Stereo Camera Geometry

Michael Starks
3DTV Corporation

Anyone shooting 3D is immediately confronted with the problem of stereo camera geometry—how to align the cameras for best results. This looks like it should be the easiest part of the entire project but in fact it's by far the hardest. Just aligning the cams perfectly in all 3 axes and locking them down is tough and keeping them aligned when changing interaxial, convergence or zoom is extremely hard. There is very little in the way of comprehensive reviews of this subject in the literature. The best list of patent and tech references I know of are still the ones I published in my SPIE papers 15 years ago and put on my page as part of the Stereoscopic Imaging Technology article. These articles, detailing hundreds of patents and papers on single and dual camera stereo, and many other areas of 3D imaging, as well as hundreds of articles by others, are on the two SPIE CDROMS containing all the papers from the Stereoscopic Displays and Applications conferences up to 2001. These CD's are mandatory for any true enthusiast.

Some may be surprised to learn that these problems are not new, nor are they unique to 3D video and photography. In addition to attention from stereographers for over 150 years, they have been the subject of intensive research in the fields of photogrammetry going back well over 100 years, and more recently in computer vision. Every book in these arenas has extensive discussions on multiple camera geometry and essentially the entire texts revolve around the problems of camera registration and image rectification for human viewing and/or computer image understanding. Algorithmic transforms for producing rectified images from single moving cameras, polydioptric (plenoptic or multiple image single lens cameras) or multiple cameras take up large sections of these books and thousands of papers, which blend into the literatures of robotics, machine vision, artificial intelligence, virtual reality, telepresence and every aspect of 3D imaging. I will cite only the continuing work from Kanade at Carnegie Mellon *www.ri.cmu.edu/person.html?person_id=136&type=publications* as I mention it elsewhere here and it is a good place to start research in this area.

One of the most pernicious problems in 3D film and television results from the use of converged rather than parallel lens axis cameras. There is absolutely no question that this causes vertical parallax and spurious horizontal parallax (even when a virtual camera is rotated for CGI stereo) and contributes significantly to

eyestrain. This is basic knowledge in stereo photography, photogrammetry, and machine vision and has also been mathematically demonstrated for the stereoscopic community many times, e.g., in great detail by Diner and Fender in "Human Engineering in Stereoscopic Viewing Devices (1993), and by Grinberg, Siegel et al in three SPIE papers a few years ago (available in the articles on the 3DTV Corp page http://www.3dtv.jp/ or at http://www.ri.cmu.edu/person.html?type=publications&person_id=285). One only has to set up a pair of cameras and view the image with parallel axes vs converged to see the problem. The closer the converged object gets to the cameras, the more eyestrain and a little closer and fusion is impossible. John Urbanic of Neotek, one of the more careful and experienced persons in the field made this comment to me recently.

"If you require a more intuitive demonstration, I suggest you take a large piece of gridded paper and use TriD to view it with, and without, convergence using shutter glasses. Try it with them on if you want, but then take the glasses off and it will be very obvious on the screen where the left and right image lines diverge in what looks a lot like spherical aberration proportional to the amount of convergence. If you do the math, it is almost the same equation to first order. The parallel cameras will give perfect overlays (assuming no regular 2D aberration)."

In truth, even parallel cameras with "perfectly" matched lenses will give serious distortions and ideally aspherical lenses should be used. If one must converge, one can do so without the distortions by horizontal shifting of the lens (rather than toeing in the entire camera) and/or imaging chip (e.g., see Figure 7 in Woods) but I don't know of cameras suitable for high quality video use that permit this to be done reliably. Another target for the serious enthusiasts with money. How about it Real D, 3ality, Pace, Imax, Sony, Panasonic, Ikegami, Philips etc? And well you are at it don't forget to add the automatic zoom convergence mechanism from Ikegami's stereo camera (I assume the patent is expired by now), automatic convergence on the principal subject etc (i.e., the stuff that is standard on consumer camcorders now) and automatic change of camera interaxials (in addition to chip and lens shifts).

As noted, there is a large literature on stereo image rectification since those doing computer stereovision or stereophotogrammetry have been dealing with these problems for over a century. One common type of rectification applies transforms to correct converged to parallel cameras. See e.g., the indefatigable Australian stereoscopist Andrew Woods SPIE article available here http://www.cmst.curtin.edu.au/publicat/1993-01.pdf , Diner and Fender Chap 9 "Reducing Depth Distortions for Converged Cameras" et passim., or chapter 10 etc. in Goshtasby's book "2-D and 3-D Image Registration (2005)-$80 from Amazon or discounted on P2P. When visiting Wood's page be sure to get all the other superb articles there since, unlike most technical work in 3D, his is of immediate practical

value. Zealots will want to download his 3D Map program which enables graphing stereo image distortions.

It is also possible to use converged optical axis cameras without (or more probably with minimal—see Diner and Fender) distortions by making a stereoscopic viewing system with correspondingly converged optical axes (Grebenyik R., Petrov V. “Stereoscopic Images Without Keystone Distortions” - Proc. Eurodisplay 2007, pp. 140-142). Also K.Grebenyuk in his PhD thesis showed that in such converged axis stereoscopic viewing systems the errors can be completely corrected. This can be done optically with a standard semi-transparent mirror systems having nonright angle with respect to the axis of one or both monitors or by using a holographic screen with two virtual mirrors recorded nonparallel. Obviously, in addition to the optical methods or computer algorithms, it could also be done in hardware via electronic image rectification using offline or realtime transforms (e.g., polar transformation-- Lee J. et al. “Stereo image rectification based on polar transformation”. - Opt. Engineering, 2008/47(8), pp.087205-1....087205-12. - and references therein). Such capabilities (e.g., correcting keystone distortion) are now available in many projectors and processing boxes. However since every shot is different it would be optimal to create metadata during filming which could be used offline for rectification that could then be projected by normal projection or display systems. As noted below, projection with dual side by side projectors might provide some rectification for converged cameras.

If you shoot converged you have to worry that objects in front of convergence will have too much negative parallax and also that those behind will have too much positive parallax, both of which cause eyestrain in low degrees and unfusable images at high ones. Parallel shooting avoids this and the only problem is lack of total image overlap in the horizontal direction as objects get close to the cameras. This is seldom a serious problem and it has various solutions as 150 years of shooting parallel photos and film shows (see below). One can crop or mask the image and/or blow up the whole frame a few percent.

This should be the end of the matter, but it seems that many, including 3ality (the makers of the recent U23D film), Peter Anderson, Jim Cameron, Vince Pace, Phil Streather and many others normally shoot converged. One even hears it said that parallel shooting gives limited depth or minimizes control over the 3D effects, but I doubt if those who say this have bothered to spend time doing meaningful comparisons. I think it’s more a matter of lack of concern and of convenience, since it’s hard to get even small cameras very close to the desired normal human 65mm interaxial, so they'd have to do alot of horizontal shifts and often blowups to eliminate nonmatched right and left edges and/or use big mirror boxes with the two cams at right angles to decrease the interaxial for close objects (as was often done in the 50’s and more recently with the immense IMAX rigs). Perhaps the biggest problem is that they are rushed and pressured in planning, on set, and in post and in any case the bottom line is that they can put almost anything they want on the

screen, 3D or 2D and get away with it, as the movie game, like all games, is about deadlines, convenience, money and power and ego and the stereoscopists are rarely in charge of the project.

Every viewer has a daily "eyestrain budget" being used up in normal life and much faster for 2D or 3D viewing of screens of any kind. It gets used up fastest by sitting in a dark theater looking at a big, bright screen, much faster when it's in 3D and very fast when the film/projection are full of errors (i.e., always), when there are fingerprints on the glasses, when one is sitting close to the front or at the sides etc. It will always be best for one's budget if one sits far in the back at the center with clean glasses and without any reflections in them from theater lights.

A major reason people get away with shooting converged is that the subject is usually not too close and the convergence mild. Also the limited depth of field leaves the background out of focus and attention is on the subject even when it is in focus. I think that nearly all films shot prior to the creation of the single camera-single projector 3D systems in the 1970's and 80's had many shots basically parallel and most others with only mild convergence. Mostly they look great –superior to later work. In fact when I viewed these films (as have thousands at the nearly complete recent retrospectives in the USA of the 50 or so films and many shorts done prior to the 60's), I was stunned at how good the images were—this in spite of such impediments as the huge blimped cameras with slow film (necessitating huge lights, large apertures and limited depth of field), lack of modern wide angle lenses and projection, lack of perfectly matched dual camera and projection lenses, and the jitter and weave of the film in the cameras, printers and the dual interlocked projectors. I am sure a good part of this is the fact that most shots were nearly parallel, as one can see by taking the glasses off from time to time. They are mostly very easy on the "eyestrain budget," in comparison with subsequent work (see e.g., my IMAX reviews). Another reason they looked good is that they had the full resolution of two 35mm filmstrips. Also, when you project with two side by side projectors this to some extent automatically compensates for the convergence of the two cameras.

The 70's invention of the single camera, single lens systems of mostly modest quality, with a convergence control on the lens, resulted in "convergence abuse", and since the single projector lenses were also limiting and screen brightness low, these systems rapidly exhausted the eyestrain budget. I did extensive work in the 80's transferring 3D film from many different formats to videotape. Horizontal shifting with blowup made both single and dual camera films much easier to watch. Subsequently, classics such as 'Dial M for Murder", "Creature from the Black Lagoon", "The French Line" and others have been released in field sequential format on video by various entities starting with the Japanese VHD disks in the late 80's and I have seen some of them many times. Even with the dramatic drop in

resolution, limited dynamic range, tiny screen, etc. they are still mostly excellent and one can see that there is minimal convergence in most shots.

In the 70's and 80's Russian workers built and used the single camera, dual lens 70mm Stereo70 system and I saw projections of some of their films when I visited NIKFI in Moscow in 1985. I made a deal with them to transfer four Russian 3D films and half a dozen short works and 3DTV Corp has sold them on video for 16 years. One can see that most shots were close to parallel. I had previously spent 12 years finding just about everything technical ever written about stereoscopy and had many of the best Russian articles translated, since they have long been among the world's leaders in this field. The results of some of this work appeared in my articles and in Lipton's "Foundations of the Stereoscopic Cinema," some 25 years ago -- now freely available online (see Woods page, Real D etc). In addition, I wrote about these issues in American Cinematographer then and posted articles on my page 15 years ago.

The stereophotographer might venture that nobody has to guess about the merits of shooting parallel as they can see it in the very common 3D slide shows or photos, virtually all of which seem to be parallel. Nearly all 3D still cameras made over the last 150 years have what look like parallel (and fixed) lenses and over 99% of all the mostly superb (and non eyestraining) 3D slides/ photos ever shot were done this way (i.e., without deliberate convergence). In half a century of viewing and discussing 3D stills I have never heard anyone say they lacked depth or realism nor heard any of the photographers say they did not have creative control over the images. Any pair of good 35mm cameras can produce slides that match or exceed the image quality, depth and comfort of anything that has ever come out of Hollywood or IMAX. Anyone who shoots and projects 3D stills or goes to a few of the many 3D slide shows knows this. And, for the higher res formats, I will be happy to match my dual 120 slides, shot with the humble, fixed lens 50 year old Russian Sputnik cameras, with anything on the big screens in film or video.

Partly this is explained by the ease with which stills can be horizontally shifted and blown up or cropped and masked to overlap the two images and manipulate the stereo window. However, the fact that one can change the fixation point of the lens by horizontal shifting of the lens or the film has been understood from the beginning, and every good stereo camera has offset the lenses so that they provide a converged overlapping stereopair at the (often fixed) focal plane. Converging in this way produces an undistorted stereopair, in contrast to the spurious H and V parallaxes unavoidable with convergence by toeing in the whole camera. Thus, most single camera stereophotography is actually converged (though usually fixed by the factory with one convergence) and this, coupled with subsequent image shifting and masking and high resolution, color and dynamic range, account for the superior look of much stereophotography. There are abundant discussions of these issues in the literature of stereo photography, photogrammetry and computer vision and an admirably clear one in Diner and Fender.

Unfortunately, most videocameras have no provision for H shift of lens or imaging chip and this leads the stereographers to the drastic measure of converging the whole camera, with attendant distortions and abuse of the eyestrain budget. A wonderful pair of papers by Prof. Mel Siegel (of the world famous Carnegie Mellon Robotics Institute) a decade ago investigated the issue of reducing eyestrain by such means as horizontal image shift to overlap the images and by reducing the camera interaxial http://www.ri.cmu.edu/publication_view.html?pub_id=2550 . He also investigated, with Shojiro Nagata and others, various means to accentuate 2D image cues, simultaneous with H shift and reduced interaxial, in order to maximize viewing comfort http://www.ri.cmu.edu/publication_view.html?pub_id=3567 . They were able to produce comfortable views with good depth by judicious manipulation of these parameters. However, as I mention in my article on stereo projection and viewing, the applicability of virtually all perceptual experiments to viewing commercial devices in real environments for prolonged times is unknown. Of course, they are hardly the first to pay attention to such issues, but they were first to attend to them all simultaneously. Curiously, though I have known them both for many years and presented papers at the same symposia and appearing in the same SPIE volumes as their own, they were unaware that I have been making use of these means since the mid 80's in the 3D videos I have sold, and made similar comments to their own in my patents and papers. I normally shift all video to minimize parallax and then eliminate problems with nonoverlap and the stereo window by blowup. My US patent 6,108,005 on 2D to 3D conversion discusses means to stimulate depth by 2D image manipulations and I employed some of these in the "solidized" videos I made beginning in 1989. A few of these ideas were incorporated in a program included in the hundreds of thousands of stereoscopic gaming kits sold by X3D Corporation for about $100 but now available for the price of a sandwich http://www.amazon.com/X3D-TECHNOLOGIES-Gaming-System-Windows-Pc/dp/B00007FY66/ref=sr_1_1?ie=UTF8&s=software&qid=1231301259&sr=8-1. A set top box including this program is still sold as the Virtual FX 2D to 3D Converter http://www.amazon.com/VirtualFX-Television-Game-Console-Converter/dp/B0006HJII2 . This is an extremely simple program, which never made it to a second generation, and its effects are modest, but it is the only consumer device of this kind to appear. Curiously, though this is by far the best known solidizing device, and it can claim priority back to 1989 and may be regarded as anticipating many aspects of work and patents done since (e.g., the well known work of DDD and of In-Three), it is rarely cited, even in patents --which are required to cite all prior art.

Siegel et al used the well known effect of wide angle lenses to enhance the depth of their shots, even at the reduced interaxials. It has been noted by Diner and Fender that if video cameras have higher resolution, this alleviates the distortions and enables the reduction of interaxials. "…if camera resolution could be increased by an order of magnitude, a stereoscopic camera system might then reach the human stereoscopic depth threshold. Then wide inter-viewpoint distances would not be

needed to increase stereoscopic depth resolution, and this in turn would reduce the distortion to resolution ratio. The inter-viewpoint distance would then not be needed as an independent variable used to control resolution, and could instead be used to control distortion" p187. What this amounts to is that higher camera resolution takes advantage of our high stereoacuity and this will permit less distortion and better stereo at reduced interaxials. Since 60hz color videocamera horizontal resolutions have increased from about 500 lines in the early 90s when they did their research, to 4K in pro cameras and even 8k in some experimental systems, this has now been realized. Consequently it looks like it should be possible to produce relatively undistorted 3D video that has good depth and is very easy to view by reducing the interaxials with high resolution cameras (and wide angle lenses when feasible). Of course clever use of lighting to create asymmetrical illumination and shadows, object placement, and textures and colors of sets and costumes will remain a subtle art.

The absolute arbiter is how the 3D looks on the display and how you feel at the end of the program, and this depends on lots of things besides how it was shot and edited including type of display, brightness, ghosting, viewing method, reflections, fingerprints on glasses, ambient light, distance from screen, the idiosyncrasies of the viewer, how bad all the other errors are, and especially on the length of the program. I am sure I could walk up to most people, including the 3D experts, after a film was shown and find fingerprints on the viewing part of the lenses up to 90% of the time. Yes, it happens to me as well!

There is a wealth of info on stereovision algorithms and camera geometry in the machine vision literature. A good starting point is the early paper by Murray et al on stereo camera mounts http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.717, and for Murray's abundant work on wearable active vision systems, telepresence and related items since that time see http://www.robots.ox.ac.uk/~dwm/Publications/index.html . An excellent review of single or dual stereo camera methodology for 3DTV from the standpoint of computer vision is given by Stoykova et al http://citeseerx.ist.psu.edu/showciting?cid=1192076&sort=cite&start=20 . For a clear explanation of the single lens approach, as used for computer vision stereo, see http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.7845 and a further short exposition at http://www-bcs.mit.edu/people/jyawang/demos/plenoptic/plenoptic.html . These are concerned with the polydioptric (plenoptic) camera which forms numerous images via a lenticular array on the chip, and has not to my knowledge been used for 3D video except in the realm of machine vision. The lenticular array will give both H and V parallax and consequently (as Adelson and Wang note in the above citation) is a method of integral photography and harks back a century to the pioneering work in autostereo by Ives and Lippmann. One can use only a vertical lenticular array and then get only H parallax and then this art blends into that of the vast literature on lenticular photography.

Regarding 3D video, there is another single lens method that can create stereo video for human viewing. If one puts two apertures inside the lens and opens them sequentially in sync with the image capture by the chip, one gets perfectly registered and converged stereo. Please see the many citations to this art in my articles. Of course the small interaxial (i.e., the distance between the apertures) means that the subject must be close to the lens. This has led to these devices being used in stereoendoscopes and microscopes by half a dozen companies in recent decades. I visited International Telepresence Corp in Canada about 15 years ago to see their stereoendoscope and camera, both of which produced 60 Hz field sequential 3D output. I put some footage of surgery and of a horse race shot with these on one of the 3DTV Technology tapes I sold for many years. As expected, the stereo of the audience between the camera and the race track was good but flattened out by 20 meters or so. Sadly they seem to have vanished without a trace. The same appears to be true for half a dozen other stereoendoscope companies that employed similar approaches. However other companies continue to pursue this method, and it is not that hard to do. The astute will realize that one could get depth this way at greater distances by using a wider lens (and hence dual apertures further apart). This has not been lost on some inventors such as Dr Maurice Tripp, whose work I have cited in my other articles, and who is pictured in them and on my page, during a visit I made to him long ago, with a very wide lens using Dove prisms, which he made for his work on a lenticular autostereoscopic tv system some 30 years ago.

One single camera 3D approach popular with engineers is 2D plus depth. The depth map is supplied by laser ranging, structured light, Time of Flight, or related means, so that each pixel of the 2D camera is assigned a depth value. This is Philips preferred approach for their lenticular autostereo displays and has been extensively researched by many including those in Europe's ATTEST stereo video program. Nevertheless, I don't see how the depth map with one picture can provide the shadow detail, sparkle, luster and texture that one gets from the horizontally asymmetrical parallax images, and it lacks monocular occlusion and transparency data, and until I see a side by side comparison or some stunning 3D footage done with depth maps, I remain skeptical. This was also the reception given Philips recent proposal, to a 3D panel in Beijing, that their method be adopted as a Chinese standard.

View synthesis enthusiasts know that it's possible to use multiple cameras with a suitable program to synthesize any arbitrary stereo view. A vast literature exists and of course it again blends into that of computer vision, artificial intelligence, robotics etc. NewSight Corp showed live 8 view synthesis from two cameras running on a laptop and displayed on an 8 view autostereo display at the FPD show in Tokyo in April 2008. This work resulted from the efforts of German image processing expert Dr. Rolf Henkel, who developed this technique initially already in the 90s to convert historic stereophotos into lenticular prints. Dr Henkel is a pioneer in this area so I let him comment directly. "The human visual system is doing itself a view-interpolation operation (compare my page http://axon.physik.uni-bremen.de/research/stereo/Cyclops/index.html). I used the same approach in the 90s in my company PixelCircus to to convert historic stereophotos into lenticular prints. To do so, I had to develop also basic algorithms for rectification and calibration of unknown camera geometries. It was at that time that I developed the "virtual

camera" concept, which allowed arbitrary changes in 3D geometry of given stereoscopic data." Currently there is research directed at creating user controlled mono or stereoscopic view synthesis some of which is called "freeview" (not to be confused with the method of viewing stereopairs, nor with the set top boxes having a package of free digital channels).

Two cams will only do the views interpolated between them in a convincing way (i.e., not views to their right or left) but dozens of cams could be used to synthesize an entire environment. The dean of this approach is telepresence and robotics guru Takeo Kanade (of Carnegie Mellon and Japan) who has created many such systems over the years. A few years ago, with assistance of colleagues from Carnegie Mellon, he created the famed Eyevision system first used for SuperBowl 2001, but only a few times since http://dev.web.cs.cmu.edu:6666/testReleases/demo/40.html . Nearly any point of view can be created realtime, as though there were thousands of cameras. This has 25 cameras mounted on robotic arms distributed around the stadium. Time and money did not permit realtime view synthesis so it was done by morphing. but it looks very good, as can be seen by the demo on his page. The pan/tilt/zoom of the robotic arms was done by supercomputer programmer and stereo expert John Urbanic of Neotek www.neotek.com . A few years ago, with Chang Lee of TJ3D Corp., we formed a company named SeeAll with the intention of updating the system to HD and stereo, instant playback etc. which we hoped to implement for the Beijing Olympics, but none of us were inclined to run around looking for funding, so it has not come to fruition. Much smaller and cheaper systems could be used for martial arts, movies, security applications etc.

So, the bottom line would seem to be that, while we wait for a modern stereo video camera with horizontal lens and/or chip shifting and other niceties, we should try to shoot as near to parallel as possible by using small cameras and mirror boxes, with image shifting, masking and blowups to overlap images and control the stereo window. When this is not possible try to avoid large negative or positive parallaxes of infocus objects to which attention is drawn. Look for and correct window errors. When possible use wide angle lenses to stress perspective. Become familiar with all the 2D depth cues and use them to maximum effect. Use lighting, sets, costumes and the environment to get shots rich in asymmetrical illumination cues and shadows. Carefully calibrate the lenses and mounts to minimize all binocular asymmetries during shooting, rather than trying to fix them in post. Use experienced stereographers from early planning to final showing in theaters and listen to what they have to say.