

Research article

The 8th Australasian Conference on Mathematics and Computers in Sport, 3-5 July 2006, Queensland, Australia

MAXIMIZING HEIGHT, DISTANCE OR ROTATION FROM REAL-TIME ANALYSIS VISUALISATION OF TAKE-OFF ANGLES AND SPEED

Richard Green

Department of Computer Science, University of Canterbury, New Zealand.

Published (online): 15 December 2006

ABSTRACT

Studies to optimise take off angles for height or distance have usually involved either a time-consuming invasive approach of placing markers on the body in a laboratory setting or using even less efficient manual frame-by-frame joint angle calculations with one of the many sport science video analysis software tools available. This research introduces a computer-vision based, marker-free, real-time biomechanical analysis approach to optimise take-off angles based on speed, base of support and dynamically calculated joint angles and mass of body segments. The goal of a jump is usually for height, distance or rotation with consequent dependencies on speed and phase of joint angles, centre of mass (COM) and base of support. First and second derivatives of joint angles and body part COMs are derived from a Continuous Human Movement Recognition (CHMR) system for kinematical and what-if calculations. Motion is automatically segmented using hierarchical Hidden Markov Models and 3D tracking is further stabilized by estimating the joint angles for the next frame using a forward smoothing Particle filter. The results from a study of jumps, leaps and summersaults supporting regular knowledge of results feedback during training sessions indicate that this approach is useful for optimising the height, distance or rotation of skills.

KEY WORDS: Gymnastics, jumping, three-dimensional kinematics, computer vision.

INTRODUCTION

Sport skills are tracked and biomechanically analysed by either requiring athletes to wear joint markers/identifiers (an approach with has the disadvantage of significant set up time) or manually marking up video frame-by-frame. Such complex and time consuming approaches to tracking and analysis is an impediment to daily use by coaches and has barely changed since it was developed in the 1970s. Using a less invasive approach free of markers, computer vision research into tracking and

recognizing full-body human motion has so far been mainly limited to gait or frontal posing (Moeslund and Granum, 2001). Various approaches for tracking the whole body have been proposed in the image processing literature using a variety of 2D and 3D body models. However cylindrical, quadratic and ellipsoidal (Drummond and Cipolla, 2001; Kakadiaris and Metaxas, 1996; Pentland and Horowitz, 1991; Wren et al., 1997) body models of previous studies do not contour accurately to the body, thus decreasing tracking stability. To overcome this problem, in this research 3D clone-

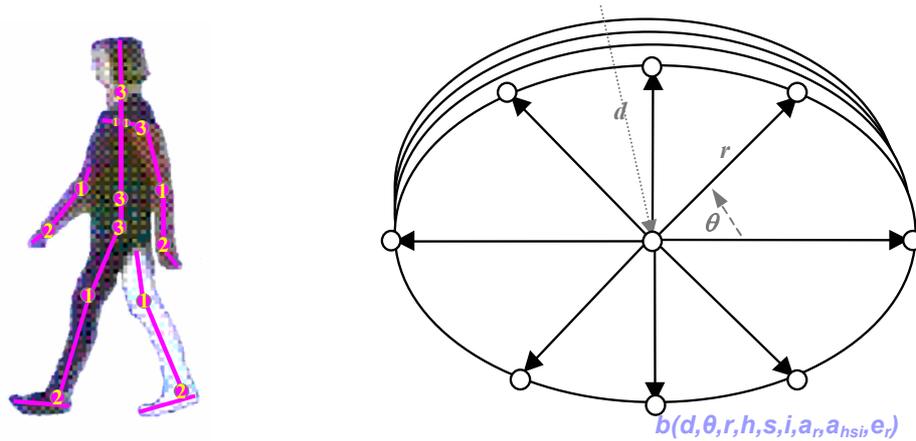


Figure 1. Clone-body-model consisting of clone-body-parts which have a cylindrical coordinate system of surface points $b()$ and up to three DOF for each joint linking the clone-body-parts. Each surface point is a vector b with cylindrical coordinates (d, θ, r) , colour (h, s, i) , accuracy of radius (a_r) , accuracy of colour (a_{hsi}) , elasticity of radius (e_r) .

body-model regions are sized and texture mapped from each body part by extracting features during the initialisation phase (Cham and Rehg, 1999). This clone-body-model has a number of advantages over previous body models:

It allows for a larger variation of somatotype (from ectomorph to endomorph), gender (cylindrical trunks do not allow for breasts or pregnancy) and age (from baby to adult).

Exact sizing of clone-body-parts enables greater accuracy in tracking edges, rather than the nearest best fit of a cylinder.

Texture mapping of clone-body-parts increases region tracking and orientation accuracy over the many other models which assume a uniform colour for each body part.

Region patterns, such as the ear, elbow and knee patterns, assist in accurately fixing orientation of clone-body-parts.

Neither joint markers nor manual frame-by-frame mark-up provide volume and 3D centre-of-mass (COM) estimates of a 3D body model – invaluable for 3D biomechanical analysis. In this study, joint angle velocities, together with the size and mass of body segments enabled more accurate optimisation of take-off angles supporting the goal of a jump whether for height, distance or rotation with consequent dependencies on phase of joint

angles and base of support.

CLONE-BODY-MODEL

The clone-body-model proposed in this paper consists of a set of clone-body-parts, connected by joints, similar to the representations proposed by Badler et al. (1993). Clone-body-parts include the head, clavicle, trunk, upper arms, forearms, hands, thighs, calves and feet. Degrees of freedom are modeled for gross full body motion. Degrees of freedom supporting finer resolution movements are not yet modeled, including the radioulnar (forearm rotation), interphalangeal (toe), metacarpophalangeal (finger) and carpometacarpal (thumb) joint motions.

Each clone-body-part consists of a rigid spine with pixels radiating out (Figure 1). Each pixel represents a point on the surface of a clone-body-part. Associated with each pixel is: radius or thickness of the clone-body-part at that point; colour as in hue, saturation and intensity; accuracy of the colour and radius; and the elasticity inherent in the body part at that point. Although each point on a clone-body-part is defined by cylindrical coordinates, the radius varies in a cross section to exactly follow the contour of the body as shown in Figure 2.

Automated initialisation assumes only one



Figure 2. Clone-body-model example rotating through 360 degrees.

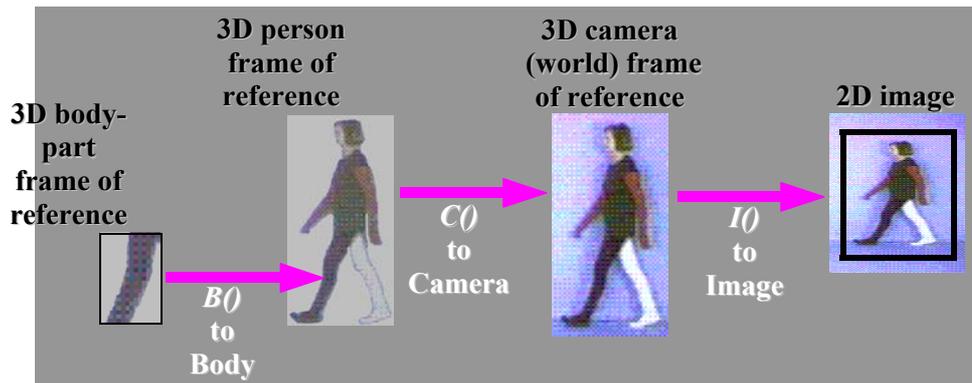


Figure 3. Three homogeneous transformation functions $B()$, $C()$, $I()$ project a point from a clone-body-part onto a pixel in the 2D image.

person is walking upright in front of a static background initially with gait being a known movement model. Anthropometric data (Pheasant, 1996) is used as a Gaussian prior for initializing the clone-body-part proportions with left-right symmetry of the body used as a stabilizing guide from 50th percentile proportions. Such constraints on the relative size of clone-body-parts and on limits and neutral positions of joints help to stabilize initializations. Initially a low accuracy is set for each clone-body-part with the accuracy increasing as structure from motion resolves the relative proportions. For example, a low colour and high radius accuracy is initially set for pixels near the edge of a clone-body-part, high colour and low radius accuracy for other near side pixels and a low colour and low radius accuracy is set for far side pixels. The ongoing temporal resolution following self occlusions enables increasing radius and colour accuracy. Breathing, muscle flexion and other normal variations of body part radius are accounted for by the radius elasticity parameter.

KINEMATIC MODEL

The kinematical model tracking the position and orientation of a person relative to the camera entails projecting 3D clone-body-model parts onto a 2D image using three chained homogeneous transformation matrices as illustrated in Figure 3 (see Equation 1 in Appendix).

Joint angles are used to track the location and orientation of each body part, with the range of joint angles being constrained by limiting the DOF associated with each joint. A simple motion model of constant angular velocity for joint angles is used in the kinematical model. Each DOF is constrained by anatomical joint-angle limits, body-part interpenetration avoidance and joint-angle equilibrium positions modelled with Gaussian stabilizers around their equilibria. To stabilize tracking, the joint angles are predicted for the next frame. The calculation of

joint angles, for the next frame, is cast as an estimation problem which is solved using a Particle filter (Condensation algorithm).

PARTICLE FILTER

The Particle Filter was developed to address the problem of tracking contour outlines through heavy image clutter (Isard and Blake, 1996; 1998). The filter's output at a given time-step, rather than being a single estimate of position and covariance as in a Kalman filter, is an approximation of an entire probability distribution of likely joint angles. This allows the filter to maintain multiple hypotheses and thus be robust to distracting clutter.

With about 32 DOFs for joint angles to be determined for each frame, there is the potential for exponential complexity when evaluating such a high dimensional search space. MacCormick (2000) proposed Partitioned Sampling and Sullivan (1999) proposed Layered Sampling to reduce the search space by partitioning it for more efficient particle filtering. Although Annealed Particle Filtering (Deutscher et al., 2000) is an even more general and robust solution, it struggles with efficiency which Deutscher (2001) improves with Partitioned Annealed Particle Filtering.

The Particle Filter is a considerably simpler algorithm than the Kalman Filter. Moreover despite its use of random sampling, which is often thought to be computationally inefficient, the Particle Filter can run in real-time. This is because tracking over time maintains relatively tight distributions for shape at successive time steps and particularly so given the availability of accurate learned models of shape and motion from the human-movement-recognition (CHMR) system. Here, the particle filter has:

3 probability distributions in problem specification:

1. Prior density $p(x)$ for the state x
 - joint angles x in previous frame
2. Process density $p(x_t|x_{t-1})$

- kinematical and clone-body-models (x_{t-1} : previous frame, x_t : next frame)
- 3. Observation density $p(z|x)$
 - image z in previous frame

one probability distribution in solution specification:

1. State Density $p(x_t|Z_t)$
 - where x_t is the joint angles in next frame Z_t

1. Prior density: Sample s'_i from the prior density $p(x_{t-1}|z_{t-1})$ where x_{t-1} =joint angles in previous frame, z_{t-1} . The sample set are possible alternate values for joint angles. When tracking through background clutter or occlusion, a joint angle may have N alternate possible values (samples) s with respective weights w , where *prior density*,

$p(x) \approx S_{t-1} = \{(s^{(n)}, w^{(n)}), n=1..N\}$ = a sample set (S_{t-1} is the sample set for the previous frame, $w^{(n)}$ is the n^{th} weight of the n^{th} sample $s^{(n)}$)

For the next frame, a new sample is selected, $s'_i = s_{t-1}$ by finding the smallest i for which $c^{(i)} \geq r$, where $c^{(i)} = \sum w^{(i)}$ and r is a random number $\{0, 1\}$.

2. Process density: Predict s_t from the process density $p(x_t|x_{t-1} = s'_i)$. Joint angles are predicted for the next frame using the kinematic model, body model & error minimisation. A joint angle, $s_t^{(n)}$ in the next frame is predicted by sampling from the

process density, $p(x_t|x_{t-1} = s'_i^{(n)})$ which encompasses the kinematic model, clone-body-model and cost function minimisation. In this prediction step both edge and region information is used. The edge information is used to directly match the image gradients with the expected model edge gradients. The region information is also used to directly match the values of pixels in the image with those of the clone-body-model's 3D colour texture map. The prediction step involves minimizing the cost functions (measurement likelihood density):

edge error E_e using edge information (see Equation 2 in Appendix):

region error E_r using region information (see Equation 3 in Appendix):

3. Observation density: Measure and weigh the new position in terms of the observation density, $p(z_t|x_t)$. Weights $w_t = p(z_t|x_t = s'_i)$ are estimated and then weights $\sum_n w^{(n)} = 1$ are normalized. The new position in terms of the observation density, $p(z_t|x_t)$ is then measured and weighed with *forward smoothing*:

- Smooth weights w_t over $1..t$, for n trajectories
- Replace each sample set with its n trajectories $\{(s_b, w_t)\}$ for $1..t$
- Re-weight all $w^{(n)}$ over $1..t$
- Trajectories tend to merge within 10 frames
- $O(N_t)$ storage prunes down to $O(N)$

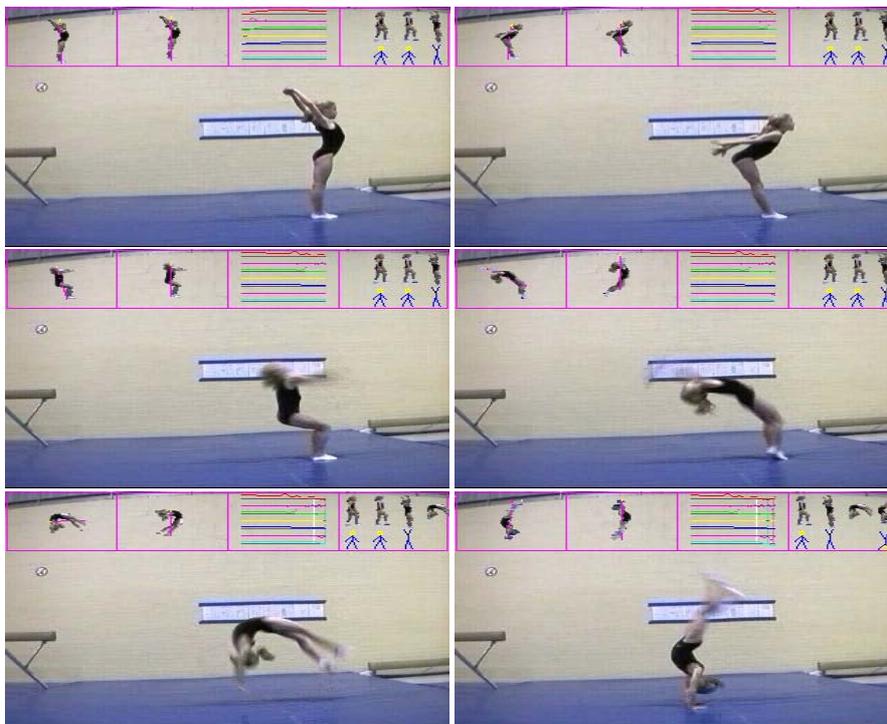


Figure 4. Tracking jumping into a flic-flac with four overlaid information tiles. Tile 1: principle axis through the body; Tile 2: body frame of reference (normalised to the vertical); Tile 3: motion vector trace (subset displayed); Tile 4: recognised skills.

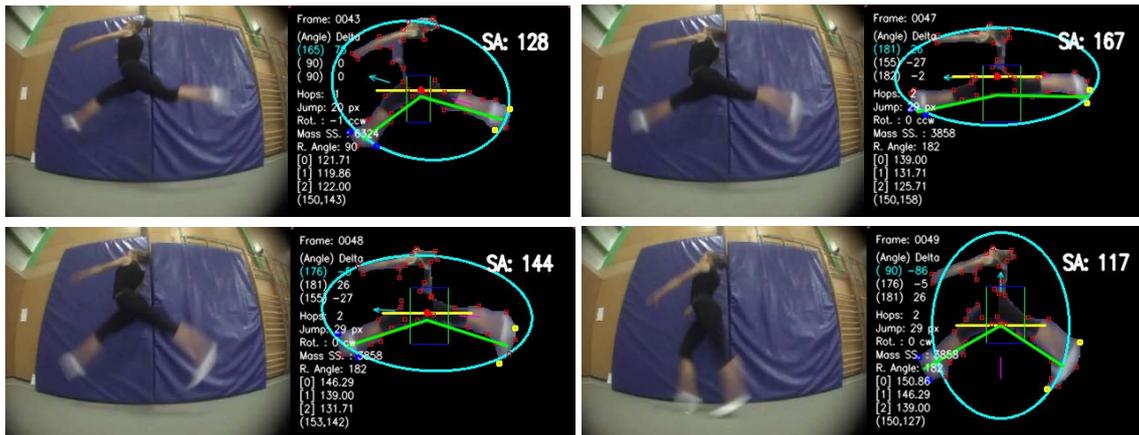


Figure 5. Tracking height, angle of splits, centre of mass and principal axis through a split jump.

In this research, feedback from the CHMR system utilizes the large training set of skills to achieve an even larger reduction of the search space. In practice, human movement is found to be highly efficient, with minimal DOFs rotating at any one time. The equilibrium positions and physical limits of each DOF further stabilize and minimize the dimensional space. With so few DOFs to track at any one time, a minimal number of particles are required, significantly raising the efficiency of the tracking process. Such highly constrained movement results in a sparse domain of motion projected by each motion vector.

Because the temporal variation of related joints and other parameters also contains information that helps the recognition process infer skill boundaries, the system computes and appends the temporal derivatives and second derivatives of these features to form the final motion vector. Hence the motion vector includes joint angles (32 DOF), body location and orientation (6 DOF), centre of mass (3 DOF), principle axis (2 DOF) all with first and second derivatives.

PERFORMANCE

Hundreds of jumps and leaps were tracked and classified using a 2GHz, 640MB RAM Pentium IV platform processing 24 bit colour within the Microsoft DirectX 9, Intel OpenCV environment under Windows XP. The video sequences were captured with a Logitech USB 2.0 camera at 30 fps, 320 by 240 pixel resolution. Each person jumped in front of a stationary camera with a static background and static lighting conditions with minimal shadows. Only one person was in frame at any one time. Tracking began when the whole body was visible which enabled initialisation of the clone-body-model.

The skill error rate quantifies CHMR system performance by expressing, as a percentage, the ratio

of the number and magnitude of joint angle tracking errors to the number of joint angles in the reference set. Depending on the skill, CHMR system skill error rates can vary by an order of magnitude. The CHMR system results are based on a set of a total of 240 jump patterns, from straight jumps and split leaps (Figure 5) to jumping backward into flic-flacs (Figure 4). These were successfully tracked and evaluated with their respective biomechanical components quantified where a skill error rate of only 3.8% was achieved.

Motion blurring lasted about 10 frames on average with the effect of perturbing joint angles within the blur envelope. Given a reasonably accurate angular velocity, it was possible to sufficiently de-blur the image. There was minimal motion blur arising from rotation about the longitudinal axis during a twisting salto due to a low surface velocity tangential to this axis from minimal radius with limbs held close to a straight body shape. This can be seen in Figure 6 where the arms exhibit no blurring from twisting rotation, contrasted with motion blurred legs due to a higher tangential velocity of the salto rotation.



Figure 6. Minimal motion blur arising from rotation about the longitudinal axis during a double twisting salto.

The CHMR system also failed for loose clothing. Even with smoothing, joint angles

surrounded by baggy clothes permutated through unexpected angles within an envelope sufficiently large as to invalidate the tracking and evaluation.

CONCLUSIONS

The 3.8% error rate attained in this research is not yet evaluating a natural world environment nor is this a real-time system with up to seconds to process each frame. The CHMR system did achieve 96.2% accuracy for the reference test set of skills. Although this 96.2% recognition rate was not as high as the 99.2% accuracy achieved, a larger test sample of skills were evaluated in this paper.

To progress towards the goal of lower error rates, the following improvements seem most important:

- Expand the clone-body-model to include a complete hand-model for enabling even more subtle movement domains such as finger signing and to better stabilize the hand position during tracking.
- Use a multi-camera or multi-modal vision system such as infra-red and visual spectrum combinations to better disambiguate the body parts in 3D and track the body in 3D.
- More accurately calibrate all movement skills with multiple subjects performing all skills on an accurate commercial tracking system recording multiple camera angles to improve on depth of field ambiguities. Such calibration would also remedy the qualitative nature of tracking results from computer vision research in general.
- Enhance tracking granularity using cameras with higher resolution, frame rate and lux sensitivity.

The results suggest that this approach has the potential to assist coaches and athletes optimise jump based skills during regular sessions by automatically displaying and logging biomechanical parameters of specific skills involving jumping and leaping.

REFERENCES

- Badler, N.I., Phillips, C.B. and Webber, B.L. (1993) *Simulating humans*. Oxford University Press, New York, NY.
- Cham, T.J. and Rehg, J. M. (1999) A multiple hypothesis approach to figure tracking. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **1**, 239-245. Available from URL: <http://ieeexplore.ieee.org/Xplore/dynhome.jsp>
- Deutscher, J., Blake, A. and Reid, I. (2000) Articulated body motion capture by annealed particle filtering. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2**, 1144-1149. Available from URL: <http://ieeexplore.ieee.org/Xplore/dynhome.jsp>
- Deutscher, J., Davison, A. and Reid, I. (2001) Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* **2**, 669-676. Available from URL: <http://ieeexplore.ieee.org/Xplore/dynhome.jsp>
- Drummond, T. and Cipolla, R. (2001) Real-time tracking of highly articulated structures in the presence of noisy measurements. In: *Proceedings of IEEE International Conference on Computer Vision* **2**, 315-320.
- Isard, M.A. and Blake, A. (1996) Visual tracking by stochastic propagation of conditional density. In: *4th European Conference on Computer Vision*. Cambridge, England. 343-356.
- Isard, M.A. and Blake, A. (1998) A mixed-state Condensation tracker with automatic model switching. In: *Proceedings of 6th International Conference on Computer Vision*. 107-112.
- Kakadiaris, I. and Metaxas, D. (1996) Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 81-87. Available from URL: <http://ieeexplore.ieee.org/Xplore/dynhome.jsp>
- MacCormick, J. and Isard, M.A. (2000) Partitioned sampling, articulated objects and interface-quality hand tracking. In: *Proceedings of European Conference on Computer Vision* **2**, 3-19.
- Moeslund, T.B. and Granum, E. (2001) A survey of computer vision-based human motion capture, *Computer Vision and Image Understanding* **18**, 231-268.
- Pentland, A. and Horowitz, B. (1991) Recovery of nonrigid motion and structure. *IEEE Transactions on PAMI*, **13**, 730-742.
- Pheasant, S. (1996) *Bodyspace*. In: *Anthropometry, ergonomics and the design of work*. Taylor & Francis.
- Rehg, J.M. and Kanade, T. (1995) Model-based tracking of self-occluding articulated objects, In: *Proceedings of Fifth International Conference on Computer Vision*. 612-617.
- Sullivan, J., Blake, A., Isard, M. and MacCormick, J. (1999) Object localization by bayesian correlation, In: *Proceedings of 7th International Conference on Computer Vision* **2**, 1068-1075.
- Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A. (1997) "Pfinder: Real-time tracking of the human body". *IEEE Transactions on PAMI* **19**, 780-785.

KEY POINTS

- Computer-vision based marker-free tracking.
- Real-time biomechanical analysis.
- Improve tracking using a forward smoothing Particle filter.
- Automatically segment using hierarchical Hidden Markov Models.
- Recognize skills using segmented motion.
- Optimize take-off angles using speed, base of support, joint angles and mass of body segments.
- Optimize height, distance or rotation of skills.

AUTHOR BIOGRAPHY**Richard GREEN****Employment**

Senior Lecturer

Degree

BSc, ME, PhD

Research interests

Computer vision, biomechanics, biomedical.

E-mail:

richard.green@canterbury.ac.nz

✉ **Richard Green**

Department of Computer Science, University of Canterbury, New Zealand

**APPENDIX
(Equations)****Equation 1:**

$$p(x, b) = I_i(x, C_i(x, B_i(x, b))) \quad (1)$$

where x is a parameter vector calculated for optimum alignment of the projected model with the image, B is the Body frame of reference transformation, C is the Camera frame of reference transformation, I is the Image frame of reference transformation, b is a body-part surface point, p is a pixel in 2D frame of video (Reh and Kanade, 1995).

Equation 2:

edge error E_e using edge information:

$$E_e(S_t) = \frac{1}{2n_e v_e} \sum_{x,y} (|\nabla i_t(x,y) - m_t(x,y,S_t)|)^2 + 0.5(S - S_t)^T C_t^{-1} (S - S_t) \rightarrow \min S_t \quad (2)$$

Equation 3:

region error E_r using region information (Equation 3):

$$E_r(S_t) = \frac{1}{2n_r v_r} \sum_{j=1}^{n_r} (i_t[p_j(S_t)] - i_{t-1}[p_j(S_{t-1})])^2 + E_e(S_t) \rightarrow \min S_t \quad (3)$$

where i_t represents the image at time t , m_t the model gradients at time t , n_e is the number of edge values summed, v_e is the edge variance, n_r is the number of region values summed, v_r is the region variance, p_j is the image pixel coordinate of the j th surface point on a clone-body-part.