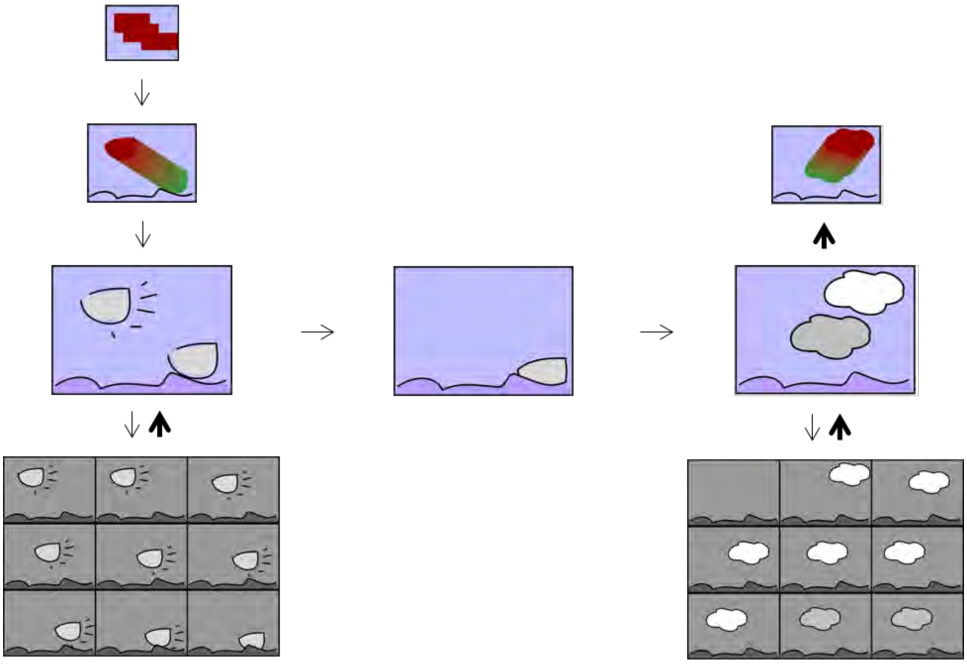
JRC Scientific and Technical Reports

VideoZoom

Summarizing surveillance images for safeguards video reviews

Scott Blunsden and Cristina Versino



EUR 25215 EN - 2011

EUROPEAN COMMISSION

Institute for Transuranium Elements

The mission of the JRC-ITU is to provide the scientific foundation for the protection of the European citizen against risks associated with the handling and storage of highly radioactive elements.

European Commission
Joint Research Centre
Institute for Transuranium Elements

Contact information
Address: Cristina Versino
E-mail: cristina.versino@jrc.ec.europa.eu
Tel.: +39 0332 786252
Fax: +39 0332 786082

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 68054
EUR 25215 EN
ISBN 978-92-79-23092-9 (pdf)
ISBN 978-92-79-23091-2 (print)
ISSN 1831-9424 (online)
ISSN 1018-5593 (print)
doi:10.2788/12093

Luxembourg: Publications Office of the European Union, 2011

Printed in Italy

Table of contents

List of Figures

List of Tables

0. Summary

Safeguards surveillance images are reviewed in batches of several thousands. In a batch, less than 0.01% of the total number of images is expected to be safeguards-relevant. Because events are to be detected and annotated by nuclear inspectors in review reports, there is a need for tools to focus the inspector's attention directly to the relevant parts of the image stream.

This report presents VideoZoom, a review tool that builds automatic summaries out of large surveillance streams taken by cameras with a fixed point of view. The purpose of summaries is to guide the review of the image stream and reduce the number of images seen by inspectors to perform the review work correctly. In building summaries, we do not rely on prior information about the visual appearance of expected relevant events. We assume the inspector to be knowledgeable about the nuclear processes she is reviewing. She is able to use the review tool in an active way, for example, by browsing the summaries to search for 'logically expected events', or decide when to expand parts of a summary to reveal the images as taken by the camera.

The report is organized in three parts. In the first part (Sections 1 to 3), the context for safeguards image reviews is set. Reference is made to review approaches in use or proposed as R&D contributions. The goal of creating a tool for the visual summarisation of large surveillance streams is defined.

The second part (Sections 4 and 5) describes R&D work that led to the design and development of VideoZoom. The key idea in VideoZoom is to create a pyramid of information layers on the image stream. From top to down, each layer reveals progressively more details about the image stream by a larger surface of summary images. The basis of the pyramid gives access to the images as taken by the camera. The reviewer can zoom through layers of the pyramid to get images' context and detail on demand.

The third part of the report (Sections 6 and 7) presents first experimental results obtained by using VideoZoom in image reviews run with DG ENER inspectors. A qualitative evaluation of VideoZoom as review tool is provided. Results indicate that the system allows accurate reviews, can save effort and is easy to learn and use. In addition the system allows detection of unexpected events which would be missed by standard review tools. The report concludes by outlining future work.

The work presented in this report is part of the JRC institutional project 'Safeguards Review Station' of the Action on Nuclear Facilities Verification (NUVER, nr. 53105).

1. Safeguards image reviews

Safeguards surveillance images are reviewed in batches of several thousands. In a batch, less than 0.01% of the total number of images is expected to relate to safeguards-relevant events. Because these events are to be detected and annotated by inspectors in *review reports*, there is a need for tools to focus the inspector's attention *directly* to the relevant parts of the image stream. For example, in a reactor, safeguards-relevant events include key steps for the processing of *flasks of spent fuel* (Figure 1, left) within a *Material Balance Area* (MBA) (Figure 1, right).

There are three locations of interest in a MBA:

- the hatch ('H')
- the decontamination area ('D')
- the pond ('P').

In a normal process a flask of fuel enters the hatch (HATCH ENTRY event, 1) and reaches the decontamination area (2), whence it is moved to the pond (POND ENTRY event, 3). From the pond (POND EXIT event, 4), the flask moves back to decontamination (5) and exits from the hatch (HATCH EXIT event, 6).
Apart from annotating these *regular and repetitive events*, any *anomalous* happening that may hint to the diversion of nuclear material is not to be missed by inspectors.

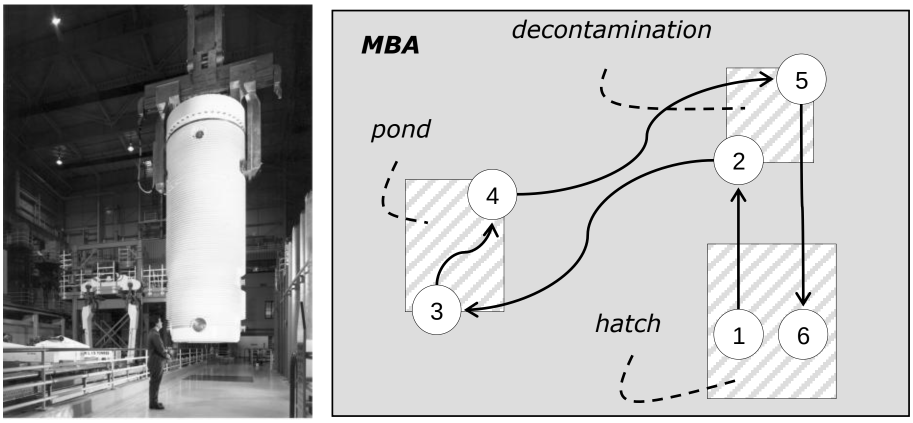


Figure 1 – Left: A flask of nuclear material (Photo credit: British Nuclear Fuels PLC). Right: Schematic movements of a flask of nuclear material in a MBA.

2. Techniques for image reviews

The current technique for reviewing safeguards surveillance images is *Scene Change Detection* (SCD) [1]. The technique in use is based on *two-frame differencing* of the average intensity value of pixels inside one or more *Areas of Interest* (AOIs). Before starting a review, inspectors draw AOIs on *reference images* of the MBA around locations of interest (e.g., hatch, decontamination and pond). Change values are computed for all images in the stream in AOIs: values breaking a threshold correspond to SCD events to be reviewed by inspectors.
In the last years, the JRC 'Safeguards Review Station' (SRS) project has proposed three techniques[1] based on *supervised pattern recognition* to be used in cascade to SCD [2][3][4][5][6][7][8]. The purpose of these techniques is to help detect regular and repetitive events taking place in an MBA *after* the SCD filter has been applied. The idea underlying these techniques is that *examples* of safeguards-relevant events annotated in archives of past review reports can be modelled and re-used as *filters* to assist inspectors in the detection of similar events in new batches of images. Exemplar images of past events act as 'teacher' and shape related event detection filters –hence the term supervised pattern recognition. By design, these

[1] These are Image Retrieval (IR), Decision Trees (DT) and Markov Models (MM).

filters are not expected to detect classes of events not seen before. They rather specialize in highlighting the 'regular' processing events, which need to be identified and annotated by inspectors as part of the review process.

3. Building summaries for image reviews

To complement previous work on pattern recognition, the SRS project is exploring ideas and techniques to build automatic *summaries* of surveillance streams *without using prior information on the expected relevance of events*.

Our task is as follows:

- To create a tool that builds automatically a *visual summary* of the content of a stream of time-stamped grayscale images taken by a camera with a fixed point of view over a scene.
- The purpose of the summary is to *guide* the review of the image stream and *reduce* the *number of images seen by the inspector* while performing the review work correctly.
- In building the summary, we *do not rely on prior information* about the visual appearance of relevant images. We assume that safeguards-relevant images produce a visual change in the scene, but not exclusively, i.e. change images can be also irrelevant from the safeguards view point.
- We assume the inspector to be knowledgeable about the nuclear processes she is reviewing. She will be able to use the review tool in an *active way*, for example, by browsing the summary forth and back as needed to search for 'logically expected events', or decide when to expand parts of the summary to the finest level of detail (the images as taken by the camera) to gain a full understanding of the processes that took place in the MBA. In other words, she will do all is necessary to understand the 'story'.

Note that a visual summary is *not* the review report itself. The report is compiled by the inspector annotating relevant images, and it is the end result of the review. By contrast, a visual summary is meant to present the inspector with the diversity of images in the sequential stream of data, minimizing redundancy to gain compactness. As such, a summary is expected to include frames that will be judged by the inspector as non-relevant. From a summarisation point of view, these frames provide continuity in visualizing the story of what happened in the MBA.

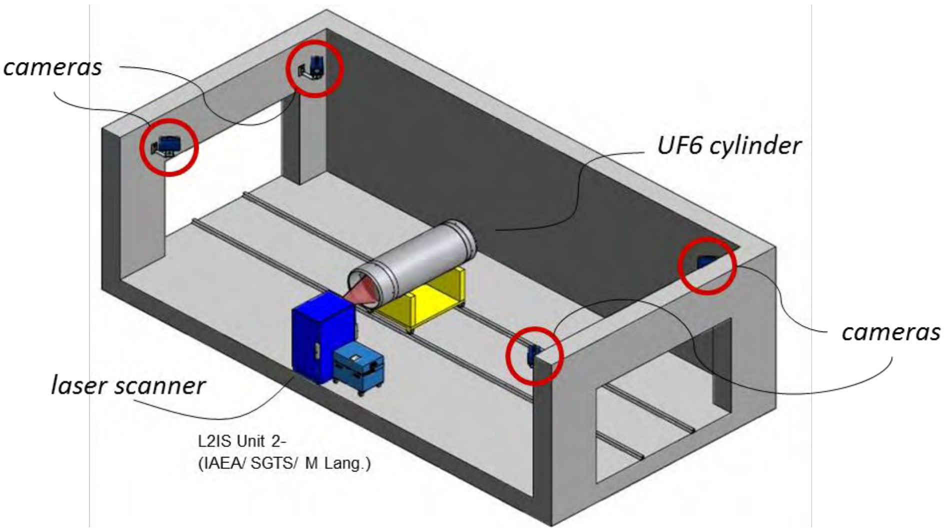


Figure 2 – Schematic view of the L2IS lab.

In what follows we describe *VideoZoom* [9], a review tool that builds automatic summaries out of large surveillance streams. To illustrate VideoZoom, we use a test surveillance stream acquired in JRC's L2IS lab (Laser Item Identification System). The laboratory hosts a mock-up of the JRC-developed system for fingerprinting (by laser scanning) UF6 cylinders in enrichment facilities

(Figure 2). Sequences of interest show a UF6 cylinder entering and exiting a process area on a trolley operated by a person (Figure 3). First qualitative experimental results on the effectiveness of VideoZoom as review tool will then be presented on a benchmark of safeguards surveillance images run with DG ENER inspectors.

Figure 3 – Images from the surveillance stream used to illustrate VideoZoom. Sequences relate to movements of a mock-up UF6 cylinder.

4. VideoZoom: A hierarchy of information layers on image streams

The key idea in VideoZoom is to create a *pyramid of information layers on the image stream* (Figure 4, top sketch). From the top to the bottom, each layer reveals progressively more details by a larger *surface of summary images*. The base of the pyramid gives access to the images as taken by the camera. The reviewer can *zoom* through layers of the pyramid to get images' context and detail on demand. Transition layers gradually blend the main layers together when zooming. On each main layer summary images are arranged on a *grid* chronologically ordered left-to-right and top-to-down (Figure 4, middle sketch) giving a view on the whole 'story'. Each summary image is built on a *grouping of images* from the surveillance stream describing a 'section of the story' (Figure 4, bottom sketch).
We now describe how image groupings are built and how image summaries are rendered at various levels of abstraction.

Image groupings

The grouping of images in a way that a reviewer can use them to navigate through the pyramid to detect events of interest is central to our approach. Any such grouping should show the information related to the images' content in a compact, readable way. The specific goal is to present large volumes of information in summarised form without missing events or objects whose relevance is not judged a priori. Approaches which remove frames or objects to form a compact representation of videos as in [10] are not applicable here.

The approach to grouping images to produce a clear and compact summary is now given. The main idea is that if a change occurs in the same location as a previous change but is different in some way, a new grouping should be started to avoid overwriting an event in that location. To this goal, *change blobs* in each image are detected and labelled by an ID. Blobs are then grouped together over time into *tubes* having a space and time coherence. Tubes are meant to represent the same object moving though time. If another change occurs in the same location as a tube and the time of the overlap is significantly different, then a new summary is created.

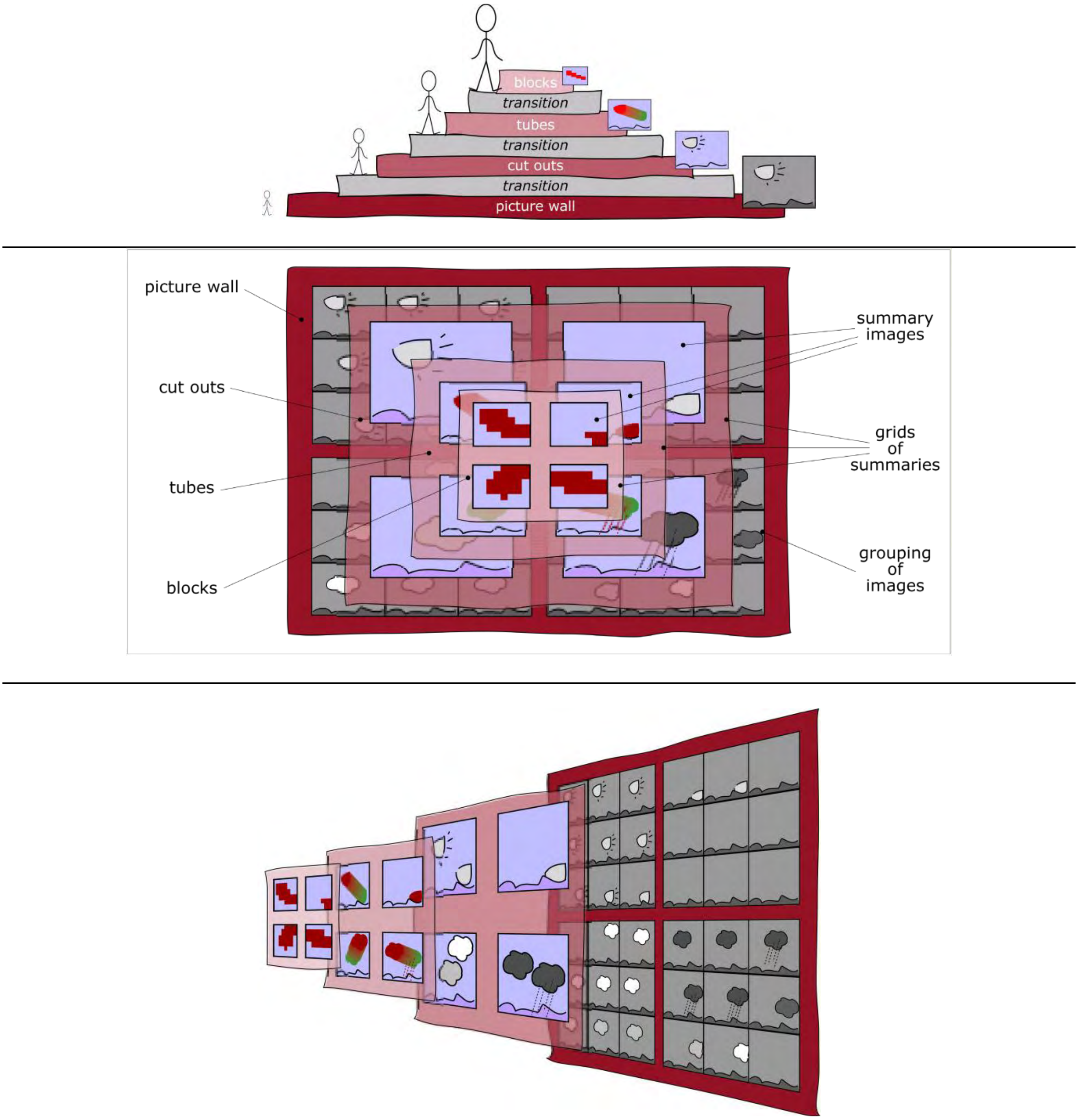


Figure 4 − Sketches on VideoZoom pyramid of information layers. Top sketch: Starting from the top and moving down, the pyramid gives access to larger surfaces of image summaries. Middle sketch: Image summaries are organized by layers in a temporally coherent grid, read left-to-right and top-to-down. Bottom sketch: View on image summaries provided by the different pyramid layers.

Change blobs. The current approach to change detection in safeguards is based on pre-defined areas of interest (AOI). In VideoZoom the change detection is performed by frame differencing over the entire image, we do not use AOIs. As the change detection process produces both negative and positive values, the absolute value is taken. Any change above a threshold of 20 units (the maximum possible change for grayscale images is 256 units) is counted as a change. This produces a binary black and white image where changes are given by a value of 1 and everything else is 0. Examples are given in Figure 5, middle part.

Tubes. For each binary image, the change regions are labelled with an ID. A region is defined as a set of connected, neighbouring change pixels (white pixels, as shown in Figure 5). Each region is recorded along with the time it occurred. These saved regions are called tubes. When change detection is performed on the next image in the sequence, it creates another change image. Each region on the new image is then compared to all the previous tubes. If a change on the new image overlaps an existing tube, the time the change occurred is compared. If the time difference is greater than a frame, a new summary image is started to show the change clearly. In this way, different events, such as a change of motion direction, are represented accurately and not compressed into a single summary image. On the contrary if the time difference is only a single frame, then the current change region is added to the existing tube because it likely refers to the same object moving in a continuous way. An example of tube generation is shown in Figure 5, bottom row.

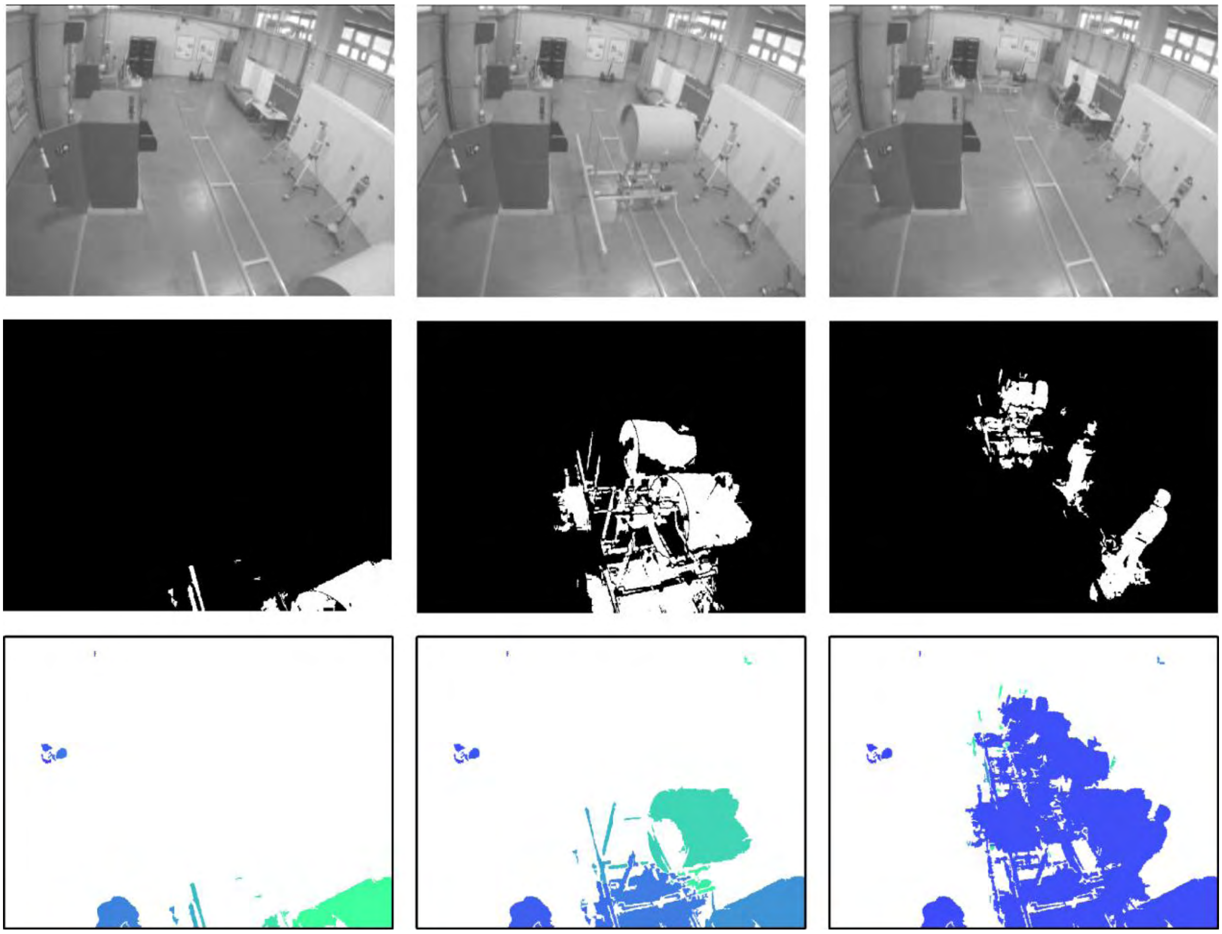


Figure 5 – Top: A sequence of frames. Middle: Change (white) and no-change regions. Bottom: Generation of a tube. Different colours refer to different tubes. Two tubes were merged in the second frame.

To visualise these images in VideoZoom different images are displayed on screen depending upon the zoom level and show different levels of detail. Here the construction of such images is explained.

Creation of summary images

Picture wall. The picture wall displays the images of a grouping in a grid arrangement without any changes to the original images. To construct the picture wall all the images which make up a grouping are first arranged by time left-to-right and top-to-down (Figure 6, left). The difference between an image and the next is then taken. If this difference is less than 5 pixels, the second image is removed from the picture wall, while the first is marked as a *compressed* image. This process is repeated for all the images in the grouping To show the viewer which images have been compressed, a blue border is drawn around them (Figure 6, right). In this way the reviewer is presented only with unique images and the picture wall is smaller and easier to understand.
The difference between a compressed and non-compressed picture wall is shown in Figure 6. One can see that in the compressed picture wall there is less visual information for the viewer to deal with. As we have only removed redundant images, the result is less overwhelming whilst

retaining all the information necessary to determine an image of interest. If one wants to see the hidden information then it is still possible to do so. This image gives access to detailed images as taken by the camera. Alternatively, it is also possible to play a video of the images composing the picture wall.

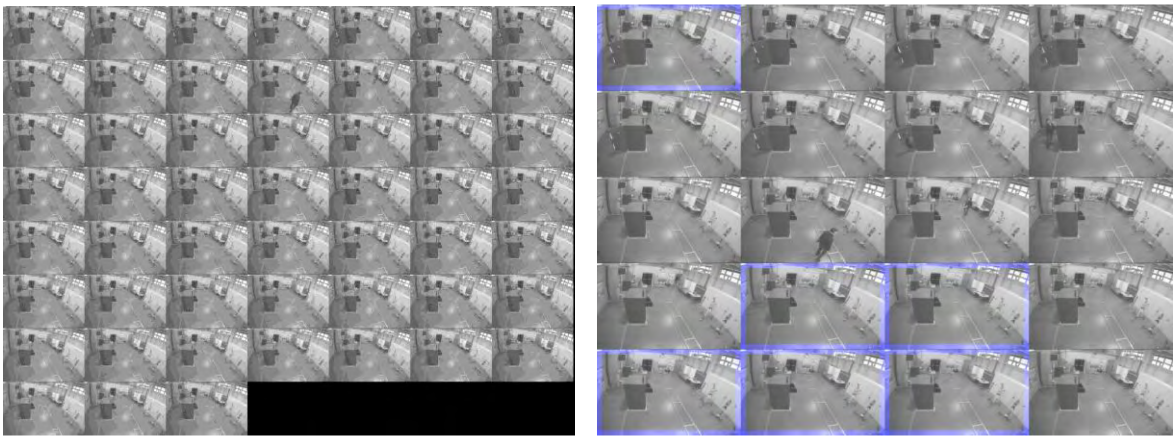


Figure 6 – Left: Initial picture wall on an image grouping. Right: Picture wall with compressed images framed by a blue border.

Cut outs. Cut outs show some of the original image photographic detail on a simplified background. The background is created by first applying a canny edge detector [11] to the initial image in the summary grouping. This produces a white image with all edges shown in black. To add some shading, and give depth information which edge detection alone may miss, the colours of the original image are quantised into 4 blue levels. The reason for the blue colour is that when dealing with greyscale images, as in safeguards, the cut out and the background appear very similar in colour, so the blue colour helps to differentiate them.
Once the abstract background is constructed, cut outs of the changes in the scene are then drawn on the background. Because there may be many cut out images, simply drawing them on top of one another would create a confusing scene. Here a heuristic method is used. The largest single image change region is selected. This original image information contained in this region is then drawn on the abstract background. The next largest region which does not overlap any of the cut outs already drawn is then selected. This process continues until it is not possible to draw any more cut outs on screen (Figure 7, left).

Tubes. An abstract background image is constructed as before. On top of this all the change regions are drawn on screen at once. To give an indication of the time a change occurred within the grouping, the time information contained in the image tubes is coloured. The colours start at red and go through yellow to green (Figure 7, middle image).

Blocks. A block image gives a very rough and coarse description which is only viewed when zoomed out. The image therefore appears very small on the screen. The image is divided up into a 10 by 10 grid. If a change has occurred in a block of the grid, then the block is coloured red; otherwise it is coloured grey (Figure 7, right).

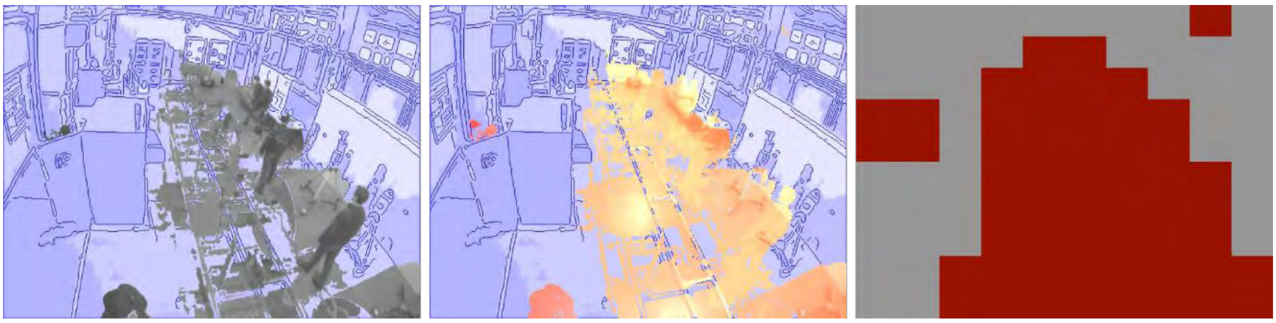


Figure 7 – Left: Cut outs. Middle: Tubes. Right: Blocks.

All together. Figure 8 shows part of the complete VideoZoom architecture for the L2IS sequence.

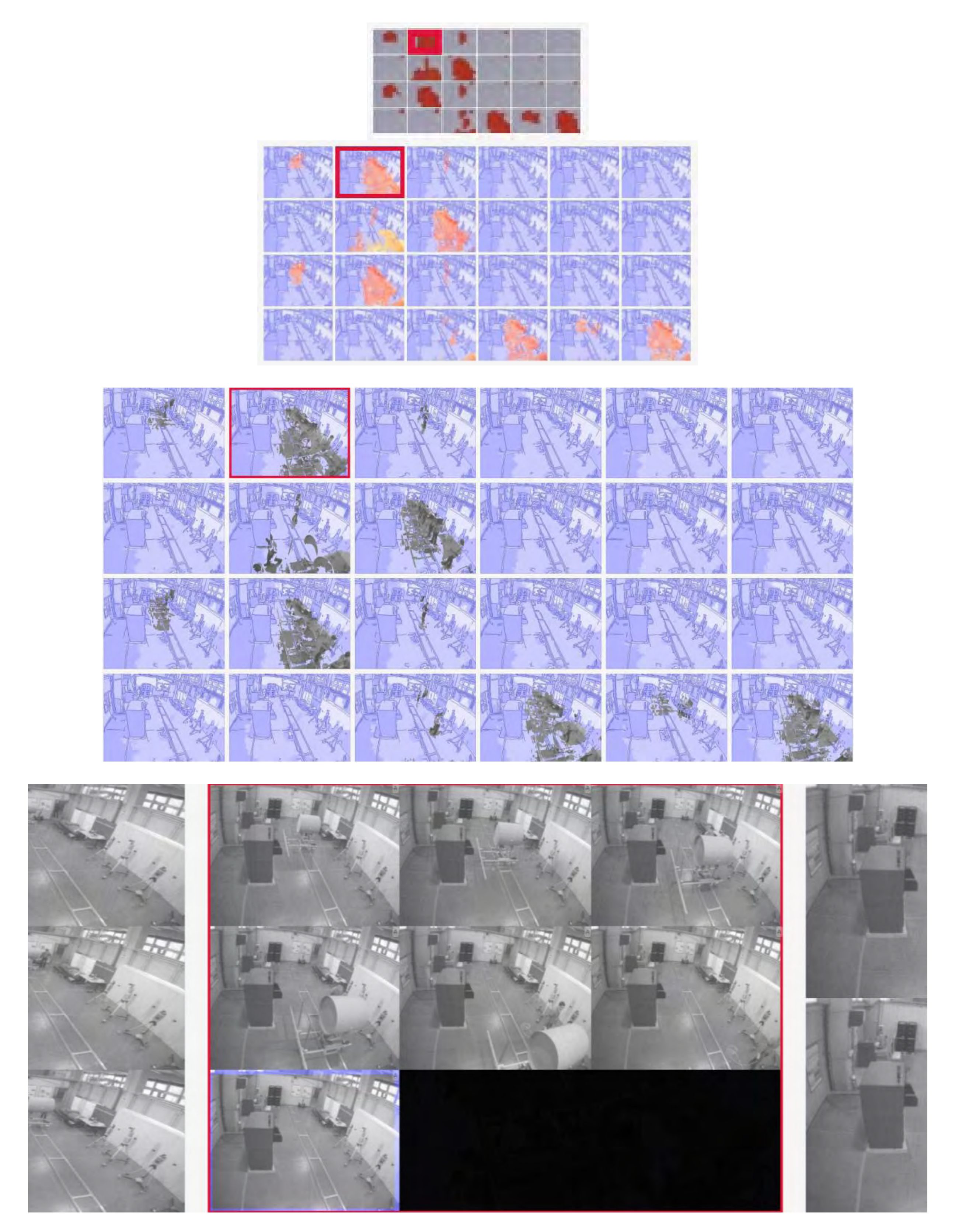


Figure 8 – VideoZoom layers illustrated on the L2IS sequence: blocks, tubes, cut outs, picture wall.

VideoZoom implementation

The implementation is divided up into two parts: preparing the images for use by VideoZoom and the VideoZoom review software itself.

Image processing

For processing the images MATLAB was used. A MATLAB script processes the images and sequentially computes the summary images. The end result of this process is a set of images at a fixed resolution (2400 x 1800 pixels) for every layer (blocks, tubes, cut outs and picture wall).
The amount of time this takes is dependent on the amount and type of changes in the video sequence. For the image sets used in the experiments of around 16,000 images, the process takes several hours[2]. The main constraint on the speed is the amount of processing required to match tubes to the changes in the current image which is exponential. Image sets with more changes or where the changes are fragmented require more computational effort to build them. In the theoretical worst case, if every pixel changed at every frame, the matching process used to build the image tubes would be exponential to the number of pixels. In practice this does not happen as there is temporal consistency in the underlying images. The result of this process is a set of summary images for all layers at a single scale (in this case 2400 x 1800 pixels). These images must be converted to a format that the visualization software can use.

Review structure

To cope with large image sets the images are divided up into screens. Each screen contains a set maximum number of summaries arranged in a grid (see Section 5). This arrangement stops the reviewer from getting overwhelmed by the sheer number of summaries on screen. This screen layout is reflected in how the summaries are stored on a computer. Each screen of summaries is stored in a directory along with an xml file containing the meta information of the summaries. The meta information contains time and date information along with a list of the images that were used to create a summary image. A single review xml file is used to keep track of how many screens there are in the complete review, the order they should be viewed in and where they are stored. This arrangement is shown in Figure 9.

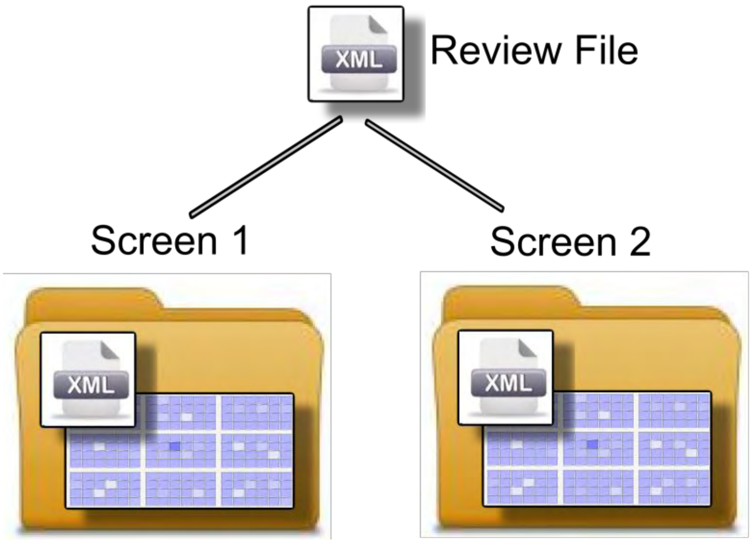


Figure 9 – Review structure. Each screen of images is stored in a folder along with the corresponding deep zoom xml. The Review file contains a list of these screens and the order to show them.

---

[2] The algorithm and the program code have not been optimized yet as the main purpose of this first prototype was to evaluate the concept of image summarization for image reviews.

To show the images on screen the DeepZoom format is used. We use Microsoft's dynamic linked library in a custom application to generate the images. For each image the DeepZoom library creates copies of the same image at smaller scale and stores them on disk. To create the layering effect, where different images are shown at different on-screen sizes, the appropriate image is scaled and stored at the desired size in the directory structure. For example the block images are shown at a small scale so only the small sized block images are stored. The picture wall is always shown when the on screen image is large and so only large size images of this layer are stored.

This whole process requires roughly 30% more storage space per image than the original. As an example an original image would take up roughly 1MB of space, the same image prepared in the DeepZoom format will take up 1.4MB.

It is not possible to calculate the storage space required by the sequence when it is fully summarised, processed and ready for use by the VideoZoom software before the algorithm is run. This is because the number of summary images that are generated depends upon the video content itself, not its size. However empirical results indicate that the storage required is roughly between the same size and 5 times as large as the original image sequence. The difference depends upon how well the original set is compressed. Sets with lots of different changes will create more blocks and thus need more storage. The barrel sequence shown throughout this report had an original image size of 145 MB (for 6945 jpeg images), the size required for the DeepZoom processed image set was 457 MB (based on 465 image summaries).

Microsoft's Silverlight version 4 was chosen as the platform for the system. The main reason for this was its ability to use the DeepZoom library for showing thousands of images in a zoomable interface. Other software is capable of doing this, but the result is not as fluid. Silverlight was also chosen for its ability to run in a web browser and use a client-server software model. All the images and data are located on a server and are retrieved by the application over a network. The application has no permanent local storage and web services, using the SOAP protocol, are used to save generated data (such as the review report) on the server. All the programming was done in C# using Microsoft Visual Studio 2010.

5. VideoZoom: Navigation, input and feedback to the reviewer

In this Section the use of VideoZoom as review tool is presented focusing on interaction modalities. Figure 10 sketches the elements of the VideoZoom interface described hereafter.

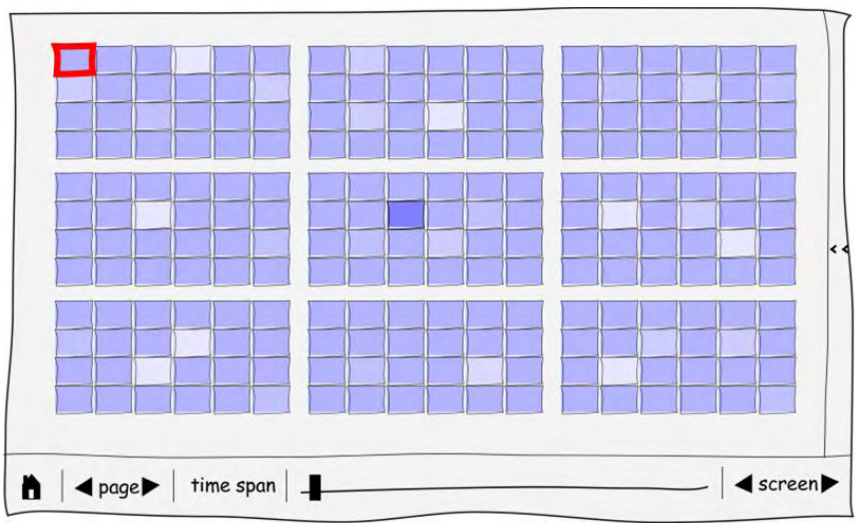


Figure 10 – Sketch of the VideoZoom interface.

Screens

VideoZoom displays summary images ordered by time on one or more screens (Figure 11). For a given image set, the number of screens required to show all summaries depends on the set size and its variability. A single screen can display up to 216 summary images.
The current screen number and the total number of screens for the image set at hand are shown in the bottom right corner of the VideoZoom window. To move between screens, the *change screen arrows* are pressed.

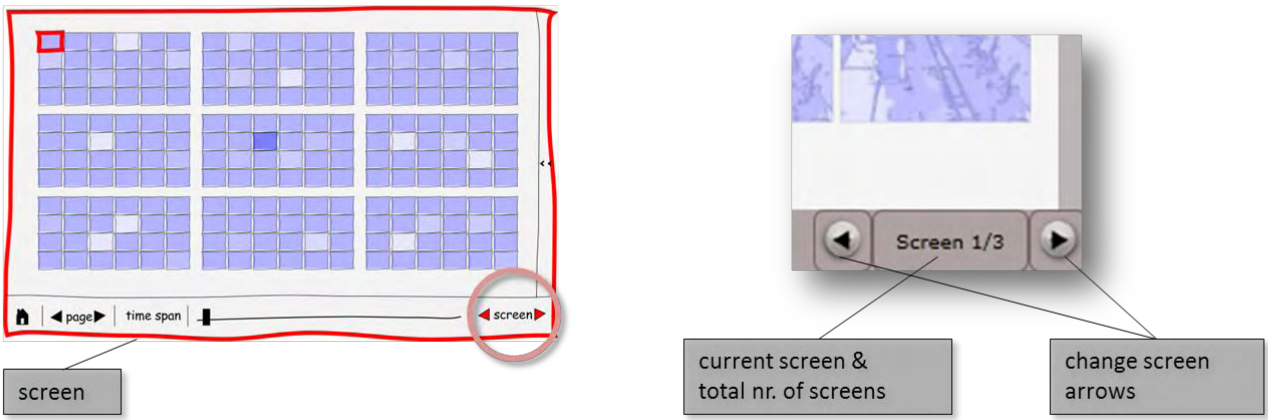


Figure 11 – Screens.

Pages

On a screen summary images are organized into 9 pages (Figure 12) comprising 24 summaries each. The current page is the one where the cursor is set (See 'Cursor' below).
Pages assist navigation: pressing the *change page arrows* centers on the screen the next/previous page maintaining the zooming level. The same effect is obtained by pressing keys *n* or *b*. Pages are useful to preview the content of a groups of summaries at the chosen level of detail.

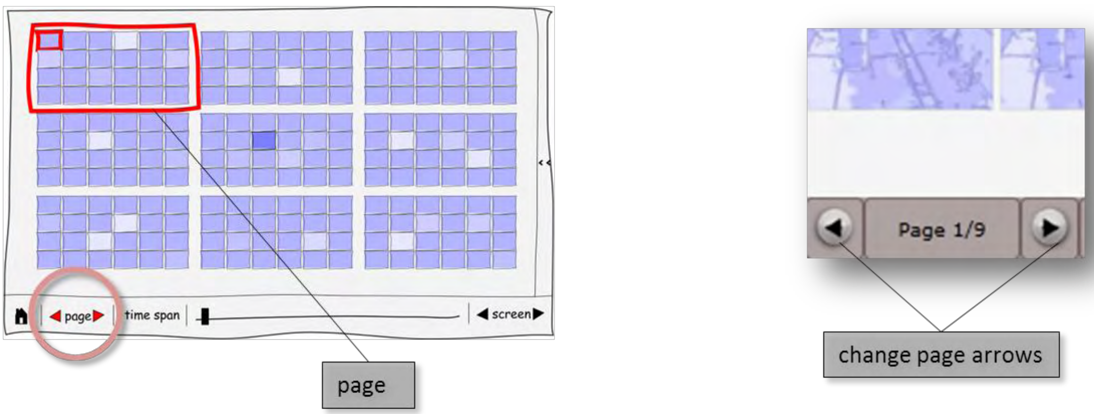


Figure 12 – Pages.

Cursor

The cursor is a red frame highlighting the summary image under review (Figure 13). The cursor position is set by *clicking* on summary images. It is moved to the left, right, top or bottom summary by using the *keyboard arrow keys*.
The cursor assists zooming on the summary image under review (Figure 14). Pressing *keys i/o* zooms-in/out the cursor image to show *directly* its blocks, tubes, cut-outs, and picture-wall. The *mouse wheel* is for *continuous* zooming. See Appendix 2 for an enlarged view of the zooming effect. The time span of this summary image is displayed on the VideoZoom bottom left area.

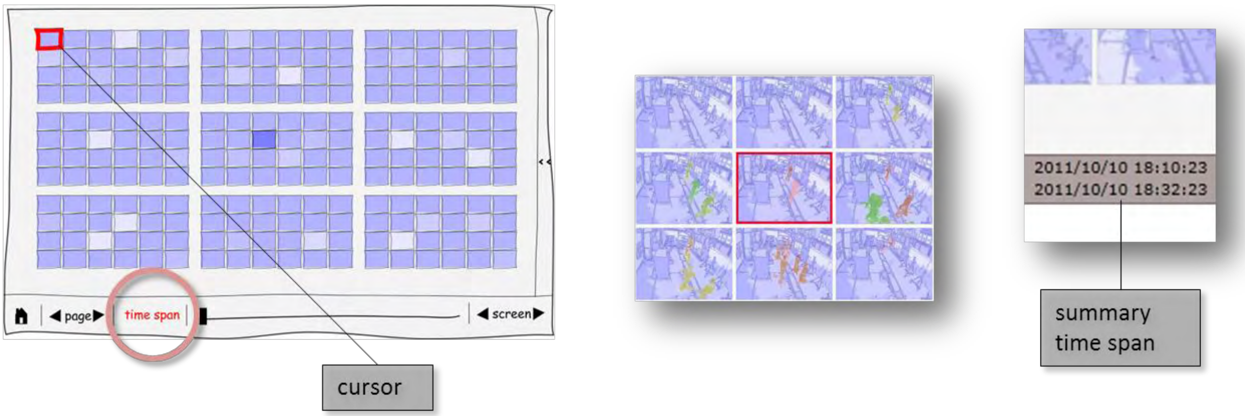


Figure 13 – Cursor.

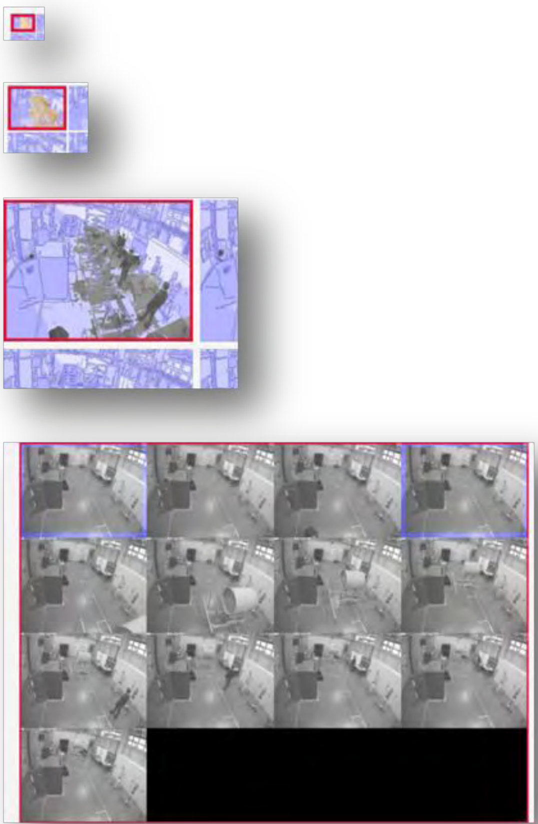


Figure 14 – Zooming-in on the cursor.

Time slider

To search images by time, the time slider is used (Figure 15).
The time slider shows the temporal position of the cursor within the screen. The start and end time for the current screen are indicated at the beginning and end of the slider line. *Dragging* the slider moves the cursor position accordingly when released.

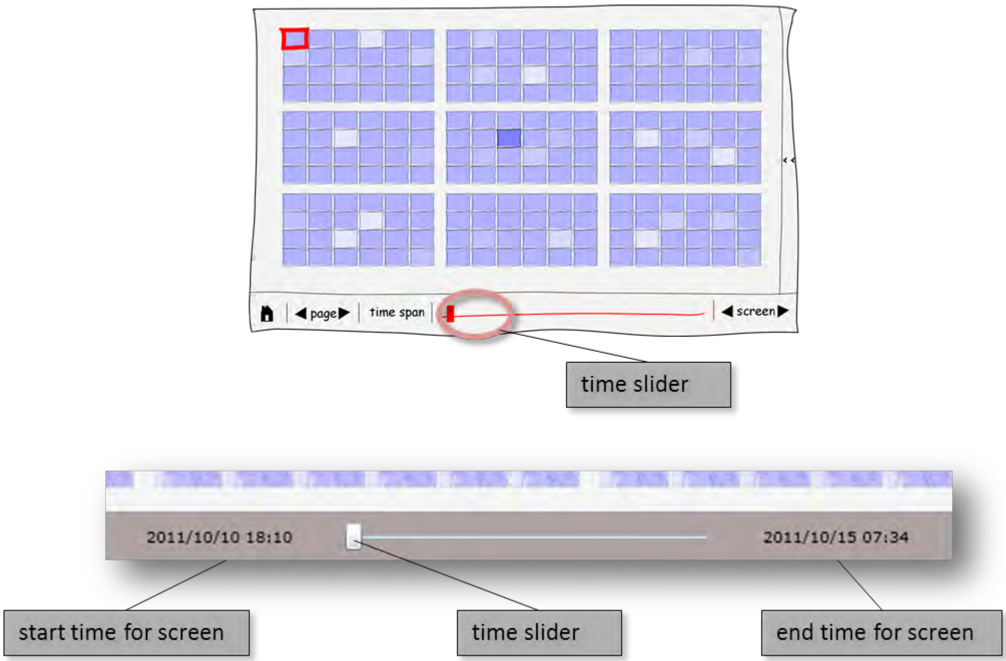


Figure 15 – Time slider.

Input and feedback information to the reviewer

The goal of reviews is to annotate safeguards-relevant images in a report in view of verifying its consistency with activities declared by the plant operator to the safeguards authority. Annotations make reference to specific images in the surveillance stream. For this reason, images in VideoZoom can be annotated by the reviewer only when at picture wall level.
Annotations are enabled by *buttons* appearing over the images of a grouping in focus (Figure 16, left). There is one annotation button per image. A button press pops-up the annotation window with a menu listing standard safeguards events' and the 'other' event (to capture non-standard events). Free text comment can be added as qualifiers to the events as required (Figure 16, right).

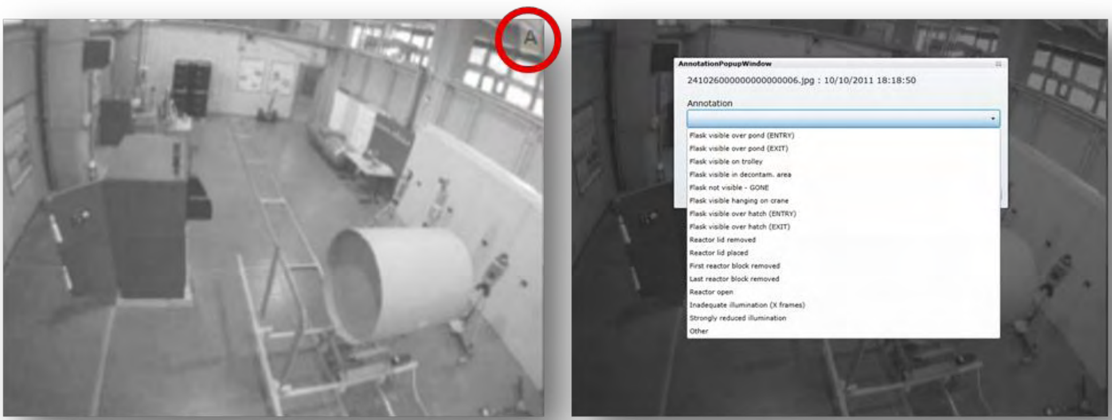


Figure 16 – Left: Annotation buttons. Right: Annotation window.

Annotated images (and related summary images) are shown on VideoZoom with a semi-transparent orange overlay (Figure 17) that can be turned on/off (*key c*). This is useful for a reviewer to check her work.

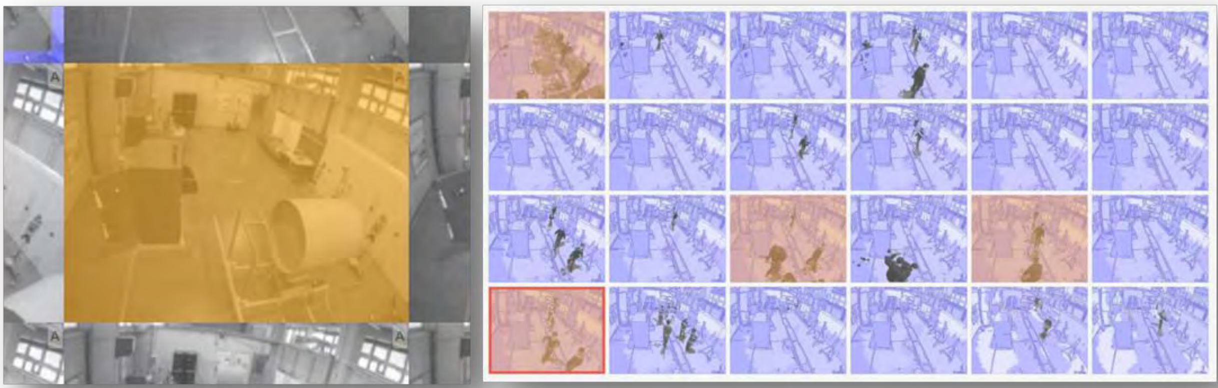


Figure 17 – Annotated images.

Annotated images become part of the *review report* (Figure 18) accessible by clicking on the tab marked **>>** on the VideoZoom window right edge. Annotations are sorted by the images' time stamps, so the reviewer can read the sequence of events and check it for 'logical consistency'. Entries in the report are mouse-sensitive: clicking on one entry shows on VideoZoom's window the part of the picture wall containing that event. In this way, one can review annotations or look for a missing event in a chain of logically incomplete flask processing.

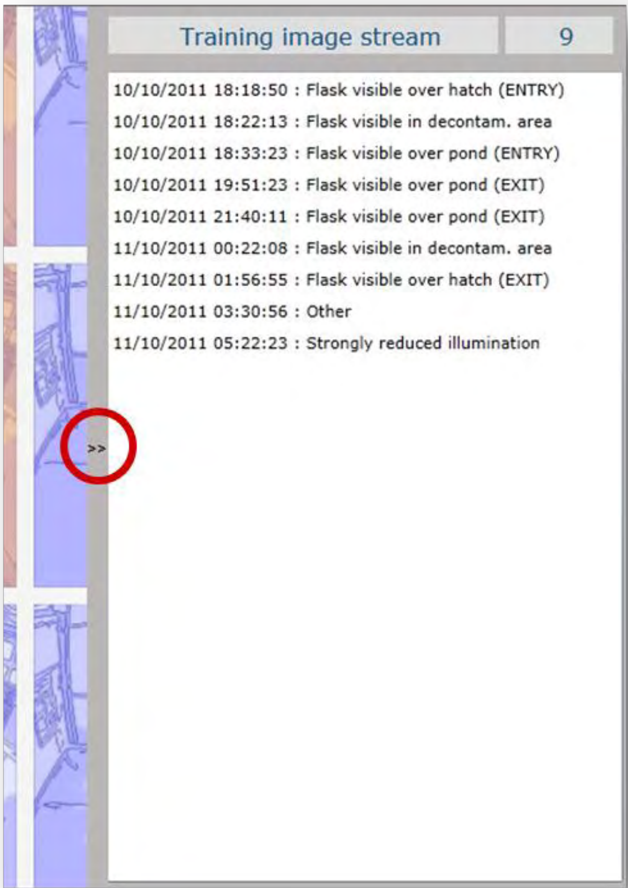


Figure 18 – Review report.

Finally it is important to provide the reviewer with information about images in the surveillance stream she viewed at a high-level of detail. VideoZoom keeps track of this information and displays it on demand by a semi-transparent white overlay on *visited groupings* (Figure 19). Type *key v* to activate/deactivate this control.

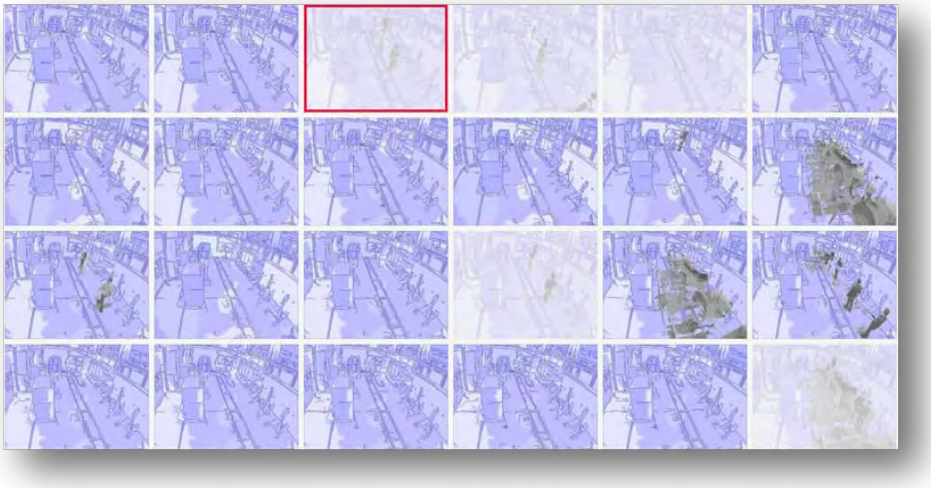


Figure 19 – Image groupings visited at a high level of detail.

The last command of the VideoZoom interface is the *home button* provided on the bottom left area of the interface to easy the re-centering of the screen if the reviewer get lost on the VZ window. When pressed, the home button displays the current screen centered at tubes level.

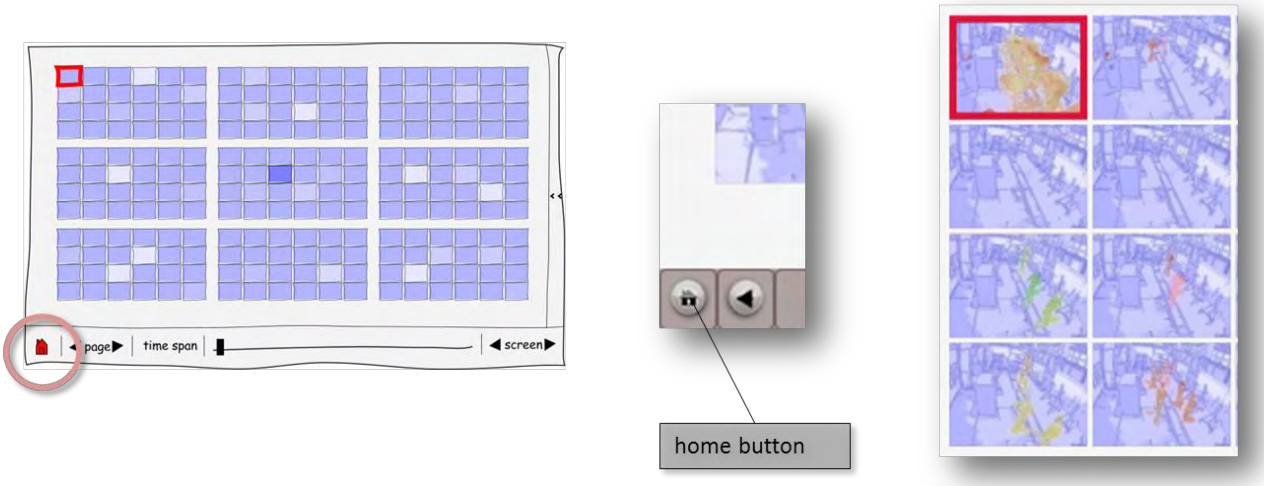


Figure 20 – Home view.

Summary of VideoZoom interface commands

Table 1 lists commands available from the VideoZoom interface to perform an image review.

Effect	Command
Screen change	Press *change screen arrows*, bottom right on the VZ window
Page change	Press *change page arrows*, bottom left on the VZ window
	Key b for page before *Key n* for next page
Move between summary layers	1. *<mouse-left>* on a summary image to set the cursor position 2. *<mouse-wheel>* for continuous zooming-in/out
	1. *<mouse-left>* on a summary image to set the cursor position 2. *Key i* or *o* or for direct zooming-in/out to blocks, tubes, cut-outs, picture wall
Move over a summary layer	*<mouse-left> + drag*
	1. *<mouse-left>* on a summary image to set the cursor position 2. *<arrow-keys>*
Search images by time	*Drag* the *time slider*, bottom center on the VZ window
Annotate an image	Press the *Annotation button* visible at picture wall level
Annotated images show/hide	*Key c*
Review report show/hide	Press *buttons <</>>* on VZ window right edge
Visited summaries show/hide	*Key v*
Home view	Press *Home button*, bottom left on VZ window

Table 1 – VideoZoom interface commands.

6. Qualitative evaluation of VideoZoom by safeguards inspectors

Previously evaluation work was completed with volunteers from the JRC's Ispra site (November 2010). Based on the findings of these evaluations and a demonstration of Videozoom to DG ENER inspectors (February 2011), the software was modified resulting in its current state as described in the previous Sections.

To investigate how expert reviewers would use VideoZoom a preliminary evaluation was conducted at DG ENER in November 2011. The purpose of this evaluation was to collect qualitative information on the software with a small set of users who have expert knowledge of reviewing image sequences for the detection of safeguards-relevant events.

Participants were given some training in how to use the system [12]. This training consisted of introducing the tool, the concept behind it and how to use it. Participants were then given some time to use the tool on a training set (the L2IS sequence). After they could use all functions of the software competently, some additional training introducing the review work itself was given. This included a detailed description of the layout and processes in the MBAs to be reviewed. The presentation of the training materials took between 1 and 2 hours depending on the reviewer.

Image sets

Table 2 provides information about the image sets used for testing. Sets A1, A2 and A3 stem from MBA *A,* B1 and B2 stem from MBA *B.* Each image set spans over several months of plant activity. For each set, the number of target events to identify is shown in the Table. The number of events is very limited compared to the total number of images in each set. MBA B is more complicated in that two flasks can be in the decontamination area at the same time. Throughout all sets is not unusual to have image sets where no or little safeguards-relevant activity needs to be reported by inspectors, as in A2 and A4.

Set A4 was deliberately tampered with. A flask was taken from another image set and was pasted in a position which was not expected. It was placed outside the hatch and pond area. Importantly this flask was artificially placed in an area which was not typically covered by areas of interest which are used in the safeguards official review tool [1]. This is significant as a review performed with the official tool would have missed the event. The purpose of this setup was to find out if reviewers could use VideoZoom to find events which are not expected. Reviewers were not told of this tampering beforehand.

Image set	Nr. of images	Hatch events	Pond events	Anomalies
A1	20160	17	17	0
A2	15661	1	1	0
A4	16022	0	0	1
B1	16020	30	32	0
B2	15446	11	12	0

Table 2 – Image sets used in the benchmark. For each set, the Table lists the number of images and the number of events to be detected over the hatch (H) and pond (P) areas.

The task

Three expert reviewers were asked to review several image sequences using the VideoZoom tool. One reviewer completed all the sets whilst the other two reviewed sets A1 and A4 due to time availability. All participants were also asked to fill out a questionnaire about using VideoZoom. The questionnaire was based on the standard systems usability scale [13] along with some open ended questions. Answers to the questionnaire are reported in Appendix 3.

The VideoZoom computer program records how a reviewer uses the system as they complete their review. The program records events related to timing, key presses and mouse movements. All reviewers were told that the program would collect this data prior to using it.

Seen images

To compare VideoZoom with GARS a count of the number of images a reviewer sees was made. A 'seen' image was counted if a new image was requested by a reviewer using one of the navigation commands. Transient images were not counted.

When a reviewer presses a key on the keyboard that will cause the image as seen on screen to change, the system will record that a new image has been seen. An example of this is when the reviewer presses the keyboard to move to the next page (key 'n'). This will cause the next page to be centered in the screen, and is counted as a new image.

When navigating using the mouse, a new image was counted as being seen when the reviewer stopped using the mouse for longer than one second. When a reviewer navigates using the mouse wheel or by dragging the mouse this causes the event to be recorded along with the corresponding time. If this difference is less than a second between events, then the reviewer is using the mouse to move (e.g., dragging or zooming) and has not stopped at their intended destination. An example of this is when the mouse wheel is used to zoom in. Typically moving the mouse wheel with a finger generates around 5-10 events but it is really only one movement. Such an event would generate many transient images which are only displayed on screen for a fraction of a second. Transient images are generated to give the impression of smooth movement but are not themselves the destination image. All such mouse wheel events would be grouped into a single unique event. Only when the event has stopped for longer than one second is the screen counted as being 'seen' by the reviewer.

In addition we also count when the new screen button or the home button is pressed as a unique screen.

These navigation events are all summed to give the number of images seen on screen by the reviewer.

Experimental results

Table 3 summarizes results obtained in running test reviews. For each image set the table shows the number of summaries generated by VideoZoom (VZ summaries) and the number of SCD events generated by running the official review software on AOIs optimized for the detection of Hatch and Pond events. These numbers give an indication on the *a priori* review effort one can expect using the two review systems, assuming the inspector will devote some attention to all VideoZoom summaries and to all SCD events. The actual review effort is then measured per image set and per reviewer (X,Y,Z) in terms of review time and images seen. Further Table 3 presents per percentage of mouse versus keyboard commands used to perform a review. This is to understand how the system was used to run a review.

Image set	VZ summaries	SCD events	Reviewer	Review time	Images seen	Mouse %	Keyboard %
A1	292	577	X	39:19	477	46	54
			Y	49:10	679	100	0
			Z	62:13	563	34	66
A2	85	179	X	10:43	232	1	99
A4	450	701	X	12:22	221	1	99
			Y	54:42	735	87	13
			Z	94:53	828	20	80
B1	418	986	X	54:24	628	1	99
B2	233	471	X	27:41	272	4	96

Table 3 – Results of the evaluation.

For set A1 all reviewers were able to successfully and accurately complete the review. The average time to complete the review was around 50 minutes. Two reviewers (X and Z) made use of the mouse and keyboard with a roughly 50/50 split. The third reviewer (Y) however did not use the keyboard at all.

After set A1 all reviewers were asked to review A4, the set containing the sequence of images artificially altered by placing a nuclear flask in a different position from what was to be expected. The anomalous event was annotated by 2 out of the 3 reviewers. The third reviewer (X) did examine the concerned summary image in detail at picture wall level, but did not make any annotation on it. This suggests the summary image enabled a reviewer to find this anomalous event but did not annotate it. The reason for not annotating the image is unclear. When conducting this second review all participants used the keyboard to perform a higher percentage of navigation tasks. Two of the reviewers specifically commented that once they knew the commands it was easier and quicker to use the keyboard. The reviewer who previously never used the keyboard did now make use of it.

Only one reviewer (X) completed a review of sets A2, B1 and B2. This was due to external time pressures of the other reviewers making them unavailable. It is evident in these reviews that reviewer X was mainly using the keyboard to perform the image reviews.

VideoZoom layers used

All reviewers spent the majority of their time in the picture wall layer. However they did spend time in both the tube and cut out layer. A detailed breakdown of this is given in Figure 21 for set A4. This suggests that the reviewers were making use of the tube and cut out layers to guide themselves to the images of interest. Once at the picture wall level the reviewers would take their time to analyse the original images in detail. Looking at the results it is clear that reviewer Z who spent most of their time in the picture wall layer saw the most images and took the most time (see Table 3) for this set. Reviewer Z took almost double the time of reviewer Y for the same review, and saw many more images. When a reviewer makes use of other layers to guide them to interesting images a large amount of time and effort can be saved.

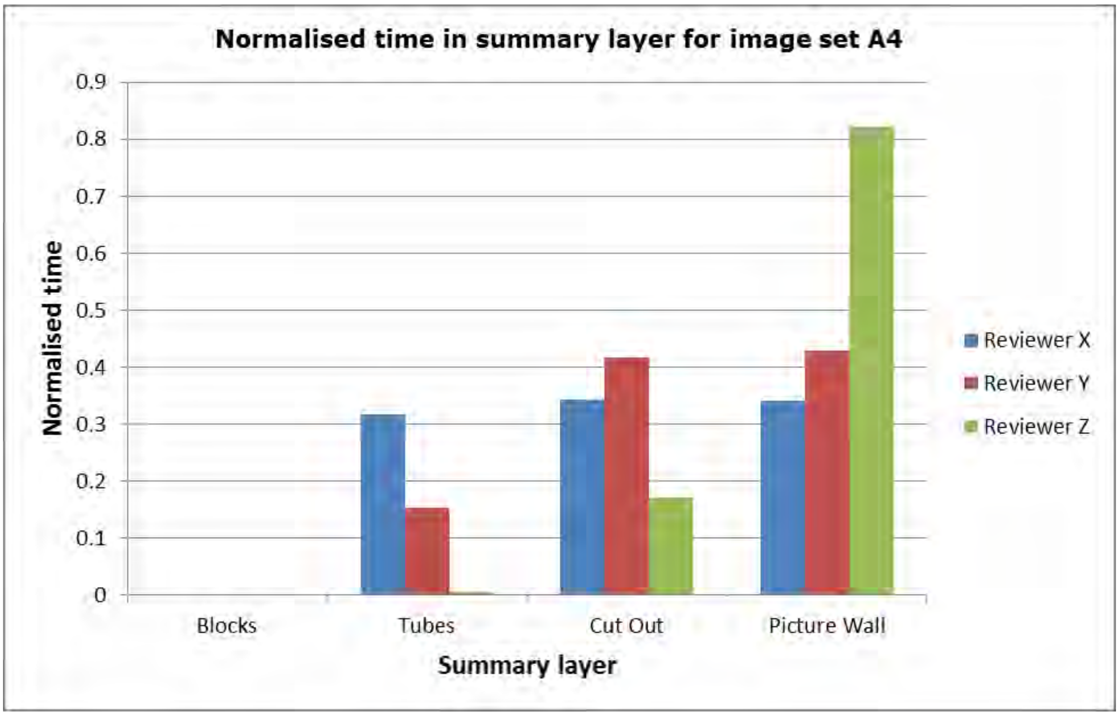


Figure 21 – Time spent in each layer normalised by time for all reviewers for set A4. It is possible to see the different review styles.

Qualitative evaluation of the VideoZoom user interface

After the review work was completed participants were asked to complete the survey (see Appendix 3). The responses show that the reviewers thought that the system was easy to use and that they did not have any major difficulties performing the review task. All reviewers said they felt comfortable using the system to perform the review work. During the review work there were only minor questions or queries about how to use the interface itself. The majority of questions were related to the review work itself mainly concerning precisely what image to annotate (it is possible to annotate one of several images next to each other in time and still be correct).

7. Discussion and future work

Because VideoZoom is self-guided there is a large variability in how people use it. Some reviewers looked at roughly four times as many images as others (e.g., for set A4, compare reviewer Z to X and Y) and took significantly longer to perform the same review. This suggests that by using the cut out and tube layers to guide the review can reduce the amount of work a reviewer has to do whilst still performing the review work correctly. *A suggested working method must be devised and taught to the reviewers before they start the work*. It is also noted that for the review work carried out in this evaluation the block layer was not used.

One of the key advantages of using the VideoZoom system is that the reviewer can see summaries of image changes irrespectively of where they occur on the image plane. Using AOIs as required by the official review system would have meant that the unusual event in image set A4 would never have been presented to an inspector and thus would have been missed. Using the VideoZoom system it is possible to detect this unusual event.

A combined approach could be designed where both review strategies coexist. AOI-based SCD events could be highlighted on the VideoZoom summaries on the inspector's demand.

Reviewer suggestions

Several suggestions were given by the reviewers themselves. The majority of these suggestions concerned the review report and how it was presented. This gives an indication that the reviewers see the report as being extremely important in any such reviewing system. It was also noted that time information was given a high priority by the reviewers. Detailed points made by the reviewers concerning the review report are:

- The MBA name and the time span of the entire image set should be visible in the review report.
- It should be possible to systematically work through events listed in the review report in an easy way.
- The review report window should show the additional annotation text entered in the annotation window (Figure 21, right) either by using multiple lines or by making the report window bigger. At present annotations in VideoZoom use a fixed lexicon, accessed from a pre-specified drop down list. However when a real review is performed inspectors modify these standard words and phrases to include more detail. This happens frequently and so the VideoZoom system should be modified to allow this.

Other comments were that it should be possible to search through the whole review by time and/or be possible go directly to a time range of interest.

There were also some issues which were not directly mentioned but came to light during the evaluation. One recurring theme was that the checking of the report should be improved to make it easier to check the logical process. An example of this is if a flask goes into the pond it must come out before another flask can go in. This type of scenario requires the reviewer to keep track of this process. Reviewers commented that there was no support for doing this in the current system.

Finally a suggestion was made that a video play of summaries is enabled by the interface. This option has already been implemented in an updated VideoZoom prototype but was not used in the evaluation. This was demonstrated to the inspectors receiving positive feedback. How to annotate images when showing a video will form part of future work.

Extended evaluation

The initial evaluation with safeguards inspectors indicates that the VideoZoom can be used to perform successful reviews on image sequences of nuclear facilities. However the number of participants was too small and to draw statistically meaningful conclusions. It is necessary to include more people in the evaluation to do so. The next step is to conduct an extended evaluation with more inspectors. An extended evaluation is needed to determine if the system is useful and easy enough to learn and that it could be accepted by inspectors.

Engineered approach

Currently work has focused exclusively on testing the feasibility of using image summarisation in a zooming interface to perform safeguards image reviews. However the time required to process the images from their original form to one that can be used in VideoZoom is prohibitively high for use in a real system. To make the transition from the prototype system as proposed here, to a real world setting, the computation times must be substantially reduced. In addition there would need to be an easy way to process image sets without specialist knowledge. An example of this would be a piece of software that takes the image/video files and computes everything necessary to begin a review in VideoZoom with a simple button press. Both of these steps would need to be taken in order to make the tool acceptable in the safeguards working environment.

Acknowledgement

The authors thank safeguards reviewers at DG-ENER for their VideoZoom evaluation work: João Rocha, Juha Pekkarinen, Victor Proenca Pascoa.
Thanks also to the JRC colleagues who evaluated earlier versions of VideoZoom: Patrice Richir, Pierluigi Taddei, Teofilo Molto.

References

[1] GARS http://www.canberra.com/pdf/Products/Systems_pdf/GARS-SS.pdf

[2] Stringa E., Versino C., Gonçalves J.G.M., Heppleston M., Tourin, L. (2004) – *Image Review Station: Assessment of Image Retrieval and Classification Technologies on Safeguards Surveillance Streams*. Proc. of the Institute of Nuclear Materials Management INMM04.

[3] Versino C., Stringa E., Gonçalves J.G.M., Heppleston M., Tourin, L. (2005) – *Towards the Integrated Review of Safeguards Surveillance Data*. Proc. 27th ESARDA Symposium on Safeguards and Nuclear Material Management.

[4] Lombardi P., Versino C., Gonçalves J.G.M., Heppleston M., Tourin L. (2006) – *Safeguards Review Station: Tools that learn from inspectors' review reports*. Proc. of the Institute of Nuclear Materials Management INMM06.

[5] Lombardi P., Versino C., Gonçalves J.G.M., Heppleston M., Tourin L., (2007) – *Safeguards Review Station: Embedding hidden Markov models in an image review tool*, Proc. of the Institute of Nuclear Materials Management INMM07.

[6] Lombardi P., Versino C., Gonçalves J.G.M., Heppleston M., Tourin L., (2008) – *MM: The Markov Model Tool for Image Reviews*. Proc. of the Institute of Nuclear Materials Management INMM08.

[7] Lombardi P., Versino C. (2011) – *Learning to Detect Event Sequences in Surveillance Streams at Very Low Frame Rate.* In: Wang L, Zhao G, Cheng L, Pietikäine M, editors. Machine Learning for Vision-Based Motion Analysis: Theory and Techniques. London (United Kingdom): Springer-Verlag; p. 27.

[8] Versino C., Lombardi P. (2011) *Filtering Surveillance Image Streams by Interactive Machine Learning.* In: Lin, W.; Tao, D.; Kacprzyk, J.; Li, Z.; Izquierdo, E.; Wang, H. (Eds.), editors. Multimedia Analysis, Processing and Communications. Springer; p. 36.

[9] Blunsden S., Versino C. (2009) – *Interactively reviewing large image sets*. Poster presentation in: NIPS 09 Workshop: Analysis and Design of Algorithms for Interactive Machine Learning, 07 November 2009, Vancouver (B.C. (Canada.)). JRC55219.

[10] Rav-Acha A., Pritch Y., Peleg, S. (2006) – *Making a Long Video Short: Dynamic Video Synopsis.* CVPR, 435--44.

[11] Canny J. (1986) – *A Computational Approach To Edge Detection*. IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698.

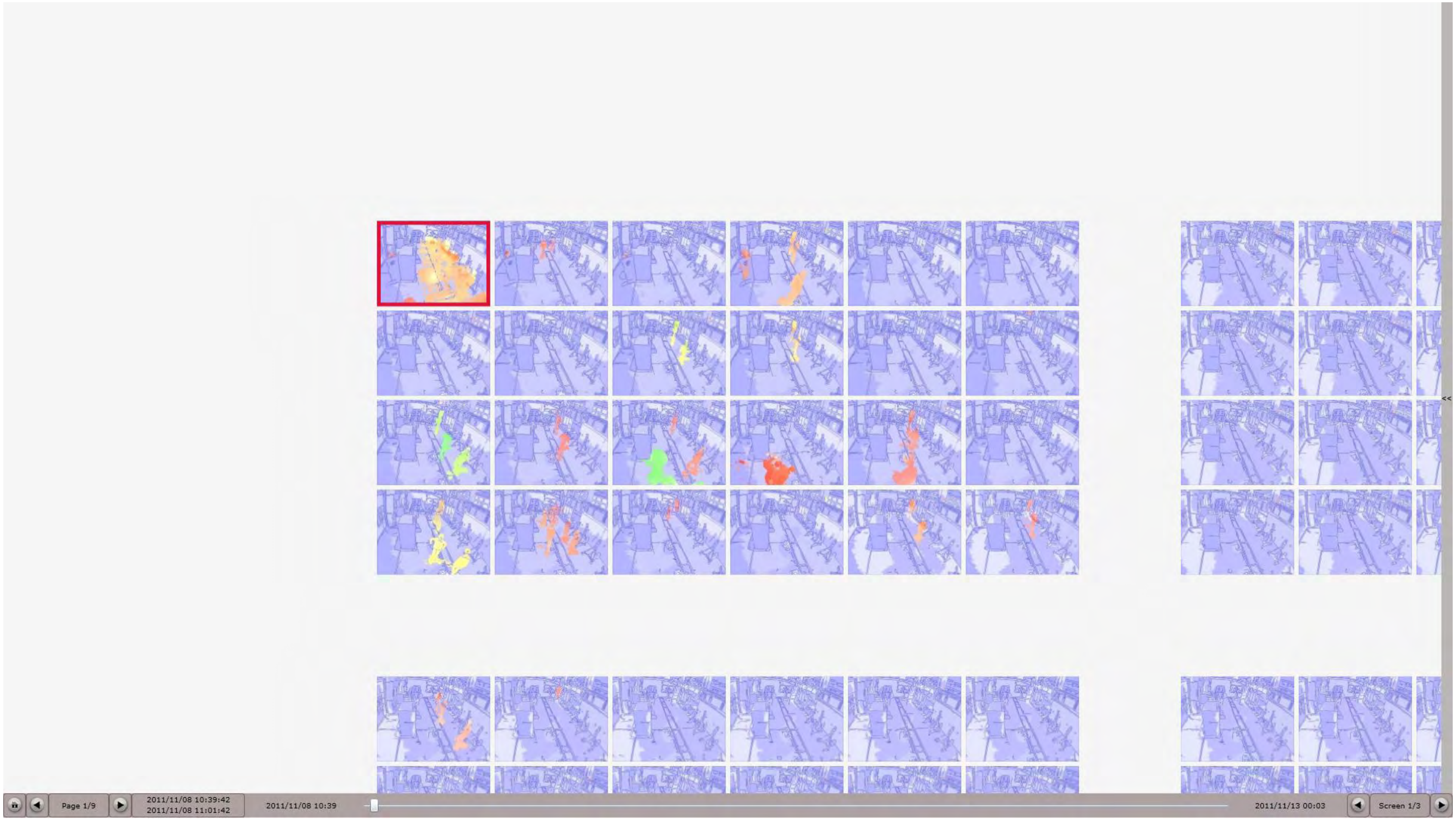
[12] Blunsden S., Versino C. (2011) – *(1) VideoZoom: A prototype tool for image reviews. (2) VideoZoom manual. (3) Evaluation of VideoZoom in the review of safeguards surveillance images.* Presentation slides.

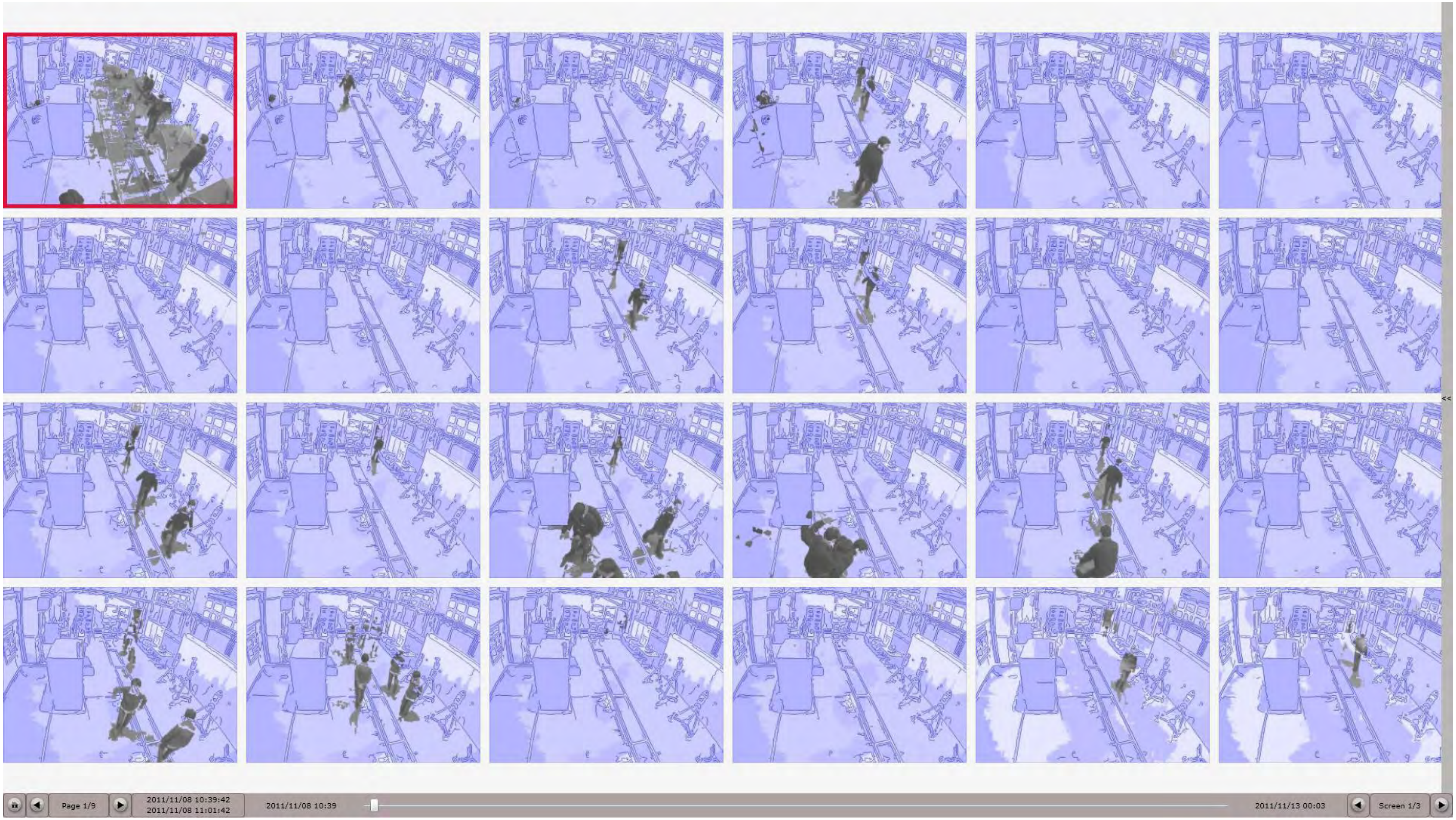
[13] Brooke J. (1996) – *SUS: a 'quick and dirty' usability scale.* In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland. Usability Evaluation in Industry. London: Taylor and Francis.

Appendix 1: Acronyms

- **AOI**: Area of Interest
- **D**: Decontamination area
- **DT**: Decision Trees
- **IR**: Image Retrieval
- **H**: Hatch area
- **MBA**: Material Balance Area
- **MM**: Markov Models
- **P**: Pond area
- **SCD**: Scene Change Detection
- **SRS**: Safeguards Review Station
- **VZ**: VideoZoom


Page 1/9
2011/11/08 10:39:42
2011/11/08 11:01:42
2011/11/08 10:39
2011/11/13 00:03
Screen 1/3


Page 1/9
2011/11/08 10:39:42
2011/11/08 11:01:42
2011/11/08 10:39
2011/11/13 00:03
Screen 1/3



Page 1/9
2011/11/08 10:39:42
2011/11/08 11:01:42
2011/11/08 10:39
2011/11/13 00:03
Screen 1/3


Page 1/9
2011/11/08 10:39:42
2011/11/08 11:01:42
2011/11/08 10:39
2011/11/13 00:03
Screen 1/3

Appendix 3: VideoZoom evaluation survey

Video Zoom Survey

Please tick in the appropriate box for each of the following questions:

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently				X	
2. I found the system unnecessarily complex		X			
3. I thought the system was easy to use				X	
4. I think that I would need the support of a technical person to be able to use this system	X				
5. I found the various functions in this system were well integrated			X		
6. I thought there was too much inconsistency in this system		X			
7. I would imagine that most people would learn to use this system very quickly					X
8. I found the system very awkward to use	X				
9. I felt very confident using the system				X	
10. I needed to learn a lot of things before I could get going with this system	X				

How easy is it to find the images you were looking for using the tool provided?

Not at all easy				X		Very easy

Compared to what you expected, how quickly did the tasks go?

Not at all quickly			X			Very quickly

How confident are you that you found all the relevant images you were looking for?

Not at all confident			X			Very confident

What works well in video zoom:

What doesn't work well in video zoom:

What is missing in video zoom:

How does video zoom compare to GARS:

SEEMS QUICKER

Any other comments:

Video Zoom Survey

Please tick in the appropriate box for each of the following questions:

	Strongly disagree 1	2	3	4	Strongly agree 5
1. I think that I would like to use this system frequently					X
2. I found the system unnecessarily complex	X				
3. I thought the system was easy to use				X	
4. I think that I would need the support of a technical person to be able to use this system	X				
5. I found the various functions in this system were well integrated				X	
6. I thought there was too much inconsistency in this system		X			
7. I would imagine that most people would learn to use this system very quickly				X	
8. I found the system very awkward to use	X				
9. I felt very confident using the system				X	
10. I needed to learn a lot of things before I could get going with this system		X			

How easy is it to find the images you were looking for using the tool provided?

Not at all easy				X		Very easy

Compared to what you expected, how quickly did the tasks go?

Not at all quickly			X			Very quickly

How confident are you that you found all the relevant images you were looking for?

Not at all confident				X		Very confident

What works well in video zoom:

Very userfriendly tool.
Should be usefull tool when the data process will be
be improved.

What doesn't work well in video zoom:

Report has to be improved.

What is missing in video zoom:

How does video zoom compare to GARS:

New approch for reviewing.
The user is directly informed of the movements in the considered area.

Any other comments:

Waiting for the next step.

Video Zoom Survey

Please tick in the appropriate box for each of the following questions:

	1 (Strongly disagree)	2	3	4	5 (Strongly agree)
1. I think that I would like to use this system frequently				X	
2. I found the system unnecessarily complex		X			
3. I thought the system was easy to use				X	
4. I think that I would need the support of a technical person to be able to use this system	X				
5. I found the various functions in this system were well integrated			X		
6. I thought there was too much inconsistency in this system		X			
7. I would imagine that most people would learn to use this system very quickly				X	
8. I found the system very awkward to use	X				
9. I felt very confident using the system			X		
10. I needed to learn a lot of things before I could get going with this system		X			

How easy is it to find the images you were looking for using the tool provided?

Not at all easy				X		Very easy

Compared to what you expected, how quickly did the tasks go?

Not at all quickly			X			Very quickly

How confident are you that you found all the relevant images you were looking for?

Not at all confident				X		Very confident

What works well in video zoom:

What doesn't work well in video zoom:

SPACE OF IMPROVEMENT IN ANNOTATION PART.

What is missing in video zoom:

SEQUENTIAL (VIDEO) PRESENTATION OF IMAGE SETS.

How does video zoom compare to GARS:

VERY DIFFERENT APPROACH BUT WITH PROMISSING RESULTS.

Any other comments:

PLEASE CONTINUE THE DEVELOPMENT!

European Commission

EUR 25215 EN – Joint Research Centre – Institute for Transuranium Elements
Title: VideoZoom: Summarizing surveillance images for safeguards video reviews
Authors: Scott Blunsden and Cristina Versino
Luxembourg: Publications Office of the European Union
2011 – iii + 39 pp. – 21 x 27.9 cm
EUR – Scientific and Technical Research series – ISSN 1831-9424 (online) -ISSN 1018-5593 (print)

ISBN 978-92-79-23092-9 (pdf)
ISBN 978-92-79-23091-2 (print)

doi:10.2788/12093

Abstract
This report presents VideoZoom, a prototype review tool that builds automatic summaries out of sequences of surveillance images taken by cameras with a fixed point of view. These summary images are then visualised in a zooming user interface allowing the discovery and annotation of images of interest.

The prototype system was used for detection of safeguards-relevant events in image sequences acquired in nuclear facilities. A first evaluation of the prototype system with inspectors from DG-ENER was performed. Results indicate that the system allows accurate reviews, can save effort and is easy to learn and use. In addition the system allows detection of unexpected events which would be missed by standard review tools.

How to obtain EU publications

Our priced publications are available from EU Bookshop (http://bookshop.europa.eu), where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents. You can obtain their contact details by sending a fax to (352) 29 29-42758.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

LB-NA-25215-EN-N

EUROPEAN COMMISSION

Publications Office

ISBN 978-92-79-23092-9

9 789279 230929