Video based swimming analysis for fast feedback

Paavo Nevalainen[1], Antti Kauhanen[2], Csaba Raduly-Baka[1],
Mikko-Jussi Laakso[1], Jukka Heikkonen[1]
[1] *Department of Information Technology, University of Turku, FI-20014 TURKU, FINLAND*
[2] *Sport Academy of Turku region, Kaarinankatu 3 20500 TURKU, FINLAND*
ptneva@utu.fi,antti.kauhanen@turku.fi,{csaba.raduly-baka,milaak,jukhei}@utu.fi

Keywords: athletics, swimming, motion tracking, camera calibration, signal smoothing, movement cycle registration

Abstract: This paper proposes a digital camera based swimming analysis system for athletic use with a low budget. The recreational usage is possible during the analysis phase, and no alterations of the pool environment are needed. The system is of minimum complexity, has a real-time feedback mode, uses only underwater cameras, is flexible and can be installed in many types of public swimming pools. Possibly inaccurate camera placement poses no problem. Both commercially available and tailor made software were utilized for video signal collection and computational analysis and for providing a fast visual feedback for swimmers to improve the athletic performance. The small number of cameras with a narrow overlapping view makes the conventional stereo calibration inaccurate and a direct planar calibration method is proposed in this paper instead. The calibration method is presented and its accuracy is evaluated. The quick feedback is a key issue in improving the athletic performance. We have developed two indicators, which are easy to visualize. The first one is the swimming speed measured from the video signal by tracking a marker band at the waist of the swimmer, another one is the rudimentary swimming cycle analysis focusing to the regularity of the cycle.

1 INTRODUCTION

This paper describes the swimming analysis system being developed at the Impivaara public swimming center in Turku, Finland. Starting a new site for swimming analysis requires usually considerable resources and our economical approach with 5-7k€ budget for hardware and software licenses should be of interest to any swimming coach considering a basic computerized real-time feedback at a local site.

Budget reasons forced us to use 3 cameras only and video coverage of 18 m. Another major constraint was to get the system up and running with no special initial procedure and without disturbing recreational swimmers. The system can be expanded in the future by a fourth camera at the grey dot depicted in Fig 1. This setup will cover the whole 25 m pool length.

We use the video image series of the light-reflective marker on the waist of the swimmer to record the movement of the swimmer. The marker moves along the tracking plane, which resides 200 mm aside towards the cameras from the centerline of the swimming lane. The distance has been chosen so that it approximates the dimensions of the pelvis of an average-sized adult male and female. The real movement of the marker is naturally a non-planar one, but the planar approximation is a useful first step to simplify the swimmer movement analysis. The tracking plane of swimming lane 7 is depicted in Fig. 1. The tracking plane has c. 1 m depth at the shallow end and c. 2 m depth in the deep end of the pool. All the calibration measurements were constrained on this plane, and the calibration result is a geometric mapping from the image pixels to global coordinates of the tracking plane. Two lanes with numbers 7 and 8 were calibrated. Lane 8 is used occasionally for swimming analysis purposes, but camera views do not cover the whole length of the tracking plane as seen in Fig. 4.

The camera positions are constrained to windowsills at the sides of the pool at the depth of 560 mm. The image mapping was constructed directly in relation to the tracking plane, and this method does not require usual camera model, camera locations and orientations. The stereo calibration method proved inferior because of very limited overlap between cones of visibility, see Fig. 4.

The design emphasizes the possibility to a fast feedback. Thus there are features which are designed to operate in real-time during the session of athletic performance. There are also some features which are

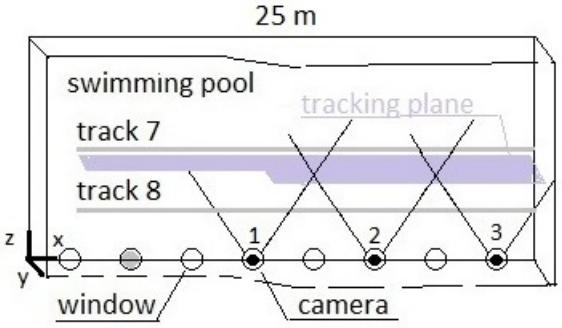

Figure 1: The general layout of the site seen from above. The tracking plane of lane 7 is emphasized.

based on the post-session phase. The forms of the feedback implemented are:

1. Real-time marker movement tracking embedded on the video.
2. Different performances presented side-by-side for visual comparison and evaluation. It can be real-time and post-session.
3. Stroke variation visualization, which is designed so that it can be monitored by the athlete in the pool. It can be real-time or post-session.
4. Geometric transformation of the video stream from pixels to global coordinates. It can be real-time or post-session.

The geometric mapping algorithm can project the raw video to a real-time 25 fps mono-color visualization on the tracking plane. The full quality color video has processing speed of c. 5 fps and cannot be performed in real-time. The geometric mapping will be a crucial part for a seamless swimmer-focused view after the swimmer detection has been implemented.

The marker tracking routine introduces both stochastic and algorithmic noise to the signal. After the pixel signal is transformed to global coordinates, the signal needs to be smoothed to eliminate the noise. The Kalman smoothing uses a basic dynamical model of a swimmer body. Smoothing requires the record of the whole performance as input and thus is a post-session step.

So far the coaching routine with verbal and video feedback has been established, but already the procedure is used on weekly basis and it requires no extra technical personnel on site.

The rest of the paper is organized as follows. Sec. 2 is a short presentation about the current research. Sec. 3 documents the architecture of the system, tracking and swimming cycle registration. Sec. 4 presents the used in-plane calibration method in detail, since this aspect is heavily dictated by the budget limitations yet opens possibilities for future research as well. Sec. 5 is about the real-time tracking visualization.

The swimmig cycle registration and comparison presented in Sec. 6 has been an important early facility for the coaching. The post-session analysis phase of Sec. 7 is one adaptation to the budget limitations. Sec. 8 summarizes the design choices made to achieve the real-time system response. Sec. 9 has conclusions and discussion about the possible future developments.

2 LITERATURE REVIEW

There are several approaches to swimming analysis. The oldest one is using wire. (Jean-Claude, 2003) reports about measuring the force in the wire while some object is dragged behind, another method is measuring the swimmer speed directly using the wire. The mechanical method is used especially to verify the video installments.

Video analysis is the dominant mode of performance analysis nowadays, see e.g. a review of the field in (Kirmizibayrak et al., 2011). A typical approach is:

- to produce the continuous video stream from multiple cameras
- and trace anchor points of the body (marked or nonmarked) and then
- combine the acquired information to a biomechanical or 3D visualization model.

There are several commercial tools available, many of them summarized in (Kirmizibayrak et al., 2011). Typical examples are Dartfish (Dartfish, 2015) and Sports Motion (Sportsmotion, 2015).

Wearable accelerometers are relatively new tools. These are developing smaller and lighter, also the operation time is increasing due the lower power requirements and increased battery capacity. (Dadashi et al., 2013) shows an arrangement with only one accelerometer to record the swimming performance over the full length of the pool. The data link is usually radio-linked in bursts like e.g. (James et al., 2011), or in the end of the performance, like in (Dadashi et al., 2013).

A typical large scale system design can be found from (Mullane et al., 2010). They provide an excellent analysis of what feedback should be provided real-time and what at post-session phase.

Swedish Center for Aquatic Studies has AIM (Athletes in Motion) system which can combine views from submerged and above-water cameras,

see (Haner et al., 2015). The calibration process resembles our approach albeit they use striped poles while we use chessboard pattern. AIM has been developed by efforts of Chalmers and Lund Universities.

Chalmers University has a multiple accelerometer arrangement, which is coupled with a video analysis and a biomechanical model. Accelerometers can be placed by suction cups to various areas of the body. One study shows how a relatively low frequency still provides adequate biomechanical modeling (Siirtola et al., 2011).

Another video technique is the virtual camera technique, where a moving viewpoint is synthetized between two adjacent cameras. It is possible to interpolate the view between stationary spots like in (Makoto et al., 2002).

Head is a popular choice of strapping the athlet with a sensor device. One can wear a colored cap, or swimming glasses with an accelerometer, see (Pansiot et al., 2010).

Relatively new approach is the video analysis without markers (Ceseracciu, 2011). From athlete point of view it is much less intrusive and enables automation of the analysis process.

The trend in research seems to be towards 3D visualization and increasing usage of biomechanical models. Actual analysis is quite developed and remaining goals are at quick performance feedback and well visualized and conceptually simple performance measures.

As a summary, existing systems are well-developed and serve the coaching activities well. Often the implementation is rather involved requiring technical assistance, set-up times and high initial and running costs. Our aim was to produce a cheap and simple non-intrusive alternative with stable basis for further improvement.

3 SYSTEM DESCRIPTION

The system consists of:

- one 2-core 3.2 GHz 64 bits computer with. 2 TB of disc space
- 3 permanently placed 50 fps cameras at the side wall of the 25 m pool. The maximum image size is 750×2044. The camera placement is dictated by the construction of the pool.
- one movable extra camera for above-water usage and one movable underwater camera. Usage of the extra cameras is just for visual observation and verbal feedback only.
- movement marker and band at the hip of the swimmer.

The cameras and computer record and store over 50 fps high resolution digital video in uncompressed format. The image size is 750×2044 pixels. An individual pixel of the geometrically transformed image corresponds to 4.0×4.0 mm^2 and 2.3×2.3 mm^2 on lanes 7 and 8, correspondingly.

The uncompressed data from three cameras amounts to about 1GB for a 10 second clip. All cameras are synchronized so that they capture images at the same time. The time stamps are stored in the video files and they can be used in determining how to stitch the tracking results.

Marker tracking algorithm utilizes OpenCV package (Bradski, 2000). Camera calibration was done with self-developed software.

Process is divided to real-time and post-session phases. Figure 2 illustrates the various steps of the process. The recorded video is stored in a raw uncompressed file format specific to the camera manufacturer and is later accessed by the post-session phase.

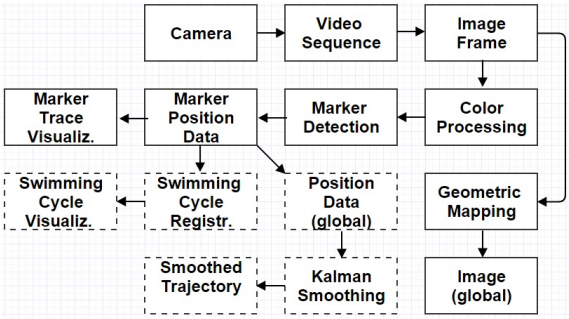

Figure 2: The processes and data flow. Post-session steps are indicated by the dashed outline.

4 CAMERA CALIBRATION

Camera calibration is a preliminary measurement process delivering either the camera model (mapping from pixels to normal vectors of the pinhole camera idealization) or the direct geometric mapping from pixels to global positions.

Three calibration methods, stereo-camera (Bouguet, 2008), mono-camera (Bouguet, 2008) and our own direct planar calibration were tested.

The stereo-calibration is an industry standard method since it is able to produce depth information (3D) and is not limited to the tracking plane G of Fig. 1. It calibrates the full camera view just like the

mono-camera method. It also provides an early quality check in the form of relative camera positions depicted in Fig. 3. The position error was c. 85 mm even after the best possible calibration measurements described in (Bouguet, 2008).

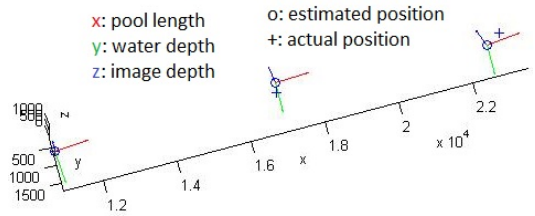

Figure 3: The camera positions from stereo-calibration.

Reason for the low accuracy is the difficult geometry of the camera positions dictated by the location of the windowsills of the pool, see Fig. 1. The amount of overlap of the camera views is only c. 22%, see Fig. 4 with refraction included, whereas the overlap ratio in a usual stereo calibration analysis is above 50%.

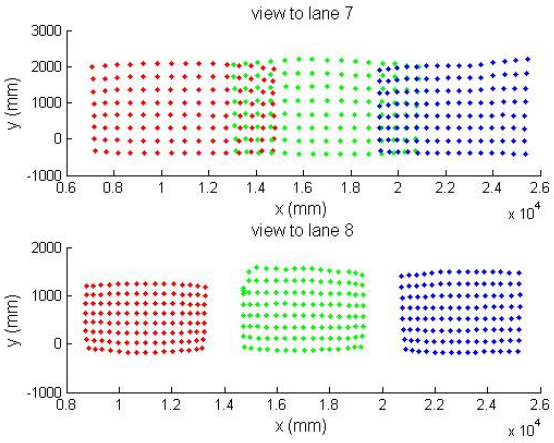

Figure 4: The area of visibility per each camera on the tracking plane. Colors (red/green/blue) correspond to cameras (1/2/3). Only c. 22% of the view is overlapping at lane 7. The lane 8 is closer to cameras, and there is no overlapping anymore.

The small overlap also rules out the homography approach described in (Chum et al., 2005) applied to all the sample points at the whole tracking plane at once. Otherwise that method would have been excellent, since it is able to use the expected location errors.

The mono-camera approach is very close to stereo-calibration, except the location and orientation of each camera is a separate subject of the matching process, when a fit is made to the tracking plane data. The mono-camera method is also close to the direct planar calibration presented in Sec. 4.1. The main difference is that the direct calibration requires no camera model (not even the camera location) and that the mapping from pixels to global locations can be arbitrarily chosen. The mono-camera calibration produces better mapping quality at the image borders than the direct planar calibration. The difference is aesthetic only, since the accurate zone of the direct calibration can be made large enough to accommodate all the swimmer motions. Also the mono-camera approach has been omitted from this presentation.

The stereo and mono calibrations were done with Matlab Camera Calibration Toolbox, see (Bouguet, 2008). The theory of the toolbox is given at (Zhang, 1999) and (Heikkilä and Silven, 1997).

The most accurate method was the direct planar calibration proposed in Sec. 4.1. This method can be categorized as an ad hoc approach answer to two constraints: sparsely placed camera array and potential for real-time video transformation. Nearest reference is (Luo et al., 2006), which uses a camera model and requires the co-planarity of the camera image plane and the tracking plane. Our method requires no camera model. The direct planar method is presented in the following.

4.1 Direct Plane Calibration

The geometric calibration was done for lanes 7 and 8 of ten available lanes. The calibration data for the direct method was gathered by floating a calibration chessboard along the surface at the tracking plane and recording its position at each picture, see Fig. 5. The chessboard had buoys at the top and weight at the bottom. The global position x_0 of the board was measured within 10 mm accuracy std.

To ease the presentation, some definitions are needed. An image $I = (P, J)$ of size $n \times m$ is a pair of set of pixels $p \in P = [1,n] \times [1,m] \subset \mathbb{Z}_+^2$ and their intensities $J = \{j(p) | p \in P\}$, where an individual pixel p has intensity $j(p)$. Images come in three varieties, source images $I_s = (P_s, J_s)$, geometrically transformed target images $I_t = (P_t, J_t)$, and calibration images.

The tracking plane $G = \{\mathbf{g} \in \mathbb{R}^3 \mid (\mathbf{g} - \mathbf{g}_G) \cdot \mathbf{n}_G = 0\}$ is defined by one insident point $\mathbf{g}_G = (0,0,z_0)^T$, where $z_0 = 5050$ mm in case of lane 7. The 200 mm aside of the center of the swimming lane. The normal vector $\mathbf{n}_G$ is a unit vector aligned with the global z axis. The marker is assumed to move along this tracking plane. Fig. 1 depicts the tracking plane G and the global coordinates x, y, z.

A chessboard corner pixel and its corresponding global positions form a measurement pair $(p, \mathbf{g}) \in U$. The calibration data set U is cumulated over all calibration images. One measurement image is depicted at the upper part of Fig. 5. The lower part shows the

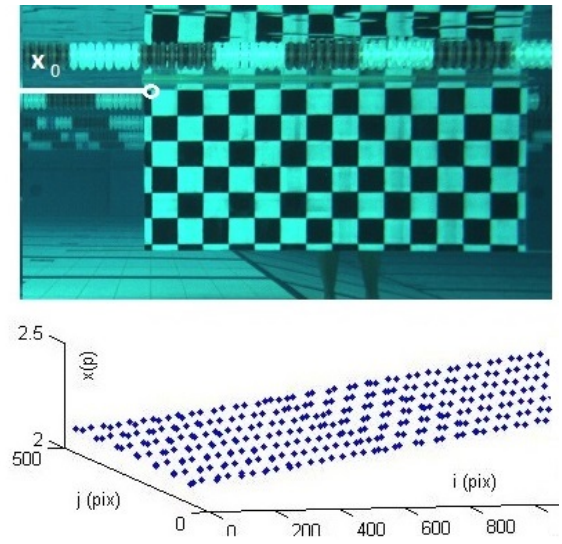

Figure 5: Direct calibration on the tracking plane. Above: an individual x_0 position of the calibration board at camera 2 view. Below: the cumulated observation set U of camera 1 consisting of corner point pixels p and global positions $\mathbf{g} \in \mathbb{R}^3$ from all images of lane 7 at camera 1. Only the x component of the global position $\mathbf{g}$ depicted.

measurement set U (only x component of $\mathbf{g}$ shown). The pixel samples of U cover only a part of the image pixels P whereas the end result of the direct plane calibration maps all pixels of the source image onto the tracking plane $P_s \rightarrow G$. In that sense this is an interpolation problem.

In the following, only the mapping of the x component is detailed. The y component has a similar treatment and is omitted for brevity. There would be some advantage to use a coupled mapping $p \rightarrow (x,y)$ e.g. a mono-camera model for interpolation but even this rudimentary approach with two independent mappings produced encouraging results.

A piecewise bilinear smoothing function:

$$x = f(p, \alpha_x^*) \tag{1}$$

with shape parameters $\alpha_x^* \in \mathbb{R}^d$ and automatic tiling heuristics is used to set a map $p \rightarrow x$ from pixels p to the global coordinate x. The shape parameter dimension $d \in \mathbb{N}_+$ varies case by case because of the implementation (D'Errico, 2006) but is in this application $d \approx 80$. The regularization parameter $\lambda_x \in \mathbb{R}_+$ controls the smoothness. The definition below uses functionals A_x for error penalty and B for non-smoothness penalty:

$$\begin{aligned} \alpha_x^* &= \underset{\alpha}{\operatorname{argmin}} A_x(f(.,\alpha)) + \lambda_x B(f(.,\alpha)) \\ A_x(f(.,\alpha)) &= \sum_{(p,\mathbf{g}) \in U} (f(p,\alpha) - x)^2 \\ B(f(.,\alpha)) &= \sum_{(p,\mathbf{g}) \in U} (\underset{q \in N_p}{\operatorname{mean}} f(q,\alpha) - f(p,\alpha))^2 \end{aligned} \tag{2}$$

In the above, N_p is the set of neighboring pixels of p at pixel radius $r = 2$. The function $f(.,.)$ is implemented as Matlab *gridfit.m* with 'bilinear' and 'laplacian' options, see (D'Errico, 2006). Values $\lambda_x = 120$, $\lambda_y = 180$ were chosen to keep the non-smoothness measure $B(.)/A(.)$ tolerable.

4.2 Precomputed Mapping

Let us combine Eqs. 2 and 1 for further treatment. The source image pixels p_s are mapped to tracking plane G by:

$$F_s(p_s) = (f(p_s, \alpha_x^*), f(p_s, \alpha_y^*), z_0) \in G, p_s \in P_s \tag{3}$$

The image of I_s on G is now $F_s(I_s)$. The target image I_t is mapped to tracking plane G by:

$$F_t(p_t) = \mathbf{g}_0 + \gamma \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} p_t^T \in G, p_t \in P_t \tag{4}$$

The target image pixel size $\gamma = 4.0$ mm for lane 7 and $\gamma = 2.3$ mm for lane 8. Image location $\mathbf{g}_0$ is specific for each camera view on each swimming line. Pixel $p = (i, j)$ has row and column indices as depicted in Fig. 6. It is now possible to interpolate the intensity value of p_t using Shepard interpolation of Eq. 6 at the tracking plane G.

First, some definitions. Let $|P_t|$ be the number of pixels in the image I_t and $M \in \mathbb{N}_+^{|P_t| \times 4}$ be a reference pixel matrix. One row $M_{p_t.}$ specified in Eq. 5, holds 4 reference pixels from P_s for a pixel $p_t \in P_t$. The nearest neighbor operator $\mathrm{NN}_k(x, X)$ selects k nearest neighbors of x from a set X. Sets are treated as vectors whenever there is a unique enumeration of the set. The location of p_t is $\mathbf{g}_t$. N_t is the set of 4 nearest neighbors of the sourcce image locations ar G.

$$\begin{aligned} \mathbf{g}_t &= F_t(p_t) \\ N_t &= \mathrm{NN}_k(\mathbf{g}_t, F_s(P_s)) \subset G, k = 4 \\ M_{p_t i} &= F_s^{-1}(N_{ti}) \subset P_s, i = 1...4 \end{aligned} \tag{5}$$

$$\begin{aligned} W_{p_t i} &= \{s(\|\mathbf{g}_i - \mathbf{g}_t)\|) \mid \mathbf{g}_i \in N_t\}^0, i = 1..4 \\ s(r) &= 1/\max(r, 0.05\ \mathrm{mm}) \end{aligned} \tag{6}$$

where $W \in \mathbb{R}_+^{|P_t| \times 4}$ is a radial interpolation weight matrix with a correspondence to same indexing as matrix M. Function $s(r)$ is the radial weight used. The

normalization in Eq. 6 happens by L_1 norm: $\mathbf{w}^0 = \mathbf{w}/\sum_i |w_i|$ for a general weight row $\mathbf{w}$.

Considering pixel intensities J_t and J_s as vectors indexed by pixels, the transformation becomes:

$$J_t(p_t) := \sum_{i=1}^{k} J_s(M_{p_t i}) W_{p_t i} \quad (7)$$

By selecting $k = 1$ one gets the real-time transformation:

$$J_t(p_t) := J_s(M_{p_t 1}) \quad (8)$$

Eq. 8 corresponds to the Nearest Neighbor referencing which can be computed at 25 fps (measured with Matlab). The case $k = 4$ of Eq. 7 can be computed at 5 fps and thus is not usable in real-time. The quality of $k = 1$ case is adequate to a video stream, see Fig. 6. If quality par the original is required, one can use Eq. 7. The balance between speed and quality can be tuned further by choosing $k = 2$ or $k = 3$.

The geometric image mapping is efficient and simple, see Eqs. 8 and 7. The formulation used also makes it possible to combine the three separate video signals accurately to one single video. This feature will be implemented when an automated swimmer targeting is added to the system.

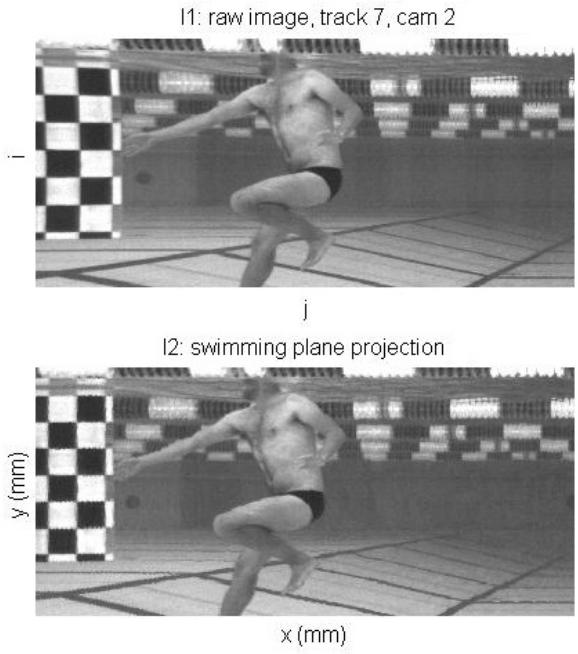

Figure 6: Quality of the fast mapping, a detail at the opposite wall. Above: the source image I_s with pixel coordinates i and j. Below: the target image I_t with global coordinates x and y.

4.3 Error analysis

The measurement points U in Fig. 5 are in approximate horizontal rows. There is c. 150 mm vertical gap between rows and c. 50 mm average horizontal distance between points. This requires the interpolant to have rather high penalty for non-smoothness.

The pixel detection was done with Matlab *detectCheckerBoard.m* function, theory of which is contained in (Zhang, 2000). The pixel detection error is $p \approx (1,1)$. The mechanical placement accuracy of the measured points $(p,\mathbf{g}) \in U$ is $\Delta\mathbf{g} \approx (10,10,10+0.01z)^T$ mm as an approximate std. The error is unbiased and the final accuracy of $F_s(p_s)$ is much better.

A pixel-wise geometric mapping error measure $e(p)$ is formulated by Eq. 9 and depicted at Fig. 7:

$$e(p) = \|\mathbf{g} - F_s(p)\|, \ (p,\mathbf{g}) \in U \quad (9)$$

Since the sample set U is of rather good quality and since the function F_s is rather smooth, the error stays almost constant even if the tuning of the shape parameters α_x, α_y in Eq. 2 is subjected to cross-validation over subsets of U. The error is largest in occasional points at the border and grows rapidly when extrapolating. The border areas are seldom occupied by a swimmer, though, and the problem is more of aesthetical nature. The border error can be eliminated in the future by applying a different interpolant instead of one in Eq. 3.

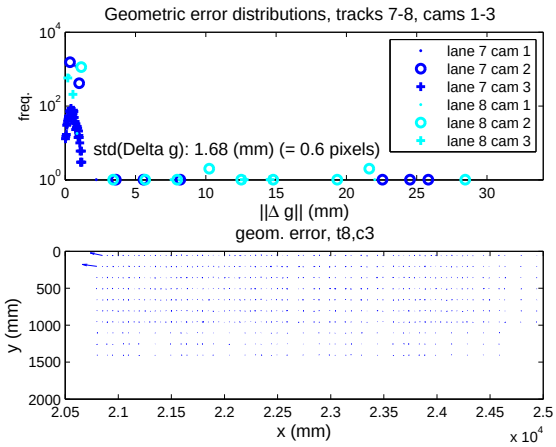

Figure 7: The geometric mapping error $e(p)$, $p \in P_s$ defined by Eq. 9. Large errors happen occasionally at the borders of the sampled area U.

5 REAL-TIME TRACKING

The swimmer tracking method is based on marker tracking using the blob tracking facilities of OpenCV

(Bradski, 2000). The marker is placed on a colored flexible band worn by the athlete in order to improve accuracy and reduce noise. The band is selected and installed so that it will not hinder the performance of the swimmer. The yellow color of the band is chosen to increase its visibility and identifiability in the environment. The band can be occluded by the swimmer's hand, or by bubbles in the water.

The color of the band is selected based on the color hue distribution of the environment. Tests on the site showed that the environment colors (water, swimmer, light, walls, etc) are between the interval $90^\circ - 270^\circ$, leaving the rest of the hue circle open. We selected a yellow color band. The choice leaves room for one or two extra markers, if needed in some future analysis.

The pixel trajectory of the marker is being visualized for the user in real-time, see Fig. 8. The pixel trace is then being mapped to global coordinates on the tracking plane using the geometric mapping described in Sec. 4. Tracking performs well with the current 50 fps speed allowing the real-time rendering.

The visualized points are referring to image pixel coordinates and are not suitable for speed analysis. The swimmer speed is calculated by converting these points using the geometric mapping described in Sec. 4.

The real-time swimmer trace is provided as an overlay curve at the video area, see Figure 8. The current frame position is highlighted. The recorded session can be replayed immediately.

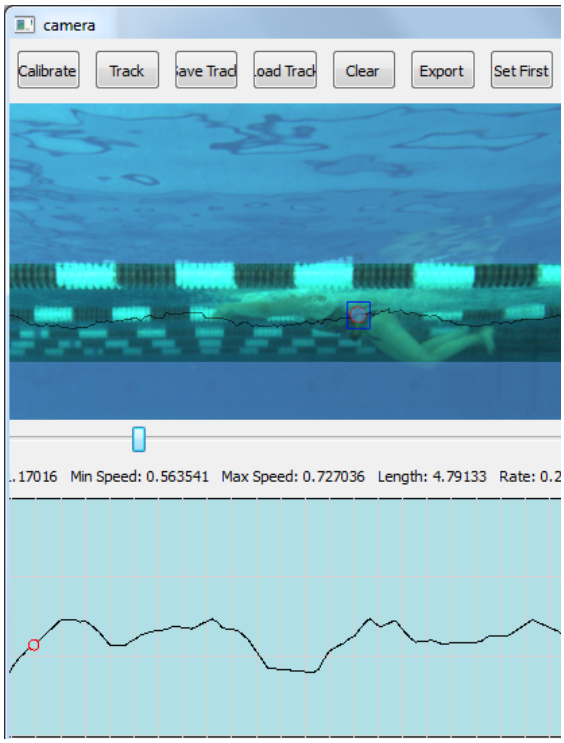

Figure 8: Presenting the tracking results to the user. The emphasised square is for the user only and it does not correspond to the tracking plane. The track is based on pixel information for reducing response time.

6 SWIMMING CYCLE REGISTRATION

A swimming cycle is the state history over a time interval during which the swimmer state returns relatively close to the initial state. The maximum performance requires rather monotonic strokes, yet the rhythm may vary based on metabolical optimum. Detecting the regularity of swimming strokes is of importance. Cycles seem to have a distinctive deceleration phase just before each hand stroke. This enables a simple cycle registration by finding a local spike heuristically in phase signal at times $t_i \in \mathbf{R}_+, i \in \mathbb{N}_+$. Using a peace-wise linear parameter τ:

$$t(\tau) = t_i(\tau - i) + t_{i+1}(i+1-\tau),\ i = \lfloor \tau \rfloor,\ \tau \in \mathbb{R}_+ \quad (10)$$

one can compare the shapes of two cycles i and j directly in a duration-invariant way on their own relative time scale $t_i + \tau$. Let us define the duration of a cycle i as $T_i = t_{i+1} - t_i$. Now, the dissimilarity d_{ij} between two cycles i and j can be defined:

$$d_{ij} = [\int_0^1 (v_x(t(i+\tau)) - v_x(t(j+\tau)))^2 d\tau]^{1/2} + \\ + \lambda |T_i - T_j| \quad (11)$$

The horizontal velocity v_x in Eq. 11 is based on pixel information with a moving average smoothing, since we noticed the raw pixel signal is enough for the cycle detection. The last summand of Eq. 11 sets a weight on the duration difference between two strokes. The duration difference penalty is open to experimentation, currently we use value: $\lambda = 4$. We have used the vertical velocity $v_x(t)$ as the target signal for similarity analysis. The target signal could be also the horizontal velocity or a vector combination of both, in which case a vector norm should be used in Eq. 11.

The swimming cycle registration is a post-session process, which will be implemented as a real-time feature in the future. A similarity matrix is cumulated from last few strokes (last 5 in cases depicted in Fig. 9). The visualization is designed to be seen directly from the pool. The colors are scaled so that black is a serious deviation from allowed, white means identical strokes. Each stroke is compared to others and no judgement is made towards the quality

of the swimming performance in general. The gray-scale used is an arbitrary choice at the moment.

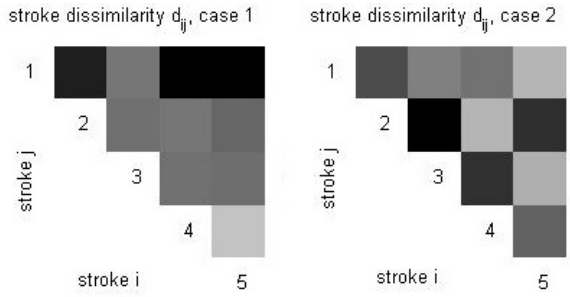

Figure 9: Swimming stroke regularity visualization. Rows and columns are individual strokes. White means zero difference and black a difference $d_{ij} = 0.2\ m/s$.

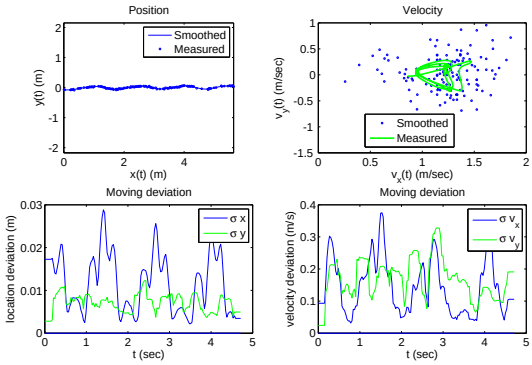

Figure 10: The smoothed position and velocity of the marker tracking. Above: Measured and smoothed signals. The current tracking brings large procedural noise component to velocity. Below: The moving deviation estimates. Only one camera view was used in this demonstration.

7 POST-SESSION PHASE

The post-session phase occurs when there is a pause in the athletic performance. First, the recorded video is stored on the hard disc. The trainer opens the video file with the tracking software and it is able to provide feedback to the coach and swimmer in a reasonable time (at most a few minutes).

The software allows the overlaying of multiple tracking results of different athletes. The trainer can use these data overlays to compare a trainee with a reference (trained) swimmer performance.

The speed graphs acquired by geometric transform of the original pixel trace are displayed in a separate area of the screen under the video frame. The graphs span the whole observed length.

A number of quantitative measures are displayed on the current swimmer performance, like average speed, distance, time, minimum and maximum speeds. A specific time period can be highlighted in the speed graph, to restrict the numeric display to measurements on this area.

There will be further experiments on visualizing various swimming characteristics. Preference will be given to the real-time feedback.

7.1 Kalman Smoothing

Kalman smoothing (J. Hartikainen and and Särkkä, 2011), is applied to pixel trace to get smoothed plots of the position and velocity components over time. Fig. 10 depicts the smoothed position and velocity history. Further swimming style analysis and movement registration operations can be based on this signal.

The marker observations are described as coming from a linear dynamical system of Eq. 12 with Gaussian noise $\mathbf{w} \sim N(\mathbf{0}, \lceil \sigma_x^2, \sigma_y^2 \rfloor)$ as the driving force component. Our numerical choice was: $\sigma_x = 10$ N, $\sigma_y = 20$ N. Other constants of Eq. 12 are specified in Eq. 13:

$$\mathbf{m}\,\ddot{\mathbf{g}}(t) + \mathbf{c}\,\dot{\mathbf{g}}(t) + \mathbf{k}\,\mathbf{g} = \mathbf{w}(t) \qquad (12)$$

The system is further discretized to given non-regular observation times and transformed to a discrete-time linear dynamical model with Gaussian noise term, see details from (J. Hartikainen and and Särkkä, 2011).

The numerical values of the swimmer model of Eq. 12 are chosen for an average swimmer, and the dampening parameters approximate the observed speed resistance of swimming. The model will be improved later e.g. using Gaussian process formulation instead of Kalman, adding more biomechanical authentity to the model and physically interesting latent forces, see e.g. (Hartikainen et al., 2012):

$$\begin{aligned} \mathbf{m} &= \mathbf{I}_2 \times 60\ kg && \text{(mass)} \\ \mathbf{c} &= \lceil 11.7\ 16.3 \rfloor\ kg/s && \text{(dampening)} \\ \mathbf{k} &= \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \times 0.05\ kg/s^2 && \text{(spring)} \end{aligned} \qquad (13)$$

The average error (std.) of the procedure is at the moment $\Delta x = 0.018$ m, $\Delta y = 0.008$ m, $\Delta v_x = 0.2$ m/s, $\Delta v_y = 0.2$ m/s for positions and velocities along x and y axes. Improvements will be made by applying a better tracking method less sensitive to bubbles.

Kalman smoothing is a post-processing task, too.

8 SYSTEM PERFORMANCE

A principal objective of the system is to provide immediate trainer feedback. To achieve that, the implementation is based on the following principles:

- The frames are processed and tracked in their uncalibrated shape (containing all camera distortions, refraction etc.). The physical meaning for the tracking signal can be attached only after the geometric mapping, which is a postprocessing step, see Fig. 2.
- The tracked area is limited manually, see the highlighted area in Fig. 8. Usually swimmers occupy only a narrow band on the screen. Limiting the tracking to this area of interest reduces computation significantly. At the moment, the tracking box (see Fig. 8) is selected manually, but there will be few prerecorded tracking areas for different swimming styles in the future.
- The dissimilarity measure of Eq. 11 is also based on the raw pixel information. The formula is computationally cheap, and it has to be evaluated on a separate processor once when swimmer passes a camera. Full real-time indicator will be implemented when a second computer and a monitor will be added to the system.
- The geometric mapping of video images uses reduced quality to deliver real-time performance.
- Seamless (combined from 3 cameras) geometrically accurate visualizations like video, stroke regularity indication and physical speed analysis are all left as a post-processing step. At this phase, the geometric mapping has been done to images and data already.

9 CONCLUSIONS

We have presented a simple video-based swimming analysis system which is easy to install, is of low cost and is simple to calibrate without any technical assistance. It can be installed to a wide variety of pool types. It is maintenance free and based on our experience so far, it can be operated by one person only. In ordinary use no technical assistance is needed.

The proposed system provides swimming speed analysis and instant visual feedback. The system is a good basis for further expansion e.g. with swimming gait analysis, biomechanical modeling etc.

The current system can be easily upgraded by a fourth camera at the location indicated by a grey circle in Fig. 1. The video monitoring would then span whole the pool length. A second video screen will be added in the future to serve the athletes better.

There are many off-the-shelf analysis systems with a wide spectrum of functionality available today. Usually these systems are much more complex and expensive than one presented here. Our choice was to implement the real-time pixel trace of the marker and swimming cycle regularity visualization.

The tracking system needs to be improved in the near future. At the moment it falls off-the-track too often, especially when a hand moment occludes the already lost marker.

The current system has been used by Finnish national swimming teams both on senior and junior level since autumn 2014. Automated tracking has made it possible to give faster and more accurate feedback to athletes. Thus it has been possible to test a large number of athletes in relatively short time during national team camps, when previously only a few of the top swimmers were able to get the service due to time investment required using the older version of the system. According to national team coach the system has been a major asset in developing technical skills of national team athletes. The findings have also been used in national coaches' education to provide insight into swimming performance.

The proposed direct planar calibration method used is aimed for efficient real-time video stream transformation. The efficiency is possible due to the restriction to 2D tracking plane projection only. There is potential for the same formulation to be generalized for 3D motion capture at the overlapping view zones (2×2 m at the current system, 3×2 m after one camera will be added). The proposed calibration method may be of use in other applications where conditions in camera placement rule the stereo-calibration out and where planar observations suffice.

The most important future goals are a reliable markerless tracking and implementing a record database with automated input from the site and a support for rudimentary searches and comparisons.

The swimming gait registration based on the profile shape of the body of the swimmer is a potential development.

Automated detection of different phases of the swimming performance remains the last goal. It is the hardest since there are a lot of different swimming styles each with somewhat differing phases, and female and male swimming costumes differ.

Acknowledgements

The project is a joint venture of University of Turku IT department and Sports Academy of Turku region and it has been funded by city of Turku, National Olympic Committee, Finnish Swimming Federation, Urheiluopistosäätiö and University of Turku.

REFERENCES

Bouguet, J. Y. (2008). Camera calibration toolbox for matlab. 3, 4

Bradski, G. (2000). Opencv. *Dr. Dobb's Journal of Software Tools*. 3, 7

Ceseracciu (2011). *New frontiers of markerless motion capture: application to swim biomechanics and gait analysis*. PhD thesis, Padova University. 3

Chum, O., Pajdla, T., and Sturm, P. (2005). The geometric error for homographies. *Comput. Vis. Image Underst.*, 97(1):86–102. 4

Dadashi, F., Millet, G., and Aminian, K. (2013). Inertial measurement unit and biomechanical analysis of swimming: an update. *Sportmedizin*, 61:21–26. 2

Dartfish (2011-2015). Dartfish video analysis tool. http://www.sportmanitoba.ca/page.php?id=116. 2

D'Errico, J. (2006). Surface fitting using gridfit. Technical report, MATLAB Central File Exchange. 5

Haner, S., Svärm, L., Ask, E., and Heyden, A. (2015). Joint under and over water calibration of a swimmer tracking system. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, pages 142–149. ScitePress. 3

Hartikainen, J., Seppänen, M., and Särkkä, S. (2012). State-space inference for non-linear latent force models with application to satellite orbit prediction. *CoRR*. 8

Heikkilä, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1106–1112. 4

J. Hartikainen and, A. S. and Särkkä, S. (2011). Optimal filtering with kalman filters and smoothers, a manual for the matlab toolbox ekf/ukf. Technical report, Dept. of Biomedical Eng. and Comp.Sci., Aalto University School of Science. 8

James, D. A., Burkett, B., and Thiel, D. V. (2011). An unobtrusive swimming monitoring system for recreational and elite performance monitoring. In *Procedia Engineering, 5th Asia-Pacific Congress on Sports Technology (APCST)*, volume 13, pages 113–119. 2

Jean-Claude, C., editor (2003). *Biomechanics and Medicine in Swimming IX*. IXth International World Symposium on Biomechanics and Medicine in Swimming, Université de Saint-Etienne. 2

Kirmizibayrak, J., Honorio, J., Xiaolong, J., Mark, R., and Hahn, J. K. (2011). Digital analysis and visualization of swimming motion. *The International Journal of Virtual Reality*, 10(3):9–16. 2

Luo, H., Zhu, L., and Ding, H. (2006). Camera calibration with coplanar calibration board near parallel to the imaging plane. *Sensors and Actuators A: Physical*, 132:480486. 4

Makoto, H. S., Kimura, M., Yaguchi, S., and Inamoto, N. (2002). View interpolation of multiple cameras based on projective geometry. In *In: International Workshop on Pattern Recognition and Understanding for Visual Information*. 3

Mullane, S. L., Justham, L. M., West, A. A., and Conway, P. P. (2010). Design of an end-user centric information interface from data-rich. In *Procedia Engineering*, volume 2, pages 2713–2719. 8th Conference of the International Sports Engineering Association (ISEA). 2

Pansiot, J., Lo, B., and Guang-Zhong, Y. (2010). Swimming stroke kinematic analysis with bsn. In *Body Sensor Networks (BSN), 2010 International Conference on*, pages 153–158. 3

Siirtola, P., Laurinen, P., Roning, J., and Kinnunen, H. (2011). Efficient accelerometer-based swimming exercise tracking. In *IEEE Symp. on Computational Intelligence and Data Mining (CIDM)*, pages 156–161. IEEE. 3

Sportsmotion (2011-2015). motion analysis system. http://www.sportsmotion.com/. 2

Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. In *in ICCV*, pages 666–673. 4

Zhang, Z. (2000). A flexible new technique for camera calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22, page 13301334. 6