Using PCA and ICA for exploratory data analysis in situation awareness

Johan Himberg, Jani Mäntyjärvi , Panu Korpipää

Abstract— **This paper presents an approach for analyzing hand held device usage situation (context) phenomena. The situation information under examination is multidimensional fuzzy feature information derived from multisensor measurements. The analysis is conducted using principal component analysis (PCA) and independent component analysis (ICA). PCA is used to fuse multidimensional feature information into a more compact representation while the ICA is applied to extract patterns containing independent low level information about the situation. The results show that a few principal components compress the situation data representation efficiently. In addition, principal component representation provides a method for visualizing high level situation information. Most independent components extracted from the usage situation data correlate strongly with some of the original signals. This suggests that the original context data already consist of relatively independent signals if the temporal relations in the data are omitted.**

Keywords— **context awareness, data mining, information visualization, principal component analysis, independent component analysis, mobile computing, ubiquitous computing**

I. Introduction

Sensor based situation awareness of a personal mobile device requires multisensor fusion. Several studies about sensor based situation recognition systems for mobile devices have been carried out. Brown et. al. propose that a situation can be recognized from processed sensor signals fused with location information [1]. Sensor based situation awareness of wearable computers has been studied by Pentland and Clarkson [2], [3], [19]. Their work discusses high level situation recognition from camera and environmental audio signals by using hidden Markov models. Also, neural networks have been applied for inferring situational information of a mobile device [10], [16]. A variety of applications utilizing situation awareness have been reported in the literature [13], [14], [16], [17]

In situation aware computing of a personal mobile device, the user carries an appliance containing multiple sensors. Information about the situation concerning for example activity of the user and state of the environmentat a certain moment is mixed in the measured signals. This problem can be viewed as analogous to the blind source separation problem, and independent component analysis (ICA) is potentially feasible [9]. Furthermore, the measurements in situation aware computing tend to contain a lot of redundant information, and statistical methods such as principal component analysis (PCA) become relevant due to the capability of compacting the representation for the purposes of storage, usage for machine learning tasks, etc. In the literature for example, PCA has been proposed for use in data compression, and ICA for inferring the state of the environment [15].

In this paper, principal component analysis (PCA) and independent component analysis (ICA) are examined for the purposes of rising the abstraction level of the usage situation information, and compacting the representation of the data. PCA is a classical method for finding a linear transformation for dimension reduction. The recently introduced ICA [4], [9] is based on a model of linearly mixed statistically independent sources. To our knowledge, PCA and ICA have not previously been applied to the analysis of personal mobile device situation awareness, which forms the main contribution of the paper. However, it must be noted that the results achieved so far are yet qualitative of nature.

II. Methods

A. Feature extraction and selection

In order to achieve sensor-based situation awareness of a mobile device, the abstraction level of the raw sensory measurements must be raised. This is carried out by quantisizing the time dimension and information content of the data by using a variety of signal processing methods. In this work, the features to be extracted are chosen according to how well they describe aspects of the real world context. Fuzzy sets [20], [21] are used as the format for these low level features (context atoms), representing meaning directly for a human and thus enhancing the understandability of the information content at the feature level, and facilitating interpretation at higher levels. Fuzzy quantization can be viewed as a granulation of information, which makes it possible to exploit the tolerance for imprecision by focusing on the information which is decision relevant [21]. The aim is to advance to manipulating perceptions instead of measurements. For example in the recognition of walking, the meaning of shifting from the action of walking in normal speed to walking fast is fuzzy, and both labels can be partially true at the same time.

We have implemented a context atom recognition system that processes a multidimensional input vector $x(t)$ from sensor measurements into a context atom vector $c(t) = [a_1, a_2, \ldots, a_k]^T$ for each quantization time step. Vectors $c(t)$, $t = 1, 2, \ldots, N$, form a dataset to be used as input for ICA and PCA. Methods used for recognition of context atoms a_k include, e.g., gesture recognition methods of mobile device user [12], frequency domain analysis and statistical methods for movement and illumination recog-

Nokia Research Center, email: johan.himberg@nokia.com
Nokia Mobile Phones, email: jani.mantyjarvi@nokia.com
Technical Research Center of Finland, email: panu.korpipaa@vtt.fi

nition.

B. Principal Component Analysis and Independent Component Analysis

The principal component analysis (PCA) and the independent component analysis (ICA) are well established statistical tools, for example, in the signal processing and data-analysis communities. In this paper they are used as a means for explorative data-analysis and information visualization. The possibility to use neural network based approaches with PCA and ICA in sensor fusion is presented in [15]. The basics of PCA and ICA as well as their applications in data analysis are discussed for example in [5], [11]. In this chapter, a brief introduction into these methods is given.

PCA is a classical statistical method for finding a linear transformation for dimension reduction. Let $\mathbf{x}$ be an n-dimensional random vector. The task is to find an orthonormal matrix $\mathbf{V}$ of size $k \times n$, $k < n$ so that the reduced k-dimensional projection $\mathbf{x}' = \mathbf{Vx}$ retains as much of the variance of $\mathbf{x}$ as possible. The rows of the matrix $\mathbf{V}$ define the principal directions of the projection. In practise, the principal directions and components can be calculated, e.g., using the eigendecomposition of the sample covariance matrix. The eigenvalues corresponding to the principal components determine the relative amount of variance that each component captures. PCA is often used for dimension and noise reduction, or for explorative visualizations. For the purpose of visualization, the data is projected into a low dimension (1D–3D) by selecting some principal component(s) (PC) and plotting the projected data points. If two dominant PCs are selected — that is the PCs corresponding to the two largest eigenvalues — the result is a projection that presents the data in 2D capturing as much of the original data variance as possible.

The model in PCA is implicitly based on an assumption of gaussian data. In the basic version of ICA the assumption is that there are n observed signals which are different linear mixtures of n statistically independent, non-gaussian source signals. The sources are elemets of an n-dimensional random vector $\mathbf{s}$. The elements of the observed random vector $\mathbf{x}$ are different mixtures of the sources $\mathbf{x} = \mathbf{As}$, where $\mathbf{A}$ is an $n \times n$ mixing matrix. The problem is to solve the mixing matrix when the observed signals, that is, a sample of the random variable $\mathbf{x}$, are known. There exists a variety of methods for estimating the mixing matrix and the independent components (IC) under different assumptions on the data [7], [11]. In this paper, we use the FastICA algorithm [6], [8] that has very fast convergence and is available as a well documented implementation[1]. ICA has recently gained lots of interest, and there are interesting and successful applications that include, e.g., blind source separation for biomedical signals [18].

It must be remembered that PCA and ICA are statistical methods that work in a similar way for any permutation of the sample data vectors. Thus, the temporal relations in the data must be inferred by some other means.

[1] http://www.cis.hut.fi/research/software.shtml

TABLE I
DESCRIPTION OF THE SCENARIO

	Activity
1	User is sitting, the device is on a table.
2	The device is hanging from the user's neck.
3	User stands up and starts to walk,
4	walks in a corridor,
5	goes up in an elevator (1 floor),
6	walks in a corridor,
7	goes out to a balcony,
8	walks in the corridor,
9	goes down in the elevator (1 floor),
10	walks in the corridor,
11	sits down,
12	and puts the device on the table.

III. Experiments and results

A. Experiment setup

A small mobile device was equipped with a set of sensors including 3 accelerometers - one for each direction, sensors for illumination, temperature, humidity, skin conductivity and audio. The device was hanging from the user's neck in front of the chest, and the user performed the activities depicted in Tab. I. Data were logged by using a laptop that the user was carrying. The scenario was repeated 25 times for two test persons. One test was rejected due to a measurement failure, so the data consist of 49 repetitions of the same scenario. Each test lasted about 3–4 minutes depending on the testee and on the environment. In result, there were approximately 10000 data vectors, since the sampling interval of the context atom data was 1 second. The 27 context atoms derived from the multisensor measurements represent movements of device and user, device orientation, humidity, illumination conditions, and level of sound pressure. An example of the data recorded from 'Test 10' is presented in Fig. 1 where the fuzzy membership of each context atom is presented as a gray level bar as a function of time.

B. Experiments using PCA

The data from all the 49 tests were pooled together, and the PCs were computed by eigendecomposition of the covariance matrix. Figure 1 shows the flow of data in one specific test, 'Test 10'. The context atoms are presented as a function of time using gray level bars. Black means that the fuzzy membership of the context atom is one, white means zero membership. The figure depicts also the two most dominant PCs, PC1 and PC2, which were calculated using the pooled data set, not only the data of 'Test 10'. The arrows in the figure point out how some phenomena in the original signals are related to the first PC (PC1). Figure 2 and Tab. II show the share of the total variance for the original variables and for the PCs. Many of the con-

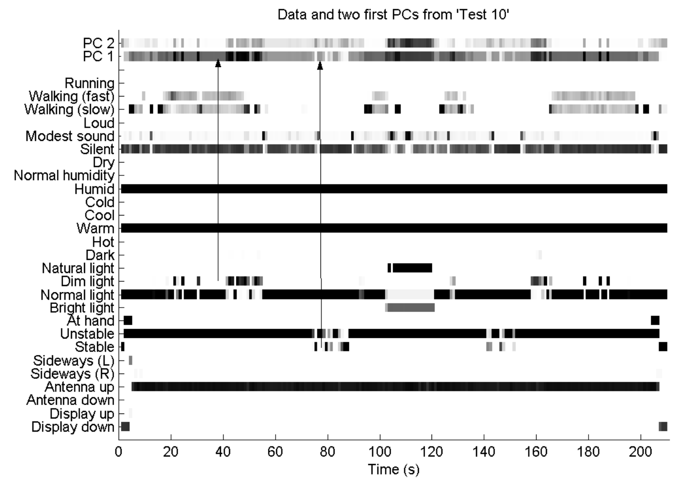

Fig. 1. One specific test ('Test 10') of the context atom data and its two most dominant principal components.

text atoms contain no information in this scenario, and the contribution of the last 10 context atoms to the total variance is negligible. Twelve context atoms out of 27 explain 96% of the data variance. By using PCA, the information content of the data can be represented more compactly. In this experiment, seven most dominant principal components explain 96% of the data variance.

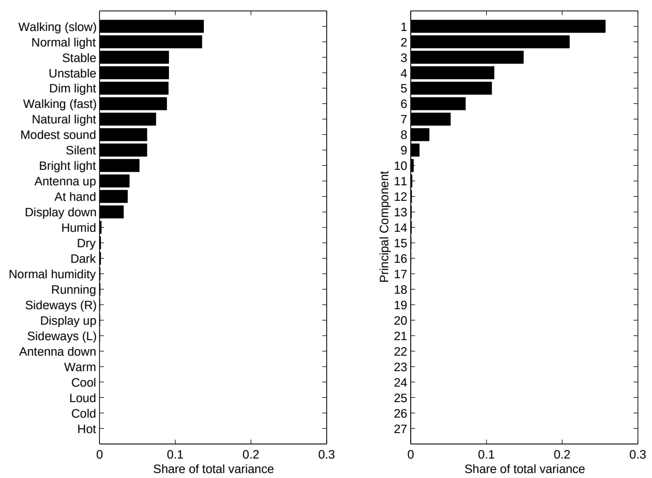

Fig. 2. Left subfigure shows the share of the total data variance for each original context atom. First 12 context atoms explain 96% of the variance. Right subfigure shows the situation for the principal components. Now, seven signals suffice for explaining 96% of the variance.

Seven most dominant PCs can be viewed more closely for example by visualizing them using the projection matrix, which is illustrated in Fig. 4. It can be used to give an overview about how strongly each PC is related to the original signals. Each column represents elements of one principal direction, and each row shows how strongly an original variable and a PC are connected. In order to facilitate comparison between PCs, the elements of a principal direction vector have been scaled by dividing the vector by the largest absolute value of its elements. Black means

TABLE II

Total variance

Context atoms	% of total var.	cum%
Walking (slow)	13.7	13.7
Normal light	13.5	27.2
Stable	13.5	27.2
Unstable	9.1	36.4
Dim light	9.1	45.5
Walking (fast)	9.1	54.5
Natural light	8.9	63.4
Modest sound	7.4	70.8
Silent	6.2	83.3
Bright light	6.2	83.3
Antenna up	5.2	88.5
At hand	3.9	92.5
Display down	3.7	96.1
PCs	**% of total var.**	**cum%**
PC1	25.7	25.7
PC2	21.0	46.7
PC3	14.9	61.5
PC4	11.0	72.5
PC5	10.7	83.2
PC6	7.2	90.4
PC7	5.2	95.6
PC8	2.4	98.0
PC9	1.1	99.2
PC10	0.3	99.5
PC11	0.2	99.7
PC12	0.1	99.8

a strong positive and white a strong negative connection between the original variable and the PC. It can be seen, for example, that especially activity in context atoms 'Dim light', Unstable', 'Walking' and 'Walking Fast' tend to rise the value of PC1. When context atom 'Stable' gets high value PC1 decreases. This can be verified in Fig. 1. To compare PCs of several tests simultaneously, a presentation shown in Fig. 3, is used, depicting the PC1s for tests 1–20. The PC1 has captured the most important phases of the scenario. The phases can be seen as a fluctuation of gray levels giving an impression of vertical stripes, which correspond to the situations in the scenario.

A more detailed view of the PC1 is given in Fig. 5 where the PC1 of 'Test 10' (in Fig. 1) is plotted as a conventional scatterplot as a function of time, and the histogram of the PC1 for all data is presented beside the plot. The scatterplot shows clear discrete levels that describe different phases of the scenario, which correspond to the peaks in the histogram of PC1 for the whole data. Peaks indicate scenario situations such as walking in the corridor (numbers 4,6,8,10 in the scenario description in Tab. I), balcony (7), and in elevator (5,9).

The similarities between events can be observed by rep-

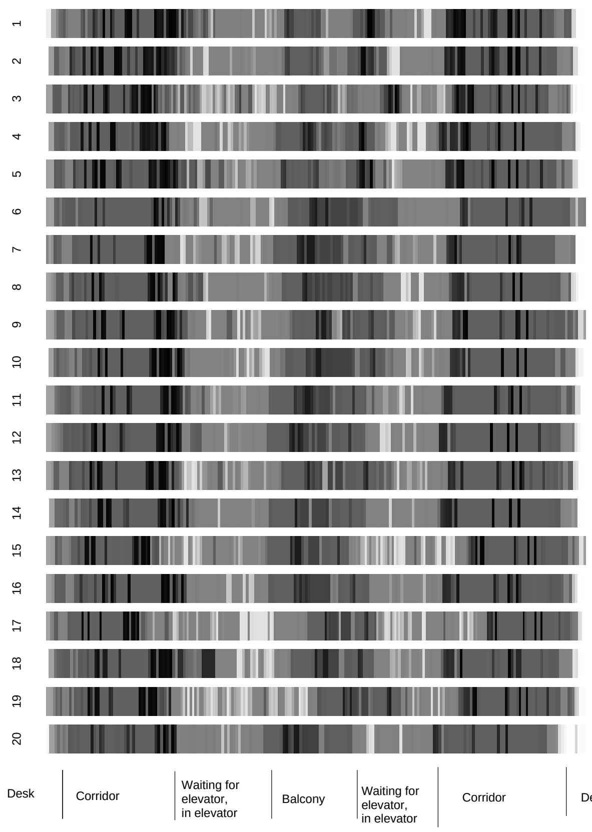

Fig. 3. The first principal component (PC1) for tests 1–20 as gray level bars (white means low value, black high).

resenting a 2D projection of the data by using two PCs, see Fig. 6. The projection reveals clustering behaviour of the data. For example, samples measured outdoors are projected into a very dense area (labeled "Outdoors" in the figure). If two clusters are close to each other, they are similar in terms of the first two PCs which capture 47% of the data variance, see Tab. II. It must be noted that there may be clusters that seem to be near each other, although they are separated in the original data space, since two dominant PCs leave 53% of the variance unexplained. However, if two dominant original signals were plotted in 2D, the representation would capture in maximum only 27% of the data variance (Tab. II).

C. Experiments using ICA

We were interested in finding ICs that would imply that a combination of the original context atom signals (observed signals) would be explained by some hidden source signal (IC). This would tell that there exists an independently appearing context that produces the combination.

The context atom data from each test was pooled together, and the FastICA algorithm[2] was used to extract estimates of the independent components. The data dimension was reduced to 19 to cancel out the strong linear

[2]The algorithm was applied using the default parameter selection proposed by the authors of the program package. Additionally, two different so called contrast functions (kurtosis and skewness) were used. Both of them produced somewhat similar results.

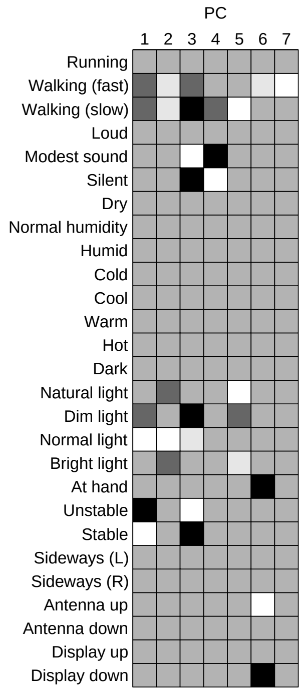

Fig. 4. The PCA projection matrix for context atom data illustrated as gray levels. The gray level scale is made coarse in order to clarify the presentation.

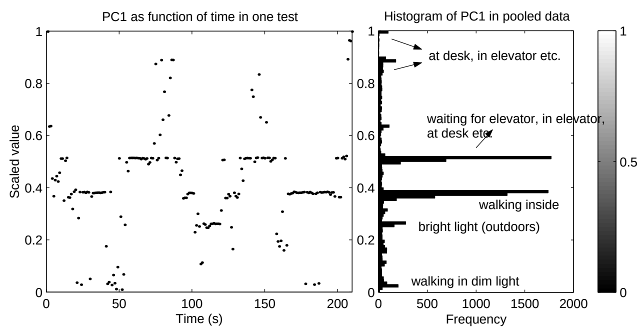

Fig. 5. The PC1 of one test ('Test 10') presented as a scatterplot and a histogram of PC1s for the whole data set.

dependencies of the original signals. In result, 19 ICs were found. It turned out that most of the estimated ICs were almost identical to some original signals. This impression was verified by calculating the correlation between the estimated independent components and the original signals.

The remaining ICs, that were not identical to original signals, originated from rare combinations of events that were mostly uninteresting. This suggests that in our dataset most of the context atoms (signals) are relatively independent when the temporal relations are omitted. On the other hand, it is not guaranteed that the linear ICA model is directly applicable to our data set which is nearly binary valued data. There was, however, one illustrative example of an interesting combination. One IC was strongly related to bright light alone and one to bright *and* natural light. When the ICs were compared to the original data it turned out that, though the only source for

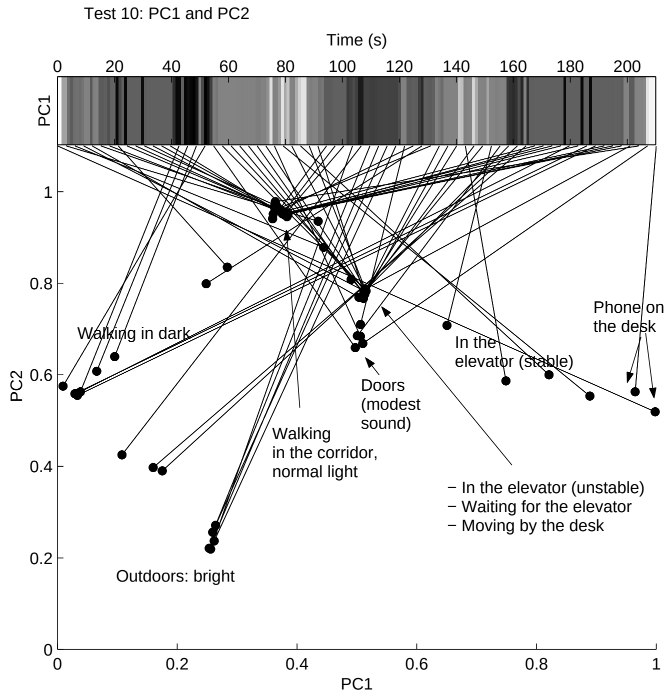

Fig. 6. The figure presents a visualization of the two dominant PCs of the data of 'Test 10' (PC1 and PC2 in Fig. 1). The lines localize each projected data point on the time axis and connect the 2D projection to a 1D gray level bar presentation of PC1 as a function of time. Some evident clusters are labeled.

bright light was natural, there existed also rare samples of "non-natural" bright light. In this case, this was due to a failure to recognize the daylight as natural. The goal of using ICA would, however, be to automatically find this kind of independent latent sources (daylight – artificial bright light).

IV. Summary

The aim of this work was to examine how PCA and ICA could be used to discover contexts that are not explicitly available in context atom data, finding higher abstraction level descriptions, and further compacting the representation of the data.

We experimented using a usage scenario data set collected with a mobile device containing multiple sensors. The use of PCA indicated that the context atom data representation for the scenario was somewhat redundant. Only seven principal components (or 12 original signals) out of 27 were enough to present 96% of the total data variance. Examples were shown on how the dominating principal components can be used to visualize the data of a usage scenario by means of one- or two-dimensional data projections. A few principal components compress the representation of the data, revealing the most important phases of the scenario compactly.

In experiments using FastICA algorithm, most of the estimated ICs were almost identical to some original signals. This suggested that most of the observed signals were quite independent, as far as the temporal relations are omitted. Some rare phenomena were spotted as separate independent components. These phenomena were in this case uninteresting, but the fact that they were found indicates that ICA is potentially useful in analysing context data.

References

[1] P.J. Brown, J.D. Bovey, and X. Chen. Context-aware applications: from the laboratory to the market place. *IEEE Personal Communications*, 4(5):58–64, 1997.

[2] B. Clarkson, K. Mase, and A. Pentland. Recognizing User Context via Wearable Sensors. In *Digest of Papers, The fourth International Symposium on Wearable Computers*, pages 69–76, 2000.

[3] B. Clarkson and A. Pentland. Extracting Context from Environmental Audio. In *Proceedings on 2nd International Symposium on Wearable Computers*, 1998.

[4] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 1994.

[5] S. Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall, second edition edition, 1999.

[6] A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[7] A. Hyvärinen. Survey on Independent Component Analysis. *Neural Computing Surveys*, pages 94–128, 1999.

[8] A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[9] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, (24):1–10, 1991.

[10] K. Van Laerhoven and O. Cakmakci. What Shall We Teach Our Pants? In *Digest of Papers, The fourth International Symposium on Wearable Computers*, pages 77–83, 2000.

[11] T.-W. Lee. *Independent Component Analysis: Theory and Applications.* Kluwer Academic, 1998.

[12] V.-M. Mäntyla, J. Mäntyjärvi, T. Seppänen, and E. Tuulari. Hand Gesture Recognition of a Mobile Device User. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume 1, pages 281–284, 2000.

[13] N. Marmasse and C. Schmandt. Location-Aware Information Delivery with ComMotion. In *Hand Held and Ubiquitous Computing*, number 1927 in Lecture Notes in Computer Science, pages 157–171. Springer-Verlag, Heidelberg, 2000.

[14] Y. Nakanishi, T. Tsuji, M. Ohyama, and K. Hakozaki. Context Aware Messaging Delivery Using Location Information and Schedule Information. *Personal Technologies*, 4(4):221–224, 2000.

[15] F.M. Salam and G. Erten. Sensor Fusion by Principal and Independent Component Decomposion Using Neural Network. In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 215–221, 1999.

[16] A. Schmidt, K.A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde. Advanced Interaction In Context. In *Hand Held and Ubiquitous Computing*, number 1707 in Lecture Notes in Computer Science, pages 89–101. Springer-Verlag, Heidelberg, 1999.

[17] A. Schmidt, A. Takaluoma, and J. Mäntyjärvi. Context Aware Telephony Over WAP. *Personal Technologies*, 4(4):225–229, 2000.

[18] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent Component Analysis for identification of artifacts in Magnetoencephalographic recordings. In *Advances in Neural Information Processing System 10 (Proc. NIPS*97).* MIT Press, 1997.

[19] C. Wren, B. Clarkson, and A. Pentland. Understanding Purposeful Human Motion. In *Proceedings of the fourth International Conference on Automatic Face and Gesture Recognition*, pages 378–383, 2000.

[20] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[21] L.A. Zadeh. *Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs: A precis. Multiple Valued Logic*, volume 1, pages 1–38. Gordon and Breach Science, 1996.