Capturing the Motion of Ski Jumpers using Multiple Stationary Cameras

Atle Nes

Norwegian University of Science and Technology
Department of Computer and Information Science
atlen@idi.ntnu.no

Trondheim University College
Faculty of Informatics and e-Learning
atle.nes@hist.no

This article presents a general description of some challenges encountered when building a computer system being used for precise motion capture of human beings moving in a large-scale environment. More specifically this article focuses on capturing and studying the three-dimensional motion of ski jumpers in Granåsen ski jumping hill in Trondheim. This is part of an on-going research project involving the Norwegian ski jumping elite. Video images are captured simultaneously from multiple synchronized digital video cameras placed strategically around a ski jumping arena. This allows estimating point positions in 3D space if the same point is identified in at least two camera views simultaneously. Ski jumpers have reflective markers serving as robust feature points strategically placed on their jumping suit and skies. Moving feature points are followed in time and space, using image processing and photogrammetric techniques, creating feature tracks in 3D space. The captured 3D data from the hill is then used to control a virtual model of a ski jumper, which is visualized and analyzed realistically inside a detailed model of the ski jumping arena. Motion analysis and visualization will be used to give feedback to the ski jumpers that can help them move towards optimal movement patterns and thereby increase their jumping length. Some guidelines and solutions for improving the accuracy of the 3D measurements will be described.

1. Introduction

One particular challenge with ski jumping, which makes it a bit different from many other sports, is that the number of ski jumps, and therefore also the amount of real training possible, is limited to only a few jumps per hour per athlete. The main motivation behind this project is that using a specialized computer system allows studying the motion in greatest detail, present it in an understandable manner and thereby give important and reliable feedback to the ski jumper and trainer. Computer visualizations modelling the ski jumper in a virtual environment and results from the motion analysis can be used as assistance complementing the regular outdoor training. Another common training situation for ski jumpers consists of indoor training, allowing for many more successive jumps onto a mattress. It is a fact that ski jumpers are usually more effective in an indoor environment than in the real hill. Virtual ski jump simulations may reveal the important clues necessary helping the individual athletes search for a better outdoor ski jump as well as transferring a good indoor jump to the real hill.

Previous work

Previous research dealing with motion capture of ski jumpers using video has had its main focus on 2D movements and ski jump take-off exclusively [1] [2] [3]. Creating a single flight track in 3D has been done recently using a GPS helmet [4], and is also available in a system developed by Cairos Technologies [5] which uses a computer chip attached to the ski jumping suit together with radio antennas in the hill. Related biomechanical research is performed in laboratories, and involves wind tunnel testing and 3D motion capture. This article focuses on 3D motion capture of a ski jumper covering the take-off phase as well as the first seconds in the air when the ski jumper stretches out to flight pose. The extension from 2D to 3D will allow studying ski jumpers twist and rotation during and after take-off. In spite of the limited research on the specific task of ski jumping, much more research has on the other hand been done in the more general field of human motion analysis. An extensive summary of status and advances showing some concepts, procedures and techniques related to human motion analysis is given in [6].

People involved

My supervisors on this project are professor Torbjørn Skramstad at NTNU, associate professor Bjørn Sæther at NTNU and associate professor Jan H. Nilsen at Trondheim University College. Co-operation has been established with several institutions and companies in the Trondheim region. Our most important co-operation partner, which possesses valuable knowledge about human movement and motion analysis, is Human Movement Science Program at NTNU, headed by professor Beatrix Vereijken. The main contact there is Steinar Bråten, which has regular contact with the Norwegian national ski jumping team through his role as an advisor for Olympiatoppen. Other important co-operation partners worth mentioning are Statoil, which has rich experience with visualization of oil and gas reservoirs and installations, SINTEF, which is using similar techniques for studying ocean currents in the project HYDRIV [7] [8], and Institute of Geomatics at NTNU, which has refined the accuracy of aerial land photogrammetry for at least a century. Worth mentioning are also several closely related student projects at Faculty of Informatics and e-Learning, Trondheim University College.

Overview

Section 2 describes how to set up for image acquisition in a ski jump arena using multiple video cameras. Section 3 describes motion capture from tracking of robust feature points. Section 4 describes how to obtain accurate 3D point positions from 2D images using photogrammetric techniques. Section 5 describes visualizing the captured dynamic 3D data in a static virtual 3D model. Section 6 describes techniques for analyzing and post processing the obtained motion data. Section 7 describes some related applications using the same type of camera equipment and similar techniques as described in this article. Section 8 gives a summary and some conclusions.

2. Image Acquisition

Three digital video cameras are used to record three parallel video streams, each of them containing their own characteristic view of a ski jumper. High accuracy on the measurements is achieved through thoroughly considered placement of the cameras in the ski jump arena, correct lenses according to the camera to jumper distance, good synchronization between the cameras and highest possible camera resolution at highest possible frame rate.

Camera equipment

The camera system used consists of three identical Marlin F-080B digital CCD-based video cameras from Allied Vision Technologies [9] [10], extra long cables, specially designed software and a powerful coordinating computer. The cameras deliver synchronous 8-bit grayscale images with a maximum resolution of 1024x768 at 15fps or 640x480 at 30fps. Another limitation of the cameras lies in the no-zoom C-mount lenses which must have the correct focal length according to their distance to the area of interest containing the ski jumper. The current lenses mounted on the cameras have a focal length of 6.5 mm which will give a visible area of 1,15 m x 0,95 m at 1 m or more interestingly 31,7 m x 26,1 m at 30 m distance in the ski jumping hill. This defines the limitations in space for an object to be observed and can be estimated easily by knowing the focal length and chip size, usually defined by the manufacturer.

$$\text{visible area} = \frac{\text{chip size} \cdot \text{distance to object}}{\text{focal length}}$$

Cables and communication

Each camera is connected to the computer through a separate IEEE-1394 FireWire cable, which delivers power to the camera as well as transmitting video and control data. FireWire is a well-established standard for high-speed two-ways serial data communication, popularly used for video, which eliminates the need for dedicated frame grabber cards. The computer is instead equipped with two identical FireWire PCI cards with two OHCI interfaces on each card. Each OHCI interface controls a physical bus (IEEE-1394a) with a theoretical transmission speed of around 400 Mbit/s (50 Mbyte/s) and a practical transmission speed of around 256 Mbit/s (32 Mbyte/s). Since each camera can produce video data at a rate up to 16 MByte/s it is advisable to use three separate FireWire buses.

In a ski jump arena one needs much longer cables than the limited 4.5 m as defined in the FireWire standard. To sustain high transmission speed over longer distances one solution is to use optical fiber as carrier for the signals and data. A set of two Opticis [11] M4-200 optical repeater boxes, in each end of the optical fiber, equipped with its own 12 V power supply, allows increasing the total distance between camera and computer up to 500 m (Figure 1). Full duplex data rate at 400 Mbit/s is maintained with this solution. Regular FireWire cables are used in each end to connect the cameras and the computer to the optical repeater boxes.

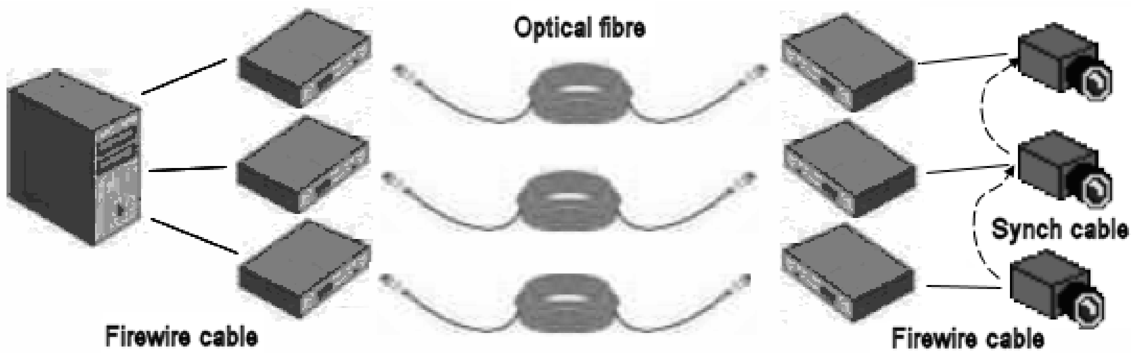

Figure 1. Multiple camera configuration

The cameras use the IIDC v1.3 protocol, also called the DCAM standard, for communication [12]. DCAM is a standard for controlling and receiving data from scientific video cameras. It defines a range of control signals that can be sent to the cameras and the format of the uncompressed video transmitted back from the cameras based on the control signals received. This differs from cheap web cameras which compress their video before transmitting it with a low bit rate at the cost of lost video quality. Compression must not be allowed in the system before reaching the image processing stages.

Computer bottlenecks

To deal with the data streams of 16 MByte/s from 3 cameras simultaneously regular hard disk drives are too slow to store that much data directly. A standard IDE hard disk is usually only able to handle one of these cameras at a time. One way to solve this is to record the ski jump, which is limited to some seconds in time, to memory and when finished capturing transfer all the video data to hard disk afterwards. A better solution which removes the speed problem entirely is to improve the hard disk performance. This is absolutely necessary if one needs to capture time sequences resulting in more data than there is memory space available. Hard disk write speed can be improved through RAID-0 solutions which use several hard disks tackling the load in parallel and/or more expensive hard disks with increased rotational speed of usually 10.000 or 15.000 rpm.

Synchronization

To be certain that the three video images are being captured at the exact same point in time a specially designed synchronization cable is connected between the three video cameras. Software synchronization can be achieved by sending out a synchronization packet, but this is only an alternative if all cameras are connected to the same physical bus. Without any form of synchronization in there will be a maximum time difference between corresponding image frames of 33 ms at 30 fps, and 67 ms at 15 fps. This is because the software is set to always fetch the last completely captured video frame. A frame that is more out of synch at any of the three cameras is simply dropped and the next frame is fetched instead. Connecting a custom-made synchronization cable supplying an external trigger pulse to each of the cameras simultaneously will improve the maximum time difference between corresponding image frames down to microseconds. This is necessary for getting accurate 3D calculations of fast moving objects. The trigger signal can originate either from one of the cameras, from a triggering box or from a generating port at the computer.

Placement

The placement of the cameras is also an important factor for the performance of the system. First of all the cameras must be placed such that they capture the desired motion of the ski jumper. This means capturing the ski jumpers motion about 10 m before and 30 m after the ski jump edge in all cameras. Different lenses are thus needed if the cameras have different distances to the volume of interest. Best triangulation performance is achieved when the camera line of sight is exactly 90 degrees perpendicular to the other cameras line of sight [13]. Best automatic feature point recognition performance is achieved when the angles between the camera views are as little as possible, which usually implies that the identical feature points look similar at all cameras. Obviously these facts are conflicting, and one has to find a compromise where the feature points can be reasonable easily detected, at least in two cameras simultaneously, and not having too bad triangulation angles. In any case one should always avoid placing the cameras on a straight line, which is one well-known troublesome camera configuration [14]. In the chosen configuration two cameras were placed close in front and on either side of the ski jump, giving nice details of the ski jumper (Figure 2). A third camera was placed further away, giving out an accurate flight path along the axis of the hill.

Figure 2. Three views of a ski jumper leaving the edge

3. Motion capture

A crucial part of this project is to be able to track feature points reliably over sequences of video frames. It is desirable to have automatic detection of feature points because marking them manually will be an extremely large and impractical task. Automatic detection is absolutely necessary if one wants a near real-time system giving feedback right after a jump. The robust feature points can be signalled human body markers, which are easy to detect, or naturally robust features, which are much more difficult to detect. Current experiments have been performed using reflective bands around legs, arms and on the helmet (Figure 2). More extremities have to be included in further research to give good connection between reality and the movable ski jumper model. Naturally robust feature points can be found by studying edges, corners, gradients etc. Harris corner detector is one commonly used and classical algorithm for finding such points [15]. Corresponding features must not only be robustly tracked over time, but also in the different camera views. Standard algorithms for feature detection and feature tracking will be applied and specialized if necessary. If a feature point is found in two cameras only, it is possible to estimate the position in the third using photogrammetric techniques described in the next chapter. This can allow time glitches in the feature tracking. In addition it is possible to use knowledge about the structure of a human body to identify the individual features and relative positions.

Blur

Blur effects can become a problem when capturing fast moving objects, like ski jumpers. This can be solved by decreasing the exposure time. The cameras used have an adjustable internal exposure time ranging from 20 μs up to 67 s, with a default setting of 40 ms. Ski jumpers moving at about 80 km/h will move 90 cm during default exposure time, causing severe blur in the image. Setting the exposure time to around 1 ms seems more appropriate, which implies a maximum movement of 2 cm during exposure. Reducing the exposure time reduces the time incoming light photons hit the CCD sensor. If the image becomes too dark it is better to adjust the lens aperture to let in more light if possible.

4. Photogrammetry

Before any 3D measurements can be performed in a scene one needs to define the geometry of the space which is being observed. Defining the geometry establishes a mathematical relationship between the image coordinates obtained from the camera CCDs usually defined by pixel positions and object coordinates in the real world usually defined in metric units. A program package for industrial photogrammetry called Straal developed by Institute of Geomatics at NTNU is used for this task [16]. Straal performs triangulation (intersection) and camera calibration (resection) simultaneously in a process called bundle-adjustment, which was first described by Brown [17]. The optimal parameters are found iteratively reusing previously calculated parameters and trying to minimize the error in a least square sense.

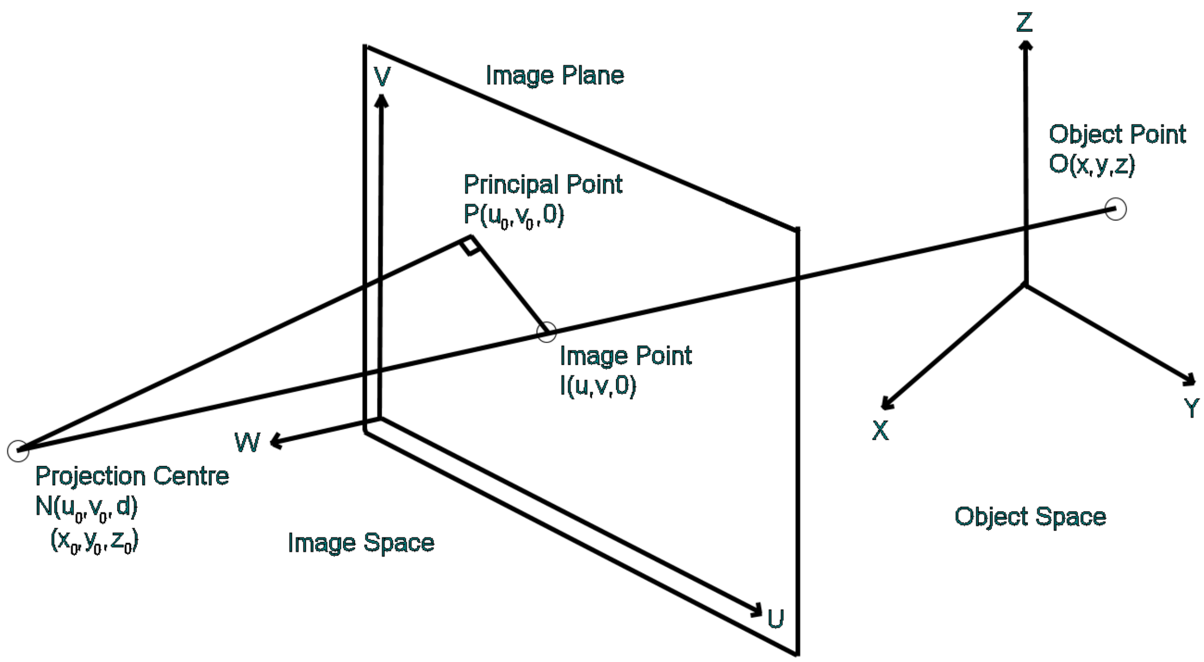

Figure 3. Projection of an object point to an image point

Direct Linear Transformation (DLT)

One method for obtaining good starting values used initially in the bundle-adjustment is the DLT method (Direct Linear Transformation), originally reported in [18] and today commonly used in industrial photogrammetry. The DLT method uses a set of control points whose object space coordinates are already known. The control points are normally fixed to a calibration frame, which can be a robust 3D metal frame with accurately measured and easily detectable spots on it. An alternative method of getting control points spread evenly in a 3D volume can be accomplished by rotating and capturing multiple images of a simpler calibration object. The flexibility of the DLT-based calibration often depends on how easy it is to handle the calibration frame. In Granåsen ski jumping hill about 60-70 known coordinates were surveyed with a mutual relative accuracy of 5 mm and absolute accuracy of 3 cm in relation to county coordinates. Some of these control points are naturally signalled like corners on tribunes and ski jump. Other control points are identified by iron bolts made visible to the cameras by placing a round ball on top. A small height representing the radius of a ball has to be subtracted in the calculations to obtain the correct coordinates. Camera calibration is done on-site with predefined coordinates in the terrain as a kind of calibration frame.

Capturing an image using a camera is equivalent to projecting an object point O in the object space to an image point I in the image plane (Figure 3). Two reference frames are defined; the object space reference frame defining the object point coordinates (x, y, z) and the image plane reference frame defining the image point coordinates (u, v). The camera points in the direction of the object to be captured, and the projection centre N defines the position of the camera lens. The line drawn from the projection centre N to the image plane, parallel to axis W and perpendicular to the image plane is called the principal axis, and the principal point P is the intersection of this axis with the image plane. The principal distance d, also called the camera constant or focal length, is the distance between the principal point P and the projection centre N. The points I, N and O will lie on a straight line and are called co-linear. This co-linearity condition is the basis of the DLT method giving rise to the standard 3D DLT equations [18]:

$$u = \frac{L_1 x + L_2 y + L_3 z + L_4}{L_9 x + L_{10} y + L_{11} z + 1}$$
$$v = \frac{L_5 x + L_6 y + L_7 z + L_8}{L_9 x + L_{10} y + L_{11} z + 1}$$

The DLT method is applied when both the parameters of the interior orientation (image plane coordinates for the principal point and the principal distance) and the parameters of exterior orientation (object space coordinates for the projection centre and the elements of the rotation matrix R) are unavailable. There are two different ways to use the equations described above; camera calibration and point reconstruction.

Camera calibration

Camera calibration will establish the relationship between the 3D object coordinates and 2D image coordinates. For each calibration point we obtain two linear equations, with 11 unknown DLT parameters. To solve out the DLT parameters we therefore need to have at least six calibration points consisting of object coordinates (x, y, z) and the corresponding image coordinates (u, v). This gives a total of minimum 12 equations. Having more than six calibration points generally increases the calibration accuracy, resulting in an over-determined set of equations. The calibration points must not be coplanar, which means that the calibration points must form some sort of volume. Camera calibration is performed separately on each camera. Approximate camera specific parameters can be estimated in advance easing the calibration process.

Point reconstruction

Point reconstruction uses multiple calibrated cameras and their DLT parameters to calculate the 3D object coordinates from the 2D image coordinates. The 11 DLT parameters are now assumed to be known. For each camera we have a different set of DLT parameters describing the relationship between the 3D object coordinates and the 2D image coordinates of that camera. To solve out the object coordinates (x, y, z) we need to have at least two cameras with the corresponding image coordinates (u, v) for that object point. Having more than two cameras generally increases the reconstruction accuracy.

Optical errors

To increase the precision even further one has to consider the imperfectness of lenses which give rise to non-linear relationships between image points and object points. The image coordinates (u, v) on the left hand side of the 3D DLT equations are then added optical errors (Δu, Δv):

$$\Delta u=\xi(L_{12}r^2+L_{13}r^4+L_{14}r^6)+L_{15}(r^2+2\xi^2)+L_{16}\xi\eta$$
$$\Delta v=\eta(L_{12}r^2+L_{13}r^4+L_{14}r^6)+L_{15}\eta\xi+L_{16}(r^2+2\eta^2)$$

where

$$[\xi,\eta]=[u-u_0,v-v_0]$$
$$r^2=\xi^2+\eta^2$$

As you can see there are five new additional DLT parameters. L_1 to L_{11} are the original DLT parameters, L_{12}, L_{13} and L_{14} radial distortion terms, and L_{15} and L_{16} tangential distortion terms [19]. The equations are no longer easy and linear and a solution has to be iterated. The 11-parameter DLT is used to get an initial coordinate estimate and 16-parameter DLT used to refine the coordinate estimates through successive runs. Experiments have shown that the first order radial distortion term L_{12} is the most important term when considering nonlinear lens artefacts [14]. Calibration performance can be improved even more if one uses intelligent techniques to remove the worst calibration points from the calibration process.

5. Visualization

When motion capture and photogrammetry has been performed using the captured video data one ends up with a very compact description including only the motion of a specific set of points in space and time (3D + t). This limited description represents a drastic reduction, and a small paradox with respect to the problems encountered earlier when trying to store the huge amount of incoming video data. The robust feature points must be wisely chosen so that they give just enough information to control a dynamic model of a ski jumper. Body markers are placed close to body joints on the suit and on the tip of the skies. An investigation has to be made into what are the best feature point positions on the jumper and how many points are really needed.

Connecting the captured feature points back onto a 3D model of the ski jumper, is similar to what is frequently done in animated movies and computer game industry. The equipment that is being used in these professional studios is usually based on infrared light which sent out, reflected of reflective markers and captured by infrared cameras. The frequencies used at such facilities are much higher compared to the scientific video cameras described here. One such system, used indoors at Human Movement Science Program in Trondheim is the ProReflex system from Qualisys [20]. Factors like expensiveness of purchasing, and difficulties setting up such a system outdoors in the hill, has motivated to try to use regular video cameras with lower frame rate instead. ProReflex is considered to be stationary installed equipment used for indoor training exclusively, but has also been tested outdoors at least once [21].

Student projects at the faculty have created an accurate CAD model of the entire Granåsen ski jump arena (Figure 4), and combined it with a dynamic model of a virtual ski jumper. We are able to use input data stored in a regular text file, or the professionally used C3D format, to control the ski jumpers motion in the ski jump arena model. One has to make sure is that all the movements made by the virtual ski jumper are allowable, which means that the jumper is not allowed to spin his leg through the other leg or twist its head five times.

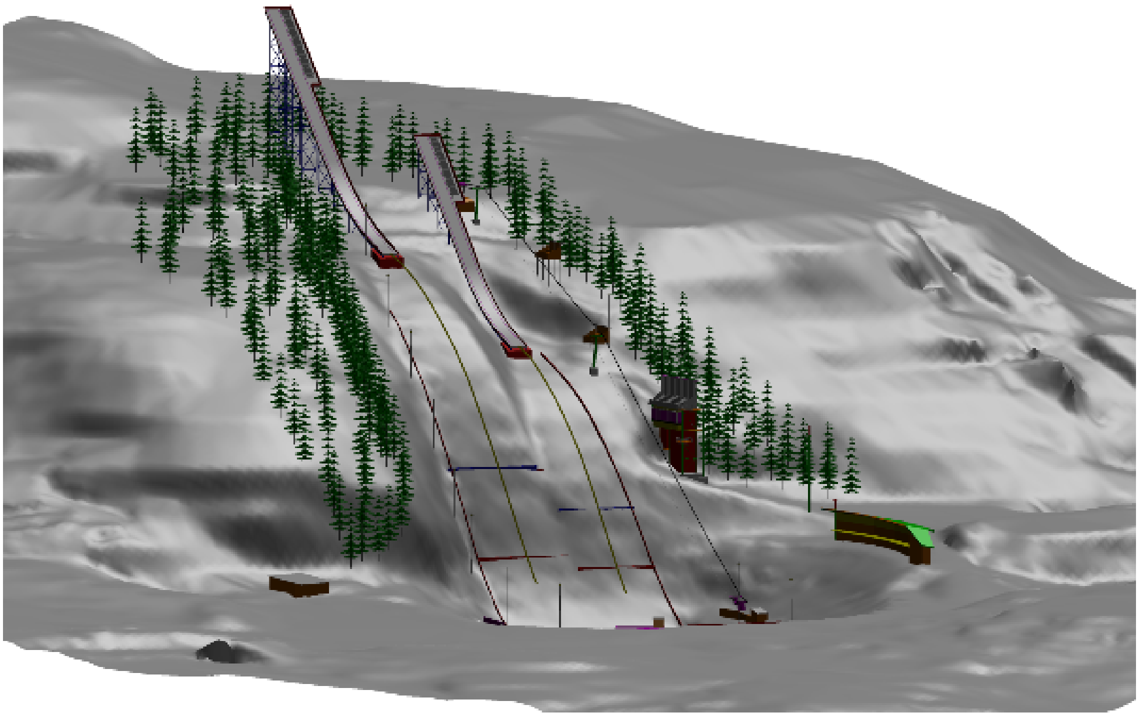

Figure 4. Rendered image of Granåsen ski jump arena CAD model

It remains to be investigated how accurate the computer model is compared to the real arena. The model may have to be fine-adjusted to fit the measurements of the terrain and also the captured video data of the ski jumper in the hill. Visualization experiments have been conducted using CAVE environments, belonging to our co-operation partners Statoil and Fachhochschule Bonn-Rhein-Sieg. This gives a much better view than just viewing the model on a regular computer screen.

6. Motion analysis

Further analysis of the motion must be done in close co-operation with the expertise at Human Movement Science Program. It is desirable to extract those movements which have greatest influence on the result. One important question is to study the motion using statistical tools or introduce prior knowledge about the movements. It should be interesting to find out what the ski jumping elite does different from a more average ski jumper.

7. Related applications

The approach described in this article is not exclusively limited to studying ski jumpers. Originally this project focused on studying infants and developing tools for early detection of brain damages like cerebral palsy by looking for irregular spontaneous movement patterns. Other medical applications are computer aided diagnosis of adult movements like gait patterns for determination of the cause of a certain problem. When it comes to athletes the system can be used to diagnose sport injuries or helping the athletes find more optimal movement patterns like in this project. Another topic that has become more interesting over the years is using video cameras for crowd surveillance and identification of possible strange behaviour in a shopping mall or airport. The idea is that terrorists and aggressors behave in a different manner compared to the people shopping or going for a travel. In such applications one can not place markers on people, and naturally robust feature detection must be employed.

8. Summary and conclusions

This article has given an overview of a system that can be used to capture, visualize and analyze ski jumpers in a ski jump hill. The system itself is not limited to studying ski jumpers, but can have other applications in topics like medical, sports and security. Since this is not a finished project, but rather a description of future work, it remains to see how well such a system can perform, and if it can help the ski jumpers improve their skills.

References

[1] M. Virmavirta, J. Isolehto, P. Komi, G. P. Bruggemann, E. Muller, H. Schwameder, 2005, ”Characteristics of the early flight phase in the olympic ski jumping competition”, Journal of Biomechanics 38(11):2157-2163

[2] B. Schmölzer and W. Müller, 2005, “Individual flight styles in ski jumping: results obtained during olympic games competitions”, Journal of Biomechanics 38(8):1055-1065

[3] L. P. Remizov, 1984, “Biomecanics of optimal flight in ski-jumping”, Journal of Biomecanics 17(3):167-171

[4] T. Blumenbach, 2005, “GPS-Anwendungen in der sportwissenschaft entwicklung eines messverfahrens für das skispringen”, Dr. ing thesis, Technischen Universität Dresden, Germany, ISSN 0065-5325, ISBN 3-7696-5030-1

[5] Cairos Technologies : http://www.cairos.com/

[6] L. Wang, W. Hu, and T. Tan, 2003, ”Recent developments in human motion analysis”, Pattern Recognition 36(3):585-601

[7] S. M. Løvås et al., 2003, “HYDRIV task 4: modernization and improvement of 3D particle tracking using three synchronous cameras for near real-time analysis”, Report, SINTEF Fisheries and Aquaculture, STF80 A038059, ISBN 82-14-01471-9

[8] J. H. Nilsen, 1994, “An experimental study of internal tidal amphidromes in Vestfjorden”, Dr. ing thesis, Norwegian Institute of Technology, Trondheim, Norway, ISSN 0802-3271, ISBN 82-7119-640-5

[9] Allied Vision Technologies : http://www.alliedvisiontec.com/
[10] Allied Vision Technologies, 2005, "AVT Marlin technical manual v1.4"
[11] Opticis : http://www.opticis.com/
[12] 1394 Trade Association, 2000, "IIDC 1394-based digital camera specification version 1.30"
[13] R. I. Hartley, A. Zisserman, 2004, "Multiple view geometry in computer vision", 2nd edition, Cambridge University Press, ISBN 0-521-54051-8
[14] J. C. McGlone, 2004, "Manual of photogrammetry", 5th edition, American Society for Photogrammetry and Remote Sensing, ISBN 1-57083-071-1
[15] C. J. Harris and M. Stephens, 1988, "A combined corner and edge detector", Proceedings 4th Alvey Vision Conference, Manchester, 147-151.
[16] I. Hådem, 1989, "A program package for industrial photogrammetry", Kart og Plan 1:24-25
[17] D. C. Brown, 1966, "Decentering distortion of lenses", Photogrammetric Engineering 32(3):444-462
[18] Y. I. Abdel-Aziz and H. M. Karara, 1971, "Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry", Proceedings of ASP Symposium on Close Range Photogrammetry, Urbana, Illinois, USA, 1-18
[19] J.S. Walton, 1981, "Close-range cine-photogrammetry: a generalized technique for quantifying gross human motion", Unpublished PhD thesis, Pennylvania State University
[20] Qualisys : http://www.qualisys.com/
[21] S. Bråten, 2003, "Understanding the timing movements in ski jumping", 11th European Congress of Sports Psychology, Copenhagen, Denmark