Response to reviewers' comments

Reviewer 1

In this revision, the authors solved the initial problems on article structure, citations, figures, and contents.

We thank the reviewer for their confirmation that we have satisfied all their comments, and would like to thank them again for their remarks, which significantly improved our paper.

Reviewer 2

Page 1. Last Paragraph. There is an enumeration of novel contributions, you mentioned the First, and the Third, to avoid any confusion I would recommend to remark which is the second.

We apologise, this was a typo. We have listed two main technical contributions.

Page 1-2. It is possible to infer the meaning of FEM, but in order to avoid any possible confusion you might indicate its meaning explicitly when the concept is first used.

We have fixed this.

I find Fig 2, 3, and 4 difficult to read, however I don't consider they should be necessarily changed.

We agree this was difficult to interpret. To aid the reader we have added some text descriptions for the most important elements to make the diagram more readable.

At the end of subsection VII. B, when the authors refer to "all elastic and plastic terms" might be illustrative to add references to the corresponding equations.

Vero?

Within the Model Overview, it would be interesting if the authors would briefly highlight the advantage of having generative models for those who are not vastly familiarized with the topic.

Thank you for this suggestion. We have added a short explanation within the section Model Overview. This explains that generative models are appealing because they allow us to solve both prediction and classification problems, whereas discriminative models are typically only suitable for classification.

When interpreting results what are the consequences when triangles in the border of the mesh overlap each other? Is there any consequence that might be significant to the result interpretation?

Vero?

In section V, during the training the segmentation assumes that the color of the object is known. Therefore, when you say that the algorithm could learn and classify previously unknown materials, at least the machine must partially know the material. Also the authors mention that the simplified active contours approach is not the most accurate, What are the consequences of this approximation? Regarding to the chosen integration time step, does the results vary considerably when 0.01 is considered?

Vero?

In section VII.E, Table III, Is there any special reason to have the same result for the training and test1 sets in the case of the sponge? It would be better if the authors offer deeper insights regarding the interpretation of the figures shown in Table III.

Vero?

Reviewer 3

ABSTRACT and INTRODUCTION

Reading the abstract and the introduction, I was expecting a framework in which the robot creates a representation of the environment (in this case the two objects) while it is interacting with it (looking at and pushing the object). As the paper shows, the relationship between the finger position and the sensed force depends on the material, so to learn correctly this relationship the robot has to create a representation of the object itself. Even if not applied to object deformation, this concept has been applied for example in

Wörgötter, F., Agostini, A., Krüger, N., Shylo, N. and Porr, B., 2009. Cognitive agents—a procedural perspective relying on the predictability of Object-Action-Complexes (OACs). Robotics and Autonomous Systems, 57(4), pp.420-432.

Antonelli, M., Gibaldi, A., Beuth, F., Duran, A.J., Canessa, A., Chessa, M., Solari, F., Del Pobil, A.P., Hamker, F., Chinellato, E. and Sabatini, S.P., 2014. A hierarchical system for a distributed representation of the peripersonal space of a humanoid robot. IEEE Transactions on Autonomous Mental Development, 6(4), pp.259-273.

Main issue in this kind of approaches is the online learning of the representation. Indeed, the robot has to identify the object in order to update the model of the object, but the model of the object is required to identify the object itself. This problem can be addressed for example by using a generative model such as

Chandrapala, T.N. and Shi, B.E., 2015. Learning slowness in a sparse model of invariant feature detection. Neural computation.

While the authors recognize the issue of the on-line learning in the introduction (page 1, col 2, line 35), they do not address that in the paper. Also they refer the problem of getting access to partial views of the objects but it seems that in this paper the object is viewed from a convenient point of view. Finally, the word "learning from data" suggests to me that the model acquires a representation of the environment based on some kind on non-parametric model, such as neural network. However, the models seem carefully designed to represent the object deformation and the stress-gain diagram. The few parameters of these models (8 in the case of the mass- spring model, right?) are obtained off-line using a genetic algorithm (GA). Despite the use of the GA, the method seems to me closer to
a calibration procedure than a learning-based approach (No on-line learning
and no adaptation). Also in the stress-strain diagram the author selected the type of feature space (linear vs logarithmic) depending on the material.

What happen if we add a new material? Do we need to add a new feature space? Even if the difference between learning and calibration is subtle, in my opinion, the use of the word "learning" creates an

expectation that is not fulfilled in the rest of the paper.

I would add a figure like Fig. 9a at this point. In this way the reader can have an idea of the setup, how vision acquire information and how the mass-spring model depends on the visual information itself. For example, it might be interesting to combine Fig 9a and Fig 2 in single sketch of the model.

Thank you for this suggestion. We have prepared a new version of Figure 2, as you suggest This incorporates elements from the old Figure 2 and Figure 9.

A consideration about the organisation of the following sections. Given that the framework consists of two models, Force prediction and Shape
prediction, I would expect section III.1 to be Force prediction and III.2
to the shape prediction (or vice-versa). However, the organization of
Sections IV, V, VI, in my opinion is misleading.

Finally, I would start describing the algorithm from the visual system. It seems that the visual system has a very important role in this paper but it is not highlighted. Indeed, the mass-spring model described is section IV should be created within the perimeter of the segmented object. Thus, it requires a visual system capable of localizing and segmenting the object.
In my understanding, this information is necessary to initialize the position of the particles (described in section IV) inside the object. Regarding this point, I would like to know how the particles are placed inside the object. What happen if we change the size of the object? Do we need to change the number of particles? What happen if the object is not rectangular?

IV. A MASS-SPRING MODEL OF DEFORMATION

My main concern about this section is Eq. 19. I really do not understand how the force could be proportional to the velocity. This point should be clarified. My understanding is that the authors assumed the acceleration (a) to be constant during the interval h, so that h*F = h*a*m => F = v*m/h.
Given that h is a constant, the force (F), the velocity (v) or the mass (m)
are recovered with a scale factor. Is that correct? If so, what is the difference between this approach and the quasi-static approach proposed by Frank. Please, explain.

Other considerations about this section. This section is quite long and I think it could be simplified notably. The authors could simply state the energy E = Sum(i=1:N) [0.5 k_i * C_i(p1, pn)^2], where N is the number of constraints. There are several constraints to preserve the length, the area, etc. and the force is given as negative gradient of the potential energy with respect to the position of the particle. Details about all the equation could be reported in an appendix to make the paper more readable. About the equation, personally I don't like the symbol composed of multiple letters such as Area, forcemag, forcedir, etc.

VII. EXPERIMENTAL RESULTS

Do the authors use just one video for the training and two video for the testing? It seems that the dataset is quite small to have a systematic
test of the algorithm. Even if there are only two objects, the authors should test different motions of the finger (i.e. pushing velocity, penetration, …)

I don’t see the results of the classification. **I would expect a confusion matrix to show how many times the sponge was classified as sponge and as plasticine and vice-versa. Table III seems to show the results about the classification of the pixels.**

Moreover, I would like to see a comparison against a baseline approach. They could some of the models proposed in the literature or some simple models. For example, the authors could assume that the material “disappears” when the finger penetrates the object and report the F-Score of this method. Alternatively, they could use a simple model with only one spring between the finger and the base of the object.

VIII.DISCUSSION

Page 12. Col 2. Line 3. Please, can you clarify the hand tuning? It seems that the genetic algorithm is not the best way to find the parameters? Have you tried to use other method like gradient descent? Is it possible to use the model to mass-spring model to predict the force exerted on the finger? In that case, the authors could use this prediction error as fitness function and use the gradient descent to find the parameters.

Other comments:

Fig. 1 is not referenced in the paper.
Page 2, Col 1, Line 17. Section VI is not mentioned.
Page 2. Col 2, Line 5: FEM is not defined. Finite-element models?
Fig 2. Instead of use symbols, I would write Vision, strain, force, etc.
Page 3, Col 1, line 57. Cretu et al. -> cite [10]?
Page 8. Col 2, Line 27. Linear snake = linear spline?
Algorithm 5. I would specify that the F-Score is the fitness function of the algorithm. Moreover, it would be nice to know the values used for the parameters of the algorithm (perhaps using a table).
Page 10. Col 1. Line 45. Please specify that there are two objects, a sponge and plasticine, with a rectangular shape and so on.