



Facultad de Ciencias,
Ciudad Universitaria,
UNAM, México, D.F.

September 9, 2016

IEEE Transactions on Cognitive and Developmental Systems
Special Issue on Sensorimotor Contingencies for Cognitive Robotics

Dear Dr. Ricardo Téllez, Dr Guillem Alenyà,

Please find attached our revised submission:

"A Multi-Modal Model for Prediction and Classification of Object Deformation during
Robotic Manipulation"

Veronica E. Arriola-Rios (UNAM, Mexico), Jeremy L. Wyatt (University of
Birmingham, UK)

We would like to thank the reviewer for their insightful comments, which have enabled us to significantly improve our paper. We have made our best effort to address them. In the following pages we explain in some detail the changes we made. We have also suggested a slight change to the title make it clear, as the reviewer pointed out, that the paper deals solely with pushing actions and not manipulation in general.

Before we detail the precise changes made, we also wish to apologise because, in spite of our efforts, we still have not been able to implement the requested comparison with another existing method, as any fair comparison requires an extension to the compared method(s). This extension is necessary because none of the methods were designed to work under qualitatively the same conditions as ours. Because of this we are still working on an extension of another method, and its implementation. We expect to complete it within a month. To this end, we wrote to you to request an extension, but we have not yet received a reply. Therefore, to conform to the deadline we were given, we are submitting the paper with all the textual revisions completed. We would like to submit a further version that includes the results of our comparison. We request an extension of four weeks to allow this. Please let us know at your convenience whether this is acceptable.

All of the changed parts of the manuscript have been highlighted in red in the re-submitted version. We also made some minor changes to the grammar in addition to the requested changes. We have not marked these.

We thank you again for your very helpful comments on our paper. They have helped us to significantly improve it. We hope that we have satisfactorily addressed all your requested changes to the text. We will submit a second revision with the experimental

comparison as soon as possible. We look forward to hearing from you.

Yours sincerely,

Dra. Verónica Esther Arriola Ríos
Profesora Asociada C de T.C.
Departamento de Matemáticas, Cub
119.
Facultad de Ciencias, UNAM
v.arriola@ciencias.unam.mx
+(52)55 5622 5426
Veronica E. Arriola

Reviewer's Comments (and responses in blue):

This article aims at tracking three-dimensional deformable objects under robot motion and predicting the force causing the deformation. For these purposes, the authors propose two models, including an extension of the mass-spring model and a regression model, where the parameters of the mass-spring model are learned with an evolutionary searching algorithm on visual data and force data, and the parameters of the regression model are learned with least-square methods on stress-strain graph. These models are tested on a sponge and a plasticine, and evaluated based on the balanced F-score. The results are presented both quantitatively and qualitatively.

However, there are several structural problems, as listed below.

1: in the abstract, the authors claim that five features are simultaneously enabled by the proposed single learning model framework, but the article actually presents a collection of different models and methods to solve different problems, thus it is inappropriate to claim them as a single unified model framework. Besides, it seems that only robot pushing is modeled, not general "robot actions".

Thank you for pointing this out. We have rewritten the model overview section, parts of the abstract, and the introduction to make this clearer. We emphasise that on the one hand we do learn the two models separately, but on the other we also emphasise that they are combined in prediction, and further that exactly the same core prediction model is used to perform classification. Thus although we do not claim them as a single model, we do want to claim them as several models linked in a over-arching framework. We have made clear the exact structures in three new figures.

2: section II is not included in the final paragraph of section I, and this section should be included in section I or section IV.

We have now mentioned this.

In addition, this section is really difficult to follow. For example, in line 27 of page 1, "the specific approach taken here separates the predictive model out into components in two ways". It will be easier to read if the authors clearly indicate the second way in line 44. And the third paragraph, on model training, should come before the second graph, which is on model testing.

Thanks for pointing this out. We agree, and have completely rewritten section II. In part this addresses this point, and in part it makes clearer the multiple machines that we have in our framework. We hope that we have made it much clearer.

3: section IV-C, "Integration Scheme", seems to be part of algorithm 1 or section IV-B, thus it should be introduced.

We are sorry this was not clear. In fact it is neither part of algorithm 1, nor of section IV-B. It is simply part of a numerical algorithm for solving differential equations. We hope the new introduction to section IV will help to make it clearer.

4, section IV-D is not explained: what is the relation between "collision detection" and the proposed model?

We have now rewritten the start of section IV-D to make this clear. Essentially, the simulation must not allow interpenetration between objects. Thus, corrections are made to the mass-spring predictions at run time, to respect these constraints. Detection of violations requires some collision checking.

5, section IV-E, "Geometric Constraints", should be integrated into section IV-A, where the energy function is built on different geometric constraints.

Thank you for the observation. We would very much like to keep the separation of the two sections, as the two constraints are not at all the same, and the additional constraints in section IV-E are both optional and separate from those in the mass-spring model (section IV-A). Those belonging to the mass-spring model are soft-constraints implemented through an energy function. The additional routines in section IV-E are independent of the mass spring model. These are hard constraints, being purely kinematic, and are rather more closely related to the collision detection techniques described in IV-D. We changed the title of section IV-E to reflect this and further explained this difference in the text.

6, section IV-F, "Many Step Prediction", explains how to recursively train and test the model. It should be included in section V, where the model training is given.

Sorry the text was confusing. Section IV-F does not depend on training. As a physics based model, the dynamics of the simulation depend only on proper entries and parameters, which could have been provided by a human expert after manual calibration or by a computer algorithm. The introduction to section IV and the last paragraph of this section was modified to emphasise this.

7, section V-A explains how to search for the optimal parameter with evolutionary algorithm based on the ground truth; section V-B explains how to get the ground truth; section V-C explains how to test the learned parameter. It seems that the correct order is B-A-C.

We have changed the order as suggested.

Besides, there are several technical ambiguities in the manuscript, as listed below.

1, in line 37 of page 1, "sensorimotor contingencies" needs to be defined or referenced. Now the first paragraph of the Introduction contains a definition.

2, in line 35 of page 1, what does "sequencing the models by mode" mean? This has been addressed as part of the rewrite of the Model Overview section.

3, in line 58 of page 1, the authors claim "In each case the models can be run in two modes, prediction and classification". But the classification mode is not explained later, but a filtering mode is included instead. Besides, please clearly illustrate the mentioned modes and processes in Figure 2.

We hope that we have now addressed this in the model overview section. We do not have a filtering mode as such. Instead we run models for each material in prediction model, and compare their predictions to the ground truth. This is the classification mode. This should now be much clearer, especially from the addition of new figures that show the prediction machine (Fig 3.) and the classification machine built on top of it (Fig 4.).

4, in section IV-A, the author needs to explain how they extend the model proposed in [16].

Since it was not clear the wording of the first paragraph of section IV-A has been modified, where the general idea is explained, the details come in later paragraphs.

5, in section IV-A, the author needs to briefly introduce the mass-spring model.

We have added an introduction at the beginning of section IV.

6, in equation (1), please explain what "particle i " is. Is it vertex i ?

The introduction added to address comment 5 covers this issue as well.

7, in line 27 of page 4, please explain what "mass particle" is.

The term was introduced in the second paragraph of section IV.

8, in line 33 of page 4, please explain why the equations are oscillatory.

We have added an explanation that it happens when/because we use C s that model springs.

9, please explain how equation (3) is derived.

It is not within the scope of the article to explain that derivation, since it belongs to Teschner's model, and it would take too much space. We have, however included a reference to the article for the reader to follow up.

10, in line 54 of page 4, please add reference to "Teschner normalised the value dividing by D_0 to make the elasticity constants scale independent."

Added.

11, in line 13, please explain "Since Teschner and Morris didn't find it helpful to add damping to the preservation of areas, it is avoided here".

We have changed this phrase to make clearer the explanation that Teschner gave for this.

12, in line 50 of page 4, please explain "the angles between adjacent edges". As adjacent edges form one angle, do the authors mean adjacent angles? Similar ambiguity exists in line 2 of page 5.

Corrected.

13, in Algorithm 1, please explain how "an integration step" leads to new length of the edge.

We have re-worded the sentence to add more detail.

14, in line 25 of page 4, please explain what "Euler or Verlet" means. Euler or Verlet?

It was "or", thanks. Corrected.

15, in equation (17), please explain if $x(t+h)$ is a position, how does $h * \frac{F(t)}{m}$ generate a position vector, since $h * \frac{F(t)}{m}$ seems to be the production of h and acceleration? And please explain "we made the velocities proportional to the forces". Which variable is the velocity? What does "m" mean? The mass?

We have explained this in more detail and changed the notation to make it consistent with previous paragraphs.

16, in section IV-D, please explain what "collision detection" means, the collision between robot and objects, or the collision between objects?

We meant the collisions between both the deformed object and the finger, and between the deformed object and the table. We have now improved the description across the whole subsection to make this clear.

17, is algorithm 2 designed by the authors?

No, it can be found in many places. We changed the word "common" in the text, for "standard", which should make this clear.

18, in section V-A, please explain why evolutionary algorithm is used. In algorithm 4, please explain what $\alpha, \beta, \gamma, \zeta$ represent.

We have added this justification at the beginning of the section. The meaning of $\alpha, \beta, \gamma, \zeta$ is covered in lines 7,9,10,11 of Algorithm 5. We have also added a short description in the main text when they are first introduced.

19, in line 60 of page 7, please explain how the object is segmented from the environment, which "simple colour segmentation is used", and how to define "edges that not likely to belong to the object of interest".

We have changed this paragraph to make it clear.

20, in line 23 of page 7, please explain what "linear snake" indicates. Is it from active contour model? Also, please add reference.

We have added a reference for this, and explained the relationship with active contours.

21, in section V-C, please note the measurement has a name, "F-score", in machine learning community. Also, please define $\mu(f)$ based on "the mean over all the frames so far".

Thanks. We now use the name "F-score" and include an equation to define $\mu(f)$

22, in line 30 of page 8, please explain how the stress-strain diagram is obtained, and how the training data set is collected.

We have added information on both these points.

23, in the experiments, please compare with other existing methods.

We are still working on this. We requested an extension, as this requires more time than we had available to do the revisions. We haven't received a reply to date. So we are submitting the paper with the textual changes now. If these textual changes are satisfactory we will submit a second revision with the comparison as soon as possible. We estimate this will take four weeks from today. We are sorry that despite our best efforts this has taken longer than we expected.

24, in Table II, please explain the meaning of the values. Are they mean values and standard deviations? If so, what does "best, worst, average" mean?

We have added explanations of these values.

25, in Figure 12, please clarify the legends. For example, why are there two lines with the same marker?

26, in line 56 of page 10, please explain "integration step". Is it the "h" in equation 17?

Yes, it is h , the name of the variable has been added to make this clear.

27, in section VII-E, please explain how the model is used in object classification. For example, what does "even for novel interactions" mean? Object classification typically tries to identify an object in a new environment. Besides, how is "Global" criteria evaluated in object classification?

We have replaced "for novel interactions" with a more concrete description.

Here we are not focusing on classifying static objects in new environments, but rather in classifying the material an object is made of when the entry is its behaviour in time. Visually, the blocks are actually very similar prior to deformation, therefore the most relevant difference lies in their dynamic behaviour. That is why the classification task looks different from traditional static datasets. Instead of a large collection of labeled images in different backgrounds, we used six videos of two different materials and asked the system to identify which material it looks at through the videos. The explanation in the article has been modified to be more specific about this.

28, in Table III, what does "test 1, test 2" mean? Please explain how these numbers are computed. Besides, object classification is typically evaluated with precision-recall curve.

An explanation about "training", "test 1, test 2" has been added.

Since there are only two materials, the number of false positives for one material is exactly the number of false negatives for the other material, that is why we considered that recall would be redundant.

29, please include more related works, since there are tons of articles on deformable object tracking.

We have added references on deformable object tracking to the background section. We have also made clear why the use of one-step prediction in those algorithms is different from our multi-step prediction problem.