Multimed Tools Appl (2014) 68:177–192
DOI 10.1007/s11042-012-1147-4

Genetic programming extension to APF-based monocular human body pose estimation

Piotr Szczuko

Published online: 13 June 2012

Abstract New method of the human body pose estimation based on a single camera 2D observation is presented, aimed at smart surveillance related video analysis and action recognition. It employs 3D model of the human body, and genetic algorithm combined with annealed particle filter for searching the global optimum of model state, best matching the object's 2D observation. Additionally, new motion cost metric is employed, considering current pose and history of the body movement, favouring the estimates with the lowest changes of motion speed comparing to previous poses. The "genetic memory" concept is introduced for the genetic processing of both current and past states of 3D model. State-of-the-art in the field of human body tracking is presented and discussed. Details of implemented method are described. Results of experimental evaluation of developed algorithm are included and discussed.

Keywords Pose estimation · Evolutionary optimization

1 Introduction

Human action recognition lies in the scope of computer vision research for years [12]. It can be utilized in human-computer-interaction methods (HCI), for gesture navigated user interfaces [10], for markerless motion capture systems, and threats recognition in smart surveillance systems [3]. This process often comprises of following stages: background modelling, detection of foreground objects, classification and tracking of objects, and finally analysis of the performed action for event recognition. In monocular vision systems a body pose to be estimated in 3D must be calculated based on a single 2D observation (video frame). For this purpose a generative approach is employed, utilizing 3D model of a human body, its pose

P. Szczuko (✉)
Multimedia Systems Department, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: szczuko@sound.eti.pg.gda.pl

being a subject to algorithmic alterations, as a result obtaining various 2D projections. Those 2D images of poses are compared with current 2D observation for best match. The pose is iteratively altered by optimization algorithms, based on matching metric values. The most popular optimization approaches are simulated annealing [4] and genetic algorithms [5]. In a single camera system (monocular) the estimation of 3D features of the object based on 2D projection is ambiguous. Therefore, multi-camera approaches are widely introduced, dealing with ambiguity by fusion of data from multiple 2D projections obtained from various observation points. Those techniques perform very well, and already have found a commercial application in markerless motion capture of actor performance [2, 6, 8, 9].

In the reported work a monocular vision is considered, as a basis for general purpose computer vision system, that can be useful in any conditions without strict requirements on the number and positions of cameras, e.g. for "smart home" applications such as health assessment based on physical activity, contactless user interfaces, or smart surveillance systems, recognizing human action for event detection.

In the paper first state-of-the-art in the pose estimation algorithms is described, then a structure and purpose of designed reference recordings database are presented, next a proposal of the new single camera method is presented based on evolutionary programming, motion dynamics and hierarchical pose estimation. Finally the approach is tested and results are discussed.

2 State-of-the-art in monocular and contactless human body tracking

Pose estimation methods utilize multidimensional parameterization of the body, where its state is described by degrees of freedom (DOFs) related to bones rotations and body position in space. Possible approaches are not consistent, as models can have 25–34 DOFs, while 3D animated models employ over 40 DOFs [4].

A 3D model of the body is considered, characterized by body proportions, number of DOFs, and angle restrictions. The optimal state of such model is being sought, the one which best matches the shape of current 2D observation of the real body and fulfils body biomechanical constraints.

Currently developed monocular vision methods try to cope with ambiguous relation between 2D observed shape and 3D pose being estimated. If an object is provided only as a shape (binary image of the silhouette, blob, and mask) extracted from the scene by typical background modelling and object detection methods, then information related to body orientation is unknown. It can either face the camera or be turned back. Other problem is related to presence of self occlusion, when limbs can be partially or fully hidden behind the torso and their state is undefined.

In case of orientation errors we propose employing well established methods of face detection, i.e. cascade of boosted classifiers working with Haar-like features [11] for resolving front-back mistakes.

In case of occlusions the unknown state of the limb is estimated as an interpolation between two well defined states (for off-line analysis, when each frame of the sequence is available) or as a continuation of previous well estimated motion, with utilization of Kalman filter for description of its state (for on-line analysis, when future states are yet unknown) [20].

2.1 Pose assessment

The model **X** to object **Z** matching degree $w(\mathbf{X},\mathbf{Z})$ can be expressed as a cumulative metric (1) proposed by Deutscher et al. [4], employing coverage of their silhouettes (2) and their edges (3):

$$w(X,Z) = w^r(X,Z) \cdot w^e(X,Z) \tag{1}$$

$$w^r(X,Z) = \exp\left(-\frac{1}{N}\sum_{i=1}^{N}\left(1-p_i^r(X,Z)\right)^2\right) \tag{2}$$

$$w^e(X,Z) = \exp\left(-\frac{1}{N}\sum_{i=1}^{N}\left(1-p_i^e(X,Z)\right)^2\right) \tag{3}$$

where:

X	model state
Z	observation (object shape or edges extracted from the video frame)
N	number of comparison points in the model image and object image
p_i^r, p_i^{e}	pixel-wise logical AND operator between model's and object's regions and edges respectively.

Optimization of model state based on matching metric $w(\mathbf{X},\mathbf{Z})$ is performed with following methods.

2.2 Particle filtering optimization

Probabilistic modelling of possible multidimensional states of the tracked body is typically performed with particle filtering approach (PF), known also as "CONditional DENSity propagATION" (CONDENSATION). In the computer vision applications it was first introduced for tracking outlines of hands [7], and later extended for whole body [4, 13]. PF method is useful for analysis of multiple hypotheses, here called "particles". Even if some model states are less probable considering previous states, they are taken into account, resulting in more robust tracking. The new particles are located densely around previous match, but also are randomly scattered in the whole range. The one with highest match (likely global optimum of $w(\mathbf{X},\mathbf{Z})$) is taken as a result. This method considers many modalities with varying probability, therefore it tests various action courses at the same time. Unfortunately, finding global optimum requires employing large numbers of particles, and long computation times are reported (e.g. 1 video frame in 18 min) [4].

2.3 Annealed particle filter

An interesting modification of PF method is Annealed Particle Filter (APF), where the optimum search is performed in M stages, so called layers. The layer m is characterized with annealing speed β_m, where $1 \geq \beta_0 > \beta_1 > \ldots > \beta_M$ and analyzed metric is $w_m(\mathbf{X},\mathbf{Z}) = w(\mathbf{X},\mathbf{Z})^{\beta m}$. The higher the β_m the more coarse $w_m(\mathbf{X},\mathbf{Z})$ is, and the search is less susceptible to

local optima (Fig. 1). In consecutive layers the particles are located in the most probable areas, where high values of $w_m(\mathbf{X},\mathbf{Z})$ occur, with some randomization introduced. This method performs significantly better than PF [4].

2.4 Benchmarking

For methods efficiency comparison several databases were created, i.e. Vlasic et al. silhouette database [19], and HumanEva benchmarks [17]. The latter consists of database of multi-camera recordings of several human actions, reference motion data gathered with high precision motion capture system, and Particle Filtering basis algorithm provided along with documentation and reference results [17].

2.5 Predefined action recognition

Other approach can be taken for distinct motions such as walking. Instead of estimating every consecutive pose, a common approach is to compare estimated pose to given presets, e.g. phases of walking [14, 16], but applications are limited to detection of predefined actions only, and extension of this set is time-consuming [3].

3 Reference data repository

For the purpose of assessment of estimation quality a reference poses database is prepared in Gdansk University of Technology, comprising of input poses of the actor performance, and reference description of actual poses. Recorded actions contain: variants of falling, tipping, fainting, resting down, balancing, tying a shoelace, sitting with crossed arms, legs, embracing torso, etc. (31 sequences ca. 6 s long). However, the captured poses data turned out to be not precise enough for the method evaluation described in Section 5, therefore supplementary input poses are synthesised employing adjustable 3D model whose state is stored in the database for reference. Nevertheless, for both real actor images and 3D model images the approach is the same, and the repository is able to include real as well as synthesised data. The body pose tracking algorithm accepts a sequence of 1-bit images containing silhouettes of actual moving body (so called "masks") obtained from the camera image. Its output comprises of state vector of 3D model, the one that was the best match for the particular 2D pose. Therefore objective assessment of the pose estimation quality requires a comparison

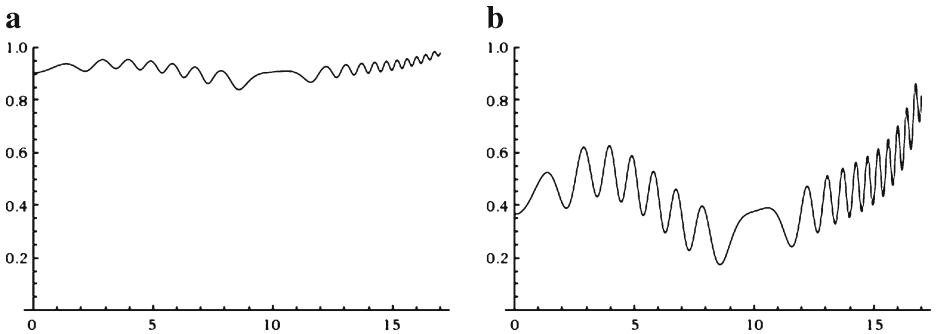

Fig. 1 APF layers of sample metric $w_m(\mathbf{X},\mathbf{Z}) = w(\mathbf{X},\mathbf{Z})^{\beta m}$: **a** β_1=0.1, **b** β_0=1. The smaller the β_m is the coarser the function, and optimization is less susceptible to local optima

between the actual 3D state of the actor's body and estimated 3D state of the model. The reference state can be acquired by Motion Capture system, measuring positions and angles of body joints (Fig. 2).

3.1 Video data acquisition

Reference video sequences for the database can be recorded with any number of cameras. It is advised to locate each camera at height of 2.0 m above the floor level, aimed at −20° elevation down, and in case of multiple cameras to acquire significantly different views of the action, e.g. 1st camera perpendicular to action (direction of a walk or fall), 2nd parallel to the action, 3rd observing action at angle 45° (Fig. 3). For the video database currently developed Canon XH G1 video cameras are used, recording FullHD image with 1920×1080 pixels resolution and 50 frames per second. A HDV (high definition video) format is used, with lossy video compression in the H.262/MPEG-2 Part 2 standard. Cameras are synchronized by GENLOCK reference signal produced by the 1st camera, and used to set internal clocks of other cameras. In this case timestamps are registered along with the image, therefore for further editing those recordings can be precisely synchronized.

Video recordings are performed on a green background (green-box), therefore correct extraction of foreground objects (so called "keying") is significantly aided by employing colour thresholding algorithm (Fig. 4).

In applications meant for real-time operation in arbitrary conditions, the background removal/object detection procedure is performed by far more complex algorithms of background estimation, e.g. modelling of pixel's color statistical properties, representing it as a mixture of Gaussians with iteratively adjustable means and deviations in a RGB color space [18], what lies beyond the scope of this paper.

3.2 Motion capture data acquisition

For registration of the reference 3D data of actor's poses a Motion Capture system OptiTrac [21] is used synchronized with video cameras. A 25 markers setup is used, aimed at reading positions of main body joints, omitting fingers and face expression.

Obtained data are exported to widely used BVH (Biovision hierarchy) format, facilitating data processing, storage, and visualization in popular 3D animation software. The file header starts with a keyword HIERARCHY and contains declaration of a virtual skeleton hierarchy, i.e. locations and lengths of bones (OFFSET), available degrees of freedom (CHANNELS) and relation parent-child between bones (JOINT). After the header a MOTION section starts, first containing length of a

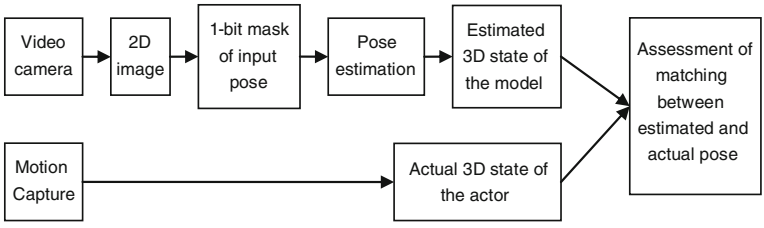

Fig. 2 Block diagram of body estimation quality assessment, based on reference data registered by Motion Capture system

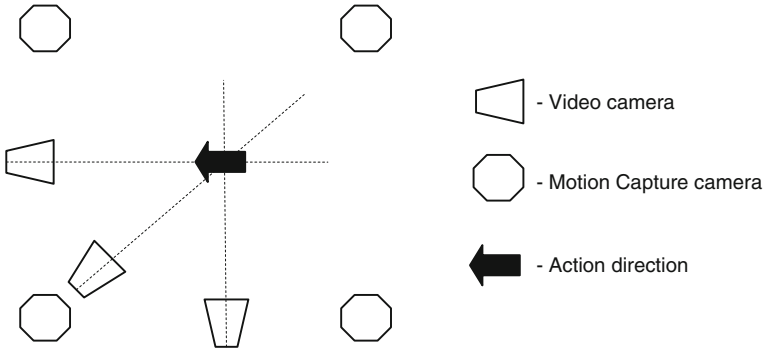

Fig. 3 Spatial configuration of the recording setup (*top view*)

sequence (e.g. Frames: 469), then length of a single frame in seconds (e.g. Frame Time: 0.016667, which can be converted to 60 frames per second). Finally, data for each captured frame for each bone DOF follow separated by a tabulator, and new line sign at the end of frame description (Fig. 5).

3.3 Database structure

For organization and storage of acquired data an Open Source MySQL 5.1 database is used, with InnoDB storage engine, running under Ubuntu Linux 11.04 operating system. Created relational database comprises of 5 tables with columns and relations presented in Fig. 6.

Table "Action" contains data from each action registered by any number of cameras. Technical parameters of cameras are stored in a table "Camera", containing foreign key "action_id", referring various cameras to particular "Action" recorded with them. Table "Frame_data_bin" for each "Camera" stores registered video frames as an original images and 1-bit images of object masks. For each "Action" and for each frame individually (identified by "frame_no") a table "Data_bvh" contains captured reference motion "data_bvh" (format compliant with BVH). Table "Skeleton_bvh" stores description of a skeleton hierarchy for particular registered "Action".

Such architecture allows for concurrent storage of actions recorded with any number and types of video cameras and any type of Motion Capture setup (as long as skeleton hierarchy and motion descriptions are provided in the BVH format). This approach also allows for integration of other datasets, both monocular and multi-camera, such as HumanEVA [17], inside one database.

```
1: for (frame from sequence)
2:    Split image into color components: red, green, blue, stored as a 8-bit words
3:    for (pixels of the image)
4:          if (green > 128 AND red < 80 AND blue < 80) pixel = 0 #pixel is the background
5:          else pixel = 1 #pixel in an object
6:    end
7: end
```

Fig. 4 Pseudo-code of image preprocessing: keying of green background and object detection, resulting in 1-bit image containing mask of the object

```
HIERARCHY
ROOT Hips
{     OFFSET   0.000000 0.000000 0.000000
      CHANNELS 6 Xposition Yposition Zposition Zrotation Xrotation Yrotation
      JOINT LeftHip
      {        OFFSET   4.740440 -3.195970-0.322709
               CHANNELS 3 Zrotation Xrotation Yrotation […]
MOTION
Frames:     469
Frame Time: 0.016667
0.030173        8.158459 0.074069  -0.154345        0.080239 -0.3143820.124953 […]
```

Fig. 5 Listing of a BVH file (abridged): section of skeleton definition and motion data

4 Genetic programming extension to APF-based pose estimation

Proposed new body pose estimation algorithm is based on genetic evolution of 3D models population tested against current 2D silhouettes from single camera. The models generation and fitness testing are performed with respect to evolutionary programming paradigm [1, 12], utilizing genetic algorithm extended with new concept of "genetic memory", and combined with APF for additional optimization.

In the proposed model-to-object matching method following ideas are employed:

– the matching procedure should be performed in two stages:

- first, the local optima of higher hierarchy of the body are sought (torso, head, i.e. the parts influencing location of arms),
- next, for each found optimum a second search run is performed, considering also lower hierarchy of the body parts (forearms, arms, legs, and hands),

– the observed motion (model and object state changes in consecutive frames) is continuous and fluent, abrupt speed vector changes are not plausible (yet still are considered as possible),

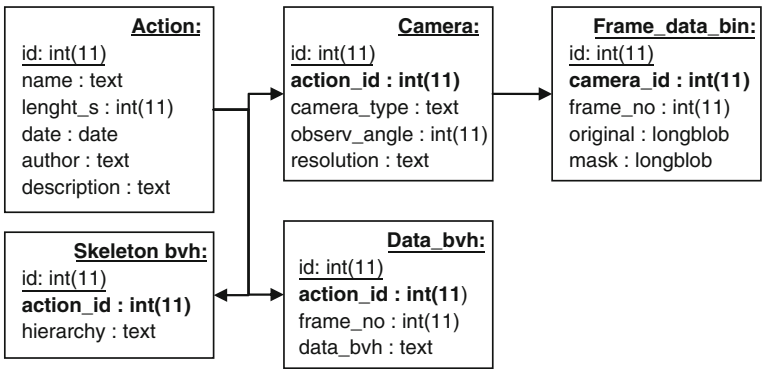

Fig. 6 Block diagram of database structure and relations. Foreign keys are marked with bold font

– therefore, for selection of best estimates during genetic optimum search, the history of motion and current estimated motion should be considered,
– the history of motion is inscribed into "genetic memory" of the object
– utilizing genetic operators of cross-over and mutation a new, possibly better, estimates can be obtained based on previous estimates,
– the process is repeated until a criteria of matching between model and observation is fulfilled.

The concepts introduced above are summarized in the following subsections.

4.1 3D model of human body

Implemented body model consists of 17 elements, modelled as balls and cuboids, some with limited Degrees of Freedom (DOF), 40 DOFs total. The structure of the model is presented in Figs. 7 and 8, and Table 1. Model state is described with g=40 values (genes) of current state, subject to modification by genetic algorithm and optimum search, and $H{\cdot}g$ values of H-long history of previous states ("genetic history"), which is neither crossed-over nor mutated. The model state can be described as:

$$\mathbf{X} = \left\{ x_{1,0}, x_{2,0}, \ldots x_{g,0}; x_{1,1}, x_{2,1}, \ldots x_{g,1}; \ldots; x_{1,H}, x_{2,H}, \ldots x_{g,H} \right\} \tag{4}$$

where: the subscript 0 depicts current moment in the history,

genes $x_{1,j} \div x_{16,j}$	higher hierarchy,
genes $x_{17,j} \div x_{30,j}$	lower body (legs)
genes $x_{31,j} \div x_{40,j}$	lower hierarchy,
$j = \{0,1,\ldots,H)$	age in the genetic history.

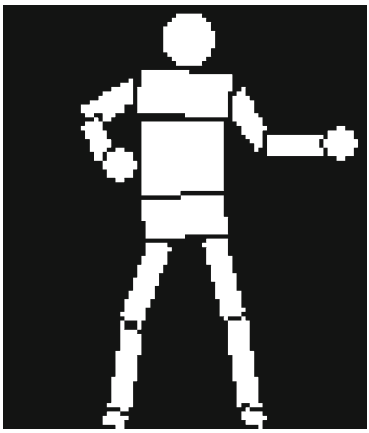

Fig. 7 3D model of human body (segment edges only for visualization, not present in actual model mask)

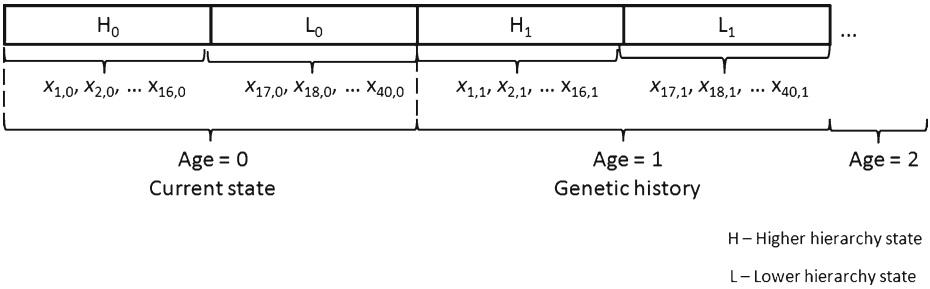

Fig. 8 Single object structure: model's state with "genetic history" of previous states

4.2 Genetic fitness function

Used matching metric is an extension of standard approach with shape and edge coverage metrics, Eq. (3). A new addition was made, considering "motion cost" $v(\mathbf{X})$, related to motion dynamics and movement speed changes (5):

$$w(X,Z) = w^r(X,Z) \cdot w^e(X,Z) \cdot v(X) \tag{5}$$

where:

$$v(X) = \exp\left(-\frac{1}{N}\sum_{i=1}^{N}\sum_{h=1}^{H}\left(v_{0,i} - v_{h,i}\right)^2\right) \tag{6}$$

where:

N number of bones analysed in current hierarchy level

Table 1 Parts of the modelled body and angle limits for joints (degrees)

		Rotations						
		α		β		γ		
Body part	Model	α_{min}	α_{max}	β_{min}	β_{max}	γ_{min}	γ_{max}	Hierarchy
Hip	Cuboid	−180	180	−180	180	−180	180	Higher
Abdomen	Cuboid	−15	15	−110	15	−15	15	Higher
Torso	Cuboid	−5	5	Fixed		−5	5	Higher
Neck	Cuboid	−15	15	−30	30	−50	50	Higher
Head	Ball	−45	45	−30	15	Fixed		Higher
Thigh ×2	Cuboid	−15	90	70	125	75	75	n/a
Calf ×2	Cuboid	Fixed		145	0	Fixed		n/a
Foot ×2	Cuboid	−15	15	−45	30	−15	15	n/a
Forearm ×2	Cuboid	Fixed		Fixed		−150	0	Lower
Arm ×2	Cuboid	0	360	0	360	−155	35	Lower
Hand ×2	Ball	−20	20	Fixed		Fixed		Lower
Hip location in space:		*x*		*y*		*z*		Higher

40 values are used to describe full state of the model

$v_{0,i}$ current angular motion speed for i-th value of state model calculated as $x_{i,0}$–$x_{i,1}$ (the time span is 1 frame, therefore no denominator for speed calculation is written)

$v_{h,i}$ historical angular motion speed for i-th value of state model calculated as $x_{i,h}$–$x_{i,h-1}$.

Presented metric (5) is used as a fitness function for evolutionary processing. $w(\mathbf{X},\mathbf{Z}) \in (0,1\rangle$, and the perfect match is obtained when $w(\mathbf{X},\mathbf{Z})=1$.

4.3 Crossing-over and mutation of the model states

All model states are aged (Fig. 9), and then selected two model states **X** and **Y** are crossed-over by exchanging one, randomly selected i-th value ($i \in \langle 1; g=40\rangle$) from current state $x_{i,0}$ with also i-th value of the other state $y_{i,0}$. The mutation is performed on randomly selected i-th value from current state, changing it by random value $\Delta \in \langle -5;5\rangle$ (Fig. 10), with the requirement that the result stays in allowed angle limit for joints (Table 1).

4.4 Hierarchical pose matching

Hierarchical matching of 3D model to 2D shape of the body is performed by successive consideration of model parts, starting high in the hierarchy, proceeding to lower levels. In each run the model is simplified to represent only the parts that are on the current hierarchy level (Fig. 11). The algorithm performs following steps (also show on block diagram in Fig. 12):

1. N random objects (model pose estimations) are generated. Initially the history contains static pose, i.e. for every i: $x_{i,0} = x_{i,1} = x_{i,2} = \ldots = x_{i,H}$.
2. Each estimate is evaluated utilizing genetic fitness function, Eq. (5) (shape and edges coverage, and motion cost) for higher hierarchy of the body.
3. $M=2$ stages of APF are performed ($\beta_0=1$, $\beta_1=0.25$, $\beta_2=0.1$) for local optima search over the fitness function. N particles in 16-dimensional space are used (genes $x_{1,j} \div x_{16,j}$ describing higher hierarchy), initialized with current state of higher hierarchy model, and adjusting only the current state (not the "genetic history"). Head and shoulders shape is very distinct and optimum search converges easily (Fig. 11b).
4. Estimates are ranked and $N' < N$ best estimates (APF particles) are selected.

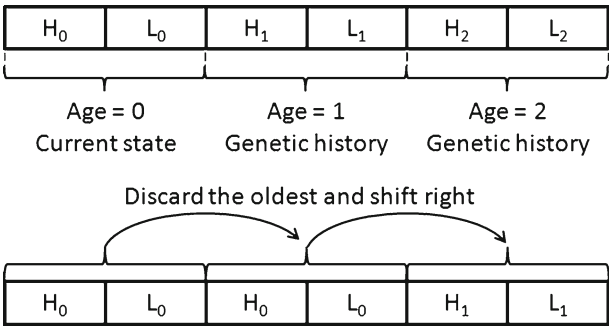

Fig. 9 Aging process

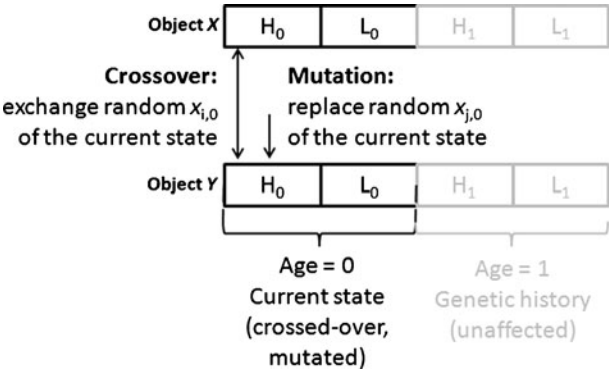

Fig. 10 Genetic crossover and mutation

5. Utilizing each of N' best estimates all (N–N') worse estimates are readjusted, by substituting $x_{1,0} \div x_{16,0}$ genes in the state with genes of randomly selected higher hierarchy estimate. Probability of selection is proportional to the value of estimate $w(\mathbf{X},\mathbf{Z})$ calculated in APF in step 3. The results are N readjusted objects with well fitting higher hierarchy of the body.
6. Optimum search is then performed with M=2 stages of APF with N particles in 10-dimenstional space of lower hierarchy bones. Local optima of those bones rotations are found (Fig. 11c).
7. Each estimate is evaluated utilizing genetic fitness function (shape end edges coverage and motion cost).
8. Best $L \geq 1$ estimates are presented on the screen to the user and compared with reference Motion Capture data for subjective rating of the pose and for algorithm benchmarking.
9. All N estimates from step 7 are aged: in the history the last state of age H is removed, other states are shifted right by g cells. New state of the model is created by crossing-over (with probability 0.5) all worse (N–L) estimates with randomly selected one of L best ones. Finally the mutation of the current state is performed (with probability 0.1). It is then taken as a starting point for matching the model to next video frame.
10. The process repeats from step 2.

Steps 3–6 are presented in a graphical form on Fig. 13.

Fig. 11 Hierarchical matching of 3D model to 2D observation of human body: **a** sample pose, **b** 3D model posed by first run (high hierarchy) and the difference between pose and model shape, **c** 3D model posed by second run (lower hierarchy) and the difference between pose and model shape

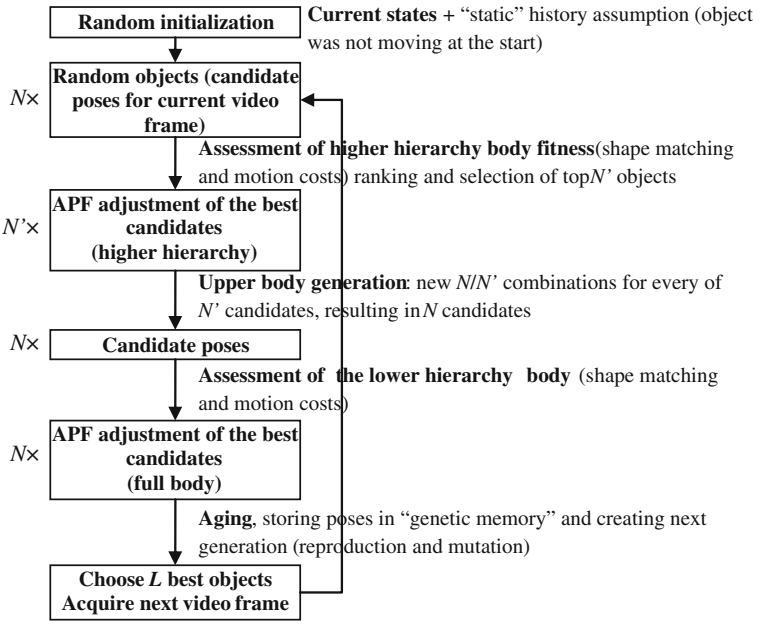

Fig.12 Block diagram of the evolutionary algorithm (description in the text)

The algorithm is implemented in C++ utilizing own pose generation library (it provides binary image of 3D model silhouette based on its state description and camera position) and

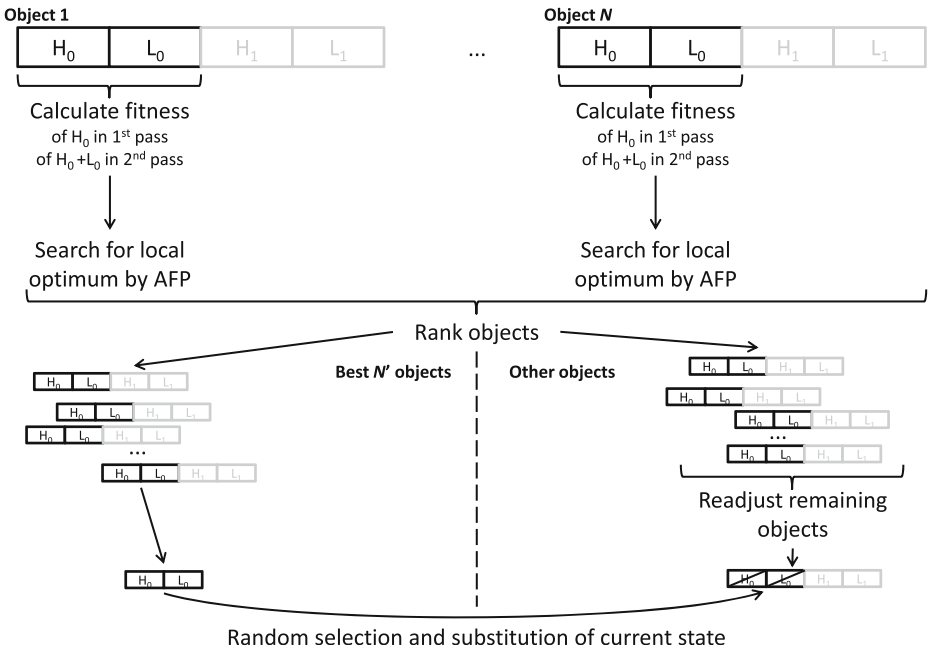

Fig. 13 APF optimization of *N* objects extended with ranking and readjustment of the worst ones

OpenCV library [15] for image processing (normalization, matching measure calculation, result visualization).

5 Algorithm evaluation

For the experiment and objective assessment of the results a set of 10 poses was prepared utilizing 3D model posed by hand, each accompanied with $H=3$ long genetic history of motion, comprised of 3 poses before the target pose (Fig. 14).

Current poses and their histories were saved in designed database as bitmap files, and supplemented with the BVH-formatted reference data in a form of bones angles values for particular poses. Then the pose estimation algorithm was initialized with T-shape pose (all angles equal to zero) and performed higher and lower hierarchy estimation with respect to designed algorithm. After the pose estimation the model state was compared to respective reference data and cumulative Sum of Squared Differences (SSD) of angles for all bones was calculated to assess the pose estimate.

In the experiment the following values were used: number of objects $N=60$, best $N'=6$ objects were selected for reproduction, motion history length $H=3$, $M=2$ stages APF was performed, and best $L=1$ object was used for comparison to reference values. The genetic optimization process was performed for 100 iterations, as this value assured reasonable processing time of ca. 60 s (10milliseconds for rendering and calculating the metric for a single candidate pose). Relation between those parameters and computation time is straightforward, therefore more precise or more coarse analysis can be performed in particular time constrains. Moreover, any number of N objects can be divided into groups processed in separate threads (parallel calculations on multi-core CPU) for a significant improvement, which is the goal for the next implementation.

Obtained results of estimation of 10 poses (Fig. 15) employing the genetic modification of APF with 3 step history, and hierarchical matching are presented in Table 2. For reference the same poses were estimated utilizing APF method, executed until the SSD error decreased below the one achieved by Genetic APF.

If the arms are stretched away from the torso (poses 1÷5 in Table 2), then the higher hierarchy matching result is high, due to precise localization of the torso and head shape. Then, for the whole body, errors of shape and edges matching higher and lower hierarchy (for torso and arms) sums, therefore decreasing total matching result.

Contrary, if in the pose shape arms connect to the torso (poses 7÷10 in Table 2), then the matching obtained in the first stage of higher hierarchy estimation is low, because of the attempt of matching "handless" model to full body shape. Then, the matching value

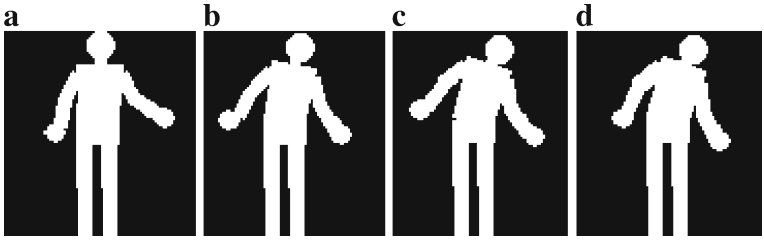

Fig.14 One of the analysed poses with $H=3$ long history: **a**–**c** 3 previous poses stored in the history, **d** current estimated pose

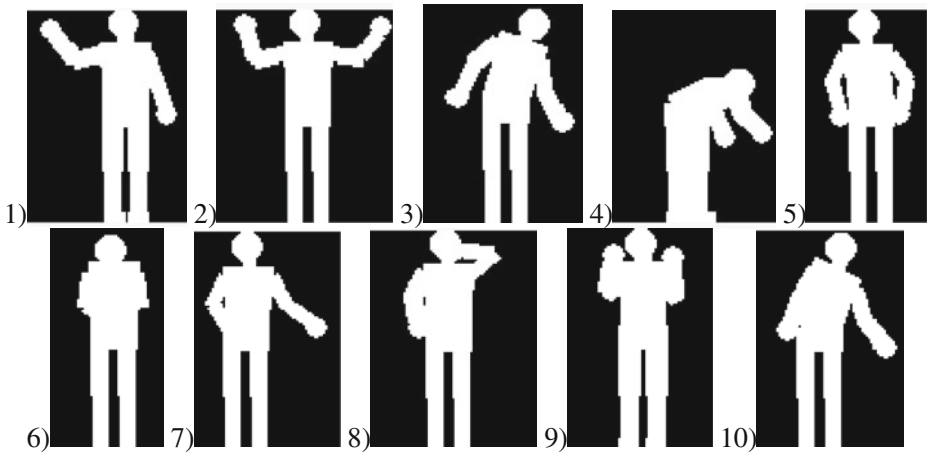

Fig. 15 10 test poses: 1÷5—arms stretch away from the torso, 6—arms embrace the torso, 7÷10 arm or two arms connect the torso

increases in the second stage, when full model is used and the hands positioned correctly provide correct matching of shape and edges.

The least effective matching process was observed for pose 6, where the arms are embracing the torso, and large degree on ambiguity is present, as very low information is contained in the shape. This type of the pose (self occlusion, limbs very close to the torso) stays currently as the main challenge in pose estimation research.

Poses 1 and 2 were created with abrupt motion change comparing to the historical poses, therefore the matching result is lowered despite average SSD values. More thorough experiments will be conducted for precise determination of the correct influence of motion cont $v(\mathbf{X})$ on total $w(\mathbf{X},\mathbf{Z})$. Currently shape, edges, and motion cost metrics are considered as

Table 2 Results of hierarchical matching of body model for various poses

Pose	Higher hierarchy matching $w(\mathbf{X},\mathbf{Z})$	Lower hierarchy matching $w(\mathbf{X},\mathbf{Z})$	Bones angles errors: SSD	Number of iterations	
				Genetic APF	APF
1	0.900	0.852	11.72	150	230
2	0.915	0.874	11.08	150	322
3	0.944	0.903	13.14	150	412
4	0.943	0.885	16.78	150	312
5	0.954	0.930	8.95	150	474
6	0.919	0.832	24.80	150	252
7	0.883	0.961	3.48	150	528
8	0.927	0.977	2.31	150	488
9	0.947	0.979	8.30	150	385
10	0.880	0.965	2.49	150	435

SSD is a Sum of Squared Differences of bones angles compared to the reference values. A number of iterations of genetically modified APF and standard APF methods are presented. APF was executed until the SSD error decreased below the one achieved by Genetic APF

equally important, while for longer sequences and histories this approach may lead to motion continuity preference over shape estimation precision.

6 Summary

Hybrid genetic-APF method was proposed and tested. New metric for model-to-object matching was proposed employing motion dynamics and movement speed changes. The concept of "genetic memory" was introduced, facilitating processing of the motion history and accounting the history in estimate fitness measurement. The genetic crossing-over operator forces APF algorithm to assess other modes of the model matching metric, and genetic mutation introduces randomness, important for avoiding local optima. The proposed algorithm can be used for other body hierarchies with more than two hierarchy levels, and various genetic history lengths H.

The future work will focus on optimization by means of parallelization of the genetic algorithm, i.e. splitting of objects set into groups analysed by separate threads for multi-core CPUs, e.g. supercomputer clusters. Also an implementation of limbs occlusions handling is planned. Moreover, the matching metric will be further extended by introducing pixel-level motions (e.g. based on Optical Flow or Motion History Imaging).

Acknowledgments Research funded within the project No. POIG.02.03.03-00-008/08, entitled "MAYDAY EURO 2012—the supercomputer platform of context-depended analysis of multimedia data streams for identifying specified objects or safety threats" subsidized by the European regional development fund and by the Polish State budget.

References

1. Bäck T (1996) Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. Oxford University Press
2. CONTOUR: Markerless Motion Capture System: http://www.mova.com (accessed: 25-04-2012)
3. Czyżewski A, Ellwart D (2010) Camera angle invariant shape recognition in surveillance systems. Proc. KES IIMSS 2010, Baltimore, USA
4. Deutscher J, Blake A, Reid ID (2000) Articulated body motion capture by annealed particle filtering. Proc. IEEE Conf. on Computer Vis. and Pattern Recognit. 126–133
5. Efros AA, Berg AC, Mori G, Malik J (2003) Recognizing action at a distance. In: 9th Inter. Conf. Computer Vis. 726–733, Nice, France
6. Gavrila DM, Davis LS (1996) 3D model-based tracking of humans in action: A multi-view approach. Proc. IEEE Computer Vis. and Pattern Recognit. (CVPR'96), 73–80
7. Isard M, Blake A (1998) CONDENSATION—conditional density propagation for visual tracking. Int J Comput Vis 29(1):5–28
8. Kakadiaris I, Metaxas D (2000) Model-based estimation of 3D human motion. IEEE Trans Pattern Anal Mach Intell 22(12):1453–1459
9. Kehl R, Van Gool L (2006) Markerless tracking of complex human motions from multiple views. Comput Vis Image Underst 104(2–3):190–209
10. Lech M, Kostek B (2010) Fuzzy rule-based dynamic gesture recognition employing camera & multimedia projector. Proc Int Conf Multimed Netw Inf Syst
11. Lienhart R, Maydt J (2002) An extended set of haar-like features for rapid object detection. Proc IEEE ICIP 1:900–903

12. Michalewicz Z (1998) Genetic Algorithms+Data Structures=Evolution Programs. Springer
13. Moeslund TB, Hilton A, Kruger V (2006) A survey of advances in vision-based human motion capture and analysis. Comput Vis Image Underst 104(2–3):90–126
14. Ong EJ, Micilotta AS, Bowden R, Hilton A (2006) Viewpoint invariant exemplar-based 3D human tracking. Comput Vis Image Underst 104(2–3):178–189
15. OpenCV Image Processing and Compute Vision Library: http://opencv.willowgarage.com (accessed: 25-04-2012)
16. Sidenbladh H, Black MJ, Fleet DJ (2000) Stochastic tracking of 3D human figures using 2D image motion. Proc 6th Eur. Conf. on Computer Vis., 702–718
17. Sigal L, Balan A, Black MJ (2010) HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. Int J Comput Vis 87:4–27
18. Stauffer C, Grimson W (1999) Adaptive background mixture models for real-time tracking. Proc. CVPR, 246 252
19. Vlasic D, Baran I, Matusik W, Popović J (2008) Articulated mesh animation from multi-view silhouettes. ACM Trans Graph 27(3). http://people.csail.mit.edu/drdaniel/mesh_animation/ (accessed: 25-04-2012)
20. Wachter S, Nagel H (1999) Tracking persons in monocular image sequences. Comput Vis Image Underst 74(3):174–192
21. WWW NaturalPoint OptiTrac: http://www.naturalpoint.com/optitrack/products/motion-capture/ (accessed: 25-04-2012)

Piotr Szczuko received his Ph. D. in 2008. Currently he is an assistant professor in the Multimedia Systems Department of Gdansk University of Technology. His research focuses on video processing for automatic human action classification and event recognition. He is also interested in computer animation, machine vision, artificial intelligence and automatic inference methods.