Analysis of Irregularities in Human Actions with Volumetric Motion History Images

Alexandra Branzan Albu[1], Trevor Beugeling[1], Naznin Virji-Babul[2], Cheryl Beach[3]
[1] *Department of ECE, University of Victoria (BC),* [2]*Down Syndrome Research Foundation, Vancouver (BC),* [3]*Vancouver Island Health Authority (BC), Canada*
aalbu@ece.uvic.ca, trjb@uvic.ca, naznin@drsf.org,

Abstract

This paper describes a new 3D motion representation, the Volumetric Motion History Image (VMHI), to be used for the analysis of irregularities in human actions. Such irregularities may occur either in speed or orientation and are strong indicators of the balance abilities and of the confidence level of the subject performing the activity. The proposed VMHI representation overcomes limits of the standard MHI related to motion self-occlusion and speed and is therefore suitable for the visualization and quantification of abnormal motion. This work focuses on the analysis of sway, which is the most common motion irregularity in the studied set of human actions. The sway is visualized and quantified via a user interface using a measure of spatiotemporal surface smoothness, namely the deviation vector. Experimental results show that the deviation vector is a reliable measure for quantifying the deviation of abnormal motion from its corresponding normal motion.

1. Introduction

Video-based motion analysis has lately attracted a fair number of researchers in Computer Vision. Their interest is motivated by theoretical challenges specific to video understanding, and also by the wide spectrum of applications in industry, surveillance, medicine etc.

The study of human motion from video sequences is mostly driven by applications in security. Therefore, some major research themes include human activity recognition, gait-based person identification and abnormal event detection. Such themes are obviously recognition-oriented. However, a contextual change of the application domain from security to health care uncovers a variety of new challenges and goals for the development of new computer vision algorithms.

This paper deals with a rehabilitation-related application about frail elderly people. The main goal is not to *recognize* the activity performed by a subject or the subject from the way the activity is performed, but to *analyze* a standardized set of human activities in order to quantify and monitor the subjects' performance over time. In this research context, abnormal motion is not considered as an outlier, but as a quantifiable deviation from the normal motion.

Motion analysis for performance quantification is not a trivial problem and poses different challenges with respect to motion recognition. Indeed, typical humans are experts in recognizing human activities and perform well at identifying familiar people from their gait [1]. Therefore, computer vision based approaches for activity recognition and gait-based identification can easily be validated against ground truth data produced by human reasoning. In fact, intelligent surveillance systems have emerged as an alternative to CCTV systems with a human operator in the loop. On the other side, a correct analysis of a standardized set of human activities can be performed only by highly trained individuals such as physiotherapists and kinesiologists. In their work they usually employ methods such as the Berg Balance Score (BBS) [2] which comprises 14 basic human activities common to everyday life assessed on a 5 point ordinal scale. Obviously, the validation of a computer vision-based approach for human motion assessment against BBS is impossible. Such an approach would have to be thoroughly validated with physiotherapists in the loop and should be designed as to allow an optimal trade-off between interactivity and automation. It is expected that an interactive tool based on computer vision algorithms for human motion analysis will efficiently complement score-based assessments as it will provide more insight into the subject's evolution over time.

IEEE Workshop on Motion and Video Computing (WMVC'07)
0-7695-2793-0/07 $20.00

This paper describes a new 3D motion representation, namely the Volumetric Motion History Image (VMHI), to be used for the analysis of irregularities in human motion. Such irregularities may occur either in speed or orientation and are strong indicators of the balance abilities and of the confidence level of the subject performing the activity. The rest of the paper is organized as follows. Section 2 presents related work in the field of vision-based motion analysis. Section 3 describes our proposed approach and section 4 contains experimental results. Section 5 draws conclusions and discusses future work.

2. Related Work

Extracting human motion information on a frame-by-frame basis can be performed using a variety of approaches, which can be either model-based or model-free, as shown in Aggarwal and Cai [3]. The recovered motion information is further used to build a motion representation appropriate for the task at hand. While motion representation is identified by Moeslund and Granum [4] as an essential component of tracking, it can also serve to higher-end purposes, such as recognition and analysis. The creation of a comprehensive and compact motion representation is not an easy task, since trade-offs must be made between the richness of the representation, the time spent to generate it and the computational complexity of the algorithms which will further use it. Such trade-offs are mostly critical for real-time applications. Other issues include the robustness of the motion representation to self-occlusion and errors occurred in background subtraction or other preprocessing steps. One may conclude that motion representation must be task-oriented in order to handle well all the constraints imposed by the application of interest.

In the specific rehabilitation context of the proposed work, one has to deal with loose clothing and a high variability of body shapes. For this reason, model-based approaches for motion representation do not represent a suitable option. The rest of this section will therefore focus on related work on appearance-based motion representation.

Appearance-based methods focus on whole-body motion rather than on the relative motion (and prior identification) of body parts. Such methods aim at extracting global patterns of motion as opposed to temporal trajectories of structural features; these patterns are captured as spatiotemporal templates.

The vast majority of spatiotemporal templates proposed in literature are 2D images, which offer a compact representation of motion information and are suitable for analysis and classification using standard image processing techniques for feature extraction and pattern recognition.

Polana and Nelson introduced in [5] the concept of temporal texture for the study of quasi-random motion such as windblown trees and ripples on water with techniques similar to those for gray-scale texture analysis. They also proposed in [6] a technique for detecting and recognizing periodic human activities from feature vectors containing spatiotemporal motion magnitudes created with Fourier image analysis. Cutler and Davis proposed in [7] the 2D inter-frame similarity plot, a spatiotemporal representation suitable for detecting periodic human and non-human motion.

While periodicity is an important feature of human locomotion, the above-mentioned approaches for motion representation are not suitable for the analysis of non-periodic basic human actions. The concepts of Motion Energy Image (MEI) and Motion History Image (MHI) introduced by Davis and Bobick in [8] are spatiotemporal templates specifically designed for non-periodic human activity description and recognition. MEI is a binary image where the non-zero regions indicate the spatial occurrence of motion throughout the video sequence. MHI is a gray-level image where the intensity level encodes pixelwise the recency of motion. Both MEI and MHI are generated over a temporal window of τ frames, with parameter τ either empirically chosen or computed after a recursive matching against a reference template.

The main strength of the MHI representation is its compactness, which makes it suitable for real-time activity recognition. However, the recognition process has to be label-based and reference templates must be available for each activity of interest. Another important limit of the MHI representation is its lack of robustness against motion self-occlusion, which occurs rather frequently in human actions.

Davis proposed in [9] a hierarchical extension to the original MHI framework which aims at eliminating previous problems related to global analysis and limited recognition. This hierarchical extension introduces a second parameter δ, the decay factor, in order to vary the length of the captured history of movement. The pyramid of MHIs allows for recovering to a certain extent motions of varying speed by exploiting spatial gradient information.

Valstar et al proposed in [10] another extension of MHI, namely the multiple-level MHI (MMHI), which aims at overcoming the problem of motion self-occlusion by recording motion history at multiple time intervals. Their work focused on the automatic detection of facial actions units that compose expressions. The experimental results shown in [10] do not clearly demonstrate the superiority of MMHI with respect to the standard MHI in the context of their application.

Weinland et al describe in [11] a 3D extension to the initial MHI concept, namely Motion History Volumes for viewpoint-independent action recognition. The proposed transition from 2D to 3D is straightforward, since pixels are replaced with voxels, and the standard image differencing function D(x,y) is substituted with an occupancy function D(x,y,z,t). The Motion History Volumes offer an interesting alternative to action recognition with multiple cameras, although the additional computational complexity introduced by the calibration, synchronization of multiple cameras, and parallel background subtraction is not discussed.

When changing the context of use of the motion representation from action recognition to action analysis, several new limits of the initial MHI framework and its extensions become obvious. To address them, this paper describes a new type of 3D extension for the initial MHI concept. Our proposed motion representation efficiently deals with the problems of motion self-occlusion, speed variability, and variable-length motion sequences. The following section describes in detail the proposed motion representation as well as a new measure for motion irregularity.

3. Proposed Approach

3.1. The Volumetric Motion History Image

This paper proposes a 3D extension of the MHI concept which also eliminates the need of a pre-specified length of the temporal window τ. The input data is a sequence S(:, :, k) k=1..N of binary silhouettes obtained with background subtraction from an initial video sequence acquired with one stationary camera.

The proposed Volumetric Motion History Image (VMHI) consists of a set of parallel and equidistant slices where the z coordinate encodes discrete temporal information as represented by frame indexes.

Let *contS* (:, :, *k*) denote the one pixel thick contour of the binary silhouette in frame *k*. The Volumetric Motion History Image is defined as follows:

$$VMHI(x,y,k) = \begin{cases} S(x,y,k)\Delta S(x,y,k+1) \\ \quad if\ contS(x,y,k) \neq contS(x,y,k+1) \\ 1\ if\ contS(x,y,k) = contS(x,y,k+1) \end{cases} \quad (1)$$

$$x = \overline{1,X}, y = \overline{1,Y}, k = \overline{1,N-1}$$

where x, y correspond to spatial coordinates in the image plane and X,Y are the frame dimensions in pixels, while k encodes discrete temporal information; Δ stands for the symmetric difference operator.

Each slice in the VMHI representation is built by integrating two types of information, related to:

a) the motion occurred within a pair of adjacent frames, captured with the symmetric difference operator between two adjacent binary silhouettes;

b) the spatial occupancy, captured with the binary contour comparison.

While the standard MHI [8] is based on motion information only, its hierarchical extension [10] can be computed using either motion information or spatial occupancy. However, it is believed that integrating information about both motion and spatial occupancy can provide a more robust representation than using one source of information only. The simple spatial occupancy, although it results in connected spatial regions in the VMHI horizontal slices does not lead to a straightforward motion representation. The use of silhouette differencing only for the extraction of motion information leads to disconnected regions in the horizontal slices of the VMHI, which are difficult if not impossible to visualize in 3D.

Fig. 1 shows an example of articulated human motion. The binary image in Fig. 1b contains motion information only obtained by upper body silhouette differencing. Due to motion self-occlusion and imperfect background subtraction, a significant portion of the arm contour is lost. These portions are successfully retrieved in Fig. 1c, which shows a slice of VMHI computed with (1). The gray level display is used for the purpose of visualization only: it shows contour information (spatial occupancy) in white, motion occurring from background-foreground transition in light gray, and motion from foreground-background transition in dark gray.

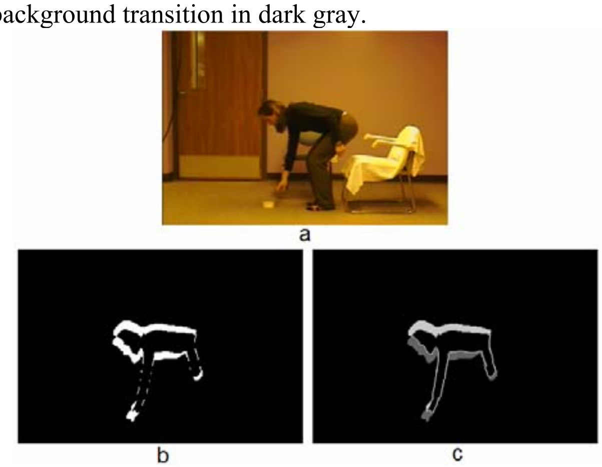


Fig. 1. a) frame in a sequence containing a picking up action; b) result of binary silhouette differencing; c) result obtained with (1).

One may observe the absence of the τ parameter encoding the length of the temporal window from the VMHI definition. In the standard (MEI, MHI) framework for recognition, τ was determined with an exhaustive research of the best correlation match between the template to be classified and a reference template. The absence of a reference template and the change of context from recognition to analysis led to the conclusion that τ is not needed in the VMHI representation. The temporal window of interest is defined over the entire motion sequence, or can be specified by the user.

3.2. Overcoming limits of the MHI representation in the motion analysis context

While the (MEI, MHI) ensemble proved to be reliable for action recognition in controlled environments [12], this compact representation fails to capture subtle details of motion essential for a successful quantitative and qualitative motion analysis. Several factors limiting the use of (MEI, MHI) representation in a motion analysis context are listed below. The capability of the proposed VMHI to overcome these limits is also proven.

a) Motion self-occlusion.

Due to the accumulative manner of generating the standard MHI, the most recent motion will overwrite all the motion information previously gathered at the same location. In the context of motion analysis, this overwriting process leads to the loss of important information and has to be eliminated.

In order to analyze the effects of the overwriting process on the MHI, we have built two sequences of identical length (58 frames) containing rigid horizontal translation with different motion irregularities. Key frames of the two sequences are shown in Fig 1a. Sequence A contains a rectangle in translation; its motion changes orientation for a certain time interval (frames 11-31), then resumes the initial orientation. Sequence B is identical to A, except for the irregularity occurring during the same time interval (frames11-31), where the object stops and remains immobile instead of moving leftwise. The computation of the MHI was performed as in [9] with τ=58. Since both sequences have identical MHIs, it can be concluded that the MHI representation fails to capture information about irregularities in motion. The choice of a smaller τ can certainly lead to different MHIs for the two test sequences, but this choice would have to be empirically made, since no *a priori* information about the occurrence and nature of irregularities is available.

Fig. 1c and 1d contain the VMHI motion representations for test sequences A and B respectively. A simple visualization of these 3D motion models allows for: a) detecting the occurrence of the motion irregularity in both sequences; b) discriminating between the two motion irregularities.

On a minor note, one may observe that capturing a temporary stop of the moving object (as in test sequence B) is possible because the VMHI model definition (1) integrates information about motion and spatial occupancy. Sudden short stops occur quite frequently in human actions and may correspond to hesitations of the subject.

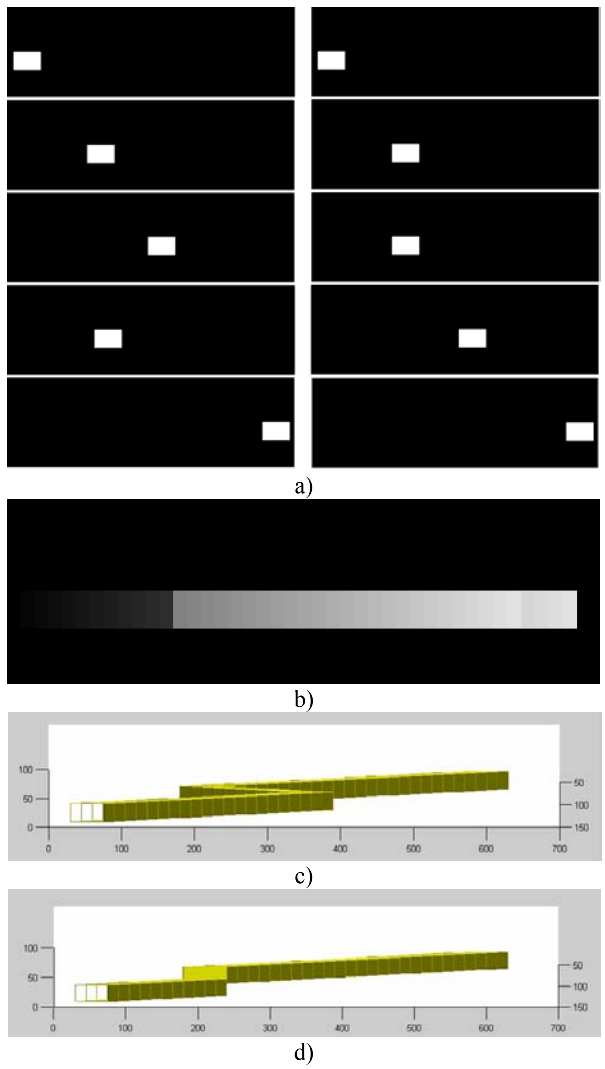


Figure 1. a) left: keyframes corresponding to the test sequence A containing a temporary change in motion orientation; right: keyframes corresponding to the test sequence B containing a temporary stop of the moving object; b) MHI is identical for both test sequences; c) VMHI for sequence A; d) VMHI for sequence B.

b) Range of the motion speed.

While fit subjects perform a given basic action in a relatively short time frame, frail subjects with poor balance take a much longer time to perform the same activity. For instance, at an acquisition rate of 30 fps, the length of a typical sit-to-stand normal action sequence is 56 frames, while an abnormal sit-to-stand lasts 148 frames. In such circumstances, it is very difficult to quantify the differences between normal and abnormal actions using the standard MHI, since it is defined over a fixed temporal window. An attempt to normalize the length of the sequences before the MHI generation has been reported in [11] for facial action units. They perform a uniform subsampling of the longer sequence. This method is not appropriate in the context of this work, since by eliminating frames in the longer sequence important information about motion irregularities will be lost.

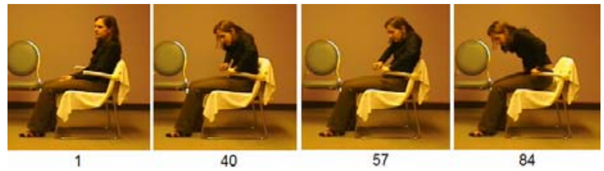
The proposed solution is the generation of VMHI models for normal and abnormal actions respectively without attempting to normalize the models for a direct inter-model comparison. Instead, the surface irregularities of each model are to be interactively observed and selected via a user interface and further analyzed with a measure of surface smoothness (see Section 4). Fig. 3 illustrates the generation of VMHI models for normal and abnormal sit-to-stand. Key frames of the abnormal and normal motion are shown in Fig. 3a and 3b respectively. For the abnormal motion, the key frames capture a major motion irregularity occurring at the beginning of the action, namely a horizontal sway used by a frail subject to start the upward motion. The VMHI model in Fig. 3d corresponds to the temporal interval where this irregularity occurs. The normal motion does not present irregularities and therefore results in a smooth VMHI.

c) Length of the motion sequences.

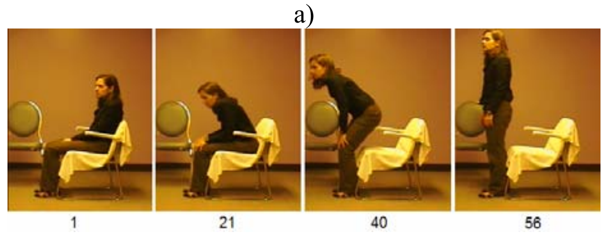
As slow motion is captured with long video sequences, an increase in the length τ of the temporal window is translated a larger number of gray levels in the MHI. Therefore, encoding motion information in a gray-level image is limited by the maximum number of gray levels. The proposed VMHI representation encodes the temporal information along the z coordinate, and therefore represents a flexible solution which is suitable for both long and short sequences.

d) Motion periodicity.

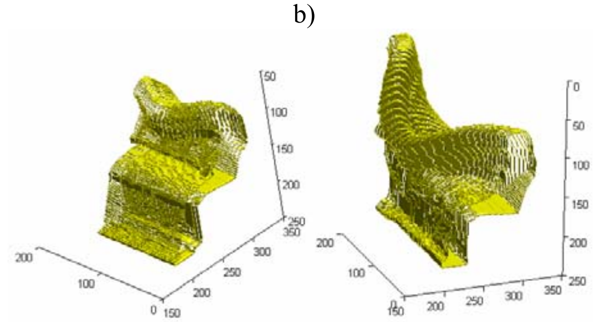
Due to the previously discussed overwriting phenomenon, the MHI template is not suitable for the analysis of periodic human motion. The proposed VMHI enables the representation and analysis of periodic motion. Moreover, it can be also used for detecting the fundamental period of motion by searching for pairs of similar slice sequences in the VMHI model separated by a minimum number of frames.



a)



b)



c) d)

Figure 3. a) key frames in abnormal sit-to-stand showing initial sway; b) key frames in normal sit-to-stand; c) VMHI for the initial sway in abnormal sit to stand; d) VMHI for the normal sit-to-stand sequence.

4. Experimental Results

4.1. Design of the experiment

The input data for the proposed work consisted in 4 pairs of video sequences containing normal and abnormal motion respectively. Each pair comprises the same human action. The actions of interest were selected in collaboration with a team of physiotherapists as follows: sit-to-stand, reaching, picking up an object placed on the floor, and stepping on one stair step. Both normal and abnormal motions were simulated by a certified physiotherapist. Working with simulated abnormal actions was considered most appropriate for the current stage of this study which focused on finding the optimal motion representation for the analysis of abnormal motion. All sequences were acquired with a monocular camera at 30 frames/second from an orthogonal view to the direction of motion. Prior to the generation of the VMHI models, all sequences were binarized with a background subtraction method based on statistical differences between foreground and background.

In the context of the work presented in this paper, abnormal motion is defined as a quantifiable deviation from normal motion. This deviation usually consists in several types of spatiotemporal motion irregularities. The sway is the most encountered motion irregularity and consists in repetitive, quick changes in motion orientation due to a temporary loss of balance or to insufficient limb strength (see Fig. 3a for an example). Other motion irregularities include variations in the speed magnitude, temporary stops in motion, as well a very limited range of motion.

While the VMHI motion representation allows for the interactive visualization and quantification of all the above mentioned motion irregularities, the work presented in this paper is focused on sway analysis.

4.2. Sway analysis and visualization

From a kinematical standpoint, human body motion is defined as an articulated motion, as it is composed of constrained relative translations/rotations of the various body parts. A smooth, consistent limb/torso translation or rotation defines a quasi-planar surface region in its VMHI template; the orientation of this surface region encodes the direction of motion. Therefore, the VMHI surface of a normal, temporally smooth motion is spatially smooth, and piecewise planar. It consists of a limited number of quasi-planar surface regions with orientations encoding the direction of motion of various body parts. The unit normals to the vertices in each quasiplanar region exhibit therefore a low variance in their orientation.

In an abnormal motion containing sway-type irregularities, the relative translation of body parts features frequent changes of speed and orientation. Consequently, its corresponding VMHI features an unsmooth appearance, since the sway translates into spatiotemporal “ripples”. The normals to the vertices in the VMHI surface region corresponding to a sway exhibit therefore a high variance in their orientation.

The above considerations led towards the adoption of a descriptor of the VMHI surface smoothness for the analysis of sway irregularities. For a given activity performed in an abnormal way (to be further referred to as *abnormal activity*), this descriptor measures its deviation from the same activity performed in a normal manner (i.e. the *normal activity*).

Let us introduce the following notations:

- $VMHI_{abn}$: the VMHI of the abnormal activity;
- $VMHI_n$: the VMHI of the the normal activity;
- (n_x, n_y, n_z): unit normal vectors defined on each vertex of the VMHI surface.
- $\text{var}(n_x, n_y, n_z)|_{VMHI} = (\text{var}(n_x), \text{var}(n_y), \text{var}(n_z))|_{VMHI}$: the statistical variance of the orientations of the normal vectors to a given VMHI surface.

The deviation of an abnormal activity from its corresponding normal activity is measured with a deviation vector D, defined as follows:

$$D = \lfloor D_x, D_y \rfloor$$

$$D_x = \frac{\text{var}(n_x)\big|VMHI_{abn} - \text{var}(n_x)\big|VMHI_n}{\text{var}(n_x)\big|VMHI_n} \tag{2}$$

$$D_y = \frac{\text{var}(n_y)\big|VMHI_{abn} - \text{var}(n_y)\big|VMHI_n}{\text{var}(n_y)\big|VMHI_n}$$

The D_x and D_y components of the deviation vector quantify motion irregularities occurring along the vertical and horizontal directions respectively (i.e. horizontal and vertical sway). Moreover, one may notice that all activities in the experimental set involve planar motion only, occurring in the (x,y) image plane.

The deviation vector has only two components, since the variance of n_z, var (n_z) is significantly influenced by the speed of motion. The following experiment is to prove this affirmation.

Three test sequences containing a 60x40 pixel rectangle in horizontal translation at constant slow, medium and fast speed were generated. As shown in Table 1, the variance var (n_z) of the unit normal vectors over the VMHI templates generated for each sequence is clearly correlated with the speed of motion.

Table 1. Correlation between var (n_z) and speed of translatory motion in three test sequences.

Speed of motion	var(n_z)
2 pixels/frame	0.3775
5 pixels/frame	0.6217
10 pixels/frame	0.7684

As mentioned in section 3.2, this work does not deal with the automatic detection of sway-type irregularities. Instead, the VMHI model is available for interactive visualization, modification and analysis of sway in a user interface which enables the user to:

a) freely rotate the VMHI of a given human action for inspection of motion irregularities;

b) simultaneously visualize and thus compare two VMHIs of normal and abnormal motion respectively;

c) generate a more precise VMHI for a spatiotemporal region of interest containing sway by specifying its temporal limits (i.e. start and end frames);

d) compute the deviation of the between normal and abnormal motion with (2).

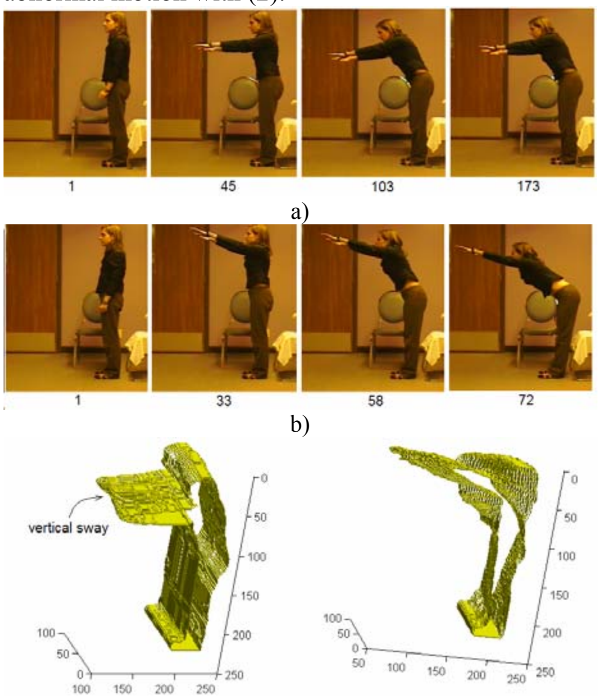


Figure 4. a) key frames in abnormal reaching; b) key frames in normal reaching; c) VMHI for user-selected sway region in abnormal reaching ; d) VMHI for normal reaching.

Figures 3, 4, 5, and 6 illustrate the visualization of VMHI templates corresponding to normal and abnormal actions respectively. Table 2 contains the statistical variance of the normal orientations to the surfaces of the VMHI models built for each activity. It is easy to notice that var(n_x) and var(n_y) are always higher in the abnormal activity than in the corresponding normal activity. This result is consistent with our initial assumption which correlates the smoothness of the VMHI surface to the motion coherency.

Table 3 uses the data in Table 2 for computing the [D_x, D_y] vector of deviations with (2). The largest deviation, D_y=76.41%, was obtained for the horizontal sway in the abnormal sit-to-stand. This result is consistent with the qualitative observation of a large-amplitude sway in the abnormal sit-to-stand video. The abnormal reaching, characterized by a more subtle vertical sway, has the D_x deviation larger than D_y. The abnormal picking up features horizontal sway, which results in D_y larger than D_x. The smallest D_x and D_y deviations are obtained for the abnormal stepping, which contains a temporary stop. This result suggests that, for temporary stops in motion, additional motion measures (e.g. speed-related), might be necessary.

Table 2. Statistical variance of the unit normals to the surfaces of the VMHI activity models

Activity	Start-end frames	var (n_x)	var (n_y)	var (n_z)
Sit to stand abnormal	1-84	0.3430	0.3373	0.3183
Sit to stand normal	1-56	0.2778	0.1912	0.5249
Reaching abnormal	102-173	0.4446	0.3477	0.2064
Reaching normal	33-72	0.2987	0.2554	0.4443
Stepping abnormal	1-98	0.3898	0.2099	0.3950
Stepping normal	1-30	0.3392	0.1819	0.4732
Picking up abnormal	1-200	0.4147	0.2872	0.2972
Picking up normal	1-50	0.3093	0.1866	0.4945

Table 2. Deviation from the normal pattern of activity

Activity	D_x (%)	D_y (%)	Dominant motion irregularity
Sit-to-stand	23.47	76.41	Horizontal sway
Reaching	48.84	36.13	Vertical sway
Stepping	14.91	10.99	Temporary stop
Picking up	34.07	53.62	Horizontal sway

5. Conclusions and future work

This paper proposes a 3D extension of the standard MHI concept introduced in [8]. This extension deals effectively with issues such as motion self-occlusion, variable speed and sequence length in the context of human motion analysis. Consequently, the VMHI representation is suitable for the visualization and quantification of several types of motion irregularities. This work focuses on the analysis of sway, which is the most common motion irregularity in the studied set of human actions. Horizontal and vertical sways are visualized and quantified via an interactive user interface using the deviation vector, which is a measure of spatiotemporal surface smoothness. The experimental results show that this measure is reliable for quantifying the deviation of the abnormal motion from its corresponding normal motion.

Ongoing work focuses on the development of methods for analysis and quantification of other types of motion irregularities. Future work will also address the issue of quantitative comparison of spatiotemporal motion templates generated from different subjects.

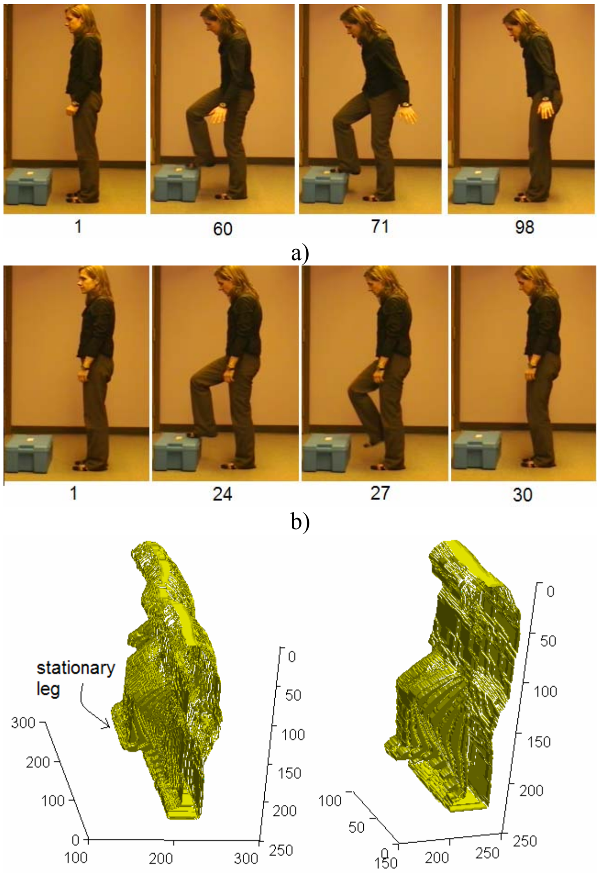


Figure 5. a) key frames in abnormal stepping; b) key frames in normal stepping; c) VMHI for user-selected region in abnormal stepping showing a temporary stop in motion; d) VMHI for normal stepping.

References

[1] J. Cutting and L. Kozlowski, "Recognizing friends by their walk: gait perception without familiarity cues," Bulletin of the Psychonomic Society, vol. 9, pp. 353-356, 1977.
[2] K. Berg, S. Wood-Dauphinee, J.I. Williams, "The Balance Scale: reliability assessment for elderly residents and patients with an acute stroke," Scandinavian Journal of Rehabilitation Medicine, vol. 27, pp 27-36, 1995.
[3] J.K. Aggarwal and Q. Cai, "Human motion analysis: a review", Computer Vision and Image Understanding (CVIU), vol. 73:3, pp. 428-440, 1999.
[4] T. B. Moeslund and E. Granum, "A survey on computer vision-based human motion capture" Computer Vision and Image Understanding (CVIU), vol. 81:3, pp. 231-268, 2001.
[5] R.C. Nelson and R. Polana, "Qualitative recognition of motion using temporal texture", CVGIP: Image understanding, vol. 56:1, pp. 78-89, 1992.

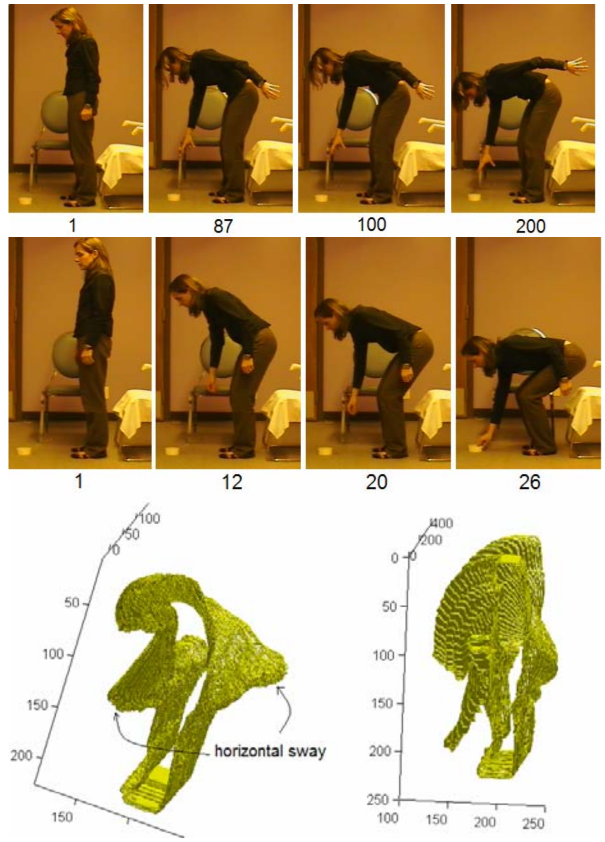


Figure 6. a) key frames in abnormal picking up; b) key frames in normal picking up; c) VMHI for user-selected region in abnormal picking up showing horizontal sway; d) VMHI model for normal picking up.

[6] R. Polana, R.C. Nelson, "Detection and recognition of periodic, non-rigid motion", International Journal of Computer Vision, vol. 23:3, pp. 261-282, 1997.
[7] R. Cutler, L.S. Davis, "Robust real-time periodic motion detection, analysis, and applications", IEEE Trans. on Patt. Anal. and Machine Intell., vol. 22:8, pp. 781–96, 2000.
[8] A.F. Bobick and J.W. Davis, "The recognition of human movement using temporal templates", IEEE Trans. on Patt. Anal. and Machine Intell., vol. 23:3, pp. 257–267, 2001.
[9] J.W. Davis, "Hierarchical motion history images for recognizing human motion", Proc. Of IEEE Workshop on Detection and Recognition of Events in Video (EVENT'01), Vancouver (BC, Canada), July 2001.
[10] M. Valstar, M. Pantic, and I. Patras, "Motion History for Facial Action Detection in Video", Proc. of IEEE Int'l Conf. on Systems, Man and Cybernetics, pp.635-640, Oct. 2004.
[11] D. Weinland, R. Ronfard, E. Boyer, "Motion history volumes for viewpoint action recognition", IEEE International Workshop on modeling People and Human Interaction (PHI'05), 2005.
[12] A. Bobick, S. Intille, J. Davis, F. Baird, L. Campbell, Y. Ivanov, C. Pinhanez, A. Schutte, and A.Wilson, "The KidsRoom: action recognition in an interactive story environment", Presence, 8(4):367–391, 1999.

IEEE Workshop on Motion and Video Computing (WMVC'07)
0-7695-2793-0/07 $20.00 © 2007 IEEE