Reconstructing Motion Capture Data for Human Crowd Study

Samuel Lemercier[1], Mathieu Moreau[2], Mehdi Moussaïd[2], Guy Theraulaz[2], Stéphane Donikian[1,3], and Julien Pettré[1]

[1] INRIA Rennes - Bretagne Atlantique
Campus de Beaulieu, 35042 Rennes, France
[2] CRCA, CNRS, 31062 Toulouse Cedex 9, France
[3] Golaem
samuel.lemercier@inria.fr

Abstract. Reconstruction is a key step of the motion capture process. The quality of motion data first results from the quality of raw data. However, it also depends on the motion reconstruction step, especially when raw data suffer markers losses or noise due, for example, to challenging conditions of capture. Labeling is a final and crucial data reconstruction step that enables practical use of motion data (e.g., analysis). The lower the data quality, the more time consuming and tedious the labeling step, because human intervention cannot be avoided: he has to manually indicate markers label each time a loss of the marker in time occurs. In the context of crowd study, we faced such situation when we performed experiments on the locomotion of groups of people. Data reconstruction poses several problems such as markers labeling, interpolation and mean position computation. While Vicon IQ software has difficulties to automatically label markers for the crowd experiment we carried out, we propose a specific method to label our data and estimate participants mean positions with incomplete data.

1 Introduction

Optoelectronic motion capture systems are among the most precise tools to track human movements. For this reason, we used motion capture techniques on small crowds of pedestrians for the purpose of experimental studies. More precisely, we captured up to 28 volunteers walking along a circular lane delimited by walls as illustrated in Figure 1 (Right). They were tracked with a 12-camera Vicon System placed all around the observation area. However, raw motion data often suffer some occlusion as each marker has to be viewed from 2 or 3 cameras to be located in 3D which led to incomplete data.

Indeed, participants concealed their markers each other when passing beside a camera. Markers also got hidden by walls. The number of cameras was limited and the experimental area was large: some parts were covered by only two cameras which is a strict minimum to enable capture. We obtained incomplete data, particularly with high density trials (approximately 33% of data loss when

J.M. Allbeck and P. Faloutsos (Eds.): MIG 2011, LNCS 7060, pp. 365–376, 2011.

Fig. 1. Left: Picture of the experimental system (empty). Right: Picture taken during the experiment with participants walking along a circular path delimited by a wall.

the 28 participants were walking along the inner wall). Moreover, due to the number of subjects, we were not able to use whole-body sets of markers to track each participant; that would have brought the total number of markers to more than 1000 markers in the trials with 28 participants, which would not have been possible to track. Consequently, we used of a limited set of 4 markers that did not enable the existing automatic reconstruction technique (provided with motion capture system) to correctly work on our dataset. This resulted in a very long manual labeling treatment and we opted for developing our own labeling technique with the aim of saving working time.

In this paper, our first objective is to automatically reconstruct the raw motion data we obtained from our experiments. We faced several difficulties: the temporary loss of some markers, the complete loss of some markers in time during the trial, the temporary loss of all the markers belonging to one participant, but also the presence of *ghost markers* (noise interpreted as markers by the system). Our second goal is to estimate the global position of each participant with a robust method given the reduced set of markers and the loss of data. Our contribution is an automatic method to meet these objectives. Details are provided in Section 3.

2 Related Work

We acquired our data using the VICON MX-40 motion capture system. Data Reconstruction with Vicon IQ software is based on a skeleton model of the tracked subject which defines the segments that connect the markers. The software is then able to detect when a segment appears and can automatically label the markers. This automatic process perfectly works when data quality is high enough and information sufficient to avoid ambiguities. In the case of our experiments, as all the subjects were only wearing four markers, the system was not able to differentiate subjects' skeletons and markers and this technique failed.

Our data also showed loss of markers due to occlusion. Different methods for filling markers occlusions have already been proposed. Li et al. [7] classify them into several categories: spline and linear interpolations (only used for short period occlusions), skeleton-based [5], dimensionality reduction and latent variables [11], database backed [1], Kalman filters [4], dynamical systems [2,7]. However, our problem differs in two main points. First, our markers have not been labeled yet and we are not able to find a marker once it has disappeared. Second we do not try to retrieve articular trajectories from a whole-body set of markers but attempt to estimate the global position of each participant from a reduced set of markers.

The objective of carrying out experimentations on pedestrian crowds are to study pedestrian behavior, enable the elaboration of pedestrian behavioral models, calibrate and validate these models by comparing real and simulated data [8]. The interest of using such a system is to acquire more precise data and to be able to track a pedestrian not only in a local area but in the whole observation system. Previously, several observations had already been carried out for crowd studies. Some natural observations have been made by Yamori [12] who placed a camera above a large pedestrian crosswalk. Experimental observations have also been processed to study pedestrian behavior at bottlenecks [6,9] or for more general situations [3]. However none of these experiments used an optoelectronic motion capture system to track pedestrians but only optic cameras. We here expect to obtain a more accurate dataset with the constraint to face challenging conditions for motion capture.

3 A Method for Automatic Labeling and Reconstruction of Crowd Motion Data

The objective of our method is to estimate the global position and motion of all the participants tracked during experiments from a soup of partial unlabeled markers trajectories. As previously said, the main difficulty is the quantity of data loss which can reach more than a third in the worst cases. Participants were wearing four markers: one on the head (called H), one on the left shoulder (called L), and two on the right shoulder (called R for the one the more on the right and N for the other one) to know their orientation, which still resulted in tracking up to 112 markers when 28 participants were captured.

From Motion Capture software, we obtain markers trajectories in a Cartesian system of coordinates which are rarely continuously tracked during a whole record. The software is not able to recognize a marker when it reappears after being lost: a "new marker" is then created which multiplies the final total number of markers and trajectories. Merging together these portions of trajectories is required to estimate continuous motion along a whole record. Two reasons explain markers occlusion: some observed areas were not well-covered by our camera system, and participants concealed their markers each other when passing beside a camera. Thereby, data quality decreases when participant density increases. We elaborated a method which is composed of the following steps:

- Alignment of motion data with the horizontal plane,
- Markers labeling,
- Estimation of pedestrian position.

3.1 Global Motion Data Transformation

We are mainly interested in the horizontal component of motion data. Before performing motion capture, the capture coordinate system is defined by tracking the so-called L-Frame the geometry of which is known to the system. Because our experimental area is large in comparison to this frame, a drift is observed: motion is not contained into a horizontal plane. To solve this question we transform motion data to align the floor as defined by the frame and the horizontal plane. We perform a multivariate linear regression of motion data to estimate the mean plane of motion. We assume this plane is aligned with the actual floor. We compute the angle formed by the vertical axis and the normal to the floor, and globally rotate data by this angle. Once this step done, motion data can be projected to the horizontal plane without biasing data by artificially introducing deformations. Next step however still consider vertical motion coordinate to automatically label markers.

3.2 Markers Labeling

The objective of this step is to regroup the markers by participant and then label each of them. From raw data, the position $\mathbf{Z}$ of a marker i at time t is described as follows:

$$\mathbf{Z}_i(t) = [x_i(t)\ y_i(t)\ z_i(t)] \tag{1}$$

As participants walk on a circular lane, we consider that their distance to the center of the system is approximately the same and constant in time. Thus, we switch into a polar system of coordinates which center is the center of the system. The position of a marker i at time t is now described as follows:

$$\mathbf{Z}_i^{cyl}(t) = [\theta_i(t)\ r_i(t)\ z_i(t)] \tag{2}$$

θ coordinate of markers' trajectory with respect to time is plotted in Figure 2. Note that discontinuity is due some marker occlusion and reappearance with reinitialized value $\theta \in [0, 2\pi]$. Plots show motion data of various qualities.

Removing ghost-markers. As well as markers occlusion problem, we observe ghost-markers, i.e. markers that are recognized by the system while they are not present. Two kinds of ghost-markers can appear: residual ones which last during a long period of time but which are very few, and temporary ones which only last a little time. We remove markers trajectories which last less than one second. This removes most of the ghost-markers and cleans the data without removing much information.

Grouping markers by participants. The goal of this step is to group markers trajectories together when they belong to a same participant. This is detected by their proximity along θ coordinate during all the time they are both detected. The distance between two markers i_1 and i_2 is defined (in m) as

$$\mathbf{d}_{i_1,i_2} = r * |\theta_{i_1} - \theta_{i_2}| \tag{3}$$

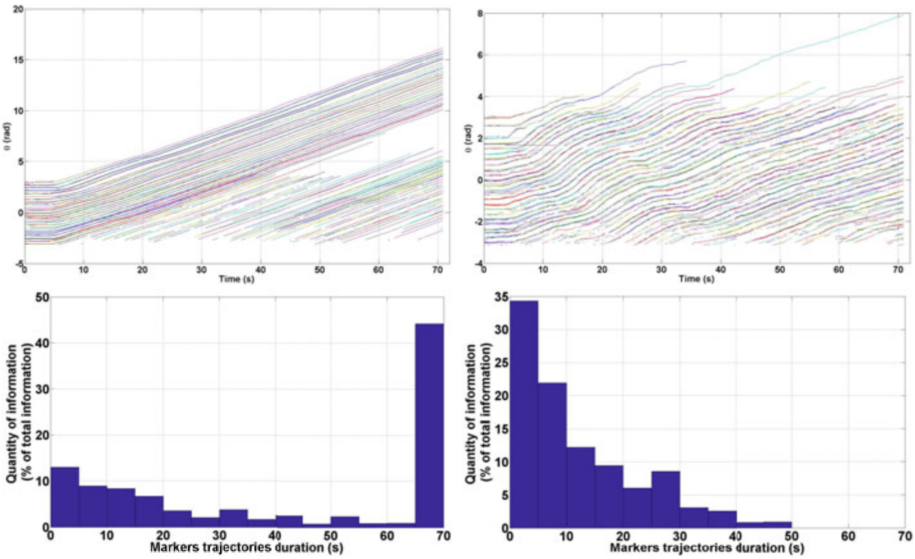

Fig. 2. Top: Markers positions along θ axis as a function of time for 2 trials with 24 pedestrians (Left: trial with good tracking conditions. Right: trial with bad tracking conditions). One can observe that many markers occlusions happen but some markers are tracked during all along the trial with no loss. Bottom: Repartition of information by markers trajectories duration for the same trials. Most of the information is contained in long (resp. short) markers trajectories on the left (resp. right) plot.

with r the mean radius of both markers. Markers should be grouped if and only if belonging to the same participant. Ambiguities occur in the noisiest parts of motion data. We apply two conservative rules to decide on markers grouping. Markers i_1 and i_2 are grouped if:

- the mean distance $\overline{\mathbf{d}_{i_1,i_2}}$ between markers during the period of time they are both detected is below $1.3 * 10^{-1}m$.
- th maximum distance $\mathbf{d}_{i_1,i_2}^{max}$ between markers is below $3.3 * 10^{-1}m$.

At the end of this step, participants' trajectory is described by the ones of a group of markers. The number of markers ranges from 1 up to 4 depending on occlusions. But in most challenging conditions, a complete loss of markers belonging to a same participant is observed. The trajectory of a participant can be described as a succession of separated pieces of group of markers. However, they are not labeled as belonging to the same participant. Thus, these pieces need to be merged and regrouped.

Matching trajectories. We merge some pieces of trajectories that do not overlap in time by using a linear extrapolation method, in time, on θ axis. We only make short-time ($< 2sec$) extrapolations to avoid risks of ambiguities.

Finally, there are still some different trajectories that must match, due to the fact that they overlap in time or do not satisfy the conditions of extrapolation. We manually merge these remaining trajectories thanks to their visual position in time on θ coordinate.

Labeling markers. Next task is to label the markers belonging to a same participant. Figure 3 (Top) shows markers positions along r (Left) and z (Right) coordinates.

First we identify H marker trajectories thanks to their upper position on z axis. Then we identify left-marker trajectories thanks to their position on r axis (This position is higher if the pedestrian walks in the anti-trigonometric way and lower if he walks in the trigonometric way). It is easier to detect L markers once head-markers have already been identified. Finally we identify N marker and R marker thanks to their relative positions on the r axis. Note that when too much information misses, these markers can be inverted, which is not a matter since they are actually close one each other and that the following estimation step palliate such inversion.

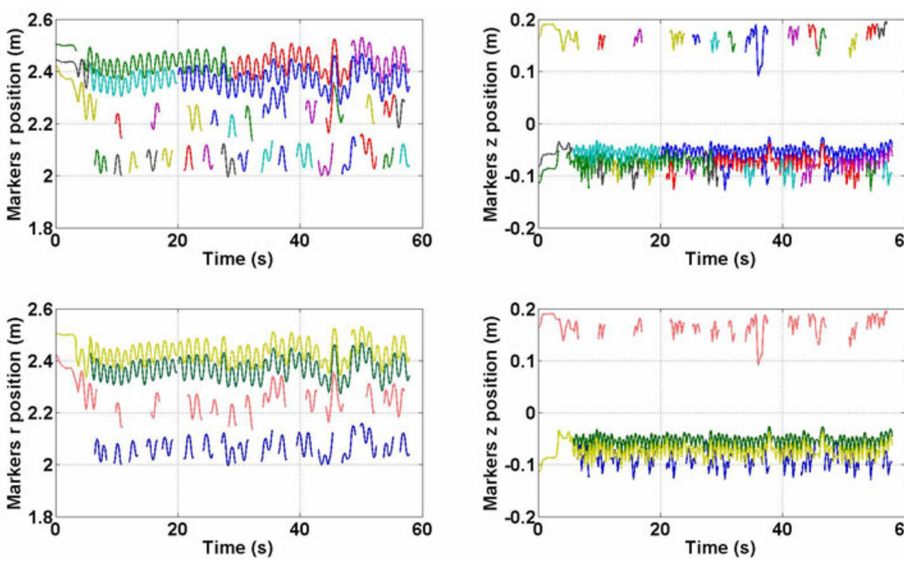

Fig. 3. Top: Unlabeled markers positions on r (Left) and z (Right) axes. Bottom: Identified markers positions on r (Left) and z (Right) axes. H, L, R and N marker are clearly observable.

3.3 Estimation of Participant Position

We aim at estimating the participant position $\mathbf{X}(t)$ at each time t whatever the number of markers describing the motion, and even if no data is available at this time. The method is based on the following hypothesis: a group of markers belonging to a same participant follows the motion of a solid in translation and rotation. Under this hypothesis, the position of a $i-$marker Z_i^*

$$\mathbf{Z}_i^*(t) = \mathbf{X}(t) + \mathbf{R}_z(\alpha(t)).\mathbf{L}_i \tag{4}$$

where $\mathbf{X}(t)$ is the pedestrian position at time t, $\alpha(t)$ is the pedestrian orientation angle i.e. the direction in which it is moving, $\mathbf{R}_z(\alpha)$ is the rotation matrix around z axis with a given angle α and $\mathbf{L}_i$ is the vector from the pedestrian position to the $i-$marker. We call the $\mathbf{L}_i$ rigid skeleton of the pedestrian. Pedestrian position, orientation and skeleton are the model's parameters and should be evaluated through the markers positions measurements $\mathbf{Z}_i(t)$. We choose to explicitly estimate $\mathbf{L}_i$ and α to obtain a well-posed linear (easy to solve) problem for finding the pedestrian position.

Estimation of $\boldsymbol{L_i}$. The skeleton is calculated in three steps:

- centering of measurements $\mathbf{Z}'_i = \mathbf{Z}_i - \frac{1}{4}\sum_{j=H,R,L,N}\mathbf{Z}_j$, where the subscripts H, R , L, N refer to the markers names given above.
- rotation of the $\mathbf{Z}'_i$ to align $(\mathbf{Z}'_L - \mathbf{Z}'_N) \times (\mathbf{Z}'_R - \mathbf{Z}'_N)$ on the y axis.
- the skeleton points $\mathbf{L}_L$, $\mathbf{L}_R$, $\mathbf{L}_H$ and $\mathbf{L}_N$ are respectively the median points of the rotated $\mathbf{Z}'_L$, $\mathbf{Z}'_R$, $\mathbf{Z}'_H$, $\mathbf{Z}'_N$.

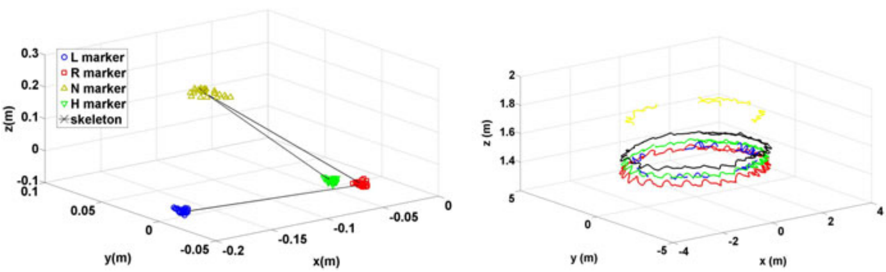

Fig. 4. Left: Example of an evaluated pedestrian skeleton using four markers. Right: A reconstructed pedestrian position over time obtained from its H, L, R and N markers. Only 5% of the markers positions are drawn for representation purpose.

An example of a pedestrian skeleton is shown in Figure 4 (Left). If the markers of one pedestrian are never visible all together at the same time, we use a mean skeleton over the other pedestrians.

Assumptions on the skeleton motion. During a given time interval T, the pedestrian position coordinates time-evolution can be approximated by a polynomial function:

$$\mathbf{X}(t+t') = \sum_{k=1}^{3}\sum_{j=1}^{N_k} U_{j+k(N_k-1)} t'^{(j-1)}\mathbf{e}_k \text{ with } t' \in \left[-\frac{T}{2}, \frac{T}{2}\right] \tag{5}$$

where N_k are the polynomial degrees in the different $k-$directions (i.e. x, y, and z), U_j its coefficients and $\mathbf{e}_k$ the corresponding unit vectors. This equation is a truncated Taylor series of $\mathbf{X}(t)$ around t, where $(U_{1+k(N_k-1)})_{k=1..3}$ is the position X at time t so that the approximation is only valid for small duration T. Introducing this expression into eq. 4 and using the explicit approximations for the skeleton $\mathbf{L}_i$ and orientation $\alpha(t)$ leads to a linear set of equations in term of U_j.

Estimation of participant orientation. It is assumed that the participants follow a rotational motion. Thus, their orientation is simply given by the direction of θ axis according to their positions. An orientation is associated to each measurement $\mathbf{Z}_i$, then $\alpha(t)$ is evaluated by a linear fitting of this orientation, assuming a constant velocity during the duration T.

Numerical resolution. We use a standard linear least square method to evaluate the U_j coefficients. The number of unknown parameters values is $\sum_{k=1}^{3} N_k$ and the number of equations is three times the number of measurements during the time interval T, which is denoted by $N_T(t)$. Several precautions must be taken in the choice of T. First, from a numerical point of view, to solve the set of equations and prevent numerical instabilities, the duration T must be small enough to ensure the validity of equation 5 and large enough to ensure that the problem is overdetermined. Second, it is well known that a least square fitting procedure could give inaccurate results in case of unbalanced data, i.e. when all available data are on the same side of the point of interest.

To circumvent such limitations, we choose an initial value of T: $T_0 = 8.33 * 10^{-3}s$, i.e. the frame rate at which the data were recorded. If all the data are on the same side or $N_T < 10 * max_{k=1..3}(N_k)$, we increment the value of T by T_0.

Sharp variations in $T(t)$ could results in sharp variation in the estimation of the pedestrian position. To avoid such a problem and ensure $\mathbf{X}(t)$ continuity, we detect the jump in $T(t)$ and overestimate $T(t)$ before or after the jump depending on the variation sign.

Interpolation using neighbors. When all the markers of one participant are occluded for a long period of time ($>$ 2 sec), previous approximation can cause major difference between estimated and real positions. In such cases, we use the fact that participants are walking in lane and prefer to use the positions of the participant neighbors (previous and next) to interpolate his position on θ coordinate. We define the ratio $R(t)$ as:

$$R(t) = \frac{\theta_i(t) - \theta_{i+1}(t)}{\theta_{i-1}(t) - \theta_{i+1}(t)} \tag{6}$$

When $t \in [t_s, t_e]$ the time gap where no data is available for the considered participant, we linearly interpolate $R(t)$ as follows:

$$\widetilde{R}(t) = R(t_s) + (R(t_e) - R(t_s)) * \frac{t - t_s}{t_e - t_s} \tag{7}$$

Thus we get participant θ position:

$$\theta_i(t) = \theta_{i+1}(t) + R(t) * (\theta_{i-1}(t) - \theta_{i+1}(t)) \tag{8}$$

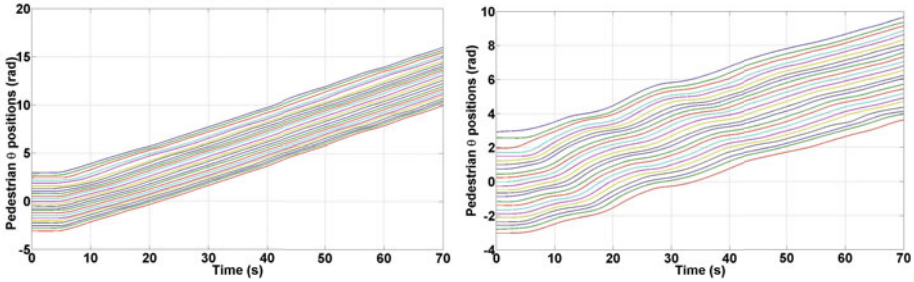

Fig. 5. Plot of resulting pedestrians trajectories corresponding to the ones of Fig 2

4 Results and Evaluation

4.1 Validation of the Pedestrian Position Estimation

To validate the method presented in section 3.3, the algorithm is applied to trials with best tracking conditions. We use an experimental record with 16 pedestrians walking in lane around a two meters radius circle. For 3 of these 16 participants, measurements of the four markers are fully available (i.e., at each time step). For these 3 participants, we compare the reconstructed position to the mean of the four markers instantaneous positions. Each trajectory of approximately 54 seconds consists in 6427 3-dimensional positions for each of the four markers. Reconstructions and evaluations are performed for all discrete time values. The method validation is performed on the accuracy of position reconstruction and on the robustness to the lack of measurements.

Comparison of results to mean positions of markers. With the following numerical parameters : $T_0 = 0.1s$, $N_1 = 2$, $N_2 = 2$, $N_3 = 1$, the mean distance between positions obtained by the reconstruction method and the mean markers one is $1.4 * 10^{-3}m$ and the maximum distance is $2 * 10^{-2}m$. The new method gives results similar to the one of the mean marker method in fully-available markers position cases. Different parameters could not significantly improve this margin which is attributed to the skeleton evaluation step, errors being of the same order of magnitude. Numerical tests (not presented here) show that this error level is reached for a large range of parameters N_k and T_0, for example, with ($N_k \in [2, 20]$ and $T_0 = 0.1$ s) or ($N_k \in [4, 10]$ and $T_0 = 0.1$ s). This robustness to numerical parameters is attributed to the dynamical procedure for the T fitting duration computation which links T to T_0 and N_k. We use the parameter values listed above in the following subsection to make the parameter fitting as local as possible to improve the robustness to errors on markers position.

Robustness to marker loss. The robustness of the position estimation method to marker loss is investigated. To mimic markers loss, we work from high quality data again and remove some measurements using the following two states Markov chain model describing the discrete time-evolution of the marker visibility: state 1 : the marker is visible and state 0 the marker is invisible. The probability of jumping from state 0 to state 1 is dt/τ, with τ the life-time of a marker disappearing. The probability of jumping from state 1 to state 0 is $a\ dt/((1-a)\tau)$, with a the proportion time of loss. The initial state of this model is that the marker is visible with a probability of $1-a$ and corresponds to the mean steady solution of the model. The accuracy of this model in the range of tested parameter has been checked (results are not presented here). This model is successively applied to the measurements of each marker of the three fully available participants' trajectories. The degraded data are then processed by the reconstruction algorithm and compared to the mean marker method with fully-available markers data which are used as a reference to evaluate the method. Twenty replicates of the three trajectories are performed to compute mean error of the reconstruction method.

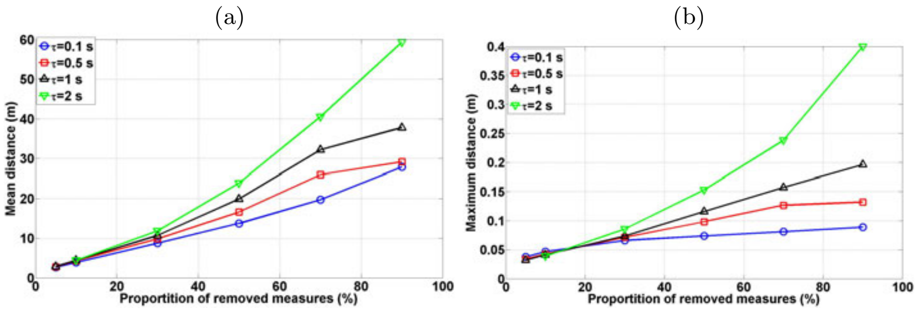

Fig. 6. Evolution of the distance between reconstructed positions and the reference method with the proportion of removed measurements for different disappearing life time τ. (a) Mean values and (b) maximal values.

Figure 6 shows the evolution of the distance between the position reconstructed using data with removed data and the reference position for different a and τ parameters: from 5% to 90% data removed during a duration from 0.1 s to 2 consecutive seconds. As expected, the distance error increases with increasing of the lack of data and with the time that a marker stay continuously invisible but its means remain lower than $6*10^{-2}m$ (Fig. 6a). Such a mean error level is smaller than a human's width (approximately $5.50*10^{-1}m$ [10]) and validates the algorithm to accurately track the participants. We should notice that the maximal distance error is larger than $3*10^{-1}m$ for $(a > 0.7)\&(\tau \geq 2s)$, so that this reconstruction is problematic at some time or under specific conditions. When the reconstruction is performed on such cases, an additional attention must be provided to access the validity of the participant trajectory evaluation. This error estimation versus the quality of the measures helps us to a priori determine whether the collected data are sufficient or too bad to be analyzed.

4.2 Evaluation of the Overall Method

As we have trials of different quality, we evaluate our method as follows:

– From the trial with good tracking conditions presented in Figure 2, we randomly remove information in a way to have both the same quantity of missing information and the same distribution of markers lengths than on the trial with bad tracking conditions.
– We run our method on this new trial.
– We compare its resulting participants trajectories with the ones obtained with the original trial.

We observe a mean difference of position on θ coordinate of $1.13 * 10^{-2} rad$ ($std = 7.30 * 10^{-3} rad$, $max = 5.02 * 10^{-2} rad$) which is very low and superimposes the original trajectories. A higher mean difference is observed on r coordinate: $6.73 * 10^{-2} m$ ($std = 6.78 * 10^{-2} m$, $max = 4.93 * 10^{-1} m$). We do not re-generate oscillations due to pedestrian stepping activity as we are not mainly interested in this study but more on longitudinal observations.

5 Discussion and Conclusion

We proposed a method that computes a human mean position from motion capture raw data. This method is composed of four steps: markers grouping, markers labeling, computation of participant skeleton and estimation of participant position. From capture with 30% missing information, we obtained continuous estimations of positions.

The main assumption in the position estimation method is the rigidity of marker skeleton motion, i.e. the distance between markers of a subject remains constant. Removing this assumption will result to a problem formulation on the form of a non-linear optimization problem which is known to be difficult to numerically handle. However, restricting our attention to cases where the definition of a rigid skeleton is valid, the position estimation method can be generalized to any planar motion by only changing the evaluation of the participant orientation. This can easily be achieved by expressing α in terms of instantaneous velocities or of a vector relying markers direction. Finally, the labeling method can be easily extended to any 1-D crowd study.

In the context of crowd study, kinematic capture is important for modeling and evaluation purpose but difficult due to acquisition conditions. Our data are intended to be available for crowd study and we detail their reconstruction process.

Acknowledgments. This work has been funded by French research agency ANR via PEDIGREE project ANR-08-SYSC-015. We also thank the M2S laboratory members, University of Rennes2, who helped us a lot in the preparation and execution of the experimental studies.

References

1. Chai, J., Hodgins, J.K.: Performance animation from low-dimensional control signals. ACM Trans. Graph. 24, 686–696 (2005)
2. Courty, N., Cuzol, A.: Conditional stochastic simulation for character animation. Computer Animation and Virtual Worlds 21, 443–452 (2010)
3. Daamen, W., Hoogendoorn, S.P.: Qualitative results from pedestrian laboratory experiments. In: Pedestrian and Evacuation Dynamics (2003)
4. Dorfmller-Ulhaas, K.: Robust optical user motion tracking using a kalman filter. Tech. rep., Universittsbibliothek der Universitt Augsburg (2003)
5. Herda, L., Fua, P., Plänkers, R., Boulic, R., Thalmann, D.: Skeleton-based motion capture for robust reconstruction of human motion. Comp. Animation, 77 (2000)
6. Kretz, T., Grnebohm, A., Schreckenberg, M.: Experimental study of pedestrian flow through a bottleneck. Journal of Statistical Mechanics: Theory and Experiment, 10014 (2006)
7. Li, L., McCann, J., Pollard, N., Faloutsos, C.: Bolero: a principled technique for including bone length constraints in motion capture occlusion filling. In: Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation (2010)
8. Pettré, J., Ondřej, J., Olivier, A.H., Cretual, A., Donikian, S.: Experiment-based modeling, simulation and validation of interactions between virtual walkers. In: Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation (2009)
9. Seyfried, A., Passon, O., Steffen, B., Boltes, M., Rupprecht, T., Klingsch, W.: New insights into pedestrian flow through bottlenecks. Transportation Science 43, 395–406 (2009)
10. Still, G.: Crowd dynamics. Ph.D. thesis, University of Warwick, UK (2000)
11. Taylor, G.W., Hinton, G.E., Roweis, S.: Modeling human motion using binary latent variables. In: Advances in Neural Information Processing Systems (2006)
12. Yamori, K.: Going with the flow: Micro-macro dynamics in the macrobehavioral patterns of pedestrian crowds. Psychological Review 105, 530–557 (1998)