An Approach to Automatic Assistance in Physiotherapy based on Online Movement Identification

Conference Paper · November 2012

4 authors:

Dennis Romero
Escuela Superior Politécnica del Litoral (ES…
12 PUBLICATIONS 6 CITATIONS
SEE PROFILE

Anselmo Frizera
Universidade Federal do Espírito Santo
113 PUBLICATIONS 255 CITATIONS
SEE PROFILE

Teodiano Bastos
Universidade Federal do Espírito Santo
240 PUBLICATIONS 836 CITATIONS
SEE PROFILE

Boris X. Vintimilla
Escuela Superior Politécnica del Litoral (ES…
37 PUBLICATIONS 81 CITATIONS
SEE PROFILE

An approach to Automatic Assistance on Physiotherapy based on on-line Movement Identification

R. L. Dennis; A. F. Neto ; T. F. Bastos ; Electrical Engineering Department, Federal University of Espirito Santo, Vitoria, Brasil ; B. X. Vintimilla; Fac. Ing. en Electricidad y Computación, Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador.

***Abstract*— This paper describes a method for on-line movement identification, oriented to patient's movement evaluation during physiotherapy. An analysis based on Mahalanobis distance between temporal windows is performed to identify the "idle/motion" state, which defines the beginning and end of the patient's movement, for posterior patterns extraction based on Relative Wavelet Energy from sequences of invariant moments.**

I. Introduction

In the human behavior analysis field, automatic identification of people movements represents a very broad research area worldwide. The initial state of the art has identified recent works that have been recognized, as mentioned in [1] [2] and [3] [4]. Initially, we studied methods to analyze people motion on video sequences. It was addressed the problem of extracting features from persons on images in different activities in time, such as walking, extending and retracting the arm, falling, etc., in order to apply it in the progress feedback for patients on physiotherapy.

In the physiotherapy and occupational therapy field, innovative concepts on rehabilitation have been found in the literature as mencioned in [5] [6] [7], where were combined modern techniques of motion analysis to monitor the patient's functional progress. These proposals take advantage of technological advances (processing capabilities, size-reduced electronic components, portability, lower power consumption, numerical methods, etc.) to contribute with alternatives that promote the rehabilitation process in all perspectives (patients, physiotherapists, rehabilitation techniques, treatments, evaluations, etc.). Among these areas we can mention: robotics for rehabilitation, cognitive rehabilitation, human-machine interfaces (HMI), brain-computer interfaces (BCI), and others. Based on this premise, the current proposal presents a technique based on computer vision for movements identification, and it's focused on global movement evaluation of patients during rehabilitation. The goal is to provide feedback to the patient when each movement has been finished, related to the performance of each activity, as parts of a whole rehabilitation exercise. Thus, patients can be informed in a quantitative way about the progress level which is being reached and about the goals that are being achieved around the whole exercise. The automatic feedback during physiotherapy would complement the therapist's intervention, by facilitating the monitoring of movement evolution during the exercise and by enriching the patient experience during the functional rehabilitation process.

The first part of the article describes the set of input data, represented by person silhouettes. Subsequently, we present an approach used for features extraction from the captured images, based on Hu moments [8]. Previous works proposed [9] described a stochastic analysis to validate Hu moments as a time series, and frequency-based features were extracted. This paper presents advantages over previous work, with an on-line approach based on wavelet transform, directly from the invariant moments signals without applying the first derivative. As result of this, were considered time and frecuency properties from the signals, improving the quantity of movements to be recognized, with a very fast response. 9-dimensional patterns are obtained in order to train a neural network for posterior movements classification.

II. Metodology

As mentioned, there are currently robust algorithms for segmentation and people tracking. In our proposal these methods could provide the input data, corresponding to images of person in motion, which it's subsequently possible to isolate it from the background. For this reason, the initial studies of this proposal were based on the database Muhavi-Mas (Multicamera Human Action Video Data - Manually Annotated Silhouette) proposed by Velastin [10], to evaluate methods of actions recognition. However, in further studies was used a Kinect sensor, which provides depth information to contribute in the segmentation process. This made it possible to generate binary silhouettes from persons for the desired goals (Figure 1).

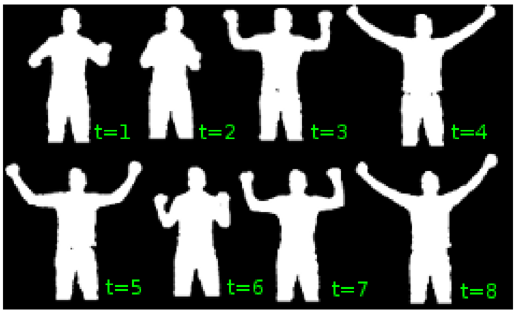

Fig. 1. Person silhouette at eight time instants.

A. By-frame feature extraction

There are several types of invariant moments. Hu's moments [8] are the most commonly used in the literature and the best in terms of their orthogonality, rotation invariance, low sensitivity to image noise, fast computation and the ability to provide a faithful image representation [11]. That's the reason they had been widely applied in image recognition [12]. Hu's moments are region-based invariant features that consider all the image pixels [13]. These seven moments have invariant properties to affine transformations, including changes in scale, translation and rotation, for these reasons we had chosen them for our work. In order to identify movement patterns, specifically were considered the 2^{nd}, 3^{rd} and 4^{th} moment for each image. The first moment was dismissed for having redundancy with the other moments and for being more sensitive to image noise. The last moment was not considered due to achieving small values in certain movement conditions. Figure 2 shows the moments mentioned during the execution of three different movements ("left arm up-down", "hello signal", "opening-closing arms").

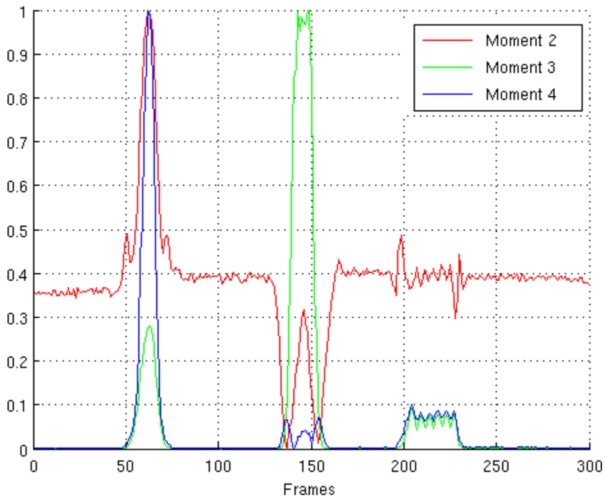

Fig. 2. 2^{nd}, 3^{rd} and 4^{th} Hu's moments, in 3 different movements

In previous studies[9], we performed a stochastic analysis of Hu's invariant moments, where it was possible to identify stationary characteristics after applied the 1^{st} derivative [14]. However, in order to preserve and facilitate the identification of detailed moments variations as well as reduce the computational cost, the application of the 1^{st} derivative was not considered.

B. "idle/motion" state detection

The extracted invariant moments for each image allowed us to represent the person's posture in time instants of his movement, on this work, in a three-dimensional space $Pos_{\delta t} = [\phi_1, \phi_2, \phi_3]$, where ϕ represents moments with order 2, 3 and 4. For a time interval τ, we consider each dimension as unidimensional signal, where $S_i = [\phi_{i0}, \phi_{i1}, \phi_{i2}, \phi_{i3}, ..., \phi_{i\tau-1}]$. This unidimensional approach is used to identify the "idle/motion" state, in which just S_4 is considered for verifying initial and ending postures of person movement, because it presents more sensibility on movement presence. Since invariant moment sequences are updated approximately at the camera's frame rate (25-30 fps), a temporal-windows analysis to identify the beginning and end of the person action was considered. The temporal window's size was defined by optimizing the Mahalanobis distance (1), which can be defined as a dissimilarity measure between two random vectors $\vec{W}_{i-1}$ and $\vec{W}_i$ with the same distribution, and with covariance matrix Σ:

$$d(\vec{W}_{i-1}, \vec{W}_i) = \sqrt{(\vec{W}_{i-1} - \vec{W}_i)^T \Sigma^{-1} (\vec{W}_{i-1} - \vec{W}_i)} \quad (1)$$

The Mahalanobis distance has been used in several applications, such as [15], where it was applied in signature recognition. In the Figure 2 it's possible to observe an homogeneous 4^{th} Hu's moment at movement absence, in this situation the Mahalanobis distance may take value lower than 0.01, under controlled light conditions. However, the tolerance's parameter for "idle/motion" detection may vary depending on light conditions or noise in the capture process. Four frames were enough to establish distance measurements that support noise in the capture and binarization process, providing a significant marking of the beginning and end of motion. These windows W are treated on-line as shown in Figure 3.

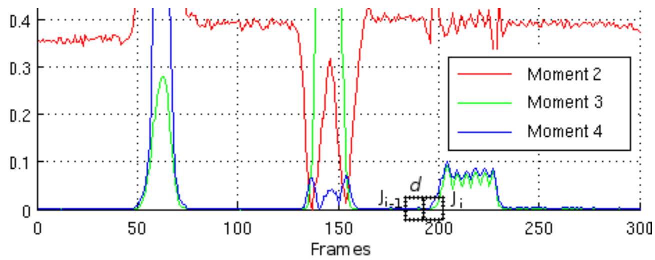

Fig. 3. Temporal windows of 4^{th} Hu's moment, in a previous instant to the movement.

The Mahalanobis distance method was compared with the auto-correlation criterion between temporal windows, getting more consistent measures with Mahalanobis. The 4-size windows allows to identify a "idle/motion" state in $size_{win}/fps \approx 0.2seg$.

C. Movement windows extraction

Once identified the movement beginning (which was determined with the 4^{th} Hu's moment sequence), the new moment's vectors (3-dimensional) are temporarily stored into dynamic growth windows, between the beginning mark $t - size_{win}$ until the end mark defined by the new "idle/motion" state. This allows that the corresponding motion data could be centered into the window, as shown in Figure 4.

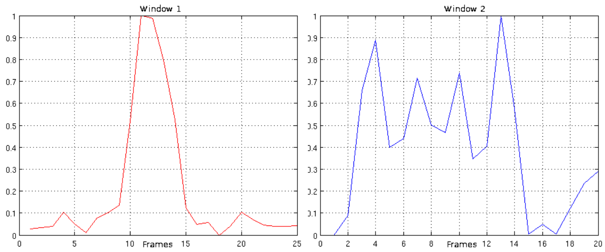

Fig. 4. On-line extracted windows during two different movements

D. Movement pattern extraction

Due to the periodic nature of certain human movements (walking, running, jumping, etc.), in the early stages of this research a frequency analysis was considered, by using the Fast Fourier Transform (FFT) after application of the first derivative to the invariant moments sequences. However, the discrete wavelet transform (DWT) specifically allows the discrimination of non-stationary signals with different frequency characteristics, so the DWT was chosen with Daubechies (db4) family, as a method for posterior features extracting of person's movement. The Daubechies wavelet family was chosen by the following properties:

a) Time invariance - If the time series is shifted in time, then its wavelet coefficients are only shifted in time.
b) Fast computation - Daubechies wavelet has self-similar type fractal that directs fast wavelet transform techniques.
c) Filter sharp transition bands - Daubechies wavelet has sharp transition bands which minimize edge effects between the frequency bands.

The DWT is a transformation of the original temporal signal into a wavelet basis space. The time-frequency wavelet representation is performed by repeatedly filtering the signal with a pair of filters that cut the frequency domain in the middle. Specifically, the DWT decomposes a signal into an approximation signal and a detail signal. The approximation signal is subsequently divided into new approximation and detail signals. This process is carried out iteratively producing a set of approximations signals at different detail levels and a final gross approximation of the signal [16].

As mentioned above, the selected wavelet $\psi(t)$ is Daubechies 4 (db4) with 3 decomposition levels, and since it is an orthonormal basis for L, the concept of energy is linked with the usual notions derived from Fourier's theory. The wavelet coefficients are given by $C_j(k) = \langle Win_{mov}, \psi_{j,k} \rangle$ which can be interpreted as the local residual errors between successive signal approximations at scales j and $j+1$, and the energy, at each level of decomposition $j = -1, ..., -N$, will be the energy of the detail signal [17],

$$E_j = \|r_j\|^2 = \sum_k |C_j(k)|^2 \tag{2}$$

where $r_j(t)$ is the residual signal at scale j, and the energy at time instant k will be:

$$E(k) = \sum_{j=-N}^{-1} |C_j(k)|^2 \tag{3}$$

Consequently, the total energy can be obtained by,

$$E_{total} = \|Win_{mov}\|^2 = \sum_{j<0} \sum_k |C_j(k)|^2 = \sum_{j<0} E_j \tag{4}$$

Finally, the normalized values p_j are defined. They represents the relative wavelet energy, where $\sum_j p_j = 1$ and the distribution $\{p_j\}$ can be considered as a time-scale density, constituing a suitable tool to detect and characterize specific phenomena in time and frequency planes [18].

$$p_j = \frac{E_j}{E_{total}} \tag{5}$$

The relative energy of three decomposition levels is calculated for each movement's window, it results in a 9-dimensional pattern. The Figure 5 presents the mentioned procedure's diagram:

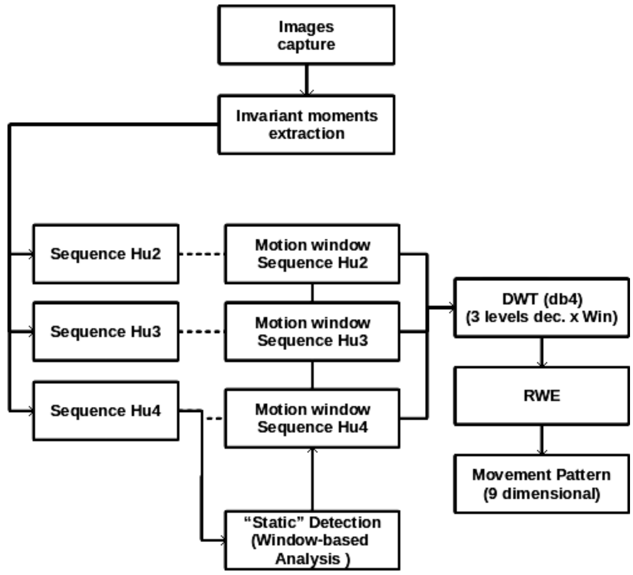

Fig. 5. Proposed methodology Diagram

III. Results

This paper presented a methodology for on-line identification of human movements. Invariant moments region-based were used to extract initial features from person's images. The post-processing applied to create silhouettes takes advantage of the depth information provided by the capture sensor, resulting in images without noise from another objects. The temporal analysis based on Mahalanobis distance allowed the quick identification of the idle/motion state (approx. 0.2 sec), making it possible to obtain dynamic growth windows containing only motion information. These motion windows are directly processed for subsequent features extraction by computing Relative Wavelet Energy (RWE - Relative Wavelet Energy). The feature extraction method based on wavelet transform has a significant computing performance, considering that processing time for short movements (gestures, or expressions by short-term) is approx. 2 ms (computer with 2GHz processor).

The Figure 6 shows a screenshot from an application with the detailed methods implemented, for on-line identification of different durations movements.

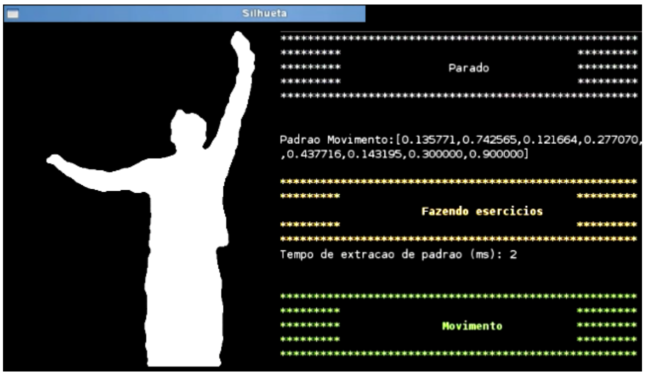

Fig. 6. Screenshot during an action recognition test

As mentioned initially, the extracted features from the person's silhouettes conforms unidimensional signals, which are analyzed by the idle/motion identification method. These signals may be substituted with others, which describe the person's movements. Thus, the feature extraction method doesn't depend exclusively on the invariant moments from the images, it being possible to include movement information from inertial sensors, for example, wich is also suitable to be used by patients on physioterapy treatment. Another advantage of the patterns extraction method is that it doesn't depend on transition modeling, as mentioned in [19], which limit the number of movements to be recognized. For testing, were considered about 32 patterns for each movement, which were divided for training (50%), validation (25%) and test (25%). As a result, 90.5% of 4 movements were correctly classified with own tests and 88.7% by using 10 movements from the Muhavi-MAS database[10].

IV. Conclusions and Future Work

This paper presented a methodology for on-line human actions recognition on video sequences, oriented to movement evaluation for physiotherapy applications. The method used for actions recognition could be applied to different multi-dimensional tracking data, in this case was applied on 3 sequences of Hu's moments, obtaining 9-dimensional patterns for classification using artificial neural networks (ANN's).

As a future work, an probabilistic approach to track different movements will be study, in order to get information around the whole physiotherapy session, to make projections about treatment evolution.

References

[1] Hee-Deok Yang, A-Yeon Park, and Seong-Whan Lee. Gesture spotting and recognition for human ndash;robot interaction. *Robotics, IEEE Transactions on*, 23(2):256 –270, april 2007.

[2] P. Chockalingam, N. Pradeep, and S. Birchfield. Adaptive fragments-based tracking of non-rigid objects using level sets. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1530 –1537, 29 2009-oct. 2 2009.

[3] Catherine Achard, Xingtai Qu, Arash Mokhber, and Maurice Milgram. A novel approach for recognition of human actions with semi-global features. *Machine Vision and Applications*, 19:27–34, 2008.

[4] Sangho Park and Mohan M. Trivedi. Understanding human interactions with track and body synergies (tbs) captured from multiple views. *Comput. Vis. Image Underst.*, 111(1):2–20, jul 2008.

[5] B.-C. Tsai, W.-W. Wang, L.-C. Hsu, L.-C. Fu, and J.-S. Lai. An articulated rehabilitation robot for upper limb physiotherapy and training. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 1470 –1475, oct. 2010.

[6] S. Kousidou, N. Tsagarakis, D.G. Caldwell, and C. Smith. Assistive exoskeleton for task based physiotherapy in 3-dimensional space. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pages 266 –271, feb. 2006.

[7] K. Kifayat, P. Fergus, S. Cooper, and M. Merabti. Body area networks for movement analysis in physiotherapy treatments. In *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*, pages 866 –872, april 2010.

[8] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179 –187, february 1962.

[9] Dennis Romero L. Anselmo Frizera Neto, Teodiano Bastos Filho. Movement analysis in learning by repetitive recall. an approach for automatic assistance in physiotherapy. In *Biosignals and Biorobotics Conference (BRC), 2012 ISSNIP*, pages 1 –8, jan. 2012.

[10] S. Singh, S.A. Velastin, and H. Ragheb. Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 48 –55, 29 2010-sept. 1 2010.

[11] Zhihu Huang and Jinsong Leng. Analysis of hu's moment invariants on image scaling and rotation. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on*, volume 7, pages V7–476 –V7–480, april 2010.

[12] Qing Chen, Emil Petriu, and Xiaoli Yang. A comparative study of fourier descriptors and hu's seven moment invariants for image recognition. In *Electrical and Computer Engineering, 2004. Canadian Conference on*, volume 1, pages 103 – 106 Vol.1, may 2004.

[13] Rafael C. Gonzalez. *Digital Image Processing, 2nd Edition*. McGraw-Hill, 2010.

[14] Andreas Antoniou. *Digital Signal Processing*. McGraw-Hill, 2005.

[15] Yu Qiao, Xingxing Wang, and Chunjing Xu. Learning mahalanobis distance for dtw based online signature verification. In *Information and Automation (ICIA), 2011 IEEE International Conference on*, pages 333 –338, june 2011.

[16] Ling Guo, Daniel Rivero, Jose A. Seoane, and Alejandro Pazos. Classification of eeg signals using relative wavelet energy and artificial neural networks. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, GEC '09, pages 177–184, New York, NY, USA, 2009. ACM.

[17] Osvaldo A. Rosso, Susana Blanco, Juliana Yordanova, Vasil Kolev, Alejandra Figliola, Martin Schurmann, and Erol Basar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105(1):65 – 75, 2001.

[18] O.A Rosso, M.T Martin, and A Plastino. Brain electrical activity analysis using wavelet-based informational tools (ii): Tsallis non-extensivity and complexity measures. *Physica A: Statistical Mechanics and its Applications*, 320(0):497 – 511, 2003.

[19] Sangho Park and Mohan Trivedi. Multi-person interaction and activity analysis: a synergistic track- and body-level analysis framework. *Machine Vision and Applications*, 18:151–166, 2007.