proceedings

MDPI

Proceedings

A 3D Visual Attention Model to Guide Tactile Data Acquisition for Object Recognition †

Ghazal Rouhafzay [1], Nicolas Pedneault [2] and Ana-Maria Cretu [1,*]

[1] Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada; ghazal.rouhafzay@carleton.ca

[2] Department of Computer Science and Engineering, Université du Québec en Outaouais, Gatineau, QC J8X 3X7, Canada; pedn01@uqo.ca

* Correspondence: anamaria.cretu@carleton.ca; Tel.: +1-613-520-2600

† Presented at the 4th International Electronic Conference on Sensors and Applications, 15–30 November 2017; Available online: https://sciforum.net/conference/ecsa-4.

Published: 14 November 2017

Abstract: Drawing inspiration from the human vision-touch interaction that demonstrates the ability of vision in assisting tactile manipulation tasks, this paper addresses the issue of 3D object recognition from tactile data whose acquisition is guided by visual information. An improved computational visual attention model is initially applied on images collected from multiple viewpoints over the surface of an object to identify regions that attract visual attention. Information about color, intensity, orientation, symmetry, curvature, contrast and entropy are advantageously combined for this purpose. Interest points are then extracted from these regions of interest using an innovative technique that takes into consideration the best viewpoint of the object. As the movement and positioning of the tactile sensor to probe the object surface at the identified interest points take generally a long time, the local data acquisition is first simulated to choose the most promising approach to interpret it. To enable the object recognition, the tactile images are analyzed with the help of various classifiers. A method based on similarity is employed to select the best candidate tactile images to train the classifier. Among the tested algorithms, the best performance is achieved using the k-nearest neighbor classifier both for simulated data (87.89% for 4 objects and 75.82% for 6 objects) and real tactile sensor data (72.25% for 4 objects and 70.47% for 6 objects).

Keywords: tactile sensing; 3D visual attention; interest points; data acquisition; object recognition

1. Introduction

Unlike the large number of sensors available and the various techniques for accurately interpreting visual and audio data, the sense of touch remains relatively less exploited for robotic applications in spite of the crucial role it can play in supporting sensing and environment understanding in many practical situations. In particular, human vision-touch interaction studies demonstrated the ability of visual information in assisting handling, grasping and dexterous manipulation tasks. In a similar manner, tactile sensing could support robot vision by compensating in situations where occlusions are present or when force estimates are required in various robotic manipulation tasks. Moreover, touching an object in order to recognize it can be inefficient, as probing requires direct contact which is lengthy and laborious to achieve in robot applications, where the sensor needs to be appropriately positioned and moved to collect quality data. This justifies the interest in guiding the tactile probing process such that the acquisition process becomes intelligent and efficient by collecting only relevant data. Drawing inspiration from the human vision-touch interaction, in this paper we evaluate the effect of human visual attention for detecting relevant areas over which tactile data can be collected in view of a subsequent recognition of the probed objects. In

particular, the main contribution of this work is to demonstrate that a series of interest points computed based on object features that attract visual attention allows for the acquisition of only relevant tactile data using a force-resistive (piezo-resistive) tactile sensor array over the surface of 3D objects in order to enable their recognition.

2. Literature Review

Tactile force sensor arrays are one of the most known and well established tactile sensors. Data recuperated by such sensors has been successfully employed for various tasks, including object recognition. Symbols in form of embossed numbers and letters are recognized in [1] using a feedforward neural network. A bag-of-feature technique is employed by [2] to classify industrial objects. The same technique is used in [3], on simulated data as returned by a tactile sensor array to estimate the probability of the object identity. Data from two tactile sensor array sensors mounted on a gripper performing a palpation procedure is classified as belonging to 10 objects by Drimus et al. [4], using a combination of a k-neighbors classifier and dynamic time warping. Liu et al. [5], classify empty and full bottles based on tactile data recuperated by tactile arrays placed on each of the three fingers of a robot hand making use of joint kernel sparse coding. To recognize a series of 18 objects in a fixed or movable position, Bhattacharjee et al. [6], extract various features from tactile sensor array data (i.e., maximum force over time, contact area over time and contact motion) that are subsequently classified by a k-nearest neighbor classifier. The authors of [7] exploit a different series of tactile features (i.e., positions and distances of the center of mass of the tactile image blob, pressure values, stochastic moments, the power spectrum and the raw, unprocessed windows centered at the points with the highest contact force) to identify objects using decision trees. The current paper aims at collecting tactile information using a force-resistive tactile sensor array only on relevant areas, represented by points of interest, over the surface of 3D objects in order to identify them.

3. Visual Attention Model for Guiding Tactile Data Acquisition for Object Recognition

The proposed solution for guiding data acquisition starts with a visual inspection of the object of interest. A 3D model of the object of interest can be obtained using an RGB-D sensor (Kinect), and a software that allows stitching data collected from multiple views in a unified 3D model (Skanect). Texture information, as recuperated by the color camera of the Kinect can then be added on the object surface to fully exploit the capabilities of the computational model of visual attention. The latter uses geometrical information (i.e., orientation of edges, curvature) and also color properties (i.e., color opponency, contrast) to identify the areas of interest that guide the deployment of attention. A novel, enhanced visual attention model and a method to extract interest points based on this model are proposed for this purpose. Taking inspiration from the human joint use of vision and tactile information for object manipulation tasks, we propose the use of these interest points for the acquisition of local tactile data (tactile imprints as collected by a force-resistive tactile sensor array). Because the acquisition of tactile data requires a direct contact with the object, the process to move and position the sensor can be extremely lengthy. We therefore use in an initial step a simulation that allows us to identify the best acquisition location and subsequently the best classifier to employ for recognizing the probed object based on the acquired tactile data. The method is then validated on real data collected using a force-sensitive tactile sensor array over a series of toy objects.

3.1. 3D Visual Attention Model for Interest Point Identification

In order to identify interest points over the surface of a 3D object, we have proposed an improved version of the classical visual attention system proposed by Itti [8]. The algorithm explores color, intensity, orientation, as the Itti model, but also capitalizes on the use of symmetry, curvature, DKL color space, contrast, and entropy to compute the level of saliency for different regions over the object model surface. All these features are known to guide the deployment of visual attention. As the visual attention model is only applicable on images, we used the virtual camera of Matlab to capture images from different viewpoints around the object. A saliency map is produced for each image based on the

previously reported features, in which the interest regions are represented by bright areas on a black background (the brighter the area is, the more salient it is). Based on this map, the interest points are determined in 2D pixel coordinates and projected back as 3D vertices on the object surface. Figure 1a summarizes the improved visual attention approach for interest point selection. In particular, 18 sets of viewpoints each containing 4 perpendicular viewpoints (to cover the whole surface of the object) are chosen to collect the images. For these, a series of conspicuity maps is created, eight for each image (one for each considered feature).

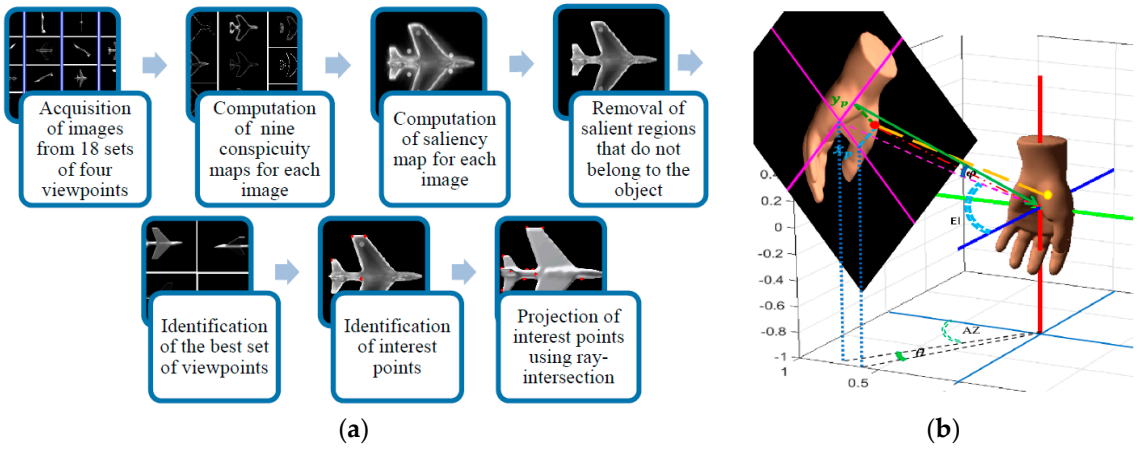

Figure 1. (**a**) Improved visual attention for interest point selection; (**b**) 2D to 3D projection.

The classical visual attention model [8] decomposes color, intensity and orientation modalities of captured images into a set of multiscale feature maps over which local spatial discontinuities are computed as center-surround differences. Further studies have proven that other features, including symmetry, curvature, DKL color space, contrast, entropy can also influence visual attention deployment. In this work, we have capitalized on all these features to compute an improved visual attention model. The bilateral and radial symmetric point detection algorithm introduced in [9] is used to determine the symmetry conspicuity map. Subsequently the center-surround operations are applied on the resulting map. Curvature information is extracted using the saliency map proposed in [10]. The color opposition model based on Derrington-Krauskopf-Lennie (DKL) color space [11] is added to provide another color feature, closer to human visual capacities. The local entropy calculation is based on a 9×9 neighbourhood of the median filtered input image and yields the entropy conspicuity map [12]. The luminance variance in a local neighborhood of 80×80 pixels as proposed in [13] is adopted to create the contrast conspicuity map. Finally, the eight conspicuity maps contribute with equal weights to produce a comprehensive saliency map for each image.

Once the saliency map is obtained for all images, the salient regions detected outside of the surface of the object are removed. Such regions can sometimes occur around the outer surface of the object due to the local contrast, intensity and color changes between the object and its background. The most salient set of viewpoints is then chosen as the set with the highest average level of saliency. The level of saliency is computed as the normalized accumulation of pixel values for the saliency maps. Interest points are then identified as the brightest points on the saliency maps for the best set of viewpoints. Figure 1b illustrates the projection geometry of an arbitrary interest point from image pixel coordinates (x_p and y_p) to the object surface. These values can be found in real world coordinate as $x = \frac{x_p}{PPWU}$ and $y = \frac{y_p}{PPWU}$, where $PPWU$ is the number of pixels per world units, calculated as $PPWU = \frac{Number\ of\ rows\ of\ the\ image}{2 \times d \times tan\frac{\alpha}{2}}$. In this formula, α is the camera view angle and d represents the distance from camera center to the origin. Knowing the spherical coordinate of the camera center (d, El, AZ), where El and AZ are the elevation and azimuth angles of camera center position respectively, the spherical coordinate of the interest point (the red dot in Figure 1b) can be calculated as $(d', El \pm \varphi, Az \pm \theta)$ where, $d' = \sqrt{x^2 + y^2 + d^2}$, $\phi = tan^{-1}\frac{y}{d}$ and $\theta = tan^{-1}\frac{x}{d}$. The

positive or negative signs of θ and φ depend on the quadrant on the image plane to which the interest point belongs to. The ray starting from the real world coordinate of the interest point on image plane in perpendicular direction to the image plane (in parallel with the vector from camera center to the origin) [14] intersects the corresponding point on object surface (yellow dot—Figure 1b).

The interest points on object surface are then sorted in descending order according to their level of saliency. The level of saliency is first determined as the number of times the point or its neighbors are identified as salient from different viewpoints and then by their corresponding pixel values in the saliency map. A series of 15 interest points with highest saliency are then selected for each object to guide the tactile data acquisition process.

3.2. Tactile Data Acquisition Simulation

Once the interest points are identified over the surface of an object, the data acquisition process is simulated over the detected points. In particular, the sensor surface is estimated as a series of quasi-tangent planes to the surface of the object at the identified interest point.

A series of four such planes are used (Figure 2a), situated close to each other in order to simulate the depth of the real sensor. The use of quasi-tangent planes is justified by the fact that research demonstrated that orienting the sensor along the local normal on the surface of the object maximizes the content and quality of acquired tactile data. To compute the first tangent plane (Figure 2b), we first identify the three closest neighbors (P_1 to P_3) to the identified interest point (P_i) over the object mesh and build a plane passing through them. The local intersection contour is then estimated between this plane and the 3D object surface. Three other equidistant planes are built along the local normal on the surface. The normal is computed as being the orthogonal vector on the first plane and it becomes the translation axis along which other three equidistant planes are placed. For each of them, the local intersection contours are identified with the 3D object. The contour information is then projected onto a plane where the gray level encodes the depth information. The procedure is repeated for each of the 15 interest points detected in Section 3.1. An example of tactile data acquired for the cup is shown in Figure 2c. Because the resulting tactile images are simulated, they have a higher resolution than the ones obtained with the real sensor. They have therefore been down-sampled to 16-by-16, to better correspond with the capabilities of the real sensor. The corresponding 16-by-16 values are concatenated to create vectors that become inputs to a classifier.

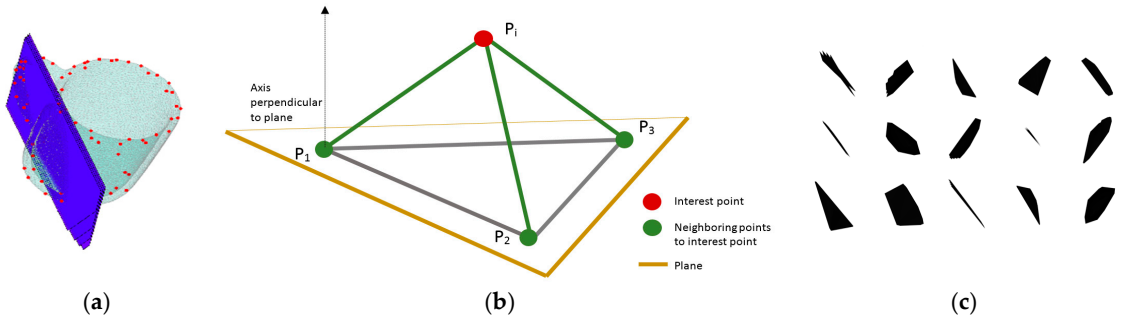

Figure 2. Tactile data acquisition simulation: (**a**) 3D object in contact with the 4 quasi-tangent planes (**b**) the first plane computation; and (**c**) series of tactile imprints obtained for the object.

3.3. Tactile Data Classification for Object Recognition

A series of classifiers, namely deep learning neural networks, Naïve Bayes, decision trees, evolutionary support vector machines (SVM) and k-nearest neighbors are then tested for the classification of objects based on the vectors encoding the tactile images. Because during experimentation we have realized that certain imprints are strongly resembling from one object to another (the ear of the dromedary and the ear of the cow), we proposed a selection of tactile imprints to be used for training the classifiers based on similarity. In particular, we have computed the similarity between the pairs of imprints of two objects at the time, as the normalized cross-correlation of the down-sampled 16-by-16 matrices. We have then eliminated all those imprints with a similarity

larger than 0.85 (equivalent to similarities larger than 85%). The results obtained over simulated data are then validated using real data collected over the interest points using a force-resistive tactile sensor array [15].

4. Experimental Results

We have performed experiments using the proposed framework for a series of four toy objects: cow, glasses, cup and hand; and then with a series of six object toys, namely cow, glasses, cup, hand, plane and dromedary, both for simulated and real data.

Table 1 displays the performance in the case in which simulated tactile imprints are used for the task of object recognition. The best recognition rates are achieved by the k-nearest neighbors classifier, followed by the evolutionary SVM classifier, with a maximum difference between them of 6.07%. The results are better when comparing the case when all the tactile imprints are used and when tactile imprints are selected for training based on similarity. An improvement of 15.95% is obtained using the selection process of imprints for the 4 objects and of 13.18% for the 6 objects. The best performance achieved is of 87.89% for 4 objects and of 75.82% for 6 objects.

Table 1. Object recognition rate with and without tactile imprints selection for simulated data.

	No Selection 4 Objects	**Selection 4 Objects**	**No Selection 6 Objects**	**Selection 6 Objects**
Deep learning	58.61%	79.95%	54.95%	56.13%
Naïve Bayes	31.81%	71.25%	40.66%	75.18%
Decision tree	43.61%	60.25%	35.16%	56.00%
Evolutionary SVM	71.53%	81.82%	61.54%	73.69%
k-nearest neighbors	71.94%	**87.89%**	62.64%	**75.82%**

The results for the real data collected over the same objects are displayed in Table 2.

Table 2. Object recognition rate with and without tactile imprints selection for real data.

	No Selection 4 Objects	**Selection 4 Objects**	**No Selection 6 Objects**	**Selection 6 Objects**
Deep learning	52.12%	59.14%	47.10%	55.13%
Naïve Bayes	50.24%	58.10%	27.54%	55.67%
Decision tree	53.17%	60.14%	22.46%	58.16%
Evolutionary SVM	61.54%	72.14%	57.25%	68.25%
k-nearest neighbors	67.23%	**72.25%**	57.97%	**70.47%**

One can notice that the best performance in terms of recognition rates is obtained using the k-nearest-neighbors algorithm, again followed in the second position by the evolutionary SVM, similar to the observation made for the simulated data. As well, it can be observed that the performance is better when using the tactile imprints selection based on similarity, both for four objects (72.25% versus 67.23%) and six objects (70.47% versus 57.97%). Due to the low results obtained both for simulated and real data, the Naïve Bayes and decision tree techniques are not appropriate for the task of object recognition based on tactile data. By comparing the best performance obtained on real and simulated results, one can notice that the first is slightly lower (max. by 15.64%). This is expected, because the real data is noisier and of very lower resolution.

5. Conclusions

In this paper we have studied the problem of object recognition based on tactile data whose acquisition is guided by an improved computational model of visual attention. The latter takes into consideration, beyond the classical features such as orientation, color and intensity, information about symmetry, curvature, contrast, and entropy to identify over the surface of a 3D object a series of interest points. It was demonstrated that tactile data collected at these points using a force-resistive

tactile sensor can be successfully employed to classify 3D objects using the k-nearest neighbors algorithm. When using the similarity measure to select the imprints for training the classifier, the recognition rate is of 87.89% for 4 objects and 75.82% for 6 objects for simulated data, while for real tactile sensor data is of 72.25% for 4 objects and 70.47% for 6 objects.

Acknowledgments: This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de recherche du Québec—Nature et Technologies (FRQNT) New University Researchers Start-up Program. The authors would like to thank Codrin Pasca and Emil M. Petriu for providing the sensor used for testing and Nicolas Bélanger for his help with the data acquisition over the real objects.

Author Contributions: G. Rouhafzay, N. Pedneault and A.-M. Cretu conceived and designed the experiments; G. Rouhafzay implemented and tested the 3D visual attention model; N. Pedneault performed the experiments for tactile object recognition; G. Rouhafzay, N. Pedneault and A.-M. Cretu analyzed the data; G. Rouhafzay and A.-M. Cretu wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cretu, A.; de Oliveira, T.E.A.; da Fonseca, V.P.; Tawbe, B.; Petriu, E.M.; Groza, V.Z. Computational intelligence and mechatronics solutions for robotic tactile object recognition. In Proceedings of the 2015 IEEE 9th International Symposium on Intelligent Signal Processing (WISP), Siena, Italy, 15–17 May 2015; pp. 1–6.
2. Schneider, A.; Strum, J.; Stachniss, C.; Reisert, M.; Burkhardt, H. Object identification with tactile sensors using bag-of-features. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; pp. 243–248.
3. Pezzementi, Z.; Plaku, E.; Reyda, C.; Hager, G.D. Tactile Object Recognition from Appearance Information. *IEEE Trans. Robot.* **2011**, *27*, 473–487.
4. Drimus, A.; Kootstra, G.; Bilberg, A.; Kragic, D. Design of a flexible tactile sensor for classification of rigid and deformable objects. *Robot. Auton. Syst.* **2014**, *62*, 3–15.
5. Liu, H.; Guo, V.; Sun, F. Object Recognition Using Tactile Measurements: Kernel Sparse Coding Methods. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 656–665.
6. Bhattacharjee, T.; Rehg, J.M.; Kemp, C.C. Haptic classification and recognition of objects using a tactile sensing forearm. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal, 7–12 October 2012; pp. 4090–4097.
7. Schopfer, M.; Pardowitz, M.; Haschke, R.; Ritter, H. Identifying Relevant Tactile Features for Object Identification. In *Towards Service Robots for Everyday Environments*; Prassler, E., Ed.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 417–430.
8. Itti, L.; Koch, C. Feature combination strategies for saliency-based visual attention systems. *J. Electron. Imaging* **2001**, *10*, 161–169.
9. Loy G.; Eklundh, J.O. Detecting Symmetry and Symmetric Constellations of Features. In *Computer Vision—ECCV 2006*; Lecture Notes in Computer Science; Leonardis, A., Bischof, H., Pinz, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 3952.
10. Loy, G.; Eklundh, J.-O. Detecting symmetry and symmetric constellations of features. In Proceedings of the IEEE European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2006; pp. 508–521.
11. Lee, C.H.; Varshney, A.; Jacobs, D.W. Mesh saliency. *ACM Siggraph* **2005**, *174*, 659–666.
12. Derrington, A.M.; Krauskopf, J.; Lennie, P. Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.* **1984**, *357*, 241–265.
13. Harel, J.; Koch, C.; Perona, P. Graph-based visual saliency. In Proceedings of the Twentieth Annual Conf. on Neural Information Processing Systems, Vancouver, BC, Canada, 4–7 December 2006; pp. 545–552.
14. Moller, T.; Trumbore, B. Fast, minimum storage ray/triangle intersection. *J. Graph. Tools* **1997**, *2*, 21–28.
15. Pasca, C. Smart Tactile Sensor. Master's Thesis, University of Ottawa, Ottawa, ON, Canada, 2004.