HISTORICAL ASPECTS AND CURRENT TRENDS IN CINEMATOGRAPHY

Joachim Wolf
Olympic Sport Training Center, Stuttgart, Germany

INTRODUCTION

The basic task of cinematography is, by definition, the measurement of the location of relevant points at known times. This can be achieved by a number of methods. To date these methods are systematized according to the measurement principle. For a problem oriented inductive selection of methods a differentiation based on reliability of the results seems much more appropriate.

A basic prerequisite for reliability and thus validity is the quality, i. e. the spatial and temporal resolution of the employed sensor. Spatial resolution is a function of the smallest discernible distance between two points and the dimensionality of the measuring system. Temporal resolution is determined by the sampling frequency. The required quality depends on the purpose of the research, which is thus the decisive criteria for the choice of sensors.

In the event, that two- or three-dimensional motion occurring at precisely predetermined locations is to be registered with high temporal and spatial resolution, cinematography is surely the adequate method. Hence cinematographic techniques play a central role amongst the methods to measure kinematic parameters.

HISTORICAL DEVELOPMENT

In 1891, the German scientists Braune and Fischer realized the necessity for describing motion not only in a two-dimensional but in spatial form. When they were looking for a suitable analysis method, it became obvious, that the problem could not only be solved with one perspective of the motion. That is why they took pictures of the movement of the leg, which was the point of their kinematic interest, from two orthogonal directions simultaneously. To avoid the problem of temporal synchronization of both cameras the scientists used a æRumkorff spark inductorÆ and metal rods emitting the electrically induced sparks fixed on three points connected with the foot of the subject. The sparks appeared simultaneously at all three points. The researchers adjusted the inductor in a way, that during the bending of the knee, which took about one or two seconds, 20 to 30 sparks flashed. Therefore 20 to 30 motion phases could be fixed. As the sparks could not be produced in exactly the same intervals, time dependent measurements were not precise, but this was not the aim of the analysis.

In 1895, exactly 100 years ago, they began to investigate the human walking pattern. They were no longer satisfied with the observation of single points to represent the segments and installed instead so called 'Geissler's tubes' parallel to the important body segments. By doing this they received further insight into the movement of the segments during walking. These tubes were filled with glooming gas turned on by electrical impulses of the 'Rumkorff's spark inductor'. The tubes were not attached directly to the subjects limbs to avoid electric shocks. Braune and Fischer were not only interested in position change but also in time parameters and their derivation. Therefore the impulses of the inductor needed to be produced in constant intervals. They used the constant vibrations of a tuning fork, whose frequency was recorded and analyzed. The scientists were aware of the possibility, that during some phases one or the other tube could be covered by the limbs of the subject. Consequently they used four cameras for their analysis of gait. The results of their cinematographic analysis warranted exact statements about the change of position of single body segments as a function of time.

The work of Braune and Fischer was continued in Moscow in 1931 by Bernstein and his colleagues. Again they devised a new method to obtain a synchronized projection of a movement from two directions by using just one camera with a rotating shutter and projecting the second view with a mirror. They reached sampling frequencies up to 150 Hz and were able to derive ground reaction-force vs. time curves using Braune and Fischer's anthropometric data. The method of synchronization was, however, not quite as precise because of the optical distortion of the mirrored image.

Many scientists were afraid of a complete 3-D-analysis because of the amount of work required and inaccurate results. Necessary determination of the external orientation of the optical system was and still is costly enough and requires a lot of time. A fundamental simplification for determining position coordinates was reached in 1971, when Abdel-Aziz and Karara developed the method of 'Direct Linear Transformation'. This well-known DLT-method not only saves incorrect and costly measurements, as they are necessary for metric cameras in photogrammetry, but also permits the use of any camera. Since the basic control point system describes only this single static camera condition, it is absolutely necessary, that camera position and zoom are precisely identical during calibration and the recording of the motion. When DLT-methods and its relating reduction of measurement expenditure for 3D-analysis were developed, a potent commercial market for orthopedic and sport science orientated computer-based analysis systems emerged. The gathering of data, administration and evaluation was more or less automized as required by the user. Common smoothing algorithms were offered and the derivation of a multitude of further kinematic and partly dynamic parameters on certain conditions were facilitated. The attempt to automatize and economize the time consuming digitizing of two-dimensional picture coordinates lead to the development of opto-electronic methods. These systems require a laboratory situation, which influences the movement. They are thus best suited for special clinical research rather than for sport-scientific investigations.

A multitude of movements, which are analyzed in sports biomechanics, occur in a horizontal plane. They can therefore be analyzed in three dimensions based on frames of two purely horizontally panned cameras. The resulting mathematical relation between known picture coordinates and sought after object coordinates can be solved when the internal and external camera orientation is known. As an example, the method of Dapena, who examined back in 1978 the high jump in this three-dimensional way, should be mentioned. The determination of both one-dimensional pan-angles occurred again with the help of the coordinates of known control points, which were placed in the field of view. Most publications dealing with a one-dimensional pan assume an exact vertical spatial arrangement of the pan-axis and the rotation point is equated with the center of the projection. But Drenk has proved, that this simplification can lead under certain conditions to considerable errors when determining 3D-coordinates. Therefore alternative methods for exact calculation of the location and orientation of the pan-axis and the relative position of the center of projection to the pivot point were discussed.

At the end of the eighties the sport-scientific research had reached a level, where a uniaxial pan or the limitation to 2D-analysis was no longer acceptable. Yeadon was one of the first, who had a critical look at the problem when he examined a variety of body positions in the course of a ski-jump in 1989. The method he developed admits both a biaxial pan of both cameras and free zooming. But this required because of proportionality of angles and digitized distances a predetermined distribution of both control points in the picture to avoid excessive errors when calculating the unknown object coordinates. Other authors tried to determine the external orientation of the camera by surveying techniques and goniometrical measurements as precisely as possible. The three-dimensional position coordinates were computed using basic photogrammetric equations. It is obvious, that these methods highly depend on the quality of orientation parameters of both cameras.

Two methods, which became quite popular in the last years, are also based on the DLT-procedure. Stivers and others (1993) proposed a method to reduce the initially 16 unknown physical parameters of the central projection by constant measurements of the pan-angle and by considering the geometrical conditions to 10. With this so called 'physical parameter transformation' (PPT) they succeeded on the one hand in receiving a higher accuracy of the determined DLT-parameters and on the other hand the pan of the recording system was made possible. Drenk determined the pan- and inclination-angle of his camera using two control points in each frame. On this basis he could compute the relevant DLT-parameters, location and position of the pan-axis taking into consideration the special geometry of the tripod. On the basis of the two known control points, it was possible to determine the internal orientation of the camera continuously, which allows a free zoom. Drenk refers furthermore to an approach, whereby the pan-angles assuming known camera orientations can be computed from a single fix point from both perspectives.

The approach of a fictitious pan of the reference system has meanwhile been integrated in commercial systems. The differences lie in the way how the pan angle is determined: Some use rather elaborate systems incorporated in the tripod, that measure electronically and store the current pan-angle onto the video tape. Others solve the problem by computing pan-angles based on reference points in the frame. This also permits variation of the focal length.

A further possibility is of course to determine the DLT parameters for each frame. This requires however a 3D reference frame visible in each picture. The advantage here is, that the cameras may be moved and zoomed as appropriate. Our method uses a combination of the surveying technique and two reference points for the determination of the camera orientation and the focal length. A DLT reference frame is not required, however, one needs to know the camera locations as well as the control point locations along the path of motion precisely. I will now discuss the mathematical basis of this procedure, which also serves to illustrate the commonly used principles of perspective projection.

BASIC MATHEMATICAL PRINCIPLES

The theory of perspective central projection, describes mathematically unambiguous, the relation between the digitized coordinates of a view and the object coordinates to be determined. Early investigators such as Braune and Fischer used this approach for their analyses. From this model the basic generalized equations of perspective projection for a tilted view can be deduced directly. When solving the equations to determine the spatial coordinates of an unknown point we run into two basic problems: There are only two equations but three unknowns and we do not know the initial orientation parameters of the camera. The first problem is taken care of by using at least two cameras. The optimum solution for specific classes of the problem, i.e. the determination of camera location as well as the orientation of the pan-axis in space and the pan-angles, has been subject of research for a number of years. The DLT method facilitates the procedure since the camera orientation parameters need not be measured but are rather approximated on the basis of a minimum of six object space-frame coordinate pairs. The determination of unknown points in the object space is performed on the basis of the image coordinates and the 11 DLT parameters. In the case of panned cameras, the orientation parameters change constantly. Hence, the need for an adaptation of the DLT method or development of new procedures arose.

Since the projection center changes only insignificantly in relation to the pan-angle, it appears sensible to determine the pan-angle in some way and then feed this data to a modified DLT algorithm. The continuous electrogoniometric measurement is rather expensive. The reliability of the results depends largely on the precision of the instrumentation. The determination of panning-angles on the basis of known reference points is problematic, since pan- and tilt-axis are coupled. To optimize results one uses special iterative algorithms. In both cases the system of equations can be simplified through a special tripod with a given relative position of the pan- and tilt- axis. If one wishes to zoom, which becomes possible as a side effect of the two available reference points per frame, one has to position and measure a set of control points along the anticipated path of motion. In this case the additional determination of the camera locations presents no problem. We have developed a method to determine camera orientation based on the measured centers of projection and an adequately positioned set of control points. Starting point is the perspective projection. The mathematical relation between the coordinates of the object point P and its projected point PÆ can be derived through simple geometry. These equations, however, are only valid for the special photogrammetric case when cameras are fixed and the frame and object space planes are parallel. A pan of the camera leads to an inclined projection coordinate system. The equations have to be adapted accordingly. The transformation of the coordinates is performed stepwise about the three orthogonal axis and is described by the rotation matrices. Due to the mechanics of the tripod used there is no twist about the optical axis. The horizontal rotation about the vertical axis (the pan) becomes the primary, the vertical rotation about the horizontal axis (the tilt) becomes the secondary rotation.

We obtain certain trigonometric functions for the elements of the resulting rotation matrix. These are integrated into the system of equations for the central-projection. There we need to know the precise projection center, the focal length as well as the orientation angles to compute the desired object coordinates. The methods for 2D panned 3D analysis were developed explicitly for this purpose.

In our case, the camera locations are measured precisely, the focal length and orientation angles are obtained in a first approximation from the control points. A precise computation is impossible, since the parameters are not independent. Thus a successive approximation is done with an iterative procedure based on the decoupled equations. It is obvious, that the object coordinates computed from these approximated values have systematic and measurement errors. To minimize the errors the mathematical method of least squares is employed. It is used in all methods for 3D cinematographic analysis.

GENERALIZATION RELATING TO THE NUMBER OF VIEWS

In order to improve the reliability of the object point determination, it is desirable to increase the number of cameras. The reasons for this are the increased likelihood of positive identification of object point, which may be obscured, and the fact, that the least squares algorithm gives better results because of the overdetermination of the system of equations, which increases with the number of available views. Hence our procedure allows for an arbitrary number of cameras. Let K be the number of cameras, then K is greater than or equal two since one perspective is not sufficient for 3D reconstruction. The number of unknowns is determined by the three spatial coordinates, the focal length and the orientation angles of the K cameras and can thus be written as 3+3*K.

The number of observations n is given by the six projected view coordinates of the three digitized points per frame: these are one unknown object point and two reference points. Thus n=6*K. The observations Ln are improved by addition of Vn such that the product of Vt and V is minimized (least squares fit as usual). The complex equations representing the observation are linearized using TaylorÆs theorem. This requires deduction of the partial derivatives. We end up with an error term V, that is input into the equalizing approximation algorithm. The resulting normalized equations give the improvement terms dX. According to NewtonÆs method for solving systems of non-linear equations, these values are added to the approximations. The process is then repeated. Iteration is halted once the improvements decrease below a predefined level.

Of course we were interested in the accuracy of our method. The first approach was an empirical validation, where two cameras and control points were used to measure a third control point. The differences between the computed and measured values was with camera distances of about 150m in the magnitude of only a few centimeters. Moreover we investigated the effect of increasing the number of views on the accuracy. We chose a theoretical approach and derived the dependence of the number of extra observations and the mean error of the approximated unknown on the basis of the least squares fit. This allows computation of the factor of decreasing mean error for an increased number of cameras. Basing the analysis on three rather than two cameras results in a factor of 0.71, increasing the number of views from three to four yields a factor of 0.82. By taking four rather than two views the error is reduced by the product of 0.71 and 0.82, i.e. by the factor 0.58. However one has to consider that this theoretical improvement is partially influenced by the concurrent increase in systematic and random errors. Hence, going beyond a certain number of views does not improve the measurement any more.