Title	The Generalist Approach to Frame Problems
Author(s)	Hidaka, Shohei; Kashyap, Neeraj
Citation	Proceedings of Asian Conference on Information Systems 2014 (ACIS2014): 311-318
Issue Date	2014-12
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/12884
Rights	Shohei Hidaka and Neeraj Kashyap, Proceedings of Asian Conference on Information Systems 2014 (ACIS2014), 2014, pp.311-318.
Description	

The Generalist Approach to Frame Problems

Shohei Hidaka[†] **Neeraj Kashyap**
[†]School of Knowledge Science
Japan Advanced Institute of Science and Technology, Japan
shhidaka@jaist.ac.jp

Abstract

The frame problem commonly appears in the field of artificial intelligence and in related philosophical literature. It pertains to the difficulty of making meaningful inferences in dynamic contexts. We discuss the frame problem and its manifestations in the practical design of artificial agents, and we set out some criteria against which the effectiveness of such an agent may be judged. Finally, we claim that technology currently exists in the form of a new clustering technique - dimensional clustering - which can enable an agent to satisfy these criteria. We justify this claim with three case studies where we apply dimensional clustering to data sets of varied nature.

Keywords: Frame problems, Dimensional clustering, Artificial Intelligence

1 Frame problems

The frame problem expresses the difficulty in deciding what counts as a meaningful basis for inference in various contexts. Human cognitive processes are regularly confronted by such problems in everyday life, but seem to overcome them with a reasonably high rate of success. One such example is the way children learn the meanings of words. Children typically start producing their first words before their first birthdays, and typically learn over a hundred words by their second birth-days. Children as young as three years old are known to show excellent inferences about the names of solid objects. These inferences can be described by the phenomenon of *shape bias*, discovered by cognitive developmentalists [1; 2] — when confronted with a never-before-seen solid object, three-year-olds associate it with the names of objects which they know about and which are similar to it in shape. For other types of objects, children generalize names based on other kinds of similarities — they use colour to generalize names for deformable objects like jellies [3], and have been shown [4; 5] to use combinations of multiple features in generalizing names for complex objects with many visible parts (like eyes and legs). These findings reflect how children naturally overcome the frame problem in world learn-ing situations. Their favoured heuristic — generalizing the scope of a name based on a particular type of similarity — is generally reasonable as far as the typical list of early learned words by three years olds is concerned [6]. It is likely, for example, that shape bias facilitates children's word learning [2]. Shape bias is *itself* learned gradually, as younger children do not exhibit the same tendency [3].

What we call a frame problem is a generalized version of the original frame problem discussed by McCarthy and Hayes [7] in the field of artificial intelligence (AI) as a logical problem concerning concise axiomitizations of logical rules determining environmental behaviour in dynamic contexts. The frame problem was reinterpreted along these lines as a broader philosophical problem by Dennett [8] and Fodor [9] (See also [10]). It is this general formulation that we address in this paper — the problem of how an agent may effectively determine the context of its operation.

2 Abstraction vs. generality

In philosophical arguments, one major obsta-cle to addressing the frame problem is the notion of *infinite regress*, which concerns the necessity of working within a higher-level frame in order to choose a frame for a given situation. Choosing the higher-level frame itself, then, requires one to work at an even *higher* level of abstraction, and this nesting of frames can never end. Thus, by the argument of infinite regress, it seems computationally intractable for one to choose an appropriate frame for a given problem or task.

The approach of the infinite regress argument, which we call the *abstractionist* approach, ignores a potential method for frame inference which we call the *generalist* approach. The abstractionist approach focuses nesting frames in a hierarchical manner. In the ab-stractionist approach, a frame at level 1 is specified by a certain frame at level 2, which is itself specified by some frame at level 3, and so on. In the generalist approach, one attempts to widen the scope of applicability of their frames within a fixed level of ab-straction so that these frames are powerful enough to solve of a majority of the problems that one is likely to face.

Going back the example of how children learn solid-object semantics, the abstractionist approach would be to state that children choose their "shape" frame because of a higher-level frame (such as their naïve understanding of the physical laws that a solid object is subject to) which gives rational justification that shape is a reliable basis for categorization. Such a theoretical account for word learning is called a theory-theory account in developmental studies (for example, Rogers and McClelland [11]).

In contrast, the generalist approach would be to state that shape bias in children of a certain age is due to the empirical and statistical robustness of shape similarity, which is gener-ally applicable to learning the specific set of words most likely to be encountered by children of that age [6; 2]. Specialists call this the statistical learning account of word learning. Note that, in this statistical account, one needs only a first- or second-level of frame, but the top level frame needs to have a broad enough scope (the scope here being words that children are most likely to encounter by age 3, say). Moreover, by identifying them as informative features of objects of each category, a generalist description of word learning through shape bias could, without any increase in abstraction, extend to a description of world learning through the colour and part similarities discussed in the previous section. An abstractionist may rather have to start from scratch and build independent theory-theories for both of these learning methods as well as a theory-theory-theory in order to connect the various theory-theories.

In support of the generalist approach to word learning, Smith and collegues [2; 12; 13; 14] have claimed that a computational model with a second-level system for inference can account for novel word generalization behaviors in children. Note that the abstractionist approach is not truly feasible if one is interested in the practicalities of word learning in children, for the theory-theory itself requires the specification of a theory-theory-theory, and so on. It is not at all clear that one can develop a formalism within which to describe the entirety of the theories regarding theories which are somewhere down the line involved with word learning, including the formalism itself in which the theories were being described. Even supporters of the theory-theory approach are being generalists, but are merely attempting to generalize frames at a higher level of abstraction than proponents of the statistical learning account of word learning.

3 The Generalist Approach in AI

Most problems in the field of Artificial Intelligence, particularly those considered in early symbolic AI studies consider solvers of problems in which a frame is already specified. Frames are typically decided upon by a researcher and are tailored to specific situations that they are interested in studying. Such studies aim to establish a frame for a specific problem, and demonstrate the type of machinery that can give reasonable solutions to it. An early and very successful example of this approach is computer chess. As the game framework can be written rigorously in a symbolic system, and as the goal of each player is clearly defined and static, a reasonable solution for the game of Chess can essentially be found by searching over the combinatorial space of series of moves, the "game tree". With sufficient computational resources and some improvements to this basic strategy, computer chess programs have become stronger even than the highest-level human players.

Agents like these, however, are not adaptive to tasks which are even slightly different from the ones they have been designed for. They are not robust against even slight perturbations in frame. For example, the chess computer Deep Blue cannot be used as it is for Shogi (Japanese chess), which has similar rules to Western chess, which has similar rules but requires a different set of specialized techniques [15].Taking another dynamic and open example, hand-coded programs designed to let a robot walk on smooth, flat surface may fail if the robot tries to walk on an unexpectedly rough surface (this is the case for a static walking framework called "flat-footed waking", but see Westervelt et al. [16] for a modern, dynamical framework). This approach, where an agent is reliant on an external frame specification, is quite limited when it comes to solving real-world problems. It is generally infeasible to exhaustively list the potential situations that the agent may face in the real world.

Many modern machine learning studies aim to give systematic methods for selecting frames. In such studies, researchers prepare a collection of frames, each of which is represented as computational model. They then choose the best model or set of models by evaluating them against experimental data. In the context of statistical modeling, this kind of frame selection is called *model selection*. It requires the formulation of the researchers' prior knowledge about the data and problems, such as smoothness, sparseness, hierarchical relationships, dimensionality, and so on. Essentially, the researchers are specifying a *hyper-model*. There is one level of abstraction between their models and their data, and one more level of abstraction between their hyper-model and their model. There are two distinct components of such methods: the data-to-model component, and the model-to-hyper-model component. We call an agent which implements model selection in this manner a *frame selector*.

A simple example of a frame selection technique is that of *sparse modelling* [17], which selects sparse regression models. The data-to-model component in sparse modelling describes how data is generated by a linear combination of some given variables, and is a standard method of data analysis. In the model-to-hyper-model component, a prior distribution imposes a preference that the data-to-model component use a small number of parameters. The decision to use sparse modelling to analyze a data set signifies, firstly, the analyst's belief that the data could be modelled well as a linear transformation of some subset of its parameters. Secondly, it suggests that the analyst believes that one could achieve an acceptable level of accuracy using very *few* such parameters. This is an *abstract* frame problem, which is solved by the analyst. However, sparse modelling algorithms solve the further frame problem of exactly *which* sparse list of parameters the data can be modelled by.

The scope of the sparse modelling extends to the analysis of all data sets which fall in line with the sparse modelling analyst's beliefs regarding their linearity and sparseness. This scope is probably not broad enough to allow the solution of many different types of problems. However, a frame selector with much broader scope would be extremely valuable when used in conjunction with more specialized agents. This is the generalist's holy grail in the field of Artificial Intelligence.

4 Building a better frame selector

The type of hierarchical approach to model selection allows for a more flexible framework and expands the applicability of the statistical techniques to a wider range of problems. The key issue in such hierarchical modeling is the formulation of one's prior knowledge of specific problems — how should one select the frames that the solver will choose from? This is essentially the problem of identifying what count as features or objects in data.

As discussed in Section 1, we can take one of two approaches, abstractionist and generalist, for this identification problem. The abstractionist approach adds a level of abstraction which provides rules about what to count as features and objects. The generalist approach is to construct as general as possible a frame at the top level of abstraction in the given hierarchy.

Regardless of approach, one needs to identify the salient objects and features in a given environment and model each object or feature at a level of detail appropriate to it. More detailed modeling of objects and features forces more complex inferences regarding salience to a given context. This makes the frame problem more serious. More coarse-grained, and thus general, modeling allows for simpler inferences and therefore relaxes the frame problem. However, the more general a model, the more it lacks in predictive power.

Going back to our example of world learning, it seems that children practically solve this problem regarding level of detail using similarities. At this level of generality, they can avoid serious frame problems while retaining predictability in generalizing to novel words. If children used more than one feature to determine similarity, for example both shape and colour, it would increase not only their inferential load but also their odds of under-generalizing their known vocabulary. Colour, for example, is frequently extraneous to the names of solid objects and is therefore indicated using adjectives. On the other hand, if children did not distinguish finely between shapes, they might make errors of over-generalization. For example, taking "roundness" as an atomic feature, they would neglect to distinguish between balls and wheels. Finding a level of generality which allows modeling of a wide range of tasks is crucial to generalist approaches to the frame problem.

The problem of identifying the salient objects and features in an agent's environment is much more difficult to address than that of determining the levels of detail at which each of these objects should be modelled. While children are certainly able to identify distinct objects in their environments, there is currently no definitive account of *how* they do so. This is an even more fundamental problem in the design of autonomous agents than the problem of levels of detail, although the two can certainly be considered as being related.

The goal of the generalist is not to build a perfect and rational solver of the frame problem but to formulate a practical criterion by which an agent may determine the level of detail at which to model various objects. Currently, most frame selectors rely on human knowledge, input, and reinforcement. The real objective in AI, however, is to build effective frame selectors which are also *autonomous*. Such frame selectors must come equipped with a method of solving the feature identification and detail problems independent of human involvement.

In summary, an *effective* frame selector must be:

1. Autonomous.
2. Able to identify the distinct features and objects in its environment.
3. Able to select models for each object and feature which are appropriate in complexity to their relevance towards the agent's task.

5 Practical Feature Detection

The biggest obstacle to building an effective frame selector according to the criteria of the previous section is the matter of autonomy, and one of the primary reasons that this is such an obstacle is that it is difficult for an agent to autonomously identify the individual objects in and features of its environment. In this section, we discuss a technique which makes a significant step towards solving this problem of feature detection. The technique in question is that of *dimensional clustering* [18].

In the absence of semantics-imposing rules, an agent can only be aware of its environment in terms of data. Dimensional clustering is a technique which operates on numerical data, i.e. data consisting of vectors in some Euclidean space. The objective of dimensional clustering is to decompose the *generating process* of such a dataset into its primitive components.

Our current algorithms for dimensional clustering involve the estimation of a fractal dimension known as *pointwise dimension* at various points in a data set. The algorithms separate the data points into clusters based on their similarities and differences in terms of their pointwise dimensions as well as a measure of the density of the data set around each point.

Dimensional clustering itself allows an agent to detect features in its environment, but the additional dimension and density information also allow the agent to model these features more effectively. This is because pointwise dimension is invariant under a large class of very natural transformations and so the similarities between points of similar pointwise dimension tend not to be mere byproducts of data representation. Rather, such similarities indicate similarities in the fundamental procedure by which both data points were generated.

We demonstrate our technique with three case studies. Although these case studies differ greatly from one another, they should not be considered independently. Taken together, these three diverse applications demonstrate the great flexibility of dimensional clustering. This is exactly the kind of flexibility that an effective frame selector must exhibit.

5.1 Case Study 1: A Random Walk

In the first case study, we examine a data set generated by two distinct sub-processes. One is a random walk along a line, and the other is a random walk in a plane. The data set is generated by running these processes alternatingly, twice each, and such that the two 1-dimensional phases operate in orthogonal directions. As these processes differ in dimension, our algorithm is expected to detect that the data points from the 1-dimensional process were generated differently than those from the 2-dimensional process.

Explicitly, the dataset in Figure 1 was generated by a random walk in the plane subject to the following rules:

1. The random walk begins at the origin and the y-coordinate remains fixed for the first 2,500 steps. The changes in the x-coordinates at each of these steps are independent and identically distributed to a normal random variable X with mean 0 and standard deviation 0.05.

2. For the second 2,500 steps, both x- and y-coordinates are allowed to vary, with the changes in each coordinate at each step being independent and identically distributed to X.

3. For the third 2,500 steps, the x-coordinate remains fixed and the changes in the y-coordinates are independent and identically distributed to X.

4. For the fourth 2,500 steps, the walk proceeds as it did in the second 2,500 steps, with no restrictions in either direction.

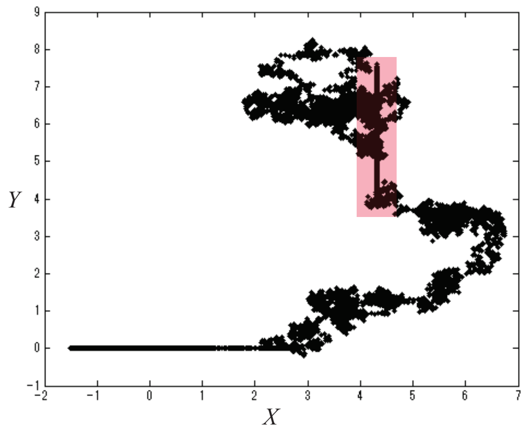

Figure 1. The points within the red rectangle make this data set particularly difficult to cluster.

The data set consists of the points in the plane that are visited on this random walk, stripped of all time information.

One could reasonably consider this generating process as being comprised of either two or three components according to whether the two 1-dimensional components are identified in terms of their dynamics or not. Regardless of the point of view one adopts, there is certainly no considerably spatial separation between the components of this process. The data generated by the 2-dimensional phases of the random walk show considerable overlap with those generated by the 1-dimensional phases. This is especially true of the data points within the region inside the red rectangle in Figure 1.

Without a description of the generating process, it would be very difficult even for a human being to make out that there is a line of 1-dimensional data cutting through the cloud of 2-dimensional data rather than two small lines of 1-dimensional data jutting out of a cloud of 2-dimensional data. To justify such a claim, one might have to resort to probabilistic reasoning — assuming that the generating process is sufficiently random, it is much more likely that there is a line cutting through the cloud than two lines jutting out of it. As difficult as this is for humans, certainly no conventional clustering algorithm is capable of differentiating between the two (or three) components of the generating process without an added temporal dimension to the data.

Figure 2 shows the dimensional clustering of this random walk data. The dimensional clustering algorithm detected two clusters in the data, and each point was assigned to the cluster to which it belonged with maximal probability. The blue cluster corresponds to the 1-dimensional component of the random walk, and the red cluster corresponds to the 2-dimensional component. Figure 3 shows the data *with* the corresponding time information and with the data points coloured according to the results of the clustering. This figure shows that the dimensional clustering algorithm was able to discern 3-dimensional information from the original 2-dimensional data set.

Note that the algorithm was able to separate the points within the red rectangle of Figure 1 which belonged to the 1-dimensional component from those which belonged to the 2-dimensional component.

5.2 Case Study 2: Balance Data

In the second case study, we analyzed time series of human movements. Specifically, we obtained high resolution time series of the center of pressure (COP) collected from subjects standing on a

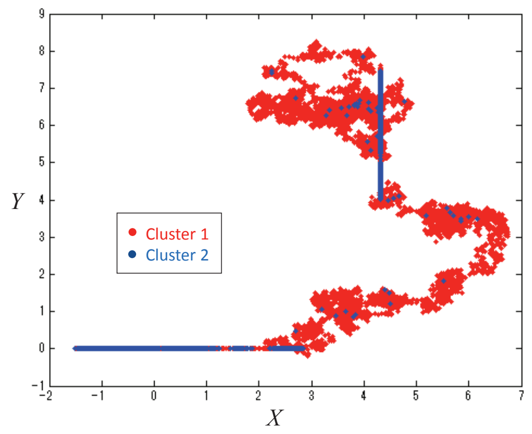

Figure 2. Dimensional clustering sharply distinguishes between the 1- and 2-dimensional components of the random walk.

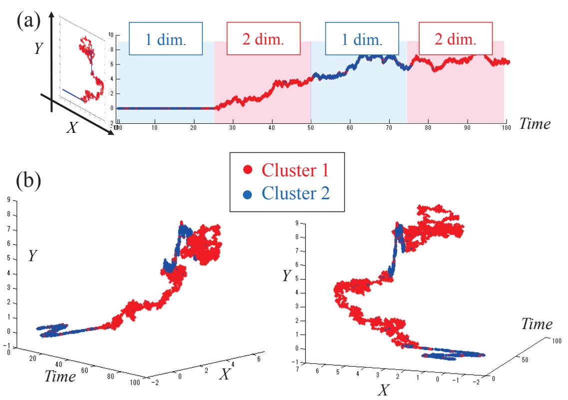

Figure 3. Dimensional clustering recovered temporal information about the random walk data which had been projected away.

pressure plate. The experiment had four healthy subjects with one session per subject. Each session consisted of three phases —

1. Phase 1: The subject was instructed to stand still for 30 seconds.

2. Phase 2: The subject was pushed sporadically by the experimenters over a period of 60 seconds. The times at which each subject was pushed were determined randomly, but were uniform over the subjects.

3. Phase 3: The subject was instructed once again to stand still for 30 seconds.

The goal of this experiment was to produce a data set with at least two different modes of posture control — maintaining posture without any external force, and a quick posture control required by non-autonomous perturbation.

Explicitly, the datasets consisted of the location of the COP of each subject presented as X(lateral)-Y(anterior-posterior) coordinates. These coordinates were sampled one hundred times per second. However, the time information was not used in the dimensional clustering analysis.

Figure 4 shows a representative analysis. We found typically four or five clusters in each data set, but two clusters seem consistently correlated to (1) maintaining mode without perturbation (red points in Figure 4) and (2) short moments right after external perturbation (green points in Figure 4). The underlying mechanism of posture control is unknown at the moment. This analysis demonstrates that dimensional clustering is capable of providing useful information about data even in the complete absence of a model for its generating process.

5.3 Case Study 3: Image Data

For the final application of dimensional clustering we present in this paper, we analyzed a few standard test images in the field of image processing. These test images are presented as (a-1) to (a-4) in Figure 5. We obtained “Lena” and “Wet paint” from Mike Wakin’s website [19], and “Airplane” and “Fishing boat” from the USC SIPI Database [20]

For each test image in Figure 5, dimensional clustering was applied to a 3-dimensional data set of vectors listing the x-coordinate, y- coordinate, and grayscale value of each pixel in the image. Our algorithm discovered three or four clusters in each data set. Among these, there was always a “feature” cluster which discerned sharp changes in hue. Images (b-1) to (b-4) of Figure 5 show the pixels in each image belonging to the corresponding feature clusters. Images (c-1) to (c-4) of Figure 5 show the results of running the Canny edge detection algorithm [21] on the same test images, and are provided for comparison.

Figure 5(b-3) highlights the region in the feature cluster detected in Figure 5(a-3) which contains the sign. This shows that the dimensional clustering algorithm very clearly extracts its primary lettering.

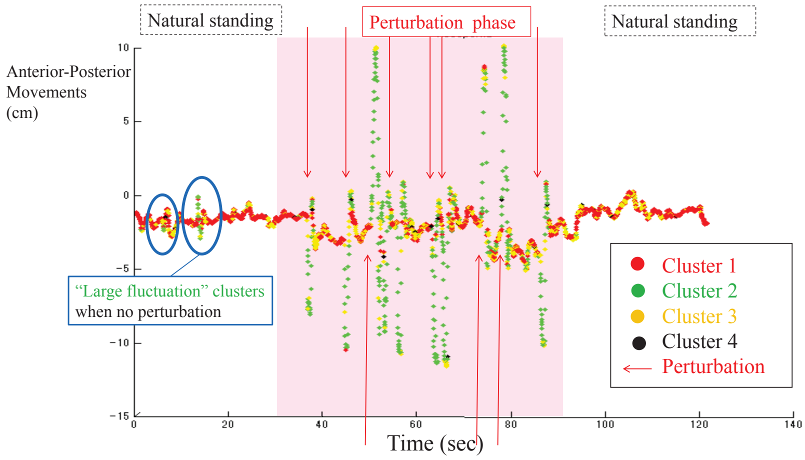

Figure 4. A typical COP time series with the spatial coordinates projected to the Anterior-Posterior axis. The four detected clusters are indicated by the four colours above.

Figure 5 demonstrates that dimensional clustering is at least capable of detecting the locations of feature pixels in the images. It is even able to detect the faint thumb print at the top of Figure 5(a-2)! However, the feature pixels only formed one cluster in the analysis of each data set. The remaining clusters detected in the image data sets seem to contain information pertaining to hue and depth. A detailed discussion of the information in these clusters would distract from the point we seek to make here regarding the scope of dimensional clustering. We therefore postpone a complete analysis of the image data sets to a future paper.

6 Conclusion

The discussion of the previous section, especially the case studies, shows that using dimensional clustering techniques along with the associated dimensional data can allow a frame selector to satisfy all three criteria for effectiveness stated in the previous section. We hope that this stimulates further investigation into our proposed generalist approach towards frame problems.

Acknowledgments

The authors would like to thank Tsutomu Fujinami and Wannipat Buated for their efforts in collecting the balance data presented in Case Study 2 of Section 5.2. This study was supported by the NeuroCreative Lab, Grant-in-Aid for Scientific Research B No. 23300099, and Grant-in-Aid for Exploratory Research No. 25560297.

References

[1] Barbara Landau, Linda B Smith, and Susan S Jones. The importance of shape in early lexical learning. *Cognitive development*, 3(3):299–321, 1988.

[2] Linda B Smith, Susan S Jones, Barbara Landau, Lisa Gershkoff-Stowe, and Larissa Samuelson. Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1):13–19, 2002.

[3] Mutsumi Imai and Dedre Gentner. A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2):169–200, 1997.

[4] Susan S Jones, Linda B Smith, and Barbara Landau. Object properties and knowledge in early lexical learning. *Child development*, 62(3):499–516, 1991.

[5] Hanako Yoshida and Linda B Smith. Shifting ontological boundaries: how japanese- and english-speaking children generalize names for animals and artifacts. *Developmental Science*, 6(1):1–17, 2003.

[6] Larissa K Samuelson and Linda B Smith. Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, 73(1):1–33, 1999.

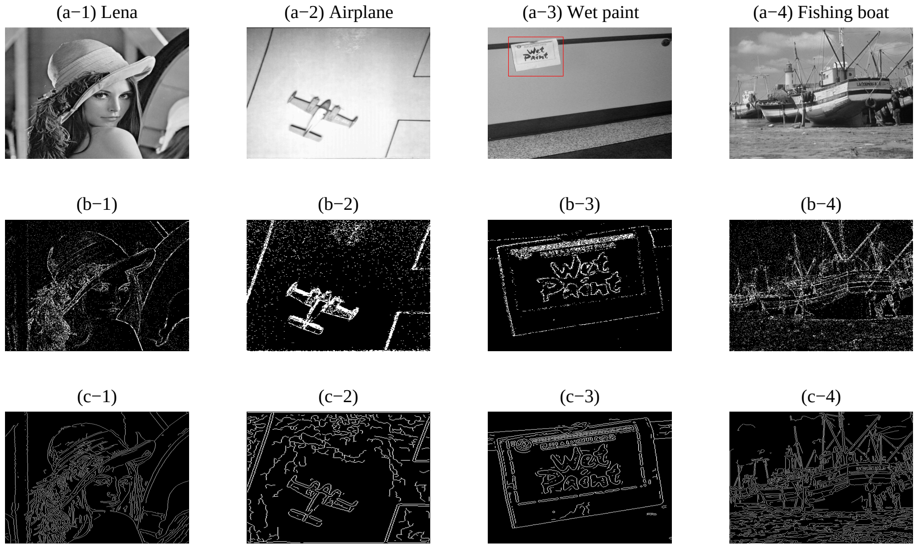

Figure 5. (a) Test images, (b) Feature clusters detected by dimensional clustering, and (c) Results of Canny edge detection on the test images.

[7] John McCarthy and Patrick Hayes. *Some philosophical problems from the standpoint of artificial intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1968.

[8] D. Dennett. *Cognitive Wheels: The Frame Problem in Artificial Intelligence*. Ablex, 1987.

[9] Jerry A Fodor. *The modularity of mind: An essay on faculty psychology*. MIT press, 1983.

[10] Murray Shanahan. The frame problem. `http://plato.stanford.edu/entries/frame-problem/`, 2009. Accessed: July 1st, 2014.

[11] Timothy T Rogers and James L McClelland. *Semantic cognition: A parallel distributed processing approach*. MIT press, 2004.

[12] Eliana Colunga and Linda B Smith. From the lexicon to expectations about kinds: a role for associative learning. *Psychological review*, 112(2):347, 2005.

[13] Shohei Hidaka and Linda B Smith. A single word in a population of words. *Language Learning and Development*, 6(3):206–222, 2010.

[14] Shohei Hidaka and Linda B Smith. Packing: A geometric analysis of feature selection and category formation. *Cognitive systems research*, 12(1):1–18, 2011.

[15] Hiroyuki Iida, Makoto Sakuta, and Jeff Rollason. Computer shogi. *Artificial Intelligence*, 134(1):121–144, 2002.

[16] Eric R Westervelt, Jessy W Grizzle, Christine Chevallereau, Jun Ho Choi, and Benjamin Morris. *Feedback control of dynamic bipedal robot locomotion*. Taylor & Francis, 2007.

[17] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

[18] S. Hidaka and N. Kashyap. Pointwise dimension estimation. under review.

[19] Mike Wakin. Standard test images. `http://www.ece.rice.edu/~wakin/images/`. Accessed: July 1st, 2014.

[20] University of Southern California. The USC-SIPI image database. `http://sipi.usc.edu/database/`. Accessed: July 1st, 2014.

[21] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.