2009 11th IEEE International Symposium on Multimedia

On the Compensation for the Effects of Occlusion in Fronto-Normal Gait Signal Processing

Tracey K.M. Lee[2,1]
[1]School of Electrical and Electronic Engineering
Singapore Polytechnic, Singapore
tlee@sp.edu.sg

Mohammed Belkhatir[2]
[2]School of Information Technology
Monash University, Sunway Campus
Belkhatir.Mohammed@infotech.monash.edu

Saeid Sanei
Center of Digital Signal Processing
Cardiff University, UK
saneis@cardiff.ac.uk

Abstract— **In surveillance applications, human gait data obtained from video contains idiosyncratic tendencies which allows it to be used as a biometric. This gait data has both time and image information. Expertise in the domain of time series analysis can be fruitfully employed in the image processing domain. In this paper, we consider the monocular frontal view of gait. In this view we track body parts to obtain time information and in doing so, complete occlusion of body parts may occur. To compensate for this, we present a novel standpoint where occluded images of objects may be considered as data missing from a time series. Thus we can consider this as a new application of the "missing data" problem studied in other fields dealing with time series data applied to the classic computer vision problem of occlusion. Using this approach, we consider three ways of compensating for occlusion - namely polynomial interpolation, autoregressive prediction and coupled time/frequency domain interpolation. We propose an experimental instantiation using a gait dataset and analyzing the motion of colored markers attached to body parts. The actual and predicted positions are compared which show our approach holds promise for complete occlusion compensation.**

Keywords-occlusion, missing data, gait

I. Introduction

The proliferation of video devices brought about by global security concerns and declining hardware costs makes readily available vast amounts of video data. Human gait data can be obtained from video and has enough idiosyncratic information to be used as a biometric, to identify people. Thus gait is an emerging biometric and is a quintessential multimedia signal, in that *both* time and image data are available. Thus domain knowledge, for example in time series analysis can be used to solve problems in another field, in this case, computer vision. Gait shows promise in its use, as it is nonintrusive and can be used at a distance. Current methods use the fronto-parallel (FP) view but Lee et al. [1] show that FN gait has several advantages. Firstly, it uses less space and requires fewer cameras. As a consequence, it is capable of unobstructed multiple subject tracking. The FN view naturally allows combined biometrics which allow for more robust recognition tasks as shown in Fig. 1 where the face and eyes are clearly seen. We also see that multiple subjects can be handled as well. Finally FN gait yields dynamic features that are useful for temporal analysis. However, the main challenges in FN tracking are the looming effect due to the camera lens, which causes the motion of the image to change in a deterministic (using camera equations) but nonlinear way and the self occlusion of body parts by the body itself as shown in Fig. 7.

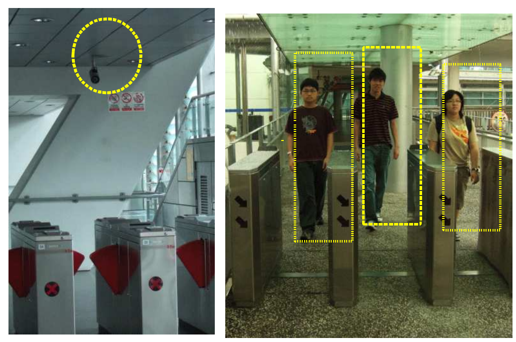

Fig. 1. Left, view of typical security camera monitoring access point. Right, tracking multiple subjects

Looming compensation has been discussed in an earlier paper and here we focus on occlusion.

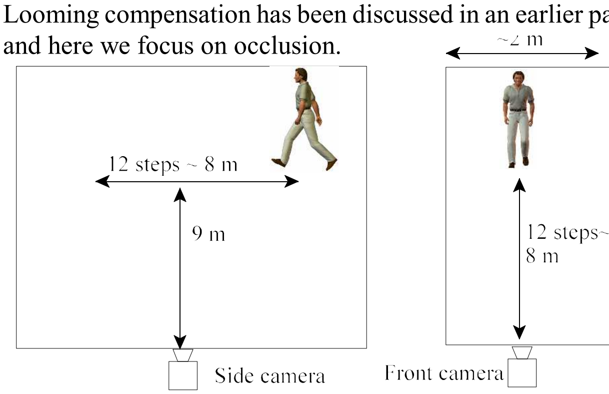

Fig. 2. FP vs FN gait - physical dimensions needed for video capture

Rather than standard approaches which use body silhouettes as ably described by Nixon et al.[2], we consider the motion of individual body parts like hands and feet. These produce biologically based spatiotemporal signal features which can be used as a biometric. As monocular FN gait has to contend with occlusion, we note that current approaches to occlusion

978-0-7695-3890-7/09 $26.00 © 2009 IEEE
DOI 10.1109/ISM.2009.130

compensation require a partial view of the object and can only handle complete occlusion for very few video frames. Many computer vision applications involve the tracking of moving objects. In real life situations these objects may obstruct other objects or themselves be obstructed from the field of view of a camera. Monocular vision is especially prone to this effect of occlusion as seen in Fig. 4. Multiple cameras can be set up to mitigate this problem, but requires alignment, calibration and synchronization among the cameras which is a significant challenge. Video data provide continuous frames of image data, from which we can track the location of an object, using methods like contours and colors, assuming that an object is in the frame. In occlusion, we have little a priori information about the object's location in the current image frame, but only that the location has been determined in earlier frames of a video stream.

In this paper, we describe the application of the "missing data" theory in the domain of time series analysis to compensate for the position of images of objects totally obscured in FN gait by extrapolation. Section II covers prior work in various fields. Section III describes the experimental setup and the problem at hand. Section IV outlines the various approaches to the problem and Section V presents the experimental setting and results. Section VI concludes the paper.

II. Prior works

Video capture of human gait is done in various fields. In computer graphics motion capture (mocap) and medical applications, several cameras are used to accurately capture motion. Occlusion is not a problem here as all the cameras will keep all the body parts in view always. With several cameras, the problems of aligning and synchronizing them are significant. In monocular systems, occlusion is more common, and refers to the blocking of view to an object. This implies the tracking of its movement and in doing so, deciding whether an object is in view or not. In the overview of occlusion tracking by Gabriel et al. [3] most approaches for handling occlusion are mainly interested when objects merge and/or split so the object location is not important during occlusion. But in our situation, we need to use the *straight through* approach for occlusion handling as we need to determine the position of the body part for every video frame. This is the case if we are using the time series generated by gait data for characterization purposes. When an object is occluded, we have to estimate its motion. Some approaches to occlusion treat it as an extension of the tracking problem. Thus, using Kalman filters or particle filters, a probabilistic model of the motion is created. In motion capture systems, the objective is to analyze human motion with a view to synthesis. In order to reliably compensate for occlusion, Liu et al. [4] use a training set of representative motions with motion markers and build a global linear model of motion. Using existing markers in a video, they predict the position of missing markers using this model. Other approaches use triangulation from other visible objects which have a fixed inter-object distance which are assumed to be constant as with Aristidou et al. [5] who also use Kalman filtering to predict motion. Koller et al. [6] use depth information to determine the order of occluding objects (cars) to remove unwanted image data. Kalman filters are used to estimate the object contour and its motion even in cases of occlusion. Blake et al. [7] use particle filtering with partitioned sampling to track multiple objects with various degrees of occlusion. However, this kind of tracking cannot be sustained for long if the object is completely occluded or the motion is complex.

In this paper, we provide a novel viewpoint to problem of occlusion by considering that the points at which the object is occluded are "missing" from the main set of data. In this approach, we do not need training images or the position of other markers. A large body of work exists in various fields which consider how to reconstruct missing or corrupted data which become outliers. Indeed, the fields of application are vast, whenever there is the need for automated data capture. Depending on how one views the problem, this approach can be used for interpolation or extrapolation, which is a more difficult problem. From the early 1980's signal reconstruction started off with ad-hoc methods based on geometrical considerations, using splines to bridge the gap between the missing samples. Splines, by their smooth nature are a natural choice for this. This led to prediction methods using AR modeling. For example Esquef et al.[8] use Autoregressive (AR) modelling to predict the gap in the data due to missing samples. As AR modeling is so often used for prediction purposes, here it is used to extrapolate from previous samples. However if the gap is large, the AR prediction becomes less effective and a backward prediction needs to be done from samples after the gap. The earliest example of using both time and frequency domain manipulation of data to obtain missing samples is the Papoulis [9]-Gerchberg [10] method. The data is alternately converted between the time and frequency domains to reconstruct a time domain signal in a typical projection onto convex sets (POCS) setting. In the frequency domain, the signal is bandlimited whereas in the time domain, the original data (existing, not missing) overwrites that generated from the frequency domain information. Next we look at our setup.

III. Experimental Setup & Problem Statement

In frontal gait recognition, we use feature points that have more motion in the image plane. This would be the hands, feet and knees. In our experimental setup, the marker designations are: *lh/rh* - left/right hand : *lf/rf* - left/right foot : *lk/rk* - left/right knee. Two additional discs of the same colour are attached at the waist and face level and are used for distance normalization. They are: *tm/bm*, the top/bottom markers, attached to the waist and neck and provide an origin for the moving body parts. They are also used to normalize the distances as the subjects will appear to grow in size as they approach the camera (i.e. the looming effect). The distance between the centers of discs is used as a scale factor. To obtain the actual coordinates, the values shown on the plots (cf. Fig. 4 for example) should be multiplied by this number. The markers are tracked using the CAMSHIFT [11] algorithm.

Fig. 4 is the plot of the movement of the colored markers for both *x* and *y* positions over time. Fig. 6 shows the trajectory of the markers in a 2D plot. The looming effect expresses itself as the amplitude of movement which increases, due to the refractive effect of the camera lens. This causes the image size on the focal plane to vary inversely with the distance of the subject from the lens. This nonlinear variation makes the analysis more challenging as we cannot make simplifying assumptions based on linearity.

As we see in Fig. 7, the left hand completely disappears from the full frontal view of the camera. An interesting observation is that as the subject approaches the camera, the hand is occluded for longer periods as seen in Fig. 3, due to the body filling a larger field of view.

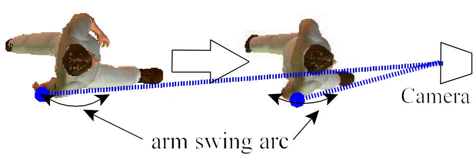

Fig. 3. Increasing self-occlusion as subject nears a camera

In considering gait, a useful characteristic is the availability of its temporal features which can be used as a biometric. To derive these features, we need a continuous flow of gait position data. So when occlusion occurs, we need to compensate, in two basic ways. First is by waiting for the occluded part to reappear and this becomes a problem of interpolating to the last observed position, which introduces a delay in processing. We prefer to compensate *during* the period of occlusion, predicting the motion of the object based on their previous motion, which makes this an extrapolation approach, which is more challenging.

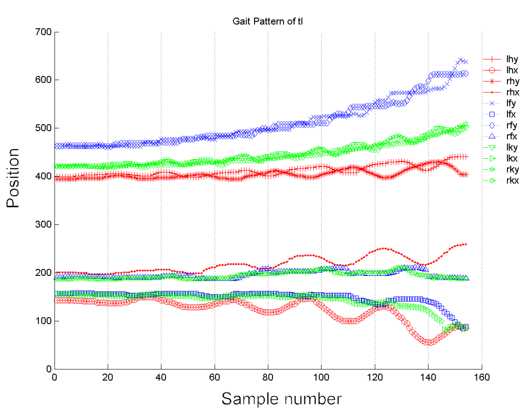

Fig. 4. Plot of all markers X and Y positions for FN walk

IV. Theory of Missing Data Handling

Although the waveform of a moving hand may seem sinusoidal, tracking is inherently a noisy process and we may not be able to justify assumptions of stationarity and linearity of the motion. There is also the effect of looming as shown in Fig. 9 where the amplitude of movement increases as it nears the camera. We hope to mitigate this effect by compensating for occlusion a segment at a time. That is, we only provide compensation up to a period of occlusion and ignore previous results due to nonlinearities.

We look at the left hand motion of subject *s01* in our dataset in Fig. 5 which shows the occluded motion of the x-axis movement of a left hand. In order to have comparative results, we use a sample sequence that does *not* have occlusion. We have *manually* noted the positions where the hand *might* have been occluded i.e. when the hand is fully swung back. These are frames 40 to 43, 67 to 73, 96 to 100, 123 to 129 and 153 to159 which are marked by dotted lines. In this case, there are five segments of data, corresponding to five possible episodes of occlusion. In the following discussion, we define the following terms for each segment of data $n+m$ samples long, where the 3rd segment is labeled in Fig. 9:

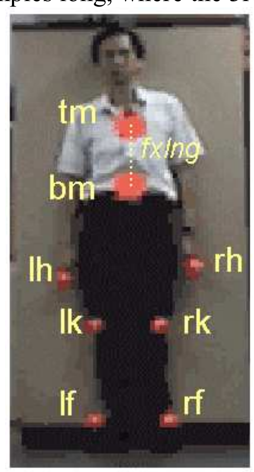

Fig. 5. FN marker designations

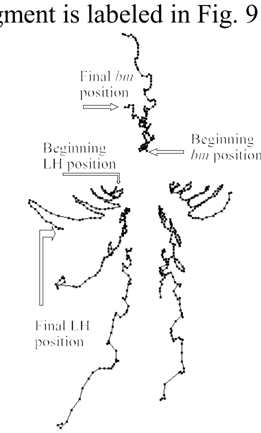

Fig. 6. 2D trajectories of markers in FN gait

Fig. 7. Left coloured marker on hand being occluded due to hand movement

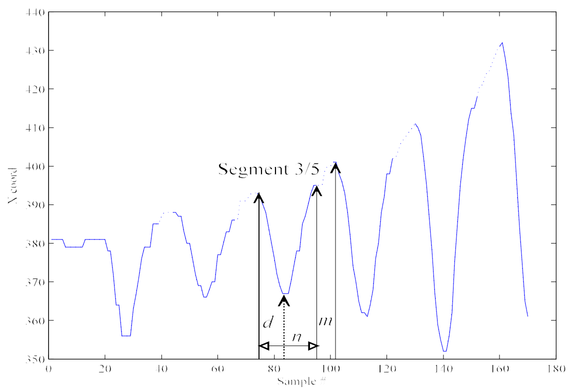

Fig. 8. Completely occluded X axis motion of LH marker (dotted lines are actual positions)

$x_{act}(t)$ - data without missing samples: $n+m$ samples.
$x_{org}(t)$ - original, available data: n samples.
$x_{miss}(t)$ - missing, or occluded data: m samples.
$x_{del}(t)$ - data deleted from segment: d samples.
$\hat{x}(t)$ - data estimated from process: $m+n-d$ samples.
$x(t)$ - data actually used: $n-d$ samples.

To be more concise, let the samples $x_y(t)$ be contained in vectors $\mathbf{x}_y$ where the subscript y refers to the data segments just described. Note that these are all of length $n+m$. They have entries where there are data, and zeros elsewhere. For example, the vector $\mathbf{x}_{org}$ will have n elements of data and m zeros. When required, we use matrices and vectors of size $n+m-d$ instead. Removing d samples of data from the beginning of the segment can help by preventing overfitting. Then we have the identity matrix $\mathbf{I}$ and a diagonal *sampling* matrix $\boldsymbol{S}$ both of size $n+m$. The elements of $\boldsymbol{S}$, s_{ii} are given by:

$$s_{ii} = \begin{cases} 0 & \textit{missing data} \\ 1 & \textit{otherwise} \end{cases} \tag{1}$$

so that $\boldsymbol{S}\mathbf{x}_{act} = \mathbf{x}_{org}$ and $(\mathbf{I} - \boldsymbol{S})\mathbf{x}_{act} = \mathbf{x}_{miss}$

Furthermore, the data in a segment are preprocessed to have zero mean and a norm of 1. They are restored to the proper scale and offset after processing. In this section, we consider four different ways of compensating for missing data.

The simplest methods for handling missing data employ a geometrical approach such as polynomial interpolation. That is, we assume that:

$$\hat{x}(t) = \sum_{i=1}^{p} a_i x^i(t) \tag{2}$$

where a_i are coefficients that are determined by the curve fitting software using a criterion like minimum least square error. However, the order p needs to be set first. While the entire segment may not fit a polynomial or by doing so cause overfitting, a small portion of a segment for example the segment (n-d) in Fig. 9 should fit a low order polynomial and be able to extrapolate beyond.

The theory for Autoregressive prediction of data is based on the fact that many natural processes can be modeled by a stochastic linear time series and can be expressed as:

$$\hat{x}(t) = a_0 + \sum_{i=1}^{p} a_i x(t-i) + \epsilon_t \tag{3}$$

where a_0 is a term to account for non zero mean data, a_i are the autoregressive coefficients and ϵ_t is a white Gaussian noise term. The order of the model is p, and this is known as an AR(p) process. The a_i terms and p can be found by a kind of optimization process, using various types of criterion which may be described in the ARFIT [12] software which also has a simulation component. This software is widely used and is capable of multivariate AR modeling.

Since our signal has a periodic appearance, it would make sense to try some kind of sinusoidal fit, even if it is only for a segment of the signal. So we have for (4):

$$\hat{x}(t) = a_0 \sin(a_1 x(t) + a_2) \tag{4}$$

where a_i are coefficients that are determined by curve fitting software. We have tried a few combinations of sinusoidal waveforms but (4) seems to give the best results.

Papoulis proposed a method of data replacement that is a type of POCS method and was designed from the outset to be used for the filling of missing data. the current data vector $\hat{\mathbf{x}}^{i+1}$ of length $n+m-d$ at iteration i, is given by:

$$\hat{\mathbf{x}}^{i+1} = (\mathbf{I} - \boldsymbol{S})\hat{\mathbf{x}}^{i}_{BL} + \mathbf{x} \quad \text{where} \quad \hat{\mathbf{x}}^{i}_{BL} = \mathscr{F}^{-1}(\mathbf{b}^{i}(\omega)) \tag{5}$$

Recall that $\mathbf{x}$ is the original data vector with missing values and with d elements removed. The bandlimited data $\hat{\mathbf{x}}^i_{BL}$ is obtained by an Inverse Fourier Transform on the Fourier Transformed $\hat{\mathbf{x}}^i$ data (which is $\mathbf{f}^i(\omega)$) bandlimited by B to give $\mathbf{b}^i(\omega)$ as shown in (6):

$$\mathbf{b}^{i}(\omega) = \begin{cases} 0 & |\omega| > B \\ \mathbf{f}^{i}(\omega) & |\omega| < B \end{cases} \quad \text{and} \quad \mathbf{f}^{i}(\omega) \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\longleftrightarrow}} \hat{\mathbf{x}}^{i} \tag{6}$$

Here $\mathscr{F}$ and $\mathscr{F}^{-1}$ are the Fast Fourier Transform and its Inverse respectively. Note that the Papoulis-Gerchberg algorithm re-substitutes the values $x(t)$ into $\hat{x}^i(t)$ at *each* iteration (see (5)) for the sampling points where $x(t)$ is known and available. In an improvement, Sauer and Allebach[13] use the previous estimated values of $\hat{x}^i(t)$ and add the bandlimited *differences* in interpolated terms with the original data $x(t)$ in one variation of the method This is done for better error control and shown in (7) and (8).

$$\hat{\mathbf{x}}^{i+1} = \Delta\hat{\mathbf{x}}^{i}_{BL} + \hat{\mathbf{x}}^{i} \quad \text{where} \quad \Delta\hat{x}^{i}_{BL}(t) \underset{\mathscr{F}}{\overset{\mathscr{F}^{-1}}{\longleftrightarrow}} b^{i}(\omega) \tag{7}$$

where $b^i(\omega)$ is defined as in (6) and

$$f^i(\omega) \underset{\mathscr{F}^{-1}}{\overset{\mathscr{F}}{\longleftrightarrow}} \Delta\hat{x}^i(t) \tag{8}$$

where $\Delta\hat{x}^i(t)$ are the elements of $(\boldsymbol{S}\hat{\mathbf{x}}^i - \mathbf{x}) + (\mathbf{I} - \boldsymbol{S})\,\mathbf{i}\hat{\mathbf{x}}^i$ and $\mathbf{i}\hat{\mathbf{x}}^i$ is the vector formed by interpolating values for missing data by the values of known data from the vector $\boldsymbol{S}\hat{\mathbf{x}}^i - \mathbf{x}$. The interpolation can be Voronoi, Line or Spline.

To compare results, we use the percentage error which is the average of all the errors for the number of segments *s* in the data and the error err_i for each segment *i* is given in (9).

$$100 * \frac{\sum_{i=1}^{s} err_i}{s} \quad \text{where} \quad err_i = \frac{\sqrt{\sum_{t \in m} [\hat{x}(t) - x_{act}(t)]^2}}{mean[x_{act}(t)]} \tag{9}$$

where *mean* is the mean operator.

V. Experimental Instantiation

In this section we describe our experimental setup, data set and results.

A. Gait Dataset

To ease the job of video analysis, we track using colored markers with the premise that this will aid in the effort for conversion to markerless tracking. In this section we briefly describe the selection and position of the colored markers used in our experiments. Spheres are best suited for our tracking purposes as it looks the same from any angle. A bright phospherent colored surface is also useful. These spheres act as markers, as described at the beginning of Section III. We look at the *x* and *y* movements with respect to the camera. Our dataset comprises video files of 12 subjects in a FN walk for training. Three of these subjects had a second video recorded a few minutes after the first to serve as a test set. For each subject, we have a set of twelve one-dimensional time series, each representing the coordinates of the movement of a body part. Thus we have 15 videos of people in a FN walk.

B. Results

We show only the figures of the original and compensated (dotted) signals from the *lhx* (left hand motion, x axis) of subject *s01*, in the interests of space. These correspond to the results shown in the first line of Table 1. Generally we see that the compensation is better for the earlier parts of the signal. This may be attributed to the smaller amplitudes of movement. It may be effective to have some kind of compensation that changes parameters according to the amplitude of the signal.

For the Autoregressive fit, we had to use a second order process because higher orders caused instability. In Fig. ? we see the rather poor fit. This is also borne out in practice, where AR processes are not used for long term predictions.

In the polynomial fit, we used the curve fitting facilities in MATLAB [14] the results are shown in Fig. 9. Again because of instability, only a second order equation was used. It must be noted that changing the value of deleted data, *d* affects the accuracy as seen in the large errors in the later part of the simulation. This points the way to a type of adaptive curve fit strategy.

We use the IRSATOL [15] software which explicitly caters for irregular sampling to perform time and frequency domain interpolation to obtain missing samples (as mentioned in section IV). Based on our setup, it provides the best results as shown in Fig. 10 and summarized in Table 1. We summarise the results from subjects *s01*, *s10* and *s08* in our gait dataset and note that similar results are obtained from other subjects in the time series.

TABLE 1 Error rates of extrapolation schemes in percent (%)

Marker-Method	AR	Polynomial	Allebach	Sine
lhx/*s01*	15.40	5.10	4.00	9.85
lhy/*s01*	24.7	17.3	13.30	8.50
lhx/*s10*	5.64	3.04	2.43	2.90
lhy/*s10*	46.5	67.30	18.61	21.70
lhx/*s08*	12.7	7.83	2.64	11.22
lhy/*s08*	29.5	18.90	10.90	22.90

VI. Conclusions and Acknowledgments

The experiments we performed demonstrate that using coupled time and frequency domain interpolation allows us to compensate for complete occlusion in FN gait, which to a certain degree, shows periodicity. We have shown that a large body of research in time series analyses exists which can be applied to compensate for visual occlusion in looming motion. We also see that a form of adaptive strategy may have to be used for such modeling in future work.

The authors wish to thank the anonymous reviewers whose comments have improved the quality of the paper.

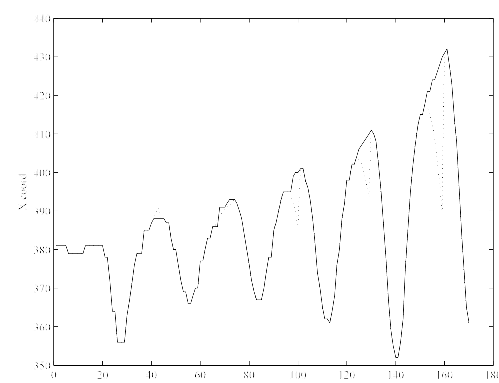

Fig. 9. Polynomial - order 2 compensation of occluded X axis motion of LH marker - dotted lines

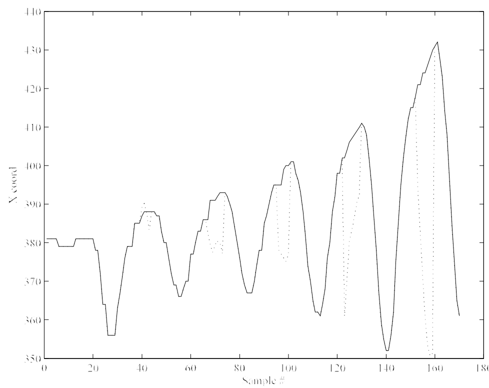

Fig. 10. Autoregressive (AR) compensation of occluded X axis motion of LH marker - dotted lines

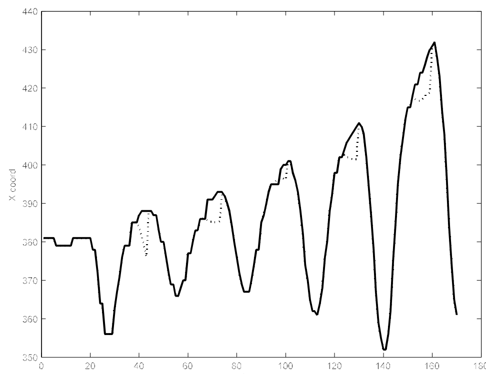

Fig. 11. Allebach method compensation of occluded X axis motion of LH marker - dotted lines

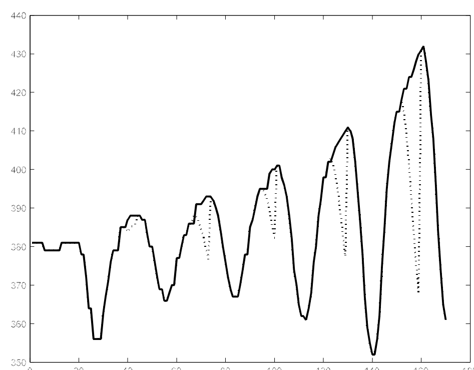

Fig. 12. Sine - first order compensation of occluded X axis motion of LH marker - dotted lines

REFERENCES

[1] T.K.M. Lee, M. Belkhatir, P.A. Lee and S. Sanei, "Fronto-normal gait incorporating accurate practical looming compensation," *ICPR 2008*, Dec. 8-11, 2008.

[2] M. S. Nixon, T. Tan and R. Chellappa , "Human Identification Based on Gait," , 1st ed, Springer, 2005.

[3] P.F. Gabriel, J.G. Verly, J.H. Piater, and A. Genon "The state of the art in multiple object tracking under occlusion in video sequences" *ACIVS '03*, pp. 166–173, Sep. 2003.

[4] G. Liu and L. McMillan, "Estimation of Missing Markers in Human Motion Capture," *The Visual Computer*, vol.22, no.9-11, pp.1432-2315, Sep 2006.

[5] A. Aristidou, J. Cameron and J. Lasenby, "Real-Time Estimation of Missing Markers in Human Motion Capture," *ICBBE '08*, May, 16-18, 2008.

[6] D. Koller, J. Weber and J. Malik, "Robust Multiple Car Tracking with Occlusion Reasoning," *ECCV 1994*, May, 1994.

[7] J. McCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *International Journal of Computer Vision*, 39(1), pp.57-71, 2000.

[8] P. A. A. Esquef, V. Välimäki, K. Roth, and I. Kauppinen, "Interpolation of Long Gaps in Audio Signals Using the Warped Burg's Method," *Proc. 6th Int. Conf. Digital Audio Effects (DAFx-03)*, London, UK, September 8-11, 2003.

[9] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans.Circuits and Systems*, vol.CAS-22, pp.735-742, 1975.

[10] R.W. Gerchberg , "Super-resolution through Error Energy Reduction," *J. Modern Optics*, vol.21, no.9, pp.709-720, Sep. 1974.

[11] G.R. Bradski, "Computer video face tracking for use in a perceptual user interface," *Intel Technology J.*, vol.Q2, 1998.

[12] A. Neumaier and T. Schneider, "Estimation of parameters and eigenmodes of multivariate autoregressive models," *ACM Trans. Mathematical Software*, vol.27, pp. 27-57, 2001.

[13] K.D. Sauer and J.P. Allebach, "Iterative Reconstruction of Band-Limited Images from Nonuniformly Spaced Samples," *IEEE Trans. Circuits and Systems*, vol.CAS-34, no.12, pp.1497-1506, Dec 1987.

[14] Mathworks Matlab 6.5 Curve Fitting Toolbox.

[15] G. Feichtinger and T. Strohmer, "IRSATOL – Irregular Sampling of Band-limited Signals TooLbox," *Conference on Computers for Teaching*, vol.Berlin, pp.277-284, 1992.