Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Hybrid Feature Subset Selection for the Quantitative Assessment of Skills of Stroke Patients in Activity of Daily Living Tasks

Gert Van Dijck, Marc M. Van Hulle, *Senior Member, IEEE*, and Jo Van Vaerenbergh

Abstract—**Stroke patients have a decreased ability in performing Activity of Daily Living (ADL) tasks such as in 'drinking a glass of water', 'turning a key', 'picking up a spoon', 'lifting a bag', 'reaching a bottle' and 'lifting and carrying a bottle'. These tasks can be quantified by measuring forces and torques exerted on the objects. However, the resulting force and torque time series represent information at a very low level of abstraction and don't inform clinicians what really distinguishes patients from normal controls in performing these tasks. We conduct an extensive quantitative analysis of these tasks and derive interesting features from the time signals that characterize the differences in behavior between patients and normal controls. We show that 'drinking a glass' and 'turning a key' are the most discriminative tasks; furthermore we show that the ability or disability to synchronize the thumb and the middle finger is one of the most important features.**

I. INTRODUCTION

It is evident that stroke patients have serious problems with the execution of an important number of tasks related to daily life activities. Brain damage caused by embolic or hemorrhagic infarction disrupts the sensorimotor organization that integrates the information from various cortical and subcortical areas [1], [2]. In some cases the consequences of the damage are limited and clinical appreciable as disturbed, jerky movements. Then kinematical analysis is practicable and leads to an understanding of the disrupted control. However a complete unilateral paralysis followed by the return of nearly invisible movements a few weeks after stroke is a customary course of events. Under these conditions kinematical analysis at task level is unfeasible. Therefore, the introduction of new assessment techniques of arm and hand functionality in clinical neurophysiologic context is of utmost importance to improve understanding of recovery of stroke patients in this phase.

As the emphasis in stroke rehabilitation is on the improvement of 'function', the new quantitative measurement tool has to use Activity of Daily Living (ADL) tasks as a principle. These tasks are thoroughly described in textbooks for physical and occupational therapists [3], [4]. Within this scope the ALLADIN project developed a measuring instrument that records the very low forces and torques produced by stroke patients during attempts of performing ADL tasks under isometric constrains[1]. The basic hypothesis is that features extracted from the movement preparation and initiation under these conditions are determinants for functional recovery after stroke and reflect important neural control parameters related to brain plasticity.

In this article we deal in particular with the challenging task of:

1. Identifying which parameters, derived from the force and torque signals, are important to describe the differences between normal controls and stroke patients. This ultimately leads to a better understanding of the processes, due to the increased level of abstraction regarding raw time series, which are responsible for the decrease in motor performance of patients.
2. Identifying in which ADL tasks normal controls and patients differ the most. This leads to important advise in developing measurement instruments for clinicians.
3. Identifying the sensors, attached to different parts of the body, that are most important to describe the differences. This leads to a reduction of the number of sensors, hence to a decreased cost of the experimental set-up.

II. METHODOLOGY

The ADL tasks need to be representative for daily living. In total six tasks were considered in this research: 'drinking a glass of water', 'turning a key', 'picking up a spoon',

Manuscript received April 22, 2006. This first author was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen). The second author is funded by the Belgian Fund for Research -- Flanders (G.0248.03, G.0234.04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), the Belgian Science Policy (IUAP P5/04), and the European Commission (NEST-2003-012963, IST-2002-016276, IST-2004-027017)).

The third author is funded by the European Commission (IST-2002-507424).

G. Van Dijck is with the Computational Neuroscience Research Group, Laboratorium voor Neuro-en Psychofysiologie, K.U. Leuven, B-3000 Leuven, BELGIUM (phone: +32 16 32 12 47; fax: +32 16 32 13 00; e-mail: gert@neuro.kuleuven.ac.be).

M. M. Van Hulle is with the Computational Neuroscience Research Group, Laboratorium voor Neuro-en Psychofysiologie, K.U. Leuven, B-3000 Leuven, BELGIUM (e-mail: marc@neuro.kuleuven.ac.be).

Jo Van Vaerenbergh is with the Centre for Multidisciplinary Approach and Technology vzw, B-1090 Brussels, BELGIUM (e-mail: jo.vanvaerenbergh@cmat.be).

[1]The ALLADIN project is funded by the European Commission under the 6th Framework Programme, IST Programme, Priority 2.3.1.11 – eHealth, IST Contract No.: IST-2002-507424

'lifting a bag', 'reaching for a bottle' and 'bringing a bottle to the opposite side'.

Eight torque and force sensors are positioned on body parts that are assumed to be relevant in performing the ADL tasks. The sensors are placed on the thumb, the index, the middle finger, the arm, the foot, the big toe, the seat and the back. Each sensor is capable of measuring 3 forces and 3 torques in X, Y and Z direction. Before the experiments start, the patient is positioned in a mechatronic platform [5]. The signals are sampled at 100 Hz.

In this study a total of 48 stroke patients and 48 normals are involved. Before the subjects try to perform an ADL task, the protocol of measuring consists of showing the subjects an example video of a person actually performing the ADL task. This helps in avoiding different interpretations of the subjects of the task to be executed, which may seriously hamper inference from the ADL tasks. The subjects are asked to perform three attempts of the ADL task subsequently. This allows us to assess whether a subject has gained some experience in the ADL task, which facilitates the identification of normal controls and patients. It can for e.g. be hypothesized that a stroke patient has increased difficulty of learning and exploiting experience from the ADL task compared to normal controls. This hypothesis will be tested explicitly.

The three goals defined in the introduction can be assessed partly by a feature subset selection (FSS) strategy [6], which aims at finding features subsets from the original set of features in order to maximally discriminate between the proposed classes (stroke patients and normal controls). However, a traditional filter or wrapper approach [7], will not be able to indicate which tasks are really relevant. Therefore, we apply our hybrid filter/wrapper approach [8], [9], which makes explicit which features can be shown to be statistically relevant. Moreover, the hybrid filter/wrapper approach outperforms the accuracy of the subsets found compared to the stand-alone wrapper approach. This will be further illustrated in section IV.

In order to get an understanding of the signals we are dealing with, we show example signals in figure 1.

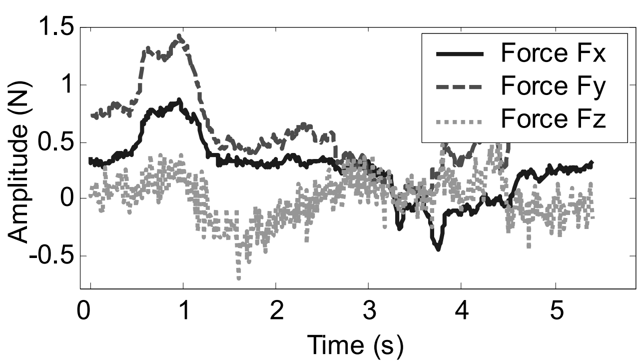

Fig. 1. Forces exerted on a glass in the drinking task. The forces in X, Y and Z direction are shown for the thumb. After approximately 0.5 s the patient tries to grasp the glass, this is shown by the increased amplitude of the forces in X and Y direction.

The relevant part of the signal to describe the differences between patients and normal controls is limited to the movement initiation part of the signal [10]. This e.g. due to the fact that beyond the movement initiation, the patient tries several attempts to move the glass, which is however fixed [5]. Hence, data beyond the movement initiation can be expected to contain very subject dependent information, while we are more interested in invariance within the patient and normal control classes. A statistically rigorous method to extract this initiation part by means of machine learning is described in [10].

III. Feature Definitions

The signals as described in figure 1 represent information at a very low level of abstraction. These time series are hard to interpret in order to describe the differences between patients and normal controls. Therefore the level of abstraction is increased by defining features (characteristics) over the relevant part of the signal (the movement initiation part as explained). In the definition of the features different hypotheses are formulated about the differences in normal controls and subjects.

The majority of the features are based on a physiological interpretation of the signals. Vector forces and torques are constructed from the signals: $\mathbf{F}[t] = (F_x[t], F_y[t], F_z[t])$ and $\mathbf{T}[t] = (T_x[t], T_y[t], T_z[t])$, where t is the time index. From this, a mean direction for the force and torque measurements can be defined by averaging the force vectors $\mathbf{F}[t]$ and torque vectors $\mathbf{T}[t]$ over the relevant time span (the movement initiation part). From now on, we denote these averaged vectors $\mathbf{F}_m$ and $\mathbf{T}_m$ respectively. Next we describe the design of the features intuitively.

- *Planning of a mean trajectory.* From experience it is clear that normal controls are better capable of 'planning' a mean trajectory in their efforts, e.g. in trying to bring a glass to the mouth. Hence it can be hypothesized that the angular deviations from the mean efforts $\mathbf{F}_m$ and $\mathbf{T}_m$ show larger deviations and abnormalities for patients. These deviations and abnormalities can be assessed by calculating the standard deviation, the skewness and the kurtosis of the angular deviations. In order to take temporal aspects into account, we propose to fit an AR-model to these angular deviations. The consistent effort of a normal control can be expected to give rise to mainly positive low order coefficients, while the anti-persistent effort ('shaky effort') of the patient is expected to give rise to some negative low order AR coefficients.
- *Continuity in effort.* The 'continuity' of the effort can be considered by computing the sequence of angles between subsequent force vectors ($\theta[t] = \mathrm{angle}(\mathbf{F}[t],\mathbf{F}[t-1])$) and torque vectors ($\phi[t] = \mathrm{angle}(\mathbf{T}[t],\mathbf{T}[t-1])$). To detect

abnormalities and persistent behaviour in this sequence of angles, the same parameters are computed as in the angles to the mean direction.

- *Velocity components.* From the force and torque values we can also extract linear velocity and angular velocity respectively:

$$\frac{m}{k+1}\mathbf{v}[k]=\frac{1}{k+1}\sum_{t=0}^{t=k}\mathbf{F}\left[t\right] \quad (1)$$

$$\frac{I}{k+1}\boldsymbol{\omega}[k]=\frac{1}{k+1}\sum_{t=0}^{t=k}\mathbf{T}\left[t\right] \quad (2)$$

However, we need to emphasize that strictly speaking, there is no real movement, while the objects are fixed. Therefore these features have no physical significance. Formula's and (1) and (2) would be correct in case of freely moving objects and time independent mass (m) and moment of inertia (I).

- *Synchronization between sensors.* Displacement of objects needs a careful synchronization between the parts of the body that are involved. Considering e.g. the 'drinking a glass' task, it is clear that one needs a good synchronization between forces and torques exerted by the thumb, the index and the middle finger. We compute this synchronization by the information theoretic measure of statistical dependency, known as mutual information [11]. The mutual information was computed as follows:

$$I(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k)\|)=$$
$$\max_{a}\sum_{k}p(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k-a)\|) \quad (3)$$
$$\ln\left(\frac{p(\|\mathbf{F}_{s1}(k)\|,\|\mathbf{F}_{s2}(k-a)\|}{p(\|\mathbf{F}_{s1}(k)\|)p(\|\mathbf{F}_{s2}(k-a)\|)}\right).$$

Hence, the mutual information between the forces of sensor s1 and sensor s2 is computed between the magnitudes of these forces. We need to take into account that signals belonging to different sensors can be shifted in time. Therefore, the delay 'a' between the sensors is searched for, such that the magnitudes of the vectors become maximally dependent.

- *Time delay between sensors*. The time delay 'a' in (3) for which the mutual information becomes maximal is considered as an extra feature.

IV. Feature Subset Selection

Feature subset selection is a means of selecting the most important hypotheses that will be useful in the description of the differences between patients and normal controls given the available amount of data.

Firstly, it is of paramount importance to identify which features really contribute in describing the differences. This problem can be assessed by a relevance analysis. Subsequently, features with overlapping information are searched for in the redundancy analysis. Both the relevance analysis and redundancy analysis constitute the filter part of the feature subset selection (FSS) algorithm. In paragraph C, we will show that preceding a wrapper search with this filter not only leads to a better understanding of the problem (as will be shown in paragraphs A and B), but also leads to superior subsets compared to the stand-alone wrapper search.

A. *Relevance Analysis*

A relevance analysis determines whether there exist a statistical dependency between a feature and the target variable, the class label of normal controls and patients. If the class is independent from the feature, this feature does not contain any information about the class. More formally this can be expressed as follows: features are irrelevant if the mutual information (MI) between the feature (F_i) and the class label (C):

$$I(F_i,C)=\int_{f_i}\int_{c}P(f_i,c)\log_2\left(\frac{P(f_i,c)}{P(f_i)P(c)}\right)df_i dc \quad (4)$$

is low.

The problem is, however, that the MI can only be estimated from the finite sample distribution (the real underlying probability distributions in (4) are not known). Hence, we need to estimate the mutual information. The actual value of the MI will generally depend on: the particular estimator, the sample size and the sample distribution. So based on the value of the MI alone it is hard to decide whether a feature is statistically relevant or not. Therefore, we introduce permutation tests [9]. This consists of computing the MI under random permutation (permutation π_k) of the class labels relative to the feature values:

$$I_{\pi_k}(F_i,C_{\pi_k})=\int_{f_i}\int_{c}P(f_i,c_{\pi_k})\log_2\left(\frac{P(f_i,c)}{P(f_i)P(c)}\right)df_i dc. \quad (5)$$

Here, C_{π_k} consists of the randomly permutated class labels, relative to the feature values. Repeating such random permutations N times, provides us with a sample distribution of the MI ($I_{\pi_k}(F_i,C_{\pi_k})$) under the null hypothesis (H_0) of feature F_i being irrelevant. From this sample distribution a statistically motivated threshold τ_p is determined, here the p subindex refers to the type I error rate, e.g. a threshold $\tau_{0.01}$ would only be exceeded by 1% of the vector samples drawn from the irrelevant feature F_i. When this threshold is exceeded by the actual MI between feature F_i and the class labels C (thus without permuting), the feature is considered as statistically relevant. In figure 1, we give the histogram of the MI for a feature obtained under 1000 permutations. The mutual information in the figure is estimated from the $I^{(1)}$ estimator of Kraskov et al. [12]. It clearly shows that the feature is not relevant, the actual value 0.044 is beneath the $P_{0.01}$ threshold at 0.1815. We performed this analysis on all features separately, using 1000 permutations for every feature and estimating the MI from the $I^{(1)}$ estimator.

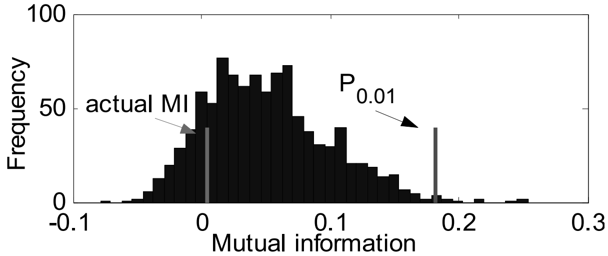

Fig. 2. Histogram of the mutual information obtained from 1000 random permutations. It clearly shows that the actual MI (without permutation) is smaller than the $P_{0.01}$ threshold. Therefore, the feature is not relevant.

In total we have 13248 features. From the relevance analysis it is shown that only a small amount (202) of the features are relevant. The majority of these 202 relevant features can be contributed to the 'drinking a glass' and 'turning a key' task, as is shown in figure 3.

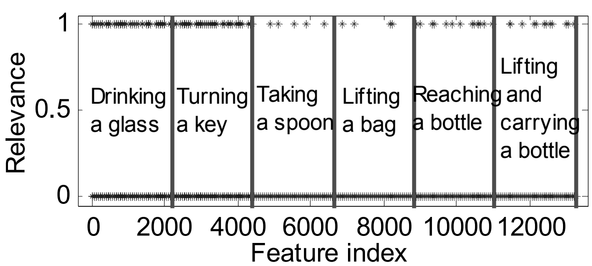

Fig. 3. Relevance analysis of the features. The feature relevance is set to 1 when it is relevant; otherwise it is set equal to 0. Only a very small subset, 202, of the original 13248 features is relevant. Most of the relevant features can be contributed to the 'drinking a glass' and 'turning a key' tasks.

A more thorough analysis shows that about 80% of the relevant features can be assigned to the 'drinking a glass' and 'turning a key' tasks, both in the torque and the force signals. Table 1 gives the relative importances of the three fingers in these tasks.

TABLE I
RELATIVE IMPORTANCES OF FINGERS

Force/ Torque	*Finger*	*Task*	
		'Drinking a glass'	**'Turning a key'**
	Thumb	23.3%	42.9%
Force	Index	30.0%	14.3%
	Middle	26.7%	17.9%
	Thumb	9.4%	51.9%
Torque	Index	40.6%	14.8%
	Middle	25.0%	22.2%

From this table we can conclude that the index plays the most important role for the torques in the 'drinking a glass' task. The thumb is clearly most important in the 'turning a key' task for both the forces and the torques.

Analysis of variance (ANOVA) learns that there is no difference in relevance between the different attempts.

B. Redundancy Analysis

The redundancy analysis is aimed at finding feature subsets which cover the same information in describing the different classes. A heuristic approach herein is to gather all features in a cluster for which the maximum distance between any features in the cluster is not larger than a given threshold, see e.g. [8] and [9]. An appropriate distance measure between features F_i and F_j is 1 minus their statistical dependency expressed by means of the normalized mutual information:

$$nI(F_i, F_j) = \frac{2I(F_i, F_j)}{H(F_i) + H(F_j)} = \frac{2(H(F_i) + H(F_j) - H(F_i, F_j))}{H(F_i) + H(F_j)}. \tag{6}$$

Formula (6) appeals to intuition, because the distance between the feature and itself (1 - $nI(F_i, F_i)$) is equal to 0, while the distance between independent features is equal to 1. Moreover it will always scale between 0 (completely dependent features) and 1(completely independent features).

This analysis shows that a limited number of features is strongly dependent e.g. if we put the normalized mutual information distance threshold to 0.3 only 4 pairs of features will have a smaller distance. Performing this analysis we learn that the MI (computed from formula (3)) between the 'thumb-middle finger' and the MI computed (from formula (3)) between 'index-middle finger' in the drinking task for the torques contain almost the same information. In figure 4 we give an example of these 2 strongly dependent features.

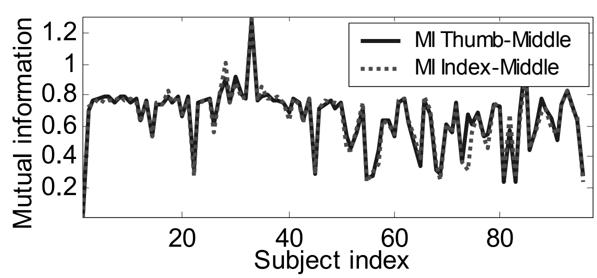

Fig. 4. Strongly dependent features. The MI value from formula (3) is shown for all subjects between 'thumb-middle finger' and 'index–middle finger' for the torques. The normalized mutual information between these features is equal to 0.271.

Finally, we perform a search among possible candidate features subsets that allow to discriminate as good as possible between normal controls and stroke patients.

C. Wrapper Search

A wrapper search in FSS (feature subset selection) involves 2 main components: the search algorithm, to search among possible candidate subsets, and the performance evaluation. A comparative study [13] has shown that the performance of genetic algorithms (GA's) for FSS is comparable to state-of-the-art search algorithms.

Moreover, GA's even outperform these algorithms for larger subsets (typically more than 50 features). Here, we focus on a genetic algorithm for FSS [8]. The fitness of the individuals (a candidate feature subset) in the GA is determined by the performance of the feature subset in discriminating the normal controls from the stroke patients.

For the performance estimation, we model every class by a widely accepted model in pattern recognition: a Gaussian mixture model (GMM). The number of Gaussian kernels in the mixture model is estimated from the MML (minimum message length criterion) [14]. The performance is then estimated by firstly estimating the class densities from a GMM using training data, subsequently test data are assigned to the class with maximum posterior probability using Bayes' rule (Bayesian classifier). To build different training sets and test sets a 10 fold-crossvalidation scheme was applied. We compare the performance of the FSS when we use the relevance filter (the hybrid filter/wrapper approach) and when we omit it (stand alone wrapper approach) and hence perform a search on the full set of 13248 features. Results are given in table 2. The parameters of the GA were set as follows: number of individuals per population = 30, number of populations = 200, probability of crossover between 2 individuals = 0.3, probability of mutation for every feature within an individual 0.01.

TABLE 2
COMPARISON HYBRID FSS WITH STAND-ALONE WRAPPER FSS

Feature Subset size	*Performance: probability of correctness of classification*	
	Hybrid filter / wrapper FSS	Stand-alone wrapper FSS
1	0.7983	0.7983
2	0.7842	0.7072
3	0.7851	0.6853
4	0.7670	0.7229
5	0.8228	0.7441
6	0.8096	0.6729
7	0.8452	**0.7645**
8	**0.8546**	0.7162
9	0.8369	0.7258
10	0.8507	0.7020

A paired t-test proofs that the hybrid filter/wrapper approach easily outperforms ($p \approx 10^{-4}$) the stand-alone wrapper approach for this large feature set.

Finally, we include a scatter plot of the data for the 2 features found in table II (for the feature subset of size 2); the decision boundaries are computed from the Bayesian classifier with a GMM. It shows a MI feature on the X axis. As was indicated in section 3, this feature defines a statistical dependency between features and therefore is capable of capturing the ability to synchronize forces and torques between different sensors. It is clear from figure 5 that this MI is decreased for the stroke patients.

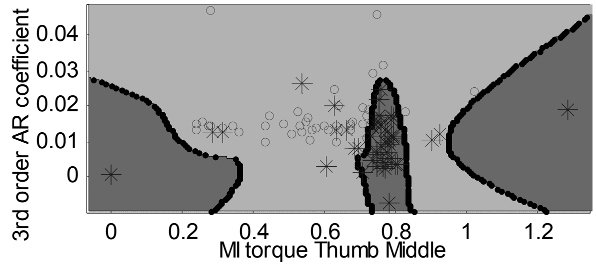

Fig 5. Decision boundaries computed from a Bayesian classifier. The stroke patients are indicated by 'o', the normal controls by '*'.

V. CONCLUSION

We have shown that the hybrid filter approach largely contributes to the understanding of the data both regarding the relevance and redundancy in the features.

From this analysis it was shown that the 'drinking a glass' and 'turning a key' tasks are by far the most important ones. The hybrid filter/wrapper approach easily outperforms the stand-alone wrapper FSS.

The ability to synchronize torques between sensors plays an important role, a significant decrease in synchronization is observed within the group of patients.

REFERENCES

[1] R. Laforce Jr. and J. Doyon, "Distinct contribution of the striatum and cerebellum to motor learning," *Brain. Cogn.*, vol. 45, pp. 189-211, March 2001.

[2] P. Haggard and A.M. Wing, "Coordination of hand aperture with the spatial path of hand transport," *Exp. Brain. Res.*, vol. 118, pp. 286-292, January 1998.

[3] J. Carr and R. Shepherd, *Neurological Rehabilitation: Optimizing Motor Performance*. Oxford: Butterworth-Heinemann, 1998.

[4] C. Perfetti. *Der Hemiplegische Patient. Kognitiv-Therapeutische Übungen*. München: Richard Pflaum Verlag GmbH & Co, 1997.

[5] S. Mazzoleni, J. Van Vaerenbergh, A. Toth, M. Munih, E. Guglielmelli and P. Dario, "ALLADIN: A novel mechatronic platform for assessing post-stroke functional recovery," in *Proc. 9th International Conference on Rehabilitation Robotics*, Chicago, 2005, pp. 156-159. Available: http://www.alladin-ehealth.org/publications/abstracts/Mazzoleni_final.pdf.

[6] I. Guyon and A. Elisseeff: "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, March 2003.

[7] R. Kohavi and G. H. John: "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, pp 273-324, December 1997.

[8] G. Van Dijck, M. M. Van Hulle and M. Wevers, "Hierarchical feature subset selection for features computed from the continuous wavelet transform," *2005 IEEE Workshop on Machine Learning for Signal Processing*, Mystic, 2005, pp. 81-86.

[9] G. Van Dijck and M.M. Van Hulle, "Speeding up the wrapper feature subset selection in regression by mutual information relevance and redundancy analysis," *submitted to ICANN 2006*.

[10] G. Van Dijck, M. M. Van Hulle and J. Van Vaerenbergh, "Statistically rigorous human movement onset detection with the maximal information redundancy criterion," *submitted to EMBC 2006*.

[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[12] A. Kraskov, H. Stögbauer and P. Grassberger, "Estimating mutual information," *Phys. Rev. E.* 69 066138, June 2004.

[13] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern recognition," *Pattern Recognition*, vol. 33, pp. 25-41, January 2000.

[14] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. On Pattern Analysis and Machine Intelligence* 24 (2002) 381-396, March 2002.