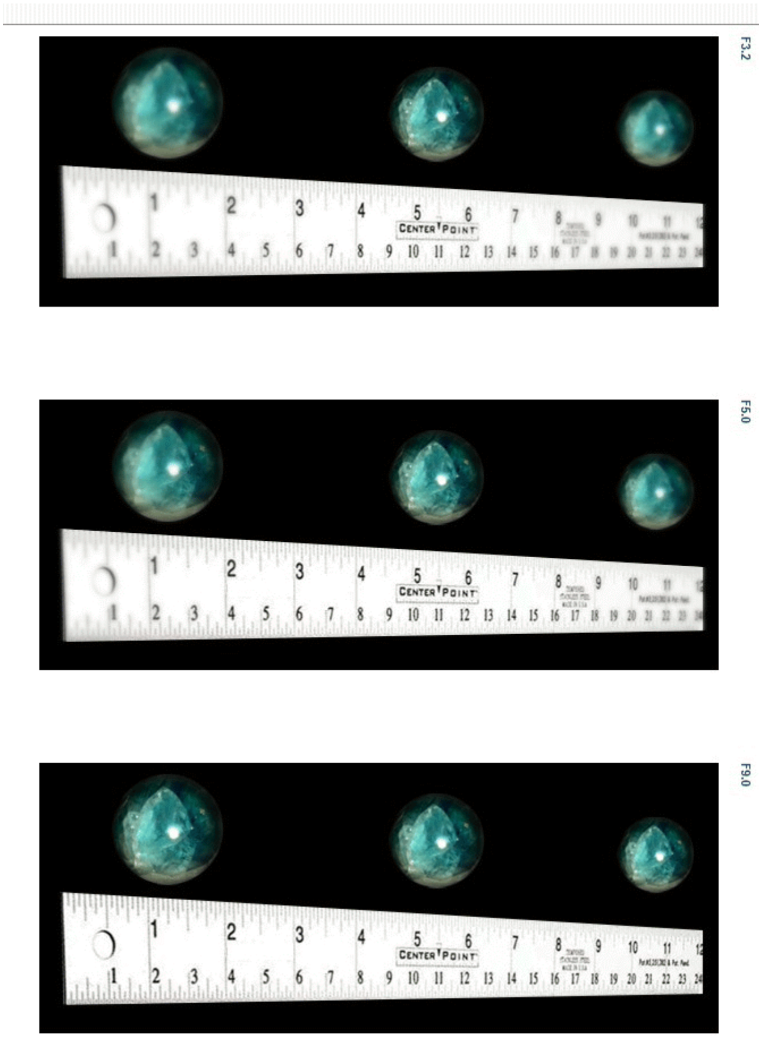



Figure 3. Illustration of DoF: top represents a small DoF, bottom a large DoF.

## 2.2 Camera and transmission

**Background - analogue television**

Though analogue television technology will be disappearing, its terminology is widely in use for CCTV-systems. In Europe, the public analogue broadcast system is based on PAL (Phase Alternate Line). Pictures are composed of 625 lines at a refresh or frame rate of 25 Hz. Out of the 625 lines, 576 lines have an image content. Each line is divided into maximal 756 points.
CIF (=Common Intermediate Format) defines an image of 288 x 352 pixels (h x w). SIF (=Source Input Format) is based on the USA 60 Hz NTSC system definition and equals an image of 240 x 352 pixels. Related definitions are:

- QCIF 144 x 176
- CIF 288 x 352
- 4CIF 576 x 704
- Full D1 576 x 720.

**Digital systems**
In digital systems, any image size may be chosen. Entertainment industry is leading in setting standards. Commonly used definitions are in a 3:4 image format: 480 x 640; 600 x 800; 768 x 1024 or 1200 x 1600.
For HD television definitions are based on a 9:16 or 10:16 image format: 1080 x 1920 pixels (2.1 Mega pixel; if indicated by 1080p, it means progressive scan/scanned line by line); 720 x 1280; or 720 x 1368 pixels.

- A camera sensor chip has a fixed number of pixels. Typically for CCTV is a 1/3" sensor (3.6 x 4.8 mm) at a resolution of 480 x 640 or 576 x 720. Recent surveillance camera specifications mention 720p or 1080p CMOS sensors. Photo cameras may have larger sensor chips including more or larger pixels (for example 9.6 x 12.8 mm; 3600 x 2700 pixels).

- The sensor-chip produces a data stream. Each pixel is specified by 24 bits of data. For example a 480 x 640 image requires 7.4 Megabits of data (307.200 pixels x 24 bits). Transmission of the data in this format at a 25Hz frequency would lead to a huge datastream. Data reduction is needed, which can be achieved by data encoding. After transmission, the encoded data will be decoded into the original format.

- Encoding is standardised by (amongst others) MPEG-2 for digital television broadcast and MPEG-4 for consumer electronics. Basically the pixel data is transformed into the frequency domain. In addition, there are several methods to encode consecutive frames. These methods are based on transmitting differences between individual frames, instead of full frames, every now and then alternated by a full frame from for referencing. All in all MPEG-2 yields a factor 30-100 data reduction.
  MPEG-4 (Part 10: H.264-codec standard) has a factor 2 higher reduction, however at a cost of image quality. I.e. small details in the image may disappear (this has been studied by Keval [3]).

- In case of large CCTV-systems and a limited transmission capacity, data reduction by encoding may not be sufficient to enable transmission of all data in time. This problem may be solved by lowering the frame rate or by compression of the original image (this may be compared to the "reduce format or size" command in image editing programmes /photo editors).
  Some examples of transmission bitrates are:
  - surveillance systems: lower limit of 32 Kbps (Kilo bits per second)
  - typical bit rate for a CCTV system: 1 Mbit/second (at a frame rate of 25 Hz, each frame can be 40 Kbit);
  - DVD-quality: 3 - 10 Mbits/second;
  - studio quality (720 x 576 pixels): 10 Mbits/s.

  At DVD and studio quality coding and transmission there will be no visible loss in image quality. Usually, cabling will not be the limiting factor (glass fibre cabling handles up to 10 Gigabit/second). The limiting factor is the encoding and decoding processing.

## 2.3 Display

At the receiver end of the CCTV system, images are decoded and the original picture can be made visible on a visual display unit. The imaging area can be defined by a pixel space and will be independent of the actual size of the display(s).

- A video processor fills the pixel space. The video card can be looked upon as a memory, mapping all pixels. Unlike progressive scan (analogue) systems, every time the content of a pixel changes, this information is immediately mapped onto the display. Images do not have a fixed refresh rate, although a fixed refresh rate might be used to refresh the video (memory) card.
  - Some manufacturers mention an image synchronisation problem in case 50 Hz frames have to be displayed on 60 Hz monitors ("Genlock"). In view of the previous remark it is not fully understood what the problem could be for digital systems.

- Projecting a 480 x 640 pixel image on a display can be done in several ways:
  - 1:1 mapping of each image pixel on one display pixel; there will be no loss of information; all available details will be shown.
  - by resizing the image, data might get lost:
    - this will be obvious, for reducing the size (less pixels)
    - by enlarging the image by a factor 2 each pixel of a frame will be mapped on 2 x 2 pixels of the display; all details will be shown, though in larger blocks; whether this is also experienced as "large block" entirely depends on the viewing distance.
    - for all other resizing ratios the resulting image quality degrades, however depending on the type of software transformation applied. Resizing principles are for example known as bicubic, bilinear or nearest neighbour.
  - The difference between a 1:1 or 1:2 mapping and other resizing ratios is illustrated in figure 4 (right: 1:1 display, left: other resizing ratios). What will be the best image may depend on its application, i.e. the left one is not necessarily the worst.

Figure 4. Different image sharpness.

# 3 Human Factors literature on CCTV systems

## 3.1 General

Human Factors literature on CCTV systems is limited. Control centre related standards [4] and guidelines [5] hardly mention CCTV as a separate topic. The EN-standard 50132 *Alarm systems - CCTV surveillance systems for use in security applications* [6] and AS 4806 *Closed- Circuit Television (CCTV)* [7] are exceptions. However, these standards do not address HF explicitly.

At the triennial International Ergonomics Association congress 2006 *Meeting Diversity in Ergonomics*, CCTV has been recognised as an important topic. This resulted in papers by John Wood [8] and colleagues, based on their experiences in rail and traffic control applications in the UK. and including some practical guidelines. In addition, there are two research proposals on HF of CCTV systems:

- Rail research UK & University of Nottingham: Optimum use of CCTV on Railway Systems (webpage: portal.railresearch.org.uk/RRUK/Site Pages/ProjectB9.aspx; should be starting in 2012).
- US Department of transportation - Federal Highway Association: Human Factors Guidelines for traffic management centres (results expected by mid 2013).

The status of these research proposals is unknown. The research groups have been approached with informative questions, however they did not respond.

In the USA, the American Public Transportation Association APTA [9] and the US Department of transportation - Federal Highway Administration FHWA [10] published preliminary HF CCTV guidelines. Other guidelines were written by Ahlstrom and Kudrick [12], issued by the US Department of transportation (Federal Aviation Administration). Finally, reference could be made to MIL-STD-1472G (USA Department of Defence; [11]).

There are many publications on HF design guidelines for (process) control centres [4,5], i.e. not specifically on CCTV Control Centres. Topics covered, are operator jobs and workload, operator workstation design, process graphics and work environment.
It was noticed that some topics, not withstanding the minor relevance for CCTV systems, get much attention in HF CCTV literature. The reason for this is probably that these topics have been well researched in the process control area. For example: character size for good readability of text labels is investigated numerous times (in the CCTV context), although well known and generally accepted guidelines are available and applicable for CCTV.

## 3.2 Recording system

The recording system refers to the functional parts of cameras, i.e. lens, sensor-chip, encoder and transmission (Chapter 2). Literature presents listings of topics to consider when designing recording systems. An overview is given in the sections below.

### 3.2.1 Camera lens

- EN 50132-7 [6] advices to consider the field of view (FoV), aperture number and internal lens reflections and ghosting;
- APTA [9] suggests to equip cameras with automatic mechanisms to ensure proper exposure under varying lighting conditions (automatic gain circuitry, day/night sensor switching and lenses with automatic iris functions).
  - The Depth of Field (DoF) is determined by the dimensions of the sensor-chip and lens focal length. In a well-lit situation with a lens set to a higher f-stop, a greater area in front of and behind the subject will be in focus. As light levels drop and the lens adjusts to a lower f-stop, less of the area in front of and behind the subject will be in focus.
  - An auto-iris lens automatically increases the size of the aperture as the view becomes darker. The DoF will decrease proportionally, and some objects in the scene may get out of focus.
  - In areas where all objects in the scene are critical, cameras designed to operate in very low light or supplemental illumination may be required.

### 3.2.2 Sensor-chip

- Colour and resolution
  - Brunt et al. [13] found in a trial with 22 railway operators supervising level-crossings, that replacing black-and-white images by colour images did not lead to significant degradation in operator performance. The researchers expected that too much image information (i.e. colour) would affect information processing and thus performance. It did not.
  - APTA [9] prefers the use of colour cameras, with a minimum resolution of 704 x 576 pixels (4CIF) and properly white-balanced (no further evidence).
- Frame rate (fps - frames per second)
  - Camera outputs should be digitally recorded at an appropriate resolution and frame rate. The table below [9] summarises recommendations for frame rate and resolution, differentiated to type of area to be observed.
  - The lower limit for experiencing a smooth flow is 12-15 fps.
  - Where cameras are covering passenger areas or areas containing any form of emergency call button, a two-speed capability is suggested, i.e. 4 fps (PAL) in normal mode, and a minimum of 12 fps in emergency mode. If a two-speed capability is not available, the 12 fps rate should be used to guarantee full-motion video recording capability.
  - The use of non-infrared, high-dynamic-range cameras and those capable of operating in low-light conditions should be considered to help improve the image quality in high contrast lighting situations.

| Area to be observed | minimum fps | minimum resolution |
|---|---|---|
| 1. Low-traffic pedestrian areas and boundary/ perimeter fences | 4 in PAL (5 in NTSC) | 4 CIF/ D1 |
| 2. Trackside operations and platform areas | 12 (15) | 4 CIF/ D1 |
| 3. Access control | 4 (5) | 4 CIF/ D1 |
| 4. Ticket office desks and pay machines | 12 (12) | 4 CIF/ D1 |
| 5. Vehicle traffic areas, parking garages, or forward-facing cameras on trains, trams and buses | 25 (30) | 4 CIF/ D1 |
| 6. On-vehicle passenger areas | 4 (5) | 4 CIF |
| 7. Vehicle passenger areas when emergency call operated or in the area of doorways | 12 (15) | 4 CIF |

### 3.2.3 Codec

The central HF consideration regarding video transmission systems is, whether the system (Codec) affects image quality. Digital systems can be designed in such a way that there is no visible quality loss at the end of the chain (refer to Chapter 2). For surveillance systems with large numbers of camera's and limited capacities for encoding, decoding and transmission, the question cas been raised to what extend image degradation is acceptable (Keval, [3]). However, no HF guidance could be found.

### 3.2.4 Environment

Image quality also depends on lighting of the scene [9]. Poor lighting is the most common factor that degrades the quality of video images.

- The lighting of the camera capture area should be adequately balanced until targets, specific reference objects and overall scene are easily identifiable. Also possible effects of unwanted and variable lighting have to be taken into account during camera installation, like flashing emergency lights, and bad weather conditions.
- The dynamic range (ratio of the lightest highlights to darkest shadow portions of the scene) should be considered. This range may not exceed the capability of the camera to record the scene. In case of low dynamic range cameras (less than 75 dB), flares or silhouettes may appear in the video as a result (figure 5).
- In case additional lighting is required [6] lists the parameters of lighting facilities one should take into consideration.

Figure 5. Illustration of dynamic range.

## 3.3 Presentation of CCTV images

CCTV images are presented in visual display units. This section is organized in four topics:

1. Text, usually superimposed on the CCTV image for labelling purposes
2. Persons and objects within images
3. Hardware, including a discussion on the number of screens and the number of CCTV images to be used parallel.
4. Information structuring, including amongst others the arrangement of images at the workplace.

Measuring image quality will be discussed in a separate section (3.4).

### 3.3.1 Text

Guidelines for text presentation do not differ from generally accepted guidelines for computer screens. Summarizing:

- Character height/ size [4].
  Maximum viewing distance to a text should not be larger than 215 x character height (of a capital character), and vice versa minimum character size should be maximum viewing distance, divided by 215.
- Apparent width/height ratio should be no less than 0.6 (60%).
- According to FHWA [10] a channel to which the operator is to attend should be made as distinct as possible from competing channels or visual signals: larger, brighter, louder, or more centrally located. It remains unclear how channel discrimination ideally should be designed.
- Video images should be labelled with a unique and adequate amount of location information [10, 4]. Even in static display arrangements (one-to-one mapping of cameras to displays), task performance appears to be faster and more accurate if screens are labelled with location names, rather than numbers.

- Labels for traffic management should include references to the cardinal direction.
- Legends superimposed on pictures should be designed such that they do not obscure essential parts of the picture, should be at a consistent position on the screen, and presented in a consistent manner.

### 3.3.2 Objects

Several attempts have been made to specify minimum measurements for objects on a screen in order to enable the operator to do an adequate job. The UK Home Office [16] and the Australian Government [15] recommend the following operational objectives in relation to the surveillance of people, objects or vehicles:

| | |
|---|---|
| – Monitor | Monitor/observe the flow of traffic or movement of people generally (not individual figures) |
| – Detect | Detect the presence of a person without needing to recognise or identify them |
| – Recognise | Recognise somebody who is known to the user, or determine that somebody is not known; monitor or track an individual person, object or vehicle |
| – Identify | Capture enough detail to identify a person, object or vehicle beyond reasonable doubt. |

In a Human Factors approach, the operational objectives could be considered operator tasks (which could also be indicated as "functions"). Aldridge [14] provides a guideline for the percentage of target-to-screen height ratio for these 4 operational objectives. In [15] the target height is specified in terms of "lines of resolution" or the minimum number of pixels on the screen. Cohen [16] correctly notes that since the use of digital systems there is variability in capture, recording, and display resolutions. For example, this implicates that a ‘recognise’ requirement can no longer be simply equated to 50% of the screen height.

| | % screen height [14] | number of vertical pixels [15] |
|---|---|---|
| – monitor | 5 % | 20 pixels |
| – detect | 10 % | 40 pixels |
| – recognise | 50 % | 200 pixels |
| – identify | 120 % | 400 pixels |

Aldridge adds some examples (see table below). Cohen [16] refers to the same tasks, adding *observe* in which a person should occupy 25 - 30% of the screen height. Some characteristic details of the individual, such as distinctive clothing, can be seen, whilst the view remains sufficiently wide to allow some activity surrounding an incident to be monitored.

| **Function and illustration** | **Screen image**<br>• monitor without zoom<br>• for objects of 1,6m height | **Examples of typical applications** |
|---|---|---|
| <br>detect | • ≥ 5% of the screen height.<br>• monitor the number, direction and speed of movement of people, providing their presence is known. | Perimeter security: Long-range images over parking lots, etc. |
| <br>monitor | • ≥ 10% of the available screen height.<br>• after an alert, an observer would be able to search the display screens and ascertain with a high degree of certainty whether a person is present. | Entrance areas: medium-range security of entrance halls, platform areas, etc |
| <br>recognise | • ≥ 50% of screen height<br>• viewers can say, with a high degree of certainty, whether or not an individual shown is the same as someone they have seen before.<br>• Recommendation also applies if target type or number of targets is important. Also they suggest ≥30% screen size for recognition that is allowed to be somewhat less reliable. | Mobile applications: Interior car and bus surveillance at door or call button area. Front-facing applications on vehicles or areas where bus or train exteriors are viewed. Short-range security for hallways, revenue and ticket areas, railroad crossings, call buttons, parking garage entrances/exits and elevator lobbies. |
| <br>identify | • ≥ 120% of the screen height<br>• picture quality and detail should be sufficient to enable the identity of an individual to be established beyond a reasonable doubt. | Mobile applications: Cash boxes, fare machines for crew safety. Short-range applications at ticket barriers, fare machines, cash rooms, garage barriers, and secure door entrances (license plate and payment machine). |

Ford [17] distinguishes four categories of criteria for image quality:

- Elements of the Action: in a very broad and general sense, identification of the series of events that took place
- Target Class: recognition of the general class of the target (e.g., person, car, type of object)
- Target Characteristics: recognition of unique characteristics of the target (e.g., gender, markings, scars, tattoos, dents, colour)
- Target Positive Recognition: recognition of a specific instance of the target (e.g., recognition of a person, a specific object, or an exact alphanumeric sequence)

In the field of cognitive sciences, over 50 years of research results are available regarding target detection, classification and identification (mainly in the Military). For example it concerns the job of radar operators. Whether an observer is able to interpret an image depends on contrast (image signal-to-noise ration), brightness, properties of the object (size, shape, colour), etc. An excellent overview can be found in Wickens and Hollands *Engineering Psychology and Human Performance* [19].

### 3.3.3 Hardware - monitors/screens

There a questions about the number of screens, the size and resolution of screens, the number of images on one screen (split screen techniques), "switching" of images on one screen (auto-cycling) and so on. Many considerations are given in literature, no hard evidence could be found.
All considerations are based on analogue VDT (visual display terminal) technology. Only one document [18] considers the idea of an independent pixel space: a predefined area for mapping images.

- **Tasks and the number of screens**
  - Howard [20] gives an overview of task performance literature for CCTV monitoring tasks. They mention some field and laboratory experiments. They believe that eye movement recording will be the key to develop models of human performance in CCTV monitoring tasks. Some eye track and eye fixation experiments including CCTV images have been carried out. However, no HF guidelines or consequences for CCTV system design haven been reported.
  - The number of screens should be sufficient to present the maximum number of simultaneous alarms as stated in the operational requirement [6] and may be determined on the basis of functional considerations, the number of operators on duty, etc. (no practical details given).
  - Dubbeld [21] found that 1100 camera views for 3 (during night 2) surveillance operators was problematic. Operators looked briefly at the views and switched a lot between views. When spotting an incident and contacting the police about it, the operators immediately turned to other camera views, instead of waiting until the police arrives. Otherwise the other views remained unattended for too long.
  - According to Gill et.al [22], operators switch to a reactive mode in case of too many camera views (>100 views). The probability of detecting an incident (surveillance) is substantially reduced when there are more than 100 cameras per operator.

- **Maximum camera views - some task related "hard" figures**
  - There is evidence [6, 23, 24] to suggest that the maximum number of camera views that can be effectively monitored is 16 or less.
  - The camera to monitor ratio should normally not exceed 10:1 [6].
  - Wood & Clarke [25] carried out some field experiments for traffic and rail:
    - an operator required to scan images to reliably detect targets, should concentrate on *one active monitor*, leaving other monitors blank until activated by an alarm or event;

    - no more than 9 simultaneously displayed pictures should be observed if pictures show *considerable movement* and the task primarily involves general surveillance;
    - no more than 16 simultaneously displayed pictures should be observed if pictures show *little movement* and the task consists of general surveillance and observation.
  - Weitenberg [26] conducted an experiment with short duration, very intense supervision tasks. The tasks concerned supervising 12 images presented in split screen on 3 monitors (3 x 4 images, screens mounted above the workstation). The workload was perceived as intense but acceptable.

- **Maximum number of screens**
  Keval and Sasse [27] found that one of the common problems in security control centres is the high screen to operator ratio. Several sources consider the maximum number of screens for one operator:
  - Workplace measurements;
    The visual area in front of the operator for primary information is limited to 70° width (for example 3 x 21" screens at 1 m viewing distance) and a 60° height [4].
  - APTA [9] indicates that 8 individual monitors, if stacked in two rows of four, is acceptable. A single operator should monitor no more than 10 screens simultaneously, although this figure may need to be reduced when the monitors show high levels of activity or detail that need careful monitoring. This is in line with the findings mentioned earlier of Wood & Clarke [25].
  - Some camera views may require constant monitoring and need a dedicated screen. Others may not, in which case a single screen could be used to cycle among several cameras [9].

- **Number of images per screen**
  Recommendations differ from rejecting multiple presentation to maximum quadruple presentation per screen [4, 9, 16, 28]. The relation between loss of resolution (image quality) by applying switches remains unclear. There is no reference to digital systems.
  - Auto-cycling should be avoided as it undermines the ability of peripheral vision to detect movement or status changes on banks of monitors [25, 4, 28].
  - Keval [3] suggest an auto cycling (or switching) frequency $< 5$ minutes as it is likely to cause visual discomfort and distract the user from the video events. FHWA [10] mentions the additional problem of recognizing at a glance the locations of the views currently on display.
  - Wood et al. [29] found in a traffic management trial with10 operators that they could detect more reliably an object on the hard shoulder, checking on one single image at a time (91x121 mm) instead of checking quadruple formats (on 19" screen at 70 cm viewing distance). Serial presentation is preferred over multi-image formats although it turned out to result in somewhat slower operator performance.

  - Howard et al. [20] report that search time for multiple targets exhibit large decrements with set size. In a multiple object tracking tasks, a capacity limit of around four objects is reported.

- **Size of CCTV screens**
  - An experiment by Wood et al. [29], showed that operators could achieve a suitable level of reliability in detecting an object of at least 10% of the image height on a road provided that larger image formats are used. The experiment included images of 91 x 121 mm, 163 x 217 mm and 284 x 380 mm presented on a 19" monitor at 70 cm viewing distance (284 x 380 mm is full screen presentation).
  - Shared (large) screens *may* be used under the following conditions [11]:
    (1) the group of users frequently refers to the same information and is required to interact as a team, based on the same information;
    (2) one or more members of a team of users must move about, yet must frequently refer to information required to make decisions - information they cannot carry with them or do not have displayed at their assigned position(s);
    (3) space or other constraints preclude the use of individual displays for each team member to call up commonly used information;
    (4) when it is desirable to have general information available to persons who shall not interrupt ongoing group operations by looking over the shoulder(s) of individual user(s) to see displays.

    Large screens shall be used only when the spatial and environmental conditions allow satisfactory observational geometry to ensure that all critical users have good visual access.
  - Monitors for close inspection are called "spot" or "incident" monitors. According to [4], wherever practical, a spot monitor should be positioned directly in front of the operator at a 0.5 - 1.5 m distance and range in size from 9" to 16" (23 - 41 cm) diagonal. Other recommendations mention viewing distances of 75 - 125 cm [7].
  - Weitenberg et al. [26] found out in an experiment that an additional monitor (hot spot) for detecting new incidents while handling incidents on a separate monitor close by the operator, did not increase performance in spite of being perceived as comfortable.

- **Display brightness and luminance contrast**
  - Display brightness and luminance ratios may be specified in line with contemporary display technology (flat screens at 250 cd/$m^2$ and 1000:1 dynamic contrast ratio [18]). Some sources [10, 11, 12] mention significant lower levels, presumably driven by (old) videowall technology.
  - Most screens will produce glare if external light sources (windows, light fixtures, etc. other displays) are located where they can reflect off of the screen. Glare may cause distraction, contrast reduction, blurring of images and so on and should be avoided [6].
  - Where only video display units are employed, illumination should be 200-500 lux. This is adequate for video display work and occasional reading of large or bold print, but is too dim for close work requiring reading of printed material [10].

### 3.3.4 Information structuring

Following a Human Factors approach it is obvious that information needs to be structured according to task characteristics. Guidance from literature:

- **Logic arrangement**
  - When large numbers of images need to be scanned frequently, monitor banks, split screen views, or switched images might be appropriate [23, 28]. To assist operators in viewing the images displayed on monitor banks, the pictures should be arranged or grouped with some underlying logic, e.g. mirroring the layout of the camera locations on a site [4, 28]. However, there is no statistical evidence to support this [30].
  - If the same coverage area is scattered over different parts of a monitor bank, this might cause problems for the operator to collate the information [31].
  - Scott-Brown and Cronin [32] found that the arrangement of monitors and the transitions from camera views can demonstrably influence overall detection levels.
  - If multiple visual channels are to be scanned, put them close together and in relative positions that reduce scanning (eye, head movement) [10].
  - Besides labelling, the content of information displayed shall be evident to a trained observer without requiring reference to display control settings [11].
  - Multiple spot monitors (varying 1-4) were advantageous because operators could passively observe an incident whilst actively monitoring other areas and observe incidents using more than one camera [22].
  - Operators have indicated that having blank screen monitors alongside the primary viewing monitor is distracting and causes visual discomfort. Ideally, these monitors should be placed in the periphery of the operator’s visual field [31].

- **Situational awareness**

  Maintaining operator awareness of the direction in which the camera is pointed is difficult, especially at night when many landmarks are not visible or when colleague operators also can manipulate cameras. FHWA [10] suggests the following for traffic control:
  - Consider using a consistent rule for camera location, like for example placing all of the cameras on the south side of the roadway.
  - Consistency in the location of images on a monitor bank, like geographical order, can be used as a means of inducing a systematic approach to sampling: “cueing” appropriate operator responses can also be assisted by the use of a stable layout of images on a monitor bank.
  - Specify consistent presets for camera's.
  - The name of the geographic location of the camera should be superimposed over the image on the monitor screen.
  - The display at the operator's station should indicate the direction of view of each camera. A rotating camera icon superimposed on the graphics display or a text display of the camera's compass pointing angle might be suitable.

- Cameras can be identified by codes, which should be logical and meaningful. Arbitrary identifier codes can be memorised over time, but increase initial operator training requirements, as well as memory demands during operations and should therefor be avoided.
- A situation map with icons depicting camera positions has proven to be a highly effective method of identifying cameras in both research and practice.

In addition:

- Good local knowledge of the area is necessary for camera search tasks [27]. It proved helpful to supply operators with a list detailing camera numbers and their locations. Knowledge of the geographical areas can also be accomplished by having frequently contact with the local police (in case of surveillance) [22].

• **Viewing distance**

Recommended viewing distances found in literature:

- 5 times the displayed picture diagonal [6].
- 3 - 5 times screen diagonal [9].
- 3 - 5 times screen height;
- 2 - 5 times screen diagonal for 3:4 aspect ratio images [16].
- 46 - 64 cm for 30 - 48 cm video screen [10].
- For direct view electronic displays (TFT) with a diagonal between 33 - 76 cm [11]:
  - viewing distance from the eye reference point of the seated user to displays located close to their associated controls < 70 cm;
  - effective viewing distance to displays shall be larger than 33 cm and preferably 51 cm;
  - when periods of display observation are short, or when dim signals must be detected, the viewing distance may be reduced to 25 cm;
- Another source [23] suggests a viewing range of 0.5 - 2.0 m for a display of 35 cm and a greater viewing distance for a display of 53 cm.
- Large screen viewing distance:
  - not *closer* than 50% of the width or height of screen (whichever is greater) [9, 10].
  - the large screen shall not be placed farther from an observer than will provide appropriate resolution of critical detail presented on the display [11]. Of course this is a correct statement, however, what is "appropriate"? Ahlstrom [12] mentions, that an analysis of information requirements will be needed.
  - ratio of viewing distance to screen diagonal: more than 2:1 and less than 8:1 with an optimum ratio of 4:1 and a preferred range of more than 3:1 or less than 6:1 [12].
  - TNO Research Institute [18] specifies without further evidence a maximum viewing distance of 10.3 times the screen diagonal of the smallest image on display.
- Monitor bank

  Video monitors placed within a monitor bank should be positioned at a minimum distance of 1.5m from the CCTV observer. Video monitors placed within a monitor bank should be between 17–28" [3, 21]; we assume one camera image per monitor.

- **Location of large screen displays**
  - A large-screen display shall be located with respect to critical observers so that the view of the display is not obscured regularly by persons moving in normal traffic patterns [11].
  - Individual viewers of a large-screen display in a fixed location should be no more than 10° off the centerline and multiple viewers should be no more than 30° off the centerline with a preferred limit of 20° [12].When multiple, large display devices are used, the normal work areas of each user should be within the acceptable off-centerline viewing area of each large display that each user must view.
  - Spatial distortion of the image and the effect of the viewing angle upon screen characteristics such as brightness and colour rendition should be taken into account. There are considerable differences between projection technology and TFT-LCD screens.

## 3.4 Measuring image quality

Several test targets have been developed in order to review CCTV system performance, such as the Rotakin test [14, 6]. The test kit includes a 1.7 m puppet with a printed test chart (figure 6). The Rotakin test is approved by the UK Home Office. There are approved procedures for testing, including facial recognition (figure 6, right side). Comparable test charts (figure 7, 8) have been developed [2, 15].

By comparing a test chart and the presentation on a screen the overall reduction of the image quality can be seen. The items on test charts provide measurements for several factors. For example: a series of parallel lines can be presented with a decreasing distance between the lines. At a certain distance between two lines (visible on the test chart), we may not be able to see two lines anymore on the screen (they melt together).
Another example: on the Rotakin a license plate may be mounted, in order to check whether the text on the plate can still be read.

Cohen [16] distinguishes four areas when determining the required image quality:

- Clarity
  Is the picture sharp enough, and is there any lens distortion? Ensure that the lens or lens / camera combination is of sufficient quality for the task in hand.
- Detail
  Is there enough detail to identify objects? Check that image quality is not compromised by trying to achieve a large FoV at the cost of image detail, and that lighting levels permit a useable depth of focus. If necessary break the scene into smaller sections.

Figure 6. Rotakin (left) and additional facial test charts.


Figure 7. Vidilab test chart.

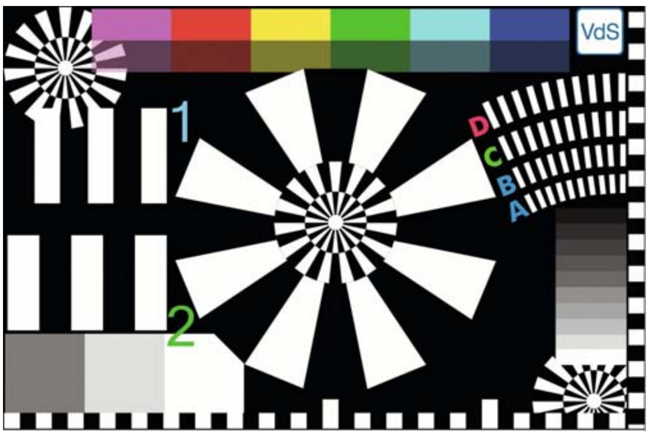



Figure 8. Test chart.

- Colour
  Is it natural? Is it necessary? If accurate colour reproduction is important then ensure the lighting is of sufficient quality and quantity to allow the cameras to achieve this.
- Artefacts
  Are there elements in the image that should not be there? And if so are they obtrusive? If this is the case then depending on the artefact, either the amount of compression needs to reduced or the camera/lighting placement needs to be addressed.

A suitable test target should be employed that contains the required level of detail to test the operational parameters of the system [16]. Field tests should be carried out to verify for example the camera's field of view, image detail, and live (and recorded) image quality. Some sample video should be recorded and exported from each camera. This can be used as a reference of image quality and camera field of view during future system maintenance operations and will highlight any change or degradation that occurs in the system over time.

Damjanovski [2] suggested to develop a key variable "Identification CCTV Unit" (ICU) to indicate the image quality after processing in the CCTV system (system = scenery (test chart) --> camera --> encoding --> transmission --> decoding --> presentation on screen). Starting point for the key variable is a value of 1.0 (100%) at a test chart. Image quality may be reduced by each component (step) of the CCTV system, assuming reduction factors per step can be determined. Unfortunately, at this point, the work of Damjanovski becomes inconsistent.

## 3.5 Controls for CCTV images

Typically, the operator has controls for camera (image) selection, assigning an image to a given monitor, pan/tilt/zoom (PTZ) and focussing. Control devices include keyboards, button boxes, joysticks, mice, touch screens, or trackballs.

- **General** recommendations for control panels [4] include:
  - controls should accommodate the needs of both right- and left-handed operators;
  - the position of critical buttons on a control panel, such as those which send cameras to pre-set positions, should be arranged such that they minimise the possibility of accidental errors;
  - where panels from different suppliers are used, a consistent design should be aimed at reducing errors.

- **Control devices [10]**
  - Mouse control for camera selection may be effective for traffic camera selection when cameras are represented by small icons (on a screen).

- Numeric keypad or keyboard to enter camera identifiers. If a keypad is used, it should contain the 10 numerals, "Enter", editing keys (e.g., Delete, Backspace), and any keys necessary to perform specialised functions associated with camera operation.
- In case of a touch screen to select traffic camera views, ensure adequate icons size and icon separation in order to reduce errors.
- The mouse is the most effective device for selecting a camera from a map display, but not necessarily for controlling the camera itself. Keyboard interfaces allow operators to make the very precise control movements that are necessary to orient cameras.

- **Camera control**
  - A nominal panning rate of approximately 12˚/second for both azimuth and elevation allows the operator to accurately control the camera without taking undue time [10].
  - Total Engineering Practices [33] specifies a minimum speed of 5˚/sec for panning and 2˚/sec for tilt.

- **Camera presets** [10]
  - Cameras with preset views are more efficient to operate than manual cameras. The best interface is one that allows the operator to step between several presets, then switch to manual control for precise camera pointing (which is called by FHWA a hybrid controller).
  - Manual control of cameras with preset pan and tilt positions should still be possible.
  - A cell of grouped keyboard directional arrow control buttons (up, down, left, and right arrows) have been shown to be highly effective control mechanisms for camera panning, tilting, and zoom & focus functions (when coupled with additional dedicated control buttons).

- **Screen control** [10]
  Screen control concerns brightness, contrast or colour adjustments.
  - Adjustment controls for brightness, contrast or colour should be located on the front of the screen and easily visible, available and labelled.
  - Control of large-screen group display systems shall ensure that critical information cannot be modified or deleted inadvertently or arbitrarily. Changes in the group display shall be controlled by designated users.

In some situations, two or more operators may be competing for control of the same camera or monitor. Methods of prioritisation need to be established.

# 4 CCTV Supervision Tasks

The ultimate goal of CCTV systems is remote supervision, related to security (surveillance, crime prevention), safety (related to hazardous environments), process control (check visual information before control actions), and service tasks. In some cases manpower reduction (remote monitoring of loading/ unloading) is an issue. Operator tasks vary from proactive monitoring to reactive screening based on alarm messages. Some monitoring tasks may be considered vigilance tasks: events rarely occur, which makes the task hard to fulfil (amongst others general sources, refer to Wickens [19].

The Department of Psychology of the University of Porthmouth published an extensive review of cognitive and ergonomic literature on the CCTV review process (Hillstrom et.al [34]). The review was limited to CCTV footage review (i.e. review of recorded CCTV images), a task which is not included in the scope of this project. Nevertheless, it is worthwhile reading, because the recommendations may be applicable to live stream surveillance CCTV review. Furthermore, regarding CCTV Supervision task, refer to the general functions of CCTV-systems in HF perspective may be considered tasks: monitor, detect, (observe), recognise, and identify (in section 3.3).

## 4.1 Workload and task performance

- CCTV images vary from simple (e.g. loading a truck) to complex (small details in a busy traffic scene). Cole & Wood [35] discovered in a explorative experiment that the amount of 'relevant' information within a picture produced the greatest associations of perceived complexity. Camera and task related criteria, like zoomed in scenes including a large number of objects, presence of many colours or camera angle, were also linked to task difficulty. They conclude that perceived image complexity correlates to task complexity and that a development of a CCTV complexity metric is needed (note: however not yet developed).

- Operators, by preference, will monitor traffic using colour CCTV rather than well-designed support systems with graphical user interfaces when both are available. This preference may result in high visual workload.

- If more channels must be monitored, the attention resources of the operator must be increased. The most obvious way to do this is to reduce the amount of attention that must be paid to other tasks. In an emergency situation in a Traffic Management Centre, the operator who must do the monitoring might temporarily have other tasks given to temporary personnel in order to attend to the important information [10].

- Neyland [36] states that operators distinguish two situations: public flow (everything normal, no attention needs to be paid) and spaces of otherness (particular incidents of accounting attention). The operators make the distinction based on time, location, materials, visuals, text (reports from earlier), and talk (with police officers in the field). For example, a group of people in the middle of the day might be normal, however might considered a space of otherness when in the middle of the night.

- It is important to state which areas are more important (and therefore should be monitored more per cent of the time) and which areas are less important [22].

- Parasuraman etal. did some experiments [37] on the vigilance performance of participants watching videos depicting intentional action (e.g. a person grasping an object, either a knife or a hairdryer) in order to detect infrequent threat related actions. Detection rate declined over time and subjective workload increased. Another study [39] looked into additional software generated cues to keep the operator focussed on his (vigilance) task. Vigilance studies can be considered "old science" applied to new types of work environment; no surprising new outcomes are expected.

- Humans are not so efficient at detecting changes in unfamiliar context or for static inanimate objects. Hence there is a difficulty in spotting unattended bags or vehicles according to Gill et.al [22].
  They also mention that people tend to fail to notice changes introduced during blinks, flickers, or disruptions. In the context of CCTV footage, the biggest disruption to motion signals is in the storage format. Even when people are aware of the conditions that make change blindness likely, they will still underestimate their own susceptibility to it [34].
  Related to this, MacCarley [38] reports that objects moving at rates up to 20 deg/sec are nearly as detectable as stationary objects. At 120 deg/sec, small objects can still be detected, but must be increased in size threefold over their threshold size at a given light level when stationary.

## 4.2 Breaks and duration

- Where CCTV is used, short, frequent breaks are recommended rather than occasional longer breaks; for example, a 5-10 min break after 50- 60 min continuous screen and/or keyboard work is likely to be better than a 15 min break every two hours [4]. This guideline seems to be based on occupational health directives for visual display work.

- Wood & Clarke [25] state that operators remain more attentive when working in 20 minute blocks with 5 minutes rest in between. Certain types of task may warrant even shorter periods, as in level crossing operation high levels of detection performance were more likely to be maintained if total duration of inspection was less than 5 minutes.

- Wood & Clarke [25] mentioned that some sources suggest 3 hours to be the maximum continuous period for reliably doing general surveillance tasks, with a cumulative total of 6 hours per day dedicated to CCTV tasks.

## 4.3 Training

FHWA [10] recommends training of the person to effectively scan information channels and develop optimal scan patterns (no details given).

- FHWA [10] states that it is often assumed that video-breed information is the best indicator of traffic flow. After all, the operator can simply observe what is happening. However, standard video is not easily quantifiable. Individuals are less accurate and reliable when their judgements are based on video images than when their judgements are based on a numerical assessment of the same situation.
  Operators tend to become overly dependent on CCTV displays, probably because of the natural presentation of information and the potential for detecting a traffic flow problem marginally earlier than would be likely with sensors alone. Scott-Brown and Cronin [33] note that one of the biggest barriers to progress is our own overestimation of our visual detection abilities. Over-dependence on CCTV for incident and congestion detection can create a significantly higher workload than a well-designed graphics display that summarises, for example, loop detector data.
  While no definitive guidelines are presently available, it would seem prudent to limit the number of monitors in use [10].

- All CCTV users who perform security observation tasks with CCTV video and images must be trained and regularly assessed (every 6-8 months) on their tasks, the surveillance environment, security procedures, and the use of tools and equipment [3].

- According to Gill et al. [22], to be effective, an (surveillance) operator requires a range of skills and a specific knowledge base including:
  - Technical and operating skills to track targets effectively, obtain evidential quality images, and use communication links to pass on intelligence;
  - Geographical knowledge of the target area and camera locations;
  - Knowledge of crime-related activity in the area, so that they can direct their surveillance;
  - Knowledge of relevant legislation such as data protection laws, so that they can operate legally.

## 4.4 Intelligent automation

Several mechanisms for intelligent automation are reported in literature. Operators could be supported in analysing image content and execution of accompanying tasks by automation of some basic standard instructions or settings of the system.

- Wood & Clarke [25] recommend that methods for directing operator's attention to particular screens should be considered in situations where several monitors are present or when operators are involved in secondary non-CCTV tasks.

- FHWA [10] is critical and states that merely providing a system that automates the incident detection task does not guarantee that an operator will interact with it as a means of obtaining incident detection assistance. Experimental results suggest that operators, when given supplemental incident detection tools, can compensate for deficiencies - such as long detection latency and low hit rate - built into an automated incident detection system. The FHWA agrees that automation can support CCTV monitor display by using established cycling patterns and pre-defined situation based priorities.

- Although the signalling of an alarm condition to the CCTV system normally has priority over other inputs, automation should be flexible so that an operator is able to take manual control of the system after alarm irrespective of the degree of automation 6, 10]. Generic guidelines on interval of alarms, strength of cues and redundancy are added by FHWA [10], but left out of scope in this review.

## 4.5 Other literature

During the literature study a report on *Requirements for Transportation Management Center (TMC) Human Factors Guidelines* (McCallum [40] was found. Apparently the FHWA feels a need to have HF guidelines developed for the design of their TMCs. The document does not specify actual guidelines. However it recommends an interesting procedure for the development of such guidelines. It also suggests some formats for presenting guidelines, for visiting existing sites (the situation analysis) and gives a preliminary table of content for a handbook. Some concepts of this document may be useful for the next phases of the CCTV research project. See for an example of a design guideline format, the reprint on the next page.

## DETERMINING VISUAL DISPLAY CONFIGURATION

**Introduction**

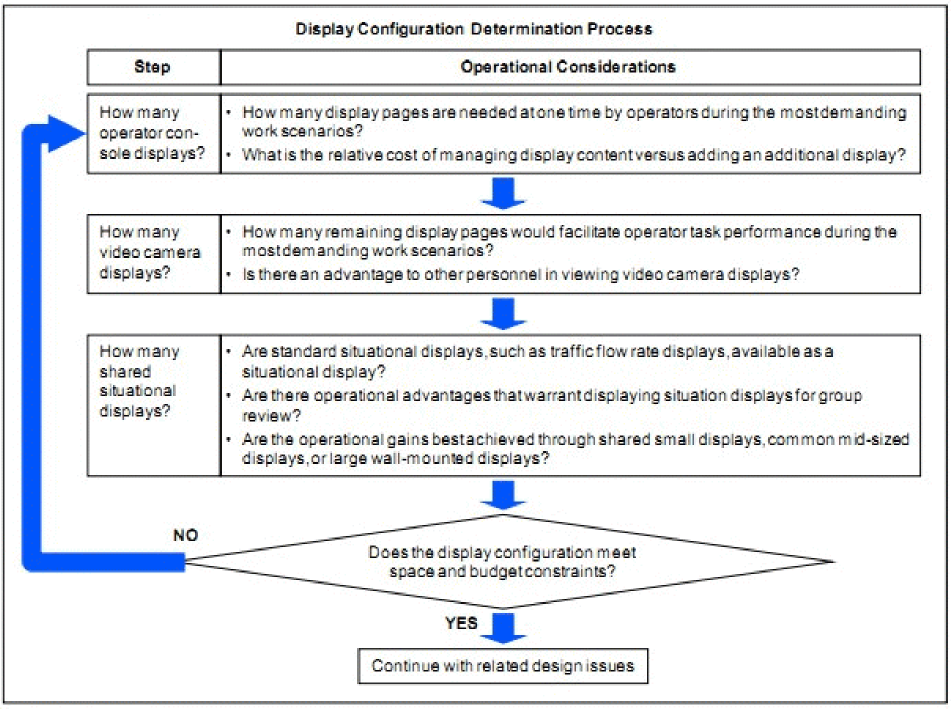
*Visual Display Configuration* refers to the number, size, and placement of visual displays throughout a TMC. The configuration of TMC displays will affect how quickly TMC operators and other TMC personnel can access the visual information that they need to perform their tasks. Very often the decisions that determine visual display configuration are made during the early stages of TMC design or remodeling. Careful consideration of the tasks performed by individual TMC operators and groups within a TMC will help to ensure that a display configuration that best supports the functions performed within a TMC is selected.

An iterative process for determining a TMC's visual display configuration is outlined in this design guideline. The basic steps reflect the consideration of key operational issues, followed by an assessment of space and budget constraints. More detailed design decisions will require a careful analysis and consideration of the TMC functions, operator(s) jobs and tasks, the visual information that they use to perform those tasks, the means of best providing that visual information, and the alternative means of managing displayed information on TMC visual displays.

**Design Guidelines**

The basic steps and corresponding operational considerations should be analyzed in sequence to initially determine operator console and situational wall display requirements.

| Based Primarily on Expert Judgment | Based Equally on Expert Judgment and Empirical Data | Based Primarily on Empirical Data |
|---|---|---|




TMC GUIDELINES 6 May 2009

# 5 Conclusions

Over 40 recent publications on Human Factors and CCTV have been found, including some standards and handbooks. In addition, it was noticed that several institutes in the UK and the USA announced research projects in this area. The outcome of the literature search can be summarised as follows (and in no particular order):

1. Literature on Human Factors of CCTV-systems is limited. There are three main CCTV application areas of interest:
   - traffic (traffic management/high way); some guidelines and some experiments are reported from the UK and USA.
   - rail and underground (UK related research), in particular safety issues.
   - camera surveillance of public areas (crime watching); predominantly UK literature.

   There are no references to process control (supported by CCTV), such as lock and bridge control, loading & unloading supervision, process unit control, and so on.

2. All documents found thus far are based on analogue systems (camera, displays). There is no HF research available related to IP-technology and the application of TFT-LCD technology in control centres.

3. There are several listings of "things you should think of" in designing CCTV control centres, however no design requirements. Any design guidance we did find concerns well established Human Factors guidelines for control centres (in general), such as:
   - workplace measurements
   - text labels on CCTV-images/screens
   - logical layout of monitor banks; logical mapping of camera's on screens
   - control devices.

4. There is some evidence regarding the number of images an operator can handle for several different types of tasks. Reports of experimental research is not very accurately documented. In many cases the size and type of screens is not mentioned, nor actual sizes of the image representation, image (or camera) resolution, or viewing distances.
   All literature indicates relative low numbers of images that can be handled by one operator (for specific tasks). The maximum number found: 16.

5. There is only one source of useful field experiments, i.e. projects carried out by John Wood and colleagues (UK), regarding traffic control (checking hard shoulder for objects), rail safety (level crossing experiments), and incident detection (in a shopping area). There is too little evidence to be able to generalise the outcomes.

6. We did not find a definition of, or experiments into, image quality. The mechanism of imaging, coding, transmission and decoding are known; the effects of control variables along this system on image quality (and in particular task performance of operators) is not known nor researched.

7. We did not find information on quantifying image complexity or image density, nor information on object sizes related to camera (and screen) resolution, viewing distances or operator task characteristics.

8. There are no human factors related case studies or situation analyses on CCTV control centres (design projects) reported in literature. This is unlike the state-of-the-art in process control room (design projects).

# 6 References


[1] Landman, R.B., Pikaar, D.N.; Pilotonderzoek details op camerabeelden - Rijkswaterstaat. Onderzoek ErgoS Engineering & Ergonomics (2010).

[2] V. Damjanovski. CCTV Networking and digital technology, Burlington, Elsevier Butterworth - Heinemann (2005).

[3] Keval, H.U.; Effective design configuration and use of digital CCTV. Thesis - Department of computer sciences, University College London (2009).

[4] ISO11064 Ergonomic Design of Control Centres - Multi part standard.

[5] EEMUA; Process plant control desks utilising human computer interfaces, publication nr. 201; The engineering equipment and materials users association, London (2002).

[6] EN 50132-7; Alarm systems - CCTV surveillance systems for use in security applications. Part 7: application guidelines (1996).

[7] AS 4806.2; Closed Circuit Television (CCTV) - Part 2: Application Guidelines (2006).

[8] Wood, J.; CCTV Ergonomics: Case Studies and practical guidance. In: Pikaar, R.N., Koningsveld, E., Settels, P. (Eds): Meeting diversity in ergonomics; Elsevier, Amsterdam (2007).

[9] APTA CCTV Standards Working Group; Recommended practice APTA-IT-CCTV-RP-001-11: Selection of cameras, digital recording systems, digital high-speed networks and train lines for use in transit-related CCTV systems (2011).

[10] FHWA - US Department of transportation - Federal Highway Administration. ITS Joint Program Office. Preliminary human factors guidelines for traffic management centers. Publication no. FHWA-JPO-99-042 (1999).

[11] MIL-STD-1472G; Department of Defence - Design Criteria Standard-Human Engineering (2012).

[12] Ahlstrom, V., Kudrick, B.; Human factors criteria for displays: A human factors design standard. Update of chapter 5 - Technical report DOT/FAA/TC-07/11. US Department of transportation - Federal Aviation Administration (2007).

[13] Brunt, M., Wood, J., Fix, C., Harding, T., Clarke, T.; Human factors in the upgrading of railway control equipment. In: Meeting Diversity in Ergonomics. Proceedings of the International Ergonomics Association congress; Elsevier, Amsterdam (2006).

[14] Aldridge, J. & Gilbert, C.; Performance testing of CCTV. Perimeter surveillance systems. Using the Rotakin Standard Test Target. Version 1.0. Police Scientific Development Branch. Publication 14/95. Home Office - Police policy directorate (1996).

[15] Australian Government; A National Approach to Closed Circuit Television - National Code of Practice for CCTV Systems for Mass Passenger Transport for Counter-Terrorism (2012).

[16] Cohen, N., Gattuso, J., MacLennan-Brown, K.; CCTV operational requirements manual. Scientific Development Branch publication. Home Office no.28/09 (2009).

[17] Ford, C.G., McFarland, M.A., Stange, I.W.; Subjective video quality assessment methods for recognition tasks. Proceedings of the SPIE, 2740 (2009).
[18] Koop, D.; Systeem/ subsysteem design description. Uniforme Werkplek Wegverkeersleiders - fase 1. Rijkswaterstaat - Dienst Verkeer en Scheepvaart (2009).
[19] Wickens, C.D., Hollands, J.G.; Engineering Psychology and Human Performance, 3rd edition, Prentice Hall, New Jersey (570 pages).
[20] Howard, C.J., Troscianko, T., Gilchrist, I.D, Behera, A., Hogg, D.C; Searching for threat: factors determining performance during CCTV monitoring. In: Waard, D. de, Godthelp, J., Kooij, F.L, Brookhuis, K.A. (Eds.). Human factors, security and safety. Shaker Publishing, Maastricht (2009), pp. 1-7.
[21] Dubbeld, L.; The role of technology in shaping CCTV surveillance practices. Information, Communication & Society, 8, (1): 84-100 (2005).
[22] Gill, M., Spriggs, A., Allen, J., Hemming, M., Jessiman, P, Kara, D., Kilworth, J., Little, R., Swain, D.; Control room operation: findings from control room observations. Home office - Research, Development and Statistics Directorate. Online report 14/05 (2005).
[23] Wallace, E., Diffley, C., Baines, E., Aldridge, J.: Ergonomic design considerations for public area CCTV safety and security applications. In: Seppälä, P., Luopajärvi, T., Nygård, C., Mattila, M. (Eds.). From Experience to Innovation. Proceedings of the International Ergonomics Association congress (1997).
[24] Velastin, S.A.; Getting the best use out of CCTV in the railways. New and emerging technologies (2003).
[25] Wood, J., Clarke, T.; Practical guidelines for CCTV ergonomics. In: Meeting Diversity in Ergonomics. Proceedings of the International Ergonomics Association congress; Elsevier, Amsterdam (2006).
[26] Weitenberg, A.I.M., Jansen, E.J.M., van Leiden, I., Kerstholt, J., Ferwerda, H.; Cameratoezicht - de menselijke factor. Kerckebosch, Zeist (2003).
[27] Keval, H.U. & Sasse, M.A.; Man or a gorilla? Performance issues with CCTV technology in security control rooms. Proceedings of the 16th World Congress on Ergonomics, International Ergonomics Association, Maastricht, Netherlands (2006).
[28] Wallace, E. & Diffley, C.; CCTV: Making it work. CCTV Control room ergonomics. Police scientific development branch. Home Office police policy directorate publication no. 14/98 (1998).
[29] Wood, J., Freer, M., Pierce, C.; CCTV User trials for the M42 active traffic management pilot project. In: Meeting Diversity in Ergonomics. Proceedings of the International Ergonomics Association congress; Elsevier, Amsterdam (2006).
[30] Stedmon, A.W., Harris, S., Wilson, J.R.; Simulated multiplexed CCTV: The effects of screen layout and task complexity on user performance and strategies. Security Journal, 24, (4): 344-356 (2011).
[31] Donald, C.; Ergonomic considerations in CCTV viewing. Hi-Tech Security Systems, 4, (3) (1998).

[32] Scott-Brown, K.C., Cronin, P.D.J.; An instinct for detection: psychological perspectives on CCTV surveillance. The Police Journal, 80 (2007).

[33] Total Exploration & Production; General specification Telecommunications GS EP TEL160 - Design and evaluation of CCTV system (2008).

[34] Hillstrom, A., Hope L. Nee, C. Applying psychological science to the CCTV review process: a review of cognitive and ergonomic literature. Project Report. Home Office Scientific Development Branche London (2008).

[35] Cole, H., Wood, J.; The relationship between image complexity and task difficulty as rated by experienced and novice viewers. In: Meeting Diversity in Ergonomics. Proceedings of the International Ergonomics Association congress; Elsevier, Amsterdam (2006).

[36] Neyland, D.; The accomplishment of spatial adequacy: analysing CCTV accounts of British town centres. Environment and Planning D: Society and Space, 24: 599-613 (2006).

[37] Parasuraman, R., Visser, E. de, Clarke, E., McGarry, R., Hussey, E., Shaw, T., Thompson, J.C. Detecting threat-related intentional actions of others: effects of image quality, response mode and target cuing on vigilance. Journal of experimental psychology: applied. 15, 4, 275-290 (2009).

[38] MacCarley, A.; Advanced image sensing methods for traffic surveillance and detection. California PATH Research Report UCB-ITS-PRR-99-11 (1999).

[39] Donald, F.M. & Donald, C.; Vigilance and the implications of using threat image projection (TIP) for CCTV surveillance operators. In: Valli, C. & Brooks, D. (Eds.). Proceedings of the 1st Australian security and intelligence conference (2008).

[40] McCallum, M.C., Campbell, J.L., Brown, J.L.; Technical Support and Assistance for the FHWA’s Human Centered Systems Team. Task Order No. 1. Requirements for Transportation Management Center (TMC) Human Factors Guidelines. Project Report (Final) prepared for Federal Highway Administration (2009).

# 7 List of definitions

**Brightness**
Attribute of visual sensation according to which an area appears to emit more or less light (ISO 9241-302:2008; Ergonomics of human system interaction - part 302: Terminology for electronic visual displays).

**CCTV system**
Closed Circuit Television System. A television transmission system in which live or prerecorded signals are sent over a closed loop to a finite and predetermined group of receivers.

**CCVS**
Composite colour video signal (CCVS): The video output signal of an analogue colour camera comprising the burst and colour information (colour) the picture luminance component (video), black reference (blanking) and the synchronisation components (synchronisation).

**CIF (Common Intermediate Format)**
Pixel resolution of a video image. For PAL (phase alternating line) video, CIF is 288 x 352 pixels; for NTCS (National Television Standards Committee) video, CIF is 240 x 352 pixels. Note: 4CIF (PAL) is 576 x 704 pixels.

**Codec**
The codec describes the encoding and decoding of video data streams. Examples of codecs are MPEG4, or H.264.

**Control**
Device that directly responds to an action of the operator, e.g. by the operator applying pressure (ISO 11064-5).

**Control room**
Core functional entity, and its associated physical structure, where control room operators are stationed to carry out centralised control, monitoring and administrative responsibilities (ISO 11064).

**Control room operator**
Individual whose primary duties relate to the conduct of monitoring and control functions, usually at a control workstation, either on their own or in conjunction with other personnel both within the control room or outside (ISO 11064-5).

**Control workstation**
Single or multiple working position, including all equipment such as computers and communication terminals and furniture at which control and monitoring functions are conducted.

**Depth of field (DoF)**
range of visual focus of images from the distance at which all images are in focus (ISO 9241-302)
When you focus precisely on a subject, a certain distance in front of and behind the subject also will be in focus, although not necessarily as sharp. Depth of field increases or decreases based on the focal length of the lens, the distance to the subject, and the aperture.

**Design viewing distance**
distance, or range of distances, between the screen and the operator's eyes for which the display is designed to be viewed (ISO 9241-302).
NOTE The distance or range of distances (specified by the supplier) between the screen and the operator's eyes for which the images on the display meet the requirements of this part of ISO 9241, such as character size, raster modulation, fill factor, spatial instability (jitter) and temporal instability (flicker).

**Display**
Device for presenting information that can change with the aim of making things visible, audible or discriminable by tactile or proprioceptive perception (ISO 11064-5).

**Display, visual**
display (in the sense of format) providing visual presentation of data, mappings or videos (NEN-EN-ISO 11064-5)

**Field of view (FOV)**

- Angular region subtending the active area of a display as observed from the design viewing direction or other eye position (ISO 9241-302).

- Related to cameras, the following definition was mentioned: The area of a scene, observed by a camera and lens combination and measured both horizontally and vertically, that can be seen through the camera.
  The FoV width in mm = Format (horizontal width of sensor chip) × Distance (from camera) / Focal Length. Format refers to optical format and in this case is the horizontal width of the camera’s image sensor.
  Examples (FoV in degrees):

| | | | |
|---|---|---|---|
| – sensor-chip | 2/3" | 1/2" | 1/3" |
| – sensor area | 8.8x6.6 | 6.4x4.8 | 4.8x3.6 mm |
| – focal length | | | |
| 4.0 mm | - | 77 | 67 |
| 4.8 mm | 83 | 67 | 57 |
| 6.0 mm | 70 | 56 | 48 |
| 8.0 mm | 56 | 44 | 36 |
| 12 mm | 39 | 30 | 25 |
| 16 mm | 30 | 23 | 17 |
| 25 mm | 18 | 15 | 12 |
| 50 mm | 10 | 7 | 6 |

**F-number**
The lens aperture, indicated by f-number or f/.. determines the amount of light on the sensor. The f-number is given by the focal length f divided by D (diameter of the effective aperture).

**Frame**
One of the many still images that compose a complete moving picture in film, video production, animation and related fields. (2011, APTA)

**Frame rate/ frame frequency / fps**
Frame rate, or frame frequency, is the measurement of how quickly a camera produces unique, consecutive images called frames. The term applies equally well to computer graphics, video cameras, film cameras and motion capture systems. Frame rate is expressed in frames per second (fps) or simply hertz (Hz). The more frames per second used, the more information is available regarding motion. Full-motion video is achieved at approximately 22 to 24 fps when viewed by the human eye. (2011, APTA)

**Human Centred design approach**
Approach to interactive system development, focussing specifically on making systems usable, and emphasizing th role of human operators as control agents who maintain authority within a working system (ISO 11064).

**Legibility**
Ability for unambiguous identification of single characters or symbols that may be presented in a non-contextual format (ISO 9241-302).

**Line of sight**
Line connecting the point of fixation and the centre of the pupil (ISO 9241-3).

**Line of sight, normal**
Inclination of the line-of-sight with respect to the horizontal plane, when the muscles assigned for the orientation of the eyes are relaxed. For design purposes, an inclination of 15° below the horizontal plane is usually assumed.

**Monitoring**
activity for the purpose of detecting deviations from normal operation (by checking variables, or their course against limits, trends or the values of other variables) to enable timely and appropriate action for response (ISO 11064-5)

**Pan and tilt (zoom) unit (PTZ)**
A motorised unit permitting the vertical and horizontal positioning of the camera equipment. (sometimes including zoom function) (NEN-EN 50132-7).

**Preset shot**
A function in pan and tilt units and/or zoom lenses, which allows automatic return to one or more predetermined positions. (NEN-EN 50132-7)

**Pixel**
A single point in a graphic image. “Pixel” is short for “picture element”. Pixel is a term used specifically with cameras and is directly related to horizontal lines of resolution.
Pixels are the actual number of light-sensitive elements that are within the camera-imaging device.

**Readability**
Characteristics of a text presentation on a display that affect performance when groups of characters are to be easily discriminated, recognised and interpreted (ISO 9241-302).

**Rotakin**
A performance test target for CCTV systems developed by the UK Police Scientific Development Board. Also known as Rotatest, it evaluates system-level performance and resolution capabilities, including playback and recordings, end-to-end (2011, APTA).

**Shared visual display device**
on-workstation visual display which needs to be used by more than one control room operator while they are at their control workstations.

**Situational analysis**
Task analysis in an existing situation to analyse all the behavioural aspects of the work system, such as revealing practical experiences, informal communication, expectations and complaints of current users and any other facts that might be useful for redesign purposes (ISO 11064).

**Switcher / multiplexer**
a switcher/ multiplexer memorises scenes from each camera, compresses them and then displays multiple scenes on a single monitor. A multiplexer adds an encoded signal that allows a picture from each camera to be viewed in succession (as with switchers) or simultaneously.

**Task**
human activities required to achieve a goal (ISO 11064-5)

**Task analysis**
Analytical process employed to determine the specific behaviours required of people when operating equipment or doing work (ISO 9241-5; 1998).

**Validation**
Confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application are fulfilled.
(ISO 11064-7).

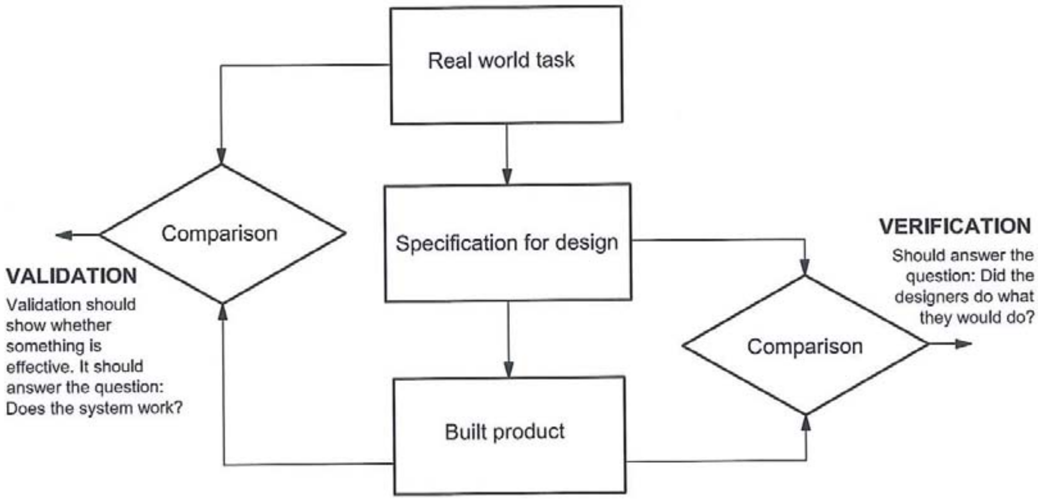



**Figure 1 — Role of verification and validation (V&V)**

**Verification**
Confirmation, through the provision of objective evidence that specified requirements have been fulfilled (ISO 11064-7).

**Visual angle**
Angle measured in minutes of arc subtended at the eye by the viewed object, such as a character or symbol (ISO 11064-4).
It represents an apparent size of an object based on the relationship between an object's distance from the viewer and its size (perpendicular to the viewer's line of sight). If an object of size h is at a distance d from the retina, the visual angle subtended (x) is equal to arctan (h/d).

**Visual field, field of vision**
Physical space visible to an eye in a given position (ISO 11064-4). Separate, distinct stimuli in the visual field will be detected even if they appear simultaneously. While the extent of the visual field is approximately 35° to both sides (left and right, total of 70°), only 1 - 2° of these are for sharp vision.