Capturing and Visualizing Large Scale Human Action

1 Introduction

A substantial part of the entertainment industry is devoted to the visualization of human activity. Frequently, these activities involve many people interacting. Examples include reality shows, dance, music and theatrical performances but most prominently sports events, especially team sports such as ice hockey, basketball and football. Today's experience of these live sports events is primarily limited to the viewing of TV broadcasts that only display the part of the action selected by the producer. This experience is substantially less than that of a spectator present at the event.

In contrast the computer games industry allows their players to experience the action from whichever viewpoint they desire. This is, of course, possible because a computer game involves only artificially generated actors and events. Unfortunately, these objects and events frequently appear very *artificial* especially with respect to the human motion. If real world actions and events could be captured using the techniques of computer vision and subsequently rendered and exploited in the game technology, this could potentially increase their realism and improve the experience of the computer game environment.

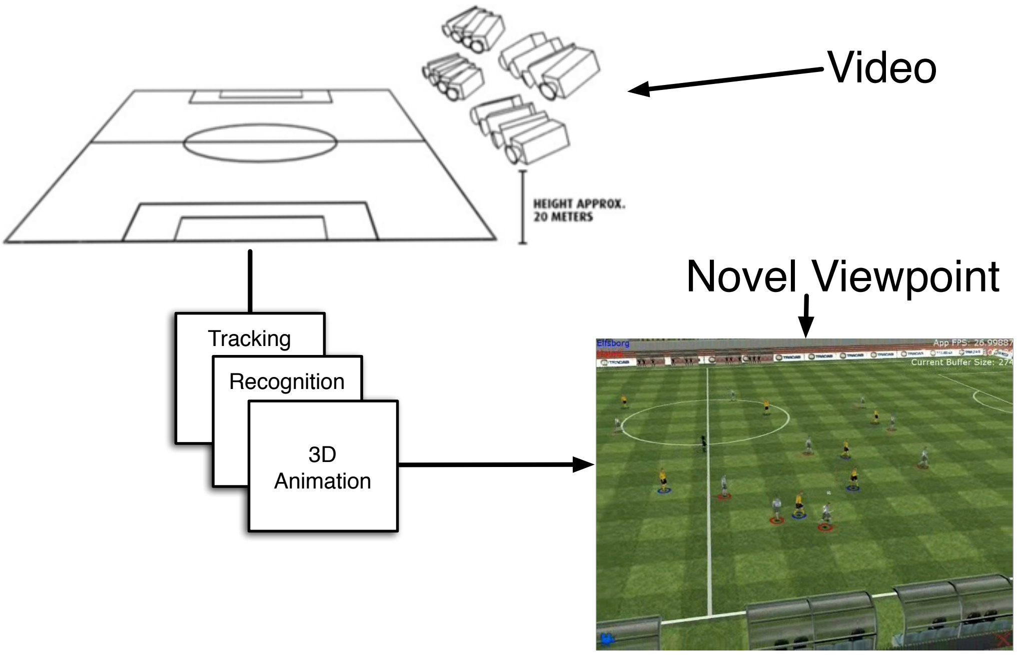


Figure 1: **Overview of the project's aims and methodology.** *We are proposing to build 3D visualizations of a football game with a particular emphasis on the motion of the players. This will be achieved by tracking the players in the original video data, recognizing their actions and poses and then reconstructing them in 3D.*

Therefore, we propose to integrate the realism of real world action events with the advanced visualization tools used in computer games. This integration will benefit and advance both the media and the game industry. Specifically we aim to produce a demonstrator with the following capabilities and commercial value:

Arbitrary viewpoint visualization. We will automatically capture the 3D motion of actors in an event such as a football game. This will allow the 3D visualization of the game and the spectator to choose any viewing angle via an application or browser plug-in on their PC or games console. Another benefit concerns transmission of data; the control parameters of a virtual game require much lower bandwidth than video data. Moreover, it is extremely costly to secure rights for TV broadcasts, an issue side-stepped by virtual reconstructions.

Technology for virtual visualizations draws on two large and rapidly growing sectors, namely the sports and entertainment industries. In 2006 the total entertainment market was a staggering USD 461 billion just in EMEA (Europe, the Middle East and Africa) with huge markets also in the Americas and Asia. European football generates EUR 11.6 billion in revenues to clubs and federations, representing half of Europe's sports business. A web-based 3D football product could additionally exploit the rapid growth in internet advertising, which has already passed 10% of the total advertising market in Sweden and the UK.

Enhanced realism of human motion in computer games. The captured 3D motion information will be used to enhance the realism of today's computer games which is often limited with respect to their rendering of human motion. This is partly due to the fact that the captured motion is commonly restricted to actions performed in a studio environment with the subject wearing markers and not those from a real game.

Within the entertainment market, gaming is also a vast sector. In the US in 2006, 17% of video games and 3.5% of computer games sold were sports games, resulting in a total market of USD 1.1 billion. The European and Asian gaming markets are of similar magnitude. The manufacturers of sports games are continuously striving to make their games as lifelike as possible, and are quick to embrace new technology which aids this goal.

Achieving these goals will require an advancement in the state-of-the-art in the visualization of events involving many humans performing a wide range of actions over an extended time. This advancement will involve automatically capturing the motion and activity of each participant to animate and render it in 3D. The increase in the realism of the visual rendering of the 3D animations will create an experience that is potentially beyond that of the spectator present an the event.

2 Partners & their expertise

The company TRACAB [18] and research group CVAP at KTH[6] will be the partners involved in this project. Both have experience in the field of real time data capture and visualization that complement each other well. The demonstrator is intended to enhance TRACAB's present commercial activity in the field. Via TRACAB, CVAP will gain access to football matches and a large amount of video and tracking data, yielding a considerable edge over equivalent projects conducted in universities elsewhere.

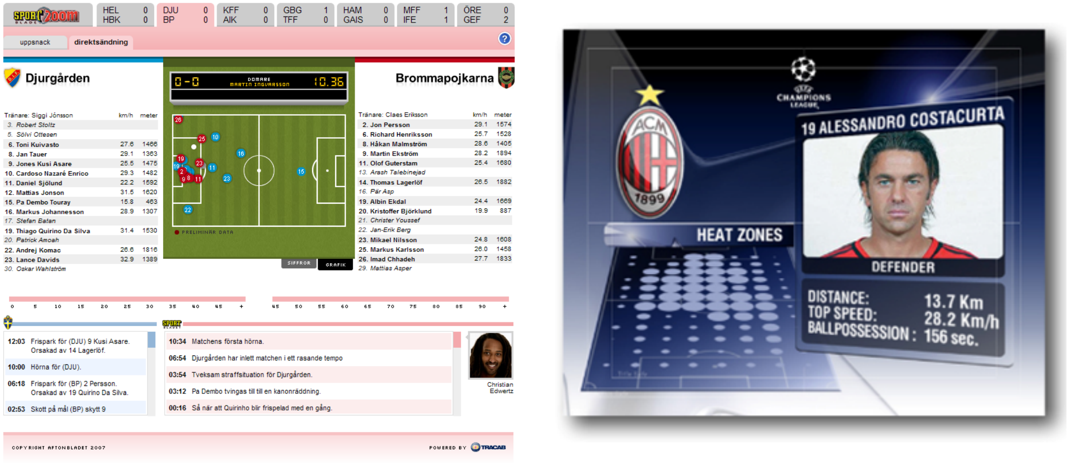
TRACAB

TRACAB's goal is to digitize sporting events so that they can be watched in real-time on digital interactive devices. The company specializes in tracking for team sports and has developed a system that locates all the moving objects (players, officials and the ball) on the playing field in real-time, 25 times per second, for football games.

This achievement is a world-first. In 2007 all matches in the Swedish premier league Allsvenskan were covered by TRACAB in collaboration with Aftonbladet, a leading newspaper whose website is the most visited in Sweden. Moreover, TRACAB is used in TV broadcasts in the UEFA Champions League, and will also be used in the 2008 UEFA European Football Championship (Euro 2008). Tracab plans for the immediate future are to expand to several other leagues.

The TRACAB system provides statistics such as the distance covered by each player and the speed of the ball at shots on goal. Additionally, on Aftonbladet's website [20], sports fans can also view the match in real-time as a *radar view* showing the position of each player as a dot on the pitch (see Figure 2). Needless to say, this view is much less appealing than either a TV broadcast or a 3D virtual game.

For this reason, TRACAB has already started exploring the possibility of 3D-graphics based visualizations of its data stream. This has been contracted to Agency9 [2], a Swedish graphics company. A snapshot from an early demonstration is shown in Figure 1. Agency9 will also be a key partner in this



HEL 0 HBK 0 | DJU 0 BP 0 | KFF 0 AIK 0 | GBG 1 TFF 0 | HAM 0 GAIS 0 | MFF 1 IFE 1 | ÖRE 0 GEF 2

uppsnack | direktsändning

Djurgården

Tränare: Siggi Jónsson	km/h	meter
3. Robert Stoltz		
5. Sölvi Ottesen		
6. Toni Kuivasto	27.6	1466
8. Jan Tauer	29.1	1363
9. Jones Kusi Asare	25.5	1475
10. Cardoso Nazaré Enrico	29.3	1482
11. Daniel Sjölund	22.2	1592
12. Mattias Jonson	31.5	1620
15. Pa Dembo Touray	15.8	463
16. Markus Johannesson	28.9	1307
17. Stefan Batan		
19. Thiago Quirino Da Silva	31.4	1530
20. Patrick Amoah		
22. Andrej Komac	26.6	1816
23. Lance Davids	32.9	1389
30. Oskar Wahlström		

0-0 DOMARE MARTIN INGVARSSON 10.36

PRELIMINÄR DATA

SIFFROR | GRAFIK

Brommapojkarna

Tränare: Claes Eriksson	km/h	meter
2. Jon Persson	29.1	1574
6. Richard Henriksson	25.7	1528
8. Håkan Malmström	28.6	1405
9. Martin Ekström	28.2	1894
11. Olof Guterstam	25.4	1680
13. Arash Talebinejad		
14. Thomas Lagerlöf	26.5	1882
16. Pär Asp		
19. Albin Ekdal	24.4	1669
20. Kristoffer Björklund	19.9	887
21. Christer Youssef		
22. Jan-Erik Berg		
23. Mikael Nilsson	24.8	1608
25. Markus Karlsson	26.0	1458
26. Imad Chhadeh	27.7	1833
29. Mattias Asper		

12:03 Frispark för (DJU) 9 Kusi Asare. Orsakad av 14 Lagerlöf.
10:00 Hörna för (DJU).
06:18 Frispark för (BP) 2 Persson. Orsakad av 19 Quirino Da Silva.
02:53 Skott på mål (BP) skytt 9

10:34 Matchens första hörna.
06:54 Djurgården har inlett matchen i ett rasande tempo
03:54 Tveksam straffsituation för Djurgården.
03:12 Pa Dembo tvingas till till en kanonräddning.
00:16 Så när att Quirinho blir frispelad med en gång.

Christian Edwertz

COPYRIGHT AFTONBLADET 2007

POWERED BY TRACAB

Figure 2: *The TRACAB Image Tracking System recovers the positions of moving objects in real-time for instance for publication on the internet (left: "Sportbladet ZOOM" on Aftonbladet's website) or on television (right: UEFA Champions League).*

Visualization research project. Their participation will be funded via Tracab's financial contribution to the project. Agency9's graphics engine, including support, will be made available to the KTH researchers. Agency9 will also be able to provide feedback concerning which parameters in the 3D model are key to obtaining pleasing visual reconstructions, and they will adapt their graphics engine based on requests from KTH, resulting from research findings. TRACAB anticipates that advances made in this Visualization project will immediately be put to commercial use.

The TRACAB *Image Tracking System*, is a camera-based system. It is powered by state-of-the-art image processing technology developed by SAAB for missile guiding systems in the military industry. TRACAB is a Swedish shareholders company, founded in 2003, which currently employs 22 staff. Its main owners are the three founding companies Hego [9], Egnus International Corp and SAAB, and additionally the venture capital advisory firm Provider Venture Partners.

CVAP-KTH

The research group at CVAP-KTH has also been involved in automatic tracking of multiple objects and actors with efforts focused on the automatic labeling of identities through the use of interaction graphs generated by multiple tracks [16, 12]. They also have a long experience in producing 3D visualizations from video, using multiple as well as single camera recordings [11, 15].

This 3D visualization is based on capturing the positions of specific body locations such as hands, elbows, shoulders over time using video captured from real sports events. If two views are available, this information can be used to calibrate the cameras and 3D reconstruction can be achieved [10]. If only a single view is available, 3D reconstruction can still be achieved by estimating specific body parameters such as length of specific limbs whose endpoints are captured. They have exploited the idea of a mapping between the image appearance of an actor and the specific body locations. By manually marking body locations in a set of images of actors, this creates examples of the mapping. Using these examples, the mapping of any image can be estimated using techniques of statistical learning. This idea of regarding tracking and visualization as a recognition and learning problem is generally considered as the most promising in this field and will be pursued throughout the project.

3 Research

The project's aim is to produce visually pleasing and faithful 3D reconstructions of the motion of the players in a football game using footage captured from video cameras viewing the game. This is an ambitious goal, but not an impossible one especially when one has leniency with respect to the definitions of *visually pleasing* and *faithful*.

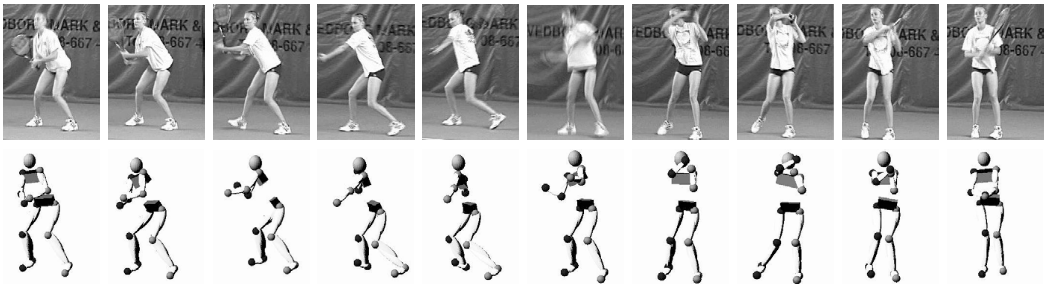


Figure 3: **Semi-automatic marker-less human motion capture results.** *A sample result from some of CVAP's work [11]. It is the reconstruction of one tennis fore-hand stroke.*

In fact, today there are already several commercial systems in use that produce such 3D visualizations of football players during a game. As already mentioned there is the TRACAB system which produces a graphical visualization of the game in which the players are represented as dots on the field at each frame. Their system runs in real-time and requires a minimal amount of manual intervention. The manual effort is mainly reserved to resolving errors in their tracking system. Another system is *Virtual Replay* used by the British Broadcasting Corporation (BBC) who display the results on their web-site [4]. It produces more sophisticated 3D visualizations of the game. Each player is represented by a generic 3D football player model which is placed at the correct location and orientation on the field. These models are also animated by the motion of the action they are classified as performing. Note the set of possible actions is quite limited and each action is represented by only one prototypical animation. So therefore all running actions look the same except for adjustments made for the speed and phase of the run. This system requires a relatively long time ($\sim$ 24 hours for the few short clips corresponding to the game's main highlights) to produce their visualizations as they have little or no automation.

With the description of these two working systems in mind, we present the goals for this project. Initially, we will try to emulate the Virtual Replay system but with a higher level of automation and therefore decreased production time. However, we would like to surpass Virtual Replay's achievement with respect to the accuracy of the visualized motion. We want to increase the number of action classes recognized and qualitatively reconstructed and then also faithfully reconstruct the variation due to person and situation specific factors within an action class. This latter task of accurately reconstructing different examples within one action class will require development of efficient but probably semi-automatic methods. Obviously, there is a trade-off between the quality of visualization possible and the amount of production time and manual effort one allows.

Figure 4 provides a graphical summary of the relative properties and merits of the current working systems discussed and those that this project will aim to produce. As is shown we intend to advance considerably the current frontier in the field of marker-less human motion capture of live sports events. Also of interest is to obtain by the end of the project a partial answer to the question - will the ideal situation of estimating accurately the 3D motion in real time with almost complete automation ever be possible? One interesting facet of the project will be the investigation of the level accuracy of reconstruction that can be achieved in real-time. Currently no-one else has access to the real-time tracking data similar to that produced by TRACAB's system so this project has the unique opportunity to assess and set standards on this front. Given this introduction, in the next subsections we highlight the key concepts and components upon which we intend to build our algorithms.

Approaches to 3D reconstruction

We want to construct accurate and detailed 3D visualizations of a football player's motion in a video clip. The input data will be provided by TRACAB, in the form of the player's trajectory along the ground plane (the football pitch) and a region of interest in each frame of the video corresponding to the player, see the left side of figure 5. It is from this data we must infer the player's 3D motion. Traditionally within computer vision there are two distinct approaches to performing this inference.

The first is based on explicitly modeling the physical process of imaging a 3D model chosen to represent

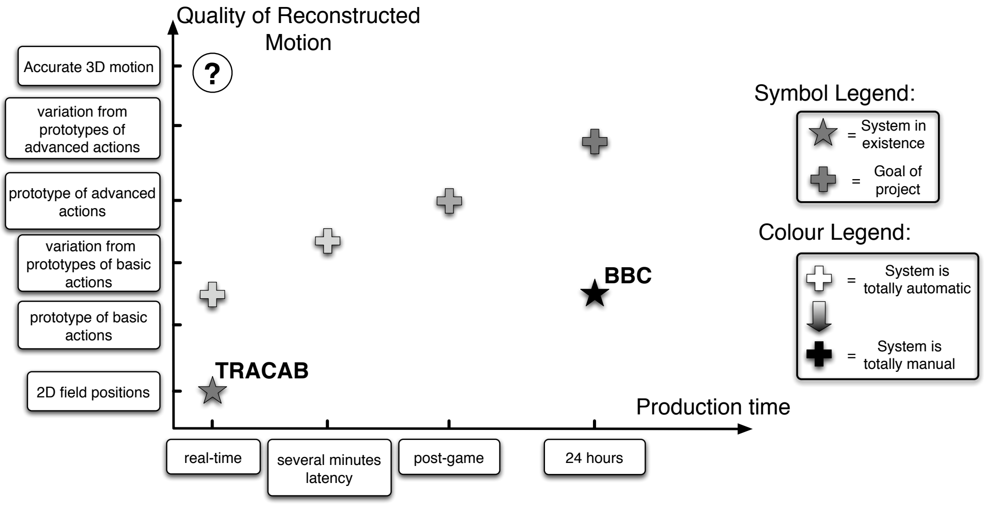


Figure 4: **Expected developments within this project relative to existing systems.** *This figure shows the characteristics - quality of the visualization, production time and level of automation of the algorithm - of systems currently in use and those we can realistically develop within this project while still advancing the frontier of marker-less human motion capture. The vertical axis displays the quality of the 3D visualization of the motion present in a game. Initially, the player's are represented by dots on a virtual field, next the motions are classified as a basic action and represented by prototypical 3D sequences of this action. At each subsequent level there is an attempt to recognize more complicated actions and to get the visualizations to more accurately reflect the specific details of the imaged motion. As can be seen from the graph, there is invariably a trade-off between the level of automation of the system and the expected degree of accuracy of the reconstruction.*

the player (e.g., skeleton or collection of cylindrical limbs) [5, 14, 7, 10]. The parameters of this 3D player model, corresponding to the player's 3D pose, are then estimated by maximizing the consistency between the appearance of the player in the video and the appearance of the projection of this 3D model onto the 2D image. One challenge, amongst many, is to find a representation of the player's appearance in the 2D image beyond the patch of pixel intensity values which can be both reliably extracted from these pixel values and easily predicted by the projection of the 3D model into the image. The representation should also be invariant to nuisance factors such as lighting conditions and clothing texture. Some commonly used representations are the silhouette edge of the player and the location of the 2D skeletal joints in the 2D image. In fact the latter case has been shown to be equivalent to having the 3D skeletal [10]. However, frequently the representation chosen involves an image processing technique that is not very robust to the presence of clutter or fast motions.

The other approach is based on recognition and learning [13, 8, 1, 19, 17]. It requires a library of training data where each entry in the library has the form - *input data*, in our case a 2D trajectory and a video clip, and its *corresponding 3D reconstruction*, figure 5. Given such a library and a novel video clip and trajectory, the 3D reconstruction of the novel clip is deemed to be similar to that of a sequence in the library with a similar 2D trajectory and appearance, figure 6. More formally a mapping from the space of input data to the space of 3D motions is learned from the labeled training pairs in the library. There are many popular regression methods for describing and learning these mappings such as linear, support vector, relevant vector and Gaussian process regression. It will be part of the research to investigate which regression method best suits our needs and data. Once again, though, there is the challenge of finding a representation of the player's appearance that can be reliably extracted and used to measure the similarity between different examples of input data. Clearly, the more descriptive the representation the more potential ability there will be to discriminate between similar 3D motions. However, the more descriptive the representation the less robust its extraction is likely to be.

In the learning approach it is not necessary to explicitly model how the appearance or trajectory data is generated by imaging the 3D model. This is an important point as it allows a much greater flexibility in the representation of the input data and it does not require this representation to be in

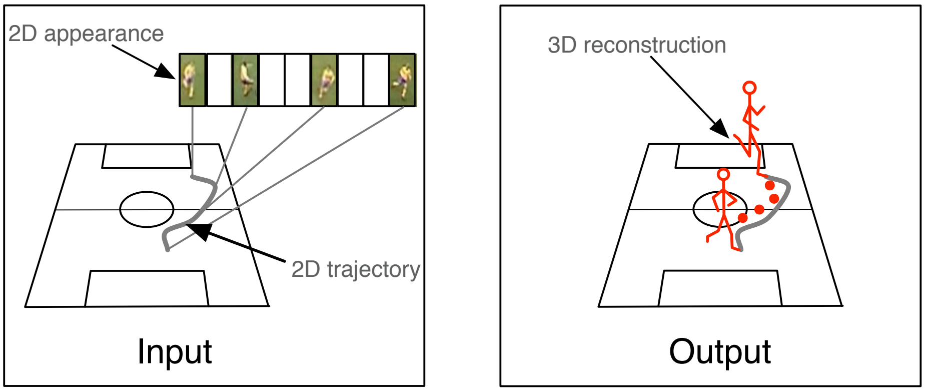


Figure 5: **An entry in the library of training data.***The input data is a 2D trajectory along the pitch and the image sequence corresponding to the appearance in the video of the player. While the output data is the 3D reconstruction of the motion of the player during this clip. This pair of input and output data is one entry in the library. From many such entries it is possible to* learn/represent *a mapping from the input space to the output space. Note that the input data may be represented and summarized at differing levels of detail depending on the required accuracy of the mapping.*

any ways explicitly related generatively to the 3D motion, but only that it is discriminative between the different 3D motions represented in the library. However, some qualities and properties of a model-based reconstruction (calculated from a detailed generative representation of the input data) are lost such as the ability to measure the accuracy of the reconstructed 3D pose and to reconstruct motions not included in the library of training examples.

As stated, extracting descriptive information about the appearance of a player, for instance localizing the skeletal 2D joints, from an image is not a robust process. However, the reliability of this process can be increased greatly when it is guided/constrained by prior knowledge. So for example locating a player's hip is much easier when it is constrained to be in a region, whose area is small compared to the image, as opposed to when no such region is specified.

These latter points highlight the reason for the approach we propose to adopt within this project. We intend to build a hierarchy of increasingly quantitatively accurate 3D reconstructions of the motion in a clip. At the lowest level of the hierarchy, corresponding to the most basic 3D reconstruction, the trajectory of the player on the ground plane will be mapped to a prototypical 3D reconstruction of a single action such as walking, running or shooting. Without any prior information, it is reasonable to assume that these trajectories will be the most reliable indication of the performed action and therefore the 3D reconstruction of the motion. This basic qualitative reconstruction will then act as a guide to the extraction of a more detailed representation of the player's appearance in the image at the next level of hierarchy. This more detailed representation will then be mapped to a more accurate 3D reconstruction. However, due to the generalization limitations of the learned mappings at some level of the hierarchy it will be necessary to apply a model based reconstruction method if a certain accuracy is required. At this stage though the measurement process, for instance skeletal joint localization in the image, may have a fighting chance of success due to the quality of the existing knowledge of the underlying 3D motion.

Acquiring training data

The discussion so far has assumed that we have access to a library of training examples from which we can learn our mappings. Crucially, this library will have be sufficiently large to represent the range of the motions we hope to reconstruct. Almost all forms of mappings learned from the labeled training data will perform poorly reconstructing motions without a close neighbour in the training set. Therefore, one of the major challenges of the project will be to acquire such a library.

The most advanced existing library of this kind has been generated by *Electronic Arts* (EA) [3] in a studio environment for the FIFA 08 football computer game which is, of course, not freely available.

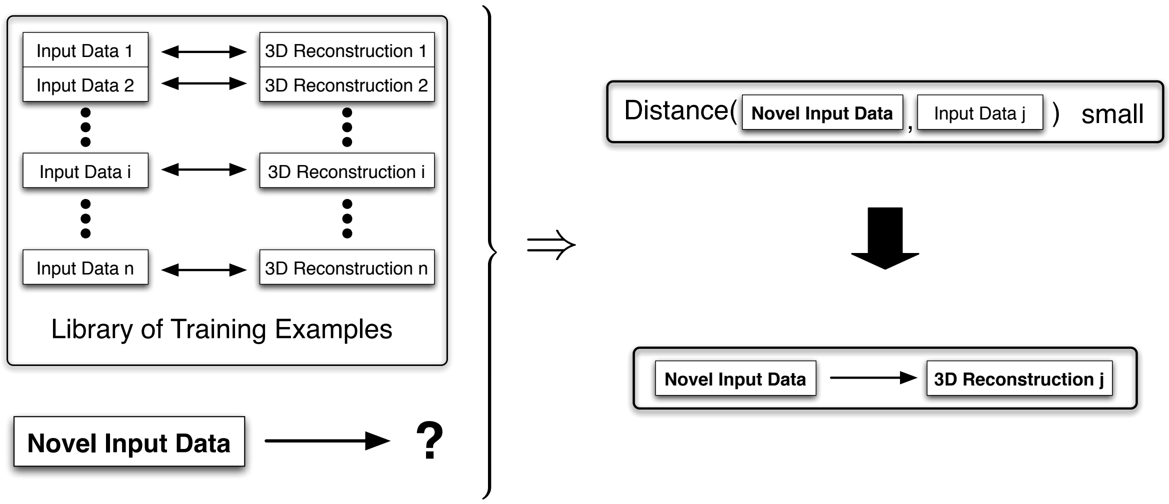


Figure 6: **Using a library of training data to reconstruct a novel video clip.***In a learning based approach to the problem, such as nearest neighbour, when one encounters a novel sequence it is compared to all the entries in the training set via the input data. The novel clip's 3D reconstruction is then deemed to be similar to the reconstruction of the motion in the library with the most similar 2D trajectory and appearance.*

But it is not obvious such a library, obtained from a marker-based human based capture system, is sufficient for our needs. The most obvious being, even given access to professional football players, one cannot recreate in a studio environment the range of motions that occur in a real competitive match. Therefore, any approach taken must involve augmentation of an existing training set, which itself has been obtained by boot-strapping from an initial relatively limited one. Augmentation will require the development of an efficient interactive scheme to accurately reconstruct previously unseen motions. This development, of an interface to perform accurate interactive marker-less motion capture, will be a key part of the project and a highly practical and useful addition to the field. To this end we plan to extend algorithms we have previously developed [10, 11]. Some of the specific factors that will be investigated when building this interface and growing the library of 3D motions are

- how to identify motion sequences that extend the library,
- how to maximize the degree of automation possible in the reconstruction given the current content of the library and the appearance of the motion,
- if a monocular video set-up be used to achieve sufficiently accurate reconstructions or if a stereo set-up is required.

For the last factor, if it is deemed that a stereo footage is necessary then such footage could be obtained by registering the TRACAB footage to the broadcast television video which will, in general, be of a much higher resolution.

Obtaining the initial library, is fairly straight-forward and reduces down to making the following choice between:

- Obtain marker-based human motion capture data of basic actions such as walking and running.
- Construct this data manually from example video sequences.

The first choice results in accurate 3D motions but it may be hard to obtain sufficient examples of football player motions while on the other hand it is not clear if sufficiently accurate reconstructions can be obtained from manual reconstructions obtained from video footage. This will be one of the first issues to be considered in the project.

4 Demonstrators

Very concrete and achievable subgoals can be defined to bridge the gap between the current state-of-the-art in automatic visual human motion capture and our ultimate goal of constructing accurate 3D visualizations of players' motions in a game. These subgoals are linked to the hierarchy of reconstructions we intend to achieve. The following explicitly describes these subgoals and the demonstrators that will show their successful completion.

Demonstrator I - The first part of the project will target the task of building a representative but approximate 3D reconstruction of the individual players' actions throughout the game or in other words the first level of our hierarchy. TRACAB's tracking system and KTH/CVAP's identity labeling algorithms will provide the trajectories and video of the players over the 90 minutes of a football. Then we will learn a mapping from these trajectories to a basic action unit (i.e., walking, running, jumping, kicking) and use a stored 3D sequence of this action to represent it in our reconstruction. This will necessitate the construction of the initial library of 3D motions and an exploration of the benefits of building it from motion capture data or manually reconstructions from video data. The demonstrator for this part will be:

> $\mathbf{D}_1$ - *Approximate/Qualitative* 3D motion capture of a football game for all players.

Demonstrator II - Next the project will focus on upgrading the approximate/qualitative reconstruction to a more *accurate* one with respect to the basic action units. This will involve a more exact recovery of a player's skeletal joints in 3D for each frame. This requires learning a mapping as previously described and this in turn requires sufficient training data. As such data is not immediately available, we will have to develop efficient interactive algorithms to perform accurate reconstruction and allow the enlargement of our library of 3D reconstructions for different actions. The best form of the *pure learned* and *modeled* mappings will also have to be explored. Ultimately the goal will be to calculate automatically, given sufficient training data, an

> $\mathbf{D}_{2a}$ - Accurate 3D visualization of sequences (∼30 seconds duration) of individual players performing common actions.

However, a necessary by-product for achieving $\mathbf{D}_{2a}$ will be the development of interactive software for the efficient reconstruction of unseen motions. Therefore another associated demonstrator will be

> $\mathbf{D}_{2b}$ - An efficient interactive algorithm for performing 3D reconstruction of motions not represented by a given library of stored motions.

Demonstrator III - Having obtained the accurate reconstruction for individuals on short sequences (∼30 seconds) we will use the methods and the library constructed to produce the next demonstrator. This demonstrator will be the accurate 3D reconstruction of a player in the periods during the game when he is performing a basic action unit.

> $\mathbf{D}_3$ - Accurate 3D visualization of the most common individual player actions for a whole football game.

Demonstrator IV - Having devoted our efforts to mainly reconstructing players when they are isolated we will concentrate on adapting the methods to more complicated multi-player actions such as tackling and marking situations. This will involve acquiring more training data and performing occlusion reasoning. Ideally we will produce:

> $\mathbf{D}_4$ - Accurate 3D visualization of sequences (∼10 seconds long) of multi-player actions.

Demonstrator V - Finally, efforts will be devoted to automatically producing an accurate 3D reconstruction of a whole game for each player over as many situations as possible. In practice this will involve implementing a full system. This can be summarized as

> $\mathbf{D}_5$ - Accurate 3D visualization of the multi-player actions occurring in a whole game.

5 Project time line

There is a differing degree of difficultly and risk involved in the completion of each of the demonstrators. The first three demonstrators will build on prior work in the fields of tracking, action recognition and human motion capture. Here the main challenge will be the scale of the problem (a whole football game, 22 players, 90 minutes) we are attempting. However, we anticipate success in these demonstrators. The last two tasks, however, are riskier, especially $\mathbf{D}_5$, requiring significant progress in the current state-of-the-art. Thus it is not guaranteed that they will be successfully completed within the time frame of the project. Table 1 gives an overview of the time expected to be devoted to each task.

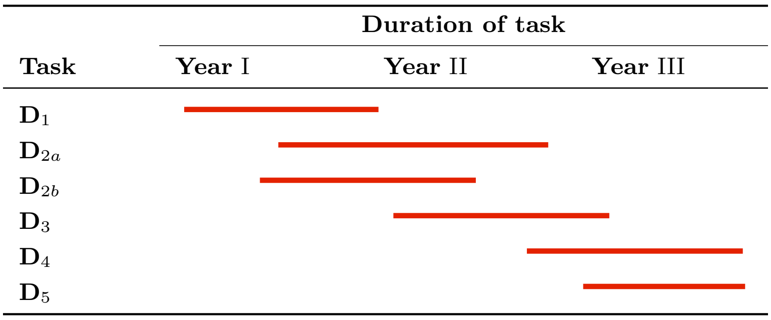


Table 1: **Time line for completion of the project demonstrators.**

6 Intellectual property

The exact form of the intellectual property rights (IPR) and its division has yet to be formalized. However, these will be negotiated between the partners along the following principles:

- TRACAB will have exclusive right to the commercialization of the results of the project,
- the KTH partners will maintain the right to use the project's results in future research,
- the KTH partners will be refunded through license fees for any commercial use of software produced within the project the levels of which will be negotiated,
- funding of IPR protection and relative ownership of patents will negotiated.

Additional Collaborations

The project will cooperate with the recently initiated VIC (Visualization-Interaction-Communication) network for visualization in research and education all over KTH, see attachment.

References

[1] A. Agarwal and B. Triggs. 3d human pose from silhouettes by relevance vector regression. In *Conference on Computer Vision and Pattern Recognition*, Washington, June 2004.

[2] Agency9. Homepage. `http://www.agency9.com`.

[3] Electrontic Arts. Homepage. `http://www.ea.com`.

[4] BBC. Homepage. `http://www.bbc.co.uk`.

[5] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Conference on Computer Vision and Pattern Recognition*, 1998.

[6] CVAP. Homepage. `http://www.csc.kth.se/cvap`.

[7] J. Deutscher, A. Blake, and I. Reid. Motion capture by annealed particle filtering. *Conference on Computer Vision and Pattern Recognition*, 2000.

[8] Alexei A. Efros, Alexander C. Berg, Greg Mori, and Jitendra Malik. Recognizing action at a distance. In *ICCV*, pages 726–733, Nice, France, 2003.

[9] Hego. Homepage. `http://www.hegogroup.com`.

[10] D. Liebowitz and S. Carlsson. Uncalibrated motion capture exploiting articulated structure constraints. *International Journal of Computer Vision*, 51(3):171–187, 2003.

[11] G. Loy, M. Eriksson, J. Sullivan, and S. Carlsson. Monocular 3d reconstruction of human motion in long action sequences. In *European Conference on Computer Vision*, 2004.

[12] P. Nillius, J. Sullivan, and S. Carlsson. Multi-target tracking linking identities using bayesian network inference. In *CVPR*, 2006.

[13] G Shakhnarovich, P. Viola, and T. Darrell. Fast Pose Estimation with Parameter Sensitive Hashing . In *ICCV*, Nice, France, October 2003.

[14] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling. *International Journal of Robotics Research*, 22(6):371–391, June 2003. Special issue on Visual Analysis of Human Movement.

[15] J. Sullivan and S. Carlsson. Recognizing and tracking human action. In *European Conference on Computer Vision*, 2002.

[16] J. Sullivan and S. Carlsson. Tracking and labelling of interacting multiple targets. In *European Conference on Computer Vision*, 2006.

[17] A. Thayananthan, R. Navaratnam, B. Stenger, P.H.S. Torr, and R. Cipolla. Multivariate relevance vector machines for tracking. In *European Conference on Computer Vision*, 2006.

[18] TRACAB. Homepage. `http://www.tracab.com`.

[19] R. Urtasun, D. J. Fleet, and P. Fua. 3d people tracking with gaussian process dynamical models. In *Conference on Computer Vision and Pattern Recognition*, 2006.

[20] Sportbladet ZOOM. Homepage. `http://allsvenskandirekt.sportbladet.se/`.