Large-scale integrating project

Deliverable I2.X
A review on object exploration through tactile and visual exploration

Project acronym:	DEXMART
Project full title:	DEXterous and autonomous dual-arm/hand robotic manipulation with sMART sensory-motor skills: A bridge from natural to artificial cognition
Grant agreement no:	FP7 216239
Project web site:	www.dexmart.eu

Due date: N/A	**Submission date: N/A**
Start date of project: February 1, 2008	**Duration: 48 months**
Lead beneficiary: SUN	**Revision: 1**

Nature: R	**Dissemination level: PU**
R = Report P = Prototype D = Demonstrator O = Other	PU = Public PP = Restricted to other programme participants (including the Commission Services) RE = Restricted to a group specified by the consortium (including the Commission Services) CO = Confidential, only for members of the consortium (including the Commission Services)

Executive Summary

In this work, a brief review about visual and tactile object exploration and recognition is presented. Until now, object recognition has been performed only through computer vision techniques. This is mainly due to technological problems related to the development of tactile sensors:

- tactile sensors are difficult to integrate beneath a compliant artificial skin
- a high integration level is needed
- wiring is a big problem not yet finally solved

The performance of artificial arrays of tactile sensors are absolutely far from that of the most simple animal. Even if most of works about this topic have been written in the 80's, in the last years the haptic exploration is slowly returning topical because new fabrication techniques have been developed, which allow to increase the integration level and the performance of the sensors. In almost all methods, haptic exploration is performed by a robotic hand equipped with tactile sensors that manipulates the object and measures a set of object features. A number of features have been proposed for tactile exploration based on geometrical properties, thermal properties, compliance, texture. A lot of methods used for robotics exploration can be applied also to human operator wearing a sensorized glove. As no system with satisfactory performance has been developed until now, the big challenge is to integrate tactile data with direct joint angular and vision data. Therefore, this document contains a review of existing methodologies for both tactile and visual object exploration.

TABLE OF CONTENTS

1 Introduction

In nature, the sense of touch assumes a fundamental importance: both the most simple and the most complex creatures are provided with a large number of mechanoreceptors which furnish tactile information. The humans use touch for manipulation, exploration and response [61]. Even fine manipulation tasks like buttoning a shirt would became very difficult without the use of touch. From object tactile exploration we can elicit a great number of material and surface characteristics, e.g. hardness, thermal conductivity, friction, roughness, that allow to recognize and classify the objects. In some cases, like distinguishing a real from a synthetic leather, the touch is much more effective than the vision. The patients with peripheral neuropathy, that reduces the response capability, accidentally do damage to themselves, because they do not distinguish a gentle touch from an impact. Even the simplest animals have thousands of mechanoreceptors per square centimeter of skin, but at the state of the art no artificial tactile sensor exist which can reach this integration level. Because of the absence of high integrated CCD and CMOS array of tactile sensors, in the last years the researchers were being more interested in computer vision techniques for performing object exploration and recognition owing to the wide range of available cameras. The tactile sensors work through physical interaction with the object to explore and have to be incorporated into skin surfaces with compliance, to conform locally to surfaces, and with adequate friction to handle objects securely. The sensors and skin must also be robust enough to survive to repeated impacts and abrasions. Nevertheless, in the last 20 years, advanced methods of fabrication have been developed, which allow to increase the integration, to decrease the number of wires, and to introduce local processing of tactile information (for more details see [35] and [34]).

One of the classical problems in computer vision is that of determining whether or not the image data contain some specific object, feature, or activity. This task can be solved robustly and without effort by a human, but it is still not solved in computer vision for the general case in which there is an arbitrary object in an arbitrary situation. The existing methods for dealing with this problem can at best solve it only for specific objects, such as simple geometric objects, human faces, printed or hand-written characters, vehicles, and in specific situations, typically described in terms of well-defined illumination, background, and pose of the object relative to the camera. When speaking about 3D object recognition one may refer to a recognition process based on 3D geometric model of the object or consider an appearance-based approach. In the remainder of this dissertation we will consider only those methods based on the visual appearance of the object. It is worth to notice that there is a difference with those view-based recognition methods that represent objects only from one particular point of view [58, 38]. Indeed, an effective 3D object recognition is based on the use of many views of the object to obtain a model which is capable to describe it from different viewpoints [38, 12, 21, 18].

The document is organised as follows. In the next Section the state of the art for object recognition by tactile information is reviewed; object recognition by computer vision techniques is reviewed is Section 3; Section 4 is left to final remarks.

2 Object Recognition by Tactile Information

In object recognition the goal is to identify one object out of a set of known objects using information derived from touch [61].

The key factor for haptic object recognition applications is the definition and detection of surface features; the surface features are important because allow to elicit and synthesize the information obtained exploring an object through tactile sensors.

The largest part of feature definition has been made by the computer vision research community, but the feature definition which are good for the visual exploration, are often not appropriate for haptic exploration. Recognition is usually carried out through geometrical features elicited through force or tactile array sensors; in [1], [64], [31] also the use of other type of information is proposed, e.g. compliance, texture, thermal properties. Many different features derived from tactile array data have been used for model-based recognition and classification. The most used geometric feature for model-based recognition are holes, edges [44], and corners [25] and object surfaces [51]. Other feature sets include geometric moments [40], [39], linear transforms [29], and sequences of surface tangents [52]. Gaston and Lozano-Perez [19] use local surface normals and contact locations as features, which could be derived from array, force, or joint sensor information.

A number of systems use statistical pattern recognition, which can improve noise immunity since only statistics derived from the sensed data are used. In the statistical approach, each pattern (object to recognize) is represented in terms of d features or measurements and is viewed as a point in a d-dimensional space. The goal is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a d-dimensional feature space. The effectiveness of the representation space (feature set) is determined by how well patterns from different classes can be separated. Given a set of training patterns from each class, the objective is to establish decision boundaries in the feature space which separate patterns belonging to different classes. In the statistical decision theoretic approach, the decision boundaries are determined by the probability distributions of the patterns belonging to each class, which must either be specified, see [27]. Unfortunately, this also means that only a few object types can be discriminated.

Kinoshita [30] and Briot [7] have proposed methods based on the statistics of tactile array sensor data, while Okada, Briot and Marik [47], [8], [41] proposed methods based on the statistics of finger joint angles while grasping the object.

A great problem in haptic exploration is that some features cannot be identified through static touch, but motion and object manipulation is needed [49]. Okamura and Cutkosky, in [48], present definition of the features based on local surface curvature; the definitions depend on both the geometry of the "viewer" (the device used for the exploration) and the object.

Several developed strategies for scheduling sensor movements are so that each additional observation decreases the number of objects which are consistent with prior observations. This is sometimes described as a hypothesize and test approach [24].

In the 80's, many algorithms have been developed to schedule the sensor movements for polygonal object recognition by contact location and surface normal measurement, for more details see the works of Schneiter [60], Grimson and Lozano-Perez [23], and Ellis [15]. Schneiter suggests that each sensing move should cross the boundary of the intersection of all objects consistent with previous sensor observations. Cole [10] shows that to determine effectively the shape of a convex planar polygon with V vertices, at least $3V - 1$ measurements are needed. Roberts [56] proposes a movement strategy for recognition that includes tracing the finger along object surfaces and edges, rather than moving the finger through free space between readings. In non-model-based approaches, Klatzky et al. [31],[32] have suggested that the same exploratory procedures used by humans in haptic exploration can be performed by robotics systems. These procedures prescribe the finger motions needed for tasks such as tracing object contours, measuring compliance, and determining the lateral extent of object surfaces. Recently, in [45] a method for shape classification in rotation manipulation has been proposed; the viewer is a universal multi-fingered robotic hand equipped with tip pressure tactile sensors, which measure the tip pressure distributions. After the rotation manipulation a degree of similarity is calculated between a kurtosis pattern calculated from the pressure distribution and a reference pattern; the experiment shows that this approach can classify three simple objects (hexagon, octagon, and circle) with a high degree of accuracy. In [66], an anthropomorphic skin developed by Takamaru and his group has been presented. It is composed of multiple layers with sensing devices detecting strain and its velocity and methods based on Self-Organizing Maps (SOM) have been proposed to perform the object recognition from a set of repetitive grasps.

3 Object Recognition by Vision

The first approaches to 3D object recognition have been tackled from the geometric point of view: the information used to characterise an object is organised in the form of a 3D model focused on geometric measurements. This approach is usually called model-based recognition. Several model-based approaches to recognition have been studied and details on possible classifications can be found in [5, 26, 22] in which the authors presented their taxonomy for model-based object recognition. A more specific work that addresses 3D object construction from image sequence, is reported in [54]. A first way to distinguish among the different approaches is dividing the recognition method on the basis whether the pose is inferred from global properties or from mappings of local model features. Usually such models are sophisticated, difficult to build, and often hard to use. Aspect graphs [33, 5], instead, are one of the first attempts to represent the 3D appearance of objects in terms of their 2D projections. With aspect graphs, though, there is no computational gain relative to computing the full 3D model of the object. Nevertheless the idea of representing objects using 2D rather than 3D models has been supported by recent advances in biological vision [9, 62], and has been adopted by various researchers.

Among view-based techniques, it is worth mentioning 2D morphable models [53, 6, 70, 28] and view-based active appearance models [14, 11], which use a selection of 2D views for the purpose of modelling a 3D complex object. The idea behind both methods (which achieve comparable results in two rather different ways) is that of synthesising a model for each object or class of objects, and then matching it with a novel image to check for consistency.

It is important to notice that techniques based on the appearance follow a considerable body of previous research in model-based vision. In fact also appearance-based methods can be divided in two categories, global or local, in case they are building a representation of the object by integrating information over the entire image [65, 43, 13, 55] or over a set of local interest points respectively [38, 42, 57, 69].

Global methods build an object representation by integrating information over the whole image and therefore they are very sensitive to background clutter and partial occlusions. Hence, global methods only consider test images without background, or necessitate a prior segmentation, a task which has proven extremely difficult or else they rely on the fact that context is as important to recognition as the object is [2, 67]. Additionally, robustness to large viewpoint changes is hard to achieve, because the global object appearance varies in complex and unpredictable ways.

Local methods counter problems due to clutter and occlusions by representing images as a collection of features extracted on local information only, thus now we focus on these approaches.

Schiele and Crowley, in [59], proposed one approach to 3D object recognition based on a framework for the statistical representation of the appearance of arbitrary 3D objects. The method proposed combines global techniques as those used in [65, 3] with object recognition methods based on local characteristics (similar to point matching tasks) [22]. The set of objects used for testing the method proposed in [59] is quite large (over 100 objects). The results reported demonstrate that it is possible to recognise objects in cluttered scenes using local approaches, but it is worth to notice that most of the objects used in the test have a planar structure, thus they cannot be considered as real 3D objects.

Lowe describes a method for 3D object recognition which is based on the use of SIFT descriptors.

After the extraction and the description of local features of the image, they perform best-bin-first search, which is a modification of the k d tree algorithm, to identify the nearest neighbours. In [38] a clustering step is added and a probability model is used to achieve more robustness also in cases of little non-rigid deformations. The performances for object recognition are increased since the model can be generalised across multiple viewpoints by linking features between different views. It is also possible to combine features obtained under multiple imaging conditions into a single model view. The object are recognised even when there are few matches between the test image and the model.

In [4], Belongie, Malik and Puzicha introduced a new shape descriptor, the shape context, for measuring shape similarity and recovering point correspondence with the aim of recognising some particular categories of objects (silhouette, handwritten digits).

There are techniques, based on the use of models, that look for relationships between the object and its projection to the image and usually these methods recognise objects by visual similarity without attempting any high level analysis of the image. For instance, Obdrzalek and Matas [46] describe an approach which aims at combining good qualities of model-based and view-based object recognition techniques: their method is based on the extraction of visual local features of the object and on the construction of a model built upon these features. This approach assumes that the deformation of the image should be at most affine. The features used in [46] are the MSER (Maximally Stable Extremal Region). Several affine-invariant construction of local affine frames are used for determining correspondence among local image patches.

The main problem with local methods is that, while observing minute details, the overall appearance of the object may be lost. Also, small details are more likely to be common to many different objects (this feature has actually been exploited in [68] to design efficient multi-class systems). For this reason local information is usually summarised in global descriptions of the object, for instance in codebooks [12, 37]. Alternatively, closeness constraints can be used to increase the quality of matches [17].

In the real world, there exists a strong relationship between the environment and the objects that can be found in it: for instance in a street we are looking for cars and pedestrians while in a room we look for chairs but not for trees. Human visual system uses these relationships for facilitating object detection and recognition [50, 67]: thus we can say that adding geometric and global information and using the context can improve object detection and recognition performances.

In [16] the authors present a parts and structure model for object categorisation. The model is learnt from example images containing category instances, without requiring segmentation from background. The model obtained is a sparse representation of the object and consists of a star topology configuration of parts modelling the output of a variety of feature detectors. This star model has the good qualities of being translational and scale invariant and it is used both in learning and in recognition steps.

Another approach to object categorisation based on the use of a model is introduced by [36]. The authors present a technique based on the use of an Implicit Shape Model, computed for each class of object, based on a codebook or vocabulary of local appearance that are prototypical for the object category. The model uses also a spatial distribution which specifies where each codebook entry may be found on the object. In other words, this approach is reminiscent of the visual vocabulary, but it has the add that it keep also spatial information.

In [57] the authors associate geometric constraints to different views of the same patches under affine projection. This lead to a true 3D affine and Euclidean model from multiple unregistered images without the need of any segmentation. This paper is based on the idea that the surface of a solid can be represented by a collection of small patches, their geometric and photometric invariants and a description of their 3D spatial relationships. It is interesting to notice that for the appearance-based selection of potential matches [57] exploits also the use of colour descriptor in association with SIFT:

it is shown that when two very similar patches have different colours their SIFT descriptors appear almost identical.

The method proposed in [18] is based on a matching procedure that no longer relies solely on local keypoints. The first step is to produce an initial large set of unreliable region correspondences. The aim, here, is that of maximising the number of correct matches, at the cost of introducing many mismatches. Then the method explores the surrounding image areas, recursively constructing more and more matching regions. This method succeeds in covering also image areas which are not interesting for feature extractor thus deciding the object identity is based on information densely distributed over the entire portion of the object visible in the test image. To integrate the information coming from multiple views of the object, the authors introduce the concept of group of aggregated matches or GAM, the use of which increases the discriminative power of the technique, but, as a drawback, the method proposed in [18] is computational expensive and the good results obtained for textured areas are not possible for uniform objects, for which is better to combine the extraction of region with the detection of edges.

In their work [20, 21], Grabner and Bischof propose an approach for extracting discriminative descriptions of 3D objects using spatio-temporal information. Similarly to [63] they extract local features that are tracked in image sequences leading to local trajectories containing dynamic information about the evolution of the descriptors. These trajectories are evaluated with respect to their quality and robustness and finally each of them is assigned a single local descriptor from a key-frame in order to obtain an object description. Their descriptors are based on the assumption that local features are always distinctive patches of an object and that this object is made of piecewise planar patches.

4 Conclusions

Object recognition by tactile information has not yet reached the diffusion of the recognition techniques exploiting vision, mainly due to the technological problems indicated above (placement and robustness of sensors, high number of wires, high integration level required for array sensors). Most of the works in the literature of the last years are about the tactile exploration performed by robotics hand equipped with tactile sensors or sensorized artificial skin. Nevertheless, these methods and manipulation strategies can often be applied also to human operators wearing a glove with a set of tactile sensors. In any case, the object recognition remains a big challenge when fusion of data from different sources (tactile sensors, joint angular sensors, and visual data) is adopted for obtaining a reliable recognition of a variety of complex objects.

Object recognition by visual information is a mature field, although the problem is still far from being solved, in the sense of having a general solution. In fact the approaches proposed so far still suffer from large changes in imaging conditions, as illumination, scale and pose. However these limitations can be easily overcame restricting to controlled experimental conditions as, for instance, an indoor environment, and with a not large number of classes of objects, making the most of the techniques usable in a number of applications.

References

[1] R. Bajcsy, *What can we learn from one finger experiments?*, Robotics Research: The First International Symposium (Cambridge) (M. Brady and R.P. Paul, eds.), MIT Press, 1984, pp. 509–528.

[2] A. Barla, F. Odone, and A. Verri, *Hausdorff kernel for 3D object acquisition and detection*, Proceedings of the European Conference on Computer Vision, 2002, pp. 20–33.

[3] S. Belongie, C. Carson, H. Greenspan, and J. Malik, *Color and texture based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval*, Proceedings of the International Conference on Computer Vision, 1998, pp. 675–682.

[4] S. Belongie, J. Malik, and J. Puzicha, *Shape matching and object recognition using shape contexts*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), no. 24.

[5] P. J. Besl and R. C. Jain, *Three-dimensional object recognition*, Computing Surveys **17** (1985), no. 1, 75–145.

[6] D. Beymer, A. Shashua, and T. Poggio, *Example based image analysis*, A.I. Memo 1431, MIT, 1993.

[7] M. Briot, *The utilization of an artificial skin sensor for the identification of solid objects*, Proc. of The 9th Int. Symp. Industrial Robots, 1979, pp. 529–547.

[8] M. Briot, M. Renaud, and Z. Z. Stojilkovic, *An approach to spatial pattern recognition of solid objects*, IEEE Trans. Syst. Man Cybernet. **8** (1978), no. 9, 690–694.

[9] H. H. Bulthoff, S. Y. Edelman, and M. J. Tarr, *How are three-dimensional objects represented in the brain?*, Cerebral Cortex **5** (1995), no. 3.

[10] R. Cole and C. Yap, *Shape from probing*, J. Algorithms **8** (1987), no. 1, 19–38.

[11] T. F. Cootes, K. N. Walker, and C. J. Taylor, *View-based active appearance models*, Proceedings of the International Conference on Automatic Face and Gesture Recognition, 2000, pp. 227–232.

[12] G. Csurka, C.R. Dance, L. Fan, and C. Bray, *Visual categorization with bag of keypoints*, Proceedings of the European Conference on Computer Vision, 2004.

[13] C. Cyr and B. Kimia, *3D object recognition using similarity-based aspect graph*, Proceedings of the International Conference on Computer Vision, 2001, p. 254261.

[14] G. Edwards, T. F. Cootes, and C. J. Taylor, *Advances in active appearance models*, Proceedings of the International Conference on Computer Vision, 1999.

[15] R.E. Ellis, *Acquiring tactile data for the recognition of planar objects*, Proc. of 1986 IEEE Int. Conf. on Robotics and Automation (San Francisco, CA), 1986, pp. 1249–1253.

[16] R. Fergus, P. Perona, and A. Zisserman, *A sparse object category model for efficient learning and exhaustive recognition*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[17] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, *Groups of adjacent contour segments for object detection*, Technical Report 5980, INRIA, 2006.

[18] V. Ferrari, T. Tuytelaars, and L. Van Gool, *Simultaneous object recognition and segmentation from single or multiple model views*, International Journal of Computer Vision **67** (2006), no. 2.

[19] P.A. Gaston and T. Lozano-Perez, *Tactile recognition and localization using object models: The case of polyhedra on a plane*, IEEE Trans. Pattern Anal. Mach. Intell. (1984), 257–266.

[20] M. Grabner, *Object recognition with local feature trajectories*, Master's thesis, Technische Unversitat of Graz, 2004.

[21] M. Grabner and H. Bischof, *Extracting object representations from local feature trajectories*, Proceedings of the I Cogn. Vision Workshop, 2005.

[22] W.E.L. Grimson, *Object recognition by computer: the role of geometric constraints*, The MIT Press, 1990.

[23] W.E.L. Grimson and T. Lozano-Perez, *Model-based recognition and localization from sparse range or tactile*, The Int. J. Robot. Res. **3** (1984), no. 3, 3–35.

[24] V.S. Gurfinkel, *Tactile sensitizing of manipulators*, Eng. Cybernet. **12** (1974), no. 6, 47–56.

[25] W.D. Hillis, *Active touch sensing*, The Int. J. Robot. Res. **1** (1982), no. 2, 33–44.

[26] D.P. Huttenlocher, *Three-dimensional recognition of solid object from a two dimensional image*, Ai lab memo, M.I.T., 1989.

[27] A.K. Jain, R.P.W. Duin, and J. Mao, *Statistical pattern recognition:a review*, IEEE Trans. on Pattern Analysis and Machine Intelligence **22** (2000), no. 1, 4–37.

[28] M. J. Jones and T. Poggio, *Multidimensional morphable models: a framework for representing and matching object classes*, International Journal of Computer Vision **2** (1998), no. 29, 107–131.

[29] J. Jurczyk and K.A. Loparo, *Mathematical transforms and correlation techniques for objectrecognition using tactile data*, IEEE Tran. on Rob. and Aut. **5** (1989), no. 3, 359–362.

[30] G.I Kinoshita, *Classification of grasped objects shape by an artificial hand with multi-element tactile sensors*, Information Control Problems in Manufacturing Technology (Oxford) (Y. Oshima, ed.), Pergamon, 1977, pp. 111–118.

[31] R.L. Klatzky, R. Bajcsy, and S.J. Lederman, *Object exploration in one and two fingered robots*, Proc. of 1987 IEEE Int. Conf. on Robotics and Automation (Raleigh, NC), 1987, pp. 1806–1809.

[32] R.L. Klatzky and S.J. Lederman, *Intelligent exploration by the human hand*, Dextrous Robot Hands (Berlin) (S.T. Venkataraman and T. Iberall, eds.), Springer, 1990.

[33] J. J. Koenderink and A. J. van Doorn, *The internal representation of solid shape with respect to vision*, Biol. Cyber **24** (1979), 211–216.

[34] M.H. Lee, *Tactile sensing: new directions, new challenges*, The Int. J. Robot. Res. **19** (2000), no. 7, 636–643.

[35] M.H. Lee and H.R. Nicholls, *Tactile sensing for mechatronics a state of the art survey*, Mechatronics **9** (1999), no. 1, 1–31.

[36] B. Leibe, A. Leonardis, and B. Schiele, *Combined object categorization and segmentation with an implicit shape model*, Proceedings of the European Conference on Computer Vision, 2004.

[37] B. Leibe, K. Mikolajczyk, and B. Schiele, *Efficient clustering and matching for object class recognition*, Proceedings of the British Machine Vision Conference, 2006.

[38] David G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision **60** (2004), no. 2, 91–110.

[39] R.C. Luo and H.H. Loh, *Recognition of similar objects using a tactile image array sensor*, Proceedings of the Intelligent Robots and Computer Vision Conference: Sixth in a Series (Cambrige) (D. Casasent and E. Hall, eds.), 1987, pp. 1249–1253.

[40] R.C. Luo and W.-H. Tsai, *Object recognition using tactile image array sensors*, Proc. of 1986 IEEE Int. Conf. on Robotics and Automation (San Francisco, CA), 1986, pp. 1249–1253.

[41] V. Marik, *Algorithms of complex tactile information processing*, Proc. of The 7th Int. Joint Conf. Artif. Intell, 1981.

[42] K. Mikolajczyk and C. Schmid, *An affine invariant interest point detector*, Proceedings of the European Conference on Computer Vision, vol. 1, 2002, pp. 128–142.

[43] H. Murase and S. Nayar, *Visual learning and recognition of 3d objects from appearance*, International Journal of Computer Vision **14** (1995), no. 1.

[44] C. Muthukrishnan, D. Smith, D. Myers, J. Rebman, and A. Koivo, *Edge detection in tactile images*, Proc. of 1987 IEEE Int. Conf. on Robotics and Automation (Raleigh, NC), 1987, pp. 1500–1505.

[45] H. Nakamoto, F. Kobayashi, N. Imamura, H. Shirasawa, and F. Kojima, *Shape classification in rotation manipulation by universal robot hand*, Proc. of 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (Nice, F), 2008, pp. 53–58.

[46] S. Obdrzalek and J. Matas, *Object recognition using local affine frames on distinguished regions*, Proceedings of the British Machine Vision Conference, 2002.

[47] T. Okada and S. Tsuchiya, *Object recognition by grasping*, Pattern Recognition **9** (1977), no. 3, 111–119.

[48] A.M. Okamura and M.R. Cutkosky, *Feature detection for haptic exploration with robotics fingers*, The Int. J. Robot. Res. **20** (2001), no. 12, 925–938.

[49] A.M. Okamura, M.L. Turner, and M.R. Cutkosky, *Haptic exploration of objects with rolling and sliding*, Proc. of the 1997 IEEE Int. Conf. on Robotics and Automation (Albuquerque, NM), 1997, pp. 2485–2490.

[50] A. Oliva, A. Torralba, M. S.Castelhano, and J. M.Henderson, *Top-down control of visual attention in object detection*, Proceedings of the IEEE Conference on International Conference on Image Processing, 2003.

[51] K.J. Overton and T. Williams, *Tactile sensation for robots*, Proc. of the Int. Joint Conf. Artif. Intell., 1981, pp. 791–795.

[52] H. Ozaki, S. Waku, A. Mohri, and M. Takata, *Pattern recognition of a grasped object by unit-vector distribution*, IEEE Trans. Syst. Man Cybernet. **12** (1982), 315–324.

[53] T. Poggio and T. Vetter, *Recognition and structure from one 2d model view: observations on prototypes*, A.I. Memo 1347, MIT, 1992.

[54] M. Pollefeys (ed.), *Proceedings of 3-d modeling from images*, 2000.

[55] M. Pontil and A. Verri, *Support vector machines for 3D object recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998), 637–646.

[56] K. Roberts, *Robot active touch exploration: Constraints and strategies*, Proc. of 1990 IEEE Int. Conf. on Robotics and Automation (Cincinnati, OH), 1990, pp. 1679–1684.

[57] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, *3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraint*, International Journal of Computer Vision (2006).

[58] B. Schiele and J.L. Crowley, *Object recognition using multidimensional receptive field histograms and its robustness to view point changes*, Proceedings of the European Conference on Computer Vision, 1996.

[59] ______, *Recognition without correspondence using multidimensional receptive field histograms*, International Journal of Computer Vision (2000).

[60] J.L. Schneiter, *An objective sensing strategy for object recognition and localization*, Proc. of 1986 IEEE Int. Conf. on Robotics and Automation (San Francisco, CA), 1986, pp. 1262–1267.

[61] B. Siciliano and O. Kathib (Eds.), *Springer handbook of robotics*, Springer, Heidelberg, 2008.

[62] P. Sinha, *Perceiving and recognizing 3D forms*, Ph.D. thesis, Massachussetts Institute of Technology, 1995.

[63] J. Sivic and A. Zisserman, *Video google: A text retrieval approach to object matching in videos*, Proceedings of the International Conference on Computer Vision, 2003.

[64] S.A. Stansfield, *A robotic perceptual system utilizing passive vision and active touch*, The Int. J. Robot. Res. **7** (1988), no. 6, 138–161.

[65] M. J. Swain and D. H. Ballard, *Color indexing*, International Journal of Computer Vision (1991).

[66] S. Takamuku, A. Fukuda, and K. Hosoda, *Repetitive grasping with anthropomorphic skin-covered hand enables robust haptic recognition*, Proc. of the 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (Nice, F), 2008, pp. 3212–3217.

[67] A. Torralba, *Contextual priming for object detection*, international Journal of Computer Vision (2003).

[68] A. Torralba, K. Murphy, and W. Freeman, *Sharing visual features for multiclass and multiview object detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **29** (2007), no. 5, 854–89.

[69] T. Tuytelaars and L. Van Gool, *Wide baseline stereo matching based on local, affinely invariant regions*, Proceedings of the British Machine Vision Conference, 2000, p. 412425.

[70] T. Vetter and T. Poggio, *Linear object classes and image synthesis from a single example image*, IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997), no. 7, 733–742.