Improving data acquisition speed and accuracy in sport using neural networks

Christopher Papic[a], Ross H Sanders[a], Roozbeh Naemi[b], Marc Elipot[c,d], Jordan Andersen[a]

[a]Exercise and Sport Science, Faculty of Medicine and Health, The University of Sydney, Australia

[b]School of Life Science and Education, Staffordshire University, United Kingdom

[c]Swimming Australia, Australian Institute of Sport Office, Canberra, Australia

[d]Australian Analysis and Research Group for Optimisation in Swimming, Australia

Corresponding author

Christopher Papic

Email: chris.papic@sydney.edu.au

Twitter: @chris_papic

Abstract

Video analysis is used in sport to derive kinematic variables of interest but often relies on time-consuming tracking operations. The purpose of this study was to determine speed, accuracy and reliability of 2D body landmark digitisation by a neural network (NN), compared with manual digitisation, for the glide phase in swimming. Glide variables including glide factor; instantaneous hip angles, trunk inclines and horizontal velocities were selected as they influence performance and are susceptible to digitisation propagation error. The NN was 'trained' on 400 frames of 2D glide video from a sample of eight elite swimmers. Four glide trials of another swimmer were used to test agreement between the NN and a manual operator for body marker position data of the knee, hip and shoulder, and the effect of digitisation on glide variables. The NN digitised body landmarks 233 times faster than the manual operator, with digitising root-mean-square-error of ~4-5mm. High accuracy and reliability was found between body position and glide variable data between the two methods with relative error ≤5.4% and correlation coefficients >0.95 for all variables. NNs could be applied to greatly reduce the time of kinematic analysis in sports and facilitate rapid feedback of performance measures.

Keywords

Swimming, digitisation, video analysis, performance analysis, applied biomechanics

Introduction

Video footage is commonly used to analyse human movement and performance in training and simulated competitive sporting environments. Kinematic analysis of video involves the identification of body landmark positions (e.g. joint centres) through the process of 'digitisation' to obtain pixel coordinates for each frame of video data. These coordinates are converted to real world metric positions and are used to derive kinematic variables of interest. Digitisation of video data in sport is commonly performed manually, where an operator estimates the position of joint centres without the need for external markers on the athlete. Manual digitisation, however, is not conducive to time-efficient performance analysis and feedback, nor for analysing large datasets, due to its laborious nature.[1] Video analysis software with image recognition algorithms can automate digitisation of body landmarks in video; however, some systems require manual intervention to improve digitisation accuracy to an acceptable level,[2] limiting the amount of time saved. In contrast, accuracy is sometimes sacrificed to increase processing speeds; for example, calculations of 2D knee angle during a drop jump from body landmark data, digitised by automatic digitising software, produced a considerable range of error (0.21-37.93%) compared to 'a gold-standard' optoelectric motion capture system.[3] There seems to be a trade-off between accuracy and processing time with video analysis software, leaving users with the decision of what to sacrifice.

Neural networks (NNs) are proven to be highly accurate and time-efficient for image recognition tasks when sufficiently trained on a large dataset.[4] For example, 10,500 images of subjects performing lifting tasks were used to train a NN to automatically digitise multiple 3D joint positions, based on annotated body landmark position data derived from an optoelectric motion capture system.[5] Mean 3D landmark position error between the NN and the motion capture system was 14.72 ± 2.96mm, highlighting the potential for automatic digitisation of video data using NNs. The NN design, however, was limited by the requirement of a motion capture laboratory to train the NN. Through a process called 'transfer learning', image recognition abilities of an existing NN are used to develop a new NN to recognise features in images, such as body landmarks, that the initial NN has not digitised previously. The advantage of this approach is that standard video analysis and manual digitisation procedures can be used to train a NN, which may be more viable for sport scientists working with athletes in training and simulated competitive environments. For instance, the NN software DeepLabCut™ utilises transfer learning and an image feature detection algorithm[4,6,7] to 'learn' user-defined features in a relatively small number of training images (<500) and digitise similar features in new videos.

NNs may be particularly advantageous for kinematic analysis in aquatic environments, which poses added methodological challenges. Manual digitisation in swimming research, for example, is necessary to minimise body landmark position error and missed landmarks by automatic methods since the identification of markers can be affected by turbulence, air bubbles, and vortices that can obscure the markers.[8] Cronin et al.[9] demonstrated that a NN could be used to digitise 2D joint positions during underwater running with comparable accuracy to a manual operator. NNs could provide a faster alternative to manual digitisation of body landmarks in aquatic video data.

The use of video analysis in swimming is practical for movement and performance analysis because swimmers' motion can be captured without manipulating technique.[10] Video analysis is often used to analyse the glide component of the underwater phase of start and turns because start time and overall swimming performance are highly dependent on the glide.[11,12] Glide performance is influenced by the swimmer's ability to minimise hydrodynamic resistance and deceleration during the glide (e.g. glide efficiency) and to maintain posture during the glide (e.g. hip angle[13] and trunk incline[14]). Given the glide remains predominantly in the sagittal plane, digitisation of body landmarks in 2D video can be used to derive glide efficiency, posture, and performance outcome measures. Deriving these measures from 2D position data, however, can amplify the magnitude of digitisation error, evidenced when calculating the first derivative of position data.[15] While markerless 2D joint position error between manual and NN digitisation methods in an aquatic setting may be acceptable,[9] the effect of digitisation error on kinematic outcomes of the glide, such as velocity and glide efficiency, requires further investigation. Athletes and coaches would benefit from an accurate and time-efficient method for glide analysis.

The emerging use of NNs for image feature detection may be applicable to kinematic analysis in sport to improve data acquisition speed and accuracy. The purpose of this study was to train a NN to digitise body landmarks in 2D video of athletes in a sporting environment and to compare the time, accuracy, and reliability of digitisation and derived kinematic variables by the NN with manual digitisation.

Method

Participants

Five male (age: 21.6±2.1years, height: 187.72±7.61cm, mass: 85.68±2.80kg, FINA score: 677±53.9) and four female (age: 20.3±2.1years, height: 172.03±6.42cm, mass: 68.98±8.61kg, FINA score: 723.5±85.7) state and national level swimmers from an Australian swimming club were recruited. FINA point scores were calculated for the swimmers' 100m long course best time of their preferred stroke within the previous 12 months. The swimmers were informed via a printed participant information statement and gave their free written consent to take part in the study.

Procedures

The testing procedures were conducted for a subsequent study to evaluate the effects of verbal cuing on glide performance. Data collection was conducted in a ten-lane 25m pool (3m depth). Swimming training attire was worn to expose the greater trochanter for body marking: briefs for males and one-piece swimsuit for females. Height and body mass were taken using a stadiometer and electronic weight scale (WS207PMSG, Wedderburn, Australia). Body landmarks were marked using black 'ProAiir Hybrid' waterproof body paint (Face Paint Shop Australia, Yamba) with 4cm diameter circles.[16] The following body landmarks were marked on the lateral aspect of the swimmers' right side: knee joint axis, hip over the greater trochanter, and shoulder over the glenohumeral joint at C7 height. The landmarks were identified by an Accredited Exercise Physiologist (Exercise & Sports Science Australia) while the swimmer adopted a streamlined position standing on the pool deck.

Swimmers performed underwater glides from the wall in the streamlined body position without upper or lower limb actions; where the arms were extended forward above the head, the hands pronated and overlapping, and the feet plantarflexed and positioned together.[17] Swimmers attempted glides until they achieved ten successful trials. A glide was deemed successful when the swimmer maintained a horizontal body position and trajectory without lateral deviation from the black lane line, which was assessed visually by two researchers.

Data acquisition

A visual representation of the experimental setup is illustrated in Figure 1. A SwimPro X underwater camera system (SwimPro RJB Engineering, Australia) captured the swimmers' glides as they pushed off the start wall. The underwater camera was located 3.5m from the start wall in the lane closest to the side of the pool at a depth of 1.0m, such that the camera was positioned 6.25m perpendicular to the direction of the swimmers' motion. The camera was fixed with a wall mount and recorded video at 30Hz and capture resolution of 1920x1080 pixels. Video data were transmitted wirelessly from the camera to a computer located on the pool deck via an antenna connected to the underwater camera by a waterproof cable. The SwimPro software (SwimPro RJB Engineering, Australia) displayed the recordings in real time and saved each glide in mp4 format. Glide trials were captured with the swimmer moving from left to right of the capture screen, with the knee, hip, and shoulder landmarks on the right side of the body visible for kinematic analysis.

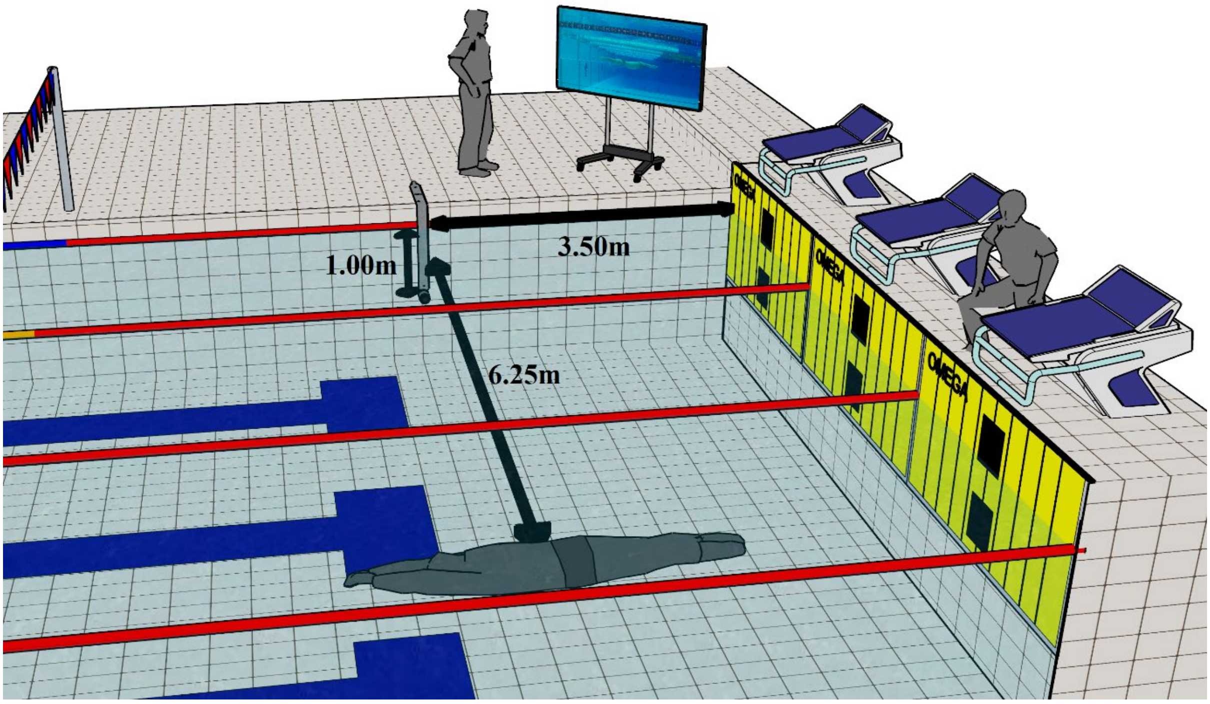

Figure 1. Experimental setup for 2D glide analysis

Data analysis

Video processing

The 'Cinalysis' software[18], was used to process the videos. Fisheye distortions were removed using checkerboard calibration (9x7, 29mm squares) as defined by Bouguet.[19] The camera lens was modelled using three coefficients to represent radial distortions and two to represent the tangential distortions, derived from the extracted corner points and known size of the checkerboard pattern.[20] Each glide trial was then trimmed and exported as 105-frame corrected glide trials: 45-frames to analyse the glide with 30-frames buffer before and after. The first frame of the glide to be analysed was when the swimmer achieved the streamlined position after leaving the wall. A calibration plane (4.98x1.00m) containing 40 calibration points, covering the entire underwater zone of interest, was used to compute the calibration coefficients applying a 2D direct linear transformation method.[21] The calibration error was assessed as the reprojection error, defined by Kwon and Casebolt,[22] where root-mean-square error (RMSE) of the reconstructed calibration marker positions were 4.7mm and 4.9mm for the x- and y-axis coordinates, respectively.

Manual digitisation

Four glides from a single swimmer were used to assess the accuracy of digitisation by the NN against manual digitisation. The four glide trials consisted of 420-frames of video data, with 1260 available body landmarks (knee, hip, and shoulder). Manual digitisation of these four glide trials was completed five times by the first author using the graphical user interface within the DeepLabCut™ software. Digitisation was performed across multiple days and the same glide trial was never re-digitised on the same day to ensure reliability was not affected by practice.[8] X- and y-pixel coordinates of the five repeated manual digitisations were averaged for each landmark in each frame of data in the four trials. The coordinates were averaged to define the most likely manually derived position for a given landmark. These data were used to evaluate the accuracy and reliability of digitisation by the NN against the manual operator.

Neural network training and digitisation

DeepLabCut™ (v2.1) was used to train a NN to digitise the knee, hip, and shoulder. Four hundred frames (following recommendations from Cronin et al[9]) were randomly extracted, using the *k*-means algorithm in DeepLabCut™, from glides performed by eight participants to train the NN. The four glide trials from the remaining participant (i.e. the trials used to assess the accuracy of the NN against manual digitisation) were excluded from the training process. The remaining six glides from this participant were set aside and digitised by the NN as part of the complete data set, as described below. The last author manually digitised the three landmarks in all 400 training frames. Inter-rater reliability between the first and last author was tested using a separate database of glide videos (see *Manual digitisation reliability*).

Image feature learning by a NN involves calculating the probability, known as a 'weight', that there is a match between the red-green-blue (RGB) characteristics for a region of an image, known as the 'input', and the RGB characteristics of the region surrounding a body landmark, referred to as the user-defined 'ground truth'. With transfer learning, training time is significantly reduced since a set of weights previously trained to identify RGB characteristics in a very large image database are used as a starting point for a new NN. Training by transfer learning involves updating the pre-trained weights by comparing the input with the ground truth for new images.

Initial weights pre-trained on ImageNet[23] served as a starting point to train the NN for 200,000 iterations using the ResNet-50 architecture in DeepLabCut™.[4] A 0.95 training fraction was used for the train/test ratio, meaning 95% of the 400 training frames were used to train the NN and 5% were used to assess the network's accuracy in estimating pixel coordinates of the body landmarks. The mean test error (that is, the output of the 'loss function') was calculated as the average difference between the pixel coordinates from manual digitisation (i.e. the ground-truth) and the NN's estimations.

The NN was trained in Google Colaboratory on a virtual 13Gb Tesla P100 GPU (CUDA v10.1). The weights were saved to a basic local machine containing a 7th Gen Intel Dual Core i5-7300 CPU (2.6GHz) with 8Gb of memory. Glide videos (n=90) from all participants were then processed on the local machine in DeepLabCut™ using the trained NN to digitise the body landmarks. The NN software output estimations for the raw x- and y-pixel coordinate of each body landmark and the probability of these estimations for every frame. The probability that a body landmark exists at a given pixel was calculated for each pixel on what is called a 'score-map'.[24] A score-map was generated for every landmark in each image of a video during processing. The location of each body landmark was determined as the pixel with the maximum probability on the score-map for that image.[4]

Glide data analysis

Kinematic data were calculated using coordinate data digitised by the manual operator and the NN from the four glide videos excluded from the training process. It is critical to note that the NN had never "seen" these images and therefore the robustness of the NN in this test setting could be evaluated. Figure 2 summarises the glide data processing stages following manual and NN digitisation of the four trials. After digitisation, raw pixel coordinate data were transformed into position data (mm) using the calibration coefficients described in the *Video processing* section. A cubic spline filter was used to interpolate missing data points, producing filled position data.

Glide efficiency is the ability of the swimmer to minimise deceleration during the glide and is reflected in a 'glide factor' obtained by curve-fitting 2D position data of body landmarks with a function based specifically on hydrodynamic principles.[25] Glide factor (m) was calculated using the hydro-kinematic method[25] in MATLAB for the 45 glide frames in each of the four glide trials. Filled position data were used to calculate glide factor to avoid over filtering. The mean position of the knee, hip, and shoulder for each frame were used to calculate glide factor due to better accuracy than using a single body landmark.[25,26] Logarithmic fitting was done by solving the differential equation of horizontal glide motion, where x is the x-axis instantaneous filled position data, C_G is glide factor, and V_{xo} is the initial velocity (Equation 1). C_G was solved using Equation 1 to determine the glide factor for each of the four glide trials.

$$x = C_G . Ln\left[\frac{V_{xo}}{C_G} . t + 1\right] \quad (1)$$

A 4th order Butterworth low-pass filter with a 6Hz cut-off frequency was applied to the 105-frames of filled position data. The 45-frames of filled and filtered position data from each of the four glides were used to calculate the following glide performance variables for each frame: horizontal velocity along the x-axis (m/s), hip angle (°), and trunk incline (°). Horizontal velocity was calculated to assess the amplified effect of digitising error on the first derivative.[15] Horizontal velocity (v) was calculated separately for the hip, knee, and shoulder using forward differentiation of the position data (x, m) with respect to time (t, seconds) for each frame (i) (Equation 2).

$$v_i = \frac{x_{i+1} - x_i}{t_{i+1} - t_i} \quad (2)$$

Hip angle was the angle of the swimmer's right thigh with respect to the trunk. The positions of the knee (k_x, k_y), hip (h_x, h_y), and shoulder (s_x, s_y) were used to determine distances between hip and shoulder (d_{hs}, cm), hip and knee (d_{hk}, cm), and knee and shoulder (d_{ks}, cm). The distance calculation is shown in Equation 3 using the hip and shoulder as an example and was repeated for the other distances. Hip angle (θ, °) was then calculated using these distances for each frame (Equation 4).

$$d_{hs} = \sqrt{(s_x - h_x)^2 + (s_y - h_y)^2} \tag{3}$$

$$\theta = \frac{180}{\pi} \cos^{-1} \frac{({d_{hs}}^2 + {d_{hk}}^2 - {d_{ks}}^2)}{2 \cdot {d_{ks}}^2} \tag{4}$$

Trunk incline (φ, °) was calculated as the angle between the trunk, defined by the hip and shoulder position data, and the external x-axis (Equation 5).

$$\varphi = \frac{180}{\pi} \tan^{-1} \frac{(s_y - h_y)}{(s_x - h_x)} \tag{5}$$

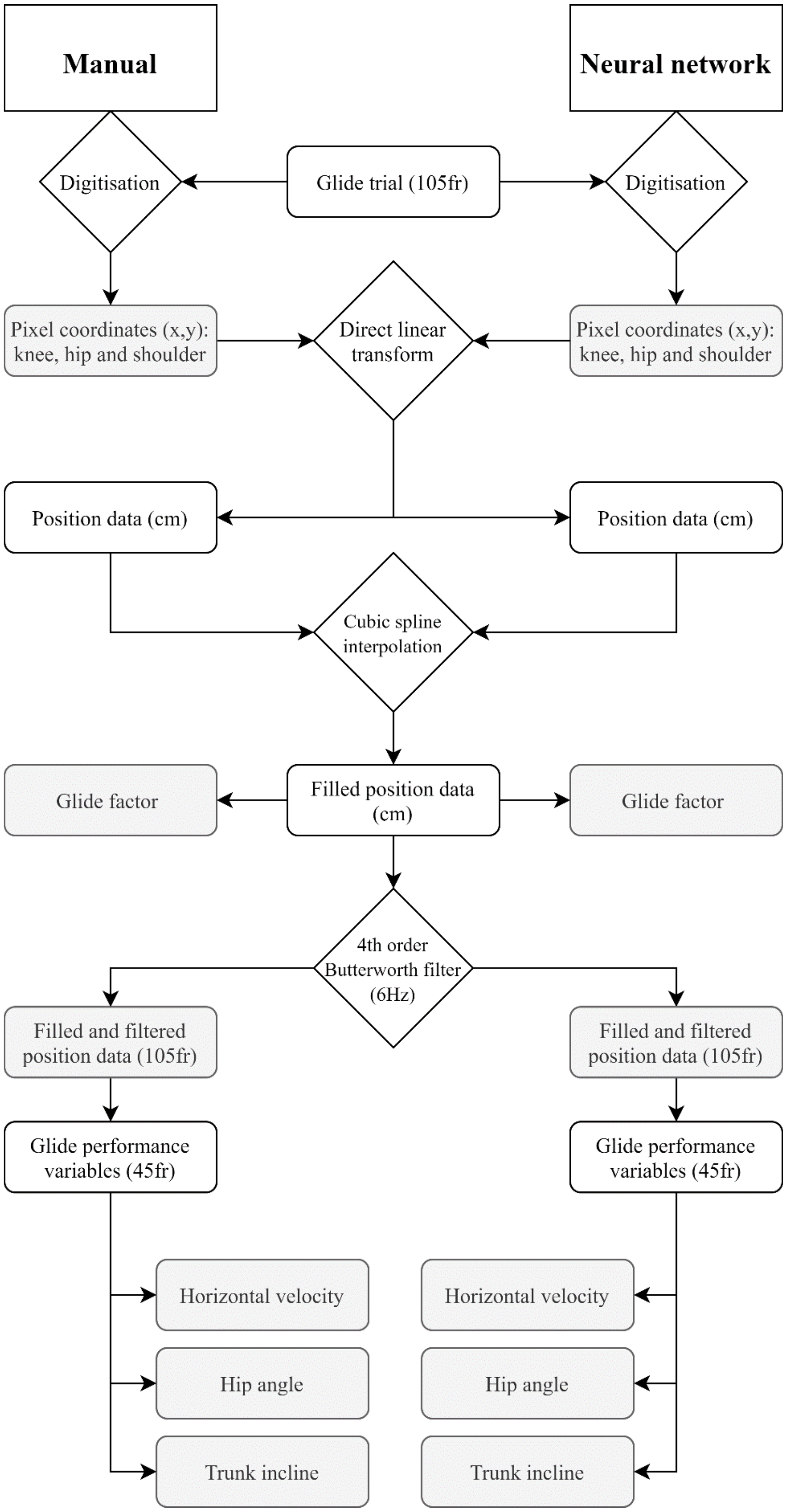

Figure 2. Data processing procedures of manual and neural network kinematic analysis of the glide phase. Accuracy and reliability analysis procedures described in "*Statistical analysis: neural network versus manual digitisation*" were carried out for the tabs shaded in grey.

Statistical analysis

Statistical analysis was performed using SPSS (Version 25, SPSS Inc., Chicago, USA), unless otherwise specified. Statistical significance was accepted at $p<0.05$ for all tests. For all intra-class correlation calculations, an absolute agreement, two-way mixed effects ICC model was used.[27] ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 were indicative of poor, moderate, good, and excellent reliability, respectively.[27]

Manual digitisation reliability

Intra-rater reliability for the first author's five digitisation attempts of the four glide trials was assessed using ICCs of raw pixel x- and y-coordinates of the body landmarks. Using Microsoft Excel, the mean of the standard deviations (mean error) of five digitisation attempts of the four glide trials (i.e. 20 datasets) were calculated for the time series data of horizontal velocity; hip angle; trunk incline; and glide factor. Ninety-five percent confidence intervals (95%CIs) were calculated for each of these variables using the *t*-distribution and the mean error. The confidence intervals were applied to the mean of each variable across the four trials to produce an acceptable range from five repeated digitisation attempts by a human operator. Inter-rater reliability of manual digitisation between the first and last authors was evaluated using RMSE and ICCs for 214-frames from ten pilot glide trials (approximately 20 random frames per trial) recorded using the same procedures in this study.

Neural network versus manual digitisation

Average time taken by the manual operator to digitise a single trial was calculated. The time to train the NN and the time required by the NN to digitise all trials (n=90) were also recorded. Similarity between digitisation by the manual operator and by the NN was assessed with RMSE and ICC for raw pixel coordinate data (x- and y-axis); filled and filtered position data (x- and y-axis); and instantaneous horizontal velocities, hip angles, and trunk inclines across the four trials (see Figure 2). RMSE was also calculated for glide factor. Relative error (%) of the RMSE for instantaneous velocities, hip angle, trunk incline, and glide factor were calculated by dividing the RMSE by the range (maximum-minimum) of each variable across the four trials and multiplying by 100. To evaluate the effect of glide velocity on digitisation accuracy, relative error (%) of the RMSE for NN and manually derived instantaneous velocities were calculated for all body landmarks (n=4 glides) within the manually derived glide velocity ranges: <1.4m/s, 1.4-1.6m/s, 1.6-1.8m/s, 1.8-2.0m/s, 2.0-2.2m/s, and >2.2m/s. Instantaneous velocity error was used to evaluate the effect of glide velocity on digitisation accuracy due to the susceptibility of error inflation when calculating the first derivative. 95%CIs were used to determine whether the neural network-derived means fell within an acceptable range of the human operator-derived average value for each variable.

Results

Manual digitisation reliability

Intra-rater reliability was 'excellent'[27] between digitisation attempts by the first author for all body landmarks in each of the four glide trials (x-coordinates: ICC=1.00, $p<0.001$ and y-coordinates: ICC>0.99, $p<0.001$). Inter-rater reliability was 'excellent' for digitisation conducted by the first and last authors for all body landmarks (x-coordinates: ICC>0.99, $p<0.001$ and RMSE=0.50 pixels; y-coordinates: ICC>0.99, $p<0.001$ and RMSE=0.45 pixels).

Neural network versus manual digitisation

The NN was trained in Google Colaboratory over approximately nine hours, without the need for monitoring by a human operator, producing a mean test error of 2.04 pixels, or 5.7mm. The NN digitised 90 glide videos consisting of 105-frames each (28,350 body landmarks) in 13.5min on the basic local machine. Average time for the first author to digitise a single 105-frame glide trial (315 body landmarks) was approximately 35min.

Frames containing body landmarks that were unidentifiable due to image blurring or that were obscured by air bubbles were omitted from analysis. Landmarks that were labelled with <95% probability by the NN were also omitted. Post-hoc analysis of the landmarks omitted from manual digitisation were found to be the same as those that were assigned <95% probability by the NN. Consequently, 3.8%, 14.5%, and 4.5% of knee, hip, and shoulder body landmarks, respectively, were filled using a cubic spline filter. Comparisons of position data between manual and NN digitisation are shown in Table 1. Agreement in raw pixel and filled and filtered position data for knee, hip, and shoulder in the x- and y-axis between the two methods was near perfect (ICC>0.999, p<0.001). RMSE of position data for all body landmarks was approximately 4-5mm.

Table 1. Comparison of digitised x- and y-coordinate and position data by manual and neural network digitisation.

Variable	**Knee**			
	x		y	
	RMSE[†]	ICC[‡], *p*	RMSE	ICC, *p*
Raw coordinate (pixel)	1.78	>0.999, <0.001	1.77	>0.999, <0.001
Filled and filtered position (mm)	5.2	>0.999, <0.001	4.7	>0.999, <0.001
	Hip			
	x		y	
	RMSE	ICC, *p*	RMSE	ICC, *p*
Raw coordinate (pixel)	2.06	>0.999, <0.001	1.50	>0.999, <0.001
Filled and filtered position (mm)	5.1	>0.999, <0.001	3.9	>0.999, <0.001
	Shoulder			
	x		y	
	RMSE	ICC, *p*	RMSE	ICC, *p*
Raw coordinate (pixel)	1.91	>0.999, <0.001	1.62	>0.999, <0.001
Filled and filtered position (mm)	4.8	>0.999, <0.001	4.0	>0.999, <0.001

†Root-mean-square error; ‡Intra-class correlation coefficient

Means, standard deviations, and 95%CIs of each glide performance variable and comparisons of glide performance variables derived from manual and NN digitisation are shown in Table 2. 'Excellent' reliability (ICC>0.95, p<0.001) was found in all glide performance variables, with relative error ≤5.4%. Mean glide variables from the four trials derived by the NN were within the acceptable range of the manual operator. Since glide factor was determined from a single swimmer, glide factor relative error was calculated using the range in glide factor (4.17–5.24 m) from a sample of 16 elite swimmers[28] of similar ability to our swimmer. Digitisation accuracy between the NN and manual operator decreased as glide velocity increased, with greater relative instantaneous velocity error at higher glide velocities (Figure 3).

Table 2. Comparative analysis of glide performance variables derived by manual and neural network digitisation.

Glide variable	Manual mean trials=4	Intra-rater 95%CIs† trials=4 of n=5 repeats	Neural network mean trials=4	Mean difference	Manual vs neural network (RMSE‡)	Relative error (%)	ICC§, *p*
Knee velocity (m/s)	1.76	1.70-1.85	1.77	0.01	0.10	5.4	0.977, <0.001
Hip velocity (m/s)	1.81	1.73-1.89	1.81	<0.01	0.09	4.8	0.982, <0.001
Shoulder velocity (m/s)	1.81	1.74-1.87	1.81	<0.01	0.08	4.4	0.984, <0.001
Hip angle (°)	166.00	164.50-167.50	166.13	0.13	0.73	3.7	0.996, <0.001
Trunk incline (°)	1.59	1.42-1.77	1.64	0.05	0.28	3.5	0.998, <0.001
Glide factor (m)	4.80	4.64-4.97	4.82	0.02	0.03	2.9	-

†Ninety-five percent confidence intervals; ‡Root-mean-square error; §Intra-class correlation coefficient

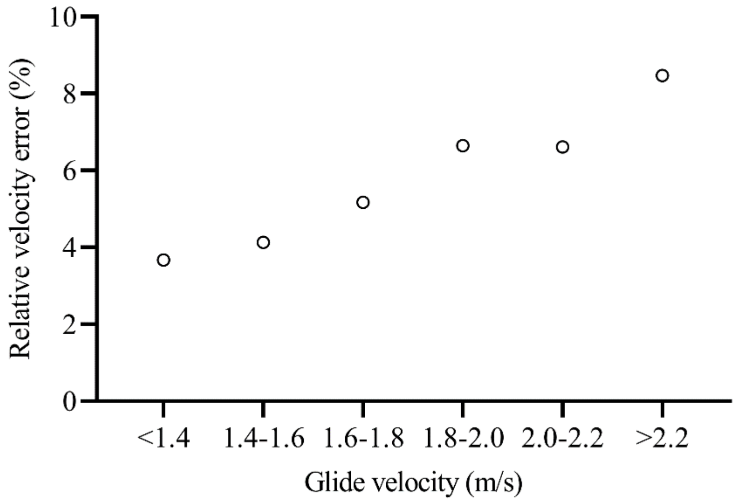

Figure 3. The effect of glide velocity on instantaneous velocity error (relative error of the root-mean-square error, %), derived from NN and manually digitised body landmarks.

Discussion

The purpose of this study was to determine the speed, accuracy and reliability of a NN to digitise body landmarks in 2D videos against manual digitisation and to assess accuracy and reliability of the derived kinematic variables from those body landmark data. The performance of the NN trained in DeepLabCut™ exceeded expectations. Not only were the relative errors within the bounds of manual digitisation (Tables 1 and 2), the NN digitised video data at a rate *233 times faster than the manual operator*. By comparison, automated digitisation methods with corrective manual adjustments have improved digitising time by 2.5 times that of manual digitisation.[2,29] In addition to significant improvements in digitising time, position data digitised by a NN can be used to compute movement and performance variables with high accuracy and reliability compared with manually-derived variables (Table 2). The findings have implications for applying NNs to digitise video data in biomechanics research to enable accurate and expedient performance analysis.

Comparison of the neural network with existing digitisation methods

For tracking programs to be useful for practical application, digitisation accuracy must be comparable to manual digitisation, as error in position data can inflate error in the calculations of kinematic variables.[30] Image processing algorithms have been used to automatically track light emitting diodes (LEDs) fixed to a swimmer's wrist in 2D video of dive starts.[31] Though the algorithm used by Slawson et al[31] allowed high digitisation processing speeds of the wrist, the estimation error was 50mm against the manually-derived wrist dive trajectory. The landmark position error in our study compared with manual digitisation was much lower (RMSE~4-5mm) than that of Slawson et al[31] and compares well with the error in landmark error from a markerless image processing system (wrist joint RMSE<5.6mm) designed by Ceseracciu et al.[32] Horizontal velocity RMSE was slightly lower in our study (≤0.10m/s) than wrist horizontal velocity RMSE in the study by Ceseracciu et al[32] (0.17m/s). Despite its relatively low error for wrist position and velocity, the markerless analysis system used by Ceseracciu and et al[32] had a runtime of 2-3hours to track the trajectories of three body landmarks for a single front crawl trial. In addition to its processing time, the system required clear images of the swimmer's silhouette during front crawl trials as well as static dry-land images, which may not be feasible for sport scientists and coaches to obtain. Another automatic tracking software showed excellent agreement with manual digitisation of LEDs attached to the anterior superior iliac spine during front crawl swimming, with a small standard measurement error of 1mm.[2] Following automatic digitisation, however, this tracking system tended to require manual adjustments to digitised data as the tracking software on its own has been found to incorrectly label between 14%[2] and 17%[29] of body landmarks. Therefore, the small digitising error of 1mm using this method may be partly attributable to corrective manual intervention.

To our knowledge, the current study is only the second application of DeepLabCut™ in an aquatic setting. 2D joint position data have also been obtained using DeepLabCut™

during underwater running, where the training digitisation error (neural network versus manual digitisation) was ~10mm.[9] The greater accuracy in our application of DeepLabCut™ than in the underwater running study may be due to different movement patterns and/or the use of black body paint to indicate joint positions in our study compared with a markerless approach used by Cronin et al.[9] Depending on the direction of the digitisation error in the 2D axis, our findings could be limited by propagation error. For example, if the shoulder was digitised 5mm above its true location and the hip 5mm below its location along the y-axis, hip and trunk incline angles would be affected. Despite the risk of propagation error, the relative error in instantaneous hip and trunk incline angles was arguably small (3.5-3.7%). Propagation error would also affect horizontal velocity calculations, as digitisation error is amplified with each derivative.[15] The NN was accurate in determining instantaneous velocities for all three landmarks when compared with manually derived velocities (Table 2). By comparison, mean differences in instantaneous horizontal velocity of the head, calculated from position data digitised by a NN ranged from 0.02-0.03m/s for all four competitive strokes,[33] producing a similar mean difference for the knee, hip and shoulder landmarks in this study (≤0.01m/s). While these two applications of NNs for digitisation of 2D video differed in their experimental approach, NNs appear to be an effective tool for digitisation when compared with a human operator. The NN in this study produced means that were consistently within the acceptable range of manual digitisation for all glide performance variables, indicating there was no loss of accuracy when compared with manual digitisation with a significant improvement in processing time.

An advantage of manual digitisation over automatic tracking methods is the decision by a human operator to omit markers that are subject to blurring or have been obscured. While the NN assigned coordinates to body landmarks in all frames, including body landmarks that were unidentifiable by the manual operator, post-hoc analysis revealed that landmarks that were given probability ratings <95% by the NN were the same ones omitted by the manual operator. The process of omitting these landmarks from the NN dataset was conducted manually in our study; however, this process can be automated using a simple computational routine in future applications to further improve data processing time. The image feature detection algorithm in the NN software appears to be robust enough to accurately determine body landmarks in underwater video that it had not been exposed to during NN training. Training, therefore, needs to be done just once for a given task, such as underwater gliding, for the NN to be valid for future data collections. NNs can also be trained with a sample from existing databases consisting of video data with painted body landmarks, unlocking the potential to analyse historical datasets in a completely new way.

Applications of neural networks in swimming

The use of a NN for digitisation in this study produced small relative error in glide factor values compared with manual digitisation. This finding was impressive given glide factor analysis is highly sensitive to decelerations and involves fitting a logarithmic function to position data. Glide factor analysis is essential to our understanding of overall glide performance because it can be used to compare glide efficiency within and between swimmers by ‘correcting’ for the swimmer’s glide velocity.[26] By correcting for velocity, factors that influence glide efficiency (e.g. posture, morphology, swim attire) can be evaluated using glide factor.[13,25] Thanks to the time-efficiency of the NN trained in this study, evaluation of glide efficiency and performance from 2D video analysis is now more viable for sport scientists and coaches.

Limitations and future research

The study was limited by the camera shutter speed that resulted in blurring of some body landmarks during the early phase of the glide when swimmers were moving at high velocities. Image distortion of body landmarks at high velocities reduced digitisation accuracy of the NN compared with manual digitisation (Figure 3). Cameras with higher frame rates (e.g. ≥120Hz), shutter speeds, and light sensitivity may reduce the amount of body landmarks omitted from analysis and provide a greater number of data points for interpolation, which may further improve accuracy of kinematic variable calculations.

The image recognition algorithm of the NN was found to be as accurate as a human operator for digitisation of painted landmarks in video captured under the same environmental conditions as the training frames. However, changes to the visual characteristics of painted landmarks in 2D video may limit the ability of the NN to recognise them, as evident with landmark distortion at high velocities. We were unable to assess whether digitisation accuracy of this NN would occur in glide video at a different location with different lighting properties, water clarity, and camera specifications, resulting in the possibility of overfitting the neural network to the training dataset. Future research would be advantageous to determine whether variability of video input in the NN training procedure improves robustness of the NN and generalisability to multiple settings. While the NN required approximately nine hours to train, once trained, the weights can be copied onto any local machine and used for analysis purposes on a basic laptop computer.

Training time could have been reduced in this study by reducing the image resolution of the training frames,[34] though it is unlikely that digitising accuracy would have been impacted because the input videos had the same resolution as the training images. Calibration time was negatively impacted because the camera setup required a field of view correction to minimise reprojection error. Where a fixed-camera setup is not viable, cameras with minimal visual distortion at the bounds of the field of view would reduce the need for a field of view correction and minimise calibration time.

Digitisation accuracy appeared to be improved by applying black body paint to body landmarks compared with markerless analysis methods.[9,32] In regards to the NNs trained in DeepLabCut™ for an aquatic setting, the use of painted landmarks improved 2D digitisation error from 10mm[9] to 4-5mm in our study. Additional time and expertise, however, is required to mark swimmers. Sports scientists and coaches should consider the trade-off between preparation time and accuracy when using NNs to digitise 2D video. The methods presented here could be used in future research involving kinematic analysis of land-based activities, especially those performed predominantly in a single plane of motion. In athletics, for instance, a fixed-camera setup and pre-calibrated area could be used to assess 2D kinematics of running, jumping, or throwing in a training environment. Kinematic analysis in weightlifting commonly involves video and manual digitisation methods to estimate barbell trajectory during lifts.[35] Barbell trajectory can be used to assess movement characteristics, provide technical feedback, and calculate critical performance variables, such as barbell velocity.[36] Automated digitisation of the end of the barbell in 2D video, however, is difficult as it can exhibit similar colour characteristics to the surrounding image.[37] Given the maximal barbell velocity of elite weightlifters during the snatch lift is between 1.5-2m/s,[38,39] NNs could be used for automated digitisation of the barbell in the sport of weightlifting.

Conclusion

To our knowledge, few studies exist in which kinematic data from video analysis have been derived in an accurate, time-efficient manner and the most effective strategies have involved the use of NNs. DeepLabCut^{TM} was found to be an accurate method of extracting kinematic data to analyse glide posture, efficiency and performance compared with manual digitisation. The use of NN software for auto-digitisation of body landmarks could be substantially beneficial to biomechanics researchers, sports scientists, and coaches. The time saving compared to manual digitising may enable rapid feedback of performance measures in training and simulated-competitive environments.

Disclosure of interest

The authors report no conflict of interest.

References

1. Yeadon M, Challis JH. The future of performance-related sports biomechanics research. *J Sport Sci.* 1994;12(1):3-32.
2. Dos Santos KB, Lara JP, Rodacki AL. Reproducibility and repeatability of intracyclic velocity variation in front crawl swimming from manual and semi-automatic measurement. *Human Movement.* 2017;18(3):55-59.
3. Adnan NMN, Ab Patar MNA, Lee H, Yamamoto SI, Jong-Young L, Mahmud J. Biomechanical analysis using Kinovea for sports application. *IOP Conference Series: Materials Science and Engineering.* 2018;342(1):012097.
4. Mathis A, Mamidanna P, Cury KM, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience.* 2018;21(9):1281.
5. Mehrizi R, Peng X, Xu X, Zhang S, Li K. A Deep Neural Network-based method for estimation of 3D lifting motions. *Journal of biomechanics.* 2019;84:87-93.
6. Pishchulin L, Insafutdinov E, Tang S, et al. Deepcut: Joint subset partition and labeling for multi person pose estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016:4929-4937.
7. Insafutdinov E, Pishchulin L, Andres B, Andriluka M, Schiele B. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. *European Conference on Computer Vision.* 2016:34-50.
8. Sanders RH, Gonjo T, McCabe CB. Reliability of three-dimensional linear kinematics and kinetics of swimming derived from digitized video at 25 and 50 Hz with 10 and 5

frame extensions to the 4th order Butterworth smoothing window. *Journal of sports science & medicine.* 2015;14(2):441.

9. Cronin NJ, Rantalainen T, Ahtiainen JP, Hynynen E, Waller B. Markerless 2D kinematic analysis of underwater running: A deep learning approach. *Journal of biomechanics.* 2019;87:75-82.

10. Stamm A, James DA, Burkett BB, Hagem RM, Thiel DV. Determining maximum push-off velocity in swimming using accelerometers. *Procedia Engineering.* 2013;60:201-207.

11. Guimaraes AC, Hay JG. A mechanical analysis of the grab starting technique in swimming. *Journal of Applied Biomechanics.* 1985;1(1):25-35.

12. Cossor J, Mason B. Swim start performances at the Sydney 2000 Olympic Games. *Proceedings of Swim Sessions of the XIX Symposium on Biomechanics in Sports* 2001:70–74.

13. Naemi R, Psycharakis SG, McCabe C, Connaboy C, Sanders RH. Relationships between glide efficiency and swimmers' size and shape characteristics. *Journal of applied biomechanics.* 2012;28(4):400-411.

14. Zamparo P, Gatta G, Pendergast D, Capelli C. Active and passive drag: the role of trunk incline. *Eur J Appl Physiol.* 2009;106(2):195-205.

15. Winter DA. *Biomechanics and motor control of human movement.* Hoboken, New Jersey: John Wiley & Sons; 2009.

16. McCabe CB, Psycharakis S, Sanders R. Kinematic differences between front crawl sprint and distance swimmers at sprint pace. *Journal of Sports Sciences.* 2011;29(2):115-123.

17. Naemi R, Easson WJ, Sanders RH. Hydrodynamic glide efficiency in swimming. *Journal of Science and Medicine in Sport.* 2010;13(4):444-451.

18. Elipot M, Dietrich G, Hellard P, Houel N. Poster Session II, July 14th 2010—Abstracts: Cinalysis: A new software for swimming races analysis. *Procedia Engineering.* 2010;2(2):3467.

19. Bouguet J-Y. Camera calibration toolbox for Matlab (2008). *URL http://wwwvisioncaltechedu/bouguetj/calib_doc/.* 2008;1080.

20. Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence.* 2000;22(11):1330-1334.

21. Abdel-Aziz Y, Karara H. Direct linear transformation into object space coordinates in close-range photogrammetry. *Proc Symp Close-Range Photogrammetry.* 1971:1-18.

22. Kwon YH, Casebolt JB. Effects of light refraction on the accuracy of camera calibration and reconstruction in underwater motion analysis. *Sports biomechanics.* 2006;5(2):315-340.
23. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016:770-778.
24. Pishchulin L, Insafutdinov E, Tang S, et al. Deepcut: Joint subset partition and labeling for multi person pose estimation. Paper presented at: Proceedings of the IEEE conference on computer vision and pattern recognition2016.
25. Naemi R, Sanders RH. A "hydrokinematic" method of measuring the glide efficiency of a human swimmer. *Journal of biomechanical engineering.* 2008;130(6).
26. Machtsiras G. *Utilizing Flow Characteristics to Increase Performance in Swimming (Unpublished doctoral dissertation)*. Edinburgh, Scotland The University of Edinburgh; 2012.
27. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine.* 2016;15(2):155-163.
28. Thow JL, Naemi R, Sanders RH. Comparison of modes of feedback on glide performance in swimming. *J Sport Sci.* 2012;30(1):43-52.
29. Magalhaes FA, Sawacha Z, Di Michele R, Cortesi M, Gatta G, Fantozzi S. Effectiveness of an automatic tracking software in underwater motion analysis. *Journal of sports science & medicine.* 2013;12(4):660.
30. Robertson G, Caldwell G, Hamill J, Kamen G, Whittlesey S. *Research methods in biomechanics.* 2 ed. Champaign, IL: Human Kinetics; 2013.
31. Slawson S, Conway P, Justham L, West A. The development of an inexpensive passive marker system for the analysis of starts and turns in swimming. *Procedia Engineering.* 2010;2(2):2727-2733.
32. Ceseracciu E, Sawacha Z, Fantozzi S, et al. Markerless analysis of front crawl swimming. *Journal of biomechanics.* 2011;44(12):2236-2242.
33. Elipot M. A new paradigm to do and understand the race analyses in swimming: The application of convolutional neural networks. *ISBS Proceedings Archive.* 2019;37(1):455.
34. Mathis A, Warren RA. On the inference speed and video-compression robustness of DeepLabCut. *BioRxiv.* 2018:457242.

35. Garhammer J. Biomechanical profiles of Olympic weightlifters. *Journal of Applied Biomechanics.* 1985;1(2):122-130.
36. Garhammer J, Newton H. Applied video analysis for coaches: Weightlifting examples. *International Journal of Sports Science & Coaching.* 2013;8(3):581-594.
37. Dæhlin TE, Krosshaug T, Chiu LZ. Enhancing digital video analysis of bar kinematics in weightlifting: A case study. *The Journal of Strength & Conditioning Research.* 2017;31(6):1592-1600.
38. Nagao H, Kubo Y, Tsuno T, Kurosaka S, Muto M. A Biomechanical Comparison of Successful and Unsuccessful Snatch Attempts among Elite Male Weightlifters. *Sports.* 2019;7(6):151.
39. Gourgoulis V, Aggeloussis N, Garas A, Mavromatis G. Unsuccessful vs. successful performance in snatch lifts: a kinematic approach. *The Journal of Strength & Conditioning Research.* 2009;23(2):486-494.