Geo-referenced 3D Video as visualization and measurement tool for Cultural Heritage

Lazaros Sechidis, Vassilios Tsioukas, Petros Patias
The Aristotle University of Thessaloniki, Department of Cadastre Photogrammetry and Cartography
Univ. Box 473, GR-54006, Thessaloniki, Greece
sechidis@hypernet.hyper.gr, tsioukas@egnatia.ee.auth.gr, patias@topo.auth.gr

Keywords: Stereoscopic vision, Epipolar images, 3D video

Abstract

Just a few years ago, a video card with stereoscopic display capabilities was assumed as a state of the technology product and only professionals had the ability to have one. Rapid technology progress combined with competition among video cards producers, have eliminated this cost so that a typical PC user can afford it. In the near future these cards are expected to be standard output devices of a PC. So, maybe it is the right time to consider of a product that can use this technology: the 3D video. Additionally, if this video includes proper metadata can be a Geo-referenced 3D video.

Architects community can use this added valued product for better presentation of their work. Archaeologists can show their sites and findings in a new visual way. And this product gives the ability to make some basic measurements on it.

Typical products of a DPS are digital elevation models, orthophotos and/or video using fly-through techniques. With the help of these products, 3D video can be easily created.

Once the rendered 3D model has been created, a virtual pair of cameras can capture images of the draped model from any place above or in front of it. Images taken from two different positions away of the object containing the epipolar geometry and can be considered stereo images. In an active polarization system using 3D glasses, these images can be interlaced producing a 3D scene. Additional 3D scenes can be produced from any other position across a virtual path that is drawn above or in front of the object's model. Finally all the scenes taken along the path defined by the user can be combined and a 3D video is generated. Additionally, the position and the rotation of each camera, for every frame, can be recorded into video file; every frame then equals to a pair of images with known interior & exterior orientation.

1. The concept of Geo-referenced 3D Video

One of the early (and still of the most popular) visualization tools has been the Video. Either in uncompressed (eg. AVI) or compressed forms (eg. MPEG) have been extensively used to animate otherwise static screens, or to allow the user to visualize (although not interactively) different parts (and in different scales/zoom factors) of the interest object. Through the time, the Video output proved to posses other useful characteristics as well :

(a) It can be used to record and visualize both real scenes and computer generated one; notably both at the same time if so wished, leading to what is called "Augmented Reality".

(b) It is a readily available output of every CAD, photogrammetric or visualization tool.

(c) Widely accessible, freeware video players can play it back.

(d) It is, by nature and practice, well-integrated in WWW applications, as well as highly compatible to web browsers (eg. VRML).

(e) It is the only way to visualize moving objects.

Apart of these obvious advantages, Video also provides huge and up to now only partially explored potentials: Construction of stereoscopic video sequences for one, or video "geo-referencing" for another; and many others too.

For instance, generally, video construction within the above mentioned tools (CAD etc) is done through the design of a flight-over or walk-through "path" and choice of the "virtual camera" positions, orientation and lens characteristics. That is both the interior and the exterior camera orientation parameters are known. By just shifting this path appropriately to accommodate parallax another path can be generated and thus another video sequence. Both sequences can now be stereoscopically viewed (since their relative orientation on a frame-by-frame basis is known by construction), creating thus a 3D video! Moreover: a "Geo-referenced" 3D video, where every pixel in every video frame could be assigned "real-world" 3D ground coordinates, by simple ray intersection.

At another instance, if video is the input of a real scene to be photogrammetrically reconstructed, then each video frame can be "tagged" afterwards (ie. after the photogrammetric processing) by information concerning their position and orientation. This potential has never been used, although this kind of information was always

available as a "spin-off" from the usual photogrammetric triangulation. The result would be again a 3D "Geo-referenced" video.

But why would we need a 3D video? For many reasons; let's name some:

- Reconstructed 3D objects can be displayed via two approaches: usual computer-generated surface rendering or image-based rendering. The first one suffers (among others) by unnatural appearance and the usually incorporated visual cues (eg. shading, color, etc., used to give an impression of depth) which more than often result in confusion and ambiguities during visualization. The second one, although much more realistic, lacks the impression of depth.

- Although the 3rd dimension was always a permanent quest in object reconstruction, and although for years different disciplines were optimizing its delivered quality, we finally "settle for less" when visualizing the final outcome: we do not provide the highly desirable and costly acquired 3rd dimension!

- Stereoscopy is for long known to give to the human visualization system (besides the most vivid impression) more resolution and more accurate estimation of object (depth) dimensions.

- Current technology permits the generation and playback of 3D videos in very low-end setups; even simultaneously by many users; even through Internet.

All this effort, however, would not fulfill its potential if we would only construct a 3D video sequence for visual pleasure. Of course, visualization alone has its merits, but having already established stereoscopy it is only natural to extrapolate the line of thinking to the development of a measuring tool based on such a stereoscopic video. This way, not only all three dimensions can be visualized but 3D coordinates of any point in any frame can be obtained (not to mention of course their natural extensions like distances, areas and volumes). It is apparent that in this case we certainly want these coordinates not to refer to some "odd" display-coordinates, but to real-world dimensions. That is why we need a "geo-referenced" 3D video.

So what is finally the point? Very simply we need to develop a stereoscopic 3D video, from where one could also extract real-world dimensions (in real-time), with a simple pointing device (ie. a mouse), a simple user interface and a minimum of hardware requirements (less than 500 euros extra to the cost of a usual desktop computer).

The purpose of this paper is to show how this can be achieved. We also wish to show, by a real example, the merits of this approach to visualization of Cultural Heritage objects. Inevitably we are bound to present technical issues and we assume that the viewer is familiar with basic video technology terminology (such as frames, FPS, codecs etc).

2. Creation of image pairs

The first step for 3D video production and probably the most important is the creation of the image pairs. At this step must be defined the number of the stereo pairs will that will be produced, the size of the images, the position of each camera (related to the observed object and to other camera), the flying path that each camera will follow. Since metadata are important for a Geo-referenced video, the position of each camera and it's rotation for every frame must also be calculated.

The number of frames, related to the video speed (FPS), would give the duration of the video. For a video that does not include moving objects (e.g. presentation of an archaeological site, of a building etc), a low rate of 15 FPS gives very good results. In other case, the frame rate must be over 24 fps (usually 30 fps) if smooth movement is wanted. Therefore, for a video with 2 minutes duration, 120x15 or 120x24 image pairs are needed. This affects not only to the disk storage room that is needed for the image pairs but on the quality of the produced video, too.

The flying path that each camera will follow is the 'art' part of the process and its definition method depends on the selected software that will be used, therefore it will not be described here. The only important thing that must be considered when creating the paths is that the two cameras' should 'aim' at the same point.

The distance between the two cameras that will produce a single stereo frame is depending on the distance of them from the imaged object and should be big enough to produce the parallax effect that is responsible for the depth illusion. A thumbrule [3] that has given good results defines the stereo basis of the cameras to have a value of 1/15 to 1/30 of their distance from the imaged object.

Fig 1. The base to distance ratio should have a value of 1/15 to 1/30.
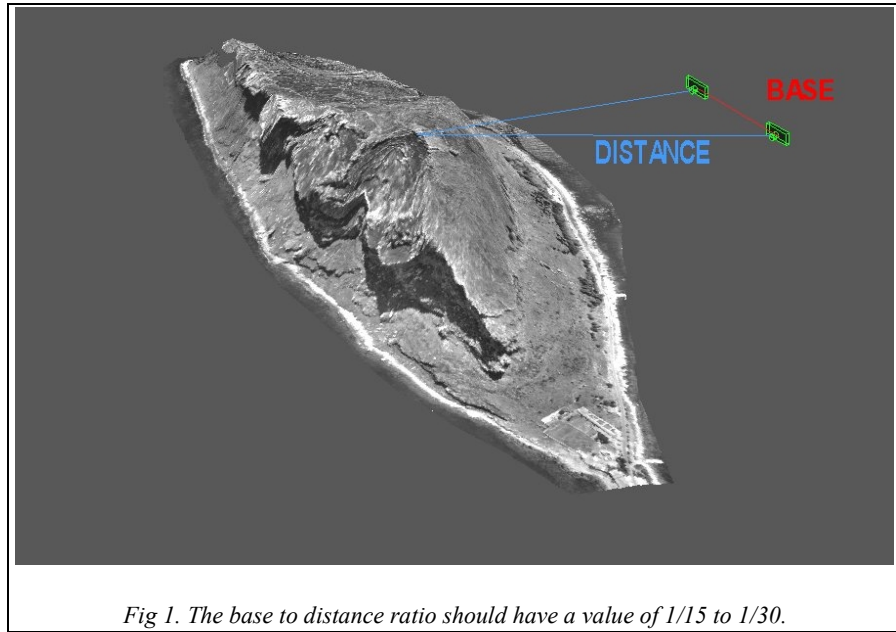
In most of cases, in order to produce the stereo-pairs, one path for each camera must be defined. Usually the two paths are parallel to each other so the distance between the cameras is the same for all frames (e.g. walking inside a building in a straight line). On special cases (e.g. walking around an object) only one path can be defined and the two cameras can follow the same path in different time segments.

Metadata of an image pair are the exterior and interior orientation of each camera. The use of virtual cameras has some great advantages: exact focus length is known, there is no lens distortion, the position of each camera and the target (subject) is known. Therefore, all the parameters of both orientations are known or can be easily calculated.

3. Interlacing images

The most famous format for 3D video is the *interlaced* format since is used from developers of the newest graphics cards chips (Nvidia etc). In this format, the odd lines on the screen content the image for the right eye and the even lines the image for the left eye. By polarizing the graphics card and the LCD glasses, the left eye sees only the even rows of the screen while the right eye sees only the odd ones. This format has two disadvantages on the stereo effect: it is totally lost if the video is zoomed by any factor and it is very sensitive to video compression algorithms. In this paper, the interlaced format will be described and used.

For other 3D video formats and their advantages and disadvantages, there is a very good description in [1].

After image pairs creation comes the interlacing of the images. In this step, each pair must create one interlaced image that will be one video frame. This way leads to the fact that the dimensions of the interlaced image will be (W x H+H) if the dimensions of the left and right images are (W x H) (fig 1). The interlaced image will have double height and the optical result will look like the image is vertically stretched by 2, not an acceptable result in most cases. It is obvious that, if the interlaced image *must* have the same height as the left and right images, only
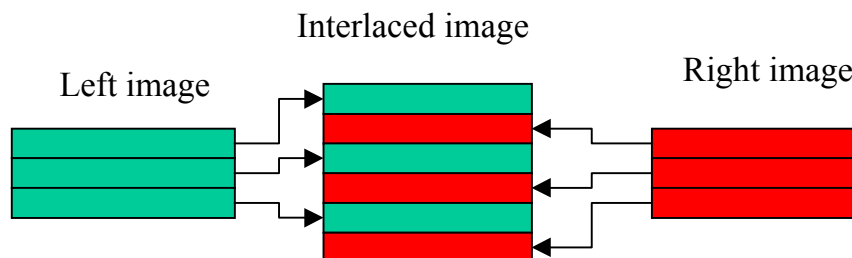

Fig 2. Interlacing images. The first line is line 0 therefore it is even and belongs to left image

half of the lines of each image must be copied. Therefore, at this point a decision must taken: which lines will be copied from the left and right images. There are four possible and logical solutions: From the left image only the

odd or only even lines will be copied. The same stays for right image. This gives four different combinations. All combinations give very good results.
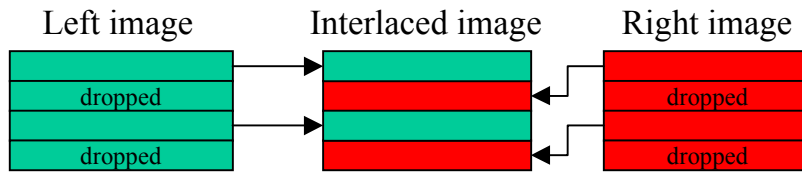
| Left image | Interlaced image | Right image |

Fig 3. Creating right sized interlaced image

Of course, this leads to loss of data since only half of each image information is stored in interlaced image.

To avoid the y-parallax, even for one pixel, it is suggested to use for both images the same format: if the odd lines will be chosen from the left image, the odd lines will be chosen for the right, too. Fig 3 shows that the even lines of each image have been used while the odd ones have been dropped.

Another way is to stretch vertically both left and right images by ½. This will lead on images with (W x H/2) dimensions. Then the interlaced image will be simply produced by multiplexing all of the left and right images lines. The sensitive point in this case is the stretching algorithm that will be used.

4. Adding metadata to 3D video – Geo-referenced 3D video

The AVI video format gives the ability to movie creators to put their metadata into movie. In our case, metadata is the interior and exterior orientations of the two cameras for each frame. These metadata can be stored using two ways:

    a. Create a new stream that has the metadata. Then store the stream into avi file as an individual stream.

    b. Create a data block with all metadata. Then assign this data block to the specific video stream.

In this research the second method was selected. Since the interior orientation needs 3 (parameters) x 8 bytes/parameter for each camera, 48 bytes needed to be stored. The exterior orientation of each frame needs 6 parameters x 8 bytes x 2 cameras = 96 bytes per frame. Therefore for a movie with 1500 frames, 1500*96+48 = 144048 additional bytes had to be stored into the video file. Since the initial (compressed although) size of the file was 55.5 Mbytes, the metadata addition increased it's size by 0.26%.

5. Video Compression

The video compression is one of the most important things to be considered when someone creates a 3D video. The algorithm that each codec uses, effects to video quality by creating small artifacts in to frame images. This may not be a serious problem when the video is 2D but in 3D it can lead to strange vision: bad vision quality, segments of frame are viewed as 3d and others in 2d. If the compression factor (therefore the artifacts) is large it can lead to total loss of depth.

Fig. 4 (Uncompressed frame, good 3d vision (courtesy of Alexander Neil)

Fig. 5 (compressed frame with artifacts that lead to bad 3D vision)

Several common compressor codecs have been tested in order to find the best compression/quality ratio of the produced video (fig 6):

Size	640x480 B/W	**Quality**	95%
No of Frames	32	**Key frame**	15

Compressor	Size (Kb)	Time (ms)	Media player	Viewed in interlaced mode by Virtual dump	Our player	3D Quality
Uncompressed	9,605	1,102	ok	ok	ok	best
Cinepack Codec by Radius	1,214	18,800	no	ok	ok	medium-
Intel Indeo 3.2	1,610	7,490	no	ok	ok	medium-
Microsoft RLE	3,179	1,130	ok	ok	ok	medium+
Microsoft Video 1	6,005	4,640	ok	ok	ok	good
Indeo Video 5.10	957	13,940	no	ok	ok	bad
Lead Video *	1,241	1,180	ok	ok	ok	very good
Microsoft MPG4	1,189	2,113	no	ok	ok	medium-
(MPEG1)	453	2,400	no	yes	no	good
(MPEG2)	375	3,000	no	no	no	good+

Fig. 6 Comparison Matrix

The suggested compression codec, especially for interlaced video is the Mpeg2 [1]. It gives hi-quality vision with a good compression factor. This is because the Mpeg2 encoder, optimized for digital video storing and broadcasting, compresses odd and even lines separately in different fields. This means that for every frame, each image of the pair is compressed separately from the other. The LEAD video codec also scored very good results.

6. Video creation

The final step is the creation of the AVI file that will include both the video stream and the metadata (stream or data block). Unfortunately, most of the software in the market has the ability to put only audio and video streams on a video file. Therefore, the use of such software can only create simple 3D video and not a geo-referenced one. For this reason, an "all to one" software for windows has been created. It can interlace 'on the fly' the image pairs (given by 2 AVI files or by a series of images), show the output frames while interlacing, give the ability to select an (available on system) compression codec and store the metadata into AVI file.
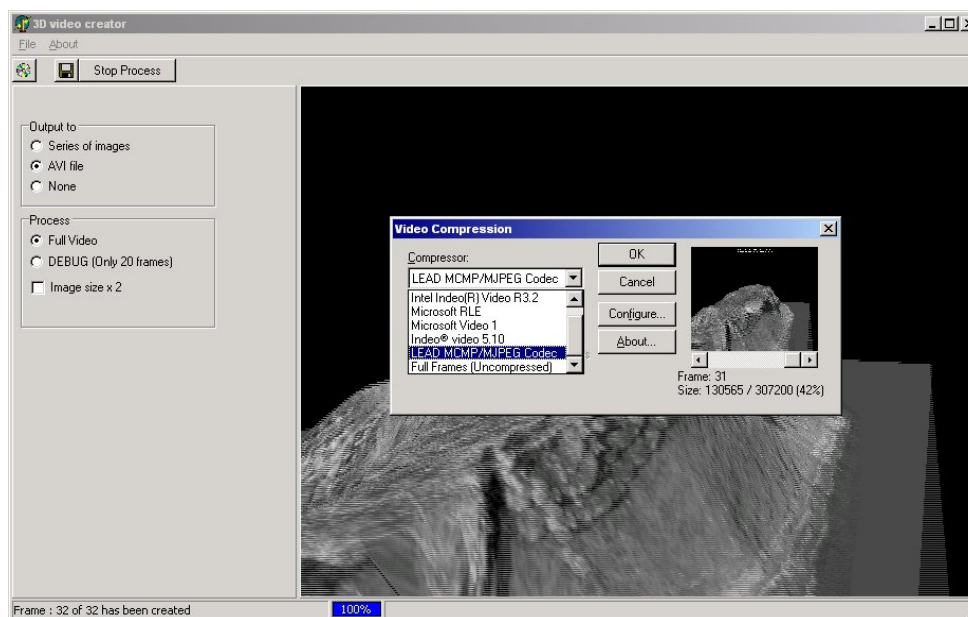

Fig. 7 3D-Video creator software

3D video creator was used to test all the codecs described in fig. 6 except the MPEG1 and MPEG2. For the MPEG1 & MPEG2 formats, the TMPEG freeware software was used but in this case metadata were not written into mpg file.

7. 3D Video playback

Unfortunately, the Microsoft's Media Player (which is the most famous video player in the Windows platform) refused to playback *correctly* all the compressed videos while the graphics card was in interlaced mode; it could play the video if they were not compressed. For this reason, hybrid video player software (with more abilities such splitting the frame into left & right images, shifting the images etc) was written in Delphi that reproduced correctly all the AVI videos. Also, a freeware but powerful video management software (VirtualDump) was tested with the produced AVI files with full success. Fig. 5 shows which formats can be viewed with the above software.

MPEG format files were viewable only with the use of commercial program (DVD software player).

There are two things that someone must take care in order to playback right the 3D video into PC:

1. Switch the graphics card in interlaced mode. Usually there is a small program that comes with every graphics card that does this. For our card (Asus 7700) it is VRViewer.
2. Move the video window in to a proper position! The first row of the playback window must be in odd vertical position on the screen; otherwise the left eye will see the right image and vise-versa.

8. The geo-referenced 3D video as measurement tool

As it was discussed above, the 3D video with metadata gives the ability to measure on it. Since each frame on video includes both left and right images and information about each camera orientations, it is easy to go back to original pair of images and to measure on them. All it needs is to de-interlace the wanted frame.

Here there is something important that needs to be noticed: while interlacing images, the vertical size of the left and right images drops to the half of the initial. This means that after reconstructing (de-interlacing) the left & right images from a frame, their size will still remain half and must vertically be stretched by 2. Only then the orientations can be used with these images.

Any available DPS then use the pair of images; but we have created a tool that uses any games graphics card with stereoscopic abilities. This tool can play the AVI file, freeze the wanted frame, split it into left & right images, read the metadata from video and do some basic measures on pair such as distance, perimeter and area. This tool is still under development.

9. Visualizing Malvasia with 3D video

As it was mentioned above, in order to create a Geo-referenced 3D video a rendered 3D model is needed.

In the case of Malvasia, a beautiful place in south Greece, six b/w aerial photos were used combined with 21 control points taken from an elevation map of 1:5000 scale to produce a DTM of about 7800 points and an Orthophoto with a pixel size of 25 cm that corresponds to a 1:5000 scale map. The Orthophoto was used as the rendered 3D model.

Two virtual cameras were used on Bentley's Microstation to shoot 1500 stereo pairs of 640x480 pixels size images, flying approximately 750 m over Malvasia and having a distance of 30 m between each other.

Four videos were produced using 3D Video Creator, having 15 FPS speed or 1 min 40 sec duration, with different compression codecs. The codecs and the size of produced videos are showed in fig.8:

Uncompressed	450 Mbytes
LEAD	244 Mbytes
RLE	247 Mbytes
MPG1	16 Mbytes

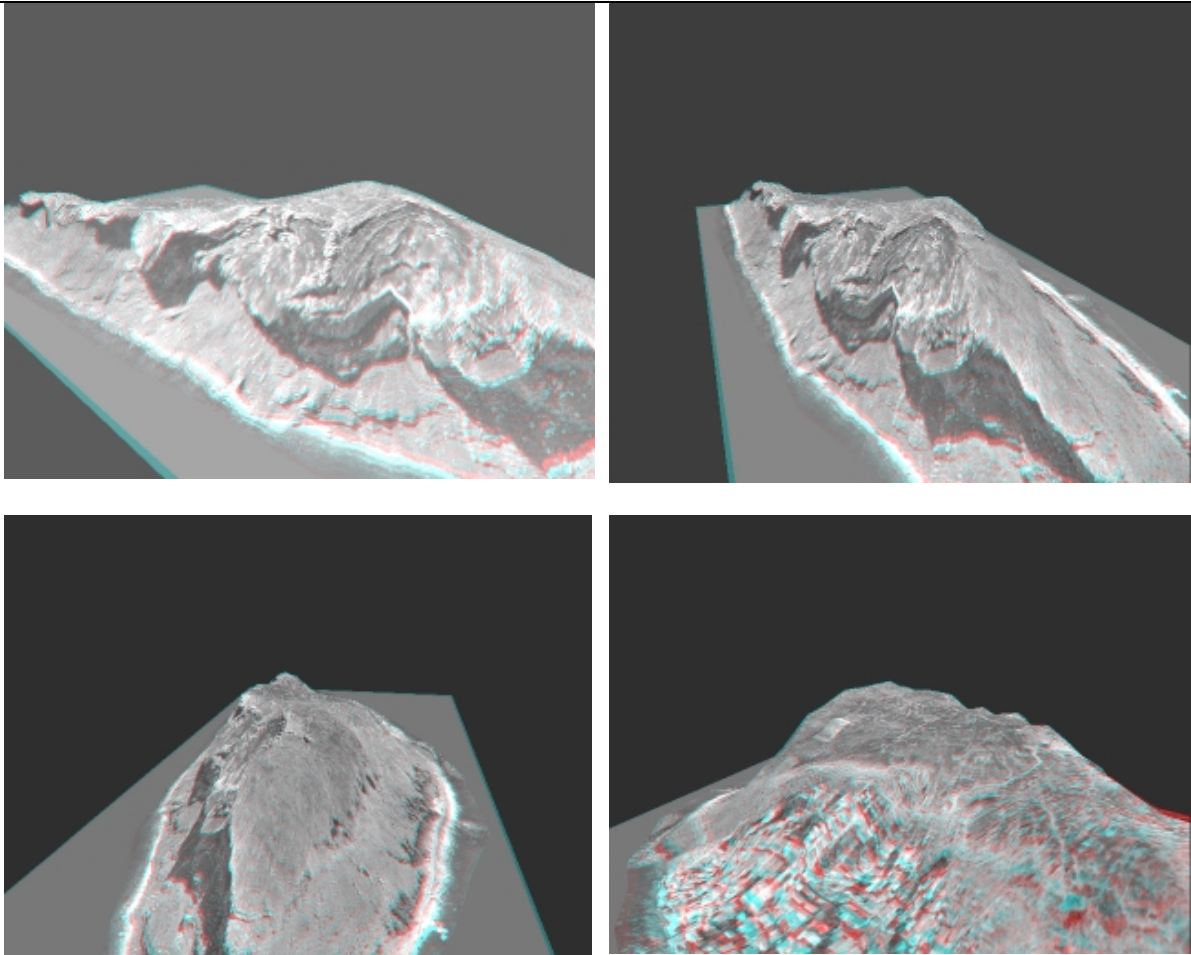*Fig 8. codecs and the size of produced videos*

Fig 9. Four framescaptured from the 3D video. Here in Red/Blue anaglyph. Original is interlaced video.
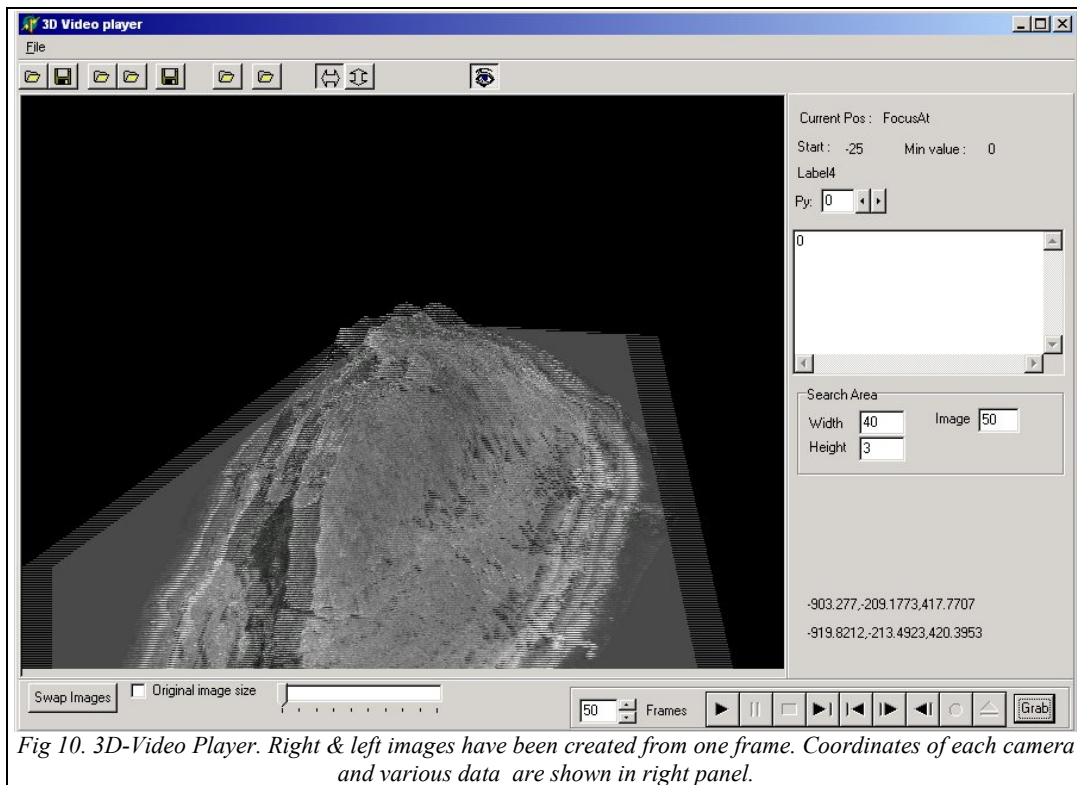

Fig 10. 3D-Video Player. Right & left images have been created from one frame. Coordinates of each camera and various data are shown in right panel.

10. Conclusions

We presented a novel technology, namely the development of geo-referenced 3D video, for both visualization and measurement of cultural heritage. The novelty of the approach (besides all the technical issues presented here, or others that are not, due to length reasons) is focused on exploiting the inherent stereoscopy of photogrammetric data and their high precision in both coordinates and dimensions extraction.

The ultimate goal has been to present a tool to both experts (architects, archaeologists, conservators, etc.) for measurements on the photogrammetric products and to general public for natural 3D stereoscopic visualization of cultural heritage objects.

Terminology (*in brief*)

Frame: *a video is a series of frames (images) that are played one after the other and give the filling of movement. Usually the frames are compressed.*
Key Frame: *an uncompressed frame that is stored in a video every N compressed frames.*
FPS: *frames per second. The number of frames played back every second.*
Codec*: (Coder – Decoder). The software driver that is used to compress and de-compress video.*
Stream: *Different areas in a video file that have different data (video, audio, metadata etc).*

References

[1] Husak, M , "Guide to making your own digital stereo-video movies in DVD quality for playing on computers", http://staff.vscht.cz/~husakm/stereopcvideo.html

[2] Stereo3d, http://www.stereo3d.com/3dhome.htm

[3] Pomaska, G, "Stereoscopic images for Visualization of Cultural Heritage", ISPRS Commission V, WG 5, IAPRS, Thessaloniki 7-9 July 1999, Vol XXXII, Part 5W11, pp.153-159