

Safe graphic insertion for stereographic material

*By Jonathan Jenkyn (3D product specialist) jonathan.jenkyn@screen.subtitling.com
and Simon Hailes (Chief Technology Officer) simon.hailes@screen.subtitling.com
for Screen Subtitling Systems*

Purpose

At Screen we have identified that the safe placement of titles using various stereographic techniques to be a major requirement as we move to support this new material type. We see significant offerings the same research and development can provide to other production and delivery chain elements. It is with this, that we propose this paper as both a documented understanding of the topics and issues, but also a solution for discussion with other implied third-parties.

Scope

The broadcast and film industry have long been aware of the requirement to produce professionally localised and captioned versions of their material for public consumption. However the inertia being applied to the stereoscopic medium has left the existing producers of 2D captions and overlays wondering what needs to change to accommodate this new format. In our previous whitepaper titled "Subtitling for Stereographic Material" (Jonathan Jenkyn, 2010) we go some way to define some of the issues, concerns and solutions to this requirement. However given the additional broadcast requirement for other graphical insertion to be made before the final display of the material is made (such as channel identifier bugs, advertising straps and other On-Screen Graphics) there is an obvious requirement to ensure that these graphics are complementary to the material they are occluding in 3D space. In this paper we will describe a proposed format and methodology that assures that a consistently matched quality of overlay is applied to stereographic material as is currently enjoyed and expected by the viewing public in 2D.

Disparity Mapping

One requirement we have identified for the correct placement of two dimensional objects in 3D space is that of the valid construction of a disparity map in relation to the target material. Disparity maps (not to be confused with Depth Maps¹) are often illustrated as a single greyscale image. This describes the computed disparity of a specific pixel with reference to the left and right channels from a central "zero parallax" position or a horoptor range. While this greyscale image format is currently an accepted form, the metadata associated with the parameters required to understand the map are currently lacking in any standardised format.

It is worth noting at this stage that while each grey value within a disparity map represents an offset in parallax, it need not represent a *whole* pixel of disparity in the original material; sub-pixel disparities can also be accommodated. Also the disparity range within the map may not be equally weighted for both positive and negative parallax disparities.

It has been suggested that certain 3D camera rig manufacturers generate disparity maps at the point of capture, and certainly this would produce the most reliable disparity map if it were retained and altered in relation to the material as it moves through the production chain. Actual examples of this

¹ Depth Maps are often captured at the capture point of the material by a camera rig and refer to the actual distances of objects from the camera position, rather than the disparities on the final presentation. They are often stored in the same greyscale bitmap form making the ambiguity that much stronger. Disparity maps and Depth maps are mutually calculable, but noise, precision and resolution loss during this calculation is inevitable.

type of material seem to be hard to come-by, and certainly those that are available seem not be adapted for additional processing within the broadcast chain.

So, in the current instance where only the left and right eye channels are available in a varying range of conditions, what computational methods can be used to generate valid disparity maps? With reference to the Middlebury Stereo Evaluation (Scharstein & Szeliski, 2010) website, several solutions to this problem can be shown for the ingestion of material and final generation of disparity maps. Oftentimes special calibration parameters need to be manually populated to allow for valid maps to be generated against a target material. In the instance of film or broadcast material these parameters are unlikely to be available and certainly would change across scene changes. Automatic calculation of these parameters or selection of an algorithm with robust adaptiveness is likely crucial in any algorithm/implementation choice.

Depth map confidence is also accounted for within many implementations, and a clip value is applied that allows for this to be adjusted as required. Low or zero-confidence is often reflected as a black value on the resulting greyscale disparity map.

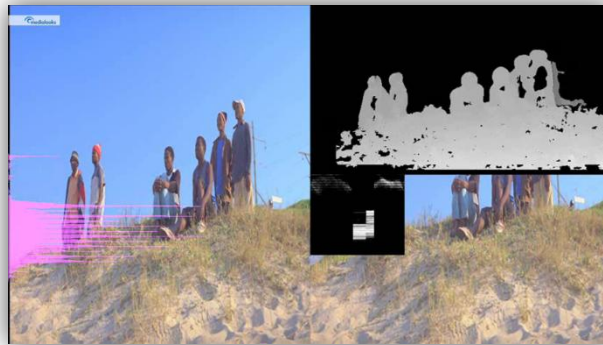


Figure 1: Disparity Diagnostic Output from MediaMate ©Copyright Screen Subtitling Systems Ltd 2010 (“Wild Ocean” material used with kind permission ©Copyright Yes/No Productions Ltd and Giant Screen Films 2010)

While film and CGI stereoscopic material is highly graded to remove various optic distortions between the left and right channels, material graded for broadcast, particularly live broadcasts, suffer heavily from an inconsistent range of issues. It is therefore recommended that for broadcast environments an adaptive algorithm is selected that is robust enough to account for the larger than usual ocular, zoom-dial, horizontal and vertical offsets/distortions we have observed. However the aim here is not to change or correct the material. It is assumed that whatever corrections need to be made to the material, indeed have been made for broadcast. In essence the disparity map reviews the stereoptic effect induced in the viewer, and while perhaps erroneous artefacts are displayed, the viewer will correct these unconsciously. However overlaid graphics will not suffer the same issue, and so care should be taken in these areas with low confidence.

The disparity map asset for a single 3D broadcast should be delivered as either a single video track, or a sequence of images totalling the number of frames within the original media. The disparity map may or may not have the same pixel dimensions as the original media, but by scaling, maps to the original material. Because of edge effects in stereographic material, areas close to the edge of the original material are generally of poorer confidence.

Information regarding the parameters used to generate the map sufficient to understand the map should also be retained as either a binary encoding into a reserved area of the disparity map frame, or as a supplementary indexed meta-data file.

Subtitle Preparation

Captioning and subtitling are often one of the first tasks undertaken upon receipt of a broadcast asset. So it is likely that any analysis that will help with the task of positioning overlays or captioning will be performed at this stage along with the generation of a suitable proxy for transcript and translation. Of course this meta-data could also be provided with the media asset, but we will assume that this is not the case simply so that we can provide an accurate workflow for current process.

We recommend that the existing 2D workflow continue as normal with a transcript being generated, timed and spotted. At this stage very minimal stylistic considerations to the titles with regards to their position, alignment or display should be made, since there is a high chance that these will change considerably during the next (stereoscopic related) phases.

Assuming the stylistic considerations of positioning titles at either the foremost disparity on screen or at the same disparity as the focal object are adhered to, a suitable depth-map can then be used in the first instance to provide a good first-cut position for the titles where the normal 2D subtitling positions are assumed. Fine tuning of these alignments can later be made by human intervention. It is our recommendation that there will be requirement for this process to be repeated for different language localisations, and may also require re-timing based on scene cuts or object occlusion.

One must also consider that as a graphic is inserted into a frame at a specified depth, the depth map for that region will then be inaccurate. An additional step whereby supplementary data is attributed to the target frames to account for this new insertion, even if the graphic is optional (such as subtitles), will allow further insertions to be made safely against the same map. As we will see shortly there may be further considerations regarding this localised disparity resource.

Graphics Insertion within the Broadcast Chain

A suitable disparity map resource should be made available to the broadcast chain as either a single media asset package or a network accessible resource via an API. Applications within the broadcast chain that need to insert burnt-in or optional graphic elements, will be required to query this resource and adapt to the underlying disparity map. The relevant cumulative disparity map returned will contain each previously inserted graphics position and occlusion information sufficient for the inserting application to make stylistic decisions on optimal position.

The same graphic insertion applications will also need to update the disparity file directly, or query an API in the instance of a network resource asset, with additional disparity map alterations. In many instances this disparity map can be queried in a time-frame basis, allowing for the inserting process to have 'ahead-of-time' knowledge of occluding objects and cumulative depth safe information, thus allowing for clean in and out cues for graphic insertions.

It is recommended that the final cumulative disparity map for a given channel (including all optional graphical elements such as all relevant included subtitle transmissions) be transmitted in a coarse form such that the resource can be interpreted by a set top box while not adding significantly to the overall transport bitrate. For example, segmentations of 16 by 9 can be packaged comfortably in to a single transport stream packet.

It may also be a requirement for supplementary data to be transmitted that includes the optional elements such as localised subtitles. This data will allow the STB to decide whether or not the data is relevant to the current display given particular overlays. Namely, an area may be occupied by a subtitle, but this occlusion is only important if the STB is displaying this optional graphical object, and so is a legitimate target for any other graphic to overlay into.

Graphics Insertion within the Set-top Box

There is certainly a requirement for set-top boxes to insert graphical overlays on top of the received video. Graphic insertions of elements such as EPGs, OSDs and Internet enabled widgets are commonplace in the 2D set-top boxes, and consumers will rightly expect the same information to be professionally rendered in 3D enabled set-top boxes also.

Current implementations often sidestep the difficult requirement to position overlay graphics in 3D space by blanket masking or simply falling back to a 2D display during the brief time of user interaction. However several STB manufacturers have expressed a requirement to position overlays in 3D space in a way that does not occlude the material running underneath. Utilisation of disparity map information will be valuable for this process.

Graphics Insertion within the Display

In several instances the set-top box *is* the display, and in these instances the televisions' on screen graphics can be handled in the same way as is described above. However it is obviously a requirement for these rendering processes to update the disparity data in a similar manner to burn in processes in the broadcast chain to avoid occlusion of STB inserted graphics (such as Internet enabled widgets).

We highly recommend that area be utilised within HDMI 1.4a (HDMI Licensing, LCC, March 2010) extensions to allow the appropriate cumulated disparity information for the entire presented frame be transferred to the display where necessary to facilitate this process within the display appliance.

Considerations for Live Subtitling

Obviously in the instance of live subtitling, various conditions change with regards to the broadcast chain. The first alteration is the position of subtitling preparation within the chain; this is moved from being one of the first elements that is prepared to one of the last. However one broadcast chain limitation exists that acts in the favour of live subtitling and object positioning in 3D; the encode delay resulting from encoding video to MPEG-2 (International Organization for Standardization, 2007) or MPEG-4 (International Organization for Standardization, 2004) provides opportunities for a period of assessment of disparity data and suitable placement of overlay objects where they are delivered directly to the multiplexer (examples being DVB-Bitmap (ESTI, 2002-2010)

and DVB-Teletext Subtitles (British Broadcasting Authority and British Radio Equipment Manufacturer's Association, September 1976)).

In the case of subtitling or captioning standards which cannot be delivered direct to the multiplexer, this time advantage is not present (examples being CEA-708 (R4.3 Television Data Systems Subcommittee, August 2008) inserted in VANC and burnt-in subtitles).

The bugs and straps are inserted prior to this encode delay. This means that they will insert to positions disregarding any possible requirement for titles to reserve a specific position. There may be certain assumptions that can be made with regards to live broadcasts, in as much as many will be sports events where subtitles and captions are largely supplementary rather than required services. In this way perhaps the fact they appear later in the broadcast chain is appropriate.

The live broadcast chain is likely to link in a similar manner to this:-

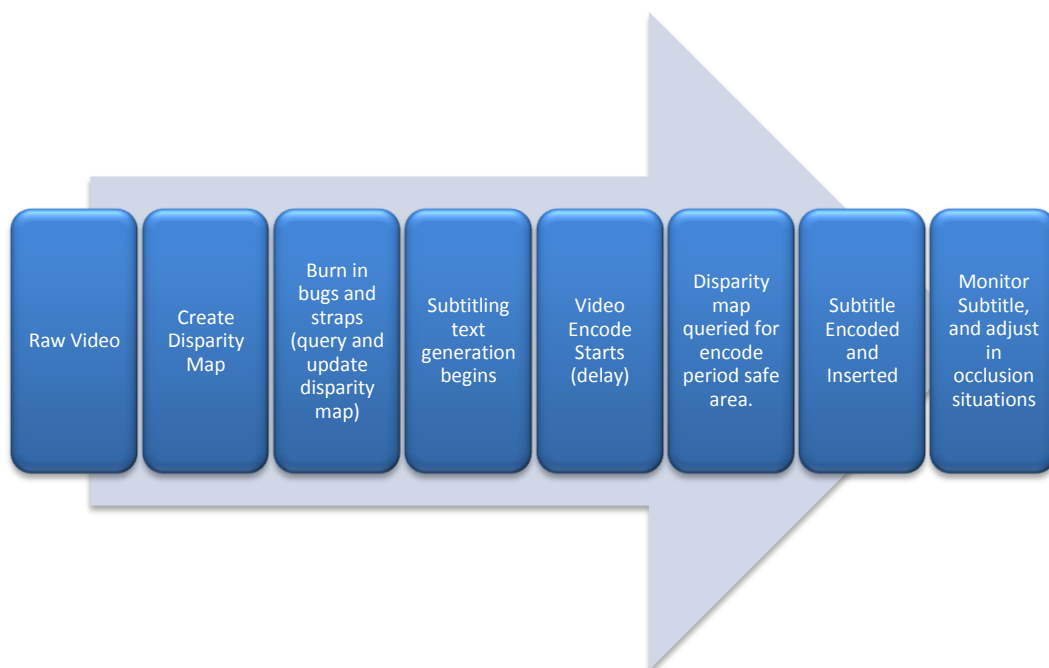


Figure 2 : Live 3D Subtitling Broadcast Chain

As you can appreciate the subtitling process begins just before the encoding delay. Using this delay it is possible to pre-empt the safe regions of the frames being encoded. After the encode delay the subtitles are positioned and encoded into the transport stream.

If during the persistence of the subtitle on screen, another object occludes or violates the 3D position of the title, a fast intervention can be made to adjust the position and disparity, allowing for analysis and re-encoding all within the encode delay.

To account for circumstances where the encode delay is too brief to be valuable for this process (or where the subtitle/caption standard does not allow) an artificial delay can be inserted into the broadcast chain. This could have the additional benefit, in the case of live subtitling, of offsetting live typing delay as well as giving 3D positional advantages. It is also appreciated that the insertion of a delay unit in a live chain is often looked upon unfavourably by broadcasters, however when one considers the benefits, it merits further consideration.

DepthSafe data

Depthsafe data is defined as a coarse representation of the original disparity map where each cell value represents the foremost disparity within the source region as a scaled value. For example depthsafe data of dimensions 16x9 consists of 144 values each representing the disparity of the closest point within the source rectangle.



Figure 3 : Depthsafe Pinboard View at 32x18 resolution, with original video texture blown on. (“Wild Ocean” material used with kind permission ©Copyright Yes/No Productions Ltd and Giant Screen Films 2010)

Depthsafe data for each frame contains a header which defines the x and y dimensions, and additional parameters which allow the calculation of disparity, in terms of pixels, from the byte values. In this way search horoptors can be fine tuned on a frame by frame basis to acquire the most relevant and accurate disparity map for the source material.

At this time, additional header parameters are being considered which could be used to aid auto scaling where relevant. (E.g. a representation of inter-ocular distance, camera convergence, depth of field, zoom factor, etc...)

The full specification for this data format will be made on the Screen Subtitling Systems website once it has completed a full review.

Bibliography

British Broadcasting Authority and British Radio Equipment Manufacturer's Association. (September 1976). *Broadcast Teletext Specification*.

ESTI. (2002-2010). *Digital Video Broadcasting (DVB); Subtitling Systems - EN 300 743*.

HDMI Licensing, LCC. (March 2010). *HDMI 1.4a*.

International Organization for Standardization. (2007). *ISO/IEC 13818-1:2007 MPEG-2*.

International Organization for Standardization. (2004). *ISO/IEC 14496-1:2004 MPEG-4*.

Jonathan Jenkyn. (2010, July). *Subtitling for Stereographic Media*. Retrieved Sept 16, 2010, from Screen Subtitling Systems: <http://bit.ly/sub43d>

R4.3 Television Data Systems Subcommittee. (August 2008). *Digital Television (DTV) Closed Captioning - CEA-708-D*.

R4.3 Television Data Systems Subcommittee. (August 2005). *Line 21 Data Services - CEA-608-C*.

Scharstein, D., & Szeliski, R. (2010). *Middlebury Stereo Evaluation*. Retrieved Sept 16, 2010, from Middlebury Stereo Vision Page: <http://vision.middlebury.edu/stereo/eval/>

Table of Figures

Figure 1: Disparity Diagnostic Output from MediaMate ©Copyright Screen Subtitling Systems Ltd 2010 (“Wild Ocean” material used with kind permission ©Copyright Yes/No Productions Ltd and Giant Screen Films 2010)	3
Figure 2 : Live 3D Subtitling Broadcast Chain	6
Figure 3 : Depthsafe Pinboard View at 32x18 resolution, with original video texture blown on. (“Wild Ocean” material used with kind permission ©Copyright Yes/No Productions Ltd and Giant Screen Films 2010)	7