Three-dimensional Motion Capture, Modelling and Analysis of Ski Jumpers

Atle Nes

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

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

Abstract

This paper presents a brief description of a computer system that can be used for precise motion capture, analysis and visualization of human beings. More specifically this paper focuses on capturing and studying the motion of ski jumpers in 3D. The results from the motion analysis 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. Video images are captured simultaneously from three synchronized digital video cameras placed optimally around a ski jump 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 markers serving as robust feature points strategically placed on their jumping suit and skies. Feature points are followed in time and space creating motion tracks in 3D space. The captured feature points and their motion tracks are then connected back onto a dynamic virtual model of a ski jumper, which can be visualized and analyzed using a virtual model of the ski jumping arena.

1. Introduction

One particular problem 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. Using specialized computer software to study the outdoor ski jumps in greatest detail or extending indoor training with virtual ski jump simulations can reveal important clues necessary for helping the individual athletes search for a better ski jump.

Previous work

Most work done on this topic in previous research has been done in 2D, and most of this has again focused on ski jump take-off. This article focuses specifically on 3D motion capture of the ski jumper both during take-off and in the air before the ski jumper has stretched 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 ski jumping in 3D, much research has on the other hand been done in the more general field of 3D human motion analysis. An extensive summary of status and advances showing important concepts, procedures and techniques related to human motion analysis is given in a recent article by Wang, Hu and Tan [1].

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 HiST. Due to this project, 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. Our main contact at Human Movement Science Program is Steinar Bråten, which has regular contact with the Norwegian national ski jumping elite 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 photogrammetric techniques for studying ocean currents in the project HYDRIV [2] [3], and Institute of Geomatics at NTNU, which has refined the accuracy of land photogrammetry for at least a century. In addition we have established a research co-operation with Fachhochschule Bonn-Rhein-Sieg in Germany, which is also working on topics like visualization and motion analysis. At our Faculty of Informatics and e-Learning there are also several student projects which deal with parts or related topics supporting this project.

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 closely related applications using the same type of camera equipment and similar techniques as described in this article. Section 8 gives a summary and conclusion.

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 purchased consists of three identical miniature Marlin F-080B digital CCD-based video cameras from Allied Vision Technologies [4] [5], 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 31,71 m x 26,07 m at 30 m distance in the ski jumping hill.

Cables and Communication

Each camera is connected to the computer through a separate 4.5 m IEEE-1394 FireWire cable. The cables have 6-pin connectors, thus delivering both power to the cameras as well as transmitting video and control data. IEEE-1394 FireWire is today a well-established standard for high-speed two-ways serial data communication, among other applications widely used for video [6]. The computer is equipped with two identical FireWire interface cards with two connectors on each card. Each FireWire card represents a physical FireWire bus (IEEE 1394a) with a theoretical transmission speed of 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 two separate FireWire interface cards are needed.

In a ski jump arena one needs much longer cables than 4.5 m. To sustain the high transmission speed over longer distances the most obvious solution is to use optical fiber as carrier for the signals and data. A set of two 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 1000 m (Figure 1). Full duplex data rate at 400 Mbit/s is maintained with this solution. Regular FireWire cables will be used in each end to connect the cameras and the computer to the optical repeater boxes.

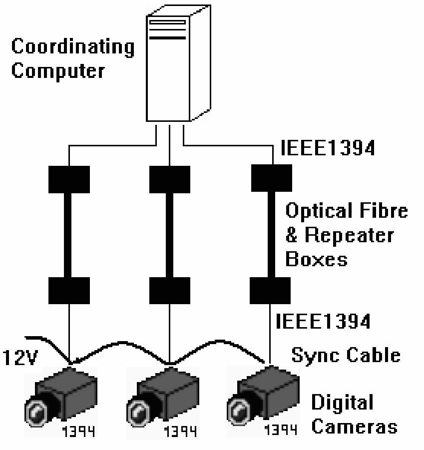

Figure 1. Multiple camera system

The cameras use the IIDC v1.3 protocol, also called the DCAM standard, for communication [7]. DCAM is a well-established 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 lower bit rate at the price of lost image quality. Lossy compression must not be allowed in the system before reaching the image processing stages.

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 more video than there is feasible memory space. 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 as high as 10.000 or 15.000 rpm.

Synchronization

To be most 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. Without the synchronization cable there will be a maximum difference in time 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 will be fetched instead. The synchronization cable is connected to each video camera with a 12-pin HiRose plug and receives 12 V DC-power from the computer using a FireWire connector. Each time an image is to be taken one of the video cameras sends out a trigger signal to all the other video cameras, which are set to listen for this external trigger signal. An external trigger signal can probably decrease the maximum time difference between corresponding image frames down to microseconds. Since this time difference can influence the accuracy considerable an experiment should be set up to determine how synchronous the cameras can get using the synchronization cable.

Placement

The placement of the cameras is important to 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 from the ski jump edge and about 30 m along the axis of motion in all cameras. Different lenses may thus be needed if the cameras have different distances from 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 [8]. Best feature point recognition performance is achieved when the angles between the cameras is as little as possible which usually implies that the identical feature points look similar in all cameras. In addition occlusion is in most cases preventing one from placing the cameras exactly 90 degrees perpendicular to each other. 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. The idea that sticks out at the moment is to put one camera on each side of the ski jump and one camera up in a light mast.

3. Motion capture

An important part of the project will be to track robust feature points over sequences of video frames [9]. It is necessary to have automatic detection of robust feature points because marking them manually will be an extremely large and impractical task. 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. The signalled human body markers can be white spots on the ski jumpers suit and skies or reflectors of some kind. Naturally robust feature points can be found by studying edges, corners, gradients etc. Harris corner detector is a commonly used and classical algorithm for finding such robust feature points [10].

It is important that the features are robustly detected by image processing techniques and that they are recognized both over time and 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. It may be necessary to also look at robust feature localization problems when having blurred images caused by ski jumpers moving too fast compared to the video frame rate.

An alternative approach for motion capture that does not need feature points to begin with, is capturing the blob motion of a ski jumper and using morphological combined with filtering operations to reduce the segmented blob to a stick figure. It could be interesting to investigate the accuracy of such an approach compared to approaches using robust feature points.

4. Photogrammetry

Before any 3D measurements can be performed in a scene one needs to define the geometry for 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 will be used for this task [11].

Direct Linear Transformation (DLT)

The method used for finding the geometric relationship is the DLT method (Direct Linear Transformation) originally reported by Abdel-Aziz and Karara in 1971 [12] 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 having control points spread evenly in a 3D volume can be accomplished by rotating and capturing multiple images of a simpler 2D calibration object. The flexibility of the DLT-based calibration often depends on how easy it is to handle the calibration frame. In the ski jumping hill we plan to use on-site calibration with predefined coordinates in the terrain as a kind of calibration frame. The coordinates will be measured in advance using known coordinates and standard land measurement techniques.

Recording images 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 2). Two reference frames are defined; the object space reference frame defining the

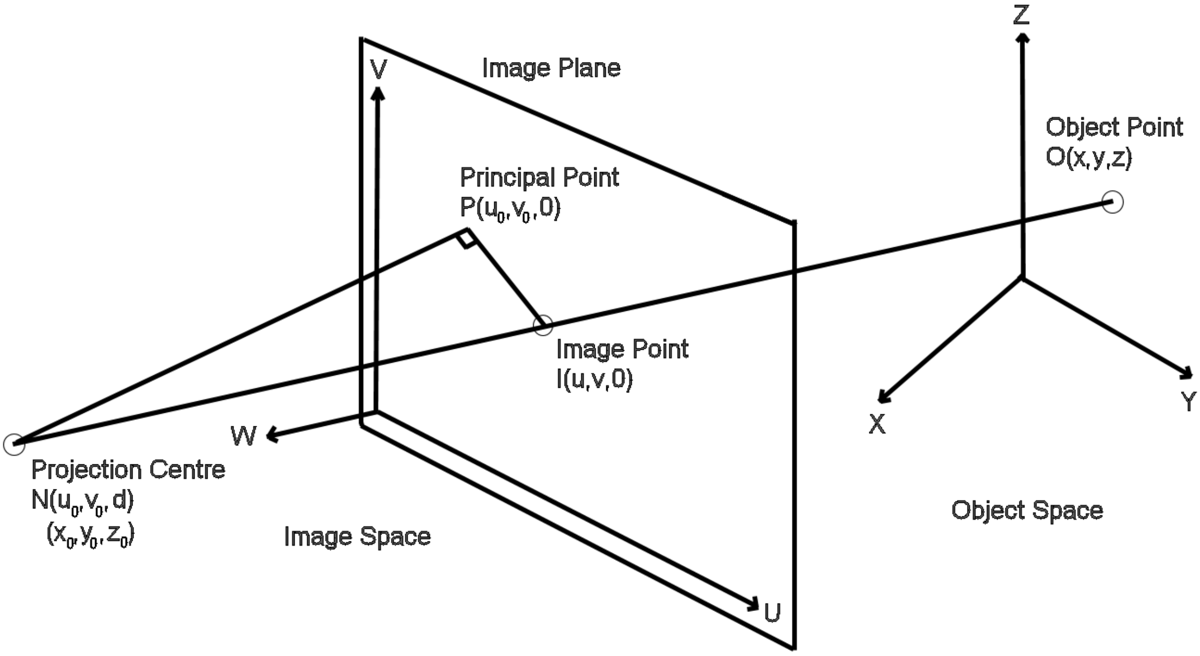

Figure 2. Projection of an object point to an image point

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 [12]:

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

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

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 overdetermined 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.

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 accuracy even further one has to consider the imperfect 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 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 [13]. 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 [8]. 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 now 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 are much higher compared to the scientific video camera described earlier. One such system, used indoors at Human Movement Science Program in Trondheim is the ProReflex system from Qualisys [14]. Factors like expensiveness of purchasing and difficulties concerning setting up such a system in the ski jumping hill has contributed to trying to use regular video cameras instead. ProReflex is considered to be stationary installed equipment used for indoor training exclusively.

Granåsen ski jump arena in Trondheim will be used as a test ground, and Steinar Bråten has a permit for doing experiments there. A student group at the faculty has obtained and modified a virtual CAD model of the Granåsen ski jump arena [15]. Another student group has been working on a dynamic model of a ski jumper and the latest news is that they now are able to control the ski jumpers motion in the static ski jump arena model [16]. It will be interesting to find out how accurate the computer model is compared to the real ski jump arena. The model must be changed to fit the measurements of the ski jump arena and the captured video data of the ski jumper in the hill.

Another thing that 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.

Some 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. It is also possible to also project some movements from the constructed 3D representation to new 2D projections.

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.

8. Summary and Conclusions

This article has given a short 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] L. Wang, W. Hu, and T. Tan, ”Recent Developments in Human Motion Analysis”, Pattern Recognition no. 36, pp. 585-601, 2003.

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

[3] J. H. Nilsen, “An experimental study of internal tidal amphidromes in Vestfjorden”, Dr. ing thesis, Department of Structural Engineering, NTH, Trondheim, Norway, ISBN 82-7119-640-5, 1994.

[4] Allied Vision Technologies : http://www.alliedvisiontec.com/

[5] Allied Vision Technologies, “AVT Marlin Preliminary Technical Manual”, 2004.

[6] 1394 Trade Association : http://www.1394ta.org/

[7] 1394 Trade Association, “IIDC 1394-based Digital Camera Specification Version 1.30”, 2000.

[8] R. I. Hartley, A. Zisserman, “Multiple view geometry in computer vision”, 2004.

[9] J. H. Nilsen, I. Hådem, “Photogrammetric tracking of tracer particles in modelled ocean flows”, ISPRS Journal of Photogrammetry and Remote Sensing no. 49, pp. 9-20, 1994.

[10] C. J. Harris and M. Stephens, “A combined corner and edge detector”, Proceedings 4th Alvey Vision Conference, Manchester, pp. 147-151, 1988.

[11] I. Hådem, “A program package for industrial photogrammetry”, Kart og plan no. 1, pp. 24-25, 1989.

[12] Y. I. Abdel-Aziz and H. M. Karara, “Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry”, Proc. of ASP Symposium on Close Range Photogrammetry, Urbana, Illinois, USA, pp. 1-18, 1971.

[13] J. S. Walton, “Close-range cine-photogrammetry: a generalized technique for quantifying gross human motion”, Unpublished PhD dissertion, Pennylvania State University, 1981.

[14] Qualisys : http://www.qualisys.com/

[15] C. Johansen, H. Homb, R. Blekken, “Ski Jumping Hill”, Student project, 2004.

[16] P. Johansen, L. A. Wik Larsen, H. Østensen, “Ski Jumper”, Student project, 2004.