

# POINT TRACKING AND IMAGE RECTIFICATION IN VIDEO ACTIVITIES

André Heck,  
AMSTEL Institute, Universiteit van Amsterdam  
heck@science.uva.nl

*Video analysis means that you capture a video clip with a web cam or digital camera, collect data of a moving object by clicking with the mouse in video frames, and finally analyse the recorded motion data. Two problems experienced in practice are that manual data collection can be too time-consuming and that perspective distortion can interfere with useful video analysis. In this paper we present how the computer learning environment Coach deals with both issues. Examples illustrate that point tracking in video clips and perspective correction of images of planes broaden the scope of video and image measurement in practical work of students.*

## POINT TRACKING

In [1] we presented learning materials for students in upper pre-university education to act like human movement scientists inside and outside the classroom. They record video clips of their own ways of walking with a web cam, collect high-quality, real-time data from these short movies, and use the computer learning environment Coach to analyse the collected data. We try to make their science learning resemble science practice, in which investigations can often be characterised as being challenging, complex, open-ended, and cross-disciplinary, and as requiring a strong commitment of participants plus a broad range of skills. In practice, this commitment of students is often put to a test when the data are collected manually in a video clip by mouse clicking. We quote two secondary school students, who investigated in a research project several gait patterns at various speeds and inclination angles:

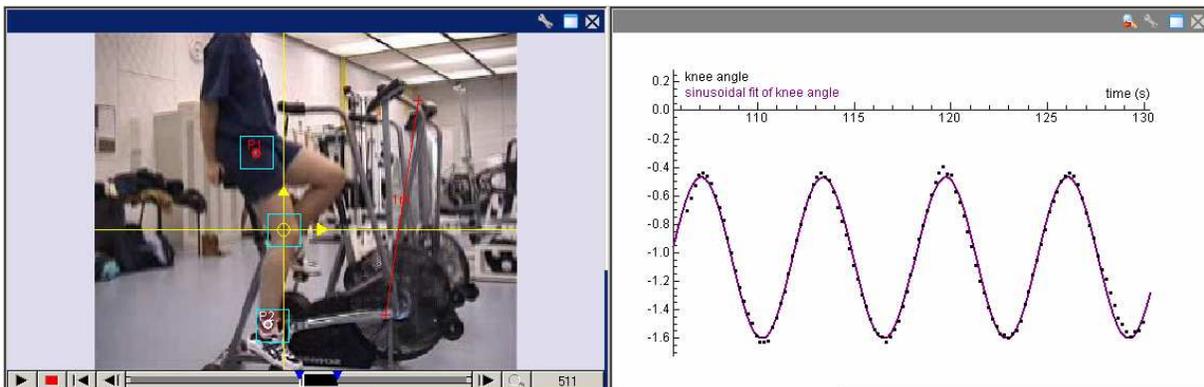
“...Furthermore we had to click three points in each video frame: the knee, the ankle, and the hip. We have looked at two steps for each variation of speed and inclination. On average  $19 \frac{1}{3}$  frames per step, times two because we do two steps, times four for the degrees, times three for the speeds, times two because we did everything twice, times three for the number of mouse clicks. This is a total of 2784 clicks per subject. Just because of the careful recording of the results it took an hour per subject. When you click for one hour, you make mistakes sometimes, so that this will also have lead to inaccuracies.”

The students' complaint is pertinent: collecting data in this kind of research work is too time-consuming, extremely boring, error-prone because the concentration on careful mouse-clicking diminishes, and it limits the number of experiments in an investigation. What is needed is the possibility to track the motion of an object, e.g., a marker or an eye-catching point, and to record automatically the coordinates of the moving object in subsequent frames of the video clip.

Point tracking has been implemented in Coach as a DirectShow transformation filter, which can be thought of as a program that connects a source, e.g., a video clip or a

streaming video of a web cam, with the video window, while at the same time passing the recorded position data to the computer application. The algorithm used in the tracking filter is composed of two parts: (1) finding the best match of a given model template in a subsequent frame, i.e., locating the area that resembles most a specified area, and (2) limiting the search area in order to reduce computing time or to avoid ambiguity. The template that will be tracked and is selected by the user at the start of the tracking process has been chosen to be an area bounded by a circle with a user-specified radius. The comparison of an area with the model template is based on pixel intensities in the three channels of the RGB colour model. More specifically, the sum of squared differences of intensities between image and model template is used as a function that must be minimized in the search algorithm. We have chosen this algorithm for template matching after testing various commonly used algorithms on sample video clips because of its performance in spite of its simplicity. The search area for the next frame has been chosen to be a rectangle of user-specified dimensions, centred on the position that was found for the current frame.

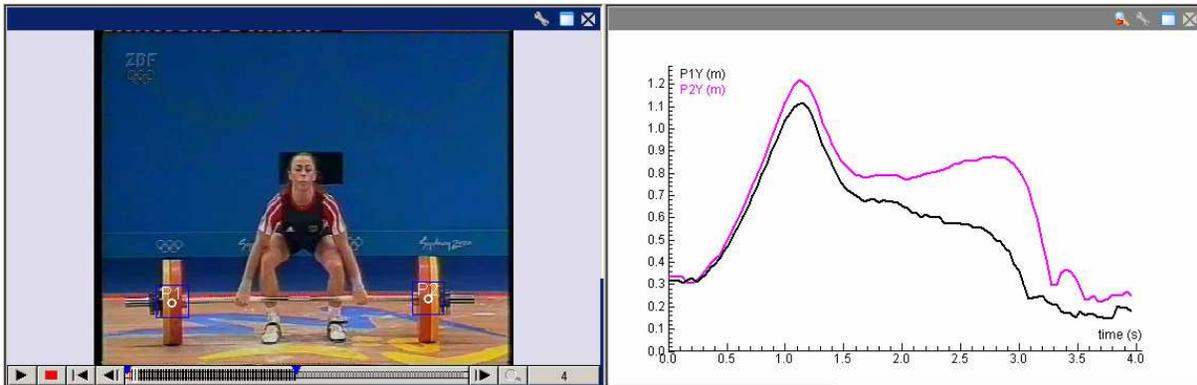
Let us illustrate point tracking in Coach by an example in which the leg motion of a student on an exercise bike in a fitness centre is investigated in much the same way as we described for gait research in [1]. The coordinates of the hip and the ankle of one leg with respect to the knee joint during cycling are recorded via point tracking. So, we work with a moving coordinate system in which the origin always coincides with the knee joint. The origin is tracked in the video clip after selection of its model template in the first frame. In the first frame we also select the hip and the ankle for point tracking. The positions of the hip and knee joint are also recorded in polar coordinates. We will only use the information about the angles to compute the knee angle during the cycling motion. Figure 1 is part of a screen shot of the video activity. In the left side you see the video window with three rectangular search areas for the points that are tracked. In the diagram to the right you see the measured knee angle versus time, together with a sinusoidal fit of the data. This fit gives food for thought.



**Figure 1.** *Screen shot of a video activity about fitness cycling with point tracking*

Another example of point tracking is shown in Figure 2. The video clip to the left comes from the women’s weightlifting competition at the Sidney Olympic Games in 2000. It shows Tara Nott failing in her 3<sup>rd</sup> attempt in the clean & jerk to lift 105 kg,

but nevertheless winning the gold medal. The weights at the end of the barbell are tracked and their heights are plotted in the diagram to the right. The students' task is to relate the graphs with what goes wrong in the clean phase of the lift. Do you see it?



**Figure 2.** Tara Nott failing to lift 105 kg in the clean & jerk

Some features of point tracking in Coach are:

- You can combine point tracking with manual video measurement.
- You can select a moving coordinate system and track its origin.
- When point tracking fails, you can interrupt it, move to the first frame where the tracking algorithm fails, reinitialise the model template or even change its settings and/or the dimensions of the search area, and continue point tracking from this frame on using the new settings.
- Point tracking can be combined with other video filters that have been implemented such as change of brightness and contrast, and correction of perspective distortion.

The main advantages of point tracking in video activities are:

- When a manual video measurement takes more than 100 mouse clicks it becomes too much work of monotonous and RSI-risky nature.
- The automatic video measurement of 3 points in 112 frames in the above example of a cycling student takes about 40 seconds. So, collection of data is done in much shorter time than ever could be achieved by manual recording of data.
- When a user collects video data by hand, he or she can be distracted or lose concentration during measurements, or get tired and make mistakes. This influences the quality of the measurements in a negative way. Point tracking is an objective way of data collection. The accuracy of measurements is the same for each frame of the video clip.

This does not mean that point tracking is free of problems. At the current stage of development of Coach, video clips must match the following requirements for successful point tracking:

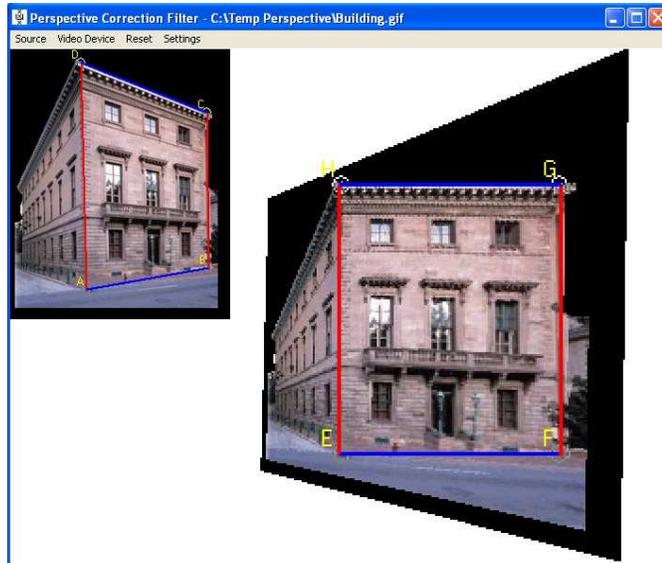
- Tracked objects may not change much in size, shape, and colour. This limits for example the usage of many television recordings of sport events.
- Markers are preferable of circular shape because of our choice of a circular-shaped model template (so that rotating objects can be tracked) and they better

have a colour distinct from the background scene in the search areas in order to minimize errors in tracking.

- Tracked points must stay within the view field as much as possible. If a tracked point temporarily disappears within the view field and later reappears again, then it is not guaranteed that the point is always automatically found again. In this case the user must reinitialise the model template and continue from the frame where the point reappears in the video clip or start over again with a larger search area.

## PERSPECTIVE CORRECTION OF IMAGES OF PLANES

The image of a world plane captured by a camera experiences perspective distortion. The picture in the left side of Figure 3 shows an image of such two planes, a photograph of two external walls of a building. Observe that the rectangular features, such as the windows, do not appear as rectangles in the image: the plane geometry has undergone a transformation that distorts Euclidean properties such as length, angles and parallelism. The picture in the right side of Figure 3 shows the perspective corrected front of the building, where the plane geometry is that which would be seen had the



**Figure 3.** *Perspective correction of the image of the front of a building*

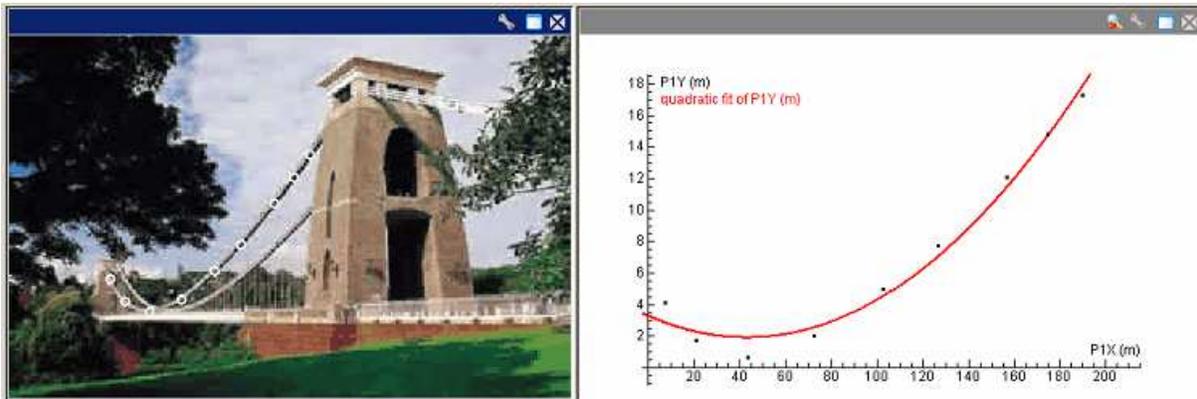
original photograph of the plane of interest been taken with the camera fronto-parallel to the wall. Such image rectification of planes is useful before data collection in a digital image or in video clip where an object moves in a (imaginary) plane.

Rectification of planes in digital images is part of the discipline that is known as computer vision. Projective geometry lays the mathematical foundation of this discipline. For example, the map between the world plane and the perspective image is a plane projective transformation, a 2D homography. Correcting perspective distortion is a matter of determining the eight degrees of freedom of the homography. The rectifying homography is computed from scene geometric information, specifically parallelism, angles between lines, and ratios of length along lines in different direction. A full description the mathematics of image rectification of planes can be found in [2,3] and in many textbooks on computer vision.

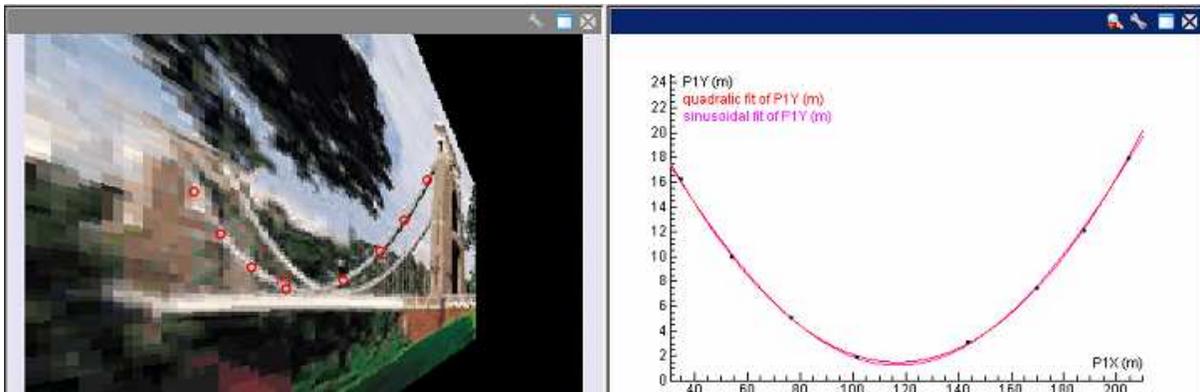
We will use Figure 3 to exemplify in what way the perspective distortion of a plane can be corrected in Coach. Our plane of interest will be the front of the building. In reality, the edges of the front wall are two pairs of parallel lines; but this is not the case in the photograph of the building. The image transformation that restores this

property is determined by mapping the 4 corners of the front wall with a projective transformation to the corners of a rectangle in a new image. Actually, the orthogonality of two pairs of parallel line determines the plane rectification up to an unknown aspect ratio; but this scaling can be specified in Coach, if necessary. In Figure 3, the result is shown on the right side. Actually, the basis of the method, which is also used in [2,3], is that any projective transformation can be uniquely written as a composition of three transformations, viz., a perspectivity  $P$  that maps a 4-gon into a rhombic figure (which consists of two pairs of parallel lines), an affine mapping  $A$  that transforms the rhomb into a rectangle, and finally a similarity transformation with isotropic scaling that rotates, translates and/or scales the rectangle. For measurements in a rectified photograph or video clip it suffices to find appropriate transformations  $P$  and  $A$ .

### WHAT SHAPE HAS THE CLIFTON SUSPENSION BRIDGE?



**Figure 4.** Screen shot of a photogrammetry activity on the Clifton Suspension Bridge



**Figure 5.** Screen shot of the bridge activity after image rectification

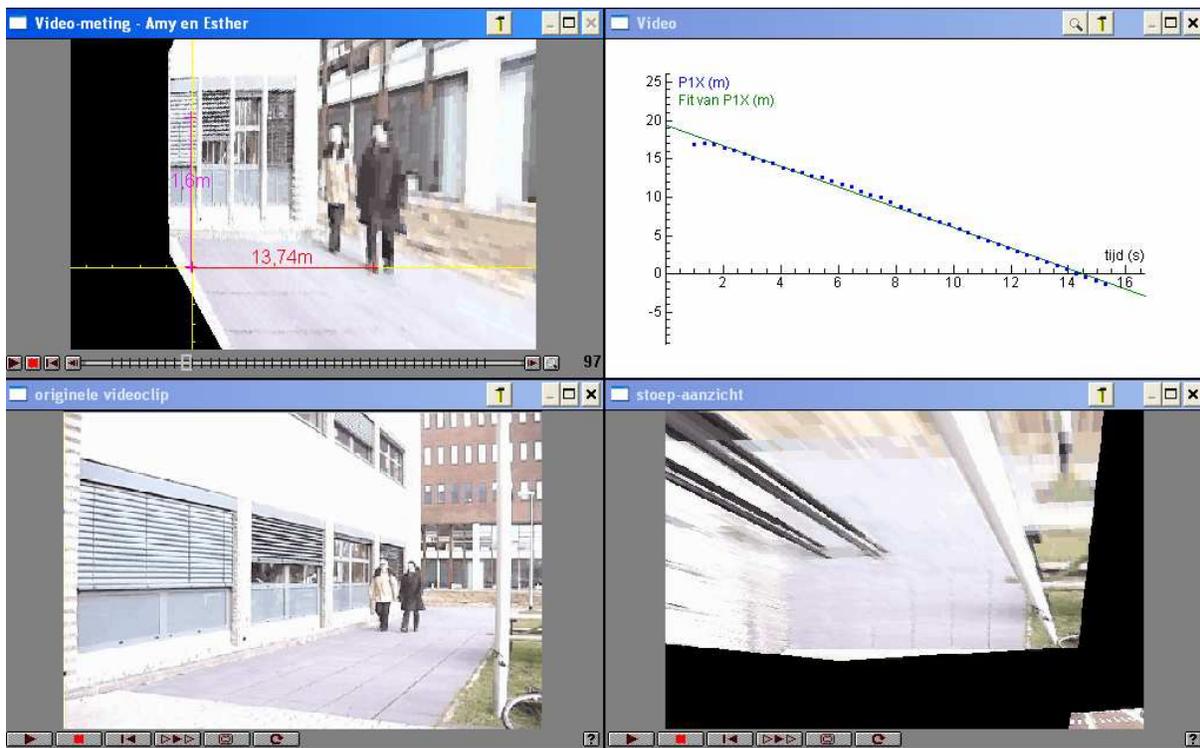
In mathematics textbooks you often see a photograph of bridge in a distorted perspective with the message that the shape of the arch is a parabola. This might be true, but not if you collect data from the photographs in the book and apply a quadratic fit to the data. In Figure 4 above you see a screen shot of an image measurement activity applied to the picture of the Clifton suspension bridge taken from [4]. The measured

positions on the cable do not lie on a parabola: the right diagram shows the best quadratic fit and it does not match well with the data.

Figure 5 above is a screen shot of the activity in which the plane formed by the piers and the left side of the road (as seen from the camera position) is rectified to a fronto-parallel view. The cable is almost in this plane and via the rectifying transformation it becomes a parabola. The diagram to the right shows the best quadratic fit. This does not necessarily mean that this is really the shape of the cable: a sinusoidal fit will do as well, and perhaps the shape can be modelled as an ellipse. Only physics and knowledge about bridge building can help students to understand which model is most appropriate (see [5]).

In the corrected, fronto-parallel view of the bridge, the scaling in horizontal and vertical direction is not the same. This causes a strange ratio of height and span of the bridge. However, this does not matter for the mathematical shape of the curve of the cable: it remains a parabola. But do students also understand this? An interesting investigation task for students is to sort out under what familiar mappings of a plane parabolic curves remain parabola.

## HOW FAST DO THE STUDENTS WALK IN THE VIDEO CLIP?

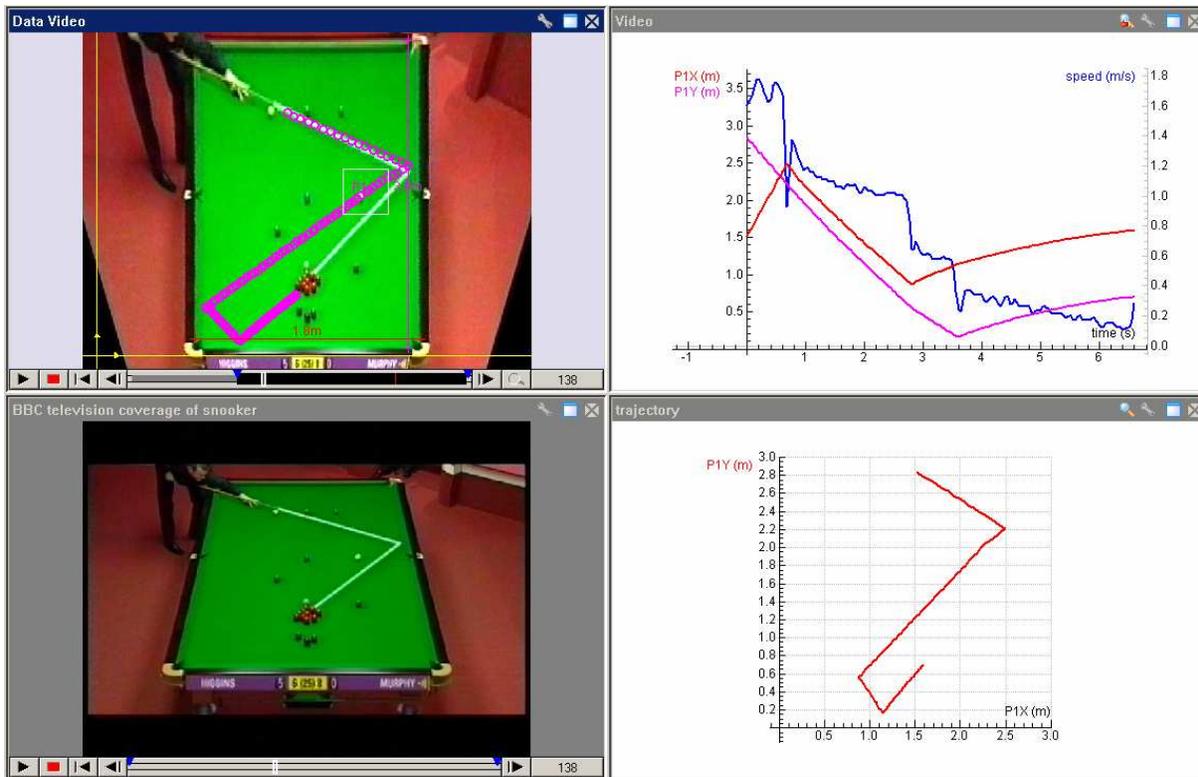


**Figure 6.** Screen shot of a video activity with students walking in a straight line

Image rectification can be used to answer the question of how fast an object is moving towards the camera in a perspective view; see Figure 6 for a screen shot of such a video activity. In the lower left corner you see a video clip of two students walking in front of a building, in a straight line in the direction of the camera. The

question is at what speed they walk. To get an answer to this question you can apply image rectification to two planes: the perspective correction of the imaginary vertical plane in which the student to the right walks and the rectification of the plane formed by the paving stones. The result of the first image rectification is the video clip shown in the upper left corner of Figure 6. As vertices of the 4-gon that is in reality a rectangle we choose the left shoe and the head of the student to the right in two video frames that are far apart from each other. The rectification of the footpath gives the alienating video clip in the lower right corner. The rectified image of the vertical plane is used to measure the distance that the student walks. The data of this measurement are shown in the diagram in the upper right corner, together with a straight line fit. Apparently the student walks at a constant speed. The slope of the line implies a walking speed of 4.8 km/hr. Students can validate this answer by using the other rectified video clip for an alternative measurement and in this way practise a good research methodology.

## MATHEMATICS AND PHYSICS OF SNOOKER



**Figure 7.** Screen shot of a video activity about snooker

Figure 7 is a screen shot of a video activity in which all features of video measurement in Coach are exploited to study the mathematics and physics of snooker. In the lower left corner you see a video clip coming from the BBC coverage of the snooker's world championship 2005 at the Crucible. The movie director has put an overlay on the clip to inform the television viewer about the prospective shot. But the player decides something else! In the rectified video clip, in the upper left corner of

Figure 7, you see the trace of the shot that the player actually makes. Because of the overlay and the cue ball are both white, the first coordinates of the cue ball are recorded manually. As soon as the cue ball is away from the overlay, point tracking is switched on and the video measurement goes further in automatic mode. In the lower right corner you see the measured trajectory of the cue ball. When students measure the slopes of the line segments, they discover that there is an increase in bank rebound angle. This might be due to spin of the ball, but basic physics can help students to realize that it would also happen without spin. We refer to [6] for an explanation of this issue and many other interesting physical aspects of snooker. But you can also ask students questions such as why the measured points are further apart from each other in the manual measurement phase than in the point-tracking phase. Of course it has nothing to do with the mode of recording coordinates, but with the speed of the cue ball. This speed can be computed as the square root of its horizontal and vertical components. The graph of the speed of the cue ball is shown in the upper right corner of Figure 7. It also reveals that the bank rebound is not an elastic collision.

This example shows that students can explore many mathematical and physical aspects of a sports game when they have appropriate video analysis tools at their disposal. The other examples in this paper illustrate that this holds for many practical work that students could do. With tools designed for education, but mimicking the tools used by professionals, we can bring the practical work of students to a higher level of quality and we can challenge them to do realistic research work.

### Acknowledgements

The author would like to thank the students Eun Mi Kim and Esther Domínguez Sol for participating in the video clip. The computer science students Erik Peeters and Marks Smids provided the prototypes of the image rectification filter and the tracking filter, respectively. The author also would like to thank his colleagues Leendert van Gastel and Marijn van Eupen for fruitful discussions.

### References

- [1] Heck, A & Holleman, A. (2003). Walk like a mathematician. In T., Triandafillidis & K., Hatzikiriakou, (eds) *Technology in Mathematics Teaching*, Conference Proceedings of ICTMT6, Volos-Greece (pp 380-387). New Technologies Publications, Athens.
- [2] Liebowitz, D. (2001). *Camera Calibration and Reconstruction of Geometry from Images*. PhD thesis (chapter 3). Retrieval at [www.robots.ox.ac.uk/~vgg](http://www.robots.ox.ac.uk/~vgg)
- [3] Liebowitz, D. & Zisserman, A. (1998) Metric Rectification for Perspective Images of Planes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp 482-488). Retrieval at [www.robots.ox.ac.uk/~vgg](http://www.robots.ox.ac.uk/~vgg)
- [4] [www.clifton-suspension-bridge.org.uk](http://www.clifton-suspension-bridge.org.uk)
- [5] Heck, A. & Holleman, A. (2002). Investigating Bridges and Hanging Chains. In: M., Borovenik & H., Kautschitsch, (eds) *Technology in Mathematics Teaching*, Conference Proceedings of ICTMT 5, Klagenfurt, Austria (pp 409-413). Schriftenreihe Didaktik der Mathematik, v. 25. öbv&hpt, Vienna.
- [6] D.G. Alciatore (2004). *The Illustrated Principles of Pool and Billiards*. Sterling Publishing. Online resources at [www.engr.colostate.edu/~dga/pool/](http://www.engr.colostate.edu/~dga/pool/)