Experimenting accuracy and effectiveness in photo-interpretation for rapid crisis response and damage assessment applications through gaze control

Castoldi, Roberta; Duță, Ana-Maria

2012
## Table of Contents

Abstract .............................................................................................................................................. 8  
Introduction ........................................................................................................................................ 5  
   The topic ........................................................................................................................................ 5  
   Objectives .................................................................................................................................... 6  
   Research challenges and their importance .................................................................................. 6  
   Limits of the study ....................................................................................................................... 7  
Methodology ...................................................................................................................................... 8  
   Architecture of the prototype ...................................................................................................... 8  
      The Limited modality (*) ........................................................................................................... 10  
      The Full modality (**) ............................................................................................................. 10  
   Architecture of the experiment .................................................................................................... 13  
      The training phase ..................................................................................................................... 14  
      The test ................................................................................................................................... 14  
Results .............................................................................................................................................. 15  
   Limited modality ......................................................................................................................... 15  
   Full modality ................................................................................................................................ 17  
Discussion ......................................................................................................................................... 22  
Conclusions ...................................................................................................................................... 25  
References ....................................................................................................................................... 26  
Annex 1 ............................................................................................................................................. 28  
Annex 2 ............................................................................................................................................. 29
Abstract
The main objective of Geographic Information and Visual Perception project (GI&VP), and of the present study, was to offer a device composed of a suite of tools and based on gaze integration to support the work of photo-interpreters involved in remotely sensed imagery exploration for rapid crisis response and damage assessment applications. The system based on eye tracking technology was conceived to improve the interpreter accuracy and effectiveness. Our research challenge was to test the accuracy and the effectiveness during a non-standard photo-interpretation procedure: in the case of accuracy, we were interested in the number of clicks and in their correct assignment; in the case of effectiveness, we evaluated the interpreter’s ocular behavior and the actions performed using the tools at his/her disposal, designed in the prototype. This experiment, and the theoretic perspective in which it was conceived, expressed the attempt to carry out a collaborative work scenario combining human and computer cooperation in order to minimize each others limitations. Therefore, an important research challenge was trying to assess how much the human relies on the computer automatic extraction, and, on the other hand, how much the computer can be a support to the human.
Introduction

The JRC is exploring the improvement of the human assessment of building damage by applying image enhancement processing methods before the photo-interpretation phase. The JRC has designed a set of experiments to assess the effect of such processing on recognition mechanisms.

In the frame of the Geo-Information and Visual Perception project (GI&VP), we applied a cognitive approach\(^1\) to the remotely sensed imagery photo-interpretation process, exploring the possibility to improve the assessment of building damage, traditionally carried out by time consuming and error prone human interpretation. This task is often performed following disasters to support the information needs of emergency rescue for humanitarian relief intervention. Therefore, while on the one hand there is a high pressure to deliver a result as quickly as possible, on the other hand it is of the highest importance to ensure the quality of the assessment.

The JRC has developed several algorithms aimed at promoting the salience of targets in complex backgrounds, with the purpose of improving semi and fully automatic image information extraction. As a rich plethora of different processing methods are at the photo-interpreter’s disposal, it has become increasingly useful to test if different processing methods impact the subjective task performance (accuracy and effectiveness) of identifying building damage.

Before carrying out the prototype and the experiment described in detail in this study, we explored the performance of photo-interpretation in a cognitive perspective by (1) a thinking aloud exercise/semi-structured interview and (2) an eye tracking test\(^2\) to assess: (a) which particular image processing, in this case the “rubble”, amongst different, put under test; and (b) evaluate the impact of that particular processing – against the standard one - on the interpreter’s performance.

The main findings of the previous tests were that the images with the “rubble processing” produced a higher rate of true positives together with a higher rate of false negatives; from the effectiveness point of view, the eye tracking data showed how the targets in the rubble processing tiles were fixated first and (time to first fixation, time from first fixation to next mouse click) of the participants showed slight differences in the time to first fixation.

A cognitive approach to damage assessment stresses the fact that the photo-interpretation process for rapid crisis response and damage assessment applications is still human centered, in spite the intensive efforts towards a fully automated possibility.

“Unfortunately geospatial products automatically derived from source geospatial data are burdened with residual errors and artifacts which should be manually inspected, cleaned and corrected. These tasks become critical in many large-scale projects that require real-time processing of immense amounts of visual information and usually require manual post-processing or visual inspection of the source and/or derived data”\(^3\).

---

\(^1\) See the work done by R. Hoffman (Hoffman 1984, 1989, 1990)
\(^2\) Castoldi R., Broglia M., *Image interpreters and interpreted images: an eye tracking study applied to damage assessment*, in valGeo workshop 2011 proceedings, European Commission, JRC
Our research challenge agreed in principle with the statement that eye movements data during geospatial imagery exploration “retain meaningful information that could be successfully utilized in geospatial image processing and interpretation”\(^4\). The prototype we designed expresses the idea that the photo-interpretation process for rapid crisis response and damage assessment applications is a human-machine cooperation process where eye movements and algorithms integrated: this is the novelty of the experience we propose. Even though for our specific application no similar challenges were faced yet, there are interesting studies in the field of satellite imagery exploration and gaze interaction\(^5\); but mainly we took inspiration from the medical field, where various experiments has been carried out\(^6\).

This work was conceived, developed and carried out thanks to the collaboration, started in 2010, between JRC and SRLabs – The Eye Tracking Company\(^7\); the GI&VP project has been partly financed by the IPSC Exploratory Research Budget 2010.

**Objectives**

The main objective of GI&VP project, and of the present study, was to offer a device composed of a suite of tools and based on gaze integration to support the work of photo-interpreters involved in remotely sensed imagery exploration for rapid crisis response and damage assessment applications.

The system based on eye tracking technology was supposed to improve the interpreter accuracy and effectiveness by:

- increasing the rate of true positives
- decreasing the rate of false negatives
- decreasing the rate of false positives

**Research challenges and their importance**

Our research challenge was to test the accuracy and the effectiveness during a non-standard photo-interpretation procedure: in the case of accuracy, we were interested in the number of clicks and in their correct assignment; in the case of effectiveness, we evaluated the interpreter’s ocular behavior and the actions performed using the tools at his/her disposal, designed in the prototype.

The results of the previous experiment led to the conclusions that it was necessary to decrease the rate of false positives, mostly present in the rubble processed image tiles. To deepen the issue we designed this prototype, based on eye tracking technology, combining eye gaze data with Computer-Assisted Detection algorithms.

This experiment, and the theoretic perspective in which it was conceived, expressed the attempt to carry out a collaborative work scenario combining human and computer cooperation in order to minimize each others limitations. Therefore, an important research

\(^4\) Ibidem


\(^7\) SRLabs – The eye Tracking Company, Milan. Paolo Invernizzi, Roberto Delfiore and Laura Guffanti worked on all the experiment phases, from the conceptual core to the data analysis.
challenge was trying to assess how much the human relies on the computer automatic extraction, and, on the other hand, how much the computer can be a support to the human.

Limits of the study

This study has several limits. First of all, it was the first study in its kind and in this field, therefore we had limited results and experiments available against which we could make comparisons and draw inspiration.

Another important limit is the number of participants: an expertise study cannot normally recruit a huge number of highly skilled participants, so we had at our disposal 10 skilled photo-interpreter with at least 3 years of remotely sensed imagery interpretation for damage assessment.

One more limit is that the sub-tiles selected as test material were not identical: the scenario was a real environment and the different zones contained a slightly different number of damaged buildings. Nevertheless the proportion between the damage classes was basically the same.

Finally, in order to optimize the exercise we left out the zoom tool, as using more tools could have confused and overloaded the interpreter.

Thereafter, in further tests we could experiment more with the active prompts time threshold with different tools (i.e. the zoom) and their visualization.

In the Annex 1 we reported the participants’ observations and comments in a comparison table (see p. 21), and, in further steps we would necessarily take them into consideration.
Methodology

Architecture of the prototype

As a study case, we chose the magnitude 7.0 earthquake in Haiti of 12 January 2010, because of the availability of i) airborne imagery, whose resolution allows for visual building damage assessment and ii) an official damage assessment, which can be the starting point to measure the performance of the task.

The images - pre and post damage - used in the experiment had a square extension of Km 1.24 x 1.24; the chosen area of Haiti was between Route Delmes and Avenue John Brown (zones Cité Numero Deux and part of Nazor), the coordinate system was the UTM Zone 18 North.

The post-damage was a Google aerial (2010, 17th of January) image with a spatial resolution of 30 cm; the pre-damage scene was a GeoEye 1 pan-sharpened satellite image with a spatial resolution of 50 cm.

These images were both resampled to an area of 10K x 10K pixels and to a spatial resolution of 15 cm.

The building damage assessment for the affected area has been carried out jointly by the United Nations Institute for Training and Research (UNITAR) Operational Satellite Applications Programme (UNOSAT), the European Commission Joint Research Centre (EC JRC), the World Bank Global Facility for Disaster Reduction and Recovery (GFDRR) and Centre National d’Information Géo-Spatial (CNIGS) representing the Government of Haiti.

The damage assessment has been conducted through the use of aerial photos provided by the World Bank (World Bank-ImageCat-RIT Remote Sensing Mission), Google and NOAA, as well as satellite imagery from GeoEye and Digitalglobe, by comparing pre-earthquake satellite imagery to post-earthquake aerial photos.

Image analysts have categorized buildings into damage classes through manual photo-interpretation as follow:

- class 1: no visible damage
- class 3: moderate damage
- class 4: very heavy damage
- class 5: destruction

More teams worked with a coordinated approach at UNITAR/UNOSAT, at the EC JRC, at the World Bank, which worked with a network of volunteer collaborators, GEO CAN (Global Earth Observation – Catastrophe Assessment Network), and with ImageCat. The results of the photo-interpretation have been harmonized in a point dataset. Each point represents a damaged building and the damage grade is classified according to the European Macroseismic Scale (EMS) 1998.

---

8 Upper left corner lat-long coordinates: 72°19’40"W18°33’20"N
The following figure shows how the interpretation process is accomplished in the architecture framework of the prototype:

Figure 1. The image shows the photo-interpretation process in the two described modalities.

(*) Limited modality
(**) Full modality

The photo-interpretation process ends when the interpreter accomplishes the interpretation of the total tile, or a given sub-tiles set.
Below a short explanation of the prototype architecture main features:

**Calibration:** we used a five points calibration. As for the fixation filter, the main parameters are the following:
- temporal window dimension (time extent during which gaze data were clustered) = 400 ms;
- saccade limit (threshold to discriminate between a fixation and a saccade) = 0.03;
- max distance (threshold determining the end of a fixation) = 0.045/3;

Furthermore, fixations are computed calculating the average on the x, y gaze coordinates inside the predefined temporal window; a fixation ends if its actual average differs from a starting point of the same fixation, of a higher value than the threshold “max distance”.

**The complete tile displaying:** after the calibration the interpreter is exposed to the complete tile displaying overlapped by a grid positioned in the center of the image and dividing it into 16 square sub-tiles (930 x 930 pixels, origin (0,0) in top left).

**Choice of a sub-tile:** the interpreter clicks on a sub-tile and it is displayed in full screen so that the exploration can start.

**The sub-tile photo-interpretation:** the interpreter explores the image sub-tile in order to mark the damaged, very damaged or destroyed buildings. Thus far, there are no differences between the Limited and the Full modalities; below a description of the specification of both.

**The Limited modality (*)**

The exploration of the sub-tiles in the Limited modality was conceived as a sort of “control” one, to be compared with the one in the Full modality where more supporting tools were available. In the Limited modality, the only tool at the interpreter disposal was the tab key, which, if pressed, displayed the pre-damage image of the whole sub-tile under exploration.

**The Full modality (**)**

The exploration in this modality was supported by two optional tools, activated by keyboard, and a hinting system of active prompts, based on gaze data. Going into details:

The pre-damage image is displayed in form of a moving window-lens (radius 225) overlapping the post image and obtained by holding down the Tab key and moved over the image by the gaze position. The centre of the lens is positioned in the gaze centre. The refresh of the lens position occurs only if the variation of the new gaze position is higher than 50 pixels on the x/y axis. (Figure 2)
Figure 2: The blue circles and the red lines are the eye tracker output respectively of the fixations and the mouse tracks. The moving lens is the circle displaying the pre-damage image.

Pressing the 2 key the interpreter activated the image processed with an automatic extraction algorithm (Rubble Detection algorithm), implemented at JRC. This image completely overlaps the post image. It is possible to switch back to the post image pressing the 1 key. (Figure 3)
The following is a brief description of the Rubble Detection processing chain:

“The Rubble Detection (RD) system is based on a modular architecture that allows the customization of the process flow depending on the crisis scenario. It operates on panchromatic images of maximum intensity resolution up to 16 bits/pixel. The system input supports additional geo-referenced sources such as the infrared channel of multi-spectral acquisitions, manually generated masks, digital elevation models (DEM), and others. Pre-disaster VHR imagery can be utilized if available, to extract the building footprints. Rubble-like structures are confirmed as rubble if detected in the vicinity of a building footprint and rejected otherwise. Rubble detected in the pre-disaster acquisition is ignored in the post-disaster process flow. The RD system, following data preparation computes an image representation structure. Image information mining and meta-processes are operated directly on this structure”\(^{10}\).

According to the automatic extraction algorithm, prompts about not fixated damaged buildings appear at a certain time, in this experiment at 20”. They appear in the periphery of the gaze, avoiding confusion in the fixation areas, and encourage the interpreter to explore not explored areas, in which, according to the algorithm, building damages could be found. If the prompts are fixated for at least 500ms they disappear, if not, they stay in the periphery as long as the image is displayed and explored. The gaze periphery is refreshed every 500ms from the last gaze movement. (Figure 4)

The sub-tile photo-interpretation ends when the interpreter decides to pass to another sub-tile (alongside the interpreted one) and clicking on it. The interpreted sub-tiles become grey in order to keep track of the work already done.

Architecture of the experiment

Participants: 10 expert visual interpreters were used in this study, with a working experience of at least 3 years.
**Material:** The material put under test is described in 1.1. According to the Grid (x,y coordinates), the sub-tiles put under test were; (1,2) and (1,3) in the limited modality; (2,1) and (2,2) in the full modality;

**Apparatus:** Eye tracking model Tobii T120; see technical specification in Annex 2.

**Procedure:** The participants were fully instructed about the aims of the test and received a complete training about the functionalities of the apparatus. The task the participants were asked to perform a photo-interpretation of the assigned sub-tiles in order to carry out a damage assessment: the participant had to search for and click on the buildings he/she detected as damaged or destroyed buildings.

**The training phase**

As for the **limited modality** the training phase was articulated as follows: the participant explored and interpreted 2 tiles clicking on the buildings he judged damaged/destroyed, without time limitation. After having explored and interpreted the first tile, the participant was asked to switch to the second one moving the mouse on the new tile and left click on it. The Tab key was active as described above;

as for the **full modality:** the participants explored 2 tiles clicking on the buildings he/she recognized as damaged/destroyed, without time limitation. After having explored and interpreted the first tile, the participants were asked to switch to the second one moving the mouse on the new tile and left click on it. The task was to find and mark, left clicking, the buildings the interpreter judged damaged/destroyed. All the functionalities above described were available.

**The test**

As for the **limited modality:** the participants explored 2 tiles for 2’ each. After having explored the first tile for 2’, detected, marked and left clicked on the buildings the interpreter judged damaged/destroyed, the participant was asked to switch to the second tile by moving the mouse on the new tile and left click on it; the task was the same. The participant had at his/her disposal the Tab key to switch to the pre-damage image. The pre-damage image was displayed as long as the key was held down.

as for the **full modality:** the participant explored 2 tiles for 2’ each. After having explored the first tile for 2’, detected, marked and left clicked on the buildings the interpreter judged damaged/destroyed, the participant was asked to switch to the second tile by moving the mouse on the new tile and left click on it; the task was the same. The participant has at her/his disposal the set of tools described in 1.2.

At the end of the test the participants were asked to express comments and remarks (see Annex 1).
Results

Limited modality

TRUE POSITIVES

Tile 12

In the tile 12, the 8 participants identified 87 true positive targets (TP). Approximately 11% of these belonged to damage class 3, while 18% belonged to damage class 4, and 70% to damage class 5.

The tool Tab key displaying the image of the pre-damage scene was used in 83% of cases. On the TP belonging to damage class 3, the this tool was used in 90% of the cases, on the ones belonging to damage class 4 in 70% of the cases, while on the ones belonging to damage class 5 in 85% of the cases.

Average fixation duration on targets:

We intersected the damage class variable with the average fixation duration variable, in order to understand if more damaged targets were fixated for a longer duration of time or not. We started by clustering the average fixation duration in 3 classes as follows: from 0” up to 1” (cluster 1); from 1” up to 2” (cluster 2); from 2” up to 3” (cluster 3).

Overall, 52% of the targets were fixated for up to 1”, 46% for a duration ranging from 1” to 2” and 2% for more than 2”. Approximately 56% of the targets included in damage class 5, were fixated for up to 1”, 43 % between 1” and 2” and only 2% for more than 2”. Within damage class 3, 30% of the points were fixated for up to 1”, 60 % between 1” and 2” and 10% more than 2”. This result is due to the fact that heavily damaged targets are easier and faster to see and do not require intense fixation.

Time to first fixation on target:

In the analysis we clustered the values obtained by this metric so as to establish a comparison between the limited and the full modality, where the main fact was that prompts on the damaged areas, extracted by the algorithm, appeared at 20”.

We divided the variable in 4 clusters: cluster 1 contained values from 0” up to 20”, cluster 2 from 20” up to 40”, cluster 3 from 40” up to 80” and cluster 4 from 80” up to 120”. The time to first fixation was less than 20” in 40% of cases, between 20” and 40” in 23%, between 40” and 80” in 22%, between 80” and 120” in 15%.

It is interesting to see the difference between the 3 damage classes. We could notice that the majority of the structures belonging to damage class 4 (44%) and 5 (43%) were fixated within 20”, while the same is valid for only 20% of the structures belonging to damage class 3.

The longer elapsed the fewer passed structures belonging to damage class 4 and 5 were seen. As for the structures belonging to damage class 3, the situation for the duration of the test remained almost constant. The percentage rose from 20% in cluster 1 to 30% in cluster 2 and 3 (from 20” to 80”) ending at 20% in cluster 4.

11 The raw data analysis was performed in SPSS
The 74% of the targets fixated before 20” belonged to damage class 5, the 20% to damage class 4, the 6% to damage class 3.

**Time from first fixation to next mouse click**

The average time from the first fixation to the mouse click was 5620ms, with 92% of the true positives being clicked within 15” since they were seen. There were no significant differences between the different damage classes; they all reflected the general trend.

**Tile 13**

In tile 13 we had a total of 240 True Positives (TP) from which 5% belonged to damage class 3, 12,5% to damage class 4 and 82,5% to damage class 5.

The tool Tab key displaying the image of the pre damage scene, was used in 83% of cases. Regarding the damage grade classes, the general trend is reflected in the 3 damage classes; no significant differences between the classes were found.

Average fixation duration on targets:

Overall, 78% of the targets were fixated for up to 1”, 19% for a duration ranging from 1” to 2” and 3% for more than 2”.

Within damage class 3, 75% of TPs received an average fixation duration lower than 1” (cluster 1) and 25% an average fixation duration between 1” and 2”. Within damage class 4, 70% of TPs received an average fixation duration lower than 1”, 27% between 1” and 2”, and 3% higher than 2”. Within damage class 5, 79% of TPs received an average fixation duration lower than 1”, 17% between 1” and 2”, 3% higher than 2”. The damage classes showed the same trend.

Compared to the previous tile we could notice that the average fixation duration decreased from 996ms in tile 12 to 702ms in tile 13.

**Time to first fixation on target:**

The average time to first fixation in tile 13 was approximately 33”. Overall 42% of the targets were fixated within 20”. In the same time, 10% of the targets were fixated after 80” but before 120”.

Intersecting this variable with the damage grade class we saw that, 43% of the targets belonging to damage grade 5 were fixated within 20”. The same is valid for 40% of the targets belonging to damage grade 4, while only 25% of the targets belonging to damage grade 3 were seen in the same range.

The changes inside each damage class are quite interesting. The majority of the targets belonging to damage class 5 were seen within 20” and as time passed their number decreased. The majority of the targets belonging to damage class 4 were seen within 20” (40%), then after 20” but before 40” their number decreased substantially (10%). After 40” but before 80” the number of targets that were seen rose again (37%), while the last ones (13%) were seen almost at the end of the test, after 80”.
The 85% of the targets fixated before 20” belonged to damage class 5, the 12% to damage class 4, the 3% to damage class 3.

**Time from first fixation to next mouse click:**

The average time from the first fixation to the next mouse click was approximately 6”. 91% of the TPs were clicked within the first 15”; taking into account the damage classes, there are no significant differences between the different damage classes, because they all reflect the general trend.

**FALSE POSITIVES (FPs)**

4% of the total clicked targets in both tiles were false positives. In tile 12, they were approximately 13%. In tile 13, they decrease dramatically to only 1%.

**FALSE NEGATIVES (FN)**

**Tile 12**

The total number of FN in tile 12 was 1769, from which 24% were included in damage class 3, around 41% in damage class 4 and 35% in damage class 5; the 38% were fixated at least once; the average fixation duration was 796ms. The tab key was used in the 76% of cases of fixated FNs.

**Tile 13**

The total number of FN was 1720 from which 29% belonged to damage class 3, 34% to damage class 4 and 38% to damage class 5; the 40% were fixated at least once; the average fixation duration was 835ms. The tab key was used in the 78% of cases of fixated FNs.

**Full modality**

**TRUE POSITIVES**

**Tile 21**

The total number of TPs in this tile was 96, from which 1% was included in damage class 3, around 13% in damage class 4 and the 86% in damage grade 5.

The difference between the full modality and the limited one is that the interpreter had at his disposal several tools:

- The use of the Tab key to display the pre-damage scene over the fixated area (moving lens);
• The use of the “2” key to display the algorithm extracted image (image with algorithm);
• Active suggestions (prompts) on the non-fixated damage areas, starting automatically after 20” from the start of the exploration.

Summarizing their use in 11.5% of the cases, none of the above tools was used; in 46% of the cases only one was used; in 38.5% two tools were used and in 4% of the cases all three tools were used.

Moving lens on TPs
This tool was used in 83% of the cases. It was used on the only target belonging to damage class 3, on 75% of the targets belonging to damage class 4, and on 84% of the targets belonging to damage class 5; in 52.5% of the cases in which the tab key was used, it was used before the appearance of the prompt. In 50% of the cases it was used after. In two cases it was used both before and after; this is why we do not have a 100% sum.

Image of the extraction algorithm
This tool was used in 12.5% of the cases. On the damage class 4 targets it was used in 17% of the cases, while on the damage class 5 ones in 12%; in 42% of the cases the key2 tool was used before the appearance of the prompt, while in 58% of the cases it was used after.

Active prompts:
The suggestions appearing on the targets were fixated in 40% of the cases. In this tile we have only one target from damage class 3. The prompt on this point appeared and it was fixated. In the same time the prompt were fixated in 17% of the damage class 4 targets, and 42% of the damage class 5 targets.

Time to first fixation:
The average time to first fixation was approximately 40”. The time to first fixation for 35% of the targets was less than 20”; for 25% between 20” and 40”; for 20% between 40” and 80”, and still for 20% between 80” and 2’. The division inside each damage class respects the general trend. The 88% of the targets fixated before 20” belonged to damage class 5, the rest to damage class 4. The only target belonging to damage class 3 was fixated after 80”.

One of our goals, in the frame of this experiment, was to test the influence of the active prompts on the interpreter’s performance, that’s why it was of crucial importance, to see what happened before and after 20” (the time at which the prompts started to appear on damaged areas not yet fixated).

Intersecting the prompt fixated variable with the time to first fixation variable, it resulted that 35% of the targets were fixated before the appearance of the prompt (0” to 20”), 25% just after the suggestions started to appear (20” to 40”), while the remaining ones during the duration of the test.
**Average fixation duration:**

The average fixation duration was 919ms. Within the damage class 4 targets, 75% were fixated for less than 1000ms, while the same is valid for 60% of the damage class 5 targets.

**Time from first fixation to next mouse click:**

The average duration from the fixation to the mouse click is 4218ms. 96% of the targets were clicked within 15”.

**Tile 22**

The total number of TPs in this tile was 111, from which 4% were included in damage class 3, around 4% in damage class 4 and the 93% in damage grade 5.

Summarizing the use of the tools in 10% of the cases, no tool was used; in 46% of the cases only one was used; in 37% two tools were used and in 7% of the cases all three tools were used. We could notice that the situation is similar to the other tile.

**Moving lens on TPs**

This tool was used in 91% of the cases. It was used in 50% of the targets belonging to damage class 3, on 100% of the targets belonging to damage class 4, and on 92% of the targets belonging to damage class 5. It was used more on this tile than on the previous one; in 49.5% of the cases it was used before the prompt and in 53.4% of the cases it was used after. In three cases it was used both before and after. According to these results we can deduce that there is no influence of the prompts on the use of the tab tool.

**Image of the extraction algorithm**

This tool was used in 13.5% of the cases. On targets belonging to the damage class 3 and 4 targets it wasn’t used at all, while on targets belonging to damage class 5 it was used in 15% of the cases. There is no big difference in the use of this tool, between the two tiles; in 33% of the cases it was used before, while in 67% of the cases it was used after. Here we could observe that this tool was used more after the appearance of the prompts.

**Active prompts:**

The suggestions appearing on the targets were fixated in 40% of the cases. The prompts were fixated on 25% of the damage class 3 targets, and 42% of the damage class 5 ones.

**Time to first fixation:**
The average time to first fixation was approximately 33”. The time to first fixation for 45% of the targets was less than 20”; for 20% between 20” and 40”; for 23% between 40” and 80”, and for 12% between 80” and 2’. The 94% of the targets fixated before 20” belonged to damage class 5, the 2% to damage class 4, and 4% to damage class 3.

Considering the appearance of the prompts on targets that appeared at 20” we might say that 35% of the targets were fixated before the appearance of the prompt (0” to 20”), 20% just after the suggestions started to appear (20” to 40”), while the remaining ones during the duration of the test.

Average fixation duration:

The average fixation duration was 998ms. Within damage class 3 targets, 75% were fixated for less than 1000ms, while the same is valid for 50% of the damage class 4 targets and 55% of the damage class 5 ones.

Time from first fixation to next mouse click:

The average duration from the fixation to the mouse click is 6021ms. 90% of the targets were clicked within 15”.

FALSE POSITIVES (FPs)

6% of the total clicked targets in both tiles were false positives. In tile 21, they were approximately 7%. In tile 22, they decrease to 6%. It is important to mention that 29% of the false positives in tile 21 and 57% of the cases in tile 22 were caused by the buffer of the algorithm that made the prompts appear on non-damaged building.

FALSE NEGATIVES (FNS)

Tile 21

In this tile the total number of FNs was 1552: 20% included in damage class 3, the 49% in damage class 4 and 31% in damage class 5. The active prompts were fixated in 18% of cases. On 62% of the FNs there was no fixation and this trend was reflected within the three damage classes.

In the case of the FNs that were fixated, the average fixation duration is 858ms; the moving lens with the pre-damage image (tab key) was used on 75% of the false negatives fixated. In 61% of the cases it was used before the prompt appeared, and in 43% it was used after. In some cases it was used both before and after.

The image with the algorithm tool (key2) was used in 10.5% of the cases. In 54% of the cases it was used before the prompt and in 49% it was used after. In some of the cases it was used both before and after. This explains the fact that the sum is higher than 100%.
Tile 22

In this tile the total number of FNs was 1377: 24% included in damage class 3, the 26% in damage class 4 and 50% in damage class 5. The active prompts were fixated in 15% of cases. On 70% of the FNs there was no fixation and this trend was reflected within the three damage classes.

In the case of the FN that were fixated, the average fixation duration is 1030ms; In tile 22, the tab key was used in 72% of the cases. In 60% it was used before the prompt appeared and in 44% it was used after; the image with the algorithm tool (key2) was used on 18% of the false negatives. In 47% of the cases it was used before the prompt and in 54% of the cases it was used after. The differences are not big enough to allow us to make any assumptions.
Discussion

a. Accuracy
Evaluating the accuracy, i.e. the correct click assignment to the targets, we had an increment of the TPs from the first to the second tile both in limited and full modality. This could be likely explained by the learning effect.

More true positives were clicked in the limited modality than in the full. In the limited modality 9% of TPs were clicked while in the full modality only 7%.

The percentages of FNs at least fixated were respectively the 35% in the limited and the 32% in the full modality.

b. Effectiveness

i. Use of tools
Overall in the first tile of the full modality in 11% of the cases none of the tools were used, while in 46% of the cases one tool was used, in 39% two tools, and in 4% all the three tools. In the second tile, of the full modality in 10% of the cases none of the tools were used, while in 46% of the cases one tool was used, in 37% two tools, and in 7% all the three tools. This result showed that probably the familiarity with the tools lead to an increment in their use. All cases in which all three tools were used, belonged to damage grade 5.

![Number of tools used](image)

Evaluating the use of the tools supporting the interpreter exploration, we noticed that as far the use of the tab key displaying the pre-damage image (moving lens in the full modality), was slightly more in the full modality rather than in the limited. Interestingly the only target belonging to damage class 3, in the first tile of the full modality, was clicked using the tab.

The image with the algorithm, only in the full modality, was used in both tiles on average the 13% of cases; the prompts were fixated in 40% of cases in TPs and were fixated on average in 31% of the FNs of the two full modality tiles.
We analyzed the “reliance” to the prompts, calculated as a ratio between prompts clicked and prompts fixated: we obtained an average reliance to the prompts of 36% for the first tile and 39% for the second. The reliance increased slightly from the first to the second tile.

**ii. Ocular behavior**

There was an increment of the average fixation duration (approximately 200 ms) and of the time to first fixation (approx. 2380 ms) from the limited to the full modality, and this trend was the same in all the damage classes. Nevertheless the time from first fixation to next mouse click decreased on average of 990 ms from the limited to the full modality. The only increment was noticed on the damage class 4 targets, probably due to the use of the tab key.

The impact of the damage grade on the interpreters’ performance was reflected in the ocular behavior as follows: we observed that targets belonging to damage class 5 were fixated first in all cases, in limited and full modality.

**iii. What happened before and after the prompts**

On average 40% of the total true positives were fixated before the prompts appeared. In the same time 36% of the targets were clicked before the prompts popped up. The average fixation duration on the true positives on which the prompts appeared increased by 31% than in the cases in which the prompts did not appear.

Intersecting the use of the tab key tool with the appearance of the prompts we saw that there was no influence of the prompts on the use of this tool.

As for the impact of the prompts on the use of the key2 tool, the situation was different. Overall this tool was used on average in 37% of the cases before the appearance of the prompt and in 63% after.

The explanation might be that the interpreter was checking the validity of the suggestion given by the prompts.
In the case of the false negatives, on average 51% of the fixations came before the prompts. Intersecting the use of the tools with the appearance of the prompts we noticed that there was no influence of the prompts on their use.
Conclusions

The two main conclusions we can draw from this experiment are: (1) within the two sub-tiles belonging to the same modality, from the first tile to the second there was always an improvement in the performance in terms of accuracy as well as an increment in the use of tools: in the limited modality the number of TPs improved of 175% in the second sub-tile and, in the limited modality, of the 15%; as for the use of tools, the use of the tab key within the two sub-tiles belonging to the limited, was used without showing improvement in 83% in both sub-tiles; nevertheless it was used slightly more in the full modality rather than in the limited. As far the two sub-tiles belonging to the full modality, probably the familiarity with the tools lead to an increment in their use: in the first tile of the full modality in 11% of the cases none of the tools were used, while in 46% of the cases one tool was used, in 39% two tools, and in 4% all the three tools. In the second tile, of the full modality in 10% of the cases none of the tools were used, while in 46% of the cases one tool was used, in 37% two tools, and in 7% all the three tools. (2) making a between-modalities comparison of the performances there was a decrease, from the limited to the full modality in terms of accuracy: more true positives were clicked in the limited modality (9%) than in the full (7%), and this could be explained by the novelty of the tools at the interpreter’s disposal.

Any new experiment should address the key limits of the study, namely the number of participants, which should be enlarged to have better statistical results; other tools o set of tools and visualizations should be experimented (e.g. the zoom); the images put under test should be more comparable in terms of number of TPs; other algorithms for the automatic extraction should be tested; different visualization strategies and thresholds for the appearing of the prompts should be set and experimented.

From a wider point of view, comparing the traditional approach to the proposed one, we consider that the main advantages offered by this last one are the integration between automated and human perspectives, which cooperate in the same phase and not separately, or one after the other, in carrying out the complex task, in a theoretic frame in which technology offers itself as a full support to the human needs.

Even though this type of approach is nowadays still more expensive than the traditional one, gaze control technology is more and more common and competitive in several fields, from the medical to the security, showing a clear trend towards a wide application of the equipment and consequently a reduction of the cost.
References


Carpenter, Roger H.S.; Movements of the Eyes (2nd ed.). Pion Ltd, London, 1988

Castoldi R., Broglia M., Image interpreters and interpreted images: an eye tracking study applied to damage assessment, in valGeo workshop 2011 proceedings, European Commission, JRC


# Annex 1

The participants’ remarks and comments

<table>
<thead>
<tr>
<th>Tools</th>
<th>Moving lens / Pre damaged</th>
<th>Image with the alg. extraction</th>
<th>Prompts</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 1</td>
<td>The prompts should start only manually</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 2</td>
<td>I think I won't use the 2 key to switch to the algorithm if I have prompts</td>
<td>The prompts are very useful, it makes the job easier and faster. I followed the nearest prompts and I followed I think all the prompts. I had trust in the prompts.</td>
<td>I would like to have different markers to define the damage class directly on the image while interpreting. I would like to have the possibility to erase a mark if I changed idea. 3' for the test would be better.</td>
<td></td>
</tr>
<tr>
<td>Part 3</td>
<td>Not useful. Maybe the image processed with the algorithm could be used at the beginning to give an idea of the damaged areas</td>
<td>The blue prompts are useful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 4</td>
<td>The lens is not useful and it's annoying.</td>
<td></td>
<td>The problem is the absence of the zoom. The post has much more info than the pre. It's a bit confusing, too many functions.</td>
<td></td>
</tr>
<tr>
<td>Part 5</td>
<td>The lens is confusing, pre is better</td>
<td>Prompt is ok</td>
<td>The image is not good. Not enough contrast. Lack of zoom is a problem</td>
<td></td>
</tr>
<tr>
<td>Part 6</td>
<td>Pre + lens would be better</td>
<td>Useful only to control the missing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 7</td>
<td>Pre + lens would be better</td>
<td>Not trustful</td>
<td>The fact that they are automatic can distract the attention</td>
<td></td>
</tr>
<tr>
<td>Part 8</td>
<td>The lens is confusing, the pre-damaged is better</td>
<td></td>
<td>Removing the clicked points could be useful. Zoom is needed.</td>
<td></td>
</tr>
<tr>
<td>Part 9</td>
<td>Difficult to move the lens</td>
<td>Useful</td>
<td>Very useful</td>
<td></td>
</tr>
<tr>
<td>Part 10</td>
<td>Useful</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Annex 2

Tobii T120 technical specifications

<table>
<thead>
<tr>
<th>Feature</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.5 degrees</td>
</tr>
<tr>
<td>Drift</td>
<td>&lt; 0.3 degrees</td>
</tr>
<tr>
<td>Data rate</td>
<td>120 Hz</td>
</tr>
<tr>
<td>Freedom of head movement</td>
<td>30x22x30cm; 12x9x12”</td>
</tr>
<tr>
<td>Binocular tracking</td>
<td></td>
</tr>
<tr>
<td>Bright/dark pupil tracking</td>
<td>Both - automatic optimization</td>
</tr>
<tr>
<td>TFT Display</td>
<td>17” TFT, 1280 x 1024 pixels</td>
</tr>
<tr>
<td>T/X firmware Embedded</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td>~9 kg (20 lbs)</td>
</tr>
<tr>
<td>User camera</td>
<td>Built in</td>
</tr>
<tr>
<td>Speakers</td>
<td>Built in</td>
</tr>
</tbody>
</table>
Abstract
The main objective of Geographic Information and Visual Perception project (GI&VP), and of the present study, was to offer a device composed of a suite of tools and based on gaze integration to support the work of photo-interpreters involved in remotely sensed imagery exploration for rapid crisis response and damage assessment applications. The system based on eye tracking technology was conceived to improve the interpreter accuracy and effectiveness. Our research challenge was to test the accuracy and the effectiveness during a non-standard photo-interpretation procedure: in the case of accuracy, we were interested in the number of clicks and in their correct assignment; in the case of effectiveness, we evaluated the interpreter’s ocular behavior and the actions performed using the tools at his/her disposal, designed in the prototype. This experiment, and the theoretic perspective in which it was conceived, expressed the attempt to carry out a collaborative work scenario combining human and computer cooperation in order to minimize each others limitations. Therefore, an important research challenge was trying to assess how much the human relies on the computer automatic extraction, and, on the other hand, how much the computer can be a support to the human.
As the Commission’s in-house science service, the Joint Research Centre’s mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.