

Function-Based Classification from 3D Data and Audio

Aliza Amsellem

Faculty of Micro Engineering
Swiss Federal Institute of Technology
Lausanne, Switzerland, 1015
Email: aliza.amsellem@epfl.ch

Octavian Soldea

and Ehud Rivlin
The Technion, Israel Institute of Technology
Haifa, 32000, Israel
Email: {octavian,ehudr}@cs.technion.ac.il

Abstract—We propose a novel scheme for fusion between two types of modalities to support function-based classification. While the first modality targets functional classification from sounds registered at impact, the second one aims classification of objects in 3D images. Using audio one can answer functional questions such as what is the material the analyzed objects are built of, if the objects are full or hollow, if they are heavy, and if they are rigidly linked to their supports. Audio based signatures are used to label parts of the object under analysis. Different parts of any object can be partitioned in generic multi-level hierarchical descriptions of functional components. Functionality, in the visual modality reasoning scheme, is derived from a large set of geometric attributes and relationships between object parts. These geometric properties represent labeling signatures to the primitive and functional parts of the analyzed classes. The fusion between both of the modalities relies on a shared cooperation among audio and visual signatures of the functional and primitive parts. The scheme does not require a-priori knowledge about any class. We tested the proposed scheme on a database of about one thousand different 3D objects. The results show high accuracy in classification.

I. INTRODUCTION

The outside world emits plenty of signals that human being must filtrate in order to extract information. Human recognition is based on multiple sensors fusion: hearing, seeing, smelling, touching etc. (see [30]). Following this idea, we designed a multi-modal approach to classify environments in a functional based framework. We employ a new fusion scheme for audio and visual modalities in a functional based framework. An audio sensor allows to recognize functionalities otherwise unrecognizable by visual sensor such as: material, emptiness, etc.

In most works, (see [1], [15], [16], [29]), analyzes of the sound properties of an object are made through a single collision of the studied object. In general, audio systems register produced sounds, create databases of their signatures, and include some testing schemes.

The audio signatures that have been most often used in databases are: the Fourier transform ([6], [16]), the spectrogram ([17]), and the decay rate ([1], [15], [16], [17], [29]). The next step after creating the database is choosing a classifier [18], [28]. According to [6], a minimum distance classifier is efficient to recognize material by colliding sound.

In [1], the authors have tested their system by trying

to differentiate four different materials: rubber, wood, glass, and steel. They show that decay rate is a convenient tool to classify these four types of material. [9] tests an audio system to recognize plastic, wood, glass, and steel. The author separates the four tested materials into two macro-categories: wood/plastic and glass/steel.

Concerning vision, the first systems using function-based classification were [5] and [32]. An impressive number of results in the function-based classification field were demonstrated with the GRUFF and OMLET systems [25], [26], [33].

Recalling [30], fusion of different sensors is at high interest in environmental classification. In [27], the authors employ fusion in order to proceed to audiovisual speech recognition. Audio and video fusion is also used in tracking [3]. [21] proposes to combine computer hardware (mouse and keyboard) in order to model human activities. Interestingly enough, applications of fusion between audio and video modalities are used in lipreading [8], [19].

We propose a novel scheme for a functional based classifier employing fusion of audio and visual modalities. While the first modality targets functional properties recognizable from audio information, the second one aims classification of objects captured in 3D images (see [23]).

The audio modality is employed to label parts of the analyzed objects with the audio signatures of sounds resulting from the impact between primitive parts and the end effector of a robotic arm. Although extensively used techniques such as FFT, spectrogram, and decay rate are employed in computing the signatures, it is the first time functionalities are inferred from sounds. Following and generalizing [23], where a visual functional based classification scheme is proposed, our classification process calls for constructing a generic multi-level hierarchical description of object classes in terms of functional components. The construction of the generic multi-level hierarchy can be thought of as a learning phase. Our scheme is able to automatically build the description of any new object class from labeled examples.

During classification, a search through a finite graph using a probabilistic matching measure is performed to find the best assignment of parts of an object to the functional structures of each class. An object is assigned to a class providing the highest matching value.

We tested the proposed scheme on a database of about one thousand different 3D objects. In so far as we know, this is the first scheme that performs function-based classification that involves a learning phase and does not require a-priori knowledge about any class. Nevertheless, it is the first scheme that performs functional based classification fusing audio and video information. Bearing in mind that today's range capturing devices are largely available at relative low prices, the use of these sensors as well as their fusion with other modalities provides the basis for a large range of classification applications.

II. FUNCTIONAL BASED CLASSIFICATION USING AUDIO

We focused on two main aspects: material and emptiness recognition. We target classification of ceramic, wood, plastic, empty and full glass sounds registered when they are struck. We also studied the stability of an object on its support, weight, and hardness (see [9]).

A. Building a Database of Sounds

The database comprises the sounds generated when striking objects built from ceramic, wood, empty glass, full glass, plastic which is rigidly-linked to its support, not rigidly-linked plastic, full plastic, hollow ceramic, and more. For each one of the first five classes we registered eighty sounds.

In order to create a database we used a robot arm programmed to strike each analyzed material. When registering sounds, a constant force was used and constant delays were detected from the beginning of the registration frame and up to the entrance of the signal.

B. Audio Based Classifier

We have implemented three signatures: FFT, spectrogram, and decay rate. These signatures are the most common used one (see Section I). At classification, a grade of matching is established. The object is associated with the class with which the highest level of matching is detected.

The first signature we have used is FFT [6]. In the period domain, the Fourier transform of the signal presents peaks at defined positions. Different classes of sounds (glass, ceramic, etc.) have different peaks in the Fourier domain. The location of these peaks represent the first signature. The spectrogram [17] also outlines a maximum that is characteristic of functions. This signature varies the less over the different samples that were registered for the database. Each registered sound amplitude decreases as a function of a constant, which is the angle of internal friction. The internal friction parameter depends of the material and causes higher frequencies to decay more rapidly. Studying the decay rate [15] of the signal gives us information about its functions.

III. FUNCTION BASED FUSING OF AUDIO AND VISUAL CLASSIFIERS

In [23], we proposed a functional based classification scheme on visual data. Here, we show a generalization of this classifier towards fusion with audio data. The proposed scheme

consists of two phases: a learning phase and a classification phase. Each of these phases receives as input segmented images (see [12]). We do not elaborate on the details of the implemented segmentation technique due to space limitations.

The objects are segmented into constituents: primitive and functional parts, together with several visual signatures. The primitive parts (also known in the literature as geons [2], [4]) that we consider are sticks, plates, and blobs. A functional part is defined as an object part that could provide a certain function, and usually comprises several primitive parts. In addition, each part is labeled with audio signatures as computed by the audio classifiers.

In the learning phase, several instances (objects) of a class are input. The learning phase computes the values of the visual signatures of the constituents, the relationships between them, as well as the values of the audio signatures.

A. Multi-Level Hierarchy Functional Structure

The classification process comprises an analysis both of the detected primitive parts and the relationships that exist among them. We call the relationships between primitive parts primitive-to-primitive connections. Each primitive part or group of primitive parts and the primitive-to-primitive connections among them that can fulfil a certain functional task are classified as a functional part. Further, several functional parts and the relationships among them can define a functional task and can form a higher level functional part. The proposed hierarchy can be as complex as one wishes. This approach is known in the literature as recognition/classification by functional parts. A relationship between a pair of functional parts is called a functional-to-functional connection. Each level in the functional hierarchy has a clique structure and each pair of functional parts (in the clique) are characterized by a relationship expressed in terms of visual or audio signatures. For example, in Fig. 1, each pair of functional parts "Back Support", "Sittable", and "Ground Support" are connected, thus forming a clique. Note that these three nodes have the common ancestor "Chair".

For any functional part f , define $P(f)$ and $F(f)$ the set of immediate primitive or functional constituents of f and $C(f)$ the set of connections between the elements of $P(f) \cup F(f)$; see Fig. 1. Note that only one of $P(f)$ or $F(f)$ is not empty for any functionality f .

For any symbolic primitive part, functional part, or connection a , we associate $AVS(a)$, a set of audio and visual signatures. If a is a primitive or a functional part then $AVS(a)$ includes, among other properties, inertia moments, material, and emptiness. If a is a connection, $AVS(a)$ includes, for example ratio of volumes. The full description of the visual signatures we have considered is relatively large and can be found at www.cs.technion.ac.il/~mpechuk/ProjectOCLS/index.html.

Each signature, either audio or visual, is associated with a histogram of measured values. The histograms are built employing B-spline functions [7].

Consider a multi-level hierarchy and let P and F be the set of all the symbolic primitives and func-

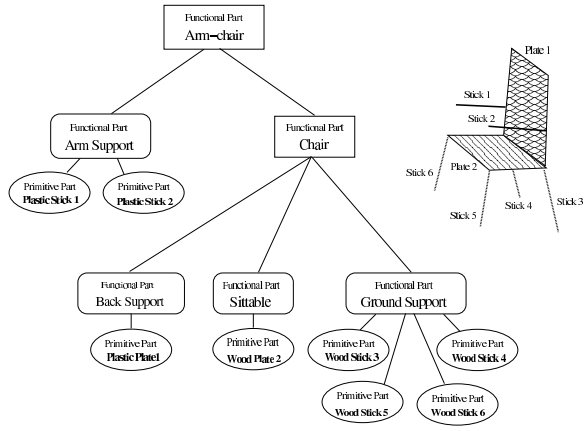


Fig. 1. The multi-level hierarchy functional structure. $F(\text{Arm-chair}) = \{\text{Arm Support}, \text{Chair}\}$. The arms represent a simple functional part with the functionality of supporting the arms, while the chair is a high level functional part due to the fact that it describes a more complex functionality.

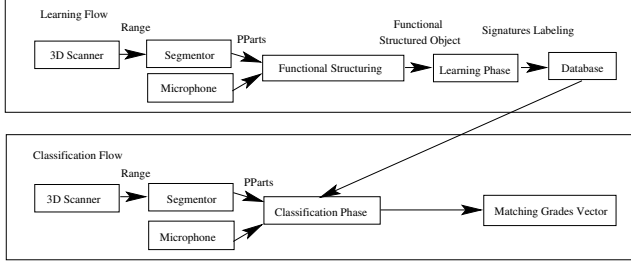


Fig. 2. Learning and classification flows.

tional parts, respectively, the hierarchy includes. Define $A = \{(a, s) \mid a \in F \cup P, s \in AVS(a)\}$ and $B = \bigcup_{f \in F} \{(c, s) \mid c \in C(f), s \in AVS(c)\}$. Then, the multi-level hierarchy of a functionality f induces a function $\mathcal{H}_f : A \cup B \rightarrow H$, where $H = \{h \mid h : R \rightarrow [0..1]\}$ is the set of all (normalized) histograms that can be implemented as B-spline functions (see next section).

B. Learning Functionalities

Fig. 2 (upper part) shows the flow of the learning phase of our scheme. The input of the learning phase is a set of functional labeled objects. Each functional and primitive part is labeled with a symbol or a generic name. For each input object, the proposed scheme calculates the values for all the pre-defined visual and audio signatures. Further, these signatures values are subject to an RBF-like (radial-based function) learning [20].

In the learning phase, the scheme builds histograms for signatures of the functional parts as well as for the connections between the functional parts. For each functional part, the set of histograms of its constituents, functional (sub)-parts and connections, represents the signature of the functional part.

C. Classification

Fig. 2 (lower part) shows the flow of the classification phase of our scheme. In the classification mode, the input consists of a set of primitive parts, the connections between them, and

the multi-level hierarchy provided by the learning phase. The classification phase computes a vector of grades that describes how an object offers class functionalities. Each element of the vector represents a grade relative to one class. The class with the highest matching grade is chosen as the best match found by our scheme.

1) *Matching Grades Computation*: Assume we want to evaluate the matching grade for functionality f and FH is a multi-level hierarchy computed for f . Let P_{in} an input set of primitive parts which are to be partitioned in the functional parts of FH in order to recover f . Let \bar{P} be the set of all hypothetic partitions of P_{in} in a FH structure and for any $r \in F(f) \cup P(f)$ let $\bar{p}(r)$ be the sub-set of the partition \bar{p} relative to r . Define

$$\begin{aligned} \text{grade}(f, \bar{p}) &= \sum_{s \in AVS(f)} \mathcal{H}_f(f, s)(\bar{p}) \prod_{r \in (F(f) \cup C(f))} \text{grade}(r, \bar{p}(r)) \end{aligned}$$

if $P(f)$ is empty and

$$\begin{aligned} \text{grade}(f, \bar{p}) &= \left(\sum_{s \in AVS(f)} \mathcal{H}_f(f, s)(\bar{p}) \right) \\ &\quad \left(\sum_{r \in (P(f) \cup C(f))} w(r, s) \text{grade}(r, \bar{p}(r)) \right) \end{aligned}$$

otherwise. Here, $w(r, s)$ is a weight function that is proportional with the standard deviation [14] of the histogram function itself corresponding to r and s (not the histogram's range) and $\mathcal{H}_f(r, s)(\bar{p})$ means the value of the signatures histogram for r implemented following partition \bar{p} . If p is a primitive part then the matching grade $\text{grade}(p, \bar{p}(p)) = \sum_{s \in AVS(p)} \mathcal{H}_f(f, s)(\bar{p}(p))$. If c is a connection, then

$$\text{grade}(c, \bar{p}) = \sum_{s \in AVS(c)} w(c, s) \mathcal{H}_f(c, s)(\bar{p}).$$

Moreover, for any functionality f , the matching grade is defined as

$$\text{grade}(f) = \max_{\bar{p} \in \bar{P}} \text{grade}(f, \bar{p}).$$

The classification phase is a search and validation like algorithm over a finite graph of partitions.

2) *Matching Partitions to Functionalities*: We define the matching of partitions to functionalities as a search in a finite graph problem. The search space is a graph with $(m+1)^n$ nodes. The first state of the graph is always the "empty" state, i.e., all the primitive parts are located in the "non-partitioned" section. The children's generation function takes a state of level $k-1$ in the graph and generates all possible realizations for the k -th functional part from the "non-partitioned" primitive parts set. It assumes that the previous $k-1$ functional parts are already realized. Thus, the last level contains all possible

partitions of the object. The goal is to find the state with the highest matching grade.

We used a heuristic search with a pruning branch-and-bound approach. Define the partial matching grade of a primitive part, a functional part, or connection a relative to partition \bar{p} be

$$partial(f, n) = \begin{cases} grade(f, \bar{p}) & \text{if } f \text{ is assigned} \\ 1 & \text{otherwise} \end{cases}, \quad (1)$$

where n is a node in the search graph. For searching purposes we use partial matching grades. From (1), it follows that when the search reaches a leaf, the partial grade equals the matching grade. Following [11], the algorithm searches for the partition that has the highest matching grade.

IV. EXPERIMENTS

We tested our scheme on a database comprising synthetic models of 200 forks, 216 spoons, 200 stools, and 200 spectacles. We also tested our scheme on a database comprising 100 forks, 100 spoons, 97 chairs, 100 spectacles, 118 airplane models, 30 cupboards, 20 mugs and cups, and 15 tables of real range images. The objects are built from wood, plastic, metal, ceramic, glass, and combinations of them. Partial sets of the forks, spoons, chairs, the spectacles, and cups and mugs that we used in experiments, are shown in Fig. 8. The objects in the range images were captured with a Cyberware range scanner (<http://www.cyberware.com>). The sounds registration was done using a robotic arm (see Figure 4 and [22]) equipped with a uni-directional condensers microphone with frequency range of 300-10000 Hz. We also considered reconstruction of 3D objects from stereo images (see Figure 3).



Fig. 3. On each row, left and right images as taken from the robotic arm in Figure 4. The middle images are their range reconstruction.

A. Experiments on Audio Classification

We tested the performance of the audio classifier for deciding if glass objects are full or empty. In addition, we tested the quality of classification of four different studied materials: plastic, ceramic, wood, and glass. The classifier was designed to decide that the analyzed object is built from a certain material (glasses being allowed to be full or empty) considering the maximum detected grade. The average of

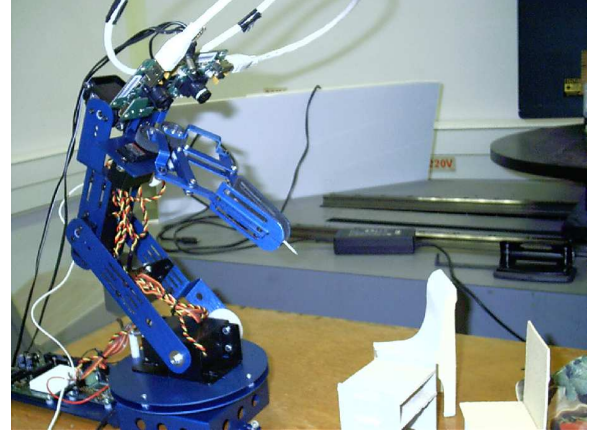


Fig. 4. Image of a robotic arm equipped with a stereo vision system and a microphone.

success of the classifier is 71.43%. Note that we built the audio based classifier for testing the quality of classification the audio signatures provide; for fusion purposes, we use the audio modalities as signatures.

1) *Receiver Operator Characteristic - ROC*: We have used a training set of fifty samples: ten samples of each one of the classes (wood, plastic, ceramic, glass, and full glass). In this case we have a multi-category classifier and the decisions are based on a unique threshold over all the five classes, that varies from 0 to 1 for ROC constructing purposes of course. For example, if a wood object is given a grade of 0.8 as wood and 0.2 as a ceramic object, it is both a hit and a false alarm for the threshold of 0.1. Unlike 0.1, for the threshold of 0.5 it is only a hit. The results of our classifier can be seen in Figure 5.

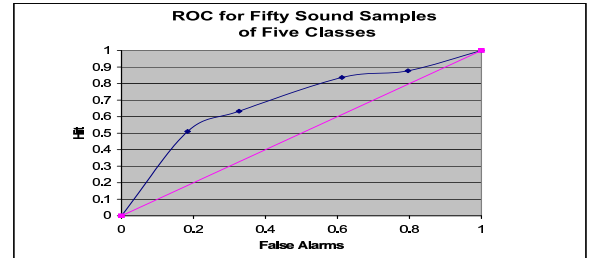


Fig. 5. A receiver operator characteristic for sounds on fifty samples.

2) *Cross Validation*: A material is recognized when the classifier recognizes it with the highest percentage from all other classes. We have tested fifty new samples. The results are represented in Figure 6.

B. Classifying 3D Data

We outline two types of experiments aiming to check model strength, and ROC and accuracy. In all the tests, the learning phase was performed on images that contain only one object. In model strength checking, and ROC and accuracy tests, we used the proposed scheme to classify objects from 3D images

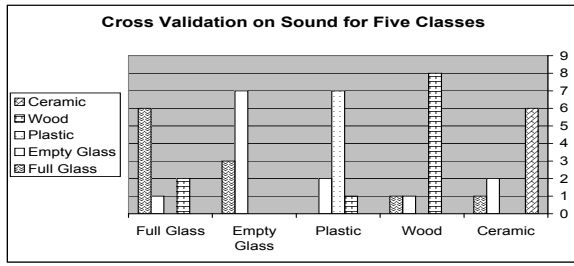


Fig. 6. Cross validation for five classes.

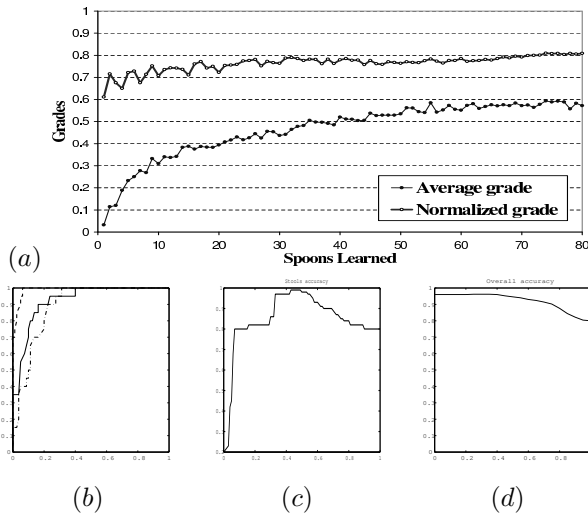


Fig. 7. Experiments: (a) Learning and classifying real spoons. (b) A ROC on the whole data base for stools (the uppermost curve), forks, and spectacles (the lowest curve). (c) The accuracy of stools versus the rest of the data base. (d) Overall accuracy.

that contain only one object. These tests show that a relative small set of learnt objects suffices for a good classification performance.

1) *Model Strength Checking*: In the model strength checking experiment, two sets of objects were used: a learning set and a test set. The graph in Fig. 7 (a) shows the test set average grades as a function of the size of the training set. Here, the lowest curve represents the average of the grades of the classified objects and the higher curve shows the percentage of the test set's average grade from the maximal grades in the test set. In the experiment shown in Fig. 7 (a) the learning sets consisted of real scanned objects. The set used for classification is constant per experiment and comprised all the scanned objects.

2) *ROC and Accuracy*: Consider classifications that work on components of the vector matching grades (the classifiers do not perform maximum on components). In Fig. 7 (b), we show the ROC superimposed curves of stools, forks, and spoons. In Fig. 7 (c), we show the accuracy of the classifier of stools versus the data base. In Fig. 7 (d), unlike in Fig. 7 (b) and (c), we show the accuracy of the classifier that performs maximum on vector grades, targeting all five tested classes and the whole data base.

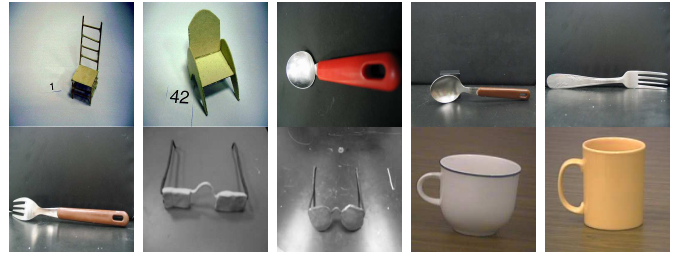


Fig. 8. Images of some objects used in our experiments.

C. Experiments on Fusion Between the 3D and Audio Classifiers

Our database include 20 cups and mugs where 16 of them are ceramic objects, two are plastic objects, and two are glasses. In addition, we have used a wood bowl. Here, we only present two experiments due to lack of space.

1) *Refined Cross Validation*: After learning all the cups and mugs in the database we provided to the visual and to the combined visual and audio classifier the wood bowl in Figure 9. Although the visual classifier recognized it as a cup, the combined one provided a very low grade, due to the fact that all our cups and mugs are from ceramic, plastic, and glass.



Fig. 9. A wood bowl recognizable as a cup in 3D.

2) *Empty and Full Cups and Mugs Classification*: We have learnt all the cups and mugs. We report that all of the 20 mugs and cups were correctly classified by the 3D scheme. In addition, we report that both of the empty and full glass were recognized. We repeated the classification process by stroking the full and empty glass mugs in several locations. As expected, the results followed the percentage of recognition provided by the cross validation scheme in Figure 6.

V. CONCLUSIONS

In this work, we have presented a novel scheme for a functional based classifier employing fusion of audio and visual modalities. While the first modality performs classification of functionalities from sounds registered at impact,

the second one targets classification of objects in 3D images. The input objects are full 3D descriptions of objects also including sound produced when struck. The proposed scheme employs an object functional structure, consisting of a multi-level hierarchy of functional parts. The multi-level approach offers a higher degree of freedom for real object modelling as compared to classical systems and can be seen as a learning phase.

Our approach was tested on a database of about one thousand different 3D objects employing several algorithms for searching and pruning. To the best of our knowledge, no other classification (or recognition) scheme was tested on hundreds of range images of real objects captured in range images. Moreover, no fusion for functional based classifiers designed for different sensors was proposed. The graphs show that our scheme represents a reliable classification method. They also provide an insight into the dimensions of the learning sets that are required so as to reach a certain degree of classification accuracy. Our work appears to be the first scheme that performs function-based classification that involves a learning phase and does not require a-priori knowledge about any class.

Part of our future work consists of introducing more modalities such as temperature or feedback force sensors. Moreover, we intend to text mobile robots equipped with arms in more complex scenes.

ACKNOWLEDGMENTS

The authors thank Diana Soldea, Michael Pechuk, Nela Gurevich, Ronen Lerner, Alexander Porotskiy, Ronen Keidar, and Yekutiell Pekar for their support.

REFERENCES

- [1] F. Avraini and D. Rocheso, Controlling material properties in physical models of sounding objects, International Computer and Music Conference, La Habana, September, 2001.
- [2] I. Biderman, Recognition by Components: A theory of human image understanding, MIT Press, Psychological Review, 94(2), pp. 115–147, 1987
- [3] N. Cheka and K. Wilson, Person tracking using audio-video sensor fusion, MIT Artificial Intelligence Laboratory, 2001.
- [4] S. Dickinson, R. Bergevin, I. Biederman, J. Eklundh, R. Munck-Fairwood, A. Jain, and A. Pentland, Panel report: The potential of geons for generic 3-d object recognition, IVC, Volume 15, Number 4, pp. 277–292, April, 1997
- [5] M. DiManzo, E. Trucco, F. Giunchiglia, and F. Ricci, FUR: Understanding functional reasoning, IJIS, Volume 4, pp. 431–457, 1989
- [6] R. Durst, E. Krotov, Object classification from analysis of impact acoustics, IROS, Pittsburgh, pp. 90–95, 1995
- [7] G. Farin, Curves and surfaces for CAGD - a practical guide. Fourth Edition, Academic Press Inc., Publishers New York, 1996
- [8] T. A. Faruque, A. Majumdar, N. Rajput, L. V. Subramaniam. Large vocabulary audio-visual speech recognition using active shape models, ICPR, Volume 3, pp. 106–109, 3-7 September 2000
- [9] B. Giordano, Material categorization and hardness scaling in real and synthetic impact sounds, The Sounding Object, Chapter 5, Edizioni di Mondo Estremo, Florence 2003.
- [10] M. Grassi and R. Burro, Impact sound. The Sounding Object, Chapter 5, Edizioni di Mondo Estremo, Florence 2003.
- [11] W. E. L. Grimson, The combinatorics of heuristic search termination for object recognition in cluttered environments, IEEE PAMI, Volume 13, Number 9, pp. 383–397, September, 1991
- [12] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, An experimental comparison of range image segmentation algorithms, IEEE PAMI, Volume 18, Number 7, pp. 673–688, July, 1996
- [13] D. W. Jacobs, Matching 3-D models to 2-D images, IJCV 21(1/2), Kluwer Academic Publishers, Volume 21, Issue 1-2, pp. 123–153, 1997
- [14] J. F. Kenney and E. S. Keeping, Mathematics of statistics, Pt. 1, 3rd ed., Princeton, NJ: Van Nostrand, pp. 77–80, 1962
- [15] R. Klatzky, D. Pai, and E. Krotkov. Perception of material from contact sounds, teleoperators and virtual environments, MIT Press, August 2000.
- [16] E. Krotkov, Robotic perception of material, Proceedings of the IJCAI, pp. 88–94, August, 1995.
- [17] E. Krotov, R. Klatzky, N. Zemel, Robotic perception of Material: Experiments with shape-invariant acoustic measures of material, Experimental Robotics IV, 1996.
- [18] L. Ma, D.J. Smith, and B.P. Milner, Context awareness using environmental noise classification, Eurospeech 2003, Geneva, Switzerland, pp. 2237–2240, 2003.
- [19] I. Matthews, T. F. Cootes, J. A. Bangham, S. Cox, and R. Harvey. Extraction of visual features for lipreading, IEEE PAMI, Volume 24, Number 2, pp. 198–213, February 2002.
- [20] T. M. Mitchell, Machine learning, McGraw-Hill Companies, 1997
- [21] N. Oliver, A. Garg, and E. Horvitz, Layered representations for learning and inferring office activity from multiple sensory channels, CVIU, Volume 96, Issue 2, pp. 163–180, November 2004.
- [22] CrustCrawler's SG5-HT/UT Series Robotic Arm Systems, <http://www.parallax.com/html-pages/robotics/roboarm.asp>
- [23] M. Pechuk, O. Soldea, and E. Rivlin, Function-based classification from 3D data via generic and symbolic models, AAAI-05, Pittsburgh, Pennsylvania, USA, pp. 950–955, July 9–13, 2005
- [24] D. Rochesso, L. Ottaviani, and F. Fontana, Size, shape and material properties of sound models. The Sounding Object, Chapter 6, Edizioni di Mondo Estremo, Florence 2003.
- [25] L. Stark and K. Bowyer, Function-based generic recognition for multiple object categories, CVGIP, Volume 59, Number 1, pp. 1–21, January, 1994
- [26] M. A. Sutton, L. Stark, and K. Bowyer, Gruff-3: Generalizing the domain of a function-based recognition system, Pattern Recognition, Volume 27, Number 12, pp. 1743–1766, December, 1994
- [27] P. Teissier, J. Robert-Ribes, J.-L. Schwartz, and A. Guerin-Dugue, Comparing models for audiovisual fusion in a noisy-vowel recognition task, IEEE Transactions on Speech and Audio Processing, Volume 7, Issue 6, pp. 629–642, November 1999.
- [28] Y. Toyoda, J. Huang, S. Ding, Y. Liu, Environmental sound recognition by multilayered neural networks, Fourth International Conference on Computer and Information Technology, pp. 123–127, 2004.
- [29] K. van den Doel, D.K. Pai, The sounds of physical shapes, Technical Report 96-03, Department of Computer Science, University of British Columbia, 1996.
- [30] R. Whitman, How to play 20 questions with nature and win, December 1982.
- [31] R. Wildes R. and W. Richards, Recovering material properties from sound, Natural computation, Cambridge, MIT Press, 1988.
- [32] P. H. Winston, T. O. Binford, B. Katz, and M. Lowry, Learning physical descriptions from functional descriptions, AAAI-83, pp. 433–439, 1983
- [33] K. Woods, D. Cook, L. Hall, K. Bowyer, and L. Stark, Learning membership functions in a function-based object recognition system, JAIR, Volume 3, pp. 177–222, 1995
- [34] <http://tosa.mri.co.jp/sounddb/nospeech/system/indexe.htm#impulse>