Chapter 2: Motion Capture Process and Systems

To appear in Jung, Fisher, Gleicher, Thingvold. "Motion Capture and Motion Editing." AK Peters, summer 2000.

By Ron Fischer, The Digital Actors Company

This chapter gives an overview of Motion Capture and various technologies used to do it.

Motion capture is a complex technology that deeply affects the aesthetics of the CG art form. Because of its complexity and the speed at which profitable production proceeds, it is a challenging process to keep under control. Stories of the failure of this work are readily available, ranging in cause from technological difficulties to the more common "operator error." Rather than build up a museum of horror stories we'll teach good procedures and specific problems to avoid. Fear-mongering does not serve anyone well.

We distinguish "motion capture process" (the topic of this chapter) from "motion application process" (the topic of later chapters). The "capture process" obtains motion data from the performer, and the "application process" retargets that motion data onto an animated character. The connection between the processes are the two skeletal models in the computer. One model, the performer model, is some representation of the human performer in the capture volume. The second one, the target model, might be a computer-animated figure onto which the motion is applied.

In the next section, we'll discuss planning and executing a motion capture production job. Most aspects are covered, from breaking down the work, to casting talent, to delivering the results to a client.

The following section introduces a practical model of motion capture technology. Several motion-capture technologies are described in terms of this "generic model." We show how these systems create raw motion data in "sensor space" and this is converted into usable motion data. Relative strengths and weaknesses are covered with an eye toward understanding which technology should be used for a given type of capture.

Production Jobs and Motion-Capture

What's the right way to do a production job and is motion capture the tool for it? One could say, if the client is happy with a result then the production was executed properly. However, clients are often forced to use cheaper means or change their mind about what they want. In this latter case Production becomes an exploration of alternatives. When searching for alternatives it's easy to mix choosing a result with how to achieve it. It's easy to get excited about new technology, seeing its advantages but ignoring important limitations in a headlong rush for "the new-new thing."

Motion-capture is often suggested because there is a need for massive amounts of motion data within limited budget and time. Hand animation might "break the bank," by taking too long or costing too much. Martial arts video games were an early example of this. It was not practical to hand animate hundreds of complex full body motions within the budget of a typical video game. In addition, there was publicity value in saying that a top martial artist's movements had been digitized. A similar example has been digital extras for crowd scenes. In both of these cases, the captured motion may be acceptable with a larger amount of error than that produced by hand animation.

However, within the motives for motion-capture's use, lie potential sources of failure. Productions that have limited time or budget imagine that new technology is a "plug and play" replacement for another more expensive technique (typically cel animation). It's critically important to recognize that:

1) Motion capture is not a transparent replacement for another animation technique. The **processes and results will be different**.
2) The differences in process and results represent potential **risks against achieving success**.
3) It is possible that a client may be trying to **close a time and budget gap that cannot be crossed**.

A client may imagine a grand project in terms of the excellent qualities and expressive characteristics of hand drawn cel animation. Then, when budget or schedule doesn't allow for cel animation, either, or both, of computer animation and motion-capture are suggested. However, using motion capture to create animation is significantly different from traditional computer animation and vastly different from cel animation. While the steps in the work processes are similar, each step is different in important ways. E.g. for **traditional CG animation**:

- Hire good CG animators
- Build character geometry
- Rough out animation
- Create skeleton
- Connect skeleton to geometry
- Animate & fix problems
- Render to completion

But for **motion capture animation**:

- Hire good CG artists
- Plan out the animation
- Determine approximate character requirements
- Audition & cast talent
- Rehearse & fix problems
- Capture motion
- Process motion data onto a skeleton & fix problems
- Build character geometry
- Place skeleton inside geometry & fix problems
- Begin to animate again, either manually or with further capture sessions
- Repeat for all shots until completion

Motion capture is also a different work process for Directors and Talent.

The Director can guide motion capture on a stage. He can talk to the performers and watch them move. Together they can critique the results and may discover better ways of performing a scene. While it is possible to interact deeply with an animation artist, the style is different. An animation artist must, in some sense, *be the actor* as well as a talented digital artist. This is a rarified skill. In addition, while an animator can render scenes quickly it still requires more time and is less direct than talking to people who are performing movements. In some sense, the Director is working with a "user interface" (actors creating a performance) with which he is familiar.

Motion-capture can be called for when a particular actor's characteristic motions are desired for a production. A walk, a wink or a nudge, if an audience recognizes it, the motion has value to carry a scene. Again, while an animator could create these motions from reference video of an actor, the actor may fortuitously string together many small motions that add up to a characteristic performance.

A few projects that started as traditional CG animation have been saved by motion-capture. This is an "about face" which can only succeed by accepting that the process and results will be vastly different. Such a change results in an entirely new project.

It is misleading, as has been suggested by some authors, that animation by motion capture can only reproduce satisfying motion on a one to one basis, e.g. of humans on humans. This is true for specialized action, i.e. only the motion of a martial artist looks convincing as the motion of a kick-fighting CG character. However, *acting ability* allows for a large range of motion adaptation. An example of this has been the increasing finesse of special effects artists such as Rick Baker [cite: CineMagic article] to both produce a visually convincing gorilla suit and then to act the role of a gorilla. Another example of highly adapted motion is real-time CG puppeteering of full-body characters. As an example, at MediaLab Paris a motion capture performer responded to the Director's call for a "cut" after a performance error by falling to the ground. On screen, the CG character humorously folded into a small pile of limbs in an odd expression of embarassment, fully as the actress intended. In this case, a human was performing a human character with unusual proportions compared to the motion actress. However, because the actress had many months of experience with direct real-time feedback she performed her own motions in ways that looked correct when applied to the CG character. An example even further along the spectrum toward pure puppeteering was the multi-legged crab character Frigate designed and executed at Protozoa for Intel. In this case, the human actor's leg motion was applied to all six of a crab's feet.

[image: kick-boxing motion capture and character]
(a)
[image: Clio character from MediaLab Paris]
(b)
[image: "Red" from Driftwood by Protozoa, with multi-leg connection image]
(c)
This sequence of images illustrates the range of acting that can be motion captured: (a) one to one mapping, (b) somewhat puppeteered human motion on unusually proportioned characters, (c) performances fully adapted to puppeteering a non-human character such as a multi-legged crab.

The expectation of one to one motion application comes from two sources. First, much of the technological drive behind optical motion capture has been focused on the "holy grail" of photo realistic synthetic humans. While this will eventually be achieved, realistic motion is only a part of the whole market of projects. Second, much early optical motion capture work was done more or less blind, in a kind of "batch mode," where the action was recorded first and then viewed later (sometimes much later). In this regime, motion actors could only imagine the weight of the bodies they were intended to inhabit through concentration, practice and guesswork. Mechanical and Magnetic capture systems (and recently optical systems) can provide real-time motion data. Moreover, this motion can be applied to a simplified real-time version of the character. By providing this feedback to the actor during both the rehearsal and production shoots much better acting will be achieved.

The difference between success and failure? At the core:

Using any complex technology in a production requires planning to reduce risk.

Therefore, with a technology as potentially complex as motion-capture, getting a client results means:

- Clarifying what they want beforehand
- Disclosing the advantages and limitations of motion capture to achieve that result
- Presenting them a process, cost and schedule needed to get that result
- Planning to test the process with them and revise it if necessary

Projects vary enormously in complexity and budget. The following methodology is for large-scale animation work, as for many scenes in a film. Motion capture for games is discussed as a delta from this process. It's better to do everything listed here, but less of it, rather than omitting steps. Time spent upfront reduces by several times the work expended "down the line" to repair problems.

Here is a suggested organization of the motion capture process based on the recommendations of two industry practitioners [cite: Alberto's book and Demian's draft course outline]. Subsequent sections discuss each phase in detail.

Pre-production planning
- Understand the desired results
  - Gather relevant material (Storyboard, Production boards, etc.)
  - Evaluate stability of the design described by the material
  - Consider working relationship with client

- Extract info about each potential motion-capture shot (Shot Breakdown)
  - Characters and dialog
  - (Inter) action, rigs, props, location, size, movement, etc.
  - Shot timing w/handles
  - Boundaries of performance
  - Framing
- Determine motion-capture requirements for each shot
  - Rough character setup and marker setup
  - Head/Tail, neutral poses, blending
  - World Axes
  - Units
  - Sample rate for type of action
- Evaluate shots for completion with motion-capture
  - Can it be performed or acted?
  - Does it fit in space?
  - Does it fit in time?
  - Number of characters?
  - Contacts? (char to char, char to prop, char to set)
  - Are the rigs feasible and safe?
  - Can data can be delivered in a compatible format?
- Audition and cast talent

- Initial Capture
  - First fitting of cast talent to motion capture markers, sensors, or suit
  - Construction of performer skeleton
  - Construction of character skeleton
  - Linking of skeletons, possibly with virtual mechanism
- Test & Rehearse (critical and often ignored)
  - Walk-through rehearsal
  - Proof of Principle motion capture shoot
  - Dress rehearsal

Production
- Preparation
- System Setup
- Calibration
- Marker placement
- Capture
- Tracking (post-processing motion data)
- Output
- Delivery

Pre-production Planning

Each step in pre-production planning is described here.

Understanding Desired Results

At some point in production it will be suggested, "Why don't we use motion capture?" This is a special moment. A responsible party must review production materials such as script, storyboard, etc. and perform a "breakdown" to determine whether motion capture is the right tool for the job. Ideally, all of this happens in pre-production, but even if not, it is critical to take planning seriously. As a rule, the better and more detailed the pre-production planning, the faster and cheaper the production work will be... or the more profitable for the company being paid to do it.

The work of planning begins with understanding the project thoroughly. This means gathering all of the relevant materials like scripts, storyboard, production drawings (characters and sets), animatics, dailies on video, scratch audio tracks, etcetera from the client. These materials are often considered highly proprietary and you may be legally bound to keep them secure. Take detailed notes, make photographs and keep them organized e.g. in a binder. Meet with others working on the project to improve your understanding of it. This includes how deeply the material is "in revision." These early meetings are also an excellent time to present the basics mentioned previously, regarding how animation by motion capture is different from other means.

Producers and Directors may know precisely what they want, and have specified it clearly, or they may know and not have specified it, or they may simply not know. When the client is unsure they may accept proposals and then change their mind after receiving (what were thought to be final) results. It's also typical that some scenes will be "locked down" and others still in revision. These are all reasonable relationships to the client and the project, none are wrong and none are right. However, each requires vastly different amounts of effort and therefore cost. Its important to plan for testing and expending more effort on shots that are vaguely specified or for a client that enjoys exploring possibilities as part of their creative process.

The next questions surround, “To whom will the captured motion go?” These describe where motion-capture fits into the client's production process.

- Will the data be used directly or as a reference ("roto-capture")
- Who on the client's side will receive and check the captured motion files? (Signs off on a delivery)
- Which team will create art with the motion? (Are likely to call with questions)
- What animation software is being used?
- What format and kind of data should be delivered? (E.g. translation or rotation)

If the client intends to hand animate their work, but wants to use motion capture primarily to lend additional realism, e.g. weight and inertia, then they may only want to motion capture for reference. This puts fewer constraints on the quality of the delivered results.

It's important to establish an approval process with the client for signing-off on delivered motion data. Ideally, one person at the client's facility will be responsible for looking at all incoming motion files and logging them into the client's production. This means that they should load the data into animation software and watch each move. It is inevitable that files will become lost or corrupted due to the usual set of human and machine errors. The sign-off process ensures that such problems are caught early and can be dealt with farther away from an important deadline.

The animation software being used by the client can provide important hints about what the delivery format should be.

The above are the outward looking questions, which clarify the world in which the work will be applied. On the one hand, their answers may seem obvious. However, they reveal where the motion capture team's work connects with other people in the highly collaborative process of production. If the answers are not agreed upon up front, then its not clear who will decide them, nor whether anyone has the right to do so later during production.

With the initial client discussions completed, it will be clear why motion capture was seen as desirable and useful. Existing materials describing the work should be in hand, to whom the final data will be delivered is understood, as well as what other constraints others might be put on the process. Of course, the production process is infinitely mutable within the bounds of finance and ego. Nevertheless, by getting the client's production team to answer these questions up front the motion capture team will have a place to start. They will also have a better basis to understand the impact of inevitable changes once the production is underway.

Shot Breakdown

The next step in planning is to examine the information describing the project and "break it down" into potential motion capture shots. Usually the client will provide this in some fashion. In larger projects it will arrive in the bid book from the Special Effects Producer.

For each shot, information relevant to the motion capture work must be extracted. This will be used to plan how to do the work and, more importantly, whether its possible at all. Note that we both extract information relevant to motion capture, and then break that information out into motion capture shots.

There is lots of information to record about each motion capture shot, preferably on a standardized sheet such as this one:

[illustration: shot planning form]
Motion capture shot planning form
(Courtesy of Performance Capture Systems).

The client's storyboards and scripts will be organized around camera shots or takes. Each take in front of the camera may require several separate motion "takes" to be shot and combined. Be sure to keep the distinction between a "take" in front of the **motion capture equipment** and a "take" in front of the **motion picture camera** clear when writing down information about shots, e.g. use prefix letters like "MC." Be sure to annotate "MC" shot information with cross references to the client's "camera storyboard" shot numbers.

If a project is bid on a "per (motion capture) shot" basis clients sometimes attempt to reduce their cost by combining a series of motions into one contiguous shot. In a combined shot each motion influences the next, which can make the motion look "wrong" in isolation. Additionally it is harder to perfect and capture a complex motion than a simpler one. It's often better to isolate motions and shoot them separately than to combine them.

The motion to be captured is usually driven, triggered or timed by external media. It could be a voice or video track, whose time code is being written into the captured motion data. This is typical when the animated characters are dancing to music or will be composited over background action that has already been shot (a background plate, usually provided on video from the production's collection of dailies). It's likely that the performers will rehearse the motion repeatedly while looking at the background plate on video or listening to the sound track. This timing material, along with suitable playback equipment, must be prepared for the motion capture shoot.

If external media is not available for timing, the Actors and the Director can choreograph the motion of a capture shot to an arbitrary soundtrack, using it purely as a timing aid. The soundtrack may be a completely unrelated piece of music that matches the energy level of the motion in the scene. Of course, if the music is lively and keeps playing in between takes Jeff Thingvold of Lambsoft has said, "Just try to stop them from dancing."

[photo: storyboard frame]
Single frame from action sequence storyboard.

The most basic information to extract about each shot on the storyboard is the names of the characters and props to be captured, along with dialog and actions. These are the heart of the acting in the shot.

Motion capture is often done for action sequences. Rigs needed to "fly" the character, hurl them into the air etcetera, should be discussed with the stunt supervisor and listed in the planning documents. Particularly fast motions or those with fast contact (or impacts) should be noted as well, because these may need higher frame rate captures.

Characters relate to an equally important environmental space (or scenery elements) in a shot. While it can be as simple as a ground plane to walk on, the environment often modifies the motion significantly. For instance, a simple ground plane could be made of ice (or polished Plexiglas walked on in silicone treated socks). Then, this "simple ground plane" modifies the performer's walk into a cautious cat creep with occasional slips. A performer will also walk differently barefooted or wearing shoes.

There will be props in the environment, e.g. heavy blocks to lift, walls to lean on, a shotgun blast simulated by a cable pulling on a harness. Every prop that has contact with the character, and every piece of the scenery that they touch, must be recorded. As well, every (inter) action should be listed in the order it occurs. It often helps the animators if reference points are captured on these props, especially near where an interaction takes place.

Another important point is estimating the length of each motion capture shot. As for film and video, the motion capture shots should include an allowance for "handles;" a little extra time at the beginning and end of a motion. During editing this allows the action to be shifted forward or back as needed.

One of the most important pieces of information to help determine feasibility of capture is in how large a volume the motion takes place. In addition, is the "action volume" oddly shaped? Or are there obstructions to the visibility of markers? Sketches of complex motion should be made to aid in planning.

Finally, where is the camera? Where will the motion-captured character be within the camera's frame of view? How close will the camera be to the character? Which side of the character will be seen? The actor who performs the motion is greatly assisted knowing to what he's playing. Even if only to know not to look in that direction, or to be careful about motions visible from a particular side. Again, for complex scenes it is helpful to sketch in the camera location on diagrams.

Characters should be listed as foreground or background. Often a foreground character will be hand animated or captured data will only be used as a reference. The Director handles characters that are telling part of the story. Their motion is likely to be revised repeatedly during production [footnote: the editing chapters to follow will say more about the tools used to do this]. When motion capture is only used for an animator's reference, it is called "roto-capture." The word is related to "roto-scoping" a technique from cel animation, where an artist traces the outlines of a real actor from reference footage shot for that purpose. Roto-capture may be useful for foreground characters, or for setting up motion cycles on background characters. In the film "Titanic," many of the synthetic characters strolling the decks of the ship had movement cycles based on roto-capture data.

Determine motion-capture requirements for each shot

Once the basic information about the motion capture shots has been extracted, that is, requirements from the client, the next step is to decide on how these requirements affect actually doing the motion capture. In effect, this step is sketches out how the motion capture will be performed to give good data.

The types of data provided by different kinds of motion capture equipment need to be considered with respect to how they will control the final character.

The raw data of an optical system is a set of marker positions, in effect a point cloud around the body of the performer. These point clouds can be used directly to animate the final character (translation data), or they can be solved into rotations on a skeleton and those rotations applied to the character (rotations on the joints of a hierarchic skeleton). Passive optical markers should be visible to three cameras [footnote: in active marker systems, two cameras suffice, usually both are mounted inside a single bar-like scanner. Passive systems require three markers to avoid marking "ghosting" (see optical section)] throughout the motion to be tracked accurately. Optical markers must be attached to the performer in a way that minimizes shifting and wobble during action.

Magnetic systems don't need line of sight, and provide both orientation and position as raw data. The actor's motion is performed relative to a large transmitter. Often the motion must remain on one side (i.e. in a single hemisphere) relative to the transmitter. Areas close (but not too close) to the transmitter usually provide more stable data, while the fringes of the capture volume (away from the transmitter) may be much noisier.

Mechanical systems have rotation and rod length as their raw data, sometimes only known internally in the driver software. This capture equipment is an exo-skeleton [footnote: "exo" from the latin for outside / external as opposed to "endo" or inside / internal] of jointed rods, fastened around the performer's body and designed to be as lightweight as possible. A key point in planning motion with these systems is that rolls and falls are generally not possible without risking significant damage to the machine.

Finally, as a general rule faster motions or harder contacts require higher sample rates, e.g. to capture the moment of contact between a karate kick and the torso of the victim. Higher final sample rates leave less room for filtering or supersampling to smooth out noise. It is critical to test setups for these fine and fast motions to be sure that they do not produce "smoothed out mush."

For human characters such as sports figures or "background people" the design of the character's jointing and the arrangement of equipment on the motion actor can follow standard setups provided by the equipment vendor (or the special effects house receiving the data). However, when the character is non-human, e.g. has extra joints or limbs, virtual mechanisms are used to connect the two [footnote: see the bird example in the following chpaters]. Likewise, situations where the equipment is affected by the motion have to be noted and planned around (i.e. optical markers may be obscured or mechanical exoskeletons may be exposed to impact damage).

For optical systems, the arrangement of equipment on the performer is called "marker setup." For magnetic systems, it is "sensor placement." For mechanical systems, there is usually only one setup of the suit. There is more information about creating setups in each of the motion capture technology sections below.

Overall, for each scene it is important to determine if the markers, sensors or exoskeleton can provide enough information to control the character in a reliable and expressive manner.

Each distinct setup should be sketched and named. The setup name should be listed with scenes in which it will be used.

[diagram: marker setup]
Sketch of a marker setup.

While it is more desirable for actors to perform together, it may be necessary to capture motion in multiple passes. If two characters do not interact significantly, it is possible to plan a shot in two passes, one for each actor. Having the first performance available as playback, perhaps projected on a wall visible out of the corner of the actor's eye, is an elaborate but reliable way to synchronize the passes. Some reasons for doing multiple passes are: needing more markers on a single character than your system can handle, being able to edit together the best two passes into a single motion "take" or simply having only one actor.

A typical problem involves two characters interacting in an optical motion capture studio. If they have lots of body-to-body contact it would be easier to capture them separately because the markers will be obscured by each other's bodies. However, this may not work because the contact between the characters makes the motion irreproducible; different in each of the two passes. This means more, redundant, markers will be needed, possibly too many to capture in one pass. Therefore, the problem goes around. Setups such as these should be part of "Proof of Principle" testing.

If the motion being captured is for a game then the organization of the motion capture shot list will be a state diagram, showing motion cycles with neutral poses that lead to other motions. In this kind of motion capture planning (and testing) of entries and exits to the neutral pose are critical. E.g. it may not be possible to capture a natural motion when a game character comes to a "dead stop" neutral pose at the end of a full out running cycle. In these cases it will be better to hand animate the transition, spending time finding a motion that looks right but would be physically impossible to perform (the magical domain of cel and traditional CG animation). Transitions between motions, cycles, and blends should all be listed on the motion shot planning sheets.

[diagram: a motion cycle state table]
A motion-cycle state table.

Some 3D software requires the Y-axis face upward, others allow this to be selected by the user. The coordinate system and the motion's orientation relative to it should be decided. Sometimes a motion may be captured facing a different direction than it is used, as when a motion picture camera is mounted pointing up from the floor to capture a "weightless" astronaut suspended on wires. In addition, it may help animators to have the origin of the coordinate system center on a prop or pre-determined position in the scene. Likewise, units of measurement are critical to interpreting the motion data. Planning these parameters avoids surprises for the animation team later.

Finally, sample rates should be suggested. Motion capture data can be altered in many ways, e.g. by resampling at a higher or lower rate or shifting data frames partial time units forward or back. However, this is akin to making a photocopy of the original data; it is possible to lose a bit of integrity with each copy.

It is best to choose a sample rate close to what will be needed based on the speed of the motion. High sample rates make larger motion files. Because larger amounts of data make loading and processing a chore it is best to try for the lowest sample rate that doesn't lose motion sharpness. This is related to choosing a taking resolution with your digital camera. Lower resolutions allow more pictures to be stored and downloaded but each is fuzzier. High resolution provides better detail, but becomes much more difficult to transfer, store and process.

While post-processing motion data doesn't always give a degraded result (as photocopies sometimes look better than the original) it is important to *always have the raw data available to return to*. E.g. for optical capture it is possible to animate directly from the marker translation data if that is sharper than the processed skeletal rotations, or it may be possible to use improved software to process the raw data in a new manner. Often motions rich in contact with props or other characters are easier to animate using translation data.

Evaluating shots for completion with motion-capture

With a breakdown of the client's script in hand (either provided by them or created), it's time to evaluate whether each motion capture shot can be completed. By this time the client should understand (from the initial meetings) that animation by motion capture is different from the other processes. Proceeding from that, their expectations should be well understood before this evaluation starts. This is also only the first level of evaluation. Further revelations will come during rehearsals, and the "Proof of Principle" shoot.

The first question to ask is whether the motion can be acted or simply performed at all. This starts with the question of whether the required motion is realistic or more of a CG cartoon-style. Beware of a need for the motion provided by traditional cel style animation, as some of these motions can't be acted, e.g. squash and stretch or the kinds of arbitrary dramatic poses that can be achieved in a hand-drawn character [cite: classic cel animation book chapter]. Of course, motion outside the human range of joint rotation isn't possible. Finally, while puppeteering techniques work for some CG cartoon-style characters, they may not be adequate for an unusual creature such as a photorealistic centaur [footnote: combined human/horse from Greek mythology]. The more puppeteering the motion requires to work, or the more complex the mapping from performer to character, the more "proof of principle" testing is required.

Does the motion fit into the capture volume that can be created by the equipment? Large area motions, such as running jumps, may not be possible with a magnetic system in a noisy capture space (one containing significant amounts of nearby metal or magnetic fields). The motion of a rider on an actual motorcycle racing around a track would be impossible to capture with anything but a mechanical suit. Check the shape of the volume that encloses the motion against what the equipment and work area can handle.

Long motion shots can require a lot of studio time before an "A" class take is captured. These are better broken down into shorter motions. Likewise, motions that are extremely short and fast could be difficult to perform and capture precisely. It can also be difficult to maintain continuity between motions captured using many passes. If it is possible to "plan around" such difficult shots to eliminate them, it is usually worthwhile.

The number of characters and markers are also important to consider. As of this writing it is technically feasible to capture about two characters in real-time using optical equipment and four or so in non-real-time. However, as the number of interacting characters goes up, and the number of markers rises, so does the chance for occlusion. Capture effort goes up with each character added to a scene.

[photo: action capture]
An action shot being captured.

Contact in a scene, either between characters, between a character and a prop, or between a character and a set piece (like a wall) should be carefully examined for occlusion issues.

For extreme action, the cost and safety of rigs must be evaluated. While hand animation may take time, it could also protect an actor from exposure to unneeded danger or be cheaper than professionally staging a stunt.

Once the motion capture shot list has been analyzed the following will emerge:

- Impossible shots to be eliminated
- Very difficult or risky shots will be identified
- Number and configuration of stage setups
- Number and configuration of equipment setups on the performer

It is reasonable to opt out of impossible or troublesome shots that cause (e.g.) 40% of a production's cost. Accepting challenge is the spirit of the production business, but "reality wins," and taking on higher risks means seeing failure more often.

Motion capture shots that remain part of the production can then be grouped. Those that use the same stage or performer setups will be executed together for efficiency (just as movies are shot out of sequence to optimize costs). There is a palpable energy or momentum during any shoot which can be lost if too much delay happens. The "dead time" during reconfiguration should be scheduled across days or over breaks to keep production flowing.

Audition and cast talent

It is surprising that many motion capture productions do little to audition and cast talent aside from choosing acrobats or stunt people. Casting motion talent should be handled as seriously as it is for film or video. The emphasis however, is different.

In this type of casting the ability to act through body motion, not necessarily through facial and voice expression is key. Talent should have excellent body control and a highly developed sense of physical awareness. Generally, "the more professionally someone does something with their body the better the motion talent they are." [cite: Demian for this quote]. E.g. superstar athletes have practiced their moves to near robotic precision. Other people, such as dancers or costume puppeteers, who use their bodies as the tool of their profession, are likewise good motion capture talent.

In no particular order, people to consider for an audition might include:

- Stunt persons
- Martial arts experts
- Gymnastics experts
- Professional athletes (specific to their sport)
- Dancers
- Mimes
- Costume puppeteers (e.g. full body puppets or prosthetic makeup suits)
- Pro wrestlers
- Fight experts
- Theatrical stage actors
- Acrobats

Ideally, a single actor is cast for each character role. But even moreso than traditional film making, multiple actors may be cast for a single motion capture role. If several motion skills are needed, few actors can handle all of them. Stunt people for stunt action, dancers for ballroom scenes, and comedic mimes for humor are typical in film or motion capture. This is especially true in motion capture for games, where the game's multi-skilled main character may have a "main actor" plus others, experts in martial arts or gymnastics, who stand in for sequences demonstrating specific physical skills.

With standard resources for casting other camera talent, such as agencies or trade publications, care must be taken to specify and cast for the skills needed to perform the role. Aside from obvious needs like martial arts skill, the posted call should say that only the actor's motion (body or face) will be used, not their appearance. If the motion to be captured is energetic be sure to indicate that in some way.

Videotape the audition for later review. Take into account whether the performer has previous experience with motion capture. An experienced performer can considerably shorten the time to capture good motion by understanding the many constraints on the work, such as avoiding slipping markers, remaining within the capture volume, etc. The best performers will have enough experience to become partners in planning.

An interesting casting technique is to handle the first call casting under normal room lighting and the second "in silhouette." [footnote: e.g. using a bright light to cast the actor's shadow onto a white sheet] This focuses the reviewers exclusively on the motion of the actors, rather than their appearance or voices, which is otherwise difficult to ignore.

Several scenarios from the script should be used to audition the actors, without giving them pages to read. E.g. to perform scenes such as a bar fight or a dock worker working the actors would be given no other instructions than "you are in a bar and get into a fight over a pretty woman" or "you are a dock worker moving heavy crates and boxes around a shipyard." The actors should perform the scenes alone and unaided. They should then be asked to modify their performances, either beforehand or during the vignette. E.g. "suddenly a bottle is smashed on your head from behind" or "a co-worker hails you while you are lifting a box." Different emotional states should be included as well, e.g. happy, sad, angry, bored, drunk, injured, heroic, evil, sneaky, etc.

The largest market for motion capture systems outside of entertainment [footnote: indeed, the largest overall] is analysis of rehabilitative exercise after injury. If an actor has old sports injuries, such as knee problems, previous surgeries, back or skeletal problems it is likely that these will (for better or worse) show up in the data. Motion capture work is physically demanding, and this is an important part of the actor's disclosure during casting. The disclosure allows these challenges to be managed if the person is cast in the role, either through acting, interaction with the Director or by motion editing.

The talent should be queried about skin allergies w.r.t. materials such as marker or bandage adhesives, etcetera. If unusual materials are used in construction of props or costumes it may be wise to test for adverse reactions.

As for film, the actors for a single role should preferably be similar in weight and body dimensions. This means that less "fitting" to the final character will be needed between actors in a single role.

Unusual CG characters require that performers adapt their motion significantly so that it looks right. Such characters have complex connections between the capture skeleton and the CG character's skeleton (different proportions, more than 2 legs, reversed knee joints, etcetera). As mentioned previously, such characters can be successfully motion-captured *if the performer is given the time and feedback that allows them to inhabit the character.* Professional costume puppeteers, those that perform inside "creature suits" have the experience to handle this type of work. When done well this acting avoids the effect of "the man in the rubber Godzilla suit." Indeed, with CG and motion capture there is no need to hide the operator inside the character, significantly enhancing the effect.

For such characters it is critical to cast talent that has operated a costume puppet, similar prosthetic suit or (ideally) a CG puppet. Also critical is providing real-time feedback of the performer's motion on the character and sufficient time for practice. The "proof of principle" shoot described in the next section is essential in these cases. A fine line is trod here, which modified human motion must cross to be believable when applied to a cartoon or fantasy character. While many successes have been achieved at Protozoa, MediaLab Paris, etcetera, the client's expectations will determine whether motion-capture "puppeteering" produces an acceptable result.

Finally, remember that, as in film, it is possible to cast an actor and then discover that they are unsatisfactory later on. It's wiser to re-audition and re-cast another performer rather than have all results be equally mediocre. Talent with difficulty adapting to the work may also lengthen sessions and cause schedules to over run. Of course, if the client requires a particular actor the schedule should be adjusted based on how well testing proceeds.

Initial Capture

"Initial capture" is the first time each actor is "suited up" and motion measurements are made of them with the equipment. This means, putting on the motion capture "clothing," checking fit, freedom of movement, and capturing rest pose data. Likewise, if props are to be motion captured, their capture setup will be tested as well. Initial capture ensures that the markers, sensors or mechanical suit fit to the actor, don't inhibit motion and produce data. The resulting data is used to configure and size the performer's capture skeleton.

Optical and magnetic motion capture systems need a means of attaching sensors or markers to the actor. These systems are usually not sold with clothing to do this. Many motion capture studios have designed and built "suits" for this purpose. These outfits have Velcro areas for marker or sensor attachment. The suit should be fitted to the actor's body so that the markers or sensors don't slip. This is a source of error during capture.

[photo: performer adjusts markers]
Adjusting markers while suiting up.

Markers should be firmly connected to the suit so that they won't fall off or be knocked off during a roll or fall. Freedom of movement needs to be sufficient for the work ahead. Also, the suit should be checked to determine whether it can hold markers in all of the positions indicated by the project's marker setups. If not, another suit may need to be constructed. Finally, at least one spare suit should be available; motion capture is physically demanding and can tear up clothes like a six year old child on a new playground.

Like the others, mechanical systems need to be fitted to the performer. In this case fitting involves eliminating slippage and measuring limb lengths for input to the driver software. With these systems it is important to test and plan motion to avoid poses in which the suit binds or inhibits the performer's motion.

Ample supplies of materials used to safely attach things to humans should be on hand, e.g. ace bandages, athletic tape, spirit glue, etc. A well identified and separate supply of harsher fastening substances should also be stocked, e.g. hot melt glue, duct tape, staple guns, etc. This will allow rapid prototyping of new marker setups or means of reducing slippage should that be needed.

An "initial fitting" into the motion capture "suit" or equipment is similar to fitting an actor for an elaborate costume. In this case the "costume" has three layers:

- the actual motion capture clothing (or mechanical exoskeleton)
- the performer capture skeleton (data in the computer)
- the character skeleton (this is called "retargeting" and is covered later)

Aside from putting the actor in the suit, an initial capture of motion information happens at this session. This data must include a "rest pose." Optical systems, which have their own capture software, use this data to create a marker network and derive a performer skeleton. Magnetic systems are not provided with capture software and so the procedure varies based on which is chosen (see later chapters of this book). Mechanical systems require the operator either measure the performer or between points on the suit for calibration. Again, these measurements size the skeleton onto which the performer's motion will be captured. [footnote: the performer and character skeletons are considered separate unless they are *exactly* the same in both limb size and jointing.]

The information about the rest pose provided by the initial capture session goes to the technician making the connection to the final character. As we'll see later this connection part of the "pipeline" (ordered sequence) of "motion mappings" that the data goes through. The pipeline converts the raw information provided by the motion capture equipment into joint rotations on a model of the performer's skeleton and finally onto the target character.

The pipeline technician uses the driver software and applications of the motion capture system to derive the "performer skeleton" from the captured rest pose. A low resolution model of the character to be animated should have been provided by the client. After setting up the pipeline, the technician can connect this model to the rest pose. This will allow real-time previewing of the motion data during the capture sessions to come.

Puppeteering setups requires more complex connections to the performer's motion capture skeleton. Rest pose and sample motion data from this session are also used to create the first version of the puppeteering setup.

Test & Rehearse

Once the talent is cast it's time to rehearse the shots with the motion capture supervisor. Several times at least.

It's common to neglect rehearsing adequately before a motion capture shoot. This is unfortunate, as actors would never get on a film set without weeks of rehearsal. Even so, many clients enter a motion capture shoot with only a script read-through. Clients should be (repeatedly) informed that rehearsal is *their most important tool* for increasing the quality and decreasing the time needed to capture motion data.

There are three stages of rehearsal:

- Walk-through rehearsal

- Proof of Principle motion capture shoot
- Dress rehearsal

All rigs, props and scenery should be constructed before rehearsals begin.

The proof of principle shoot is a technical rehearsal, to show that each configuration of markers, sensors, props, rigs and performer to character connections will produce good motion. Several example shots should be selected for each distinct configuration. The captured motion should be applied to a test character. Regardless of whether this type of marker, performer, rig configuration has been captured before it should be tested again. Conditions may have changed since last use, and improvements may be realized during testing. Again, an assistant should take notes on materials that were needed but not present, problems that occurred, and solutions proposed. This rehearsals should be scheduled to leave time to resolve problems before the shoot. Members of the team should be assigned responsability to fix problems before production begins.

The walk through rehearsal has the talent act every scene under the motion capture director's supervision. Props and scenery are used, but no motion is captured. It helps to have an assistant take notes of the opportunites and problems found along the way. This rehearsal should be videotaped and made into a "video storyboard" for the motion.

[Photo: Videotaping]
Videotaping a dress rehearsal.

Finally a full dress rehearsal should be done while capturing motion. During this rehearsal, the talent wears all motion capture equipment and rigs. Each shot is acted and captured. However, the captured data should be considered "disposable." The emphasis should be on resolving problems rather than achieving a perfect performance. Again, an assistant should take notes, especially recording missing items and techniques that the director would like to use during the production capture sessions.

Production

All of pre-production to this point should have prepared the talent, client supervisors and motion capture team to efficiently carry off actual production. It should be possible to enter the studio on the day of production, calibrate the system, place the equipment on the talent, calibrate the performer and start shooting. There should be no need to search for things; they should be laid out beforehand within easy reach. If it isn't possible to get up and go, then the preparation was inadequate.

When one considers the amount of money a client spends during production, which can run as high as hundreds of thousands of dollars a day, the importance of the pre-production becomes clear.

Now, on to getting ready for and carrying out production...

Preparing the Studio

Preparation occurs in the days and hours before the client and talent arrive for the actual shoot. Consider this analogy: it is like preparing an operating room for surgery, all personnel and equipment should be neatly laid out and ready for the client who's depending on the success of the work.

[list: pre-shoot checklist]
A pre-shoot checklist
(courtesy Performance Capture Studios).

The areas that need to be prepared are:

crew training
stage
    cleaned
    props
dressing room
    motion capture suit cleaned and ready
    maintenance items for suit
motion capture control
    equipment powered, tested, orderly
craft service
    food and drink for talent, crew and client

Few things will raise doubts in a client's mind faster than a dirty and disorderly studio. Think about what your reaction would be if the facility or equipment you were renting was a mess. Enough said.

First and foremost, it is likely that somenew crew, grips, dressers, managers, etcetera will be joining the team. Work in a motion capture studio is not like typical production in many ways. The stage areas are less clearly defined, lighting is avoided, and moving calibrated equipment (such as cameras) is nearly a punishable offense. It is critical to hire experienced crew and practice with them during the previous rehearsal phases to be sure they're ready to work efficiently in a motion capture studio.

The stage area should be have props laid out for easy access by the grip crew, who should know where every prop is stored. If multiple camera setups will be used, tape marks should be prepared beforehand to indicate approximate positions. Measuring equipment should be ready, tape measures for setup and recording of prop positions on the floor, folding rulers, large calipers, etcetera for calibrating the equipment on the performer. If using magnetic capture, the transmitter should be locked down and the volume already calibrated for distortion cancellation [footnote: see the subsequent section on magnetic motion capture for more information on what this is and why it is needed].

The motion capture suit should have been dry-cleaned or washed before the shoot day. All equipment that is to go onto the talent should be clean and dry. Tape, Velcro, bungee bands, etcetera should be laid out and ready. There should be no need to prepare materials on the day of the shoot. Ideally, even tape strips should be pre-cut. Likewise, maintenance items for the suit, extra markers for optical capture, lubricant and cleaners and tools for mechanical systems, should be ready. Specifically for optical motion capture, marker balls should be cleaned or replaced if damaged. The reduced reflectivity of dirty markers causes tracking and sensitivity problems. Materials should be on hand to make the talent comfortable, e.g. towels for sweat, extra padding to fix chafing, ice packs, and support bandages.

The control area of the studio where the motion capture equipment is located should also be prepped. Power cables should be taped down to avoid being kicked out. Power (and data) loss during capture is an obvious nightmare. Manuals and equipment spares should be at hand. At least one telephone line should be devoted to the motion capture system operators. The telephone support number of equipment and software vendors should be at hand. The operators should have easy line of sight to the capture volume and the client's waiting area.

Finally, "craft service" is an item that should never be overlooked. Amenities like food and drink for the client, talent and crew should be ready. An assistant should check dietary restrictions.

Once preparations are completed, the motion capture supervisor should make a final walk-through, taking the viewpoint of talent, crew and client in assessing readiness. This phase ends with the crew and all equipment ready to work.

Setups for a series of Shots

Planning should have listed the motion capture shots according to how the stage is set up, so that shots using similar setups are all captured at once. This is similar to the way shots are filmed out of order when a movie or video is made.

While many projects may be planned so that only one setup is used, it's not unusual to change it.

There are three kinds of setups:

- stage setup, only props, rigs or scenery are moved
- performer setup, where equipment is re-arranged on one of the actors
- capture volume setup, where the motion capture volume is reconfigured

The grip crew prepares the stage using one of the planned configurations, including props, trampolines, landing pads. The stage can normally be reset quickly.

[photo: stage setup]
Grips set up props on the motin capture stage.

Motion capture technicians reconfigure the capture equipment on the performer. This takes between ten and thirty minutes on average. The performer's calibration needs to be checked after this.

Moving to a new capture volume setup can take hours. This requires that the motion capture equipment, whether it be optical cameras, magnetic transmitters and sensors or a mechanical suit, are moved to new marks. At the end of setup the grip crew and technicians should be ready to start calibration.

Because optical systems have longer volume reconfiguration times, different setups should be done on different days so that the equipment can be prepared beforehand. Likewise, Polhemus magnetic capture systems need significant time for calibrating out distortions caused by nearby metal.

Running the Capture Session

The capture session is where all the planning comes together to produce motion data. It is organized and pushed forward by the team, who are:

- Motion Capture Supervisor (oversees shoot, pre- and post-production)
- Motion Capture Director (directs talent, picks A & B takes)
- Capture Technician (motion capture system operator)
- Tracking Technician (data checker)
- Script Supervisor (continuity and logging)
- Preview Artist

In addition there will be Grips, who manage the stage setups, Videotaping crew, Production Assistants (PAs) who handle running small errands, fetching missing items or dealing with a client's cappuccino needs. Not mentioned are the owners or managers of the motion capture facility and the Producers who package financial deals. Other authors are left to describe the amazingly complex deal-making aspects of the business.

The listed titles will vary from company to company. Smaller projects will collapse multiple responsibilities onto one person to reduce staffing.

The Motion Capture Supervisor oversees the work from the pre-production planning, shot breakdown, the shoot and post-production. This person maintains the vision of the complete work and how it is being accomplished by the team.

The Motion Capture Director is sometimes the Supervisor, but is often someone from the client's team, usually a Second Unit Director or the Visual Effects Director. They are working directly with the talent and judging the performances to pick A & B takes. They call setups to the grips, clear meals and other breaks to start. Most of the rest of the team support this person. Because the Director is calling (judging) the shots as they are made he or she also sets the pace of work.

[photo: Director during shoot]
The Director of a motion capture shoot.

The Capture Technician is directly controlling the motion capture equipment, such as the optical camera controller, managing calibration of the volume, etc. With real-time systems they're also ensuring the continuous transmission of real-time motion data and it's formatting with the marker setup information from the performer.

The Tracking Technician is a post-processing operator, who does a cleanup pass on the data, filling tracking gaps, etc. Their data is passed to the Preview Artist or in some cases they may also act as the preview artist.

A Script Supervisor handles continuity and logging of shots as they are completed. Attached is an example form for completed shots. Continuity is always important, and in capture of motion cycles for games, this person's role in locating neutral poses between actions is doubly important.

The Preview Artist brings up displays of the completed data, ideally on a moderate resolution version of the final character. In more modern setups they will be using software like FiLMBOX from Kaydara to animate a low-resolution version of the character in real-time. The character display may be on a set of monitors in the studio or perhaps projected in large format onto a wall (as is done at PCS).

If the primary staff must be reduced, a bare minimum of three people are needed: Capture Technician, Tracking Technician and Supervisor/Director. This assumes some number of assistants will be available.

While is it is best to keep the number of other people present small to encourage focus, any number of additional bodies may be in the room. The Motion Capture Supervisor should make an announcement to those present before each shot. Action should be described to avoid startle reactions. If onlookers are "out of bounds" the off limits area should be demonstrated again. Onlookers should be asked to keep quiet. The technicians should be able to hear the Director and respond "ready" when asked. Any media should be started with a call for "playback." The shot should be called with "action" and "cut" as per standard practice. If there is a videotape crew shots should be visually logged with a slate and use the same timecode base as the motion capture system.

Capture sessions vary according to the technology being used (optical, magnetic or mechanical). In the immediately following section, these steps are described.

As previously mentioned, a procedure should have been agreed upon for delivery of the motion data to the customer. Again, this should include a review and sign-off on their side to ensure that the process is managed from end to end. Every chance should be taken to detect problems as early as possible in the schedule.

Technology Specific Motion Capture Work

In previous sections we described the production processes that surround motion capture, alluding to the work of a of motion data client, and describing in detail the preparatory work done by the motion capture team at the studio. The planning readies them to efficiently capture "hero takes" of the talent's performance with the equipment. The four core steps of the motion capture work process which are tightly tied to a technology are:

- calibration
- setup of the equipment on the performer (performer setup)
- capture
- tracking or data post-processing

Aside from pointing out differences, the following sections will also show ways in which the technologies are similar, by laying out a generic model of motion-capture technology. We'll show how each technology is explained by the generic framework.

Generic Model of Motion Capture

To better understand the technologies of motion capture we present a generalized model of a motion capture system. In subsequent sections, the parts of this model, both physical and mathematical, will be actualized for each of the different motion capture technologies.

Our technical definition of motion capture is:

> *Recording data that encodes the performed motion of a creature or object with the intention to map that motion onto similarly structured mathematical objects in the computer.*

For example, an instrumented body suit encodes an actor's performed motion so that it can be mapped onto a computer character. With this in mind the following sections mostly ignore the history of 3D input for digitizing shapes, or dial-spinning interfaces to motion programming devices such as robotic camera platforms.

The goal of the generic model is likewise to take the motion from a moving person or prop and put it onto a computer model, possibly one with different proportions or jointing than the original.

The generic model begins with the physical parts, which gather sensor information from performers and props:

- An exo-skeleton, a structure that connects a performer or prop to...
- Sensors or markers, transmitters and receivers, used to measure motion, which are connected via...
- Sensor data link to...
- Sensor data sampling unit usually an embedded computer system

[diagram: GenericModel.ai]
A generic model of the motion-capture process.

The exo-skeleton can be an articulated machine, as for the "armature" of a mechanical system. However, for all of these mechanical, optical and magnetic systems the exo-skeleton is an acknowledgement of the way sensors are connected to the body of the performer. This forces us to consider the effect of the sensors, markers or armature slipping or changing position relative to the performer's body during capture.

Sensors are either some form of transmitter/receiver pair or a direct measurement device. In passive optical systems, the transmitter is a ring of LEDs mounted around the lens of a camera. A retroreflective marker returns an image to the cameras (receivers). In an active marker optical system, the markers are light emitting (transmitters) with the cameras still receiving. In magnetic systems, there is a large electromagnet which sets up a magnetic field in the capture volume (transmitter). The magnetic field is received by small sensors worn by the performer. Mechanical systems use potentiometers to directly read rotations within their articulated armatures.

The next part of the generic model is the head end of the motion data processing pipeline. It has three motion mappings, typically implemented in software, which determine how the skeleton of the target character should move based on the sensor data.

The first motion mapping:

- Raw, unspatialized, data in the format of "sensor space" is mapped to...
- Spatialized sensor data, e.g. 3D positions using sensor calibration information

Each motion capture technology has some kind of "raw data" that its sensors provide. For optical systems, these are cameras images with "spots" where the markers are located. For magnetic systems, raw data is 3 axis magnetic field strength information. For mechanical systems, raw data is joint rotation in the armature. This is "first light" in any technology's detection of motion.

We define calibration as the discovery of parameters to map data from one mathematical measurement space to another. By "sensor calibration," we mean finding the relationships needed to convert the raw sensor data to spatialized data. In an optical system the sensor calibration is the position and orientation of each camera and their relationships. In a magnetic system the sensor calibration encodes distortion of the magnetic field, so that it can be factored out for a "clean" capture volume. In mechanical systems the sensor calibration information is limb lengths in the armature.

For example, in an optical motion capture system the unspatialized "raw sensor data" comes from several video cameras looking at reflective dots attached to the performers. After initial image processing step each camera provides the 2D position of white dots against a black background. The mapping operation that spatializes this data is intersecting rays mathematically projected from the camera position. These calculate the 3D position of each white marker dot. The calibration information is the fixed position of the camera, the lens parameters that indicate it's field of view, lens distortion, etc. Discovery of this information is called "camera calibration."

Mechanical systems output very structured spatial data. Their raw sensor data are rotations of the mechanical joints, which must be combined with a model of the armature (exo-skeleton) before it "makes sense." Measurements of the limb lengths between the mechanical joints are the calibration information. The resulting data is a hierarchic skeleton model of the armature, with limb lengths, joint rotations and a root position in space. in the case of a PuppetWorks-style system, the "sensor space" joint rotations of the armature animate the computer's armature model directly. More complex driver software derives and animates a model of the performer's internal skeleton in the Analogus Gypsy's driver.

Data is also usually filtered at this early stage of the pipeline. It is important not to confuse filtering with calibration and mapping. Filtering takes motion data at any stage in the pipeline and eliminates erroneous samples through techniques that will be described in later chapters. These techniques can include oversampling, out of range detection, etc. (see the appendix on filtering)

In summary the result of the first motion mapping is spatialized sensor data. This comes in three forms:

- 3D positions, as of markers in optical systems
- 3D positions and orientations, as of sensors in a magnetic system
- An articulated skeleton with joint rotations, limb lengths and root position, as for a mechanical system

The first two, simpler, types of data can be passed through a continuation of the motion processing "pipeline" to transform them into joint rotations on a character's skeleton.

But, it isn't necessary to pass translation data through the entire pipeline to apply it to a character directly. Raw point clouds representing translating markers can be used to animate a character's skeleton. Briefly described, the character's skeleton is equipped with handles for inverse kinematic solvers at the ends of its limbs. Point constraints are used to drag the ends of the limbs (and the root of the figure) to follow the point cloud.

If joint rotations on a character skeleton are desired, then two more motion mappings are performed in the pipeline.

The second motion mapping:

- 3D translations (possibly with rotations) are
  mapped onto...
- Model of the performer's skeleton, becoming joint rotations and root translation
  using the performer calibration information.

The performer calibration controls mapping of spatialized sensor data (the "exo-skeleton") to the performer's (internal) endo-skeleton.

The second motion mapping starts with sensors translating around the performer. Let's call them markers, as this is a common term in optical systems. The markers have to be identified with segments of the performer's skeleton. E.g. pelvis markers connected as the root segment, which connects to the left and right upper leg segment, etc. This segmenting of the markers produces a hierarchy. This "skeleton-like" segment hierarchy is the "exo-skeleton" of the generic model. The performer's skeleton is scaled and rotated into the segments and connected by fixing "offsets" between the segments (in the segment hierarchy) and the bones (in the skeleton of the performer). These offsets are therefore the calibration information for the mapping. As this example hints, placement of sensors on a performer is an art on its own. This will be described in Chapter 3.

The third motion mapping:

- joint motion of the performer's skeleton is mapped onto
- the character's skeleton in a similar format
  ...using the CG character calibration information

The third motion mapping puts motion data onto the target CG character. If the CG character and performer skeletons are exactly the same (limb lengths, shape, etc.) then this third mapping is not needed. The joint rotation information from the second mapping can be directly applied to the character skeleton. However, target characters will very commonly be of a different size from the performer.

Consider the Barbie problem. Lamb & Co. had a dancer perform the motions of Barbie for Mattel, Inc. [cite: Lamb & Co studio] Barbie's non-normal human proportions are well documented [cite feminist literature]. Transformation of data from the space of human motion to that of "Barbie-space" involves, e.g. dealing with longer than normal legs, shorter torso, etc. Transforming data from the performer's skeleton onto the target model is one example of calibration.

A bad calibration for mapping from performer to the character causes a problem so common it has the name "foot sliding." This is caused either by inaccuracies in the upstream capture process (bad calibrations mapping motion data onto the performer skeleton) or bad (or no) mapping from performer to character. If two characters have different leg lengths, and joint rotations from one are applied to the other, then the sweep of the target character's foot along the ground will be a different distance. Hence, foot sliding.

The third motion mapping adapts the performer's motion to the (possibly) differently sized character. This will be discussed further in Chapter 5.

The job of mapping from the performer's skeleton to the character becomes even more interesting when the target figure has a different number of limbs than the performer.

[diagram: GenericDataFlow.ai]

Data movement in the generic model

The generic motion capture process described here is intended to be industrial strength, ready to describe most situations. However, many production jobs are completed optimally by omitting un-needed work. In cases where a character's feet will never be seen interacting with a floor, foot-sliding will not be an issue and concerns about the third mapping are alleviated.

Next, we'll describe the various motion capture technologies, how they work, how to work with them and when to use them.

Motion Capture Technologies

There are several commercially available motion capture technologies. Each technology has advantages and limitations. Each technology is based on a different kind of sensor system that has its own way of responding to motion. These peculiarities of response ultimately determine to which jobs a motion capture technology applies.

This section describes:

- How does each type of motion capture technology work?
- How is the equipment set up in the volume and on the performer?
- What kinds of calibration are done?
- What post-processing is done to each kind of data?
- What kind of motion is each technology best for?

Mechanical Armatures

Mechanical motion capture systems seem like an obvious way to encode human motion. A mechanical suit with wires and joints is easy to imagine. But, while mechanical motion capture has a long history, its applicability as a motion capture solution has been limited. It enjoys a cost advantage against other types of systems, but limitations have kept it a niche product with a promising future.

History

Milestones in the development of mechanical motion capture:

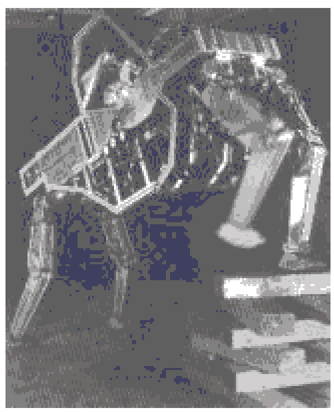
1940s	Nuclear material handling arms (mechanical linkage)
1950s	Handling arms (electrical linkage)
195x	Master/Slave Feedback mechanisms
1965	Handiman & Hardiman are developed at General Electric
196x	Disney Imagineering uses mechanical "motion programming" for audio animatronics
1983	Simon Frasier Univ., Tom Calvert, potentiometer-based Goniometers for computer analysis
1983	Atari Research, Jaron Lanier develops his "Data Glove"
1985	VPL Research founded to develop "Virtual Reality" products. MoCap is a core technology
1988	Protozoa's Mike the Talking Head puppeteered with mechanical hand controller
1988	Pacific Data Images, Jim Henson's Waldo C. Graphic character mechanically puppeteered
1988+	Pacific Data Images, builds a lightweight plastic exoskeleton using potentiometers
1992	SimGraphics, Mario, "Face Waldo" mechanical face-tracker is glued to performer's skin
1993-4	Stan Winston Studio & ILM, Dinosaur Input Device (DID) digital stop-motion armature
1997	Puppetworks delivers its digital stop motion armature product

[Need PuppetWorks date of body tracker]
[Need Analogus founding date]

[Need to mention early mathematicians who formalized the equations of forward kinematics. This section is designed to give due for the debt owed to the field of robotics. Also mention the CG scientists who came up with the hierarchic skeleton.]

Mechanical motion capture systems have deep roots. The ideas behind mechanical motion capture go back to the remote manipulators used since the 1940s in the nuclear industry. There, human hand motion was electromechanically encoded, transmitted and then repeated many feet away behind leaded glass where highly radioactive components were assembled. General Electric commercially built the earliest handgrip manipulators in 19xx.

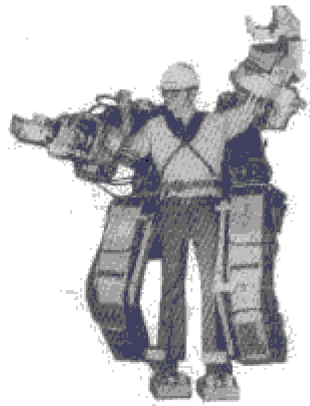
Modern mechanical motion capture systems are more closely related to early robotic control systems where the goal was "motion control" of a robot using "human motion capture." The early hand grip work at GE eventually lead in 1965 to the well known "walking truck," and the less well-known Handyman and Hardiman strength augmentation systems. These were the forerunners in reality of the imaginary "forklift" featured in the movie "Aliens." Much effort was expended on how to bring force feedback into these systems. Interestingly, force feedback has not been explored much as an aid to motion capture in entertainment.



The General Electric “Walking Truck”

In the 1960s, researchers at General Electric--cooperating with the U.S. Navy--built the "Hardiman." A force-amplifying exoskeleton, Hardiman gave its users a lift capacity 30x normal. Proponents envisioned all sorts of maintenance missions, from the loading of aircraft ordinance to underwater salvage.

[From Global Design News]



The Hardiman strength amplifier

> *In 1965 General Electric moved the concept off the drawing board with a military-funded exoskeleton called Hardiman, a 1,500-pound, 30- jointed contraption that looked as if it had stepped out of a Japanese cartoon. Hardiman had two massive load-bearing legs, two arms, and a hip girdle, all attached to a smaller, jointed framework. The user was supposed to strap himself into the framework, and a network of complex hydraulic and electric linkages would power-assist his movements to drive the surrounding exoskeleton. Dean Martin's sexy sidekick, Janice Rule, had no problem fighting off villains in a Hardiman knockoff in the 1967 film The Ambushers, but the GE engineers barely got a single arm to function properly. Hardiman was simply too large and bulky to maneuver easily*
>
> [From Discover magazine]

In the more cost conscious realm of entertainment Disney Imagineering built an instrumented upper body harness in the 1960s. It was used to program the motion of its animatronic characters, saving hours of work using individual joysticks and potentiometers. Even mechanical facial motion capture, using linear encoders and pads glued to the performer's face was part of these early experiments [cite: PDI work]. These control systems have been used sporadically by the entertainment industry to control mechanical puppets. The easy training but lower reliability of a single performer in a complex harness is weighed against the time needed to train a team of operators on a set of simple levers and joysticks.

Mechanical motion capture finally joined computer graphics for entertainment when Jim Henson and Associates worked with PDI to create a mechanical arm in 1988 to preview and record motion for the character Waldo C. Graphic, which was then offline rendered. This system was a simple affair, more of a puppetry rig. The arm gave the flying character's position and orientation in a small 3D volume, while a glove-like appendage at the end of the arm allowed the mouth to be moved.

Theory of Operation

In mechanical motion capture systems the exo-skeleton is literal. The performer wears a mechanical armature fitted to their body. The fitting is usually done with elastic straps and belts, which hold plastic plates against the body of the performer. The plates anchor instrumented joints with rods connecting to the other plates. The joints of the exo-skeleton respond in a complex manner to the motion of the performer's real endo-skeleton.

[Photo image PIC00014.jpg]
Upper back section of mechanical capture suit

The sensors in a mechanical armature are usually variable resistance potentiometers or digital shaft encoders. These devices encode the rotation of a shaft as a varying voltage (potentiometer) or directly as a digital value.

A cable harness carries the data from the encoders, riding the armature to some central point on the exoskeleton where it is either taken off on a larger cable or encoded into a radio signal for transmission to a base station. One mechanical exo-skeleton used on onboard data recorder to make it entirely self-contained [cite Analogus unit].

Mechanical exoskeletons require care to fit to a performer. All the components are mechanically linked together. This requires a kind of "universally tailored mechanical suit." The suit must fit the performer precisely and not slide around to achieve high precision. Given the weight of a mechanical exoskeleton and stiction in its joints this is a daunting design task on its own. The solutions have been ingenious and have solved this problem for a good range of body sizes.

Analogus and Puppetworks both use systems of sliding rods between the joints of the suit (rod & joint type). These rods are either grooved or prismatic and transmit torque to the encoders. Puppetworks in particular uses aircraft grade parts designed to reduce stiction and "play" at the joints.

Current mechanical systems use rotation information as the raw data of their sensor space. To this is manually added information about the distance between the rotating mechanical joints.

It might be expected that the sensor space of a mechanical motion capture system is the simplest of all, but in fact, it is not. The joint sensors of a mechanical motion capture system do not directly map onto the joints of a performer. There is a correlation, but its not quite linear and not quite one to one. Multiple sensors respond in a non-linear fashion to the linear motion of a single bone joint within the performer's body.

The best way to understand the sensor space of these systems is to remember that it is a model of the mechanical motion-capture system itself and not a model of the performer's joint system. Recall that in our generic model of a motion capture system we start with the sensors, model the exo-skeleton (armature) using calibration data and then extrapolate the performer's skeletal pose from that.

The rod and joint mechanical systems are modeled as a chain of joints on fixed or variable length arms: a classic "forward kinematic system." The system is considered fixed to global coordinate system at it's root node, usually the upper or lower back. E.g. the lower back is considered fixed and the forward kinematic chains extend from there, along the upper back, to shoulders, to elbows, to wrists, etc. In such a system the figure would stay in one place, arms and all moving as though rigidly stuck on a pole attached at their lower back. To provide translation data for the entire armature another device can be attached at the root to measure global position and orientation.

One version of this idea is the Puppetworks roving arm. A long, two-jointed mechanical arm attaches between the shoulders of the performer, which indicates in 3D where the figure is located (where its upper back is located).

[Photo image SIGGRAPH16.jpg]
Early PuppetWorks BodyTracker with mechanical positioning
(Note mechanical arm extending from shoulder pad going upward)

A more complex and versatile extension of this idea is the "Gypsy" technology from Analogus. In this system foot to ground contact is determined, with foot switches in early versions of the system or with inertial sensors in later versions. The "planted" foot is considered fixed and the rest of the exoskeleton moves around it, until the foot is lifted. The Gypsy concept is powerful because, while it has low spatial accuracy, it allows the performer to move through an unbounded area.

Mechanical capture systems have a "distinct flavor" relative to the other technologies. The revival of their use for body motion capture started with "pose capture machines" that is, they recorded the changing poses of the person to whom they were attached, but could not tell how the entire figure was moving in space. This "base position" problem was later solved in two interesting ways. The Puppetworks BodyTracker began by using a mechanical arm connected at the hips to the performer's armature. The Analogus system uses their "Gypsy" technology, which provides a kind of "dead reckoning" of the performer's position in space relative to their feet being planted on the "ground."

On their higher-end products both of these companies now offer baseline positioning with the interSense gyroscope and accelerometer system. InterSense offers a less expensive gyroscope-only unit to determine relative orientation, or a higher-end unit that also incorporates a 3-axis accelerometer. The accelerometer is very accurate and can accumulate delta-V to determine speed and position. The high-end interSense units (IS-600) have a ceiling mounted ultrasonic transmitter array to manage the slow drift of positions accumulated by the accelerometer attached to the motion capture suit.

Setup & Calibration of the Capture Volume

Unlike other forms of motion-capture, to be described later, mechanical systems do not measure from discrete points in or around a volume. Instead, they are attached directly to the performer and measure relative motion.

In a mechanical arm system, the capture volume is the limit of the arm's travel. Calibration involves planning motions around the limitations of the arm's encumbrance.

In a “dead reckoning” system based on detection of “foot planting” (when a foot firmly touches the floor) calibration of the capture volume is in some sense dependant on the accuracy of the foot contact sensor's calibration. This, along with the gyro that determines the orientation of the suit's root node, relates the suit to the space around it.

Versions of these systems that use an inertial sensor for global positioning follow the calibration procedures for that equipment.

Setup & Calibration of the Performer

For mechanical systems calibration provides us with two kinds of information: about the exoskeleton and indirectly about the performer wearing the exoskeleton. Calibration requires measuring parts of the exoskeleton as well as the performer.

In a rod and joint system the goals are to understand the length of the sliding rods, i.e. what size has the exoskeleton taken to accommodate the wearer, and position, i.e. where has the exoskeleton been attached to the wearer.

Especially critical is hip and shoulder attachment of the armature. These parts of the exoskeleton need to be very stable and tightly attached to the performer. This is because these form the root of the long kinematic chains extending to the wrists and feet. Inaccuracies or slippage of the armature at these points affects all the limbs down the chain.

Noise and Accuracy

The primary sources of inaccuracy in mechanical motion capture systems are electrical noise from the potentiometers and slippage of the suit on the body of the performer.

Systems based on digital shaft encoders rather than analog potentiometers are more precise. However, digital shaft encoders have limited radial resolution, that is, they have a fixed number of discrete rotational steps they can measure. This results in a kind of "digital joint play." This is especially true of the more compact digital shaft encoders that are favored for use on mechanical motion-capture systems. Fortunately, for most entertainment applications this "play" is not an issue and the advantage of a noiseless sensor outweighs the need for extra resolution.

Advantages

Overall the advantages of mechanical systems can be summarized as:

- Relatively low cost
- Real-time data
- No line of sight "capture environment" needed
- Unlimited capture volume (Analogus Gypsy)
- Untethered motion or performance is possible (Analogus Gypsy with radio link or onboard recorder)

A mechanical system for capturing full body motion costs about a quarter of a comparable optical system or about 50% less than a comparable magnetic system.

Mechanical motion capture systems easily provide real-time data. This is essential for applications such as computer puppets performed for a live audience.

Aside from an optional intertial base positioning device, mechanical system does not use a specially prepared capture volume. Magnetic systems must operate within range of their transmitter, and optical systems can only capture what their cameras can resolve. A mechanical system has been strapped to a motorcycle rider while he was zooming around a racetrack. While the data did not record the motion of the rider relative to the ground, all of his leaning and pose movements on board the bike were captured.

The sample rates of mechanical systems are less constrained than, e.g. magnetic systems. While most mechanical systems provide 120 Hz sample rates this limit is chosen for convenience of connection to a computer (usually by serial port), rather than an electronic constraint. Recording at a high rate and then filtering the samples down to a lower frame rate, i.e. over-sampling, is one way to detect and eliminate transient electrical noise that might be induced. Potentiometers are typically tested and culled from commercial production runs, since average quality is low. Some "pots" will generate transient noise that super-sampling can cure.

As mentioned earlier the Analogus Gypsy uses a unique "foot planting" technology, a kind of "dead reckoning" navigation that indicates where the armature is in space. This is accomplished using a hip mounted gyroscope provided by interSense in conjunction with software. Using a calibrated "standing on the ground" position, the gyro lets the armature modeling software calculate when a foot is planted on the ground and therefore when the armature is moving forward relative to a "planted" foot. The software makes a number of assumptions and is susceptable to drift, but generally works well.

Disadvantages

To be sure, mechanical systems have some disadvantages as well, adding more "flavor" to their use.

These can be summarized as:

- Limited ability to reconfigure the sensors on their armature
- Armature would not survive a performer’s roll or fall motion
- Encumbering (like wearing a bicycle or at least a stiff body suit)
- Some units have mechanically tethered capture volume (older Puppetworks design)

The greatest issue with mechanical systems is that they are almost single purpose, i.e. a body motion-capture system will only do that; capture the motion of a single performer's body. The motion of props held by the performers cannot be captured using the technology. It would require development of mechanical linkages to the props or, more practically, the use of another capture technology. Overall, cost sensitive jobs like computer games and sports training applications have made use of mechanical systems.

Another trade off is that mechanical systems cannot readily handle rolls and falls, a common type of motion to be captured. The current, somewhat delicate, mechanical armatures would be crushed or bent, possibly injuring the performer as well.

Mechanical systems are also somewhat encumbering. Putting on the Analogus and Puppetworks machines could be thought of as similar to "wearing a small bicycle." The first moments are confusing, with parts swinging and hanging all around the performer, but then, as the straps and belts are tightened, the suit nearly disappears. One must still be careful to avoid extreme movements in the current generation of suits. It's possible to bind joints or stress them when reaching for a big pose (e.g. overhand tennis smash).

Finally, mechanical absolute positioning systems, like that of the earlier version of the Puppetworks machine, are tricky to work with, being a notch above the difficulty of avoiding the sensor cables that trail from electromagnetic suits.

The future for mechanical systems lies in making the most of their greatest strength: the potential for low cost.

Magnetic Fields

Magnetic motion capture technology is a wonder. "Exactly how can that work?" is a common first reaction among users and viewers of these systems. Of course, knowing something about the technology is an important precursor to recognizing its limitations and advantages. The following section provides a solid basis for understanding this relatively mysterious technology, its excellent flexibility and interesting quirks.

History

Milestones in the development of electromagnetic motion capture technology:

1960s Bill Polhemus develops magnetic tracker for H.M.D. at Harvard University
1970 Polhemus founded to build head mount aiming device for helicopter pilots
1986 Ascension founded
1988 Polhemus first commercial device the 3Space digitizer (singe channel wand)
1991 Videosystem (later MediaLab Paris), puppeteers "Mat the Ghost" with Polhemus, gloves and MIDI
1992 deGraf, Moxy full body puppet premiered Alive! Performance animation system (using Polhemus)
1994 FasTrack first commercial multi-channel electromagnetic motion-capture system
1996 Polhemus wins Academy Award (Oscar) for contribution to Disney's "Beauty and the Beast"
198x? deGraf & ?, character for Spanish exhibit at World's Fair uses one Polhemus to move head

Bill Polhemus developed Electromagnetic (EM) motion capture technology at Harvard University in the 1960s as a component of a Head Mounted Display system. He founded Polhemus Incorporated in 1970 while consulting for the military. Helicopter pilots needed a means of aiming without taking their hands or feet from the complex controls of their aircraft. In 1988 the first commercial use of the technology appeared as the Polhemus 3Space digitizer. A single channel system, it was a pen mounted sensor that would be touched to points on a real object, providing raw 3D data points for modeling. Polhemus won a technical achievement award in 1996 from the Academy of Motion Picture Arts and Sciences (the "Oscar"), for the use of the 3 Space digitizer in creating the chandeliers in Disney's "Beauty and the Beast" ballroom dance scene. In about 1994 the first multi-channel system was introduced; the four channel

FasTrack. It was designed for medical and sports applications but became the basis of the many times' larger full body motion digitizing systems we have today. Polhemus sells approximately 25% of its systems into medical applications, with the remaining 75% going to entertainment and industrial customers. About 15% of their business is still from the military in the original weapons pointing application.

For years Bill Polhemus and his company sat atop a small mountain of patents and continuing research. Several indivuals left his company and founded Ascension in 1986 with a new (patented) means of pulsing the magnetic transmitter's field (more below). This provided some competition in an otherwise slowly developing market.

Most recently the controllers that capture data have been refined. Instead of serial ports systems now depend on Ethernet to transmit data. Some suits are also wireless, using spread-spectrum radio to get data to the controller.

Theory of Operation

The basic idea behind EM capture technology is simple enough: generate a magnetic field with a "transmitter" and then detect the resulting field using a "sensor."

The transmitter is a moderately large object. The Ascension system's is a solid one foot square black cube weighing a hefty 50 pounds. The Polhemus transmitter is a somewhat whimsical transparent plexiglas sphere about 1.5 feet in diameter which must be mounted in a stable fashion for calibration.

Sensors are small units about an inch square, although there are now micro-miniature units designed for surgical applications. These are attached to a performer's body using elastic straps and Velcro.

Cables run from the receiving sensors either directly to the system controller, or to a radio transmitter package strapped to the performer. The system controller's antenna receives the radio signal. A handy feature of the wireless transmitter package are input plugs for two serial ports. Data gloves, encoding the performer's hand and finger motions, can be plugged in there for radio transmission back to the base station.

Older style magnetic systems communicated with a computer via a large number of serial data cables, one for each sensor. This lead to some unusual setups, including ethernet serial port concentrators, which provided 16 or more serial ports on a box connected to the network. The receiving computer would run a driver program communicating over the ethernet which made the ports appear to be directly connected to it.

All modern magnetic systems use a direct ethernet connection to communicate with the receiving computer. Each system uses its own ethernet protocol.

With this, we have a complete user's view of the parts and operation of a magnetic capture system. But, how does it really work?

[Photo image Back.jpg]
Wireless sensor magnetic tracking system
(Transmitter ball in foreground)

Starting from basics: a varying electric current in a wire creates a varying magnetic field around that wire. An electromagnet is a wire repeatedly wrapped around a core of metal (which organizes the field to strengthen it). The process is reversible, so that a varying magnetic field will also induce a varying current in a wire.

This is the basis of a simple transmitter and receiver. It cannot act as a radio, because a pure magnetic field loses strength quickly as it moves away from the transmitter. In fact the field strength drops in proportion to the change in the surface area of a sphere centered on the transmitter. Said another way, imagine a balloon centered on the transmitter. As it blows up the surface stretches apart, "weakening" it. The field is more or less like that. As it expands away from the transmitter it loses power.

[Diagram MagTransRecv.ai]
Transmitting and detecting a magnetic field

It's possible to tell how far away the transmitter is from the receiver by the strength of the signal. This is the first way to get information in an electromagnetic motion capture system. But, signal strength alone only indicates the sensor's distance from the transmitter, a vector length without orientation of the vector (see below).

Magnetic fields are bipolar (e.g. have a North and South side) and flow between the poles of a magnet. An example of this is the high school experiment where iron filings are scattered on a piece of paper. A magnet is placed under the paper. The iron filings react to the field, aligning to show its organization between the poles.

The sensors in an electromagnetic motion capture system use this property of a magnetic field to determine orientation relative to the field. Each sensor contains a delicate detector coil (omitting a core because of the distortions it causes), which reports how well aligned it is with a magnetic field line. Imagine placing a transmitter and a sensor next to each other on a tabletop with the axes of their North-South poles in parallel. As they are moved apart, signal strength at the sensor drops. However, if the sensor's North-South axis is turned out of alignment with the transmitter, the signal strength also drops.

[Diagram MagOrient.ai]
Determining orientation in a magnetic system (2D example)

In our simple example then we can now report a signal strength that indicates a combination of distance from the transmitter and how far off the axis of the magnetic field lines the sensor is oriented. This is a start, but now we must pick apart this combined orientation and strength information to determine the actual position and orientation relative to the transmitter. This is possible because magnetic fields are not spherical, like point source radio waves.

We'll now add two more dimensions to the example. Our transmitter now has three coils at right angles to one another, and so does our receiver. The controller pulses each of the transmitters, one at a time, each time briefly creating a magnetic field oriented in its direction. Our three dimensional sensors will take readings on all their axes during each pulse. This is the real key to the working of the system. By measuring along multiple axes its possible to mathematically derive 3D orientation.

Each time the field turns on (each pulse) it's possible to calculate a surface on which the sensor must lie. The trick is that the surface is not a sphere, it's a toroid centered on and aligned with the core of the transmitter. After doing this three times the three toroidal surfaces are intersected, providing two possible places where the sensor is located (see the diagram). This can also be correlated with the field intensity, which gives a check of the distance to the transmitter.

To summarize, magnetic fields have three properties that make it possible to build a motion capture sensor. First, they can be created with a varying electric current and vice versa. Second, the field is oriented along the axis of the electromagnet's poles. Third, the field is toroidal in shape, not perfectly spherical. A magnetic motion capture system creates data by doing two things with the magnetic field. First, it uses three coils (one for each axis) in both the transmitter and the sensors. Second, each transmitter axis is pulsed one at a time, and all three axes in a sensor detect each pulse (at varying strength). The system finally yields 9 data points for each sensor (3 axes times 3 transmitter pulses). These data are integrated to produce two solutions. The orientation solution is given by solving for the best orientation (two solutions). The position is given by intersecting the three toroids that also yields two solutions.

As it is easy to imagine, these are only the basic principles underlying the highly evolved components and software found in commercially available equipment.

[cite: Vlad Kogan from Ascension for information in this section.]

[Diagram MagPosition.ai]
Determining Position with a magnetic system

The two leading producers of magnetic motion capture systems are Polhemus and Ascension. Both companies pulse the power to the transmitter coils differently. These differences are subtle and result in very different noise response.

The Polhemus system is referred to as "AC magnetic." Like the house current powering your home appliances Alternating Current (AC) varies continuously over time in the manner of a sine wave. Each AC "wave" is considered a transmitter "pulse" in this system.

The Ascension system does not vary current during a pulse (DC), i.e. the transmitter is turned on, the current remains steady, and then it is turned off. The shape of the current pulse over time looks more like a square wave than a smoothly curved sine wave.

Each of these approaches delivers a unique set of tradeoffs, problems and solutions. To understand them recall that a varying magnetic field induces a current in a wire, or any other piece of metal. This undesirable induced flow is called an "eddy current" after the small, sometimes reversed flow of water at the edge of a stream or river. When the magnetic field stops, or begins varying the other way, the eddy current collapses in the metal and its energy is again converted into a magnetic field.

What this means is that within an AC varying magnetic field certain metals become secondary transmitters because of induced eddy currents. In practice this causes measurements to be "bent" near large metal objects, e.g. a sensor slid across the floor will appear to rise or turn as it approaches a metal object.

A DC system also induces eddy current but the eddies settle to zero during the sample time such that their effect is minimized.

As described above, the sensor space for a magnetic system is complex. However, both Ascension and Polhemus hide most of this complexity from the user, instead providing them with control of sample rates, filtering parameters and a set of guidelines for preparing the motion capture area.

Setup & Calibration of the Capture Volume

Setting up a magnetic capture system requires reducing the amount of magnetic noise in the capture volume. There are three ways to do this. The first is to learn about the materials that make up the building and any equipment that might create a magnetic field. The second is to use the magnetic capture system itself, with all filtering turned off, to directly watch for interference when a sensor is placed (motionless) at various points within the proposed capture volume. The third is to work closer to the transmitter, improving the signal to noise ratio of the detected signal.

The most typical sources of passive magnetic noise are the construction materials that make up the floor and walls around the motion capture area. It helps to have the building plans for the motion capture area. These plans will indicate major electrical lines, large steel beams, and other sources of interference. But, even without building plans it's not difficult to "tune up" a performance space. For unknown metal objects it is helpful to keep that bane of diskettes and credit card stripes, a permanent magnet, handy for testing. Rebar (reinforcement bars) are steel rods used to reinforce concrete floors and walls, is one of the easier to find sources of interference. Always ask what's on the other side of the wall. In a high-rise building, near the roof, there may be air conditioning or elevator motors.

With this information in hand it's easier to understand why most magnetic motion capture studios build an 18" tall stage on which to place their performers. This gets them off the floor and away from rebar.

The magnetic motion capture system itself is an excellent means of detecting magnetic interference. Due to the "measurement bending" distortions mentioned previously the Polhemus system includes a calibration step to cancel this out. A single sensor, mounted on a stable, non-metallic pole, can be placed at various points in the capture space. The system's filtering parameters are removed or reduced as much as possible. Samples are taken for a period of time similar to a capture shot. The noise characteristics of these samples are analyzed. Sampling proceeds in some pattern through the entire capture volume (usually a gridded cube). Samples should be taken at all positions that a sensor is likely to be in the volume, from the top of a foot or prop to the highest arm reach in a tennis swing. Some software systems can graph the noise patterns in 3D. While it takes significant time to survey an entire capture volume sources of potential trouble are made clearly visible.

The best data are usually captured near the transmitter, with accuracy dropping off in the same inverse cube relation mentioned earlier. Typically the transmitter is placed at about the performer's hip level or slightly higher. With this placement the sensors are about equidistant from the transmitter when the performer is in a rest pose. If there is active magnetic noise from above the transmitter can be stationed higher it to put more field strength there to counteract the noise. The transmitter can also be mounted higher if passive magnetic noise is coming from below (see the next section regarding the types of noise).

It is difficult to recommend a standard setup, but MediaLab Paris has a good start. They construct a two-tiered stage. The first 18" produces a platform on which a second, slightly less wide and deep 18" tall platform is constructed. Between the layers of this "sandwich" is placed the transmitter, whose active hemisphere faces up. This works well for a high ceiling room where most of the active and passive magnetic noise is down at floor level.

An alternative is mounting the transmitter overhead.

[photo: overhead transmitter]
A magnetic capture stage with overhead transmitter.

A specific worst-case volume existed in the now defunct sgi virtual studio, in Mountain View, California between 1997 and 1998. A blue screen television studio and real-time motion capture area were placed opposite a large glass window looking into the control room. There was only one place in the room for the motion capture stage. This was next to the windows, behind which were three large tube-based video monitors. In addition, a ferrous metal cable trough ran around the custom-made wooden stage, carrying high current power cables. The control computer for the motion capture equipment was placed at the edge of the stage (with another tube-based monitor). Finally and most subtle of all, a custom-made wooden stage had been constructed by sgi to avoid the steel rebar in the concrete floor. The stage plans specified assembly with non-ferrous nails and glue. In spite of this the performer's feet were "wobbling" when planted in certain places on the stage. Turning on the system and doing a sensor survey over the floor allowed us to create a "map" of the noisy areas. We enlisted the help of six other people to split the 400-lb. stage in half and turn it upside down. Testing the protruding nails with a permanent magnet confirmed the sweaty hypothesis: one third of the nails in the stage's mouldings were ferrous. The stage craftsman had reverted to steel nails when he ran out of aluminum ones. With this evidence in hand the builder was given a considerable repair job.

Setup & Calibration of the Performer

Magnetic systems provide both orientation and position info, are very versatile and can be set up on the performer and props in a number of ways.

While other systems (passive optical and mechanical) are sold with software packages to load motion onto computer characters, magnetic systems don't usually include this. The following section discusses general ideas about sensor setup on a performer. Motion editing software used to apply "magnetic data" (orientations and positions of sensors) to characters is described in later chapters.

Partly because motion application software was not provided, early in the evolution of these systems it was typical to use only orientation data to directly rotate the joints of an animated figure. While this can be inaccurate, due to size differences between the performer and character, it is easy to set up and program. Skilled performers can puppeteer a character in spite of the inaccuracies by adapting the motion they are acting to the character through practice with a real-time display. A typical 15-sensor setup that uses only orientation information would be:

- Head
- Between shoulder blades on spine
- At top of hips on spine
- on upper arms
- on lower arms
- on back of palms
- on upper legs
- 2 on lower legs
- 2 on top of feet

This configuration is very sensitive to slippage of sensors on the performer's body. In particular it is sensitive to slippage around a limb, which is unforunately very common because sensors are attached with a strap around the limb.

In order to reduce cost and use fewer sensors, if the software system can perform inverse kinematic translations there is a simpler 7-sensor configuration:

- Head
- Between shoulder blades on spine
- At top of hips on spine
- on backs of palms
- on tops of feet

It is easy to attach sensors to these points because they are relatively stable places on the body. The trickiest tends to be the spine. While the spine articulation can be inferred and calculated, additional sensors can be added to the spine if more accurate articulation is desired.

This configuration depends on the motion capture software to calculate joint rotations between the sensors. The sensors provide the goal position for the ends of the limbs in an inverse kinematic solution. IK solvers can be mathematically unstable, producing unusable solutions if the goal position moves the limb unexpectedly or wildly due to noise. For this reason this configuration is very sensitive to noise in the position or orientation of the sensors.

As a general rule: the fewer sensors on the performer's body the more dependent the solution is on those sensors for good quality data. More sensors allow dropping bad data, but still having enough to compute a solution.

Some newer systems continue to use 15 sensor setups but calculate an additional inverse kinematic step. The joint rotations thus derived are used to double check and increase the accuracy of the pure rotation data. This has come about because of improvements in computer power, stability of inverse kinematic algorithms and the fact that position data from electromagnetic systems is somewhat more stable than orientation data.

Noise and Accuracy

There are two kinds of sources of magnetic noise: passive and active. Passive sources are ferrous metals, such as steel nails in a stage floor. Active sources are current carrying wires and electromagnets that generate varying magnetic fields.

For active magnetic noise sources, the more current (measured in amperes) flowing through a wire, or the closer the rate at which it varies matches the sampling rate, the more likely it is to add noise to the motion capture data. This means that 60 Hz (or 50 Hz in Europe) sample rates are likely to be affected by wall current in a wire. A coiled extension cord is a low quality electromagnet. Often left near the motion capture area, these are a common source of interference. Many types of electronic or computer equipment have strong electromagnets in their power supplies. The worst offender in this regard is tube-based computer and video monitors. A strong electromagnet is used on the "neck" of the video tube to deflect the electron beam drawing the screen. Another common mistake is to place a computer monitor near the edge of the motion capture area. This is such a strong varying magnetic field that it easily renders data taken in its vicinity useless. Finally, the industrial buildings often have large electrical distribution panels and heavy gauge electrical wire routed around them. These are also to be avoided.

From the Ascension web site:

> *A CRT located within a foot or two of a sensor may produce noisy outputs unless compensated. The problem occurs because CRTs generate electromagnetic fields, composed of both horizontal and vertical signals. While immune to horizontal interference, vertical scan rates often operate within our tracker bandwidth. The amount of noise will vary depending on the operating frequency and shielding of the CRT. Using the CRT SYNC command along with the CRT synchronization pickup cable can minimize the amount of CRT noise picked up by the sensor. When properly implemented, this technique significantly reduces noise produced by a nearby CRT.*
>
> *A sensor should not be located near power cords, power supplies, or other low-frequency current-generating devices.*

Calibration of filters results in an overall tradeoff between sample rate, "sharpness" of the motion and noise reduction. There are two sample rates, usually one at which the system is operating and another at which the end user is actually receiving the date. High sample rates may result in noisier data. Low sample rates allow filtering algorithms to use oversampling to reduce noise.

In both the Polhemus and Ascension systems there is a handy technique for analyzing noise. A single sensor is placed on a non-magnetic pole. A 3D grid is established in the capture space with about 1' per division. A sensor is clipped onto a non-magnetic pole and moved through the grid step by step. The motion capture system's filters are reduced to their minimum settings. The sensor is placed at each position in the grid, including on the surface of the motion capture stage for approximately a half minute. At each position, data is recorded and the noise analyzed. This will produce a 3D "noise map" of the motion capture space. If it's possible to plot the noise data in 3D it becomes very easy to see where problems exist. Sometimes it is easier to show the map to the performer and let them avoid the problem areas rather than eliminating them. E.g. stay away from the back wall, there's a refrigerator on the other side.

Some Characteristics of the Polhemus system

Mounting the transmitter and mapping distortions in the field. Tuning sample rate to eliminate AC power system noise.

While the continuously varying AC signal of the Polhemus transmitter tends to induce magnetic fields in nearby metal objects these induced fields are stable; i.e. the noise produces a simple linear change in the sensor's data. As the manuals for the Polhemus system show, this results in a "skewing" of data inside the capture volume. To compensate for this Polhemus uses a 3D mapping technique (on which the noise map mentioned previously is based). The transmitter is fastened down in what is believed to be an optimal location. Distortion within the entire capture volume is mapped. The distortion map of "skewing" is saved and an "unskewing" correction is applied to further data recordings. A very important point is that for these high accuracy setups of the Polhemus system the transmitter must be fixed in place. If it is disturbed then the “unskewing” compensation will work against the distortions present in the new position of the transmitter, producing even more distortion of data across the capture volume. It is useful to attach a sign to the Polhemus transmitter after calibration, to encourage passersby to leave them undisturbed.

[photo: "life threatening high voltage" sign on the transmitter at Protoza]
A well-labelled transmitter.

Since the nominal sample rate of the system is 50/60 Hz it can easily run afoul of noise from the standard electrical systems in buildings. To avoid this the Polhemus transmitter’s frequency can be varied in small steps above or below the sample rate.

Some Characteristics of the Ascension system

From the Ascension web site:

> *An AC electromagnetic system is sensitive both to electrically conductive metals, such as stainless steel, copper, aluminum, titanium, and to ferromagnetic metals, such as iron and carbon steel.*
>
> *Ascension trackers have a built-in software command that lets you decrease their measurement rate, which, in turn, helps control errors from conductive metals. Depending on your needs, you can slow your Ascension trackers down to approximately 12 measurements a second. The slower the operation, the less output errors will be caused by nearby conductive metals.*

As mentioned in previous section there is an internal sampling rate and an output rate. The Ascension system internally measures each axis 3 times (for each of the three axes).

The Ascension system loses data if the sensors are overwhelmed (saturated) by the transmitter’s signal. To solve this Ascension varies the power of the transmitter depending on the proximity of the nearest sensor on the performer. This distance is about 18”. If e.g. a performer’s hand swings near the transmitter the power will be reduced, making data from the farthest away sensors suddenly noisier. Also, the power variation is in two steps: high or low. A momentary glitch in the data sometimes occurs during a switch between power levels. This feature can be disabled, leaving the system operator to manage saturation problems.

To deal with this it is a good idea to plan the placement of the transmitter and the path of the performer’s motions so that no sensor swings within 18” of the transmitter. Movements in a flat “plane of action” should occur at right angles to the transmitter. A worst case would be an arms outstretched movement, with one hand at the transmitter and the other far away. This motion would be better recorded with the performer turned to face the transmitter.

Advantages

When magnetic motion-capture systems first appeared they promised two things: real-time data and orientation information along with the 3D position of the sensors. There are several more to add to that:

- Real-time data
- Position and orientation data from sensors
- Sensors are easily reconfigured
- Trade for lower noise with lower capture speed
- Roll and fall captures possible

- Untethered motion or performance is possible
- Encumbrance comes from cable, not sensors

Real-time data is the key reason that magnetic systems are used to control real-time computer puppets. While mechanical systems have made inroads, magnetic systems still dominate this application with their easy reconfigurability.

Orientation data in addition to 3D position allows for easier programming. Even without compensating for the difference in size between a performer and a computer puppet, orientation data alone can be used for control. In early computer systems there was not enough compute power to calculate an inverse kinematic solution in real-time. Orientation information was an answer to this problem. While this is no longer the case, even the smallest PC has enough power for real-time IK, the orientation data can be made part of the IK solution, helping to stabilize the algorithm's solutions.

Magnetic sensors are easily reconfigured. E.g. they can be attached to a prop to immediately provide data on how the performer is interacting with it. It is prudent to purchase a system with a few extra sensors for this purpose (in addition to providing redundancy should a sensor fail).

Noise is a reality for users of magnetic systems, however, lower capture speeds can help to smooth out data. This remains problematic for fast transient motions, like karate kicks. It makes sense to capture these moves in the "quietest" part of the capture volume.

Motion capture is often used for action sequences. Rolling and falling motions can be captured with magnetic systems if the sensors are padded and cable motion is planned. Untethered systems make this much easier.

Finally, the sensors in these systems are light and unlikely to hinder a performer's motion. Any problems usually come from the cabling running across the limbs and off of the performer (when using a tethered system).

Disadvantages

Common complaints about magnetic motion capture systems are their small capture volumes (must stay near the transmitter) and the potential of noisy data when used in a badly prepared space.

- Small capture volumes
- Disturbed by varying magnetic fields (fly-back coil in tube-based computer monitors are worst)
- Very sensitive to metals (ferrous & to a lesser degree non-ferrous)
- Very non-linear noise characteristics (hard to filter out)

Magnetic motion capture requires the performers remain relatively close to the transmitter. This is where noise is reduced to a minimum. This can severely limit the usefulness of the system. E.g. pole-vaulting motions are unlikely to be captured using a magnetic system. A promising development has been multiple "cellular" transmitters. However the sample rate was reduced because each additional transmitter reduced the sampling rate, i.e. two transmitters would halve the rate because they would have to be pulsed separately.

Varying magnetic fields that are actively generated by electrical devices, such as a tube-based computer monitor, can render large areas around them useless for magnetic motion capture. Flat panel monitors are an alternative.

Sensitivity to metal, while problematic, can be reduced through avoidance or calibration in the case of the Polhemus system.

Some types of noise encountered by these systems is more or less random, and therefore hard to filter out. Being infrequent, this noise can be edited out by hand if the motion capture is intended for a post-process.

Passive Optical Markers

Passive Marker optical motion capture systems use reflective dots attached to key points on a performer. These dots are tracked using multiple video cameras. Special illuminators cause the reflective dots to glow brightly as seen from the cameras.

[Photo image Dancers1.jpg]
Reflective markers on performers

In the sections below we'll provide a brief history of the many technologies that combined to enable optical motion capture. As will be clear, passive optical motion capture technology has greatly benefited from the exponential increase in computing power in personal workstations and computers.

Early History

We present a "pre-history" of the diverse technologies that were combined to create commercial optical motion capture systems. Milestones in the development of these precursor technologies are:

1774 Johann Heinrich Lambert develops "spatial resection" [footnote: Lambert is better known for proving that pi is an irrational number and for developing the lighting model bearing his name.]
1840 Arago suggests use of daugerrotype for cartographic measurement
1849 Laussedat (Father of Photogrammetry) uses photographs to make maps
1856 Porro invents the first means of correcting lens distortion [footnote: better know for the Porro Prism used in standard binoculars to widen the stereo view.]
1867 Laussadat exhibits Photothéodolite and plan of Paris made from rooftops
1872 Muybridge begins work on sequential photography
1893 Meydenbaur coins the term "photogrammetry"
1899 "The Geometric Fundamentals of Photogrammetry" published by Finsterwalder
1915 Max Fleischer develops rotoscope technique
1924 Gruber derives projective equations of photogrammetry and their differentials
1937 Snow White and her Prince are animated using rotoscope
1940s? Church contributes equations for resection, intersection, etc.
1953 Schmid develops matrix equations for multi-station photogrammetry, with error analysis
? DLT optimization [Abdul and Aziz??]
? television initiates video technology
? digital video
? Automatic dot feature isolation from machine vision
1983 Ginsberg & Maxwell's "Graphical Marionette" developed at MIT
1985 Robert Abel and Associates create "Brilliance" using "digital rotoscoping"

The earliest work related to optical motion capture was the development of the "space resection" technique by Johann Heinrich Lambert in 1774 [footnote: better known for proving that pi is an irrational number and for developing the mathematical model of surface illumination used in computer graphics]. Resection allowed estimation of the viewer's position using geometric perspective. Knowing where the viewer (camera) is located is critical to measuring where image features (markers) are in 3D.

[photo: historic profile of Lambert]
Johann Heinrich Lambert

By 1840 French geodesist Dominique François Arago had suggested using Daugerreotype images for cartographic measurement [footnote: cartographic means related to making of maps]. By 1849 Aimé

Laussedat [footnote: the "Father of Photogrammetry"] became the first person to use photography to create maps. In 1867 he exhibited the "Photothéodolite." It used resection techniques to join photography with the instruments of plane table surveying. At the same time he displayed a plan of Paris made using photographs taken from rooftops. Thus, cartography started to drive development of the mathematics that underlie photogrammetry.

[photo: historic plane table]
The instruments of plane table surveying.

From rooftops, cameras quickly moved up to balloons and finally "aeroplanes." Aerial survey cameras developed into “metric cameras.” Modern versions are constructed to tight tolerances and the optical system is measured so that it can be mathematically modelled. For example, commercial “metric cameras” allow measurements of 2-5 mm precision at a 50-meter distance.

In addition, the popularity of the "stereopticon" around the turn of the 19th to 20th century saw the development of the "floating mark," a means of measuring 3D depth. Eventually stereo plotting machines were built by Zeiss and others to make contour maps, e.g. of mountainous regions, from aerial stereophotos.

[photo: stereo plotting machine]
Stereo plotting machine

The first images created to measure movement were taken by Eadweard Muybridge [SIC] in 1872 [cite]. Initially sponsored by former California governor Leyland Stanford [footnote: of the railroad fortune and founder of Stanford University in California, USA] Muybridge began essentially helping to settle a wager over whether a horse's feet all left the ground at the same time during a trot. This lead to the development of motion pictures and (eventually) video systems during the 20th century.

[photo: historic muybridge]
Running horse by Eadweard Muybridge.

The first motion measurement technique directly related to entertainment was Max Fleischer's "rotoscope," developed in 1915 (patented in 1917). It was used by the Disney Studio in 1937 for the characters of Snow White and the Prince.

The field of Civil Engineering eventually developed a variant technique called "Close Range Photogrammetry." It uses multiple still photographs to measure, e.g. buildings or industrial plants. Features common to two or more photographs are identified by an operator and located precisely in 3D.

[photo: close range photogrammetry]
Typical use of photogrammetry for as-built diagrams.

Computers influenced all aspects of photogrammetry. Stereo plotting machines began using potentiometers and servos connected to a computer instead of mechanical gears. Later, photographs were scanned, digitized and analyzed directly in the computer in a process called "Soft image photogrammetry." An operator was still needed to identify common features in each of the images. For the process to be further automated, features in the image would need to be automatically identified.

Research in Robotics and Machine Vision yielded a means of finding discrete, high contrast markers (like white marker dots) in a digital image and calculating the 2D position of their centers (centroids) with subpixel accuracy. Motion Analysis, one of the leading companies in this area, was founded in 1983. At its founding the company had a 2D tracking technology for living cells seen through a microscope. Their video edge detection technology became the basis for their 3D tracking system in about 1986.

Machine vision also used close-range photogrammetry [cite: Atkinson90]. The math was simplified (with known error) in a paper on the Direct Linear Transform (DLT) [Cite: Abdul and Aziz?]. The DLT computes camera "metrics," including position and orientation, from the image of a reference “cube” seen through those cameras.

With this, the means were in place for automated determination of 3D marker trajectories. As the power of smaller computers increased, it became possible to label the marker trajectories and connect them to a hierarchic skeleton.

The first systems combining these techniques were used in the field of biomechanics for measuring human motion. This market for motion capture equipment remains larger than that of entertainment [footnote: the ratio is around 70% biomedical and 30% entertainment-related.]. The active marker optical systems show this heritage by being: (a) very easy to set up, with two cameras mounted on a single bar, and (b) having capture volumes optimized for biomedical use, that is easily suited to capturing a single person walking.

[photo: biomedical capture]
Biomedical motion capture using an optical system.

The first research use of optical motion capture for computer graphics occurred in "Graphical Marionette" (using an active marker system called SELSPOT). In the commercial realm Robert Abel and Assocites used rotoscope-like techniques to capture motion from 16mm film frames in 1985 for the TV commercial "Brilliance" [footnote: AKA "Sexy Robot" made for the National Canned Food Information Council.] Kleiser-Walczak later created a SIGGRAPH demo called "Dozo" in 1989 using 30 seconds of motion data captured with an early version of the Motion Analysis system.

[photo: frame from brilliance]
Brilliance, by Robert Abel and Associates.

History

Milestones in the development of commercial optical motion capture systems using passive markers:

?	Vicon dates?
1983	Motion Analysis founded
1986	Motion Analysis presents paper on 2D tracking to 3D ray tracking technology
1985	Sun-1 computer takes 17 hours to compute 8 points from 4 cameras on 3-second trial
1989	Kleiser-Walczak, Dozo, Motion-Analysis system tracks 30 seconds of a singer
1993	Acclaim, SIGGRAPH ’93, 2-character animation, 100+ passive markers, proprietary system
1993	Motion-Analysis, trc file, translational data format established
1994	First entertainment sale by Motion-Analysis (to US game developer)
1995	Motion-Analysis Segro, solves translation data to skeleton rotation on scaling limbs
1996	Motion-Analysis Mocap Solver, translation data to rigid skeleton
1996	Motion-Analysis auto-marker labelling software
1999	Dual 450 MHz Pentium compute 70 markers from 10 cameras in real-time

Automated capture of passive marker 3D position by optical motion capture was achieved by combining several technologies. These include: (most importantly) Close Range Photogrammetry, retroreflective material technology (e.g. 3m ScotchLite™), video cameras and feature detection techniques from machine vision.

[cite: information on the development of motion capture software at Motion-Analaysis corporation was provided during personal conversations with Mr. Pat Miller and Mr. John Greaves March of 2000.]

[cite:Mr John Greaves provided detailed information about early image processing work and use of the Walton thesis at Motion-Analysis in a personal conversation in May of 1999.]

Much of the work at Motion-Analysis in the late 1980s was inspired by a very detailed Phd thesis [cite: James Walton, Phd thesis, 1981 Pennsylvania State University, "Close-range Cine Photogrammetry: a Generalized Technique for Quantifying Gross Human Motion." Collegeof Health, Physical Education and Recreation. #8120471]. Motion-Analysis hired the author away from General-Motors research lab.

[Early Vicon history should go here.]

The machines of the early 1990s differed from modern systems in several important ways, each improvement having had an enormous impact on usablity. Camera calibration was more elaborate and less precise than today. Data that emerged from these early systems was strictly translational.

Driving a character with translational data was only possible with custom programming or in a few high-end software applications, such as Wavefront Kinemation, Softimage and Alias Power Animator [footnote: Alias and Wavefront were later both bought by Silicon Graphics and merged into Alias|Wavefront in 1997(?).] Also, given that the calibration process was less accurate, the data that resulted was more error-prone. Between the effort required to clean up the data and laboriously connect it to a character, it is clear why early complaints arose about motion capture. The data was precious because it captured a real performance, but the labor barrier set before its use was formidable. Early projects such as Kleiser-Walczak's "Dozo" needed to recycle 30 seconds of good capture data to fill several minutes of running time.

Because translational data was too difficult to use, the vendors started to develop a means to output joint rotations on full skeletons.

During an eight month period in 1995 Bart Gawboy (then at Nichimen, now at Anahana or Dreamworks Feature Animation) helped Motion-Analysis develop software that would map 3D trajectories to skeletal joint rotation. They incrementally built-up the technique by deriving rotation from combinations of 3 markers. Eventually their Segro technique used "virtual markers" offset onto the joints in the performer's body. The virtual markers were calculated as an offset from the actual markers. The difficulty with this technique was that as the real markers slipped on the skin of the performer the virtual markers would move the joint centers, causing limb lengths to change or "scale." Aside from the possibility of this looking "wrong," limb scaling had two results. Some kinds of scaling caused the infamous "foot sliding" behavior [footnote: discussed later in the Noise and Accuracy section.]. On the positive side, analyzing limb scaling provided a means of determining data quality. Fewer or smaller scaling changes meant better data had been recorded.

Optical markers have to be identified to the system through the course of a shot. Early on, after capturing the motion shot an operator would choose a pointon the timeline to designate which marker was which. While the early systems used various means of holding on to this identity information from frame to frame it was often lost, causing "gaps" in the data. As a result, re-identifying markers contributed significantly to the labor of optical data "cleanup." In about 1996 Motion-Analysis developed their first "statistical best fit" algorithms to automatically identify markers. This not only helped in establishing initial marker identity, but also in re-establishing it after the data dropouts when markers disappeared.

Finally in 1997 Motion-Analysis went after rigid skeleton solving. Their Mocap Solver software performs a best-fit of the skeleton's pose to the point cloud around the performer. The techniques behind this are still proprietary, but the result also allows a basic form of motion retargeting [footnote: retargeting is described in a later chapter, it entails mapping the performer's motion onto a character of different size or proportion.]

Back in the early camera systems arrived "pre-calibrated" from the factory for distortion. Over time their accuracy would degrade as they were bumped during transport or reconfiguration of the capture volume. Eventually customers were given the means to recalibrate their cameras, using precisely marked grids and measurement software. Even so, the process requires bringing the camera to a "lab" area, reducing system up time.

A significant advance in this is wand-based distortion calibration, commercially released in 1995 by Motion-Analysis. Lens calibration is carried out with the cameras already mounted around the capture volume. This is both convenient and more accurate.

Today's optical motion capture systems are among the most highly regarded (and highly priced) motion capture solutions available. They require significant training, but are then easy to setup, calibrate and use effectively. While the following sections do not provide system-specific training they do give a basis for understanding how optical systems work.

Theory of Operation

An optical motion capture system has the following parts:

- Reflective markers on the performer
- Cameras with co-axial flash illuminators around the lens
- Image Processing hardware to detect and isolate the marker dots in the video image
- An accurately measured "reference cube" which is reflectively marked
- A calibration wand with reflective markers

The performer wears reflective marker balls attached to their body. The markers are covered with a material from the 3M Corporation called "Scotchlite." It is a thin layer of thousands of tiny glass spheres attached to an adhesive backing . Each of the tiny glass spheres acts as a retroreflector through internal reflection. A light ray striking a sphere bounces along its curved inner edge until it emerges to return from the direction it entered. Scotchlite is the same sort of material used to make reflective traffic signs, which glow brightly in a car's headlights at night.

A set of specialized video cameras record sequential image frames of the capture volume. To do this:

- The illuminator ring around the camera lens flashes brightly as a video frame is being recorded
- The marker dots on the performer reflect the light almost directly back at the camera
- The camera records a monochrome image with the markers appearing as bright dots
- Image processing hardware isolates the marker dots in the image and records their position.

Isolating the marker dots in the video image from the camera is done using several techniques. These are background thresholding, coaxial flash illuminators with frame subtraction and dot motion prediction.

The first technique is giving the cameras a "threshold" or background value. This means they are set to ignore areas of the image below a given brightness. Some systems set an overall threshold, and some take a "snapshot" image of the empty capture volume and use that as the threshold. Light in the room is "ambient noise." The less noise, the greater the distance at which a marker can be clearly detected. Motion capture studios are kept dimly lit for this reason. Background features are less likely to be mistaken for markers. Setting a lower (darker) threshold value, allows a greater number of gray values for each pixel in the image. In a moment it will become clearer how having more gray levels improves subpixel resolution.

[photo: threshold off and on]
Thresholding (1) off and (2) on

The next two techniques use a flash illuminator. Recall that the markers are retroreflectors. The closer the light source is toa line drawn down the axis of the lens of the camera, the more light the retroreflector returns to that lens. A ring of LEDs is mounted around the lens of the camera (co-axially). Like a traffic sign at night, when the headlamps of the car line up, lots of light is returned to your eyes (which are somewhat co-axial with the headlamps).

[Photo FalconOn.jpg]
Co-axial illuminator on lens of motion capture camera

The illuminator's flashes are very short and synchronized with the frame rate of the camera. By making the flashes very short they can also be made higher power (brighter) without burning out the LEDs.

In addition, the LEDs are only flashed on alternate frames and the images are subtracted from one another. Subtracting the images from alternate flashes leaves only differences in the frames. In this case, the differences are the reflector dots, brightly lit in one frame and relatively dark in the next. Note that chrome spheres (or glass or mirrors at exactly the right angle) seen by the motion capture camera may be detected as markers.

[photo: frame 1, frame 2, substracted]
Frame subtraction: (1) first frame (2) second frame (3) frame difference

The "cleaned up" images are then searched for distinct bright blobs. This is sometimes done with proprietary parallel processing hardware, often based on one or more high speed Digital Signal Processor chips (DSP). The center (or "centroid") of each blob is calculated and reported. The position of each pixel in which the blob appears is averaged, using the pixel's brightness to weight the average. In this way it is possible to achieve "subpixel accuracy." There are additional aspects to this problem which go beyond what we can discuss here, such as correctly calculating the center of partially obscured markers.

The accuracy of a motion capture camera can depend on its ability to resolve gray levels as well as how many pixels across and down the camera imager has. As mentioned before, a high threshold value (needed in a brightly lit studio) will reduce the accuracy of positions reported for small markers. Calling a marker "small" here refers to moments when the bright blobs it produces covers only a few pixels on the camera's imager.

To this point the position of the blob's centroid has been calculated with subpixel accuracy on the camera imager. But, lens distortion hasn't been considered. From the BTS site:

> *Owing to the particular processing performed, the centroid of the marker is computed with resolution of 1/65536 of the field of view. The final accuracy, accounting also for distortion corrections over the whole field of view, is experimentally proved to be 1/2800 of the field of view, even in case of detection of very small markers (i.e. the accuracy with a 280 mm field of view and a 0.8 mm diameter marker is 0.1 mm).*

As we'll see shortly, measuring lens distortion is an important step in calibrating cameras.

Overall, the resolution of the system depends on enough light energy reaching the camera lens, and the marker blobs covering enough pixels on the camera's imager. Light energy is dependent on the power of the illuminator, the reflectivity of the marker, the lens opening & shutter speed of the camera. Covering enough pixels depends on the focal length of the lens & the depth of the capture volume.

Once the blob produced by a marker is located there are several ways to continue to track it in subsequent frames. The simplest method is to look in a box where the dot was last seen. In addition, people and their markers tend to make smooth continuous motions when observed 60 times per second. For that reason the dot's frame to frame position can be plotted and used to predict where it's going.

It is possible to scale up the size of the capture volume using larger markers. I.e., if the cameras are moved out, the studio darkened and larger markers are used, it is possible to capture bigger motions. Contrariwise, fine facial motion is captured using smaller markers and by bringing the cameras closer. The system's resolution is not changed, but it is spread out in the first case and compressed in the second. Applying resolution where it is needed in a capture volume is a critical part of a setup.

Once the marker dots have been isolated at each camera data about them enters the processing pipeline.

Here is how this will be described in the next two sections:

First, parameters of the capture volume are derived for the pipeline with:

- Process to calibrate camera lens distortion
- Process to calibrate relative camera positions & orientations

Then the marker dot information from the cameras is transformed into 3D trajectories with:

- Process to integrate 2D dot positions from multiple cameras into 3D trajectories
- Process to label trajectories and cross gaps of missing data

In the next chapter the 3D trajectories will be converted to skeleton motion with:

- Calibrating a skeleton to the segmented markers in the rest pose
- Process to keep the model of a performer's skeleton connected to the segments

The following sections cover the theory behind the steps outlined above.

Setup and Calibration of the Capture Volume

Setting up an optical motion capture volume entails:

- Calibrating lens distortion beforehand (if using the "grid method")
- Basic Setup of the capture volume by placing the cameras
- Adjusting each camera and Refining the volume
- Calibrating each cameras' lens distortion (if using the "wand method")
- Calibrating to determine each cameras relative position and orientation

Lens Distortion by Grid Method

From the history section above recall that correcting lens distortion is an important means of increasing accuracy. In commercial systems There are two popular procedures to do this: theolder grid method and the newer wand method. The principle behind both is recognizing how an image of known geometry is being distorted. A correction is filed with the camera to be applied to future images seen through its lens. Vicon calls this "linearization," Motion-Analysis calls this "camera calibration." In photogrammetry "squaring" images into an orthographic view is called "image rectification."

The lenses to be used need to known and mounted on the camera before this step. Unscrewing a lens and then reattaching it (even the same lens) changes the calibration of the camera. While the details of "back focus" are beyond this book, it also needs to be completed before distortion is calibrated. This is adjustment of the imager plane relative to the lens mount.

The grid technique is performed with the cameras away from the motion capture studio. The camera views a grid of precisely made marks, to which it is positioned orthogonally and exactly centered. Software examines the marks seen by the camera and determines the lens distortion. When done properly the technique works well enough. Its chief disadvantage is the that cameras have to be taken to a "lab area" where the grid is set up. If the cameras are significantly bumped (e.g. during transport), or the lenses unscrewed, the calibration needs to be done again, requiring another trip to the lab. [cite: Vicon manual] Systems using this technique ship their cameras pre-calibrated and recommend re-calibrating at three month intervals [cite: Vicon manual].

[photo: camera and calibration grid]
Calibrating a camera with a grid.

The newer "wand technique" for calibrating lens distortion is described below at the point in the setup process where it will be performed.

Basic Camera Setup Around the Capture Volume

Setting up a volume is easiest if at least three people are available:

- a grip who moves from camera to camera making adjustments
- another grip to place and hold markers at points in the volume
- a technician to "look through the cameras" on monitors and call out adjustments

If more crew are available then multiple cameras can be adjusted at once.

Basic setup of the volume proceeds with these steps:

- Use a recommended setup or plan a custom volume around the motion envelope
- Place cameras around the volume
- Background check
- Create "worst case" marker pair
- Test coverage & range by walking markers around the volume
- Optimize camera pointing and position

Most motion capture is done using a simple square or rectangular volume. Manuals provided by the manufacturers have recommended camera positions and lensing based on the number of cameras available. These basic volumes can also be turned or tilted to accomodate motion that rises or falls in a simple way. Something important to note when judging whether a default setup will work is that they usually provide optimal coverage in a pyramidal shape at the intersection of cameras.

[Diagram: 3D capture area]
Volumes covered by default camera setups tend to be pyramidal.

Variations on the basic capture volume are the next option. A common one is reserving 2 or more cameras from the total available to handle occlusions. These additional cameras can be added later, e.g. looking under an overhanging platform or near a wall, essentially where ever occlusion or ghosting problems indicate.

Designing a custom volume require time be set aside for trial and error testing to perfect the setup. This can be reduced with planning that avoids the obvious problems. Some important things to consider are:

- 2-way (or better 3-way) camera coverage of motion in the capture volume
- Visibility of a central area to all cameras (or a series of these, see below)
- Maximizing angles between cameras (see below)
- Where occlusions with props or scenery may occur

Manufacturers can provide charts that show camera viewing angle based on lens focal length. These can be used to draw out the custom volume on paper, or better, prototype it using carefully scaled models in a 3D package.

Once there's confidence that the design of the volume will work "on paper" the cameras can be set in place in the studio. Some general principles to keep in mind when setting the cameras are:

- **That the cameras will be undisturbed**
- That no reflective surfaces are visible to the cameras (especially spherical or curved ones)

- That the cameras see dark, low-contrast backgrounds beyond the capture area

As mentioned, optical motion capture makes precision measurement of camera position and orientation. In a studio setting cameras are best attached to a ceiling grid. This is an excellent way to keep them out of harm's way. However, if the cameras will be set up on high tripods additional precautions should be taken to keep them from being disturbed. The legs can be sandbagged and a zone cordoned off around them. A bumped tripod will ruin calibration.

[photo: camera on tripod]
A properly guarded camera tripod.

The most important thing to do in setting up a volume is to look critically through the cameras. On this first pass, check to make sure the cameras see only dark areas beyond the capture volume, and that there are no reflective surfaces in view. Use black staging cloth such as duvoteen to cover problem areas. Windows and doorways showing daylight should be blocked.

The next item to be checked is how many cameras are covering the areas in the volume.

The first way to test for camera coverage is to place pairs of markers in a large circle within the borders of the capture volume. The markers should be the same size as the ones to be used for capture. Each marker pair should be spaced apart as closely as the closest two markers on the performer. This is often the distance between the ankle and foot markers. We call this a "worst case pair." Seen from each camera, the two markers in the pair should be distinct and not "blob together." Additional refinement of camera settings will affect this in the next steps [cite: Vicon manual 2-4]

[photo: marker pairs around border]
Marker pairs around the volume.

Next, a pair of markers should be "walked around" the volume to test that all points inside the space are covered. This includes holding the marker pair overhead. The markers should continue to be viewed through the cameras. At this point the active part of the capture stage should be marked with tape for the performer.

It is adequete for all areas inside the capture volume to be visible to two cameras and at best three cameras[footnote: But see the comment on "ghosting" in the "Noise and Accuracy" section below]. At the time of this writing, real-time optical preview operates at better frame rates with *no more than* three cameras covering all areas of the volume. More cameras may solve an occlusion problem, but could cause real-time frame rate to drop, requiring a return to "batch mode" post-processing of the data instead of capturing the real-time data.

While a marker will remain bright at some distance from the camera, eventually it will be so small that it appears as a single camera pixel. When this happens the marker blob image analysis no longer calculates a sub-pixel accurate "centroid." The general rule is to ensure that markers are *never less* than 2 scan lines tall when at their most distant point in the capture volume. Markers larger than six scan lines do not add additional accuracy to centroid calculation [cite: personal conversation with Pat Miller at Motion-Analysis]

In addition to being panned and tilted, the cameras should be rotated to avoid "dead areas" in their view. Ideally the entire area viewed by the camera will be an active part of the volume.

[diagram: bad and good camera view of volume]
Cameras should be turned to reduce dead spots in view.

If there are problems with camera placement this is the time to change tripod position, lens focal length, etcetera. As the refinement continues these gross changes will be harder to make without affecting the set up of at least two other cameras. Bear that in mind during volume setup: changes never affect one camera, but always groups.

Adjusting Camera Settings & Refining the Volume

Once the basics of the volume are set the cameras can be adjusted. While we are describing them separately, in practice this step and the last one will be performed iteratively as refinement proceeds. Camera adjustment steps entail:

- Focus check
- Exposure check
- Coverage check

The motion capture cameras are usually fitted with very wide angle lenses [footnote: a lens with a lower focal length number has a wider viewing angle.] Two basic points about focus. Focus primarily needs to be considered for markers at long distance from the camera. Focussing the image of a distant marker gives a longer range in which it will be properly tracked. The range of distances where an object remains focussed is called the "depth of field." This is affected by focal length of the lens and the aperture (aka lens opening or f-stop). For full body capture in typical volumes with recommended aperture setting the depth of field will be infinite and the lens can remain focussed at infinity.

[photo pair: defocussed marker, focussed marker]
At a distance, a defocussed marker is no longer visible.

In adjusting the camera's exposure controls "overexposure" is the primary concern. Starting with each camera's lens aperture set to the manufacturer's "default" watch the monitor as a marker make a close pass. If the image of the marker "blooms" in size because it is too bright it can occlude other markers (sometimes all other markers). Although a bright ring flash and wide aperture seems like the best solution longer range this reduces depth of field and worsens blooming.

[photo: overexposed marker]
An overexposed marker "blooms" to occlude others.

A final coverage check should be made of the volume to ensure that no matter where a marker is, it is always seen by two to three cameras.

Once this refinement step is completed, and it may take some time to complete, the capture volume should be marked out with tape on the floor of the stage.

Wand Calibration for Distortion

If the "wand technique" will be used for lens calibration it is done at this point in volume setup. The “calibration wand” is a rod with a set of reflective markers along its length set in a non-symmetric pattern. The technician setting up the motion-capture space activates the calibration software, which starts the cameras recording. The technician walks through the entire capture volume several times, waving the wand so that the markers pass through most of the space. The calibration software uses the changing perspective on the wand’s markers to determine each camera’s parameters, e.g. focal length, lens/imager alignment, and (critically) lens distortion. [cite: Motion-Analysis manual]

Note that while both Vicon and Motion-Analysis use a wand during calibration, they use it for different purposes. [footnote: as of this writing; but software develops rapidly.] Vicon uses the wand strictly to set up the camera position/orientation relationships, whereas Motion-Analysis uses the wand to determine lens distortions and refine camera position/orientation relationships after the "cube calibration" mentioned below.

In addition, the Motion-Analysis software also does a continuous calibration, which uses the marker movement during capture to make small improvements or corrections to the camera metrics. [cite: Motion-Analysis manual]

Calibrating Camera Position and Orientation

Once the camera's are calibrated (or linearized) a procedure is used to determine their position and orientation relative to one another. Once this calibration begins the cameras can no longer be moved. They must be kept rigidly fixed in position and orientation. Again, both Vicon and Motion-Analysis calibrate the volume in different ways.

Motion-Analysis uses a "reference cube." This is a precisely measured open metal frame with non-symmetric markings. Reflective markers are precisely attached to the object. The cube is placed in a spot such that it is visible to all the cameras simultaneously. Each camera has a unique perspective image of the cube's markers. This determines both the scale of the capture space, because the cube is of a known size, and the position of each camera relative to the other.

[photo: reference cube]
Typical reference cube.

Rather than requiring all cameras to see a reference cube at once, Vicon uses "wand waving" to calibrate the position and orientation of the cameras. At any point in this process the wand must be visible to at least three cameras, and it must pass before all cameras at least once during the calibration. The calibration software "stitches together" overlapping capture volumes into one, possibly oddly shaped, fully calibrated volume. While the Motion-Analysis system can also create oddly shaped capture volumes by "reference cube shifting" the Vicon technique is very convenient.

[photo: wand-waving]
The wand-waving technique.

Given the two kinds of calibration information, (a) metrics of each camera's optical path, and (b) each camera's relative position and orientation to the other, processing software can now make the conversion from 2D dot positions to 3D marker positions. To do this, a line is mathematically extended from the lens of the camera toward one of the glowing reflective markers. When lines are extended from several cameras toward the same marker the lines will intersect, or at least have a point of closest approach. This is the point in space where the system reports the position of the marker.

[diagram: triangulating marker position]
A marker position is optically triangulated.

At a minimum two cameras are needed [footnote: See "Noise and Accuracy" section below about why 3 cameras are needed to avoid the appearance of "ghost markers."]. The most accurate intersection is one where the lines meet at 180 degrees divided by the number of cameras viewing (for two cameras: 90 degrees or perpendicular). If the cameras viewing the same marker are very close together the intersecting lines will be almost parallel, the "angular difference" will be low and accuracy reduced. While a very wide angle between cameras is usually impossible, spread them out around the capture area as much as possible.

Position and orientation calibration should be done before the start of capture, after breaks or pauses, whenever cameras are bumped and after the end of capture. Optical equipment requires a "good" calibration to process a shot into usable data. The worst case would be to have a an otherwise excellent take lost because the calibration was "off."

Setup and Calibration of the Performer

Optical motion capture's basic output is 3D markers moving in time, called "marker trajectories." This is also referred to as "translation data" and it is possible to directly animate a character with it. This is somewhat laborious and is done by constraining parts of the character to follow the markers. The technique can be advantageous for unusually constructed target characters or motions involving precise contact with objects. [footnote: more about this appears in section X on page Y.]

Optical motion capture systems also process their translation data into rotations of bones in a hierarchic skeleton model of the performer. This process will be described in the next section on "Skeletons, Motion and Motion Mapping."

In either case, the goal is to put a minimum number of markers on the performer to capture their motion. This means that a skeleton must be designed, similar to that of the performer, along with a "setup" of markers on and around the joints.

In this section we will show a typical optical marker setup for capturing human body motion and explain why it works.

Setting up and calibrating the performer requires:

- Creating a capture skeleton and defining its joint motion limits
- Designing a setup of reflective markers around joints to capture all their motion
- Defining segments (or groups) of markers that reveal the joint motions

In practice, this step is tightly coupled to the segmenting and calibration processes described in the next section on data processing.

If basic human or animal motion is being captured there are standard marker setups available.

[diagram: human marker setup on performer]
A 37 marker setup for capturing human motion
(courtesy of Motion Analysis)

It's useful to know why the standard setups work for two reasons: (a) if it is failing to provide enough data there is a chance of knowing why and (b) when the need arises to create a new marker setup the principles are understood.

While a hierarchical skeleton has many joints, they don't all have complete freedom of motion. They can be classified into several types: [cite: Motion-Analysis EVa Mocap Solver Manual section 2 page 3]

- fixed - a non-rotating "joint"
- hinge - such as the elbow
- universal - like the ankle
- ball - for example, the shoulder

A so called "fixed" joint represents translation in space. The root of the skeleton in space is a fixed "joint."

[diagram: root joint]
The fixed "joint" at the root of a skeleton.

A hinge joint joins two parts with one degree of rotational freedom. The elbow is a good example of this.

[diagram: elbow joint]
The hinge joint at the elbow.

Universal joints allow two degrees of rotational freedom.

[diagram: ankle joint]
A universal joint at the ankle.

Ball (or spherical) joints such as the shoulder allow rotation on three axes.

[diagram: shoulder joint]
The ball joint at the shoulder.

Categorizing joints helps in designing where markers should be placed on the performer. Sets of markers, taken together as a group called a **segment**, reveal the motion of a joint. The more complex the motion a joint performs, the more markers are needed around and on it in the segment to reveal that motion.

A single marker attached to a performer can return information with 3 degrees of translational freedom. A single marker can position the root of the figure, or indicate where the edge of a prop is located in the capture volume.

[diagram: single marker freedoms]
A single marker segment gives data with 3 degrees of freedom.
(courtesy Motion-Analysis)

Two markers can of course provide two separate 3D positions. However, if they are related to one another as a segment, e.g. along an arm, they give a position, a distance and two rotations. A two marker segment can be used to position and orient a cylinder (but a sword needs more info as will shortly be shown).

[diagram: two marker freedoms]
Two marker segment gives data with 5 degrees of freedom.
(courtesy Motion-Analysis)

Imagined from the viewpoint of the cameras, a two marker segment does not provide enough data to determine rotation of, e.g. a sword in the hand or of the forearm of a performer. For this, a three marker segment is needed.

[diagram: three marker freedoms]
Three marker segment provides data with 6 degrees of freedom.

It would be ideal if markers could be attached directly to the bones of the performer. But, this surgical approach is only used in the most serious biomechanical research [cite: paper where that was done]. We must settle for second best and attach the markers to the skin of the performer. The best places will be areas where the skin doesn't move too much due to motion. Some examples are over the points of the spine, on the bumps made by the pelvis on the stomach, at the elbows, etc.

In the early days of motion capture markers would be arranged in "triangles" along each bone. Taking this approach with modern systems adds unneccesary processing overhead.

Using fewer markers means faster processing, or more marker capacity left over for other parts of the scene. The typical human marker setup shown earlier captures motion with a minimum of markers. In part, it does this by sharing markers between more than one segment. A marker which indicates how the elbow is swinging off the shoulder is also used as the base of the forearm's rotation. The shared "segmenting" of the marker network naturally matches the jointing of the limbs to which they're attached.

[diagram: marker joined into segments]
Elbow markers are shared to both indicate shoulder rotation
and be the base of forearm rotation.

Here is a diagram that shows marker segments relative to the skeleton they will drive.

[diagram: skeleton and segmented marker set]
A skeleton positioned inside a segmented marker set.

The diagram above clearly shows how the segmented markers form an "exo-skeleton" around the performer which is related to the "endo-skeleton" model of the performer [footnote: as described previously in the generic model section].

After post-processing, described in the next section, markers have no inherent identity and need to be "labelled" so that the segments are known. Modern software systems can perform an "auto-identification" to label the markers. To do this, a few asymmetric markers are placed on the performer to distinguish left from right automatically.

[diagram: asymmetric markers]
A few markers are placed asymmetrically to show right from left.

When high contact motion is being captured, and markers are likely to be hidden, redundant markers may be needed. Markers can also be extended on stalks, as is sometimes done around the wrist. Wrist stalks avoid markers being hidden when a performer holds a prop like a weapon.

The relationship of markers, segments of markers and the joints of the performer's skeleton should now be clear. While markers and their segments are outside of the body [footnote: the exo-skeleton mentioned in the generic model section], the joints being measured are inside. In the next chapter we'll describe the process of applying translation data to a skeleton in more detail.

Once the performer's marker setup is complete and on their body both Vicon's BodyBuilder and Motion-Analysis' Mocap solver create a "template" before post-processing. The performer stands in a reference pose, the same neutral pose in which their capture skeleton was built. Camera data is recorded as they do a quick series of movement designed to show the range of all limbs. Markers are identified. The 3D cloud of points around the performer is segmented. The previously built model of the performer skeleton is aligned inside the segment exo-skeleton. The software then builds the template using the motion data. Offset are calculated to relate the exo- and endo-skeletons, as well as joint motion limits. This template will be used to map marker trajectories onto the performer skeleton.

Post-processing of Optical Motion Data

In the generic motion capture model the system hardware generates "raw sensor" data in a native format. For an optical system this is 2D image data from the cameras. Unlike mechanical and magnetic systems, processing optical data requires more interaction with an operator.

Converting camera data to a set of marker trajectories is called "reconstruction" by Vicon. The trajectories are passed into the rest of the motion mapping pipeline described in Chapter 3 "Skeletons, Motion and Motion Mapping" in the section on "Motion Mapping Translational Data to a Skeleton."

The software tries to follow the marker's motion for the length of the entire shot. If all goes perfectly the system locates a number of tracks equal to the number of markers on the performer. But, various problems cause the markers to disappear and re-appear during the shot. This breaks the tracks into pieces. Heuristics can sometimes "reconnect" these pieces automatically. But the heuristics can also cause a marker "exchange" when two markers disappear and then are mis-identified when they re-appear. Thus, after processing the typical problems remaining are gaps and exchanges.

[snapshot: marker tracks with gaps]
Marker tracks on a timeline with gaps.

As a marker moves the software searches in a box near the last position it was seen and along the path it was travelling. Box size can be lowered when smaller markers are used. Likewise an acceptable maximum marker velocity can be set.

[image: tracking box]
A tracking box and motion prediction are used to cross small gaps.

If a marker disappears, the system calculates a curved track from its last few visible positions. The track is used to predict where the marker is likely to be. If a "new" marker reappears where the curved track predicts, it is assumed to be the same marker. If the marker stays "gone" for too long it is taken to be lost and the software creates a new track if the marker reappears.

Markers are shaped like a sphere so that when partially obscured it is still possible to calculate an accurate centroid.

[image: partially obscured marker]
Round markers provide a known centroid even when partially obscured.

Further problems related to how tracks can be broken up are covered in the next section on Noise and Accuracy. Editing tools used to repair gaps and exchanges will be described in a later chapter.

Noise and Accuracy

Different kinds of noise and reductions in accuracy appear at different stages of the data processing pipeline. In this section we discuss noise and accuracy reductions that occur during reconstruction of marker trajectories. Later sections will cover problems that occur, e.g. during mapping of motioin onto a performer skeleton using an inverse kinematic solver.

Diagnosing noise and accuracy problems means paying close attention to where in the volume, or where on the performer the problems are happening. E.g. if trajectories are consistently clipped off or noisy when a performer reaches for an overhand tennis shot it is likely that not enough cameras are seeing the markers at the top of the volume, or perhaps the markers are dirty on the side that's exposed during that move.

Primary sources of trajectory noise:

- occlusion
- ghosting
- markers leaving the volume
- bad threshold level
- bad ring flash or exposure setting
- dirty markers
- reflective surfaces in view
- stray light

Primary sources of reduction in accuracy:

- marker slippage
- bad camera coverage of the volume
- bad volume calibration

While the retro reflection from the markers is bright, visual noise can occur due to shiny reflective surface, especially rounded shiny objects (a bowling ball or motorcycle helmet would be examples). These surfaces can produce "false markers."

The background image seen beyond the capture area should be neutral. Bright lamps, especially flickering ones, or ones that people walk in front of, will generate false markers in the data.

Other solutions for threshold, exposure, marker slippage and camera placement problems are described in earlier sections.

If fewer tracks than the number of markers are consistently produced it is likely that the markers aren't visible to the system. They may be consistently hidden from camera view, behind the performer or a prop, or outside the volume. In addition, there could be a weakness in the chain of illuminator, reflective marker, lens, f-stop, shutter speed. If the problem happens consistently it is time to retest the volume by looking at the markers through the cameras.

More tracks than markers are caused by two common problems in identifying markers for the length of a shot: occlusion and ghosting. These are also called "correspondence problems" because when they occur we don't know which blob in a camera image "corresponds" to which marker.

The markers are just dots on an image. When the performer turns around and two or more markers are eclipsed (or occluded) by their body, when they re-appear it is no longer certain which is which. This is the "correspondence problem." Which marker corresponds to which dot as seen before the dots were occluded? While several means are available to do so, it may not be possible to automatically re-establish their identity when they reappear. The operator may need to connect marker tracks across the moments of occlusion.

[photo: occlusion]
Occlusion of markers.

Serious occlusion problems are solved by adding redundant markers to the performer, or more cameras to the capture volume.

Ghosting occurs when two markers are seen by only two cameras, and the cameras and markers lie in the same plane. In this case there is an ambiguity about exactly where the markers are. Again, while several means are availible, it may not be possible to automatically re-establish the identity of the markers.

[diagram: ghosting]
Ghosting occurs in a specific two-camera configuration.

Ghosting is solved by making sure that each marker is seen by three cameras during capture. Note that adding more than three cameras does not always significantly increase precision, but will certainly increase processing time.

In an long motion capture shot it was common for several of the 32 markers on a performer to be occluded or ghosted several times. Re-establishing marker identity (or "correspondence") across the gaps of data required hours of manual labor doing what was referred to as "cleanup." Also, while adding cameras was desirable for accuracy and enlargement of the capture volume, it tended to increase the opportunities for ghosting or occlusion, and made re-identifying the markers to each camera harder. This optical data cleanup problem was the source of the early (and correct at the time) belief that it took almost as long to capture *and clean up* motion data as it would to hand animate it.

Much of this difficulty predicting marker positions was due to inaccuracies in modeling the optical chain from illuminator, to the marker, through the lens and onto the camera's imager chip. Early camera systems were noisier as well. Many of these problems have been eliminated through better camera placement and much-improved modeling of camera lens distortions. Recall that in the field of photogrammetry, from which optical motion capture derives, specially calibrated "metric cameras" are used to take high precision photographs. A part of the advancement in optical motion capture has been the creation of video "metric cameras" by using software calibration techniques.

Modern optical post-processing software loses marker identity much less often, and is able to automatically re-establish it more often as well. While the data cleanup problem was startlingly large, steady progress toward its solution means that capture is now much more efficient and reliable than in those early days.

Advantages

Passive marker optical motion capture technology has significantly improved from the early days of its use.

Its primary advantages include:

- Accuracy
- Highly configurable marker setup on performer and props
- Large, configurable capture volumes shapes
- Real-time preview

Optical motion capture with a properly calibrated system is very accurate. For this reason it has been used for many years in biometric applications such as gait analysis for physical therapy.

Disadvantages

There are two negatives to consider with optical systems:

- Cost
- Complexity of setup and calibration

Optical motion-capture systems, while extremely capable and versatile, are also at the top of their food chain in terms of cost. An entry-level system costs $250,000 USD (circa 2000).

Also, while it is not difficult once one is trained, the setup of optical systems is involved, and requires care at each step. Experts such as Pat Miller of Motion Analysis can set up an eight-camera system in less than two hours. The whirlwind of activity is quite something to see. In practical use what this means is that the shape and layout of the capture volume relative to the cameras must be carefully planned, as it takes significant time to reconfigure.

Active Optical Markers

Like passive marker optical systems, the active marker systems also use illuminated markers attached to the performer. The difference is active markers are powered to provide their own illumination and each flashes in a distinctive pattern that uniquely identifies it to the “camera.” This eliminates marker correspondence problems such as exchanges or ghosting (see above under passive optical motion capture). From a purely technological standpoint these systems are an ultimate evolution of optical capture. But there are interesting trade-offs.

Early History

Please see the "Early History" section above under "Passive Markers" for the development of the mathematics and technologies leading up to optical motion capture.

History

1975 Selcom, SELSPOT
1982 Selcom, SELSPOT II

1983 MIT, Ginsberg & Maxwell, Graphical Marionette, Op-Eye active markers track body position
1983 Northern Digital, WATSMART
1990 Northern Digital, Optotrak 3020
1994 Charynwood Dynamics, CODA mpx30
1999 PhoeniX Technologies, Visualeyez VS-2000

The SELSPOT, from Selspot AB in Sweden, was the first commercially available active marker system [cite: conversation with Manfred Berger, Innovision Systems March 24, 2000]. The imaging system was unique. Instead of a vidicon tube, the "camera" used a patented 2D photodiode as its sensing element. These non-imaging "cameras" were called "detectors" or "receivers." A flashing marker would register a pulse along the two dimensions of the diode surface. The diode could only register one marker at a time and tended to get errors from other infrared light sources in the field of view. Also, as the number of markers rose, the exposure time for each one dropped, causing the signal to noise ratio to worsen. This made it difficult to track a person wearing dozens of targets at once. The LED's were tethered to a controller that turned them on one at a time. This limited the wearer's motion somewhat due to the cables. [cite: paraphrased from Tracy McSheery's Phase Space competitor analysis documents.]

This early system set the basic design. Markers would be flashed one at a time so that they could be easily distinguished. However, this meant that with more markers either (a) each one would be "on" for a shorter time, or (b) the sample rate would drop as it would take longer to cycle through them all. It is also interesting to note that this early system used a separate detector (versus later designs using multiple detectors mounted in a bar).

The SELSPOT II was released in 1982. It used the same unique detector element in up to 16 cameras. It could view up to 120 markers and used many of the same calibration processes as passive marker systems.

In about 1983 the Op-Eye system was used in the Ginsberg & Maxwell "Graphical Marionette" project in the Architecture Machine Group at MIT. The Op-Eye used flashing, synchronized LEDs viewed by a pair of video cameras.

In 1983, Northern Digital was founded and released their first product, the WATSMART. The sensor was similar to that of the Selspot and 2 cameras were used to provide 3D data.

In 1990 Northern Digital achieved a significant advance with the Optotrack 3020. For ease of set up, three linear sensor arrays were mounted behind lenses on a rigid bar. This was the first system that eliminated the need to calibrate, greatly simplifying setup.

In about 1994 another variant design was released by Charynwood Dynamics. The CODA mpx30 uses a scanning mirror assembly to image markers onto a linear sensor without lenses. This eliminated the need to correct for lens distortions, and provided a field of view greater than 90 degrees wide.

One of the key reasons for using line sensors is that they reduce computational load (and cost) of the system. By 1999, decreasing chip costs allowed PhoeniX Technologies to build their Visualeyez VZ2000 system using normal camera CCDs. While a commercially availible CCD is used for imaging, custom hardware makes it act as a linear sensor with much better noise rejection characteristics. Three of these sensors are built into a bar and as an additional convenience a protective cylindrical outer shell splits to become its mounting tripod.

Thus far, active marker systems had used all manner of image detectors, from unusual 2D photodiodes, to line scan CCDs, to standard video cameras. But, in all cases markers were being flashed and scanned sequentially.

Northern Digital patented a system to flash up to 4 markers on at once. Calibration occurs with one marker on, then two, then three, then all four flashing together. While this seems a significant patent, it has not been demonstrated in a product yet.

Phase Space, founded in 1994, has demonstrated an interesting combination of features in their as yet unreleased product. Using separate cameras, with normal 2D CCD imagers, markers are tracked using digital signal processors. To establish and maintain marker identity Phase Space took the flashing concept one step further. They flash all markers sequentially once to establish identity. After this, all markers are flashed on every camera frame. Markers which are lost due to occlusion are sequentially flashed to re-acquire them. This eliminates much of the frame rate penalty usually incurred by active marker systems.

However the high cost of the systems, easily the most expensive available, kept customers in search of alternatives. Still, these systems are very easy to set up and modern units provide very clean data.

Theory of Operation

Active marker systems have the following parts:

- Illuminated markers (typically LEDs)
- A marker controller that coordinates flashing the LEDs in a timed sequence
- A set of "detectors" (similar to cameras) mounted separately or together on a bar

Active markers are coordinated to flash one at a time. This means that, while the markers can never be misidentified (only one is visible at a given moment), a given marker must wait for all others to be sampled before it can be examined again. This is the first tradeoff in an active marker system. While the sample rates for these systems range as high as 2,400 Hz the sample rate for a set of “n” markers is actually max sample rate divided by n.

For a small set of markers there is no need to coordinate flashing with the camera’s frame rate. But, as the number of markers becomes greater than 32 a system is needed to synchronize the marker flashes with the camera’s sampling. This can be done with various kinds of links to the LED markers, including radio, or a triggering IR flasher on the camera.

While passive marker systems carefully refined their hardware design, radical change was happening "under the hood" in software. In contrast, active marker technology has continued to evolve its hardware.

“Cameras” for active marker systems are configured either separately, as for passive optical systems, or with several mounted inside a bar. Systems with separate cameras are substantially similar to passive systems in calibration and theory of operation, except that markers never lose identity. Bar mounted systems are an interesting variation.

[Illustration: bar mounted optical receivers]
Determining position with bar mounted optical receivers.

A bar-mounted system has three camera-like optical receivers mounted along a rod about a meter long. Relative to the rod, the two receivers at the ends determine depth and yaw of a marker, while the center receiver determines pitch. Each of these signals can be visualized as defining a plane. Combining the yaw, pitch and depth information is in effect intersecting the 3 planes to result in a single point in 3D. [cite: conversation with Dr. Ma from PhoeniX Technologies]

The optical receivers can be 2D CCDs (i.e. like the imager in an electronic camera), a 1D line CCD. While a line sensor can be cheaper and its output somewhat easier to process, the 2D imager provides some redundancy.

The optical receivers use a lens, pinhole or scanning mirror arrangement to focus light on a detector. With a lens, a narrower field of view is created, and the targets can be less bright or detected to greater depth. Without a lens, the reverse is true, a wider capture volume is available but the markers must be brighter (higher powered).

Because the detectors all lie in the same plane a bar-mounted setup is susceptable to "ghosting problems" (see the "Noise and Accuracy" section under "Passive Optical Capture"). However, because the markers are self-identifying, this cannot occur.

A forthcoming system from a small company called Phase Space uses active markers with separate “smart cameras.” As mentioned previously most active marker systems flash their LEDs in sequence, one per frame. This causes the sampling frame rate to drop to max rate / n where “n” is the number of markers being sampled. The Phase Space system uses “smart cameras” with built-in digital signal processors (DSPs). At the start of a capture session the system flashes the markers sequentially to establish their identity. Once this is done, all markers flash on at every frame and the DSPs track each marker until they are occluded or the capture ends. If one or more markers is occluded those markers are flashed to re-establish their identity to the cameras. In this way, the temporal resolution of individual markers may drop for a moment after they are occluded, but on average the frame rate of capture remains as high as the maximum possible frame rate of the cameras. The Phase Space system’s markers are coordinated using a radio link.

Setup and Calibration of the Capture Volume

Active optical systems are not usually packaged with extensive software to support motion capture for entertainment. This is partly because bar-mounted units have almost no need for calibration, and partly because translation data mapping is now supported in several 3D animation software packages.

Using a single bar-type receiver a "camera facing" capture volume needs no setup at all. The unit is powered up. There is usually a minimal calibration step, such as placing a single marker where the origin of the coordinate system should be located.

Northern Digitial supports coordinating multiple receivers to create larger or oddly shaped capture volumes. The calibration process for this is still extraordinarily simple, requiring capture of 75 frames of a "reference frame" object in view of all receivers.

Setup and Calibration of the Performer

Like passive optical systems active systems output 3D marker trajectories (translation data). Marker setup on the performer is substantially similar to that of passive optical systems. Please see that section for more information.

Noise and Accuracy

Active marker systems have noise and accuracy issues similar to passive markers.

Occlusion of markers and markers passing out of the capture volume (line of sight requirements) are the primary source of data gaps. Systems based on 1D optical line sensors are more sensitive to ambient lighting, which adds noise. 2D image sensors are less sensitive to noise, but require more processing power to resolve data and may cost more. Finally, as for camera-based passive optical systems, active marker machines with lenses must compensate for distortion, which again requires the cost of more processing power.

Ghosting is impossible because the markers are self-identifying, hence only two receivers are needed.

Advantages

Real-time data marker position data is provided by these systems.

Camera-facing capture volumes can be setup with almost no calibration. A single marker is placed to determine the origin and facing of measurements made in the volume. Because the cameras are rigidly mounted at fixed positions inside the bar they need not be calibrated to one another.

Very high frame rates are achieved when a small number of markers are used.

Non-lens systems need no distortion correction. Lens-base systems can be factory calibrated against distortion.

Reasonable capture frame rates can be achieved with the number of markers on a single performer.

Large capture volumes can be created with multiple receivers.

Disadvantages

Precision, bar-mounted receivers are more expensive than individual cameras.

Wiring on active markers may be encumbering.

Single receiver systems have shallow "camera-facing" capture volumes.

Other Technologies

In the following sections we offer a quick look at some existing or forthcoming technologies. While these are not currently in broad use for human motion capture in entertainment (some are used in the medicine) they may help understand forthcoming developments in the field.

Bend-sensors

Taking an entirely different tack, it is possible to construct a motion capture system using bend sensors. One technology that has been used is a fiber optic cable whose exterior has been roughened [cite: VPL glove]. When such a cable is bent it modulates the amount of light returned by an illuminator. Certain types of metallic strips will change their resistance when bent. In this latter type of device reducing sensitivity to temperature is a factor [cite: Measurand ShapeTape™].

It is also possible to develop a completely "soft suit" with flexible bend sensors built into body hugging Lycra clothing [cite: Virtual Technologies]. Using such a body-hugging suit does not solve fit problems. While the sensors remain close to the body they will be in different positions for persons of varying proportions and, as for a completely mechanical system, the bend sensors respond several at a time to the motion of the performer's skin surface in a complex manner. In a body suit and bend sensor system the goal is to understand how the suit is stretched over the performer's outer skin, where the sensors are, and ultimately how the skin of the performer bends with respect to their joints.

As was mentioned by Seth Rosenthal of Lucas Digital [cite: VR News interview] it would be desirable to capture the motion of actors while they were being filmed on a set. Virtual Technologies' body suit and bend sensor system could be worn under an actor's clothing (it was originally designed to fit inside NASA space suits). In this way it could unobtrusively record a thespian's movements while keeping out of sight.

If there is ever a use for motion capture technology in consumer electronics it will likely be a body suit and bend sensor style system, rolled up in a bag, that connects to the Nintendo game systems of the next decade.

Inertial Sensors

As of this writing there are no commercially available body motion capture systems based exclusively on inertial motion capture technology. Individual sensors are available from interSense and Ascension. But, inertial sensors' high cost makes a large number of them unsuitable for use as a body motion capture system.

This having been said most professionals in the entertainment industry look forward to a time when this very promising technology is developed into a cost effective means of body motion capture.

Inertial systems could achieve the best characteristics of a magnetic motion capture system, i.e. real-time operation, orientation as well as position data, but without accuracy degrading sensitivity to magnetic fields and metal objects.

In the meantime, inertial sensors are providing absolute position and orientation data for: mechanical capture suits, virtual set cameras and head mounted virtual reality displays.

Acoustic Systems

Acoustic position measurement systems measure the flight time of a sharp audio pulse from a transmitter to a receiver. The speed of sound in air makes this type of measurement relatively easy.

Science Accessories introduced a sonic digitizer in 1969 with a significant effect on the digitizer field. At that time, a tablet digitizer had an average price tag of $100,000, compared to SAC's two-dimensional GRAF/PEN around $5,000. In 1972, SAC filed its first patent for a three dimensional device using three linear microphones arranged at the corners of a cube. It made it possible to digitize three-dimensional objects to obtain all external measurements and to track the motion of moving parts; both animate and inanimate. Applications have included animation, athletic performance measurement, shape and position digitizing of archeological findings, molecular model creation, rigid body analysis, architectural design and drawing preparation and drawing topological relief maps. When used in conjunction with various software packages the digitizer becomes a general scientific measuring tool with broad uses in medicine, industrial design and graphics. [cite: Tracy McSheery competitive analysis documents from Phase Space biz plan.]

At a consumer level this technology has been used to create a 3D mouse [cite: "Red Baron" from Logitech] and a theatrical spotlight to performer tracking system [cite: Wybron Autopilot].

Systems of this type are subject to echo problems. The sharply defined audio pulses easily bounce around a room.

While we mention this technology here, there are currently (2000) no commercial systems for performer motion capture based on it. This technology is primarily used in medical biometric applications [cite: Zebris, Germany].

Radio Frequency Systems

Within ten years we'll have the technology to perform tracking using radio frequency signals [cite: conversation with Tracy McSheery 1/27/2000]. The receivers would be non-co-located so that a receiver/marker on the performer could intersect power spheres to determine position.

The speed of light makes this a very high precision task for an analog to digital (A to D) converter. Given the speed of radio signals, currently available 1 GHz A to D converters would provide 30cm accuracy, not good enough. With a 10 GHz converter 3cm accuracy is possible, interesting but not precise. At 100 GHz 0.3cm accuracy is achieved, which could be the basis of a commercial product. Such a system's sample rate would be limited as is an active marker system, i.e. only one marker receiver can be turned on at a time.

There are many further technical details, with the transmitter and receiver having to send signals back and forth, etc. However, it will be interesting to see what happens when the underlying technology is commercially available, and whether it will be needed with the improvement of existing technologies.

Feature-based Systems

[insert Ioannis' paper here]