TaPRA 09 – University of Plymouth

Directors Working Group

M Delbridge

Queen Mary, University of London

Glossary of Terms for capturing the 360-degree (or Omniscient) Frame:

As a means of beginning to frame a discussion around the methods and challenges facing directors working in 360 degree (or Omniscient) framing I very quickly discovered that the ream of terms I would be using in this work would for the most part be completely foreign to those reading it. There is no existing glossary of terms available for this sort of work and so I have sought top begin to address this issue with the terms I have classified below. I will leave Omniscient framing til last as this poses the greatest challenge to directors/actors and animators working with actors in the performance capture environments of the present (see glossary for description of this term). For the most part these terms are quite technical and are provided to describe the physical environments required to work in the areas of mediated performance, motion capture and contemporary film making (or image capture). It is clear that a new methodology and language is required to begin to understand this contemporary performance practice, and as this work becomes more mainstream and industry adopts these techniques as the norm it is essential that these ideas begin to filter down into our academy and vocational areas so that we are able to not only "supply" industry with trained and conversant personnel, but also able to discuss this new performance paradigm in a meaningful manner. It became very clear that this was the most obvious place to begin the discussion around directing for the 360-degree frame and it is my hope that these ideas ignite a discussion around the current language we employ in our studios and practice.

Motion Capture

Motion Capture (or mocap) is a term used to describe the process of digitally recording movement in 360 degrees and translating that movement onto a digital model in projected 3D space. While in many ways it is similar to, and borrows a lot from traditional film making, the major distinction is that it is not used to record what would be traditionally referred to as the framed moving image (the translation of the 3D to the 2D) but more accurately to record an accurate impression of plotted movement in 3D space that can then be transferred to a screen based 3D impression. For the purposes of this discussion, while mocap is used extensively in military and biomedical applications, it is the predominant use of mocap in the creative industries that I am seeking to define. In *filmmaking* (this term in and of itself is becoming increasingly obsolete) it refers to recording actions of human actors, and using that information to animate digital character models in 3D animation. When it includes face, fingers and captures subtle expressions, it is often referred to as *performance capture*.

In motion capture sessions, movements of one or more actors are sampled many times per second, although with most techniques (recent developments from ILM use images for 2D motion capture and project into 3D) motion capture records only the movements of the actor, not his/her visual appearance. This *animation data* is mapped to a 3D model so that the model performs the same actions as the actor. This is comparable to the older technique of rotoscope where the visual appearance of the motion of an actor was filmed, then the film used as a guide for the frame-by-frame motion of a hand-drawn animated character.

Camera movements can also be motion captured so that a virtual camera in the scene will pan, tilt, or dolly around the stage driven by a camera operator, while the actor is performing and the motion capture system can capture the camera and props as well as the actor's performance. This allows the computer-generated characters, images and sets, to have the same perspective as the video images from the camera. A computer processes the data and displays the movements of the actor, providing the desired

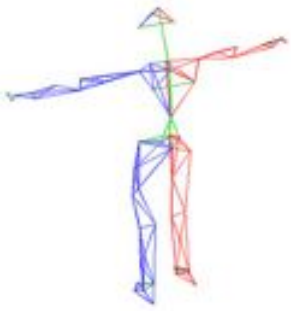
camera positions in terms of objects in the set. Retroactively obtaining camera movement data from the captured footage is known as match moving.

Performance Capture

Performance Capture is a term first employed by the Director/Producer Robert Zemeckis (Back to the Future, The Polar Express, Monster House) that is used to describe the recording of a performance, either human or animal, with an input system in motion, like Motion Analysis or Vicon. While there are many types of Motion Capture devices available; exoskeletal, magnetic, radial, in performance capture scenarios it is an optical motion capture system that is used as this is the only system that allows for multiple objects (actors) to be captured at once (up to 10 pp) and is the only system that is free of external wires and devices that inherently limit the movement and performance of the actors. Performance Capture is inherently theatrical in that it allows for a performance in its entirety to be captured in one take allowing for all traditional framing questions and dramatic devices to be employed after the performance has been recorded. Optical Motion Capture employed in animated feature production and the video game industries allows for a freedom of performance for actors that is not hindered by the constant hurdles encountered in film production where actors are continually repeating small sections of dramatic storylines or waiting for physical environments to be reset or reframed.

T- Pose

To begin with we'll jump straight in and discuss the T-Pose. The T-Pose is the standard physical pose adopted by participants being captured in performance and motion capture environments. It is also the standard character pose employed in the creation of characters in the CGI of biped avatars. The T-Pose is the internationally recognized and employed static pose for all animation and performance capture environments used for the construction of a marker data set into a recognized object or template. It has been adopted as the standard as it keeps all markers at a relative distance from each other to prevent marker swap in the construction of a template. The other vital function of the T-Pose is that it will be used throughout the capture session as a means for the markers to realign them selves as the template slowly breaks down during the day.



Markers

Passive optical Motion Capture systems, such as Motion Analysis and Vicon, use markers coated with a retroreflective material to reflect light back to a camera that generates its own near infra red light source through an array that encircles the cameras lens. The camera's threshold is routinely adjusted through a calibration process so that only the reflective markers will be sampled as points in space thus ignoring the skin and fabric and other materials in a capture volume.

The centroid of the marker is estimated as a position within the 2 dimensional image that is captured. The grayscale value of each pixel can be used to provide sub-pixel accuracy by finding the centroid of the Gaussian.

An object with markers attached at known positions is used to calibrate the cameras and obtain their positions and the lens distortion of each camera is measured. Providing two calibrated cameras see a marker, a 3 dimensional fix can be obtained. Typically a system will consist of around 6 to 24 cameras.

Systems of over three hundred cameras exist to try to reduce marker swap. Extra cameras are required for full coverage around the capture subject and multiple subjects.

The greatest benefit of Passive Optical Motion Capture is that unlike active marker systems and magnetic systems, passive systems do not require the user to wear wires or electronic equipment rather hundreds of rubber balls with reflective tape, which needs to be replaced periodically. The markers are usually attached directly to the skin (as in biomechanics), or they are velcroed to a performer wearing a full body spandex/lycra suit designed specifically for motion capture. This type of system can capture large numbers of markers at frame rates as high as 2000fps. The frame rate for a given system is often traded off between resolution and speed so a 4-megapixel system runs at 370 hertz normally but can reduce the resolution to .3 megapixels and then run at 2000 hertz. Typical systems are $100,000 for 4 megapixel 360-hertz systems, and $50,000 for .3 megapixel 120-hertz systems.

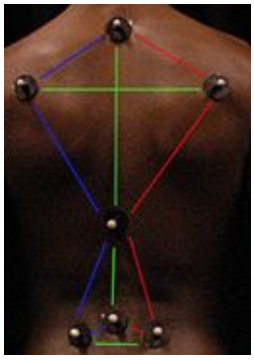
Marker Set

The marker set is the group of markers that are allocated to and placed on the human figure or object in a motion capture or performance capture environment. Traditionally for a human figure we would place between 35 and 50 markers on the body at designated areas that once used to create a template will drive the designated components of a digital skeleton in an avatar. This marker set requires uniform placement depending on the software and system that is used but here is a sample list below:

iQ_ACTORExample_V5.vst contains the following 47 markers:

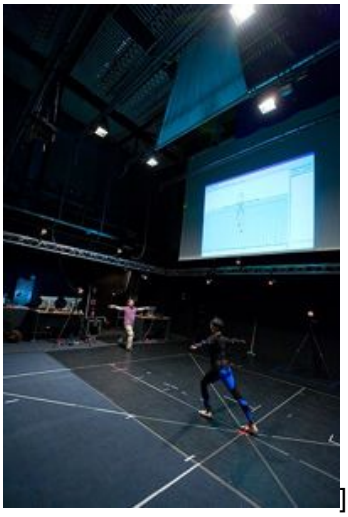
5 Head	LFHD: left front head RFHD: right front head LBHD: left back head RBHD: right back head ARIEL: top of the head
4 Shoulders	LFTShould: left front shoulder (clavicle) RFTShould: right front shoulder (clavicle) LSHO: top of left shoulder (place on bone) RSHO: top of right shoulder (place on bone)
4 Upper Arms	LUPA: middle of left upper arm (not on muscle) LELB: left elbow (find the part of the bone that does not rotate) RUPA: middle of right upper arm (not on muscle) RELB: right elbow (find the part of the bone that does not rotate)
4 Lower Arms	LFRM: left forearm (place directly on ulna) LWRIST: top of left wrist just before rotation point

	RFRM: right forearm (place directly on ulna) RWRIST: top of right wrist just before rotation point
4 Hands	LTHUMB: base of left thumb LPINKY: base of left pinky RTHUMB: base of right thumb RPINKY: base of right pinky
5 Upper Back	TopSpine: top of the spine MidBack: on spine where the ribcage ends STRN: centered on sternum LRRShould: left shoulder blade RRRShould: right shoulder blade
7 Lower Back / Pelvis	Root: the base of the spine LFWT: left front waist RFWT: right front waist LBWT: left back waist RBWT: right back waist Pelvis: just above the waist, skewed to one side LowerBack: just below the middle of the back, skewed to one side
4 Upper Legs	LTHI: outside middle of left thigh LKNE: outside of left knee RTHI: outside middle of right thigh RKNE: outside of right knee
4 Lower Legs	LSHIN: left shin LANK: left ankle RSHIN: right shin RANK: right ankle
6 Feet	LHEE: left heel LTOE: left foot just before big toe starts LMT5: outside of left foot where toes start RHEE: right heel RTOE: right foot just before big toe starts RMT5: outside of right foot where toes start



Range of Motion

A Range of Motion is a series of movements captured that are used to “teach” MoCap systems and related software the behavior and movement of a particular set of markers when attached to a moving figure. This range of motion always starts and ends with a T-Pose and is built up over 5-6 individual takes that increase in complexity in the construction of a template. Each Range of Motion is inextricably unique to every individual performer and while based in the first stages on a prescribed set of movements, needs must adapt at the later stages to match the individuals movement style and/or limitations and the specific characterized movement of the character to be driven and captured

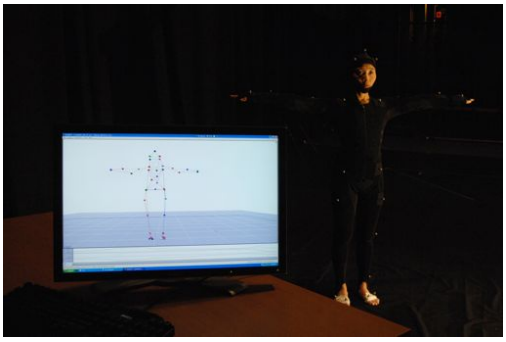


]

Template

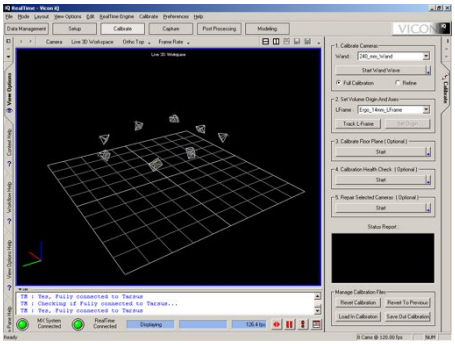
A template is created when the individual markers within a particular marker set have been allocated properties in relation to each other. Sometimes also called an object, the template allows for the individual markers tracked in 3D space to be allocated specific relational characteristics that form a cohesive set or object. A template once created can then be used to drive/control an avatar that in turn can drive a character.

The first phase of creating a template is to “capture” a performer in a T-Pose. This is achieved by recording the actor in the pose for 10 seconds. This static capture is then used to label a marker set and create a set of relationships between the individual markers. Once the relationships have been established the first phase of a template can be individually named and then used to create what is known as a more robust template using a series of Range of Motion captures. It is this individualized template that that is unique to each and every performer and needs to be reestablished at the beginning of each capture session. Put simply, the template of today will not necessarily work as the same individuals template on subsequent days depending on marker placement, posture etc.



Capture Volume

The motion capture volume is the amount of 3D space that the Mocap system can see. This translation of screen based 3D space to physical space is determined by the placement and settings of the capture devices (cameras) and their distinct relationship with each other as separate units. Depending on the capture that is being undertaken the volume will be adjusted. These variables could include: the amount of objects to be captured, the nature of the performance that is to be captured or the physical properties that are required in the space for performers to interact with. On this point it is worth noting that if in effect a particular character needs to be captured sitting at a desk writing, or climbing a rope then the best way to achieve this is to physically have them sit at a desk or indeed climb a rope, remembering that it is only their movement in space that is recorded and not a visual image of the physical object. The establishment of the volume is a vital early step in the profilmic setup as any character or object performing outside of this volume in whole or in part will either not be captured at all or their individual template will break and turn into an unmarked data stream or cloud of ghost markers.



Cartesian Co-Ordinate system

Two-dimensional space is traditionally represented with an X (width) and Y (height) axis, Performance Capture environments and 3D animation adds depth (or action) to this equation and this third parameter is known as the Z-axis. This three-point reference system is known as the Cartesian Coordinate system. It has been somewhat re-appropriated in the Animation industry and is generally used to define many properties of a 3D object or character including position, rotation and scale. As a means of plotting an object in 3D screen space it is a vital tool to exactly place an object in a precise location. A television term often associated with the Z-axis is its reference as the action axis, or depth axis. It is the axis that is manipulated the most in screen-based work to give the impression of scale within the 2D frame and to suggest a line of action for the viewers gaze.

Rotoscoping

Rotoscoping is an animation technique in which animators trace over live-action film movement, frame by frame, for use in animated films. Originally, pre-recorded live-action film images were projected onto a frosted glass panel and re-drawn by an animator. This projection equipment is called a rotoscope, although this device has been replaced by computers in recent years. In the visual effects industry, the term rotoscoping refers to the technique of manually creating a matte for an element on a live-action plate so it may be composited over another background.

Avatar

Avatar as used for a computer representation of a user or character, is a term that dates at least as far back as 1985, when it was used as the name for the player character in the *Ultima* series of computer games. The use of *Avatar* to mean online virtual bodies was popularised by Neal Stephenson in his cyberpunk novel *Snow Crash* (1992). In *Snow Crash,* the term *Avatar* was used to describe the virtual simulation of the human form in the *Metaverse*, a fictional virtual-reality application on the Internet. Social status within the Metaverse was often based on the quality of a user's avatar, as a highly detailed avatar showed that the user was a skilled hacker and programmer while the less talented would buy off-the-shelf models in the same manner a beginner would today. Stephenson wrote in the "Acknowledgments" to *Snow Crash*:

> "The idea of a 'virtual reality' such as the Metaverse is by now widespread in the computer-graphics community and is being used in a number of different ways. The particular vision of the Metaverse as expressed in this novel originated from idle discussion between me and Jaime (Captain Bandwidth) Taffe...The words 'avatar' (in the sense used here) and 'Metaverse' are my inventions, which I came up with when I decided that existing words (such as 'virtual reality') were simply too awkward to use...after the first publication of 'Snow Crash' I learned that the term 'avatar' has actually been in use for a number of years as part of a virtual reality system called 'Habitat'...in addition to avatars, Habitat includes many of the basic features of the Metaverse as described in this book"

Today the term Avatar is widely used as a generic descriptor for a digitally animated and manipulated character. No longer solely based in the gaming sphere of real-time interaction it is now accepted in many circles as an uncharacterized digital “actor” that can be prescribed the functions and capabilities of a characterized biped in the live performance, film and gaming industries.

3D space

While it is widely accepted and understood that the work we undertake in traditional performance spaces employ the standard rules of the three dimensions, as soon as the physical 3D space is transformed into the Computer Generated Space a higher level of understanding necessarily comes into play. So a traditional understanding of the classification of 3D space and/or objects would be something similar to:

An acceptance of an object having parameters that can be measured in the basic terms of length, width and height

As soon however as we begin to translate this basic depth perception understanding into the mediated, screen based or digital environment (that essentially replicates 3D space), and as soon as we are “capturing” a performance or action to be translated into one of these representations of 3D space, (keeping in mind that the screen is essentially a 2D environment containing only a vertical and horizontal plane) we start to enter into an interrogation of the 4D space which adds into the equation time.

4D space

A useful application of dimensional analogy in visualizing the fourth dimension is in graphical projection. A projection is a way for representing an n-dimensional object in $n - 1$ dimensions. For instance, computer screens are two-dimensional, and all the photographs of three-dimensional people, places and things are

represented in two dimensions by projecting the objects onto a flat surface. When this is done, depth is removed and replaced with indirect information. The retina of the eye is also a two-dimensional array of receptors but the brain is able to perceive the nature of three-dimensional objects by inference from indirect information (such as shading, foreshortening, binocular vision, etc.). Artists often use perspective to give an illusion of three-dimensional depth to two-dimensional pictures.

Similarly, objects in the fourth dimension can be mathematically projected to the familiar 3 dimensions, where they can be more conveniently examined. In this case, the 'retina' of the four-dimensional eye is a three-dimensional array of receptors. A hypothetical being with such an eye would perceive the nature of four-dimensional objects by inferring four-dimensional depth from indirect information in the three-dimensional images in its retina.

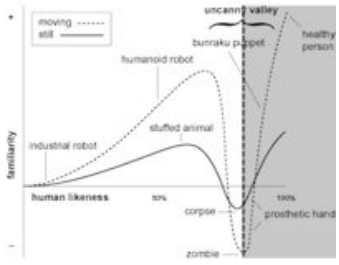
The perspective projection of three-dimensional objects into the retina of the eye introduces artifacts such as foreshortening, which the brain interprets as depth in the third dimension. In the same way, perspective projection from four dimensions produces similar foreshortening effects. By applying dimensional analogy, one may infer four-dimensional "depth" from these effects.

Stereoscopic Projection

Stereoscopy, stereoscopic imaging or 3-D (three-dimensional) imaging is any technique capable of recording three-dimensional visual information or creating the illusion of depth in an image. The illusion of depth in a photograph, movie, or other two-dimensional image is created by presenting a slightly different image to each eye. Many 3D displays use this method to convey images. Stereoscopy is used in photogrammetry and also for entertainment through the production of stereograms. Stereoscopy is useful in viewing images rendered from large multi-dimensional data sets such as are produced by experimental data. Modern industrial three-dimensional photography may use 3D scanners to detect and record 3 dimensional information. The three-dimensional depth information can be reconstructed from two images using a computer by corresponding the pixels in the left and right images. Solving the Correspondence problem in the field of Computer Vision aims to create meaningful depth information from two images.

The Uncanny Valley

The uncanny valley' is a hypothesis introduced by Masahiro Mori in 1970 that initially referred to robotics but has now been appropriated to refer to computer animation. It suggests that initially as the artificial becomes more realistic and humanoid, the more empathetic response human beings will have. However, when an almost realistic point is reached, the response will turn into repulsion. It is only once the artificial then becomes not at all distinguishable from the real, having crossed the uncanny valley, that the response will become empathetic again and may even approach a human-to-human empathy level



Human emotional response plotted against the anthropomorphism of a subject, according to Mori's hypothesis

180-degree rule

The 180° rule is a basic guideline in film making that states that two characters (or other elements) in the same scene should always have the same left/right relationship to each other. If the camera passes over the imaginary axis connecting the two subjects, it is called crossing the line. The new shot, from the opposite side, is known as a reverse angle.

30-degree rule

The 30° rule is a basic film editing guideline that states the camera should move at *least* 30° between shots of the same subject. This change of perspective makes the shots different enough to avoid a jump cut. Too much movement around the subject may violate the 180° rule.

As Timothy Corrigan and Patricia White suggest in *The Film Experience*,

"The rule aims to emphasize the motivation for the cut by giving a substantially different view of the action. The transition between two shots less than 30 degrees apart might be perceived as unnecessary or discontinuous-- in short, visible." (2004, 130)

Following this rule may soften the effect of changing shot distance, such as changing from a medium shot to a close-up.

270 degree framing

In theater, a thrust stage (also known as a platform stage or open stage) is one that extends into the audience on three sides and is connected to the backstage area by its up stage area. The 270 degree frame is widely used and employed in modern theatre practice and requires a particular style of acting, performance and direction that allows for scenographic elements and particular blocking to be viewable from all degrees of spectatorship. The 270-degree frame can also be employed in filmic space by the use of a camera array (multiple synched cameras) as employed by the Wachowski brothers in the Matrix trilogy and by Justin Lin and others in films like the Fast and the Furious. The key to the success and safety of the 270-degree frame is the upstage area of the film/performance space that remains a stage for performance.

The Omniscient Frame

The Omniscient (or global) Frame is a mechanism that can be employed in the capture of live performance for filmic, game and theatrical production. When performance is captured in a mocap environment, dependent on the established capture volume, all framing decisions can be made after the capture (in game or film environs) or during the capture real time in live performance. The Mocap system (essentially capturing a point cloud of movement globally that has in turn been made into an object in 3D space) allows for all framing decisions to be achieved during or after the event. Unlike traditional film making or the staging of performance where all of these intentions necessarily need to be confirmed by the director in the production or rehearsal stage in the global capture of performance these decisions can be made after the shoot, in several different permutations or indeed by the end user depending on the user interface. This in and of itself presents a unique challenge to the maker (the director/animator/actor) in that there is no specified performance frame intention apart from a direct concentration on the actual scene.