A Preliminary Approach to Study the Causality of Freezing of Gait for Parkinson's: Bayesian Belief Network Approach

A. Saad[1,3], A. Zeineldine[2], I. Zaarour[2], M. Ayache[1], D. Lefebvre[3], F. guerin[3], P. Bejjani[4]

[1]Islamic University of Lebanon, Engineering Faculty, Department of Biomedical, Beirut, Lebanon

[2]Lebanese University, Faculty of business and economical sciences, Doctoral school of Science and technology

[3]Laboratoire Groupe de Recherche en Electrotechnique et Automatique du Havre, Université du Havre France

[4]Director of Parkinson Center, Notre Dame de secours University Hospital, Beirut , Lebanon

Abstract - *Parkinson disease patients suffer from a disabling phenomenon called freezing of gait, which can be described as if their feet are 'frozen' or stuck, but that the top half of their body is still able to move. In this paper, we make a graphical probabilistic modeling study, "Bayesian Belief Network (BBN) approach" of a previously collected dataset that represents the measurements of acceleration sensors placed in the ankle, knee and hip of PD patients during their march. In an attempt to know if this is a traditional BBN model or a causal one, we built a FoG Model and tested it, first by forming an Epidemiological Approach, then, by inferring causal relations based on Additive Noise Models (ANM). Consequently, we built a Bayesian Naive Classifier Model related to FoG. The Bayesian belief Network classifier had the ability to identify the onset of freezing of PD patients, during walking using the extracted features. Promising results appeared into evidence when testing the BNC classifier models*

Keywords: Parkinson Disease, Freezing of Gait, Bayesian Network, Causality, Data Mining.

1   Introduction

Parkinson's disease (PD) is a common neurodegenerative disease. One of PD symptoms is freezing, which may occur during gait, speaking or a repetitive movement like handwriting. Freezing of gait (FoG) can be defined as "a brief, episodic reduction of forward progression of the feet despite the intention to walk", and is often described by patients as if their feet are glued to the floor for a short period of time [1]. FoG aspects of PD do not respond well to dopaminergic drugs, as it is one of the symptoms that often result from non-dopaminergic pathology [2]. Recent studies, investigated measuring features that may evaluate patterns of the handwriting and speech of PD patients and school children [3, 4], which can be used to detect writing and voice freezing episodes for PD patients. This study is oriented to the freezing of gait phenomenon of PD patients. Our

proposal is a modeling approach that focuses on a specific class of Probabilistic Graphical Model (PGM), the directed[1]one, i.e. Bayesian Belief Network (BBN).The followed methodology consists of: (1) assessing the framework of the BBN model, we tried to identify if this is a traditional BBN case [5, 6] or a causal one [7, 8]. (2) By means of the assessed model a classification tool is built, to judge the FoG episodes of PD patients. This classification model can be inferred to diagnosis or forecasting issues. The following part of this paper discusses the explanation and background of the pre-collected dataset, and it gives a clear explanation of the modification done on the dataset. Next, we illustrated a brief state of the art in theories and concepts surrounding the *causality, as a background,* in order to assess a causal link between variables of interest, before building our BBN model. The machine learning approach is described in the fourth part, whereas the obtained results are described and illustrated in the fifth part of the paper. Finally the last part holds the general conclusion that is accomplished.

2   Data Preparation

2.1   Native Dataset

In previous studies, Marc Bächlin et al developed a wearable assistant for Parkinson's disease patients that detects FoG by analyzing frequency components inherent in the body movements, using measurements from on-body acceleration sensors [11]. They used three acceleration sensors positioned in different body parts (*ankle, knee and hip*) each sensor measures three components of acceleration(x: *horizontal forward axis*, y: *the vertical axis and z: the horizontal lateral axis*).Their detection algorithm was based on the principle illustrated by Moore et al that introduced a freeze index (FI) to evaluate the gait condition of PD patients. The FI is a ratio defined as the power in the 'freeze' band [3-8Hz] divided by

---

[1]The alternative classes of Probabilistic Graphical Model are Undirected Markov networks and Hybrid graphs [9], those families of classes are more adapted to statistical physics and computer vision [10].

the power in the 'locomotor' band [0.5-3Hz]. The FoG detection is performed by defining a 'freeze' threshold, where values above this threshold are considered as FoG events [12]. Referring to the data obtained by Marc Bächlin et al from 10 PD patients, we incorporated these values into our probabilistic model in an attempt to predict upcoming FoG episodes. The dataset is composed of separated files for each patient, although some patients have multiple files for each test done. Each file is composed of a matrix that contains measurement data of the three sensors in x, y and z directions. The last column contains the annotation, weather FoG occurred or not. These annotations was labeled by synchronizing the data by a video that recorded each patient run, which allowed to identify the exact start times, durations and end times of FoG episodes.

2.2 Employed features

Starting from the above described dataset, the freezing index for each acceleration measurement is calculated, using a sliding window that calculates the FI of a 256 samples of acceleration data. So we mapped the dataset from raw data to normalized data for generalization purposes in future work. The second step was to eliminate the data which is irrelevant to experiments done (Annotation 0), in order to constrain the classification between occurring of FoG and or NoFoG. Then we calculated the magnitude of the three components of the FIs. Accordingly, all of the measurements taken are represented in a low dimensional dataset, that it is ready to be introduced to our proposed machine learning model.

3 Causality

3.1 Epidemiological Approach

Inferring the causal structure of a set of random variables is a challenging task. In the causality domain, the variables of interest are not just statistically associated with each other, yet there is a causal relationship between them. The famous Slogan "correlation does not imply causation" is recognized and seems approved by researchers in empirical and theoretical sciences. For example, in analyzing a demographic database, we may find that the attributes representing the number of hospitals and the number of car thefts in a region are correlated. This does not mean that one causes the other. Both are actually causally linked to a third attribute, namely, population [2][13]. Formerly, authors in [14] quoted that "one *of the common aims of empirical research in social sciences is to determine the causal relations among a set of variables, and to estimate the relative importance of various causal factors*". Recently, the philosophical wise of this quote is broadly discussed, specifically in the medical and health science, more precisely in the context of Symptoms/Disease episodes [15, 16, 17, 18]. In particular, Lagiou et al. (p 565), mentioned that: *"A factor is a cause of*

a certain disease when alterations in the frequency or intensity of this factor, without concomitant alterations in any other factor, are followed by changes in the frequency of occurrence of the disease, after the passage of a certain time period (incubation, latency, or induction period" [17]. In order to highlight the causal trends of our FoG problem, and from an epidemiological point of view, explicitly we will illustrate the FoG Model (Figure 1) by applying what so-called Hills Criteria of Causation [19], which is an old approach that outlines the minimal conditions needed to establish a causal relationship between two items. Hill's work has been recently validated by, Kundi (2006, p. 970) as a valuable tool, since both mechanistic and probabilistic aspects were considered [20]. Kundi applied Hill's criteria to the classic case of smoking and lung cancer. The first step for examining our causal proposal was to test if our study is consistent with Hill's criteria. Table I summarizes the nine criteria defined by Hill and the observations when applying it on the FoG case with respect to freezing index. It can be clearly observed that not all of the criteria hold in our case, where criteria (4 and 9) weren't applicable. On the other hand, the other criteria weren't as satisfactory as expected.

TABLE I. Observations based on Hill's criteria for FoG

Criterion	FoG correlation with freezing of index
1.Strength of Association	As FoG episodes occur, the value of the freezing index in higher than that when normal gait is happening.
2.Temporality	FoG in the vast majority of cases occurs when the freezing index increases.
3.Consistency	Several studies were applied on different patients, which produced the same results. The relationship also appeared for different genders.
4.Theoretical Plausibility	We don't have an explained biological theory stating a theoretical relationship between freezing index and FoG.
5.Coherence	The conclusion (that accretion of freezing index causes FoG) "made sense" given the knowledge about the algorithm for calculating the freezing index with respect to FoG occurrence.
6.Specificity in the causes	Freezing index is one of the clinical features (not the only one) that can be used to predict FoG.
7.Dose Response Relationship	Extracted data showed that there is a direct relationship between the value of the freezing index and the occurrence of FoG episodes.
8.Experimental Evidence	The experimental data collected clinically from patients made certain that FoG occurs when the freezing index increases.
9.Analogy	In this case, contrasting similar phenomena could not be applied, due to the fact that the approach of detecting causality of FoG is novel.

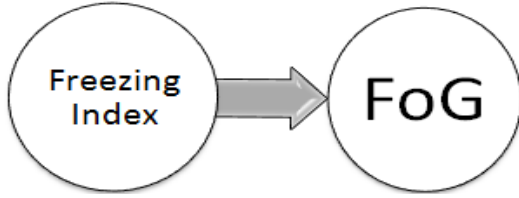[2] This example is fully inspired from [13] p. 68

Figure 1. FoG causal model

3.2   BBN approach

The controversial debate on causality is still widely discussed in machine learning, probability theory and artificial intelligence. Several studies proposed causal discovery methods in the framework of BBN [8, 21, 22, 23, 24]. In this context, the causality issues have been studied by discovering the structure of BBN and it needs interventional data in cases where purely observational data is inadequate [10]; In general, the relation between causality and probability is based on a set of assumptions that allow the causal inference [25], and those assumptions are: (1) Causal sufficiency, (2) Markov, and (3) Faithfulness (definition of these assumptions are briefly mentioned in [10]). One of the known approaches to causal discovery is the So-called constraint-based approaches [8, 26], that select all direct acyclic graphs (DAGs) which satisfy the second and third assumption. In order to evaluate the causal link between our employed features we refer to a recent study that infers causal relations based on additive noise models (ANM). Jonas et al[27] published an algorithm that "able to distinguish between cause and effect, for a finite sample of discrete variables, and works both on synthetic and real data sets. The principle is that whenever the joint distribution P(X; Y) admits such a model in one direction, e.g. but does not admit the reversed model, one infers the former direction to be causal (i.e. X →Y)". Briefly, this algorithm tests whether the data admits an additive noise model by checking all possible functions and test whether they result in independent residuals. Applying Jonas et al causal inference method resulted that no causal relationships can be applied between any of our variables and between FoG.

4   Bayesian Naïve Classifier

Data mining is the science of extracting useful information from large data sets. It covers areas of machine learning, pattern recognition, artificial intelligence, and other areas [28]. One of data mining main objectives is prediction, which involves using some variables in data sets in order to predict unknown values of other relevant variables (*e.g. classification, regression, and anomaly detection*) [29].We already initialized the process of building a BBN model (section3.1 and 3.2) by studying the type of relationship between the Freezing Index concept and FoG episode via Hill's rule, and among features themselves via (ANM) model. Those two methodologies didn't validate the causality

behavior between Freezing Index and FoG. Hence, we assume that BBN structure will depict a simple correlation between variables and FoG, and we will study the FoG episode via the simplest and traditional way of Classification Model where the FoG can be simply inferred to diagnosis or forecasting issues, specifically we tended to use the Bayesian Naïve Classifier (BNC), which is one of the most effective and popular classifiers in data mining techniques [13, 30]. It has been successfully applied to the different problem domains of classification task such as intrusion detection, image and pattern recognition, medical diagnosis, loan approval and bioinformatics [31].

4.1   Classification protocol

4.1.1   Learning

The first step of our learning protocol was to divide the previously described datasets (section 2.1), some for learning (9 datasets each for different patient) and the rest for testing. Thus, we built 9 BNC Models for nine different patients. For this purpose, 9 Belief network graphs were constructed (Figure 2), where the class node (FoG) will be the parent of the three FI nodes (FI nodes represents the magnitude of each acceleration sensor). Although, the data intended to learn each BNC model, was divided into 70% learning data and 30% testing data. The difference between the 9 BNC models is the conditional probability that will be learned according the training set introduced to the BNC model. Continuous variables have been discretized based on Akaike's criterion [32].The learning experiments were conducted with a random 10-fold validation; each fold takes a random 70% from the data set for learning and the remaining 30% for testing.
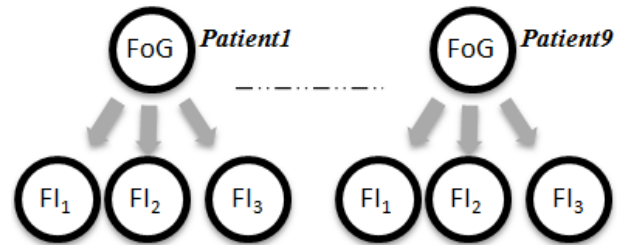

Figure 2. Nine BNC Models for each PD patient

4.1.2   Testing

After learning each fold, a confusion matrix was calculated (Table II) using the test data, the table represents the ***true positives*** (TP), ***false positives***(FP), ***false negatives***(FN) and the ***true negative*** (TN).From the confusion matrix, we evaluated three important values: FoG-precision, NoFoG-precision and Accuracy.

TABLE II. Confusion matrix calculated for each random fold.

Real classification	Model classification		
	FoG	True	False
	True	TP	FN
	False	FP	TN

Subsequently, and after calculating the above listed values for each fold, we choose the learning that holds the highest three values by referring to the priority of each value (starting by FoG-precision as highest priority followed by NoFoG-precision and finally Accuracy).After learning the nine BNC Model for 9 different patients, the rest datasets was introduced to each BNC model as testing datasets, for the purpose of testing the degree of generalization of our models. Also from each test dataset the confusion matrix, FoG-precision, NoFoG-precision and Accuracy were evaluated. In addition we made another testing approach, which is to enter each data sample as a parallel input to every one of the 9 BNC models, and the final decision that classifies whether a FoG or NoFoG is occurring, is based on the most likely decision made by all BNC models individually. For example, if 5 models decide that this sample is FoG and the rest do not, the final decision is taken as FoG.

5   Results

Following the learning and testing protocol, figure 3 summarizes the obtained result as function of FoG precession and NoFoG precision. Datasets were represented by "S<patient number>R<test or run number">.It is noticed when testing S01R01 (patient 1, first run) the FoG precision value was apparently high in all nine classifiers. Although the NoFoG precision values were low for some patients, yet this result showed that our classifier was able to detect every FoG episode with high precision with average of FoG precision 79.5%. In addition, if we take into consideration both FoG and NoFoG precision values; we can see that the best results were for datasetsS01R02 (FoG precision=70.67% and NoFoG precision=84.74%) and S03R01 (FoG precision=73.68% and NoFoG precision=79.13%), where the first dataset is for the same patient but on a different run while the second dataset is for another patient, this shows that both patients maybe correlated in freezing behavior. As for dataset S02R02, some results had low FoG precision; this may be due to the different walking behavior of patients, knowing that S02R01 (same patient but different run) showed an acceptable result for NoFoG precision and a very high result for FoG precision(92.85%).

Results for S05R02 showed that this patient may have a unique freezing behavior among the other learned patients, that's why none of the nine BNC models were able to differentiate this patient's freezing episodes from normal gait with high precision. Finally, for the dataset S06R01, some results had very good FoG precision about 89%. The best results were for S07R02 and S05R01 since they have moderate FoG and NoFoG precision. This may be due to the

similarity in FoG behavior between the two patients. The testing results of each dataset can be summarized by calculating the *average* for Accuracy, FoG precision and NoFoG precision (Table III). We can see that our system's accuracy is about 66.87 % with FoG precision 59.34% and NoFoG precision 69.24%.
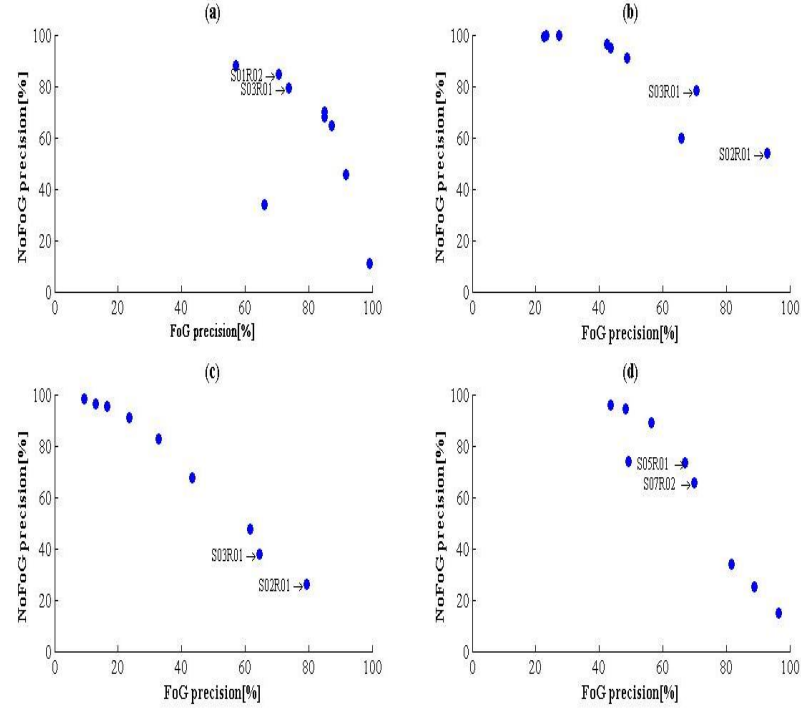

Figure 3. FoG precision vs. NoFoG precision results for testing datasets, (a)S01R01, (b)S02R01, (c)S05R02 and (d)S06R01.

Table IV shows the results for the second approach of testing which is making a decision based on the most likely one made by all BNC models individually. We can see that the accuracy and NoFoG precision increased and the FoG precision slightly decreased.

TABLE III. First approach averaged system accuracy.

Average	NoFoG precision (%)	FoG precision (%)	Accuracy (%)
S01R01	59.07	81.20	60.10
S02R02	85.22	50.31	80.87
S05R02	71.42	39.20	65.10
S06R01	61.23	66.66	61.40
System accuracy	**69.24**	**59.34**	**66.87**

TABLE IV. Second approach averaged system accuracy.

Average	NoFoG precision (%)	FoG precision (%)	Accuracy (%)
S01R01	65.21	85.71	66.16
S02R02	94.97	43.65	88.58
S05R02	83.03	33.74	73.36
S06R01	69.23	67.85	69.13
System accuracy	**78.11**	**57.74**	**74.31**

6 Conclusion

We have described a way for modeling freezing of gait phenomena of PD patients, based on BBN formalism. We made use of a dataset available online extracted from real PD patients while walking and having freezing episodes. The first approach, was studying the causality in the FoG/freezing index system, this was done by making an Epidemiological study followed by Causal inference one. This approach resulted in weak or no causality in FoG/Freezing Index system. Although, this can be further studied in future by calculating more features that may define FoG better. This result lead to a second approach which was applying Bayesian Naive classifier model to represent the datasets, we built 9 different BNC models for different patients, and the remaining datasets were introduced to each BNC model as testing datasets. This approach showed a fluctuating percentage of accuracy, FoG precision and NoFoG precision. Our classifier had the ability to detect FoG up to 99 %(FoG precision) if tested on the 9 BNC models locally, and up to 86% if tested globally. Some testing results were not as expected, we assume this was because of the different freezing behavior in different patient, knowing that when testing a specific BNC model related to a specific patient, with a dataset extracted from the same patient the result was significantly improved.

7 References

[1] N. Giladi, D. McMahon, S. Przedborski, E. Flaster, S. Guillory, V. Kostic, and S. Fahn, "Motor Blocks in Parkinson's Disease", PubMed, Neurology, Vol. 42, No.3, 1992.

[2] Rajesh Pahwa, Kelly E Lyons. Handbook of Parkinson's disease. 4thedition (2007).

[3] Ali Saad, IyadZaarour , Paul Bejjani , and Mohammad Ayache. Handwriting and Speech Prototypes of Parkinson Patients: Belief Network Approach. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012

[4] I. Zaarour, L. Heutte, PH. Leray, and J. Labiche, "Clustering and Bayesian Network Approaches For Discovering Handwriting Strategies of Primary School Children", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 18, No. 7, 2004.

[5] Judea Pearl, "Fusion, Propagation, and Structuring in Belief Networks", Artificial Intelligence, Elsevier, Vol. 29, No. 3, pp.241-288, 1986.

[6] Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers Inc., San Francisco, A, USA, 1988.

[7] C. Glymour, and G. Cooper, "Computation, Causation, and Discovery", MIT Press, Cambridge, MA, 1999.

[8] Judea Pearl, "Causality: Models, Reasoning and Inference", Cambridge University Press, Cambridge, MA, 2000.

[9] Steffen L. Lauritzen, "Graphical Models", Clarendon Press, Oxford,UK, 1996.

[10] Montassar B. Messaoud, "SemCaDo: An Approach for Serendipitous Causal Discovery and Ontology Evolution", PhD thesis, Ecole Polytechnique de l'Université de Nantes, 2012.

[11] Marc Bächlin, Meir Plotnik, Daniel Roggen, InbalMaidan, Jeffrey M. Hausdorff, NirGiladi, and Gerhard Tröster, "Wearable Assistant for Parkinson's Disease Patients with the Freezing of Gait Symptom", IEEE Transactions on Information Technology in Biomedicine, Vol. 14 No. 2, pp 436-446, 2010.

[12] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory Monitoring of Freezing of Gait in Parkinson's Disease," Journal of Neuroscience Methods, Vol. 167, No. 2, pp. 340-348, 2008.

[13] Jiawei Han, and Micheline Kamber, "Data Mining: Concepts and Techniques", 2nd Edition, University of Illinois at Urbana-Champaign, Morgan Kaufman Publishers, Elsevier Inc., USA, 2006.

[14] Spirtes Peter, Glymour Clark, and Scheines Richard, "Causality from Probability", Department of Philosophy, Report Paper 236, 1990.

[15] Federica Russo, and Jon Williamson, "Interpreting Causality in the Health Sciences", Research Project, International Studies in the Philosophy of Science Vol. 21, No. 2, pp. 157-170, July 2007.

[16] H. Frumkin, "Causation in medicine, Emory University, Rollins School of Public Health, Emory University, Atlanta, Georgia, 2006.

[17] P. Lagiou, H.O Adam., and D. Trichopoulos, "Causality in Cancer Epidemiology", European Journal of Epidemiology, Vo. 20, pp.565-574, 2005.

[18] P. Thagard, Explaining Disease: Correlations, Causes, and Mechanisms" Journal of Minds and Machines, Vol. 8, pp.61-78, 1998.

[19] B. Hill, "The Environment of Disease: Association or Causation?" Proceedings of the Royal Society of Medicine, Vol. 58, pp.295-300, 1965.

[20] M. Kundi, "Causality and the Interpretation of Epidemiological Evidence", Environmental Health Perspectives, Vol. 114, pp. 969-974, 2006.

[21] R.E. Neapolitan, "Learning Bayesian Networks", Prentice Hall Series in Artificial Intelligence, 1st Edition, 2003.

[22] J. Zhang, and P. Spirtes, "Detection of Unfaithfulness and Robust Causal Inference", Journal of Minds and Machines, Vol. 18, No.2, pp.239-271, June 2008.

[23] J. Pellet, and A. Elisseef, "Using Markov Blankets for Causal Structure Learning", Journal of Machine Learning Research, Vol. 9, pp.1295-1342, 2008.

[24] S. Nagl, M. Williams, and J. Williamson, "Objective Bayesian Nets for Systems Modeling and Prognosis in Breast Cancer", studies in computational sciences, Vol. 156, pp 131-167, Springer Berlin Heidelberg, 2008.

[25] Marek J. Drudzel, and Herbert A. Simon, "Causality in Bayesian Belief Networks", In Proceedings of the 9thAnnual Conference on Uncertainty in Artificial Intelligence (UAI),1993.

[26] P. Spirtes, and C. Glymour, and R. Scheines, "Causation, prediction, and search", Lecture notes in statistics, New York, NY, Springer-Verlag, 1993.

[27] Jonas Peters, Dominik Janzing, and Bernhard Scholkopf, "Identifying Cause and Effect on Discrete Data using Additive Noise Models, Appearing in Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS) ,Vol. 9,2010.

[28] Adem Karahoca, "Data Mining Applications in Engineering and Medicine", 2012.

[29] F. Gorunescu, "Data Mining: Concepts, Models, and Techniques", India, Springer, 2011.

[30] R.Duda, and P. Hart, Pattern Classification and scene analysis, John Wiley & Sons, 1973.

[31] R. O. Duda et al, Pattern Classification, 2nd ed. Chichester, U.K.: Wiley-Interscience, 2000.

[32] D. Song, C. E. Henrik, K. Huebner, and D. Kragic, "Multivariate Discretization for Bayesian Network structure learning in robot grasping", in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011, Vol. 11. pp. 20-27.