Entropy-based Gaze Planning

Tal Arbel and Frank P. Ferrie
Electrical Engineering
McGill University, Center for Intelligent Machines
Montréal, Québec, CANADA H3A 2A7
tel. (514) 398 2185 fax:(514) 398 7348

Abstract

This paper describes an algorithm for recognizing known objects in an unstructured environment (e.g. landmarks) from measurements acquired with a single monochrome television camera mounted on a mobile observer. The approach is based on the concept of an *entropy map*, which is used to guide the mobile observer along an optimal trajectory that minimizes the ambiguity of recognition as well as the amount of data that must be gathered. Recognition itself is based on the optical flow signatures that result from the camera motion - signatures that are inherently ambiguous due to the confounding of motion, structure and imaging parameters. We show how gaze planning partially alleviates this problem by generating trajectories that maximize discriminability. A sequential Bayes approach is used to handle the remaining ambiguity by accumulating evidence for different object hypotheses over time until a clear assertion can be made. Results from an experimental recognition system using a gantry-mounted television camera are presented to show the effectiveness of the algorithm on a large class of common objects.

1 Introduction

Many tasks in mobile robotics involve the recognition of known objects and structures in potentially unstructured environments. Such tasks include the identification of landmarks, retrieval or inspection of particular objects, and the determination of the relative pose between the robot and objects in the environment. Relative to fixed installations, computational and sensory capabilities are often limited as is the amount of energy that can be expended in performing a particular task. For this reason we are interested in algorithms that are computationally inexpensive, robust, and operate with relatively modest sensory requirements. In this paper we investigate such an algorithm - one that uses the optical flow patterns that result from the motion of a single monochrome television camera to recognize the identity and determine the pose of an object that is known a priori to the system. The novelty of the algorithm lies in how gaze planning and sequential estimation are used to resolve what is an inherently ambiguous task. It is also computationally efficient and robust, largely through the use of appearance-based techniques and a sequential Bayes estimation procedure.

The apparatus used in our experiments is shown in Figure 1 and consists of a monochrome television camera mounted on the end effector of a six degree-of-freedom gantry robot. Using this set-up, we can generate arbitrary trajectories through a workspace of approximately $1m^3$. As the camera moves, time-varying intensity patterns are induced on the camera retina and captured as a sequence of television images. The recovery of the velocity field associated with these patterns, the optical flow field, is a longstanding problem in the computer vision literature [16, 11, 3, 17, 1, 13] because of its relationship to the motion and structure of objects in the scene [12]. Over the years a number of algorithms have been developed for the recovery of optical flow from image sequences and in this paper we use the one described in [4] because it can be made to run in real-time. Unfortunately the recovery of structure and motion from optical flow estimates is still an open problem in the literature as it is fundamentally ill-posed.

In this paper, the environment is assumed to be stationary and the observer mobile so that the flow velocity is a function of observer motion, object shape, and camera imaging parameters. Our strategy will be to factor out of the optical flow field a signature related to shape by exploiting a priori knowledge of motion and imaging parameters learned through training. Others have demonstrated that this factoring problem can be solved in the presence of suitable a priori constraints [2, 19, 5], usually for the case of a stationary observer where the task is to identify the motion of a moving object. Here, we are attempting to recognize the object itself by driving the camera through a sequence of trajectories with the purpose of extracting a component of the flow related to shape (i.e. a shape signature). We use the signature to compute, for each object in the database, the probability that the signature was

Figure 1: Experimental Set-up.

generated by a given object. Because the problem is ill-conditioned, it is expected that several objects could give rise to the same signature. However, this is not expected to be the case over a sequence of observations, particularly if viewpoints are chosen that maximize the discriminability between objects.

We use principal components analysis (PCA) [15, 18] to compress the set of optical flow estimates obtained during training and allow representation in a lower dimensional space. On-line, optical flow estimates computed from motion about the unknown object are projected onto the PCA basis and Bayesian analysis is applied to compute the support for the different object hypotheses in terms of conditional probability distributions. For brevity we refer to the latter as *belief distributions*. This analysis can be taken one step further by updating the belief distribution for each hypothesis as new data are acquired by applying Bayesian chaining. As we shall demonstrate in later in Section 4, accumulating evidence in this manner generally leads to an unambiguous assertion in a short sequence of steps provided that motions are controlled so as to be reasonably approximated by the training set and ambiguous viewpoints are avoided. Furthermore the computational load associated with these operations is minimal relative to optical flow estimation.

These considerations lead naturally to the questions of how best to store information relating ambiguity to camera (viewing) position, i.e. *entropy maps*, and how to use them for gaze planning (navigation). During training, each object is placed at the origin of a tessellated viewsphere and sampled by sweeping each facet with a pair of short curvilinear arcs. The optical flow fields induced by these motions, which form the training set, are also used to compute a belief distribution for each coordinate of the viewsphere. In Section 3.1, we use the concept of Shannon entropy to define a measure of ambiguity for these belief distributions. The resulting *entropy map* is the parameterization of this measure in viewsphere coordinates. The map then serves as the basis for our gaze planning system. Like others [8, 7] the system operates by choosing locations that maximize information gain, reducing the number of observations required to make a confident assertion in the process. The key difference in this work is that we maximize the a priori information available, by building the entropy maps *off-line* and use them to guide the on-line navigation, further contributing to the computational efficiency of the method.

The remainder of this paper is as follows. We begin in Section 2 with a description of how optical flow signatures are used for recognition, leading to the determination of belief distributions corresponding to on-line measurements. This section also describes how these distributions are sequentially updated using Bayesian chaining. Section 3 describes the procedure for computing the entropy map and a navigation strategy based on it for gathering new data. Experimental results are presented in Section 4 which show how the strategy performs overall in correctly recognizing objects from their optical flow signatures. It also compares the entropy-driven approach to naive exploration. Finally Section 5 concludes with a brief discussion and pointers to future work.

2 Object Recognition Based on Optical Flow Images

We begin by considering how the optical flow fields computed from camera motion can be used to identify objects and describe a Bayesian strategy for associating probabilities to measurements, where probabilities relate the support for the different object hypotheses.

2.1 Why Optical Flow?

As a suitable input for the recognition system, we seek a reliable measure related to object shape in conjunction with an appearance-based strategy as in Nayar et al. [15] and Turk et al. [18]. Differential properties, such as optical flow images, offer some advantages with respect to minimizing sensitivity to illumination variations and background conditions (in contrast with standard appearance-based approaches). For this reason we have chosen optical flow as the input to our recognition system.

Now the problems associated with extracting structural information from flow images are well known. The difficulty in the task lies in the fact that the shape of the object, the relative motion between camera and object, as well as camera geometry are confounded in the resulting flow pattern. Our current goal is not to extract detailed structural information, but rather to extract features from the flow images that are repeatable signatures of object shape. In this work, we make the following assumptions:

1. *Camera Constraints.* Camera to object distances are bounded and scaled orthographic projection is assumed.

2. *Motion Constraints.* The same motion model can be used to account for an object moving about a fixed observer provided that rotations are limited to axes that are approximately parallel to the image plane.

3. *Motion Decomposition.* The trajectory of an observer moving through a stationary environment can be decomposed into a sequence of short, curvilinear segments. This motion model can be guaranteed in the case of an active vision system or mobile robot equipped with a suitable tracking system.

We claim that the above assumptions are not overly restrictive and can account for a reasonably wide range of viewing situations. A mobile agent can then control the sensor trajectories and camera positions so that the resulting flow patterns associated with different objects can be *learned* by the training procedure off-line. Further, by carefully choosing viewpoints that minimize ambiguity and by accumulating evidence over time, further robustness to flow estimation errors and confounding signals is achieved.

2.2 Bayesian Recognition

An optical flow algorithm is applied to the raw image sequence acquired by the camera yielding a second set of images that encode the optical flow. We refer to the latter simply as optical flow images, and for the remainder of the paper "image" should be taken to mean "optical flow image." In fact, we use the magnitude of the flow image denoted by $\mathbf{x}$ as the input to the system. For each of the image in this sequence, the recognition strategy computes a degree of likelihood in matches with each of the objects in the database. This leads to the formulation of a Bayesian recognition strategy whose goal is to represent the posterior beliefs over the entire set of n object hypotheses, $\{O_i\}$ where $i = 1 \dots n$, given a single flow image, $\mathbf{x}$, by a posterior probability distribution of the form $P(O|\mathbf{x})$, with discrete (conditional) probability density function $p(O_i|\mathbf{x})|_{i=1\dots n}$. As the sequence of such images is presented to the system, the system will then be able to gather evidence in the various hypotheses over time.

During the off-line (training) phase the camera is moved about the objects of interest along trajectories that reflect the expected on-line motions. Each coordinate of a tessellated viewsphere surrounding an object is sampled with a sequence of short, curvilinear sweeps along different directions. The resulting set of flow images, acquired from all viewpoints for each object, is used to determine a basis for representation using Principal Components Analysis (PCA) [18]. In this fashion, an appearance manifold for motion is built, with the hypothesis that on-line recognition from a larger set of motions will be possible. A representation for each object is then constructed by projecting its corresponding flow image vectors $\mathbf{x}_j$ onto this lower dimensional basis. The projected vectors $\mathbf{m}_j$ are subsequently parameterized by a multivariate normal distribution with density $p(\mathbf{m}|O_i)$. This represents the physical theory predicting possible variations in parameters given each object in the database.

On-line, an image, $\mathbf{x}$, corresponding to the unknown object is projected onto the basis determined during training, resulting in the parametric description $\mathbf{m}_x$. Using standard Bayesian techniques to determine the data support for each object hypothesis gives:

$$p(O_i|\mathbf{x}) = \frac{1}{K}\, p(\mathbf{m}_x|O_i)\, p(O_i), \quad 1 \dots n \qquad (1)$$

where $p(O_i)$ defines the prior probability for each object hypothesis, O_i, $p(\mathbf{m}_x|O_i)$ is the multivariate normal distribution derived during training evaluated at the location in space defined by $\mathbf{m}_x$, and K is the normalization constant such that:

$$K = \sum_{j=1}^{n} p(\mathbf{m}_x|O_j)\, p(O_j). \qquad (2)$$

The result is a discrete conditional probability distribution describing the belief in each of the models in the database, given the flow data.

However, recognition from a single optical flow image can be ambiguous and even erroneous. We therefore formulate the problem as a sequential estimation problem, where a more robust solution is attained by accumulating evidence in the various object hypotheses over time, as each of the flow images in the sequence is gathered by a mobile agent. This is accomplished efficiently at the level of the probabilities, by using a Bayesian chaining strategy that assigns the posterior probabilities at time t, $p(O_i|\mathbf{x}_t)$, as the priors at time $t+1$:

$$p(O_i|\mathbf{x}_{t+1}) = \frac{1}{K}\, p(O_i|\mathbf{x}_t)\, p(\mathbf{m}_{xt+1}|O_i), \quad 1 \dots n \quad (3)$$

where $\mathbf{x}_t$ is defined as the data set at time t, and $\mathbf{m}_{xt+1}$ is the parametric description of the measured flow, $\mathbf{m}_x$, at time $t+1$.

3 Navigation Based on Entropy Maps

3.1 Entropy Maps

With the recognition strategy in place, a mobile agent can move around a scene, gathering evidence in the various object hypotheses until a satisfactory confidence level is attained. As the object can be used as a landmark for global position estimation, it is essential that the strategy minimize the chances of the system of arriving at an incorrect recognition result. Furthermore, as resources are limited, the system needs to converge to a solution in a short number of steps.

We propose a strategy that takes maximal advantage of a priori information available in order to attain a fast and reliable on-line solution. Specifically, we propose building *entropy maps* off-line during training to relate recognition ambiguity to viewing position. Once a map is built for each object in the database, the system can store the locations that are maximally informative in terms of disambiguating between the objects in the database. This information can then be made available to the agent on-line to aid in recognition.

This leads to the problem of building these maps in practice. We wish to obtain a measure that predicts the likelihood of ambiguous recognition results as a function of viewing position. A suitable measure is defined in terms of the Shannon entropy [9],

$$H(P(O|\mathbf{x})) = \sum_i p(O_i|\mathbf{x}) \; log \; \frac{1}{p(O_i|\mathbf{x})}, \qquad (4)$$

which is a measure of the ambiguity of the posterior distribution produced by a recognition experiment. Higher entropies reflect greater ambiguity.

Entropy maps are built off-line, for each object in the database as follows:

1. During training, each optical flow measurement, $\mathbf{x}$, is stored along with its coordinates of acquisition on the viewsphere which, for this type of measurement, refers to three parameters: latitude, longitude and relative angle of motion between camera and object.

2. Recognition is then performed on each training measurement, resulting in the association of the posterior distribution, $P(O|\mathbf{x})$, to each coordinate.

3. The entropy for each measurement, $H(P(O|\mathbf{x}))$ is computed and stored at its associated coordinate. In this context, this implies that several entropy values are stored at every (latitude, longitude) position, each associated with a different relative motion.

Prior to using the entropy map for navigation purposes, further processing is required, namely in the enforcement of particular smoothness constraints which are essential for stability. For example, a gaze planner using the map to determine minimal entropy viewing positions would seek to avoid locations corresponding to singularities or discontinuities in the entropy field. Slight errors in positioning (or equivalently in determining the pose of the entropy map relative to the data acquired) could result in sampling at precisely the *wrong* locations. These constraints are made explicit by applying the following non-linear smoothing operator,

$$H(P(O|\mathbf{x}_i)) = \frac{\sum_j cos(\theta_{ij}) \times H(P(O|\mathbf{x}_j))}{\sum_j cos(\theta_{ij})}, \qquad (5)$$

where $\mathbf{x}_i$ is the data vector gathered at viewsphere location i of the operator, $\mathbf{x}_j$ are the points in the local neighborhood indexed by j, and θ_{ij} the angle subtended at the center between $\mathbf{x}_i$ and $\mathbf{x}_j$. The minimal entropy location on this map will correspond to an optimal location which is stable with respect to localization errors.

The resulting entropy map can be very informative in the context of planning gaze for object recognition. It provides a *quantitative* prediction of the level of difficulty of recognizing each object in an on-line experiment. In contrast to human-generated aspect graphs, e.g., [10, 14, 6]), by linking location and discriminability using entropy maps, a set of such characteristic views can be automatically generated off-line.

Two viewpoints of an entropy map, one "good" and one "bad", can be found in Figure 2. In Figure 2(a), one can see the raw entropy values at these locations, and the corresponding smoothed entropy maps can be found in Figure 2(b). Each tile corresponds to a particular camera view of the object at the origin. The tiles are shaded in accordance with their entropy values: from low entropy (dark) to high entropy (light). Larger areas with no tiles indicate viewpoints that lead to a false identification, i.e. a maximum a posteriori result (MAP) indicating a different object. Only the best entropy result (among all those generated from different movements at that location) is shown at each location.

Examining the maps, one can see that their structure is such that areas that result in inter-class confusion are found in isolated patches on the viewsphere. In addition, the ambiguity increases in the areas surrounding the "worst" locations. The structure of the map lends itself to our navigation strategy as generally large patches are comprised of optimal viewpoints, in terms of high confidence in the correct model. The smoothed versions of the entropy map at the same location, as seen in Figure 2(b), illustrate both the best and worst locations (in terms of location and movement) for discrimination. Notice that the best location is maximally far from the areas of high confusion or strong belief in the wrong object. This leads to the hypothesis that moving towards this location will lead to a correct solution with relative safety.

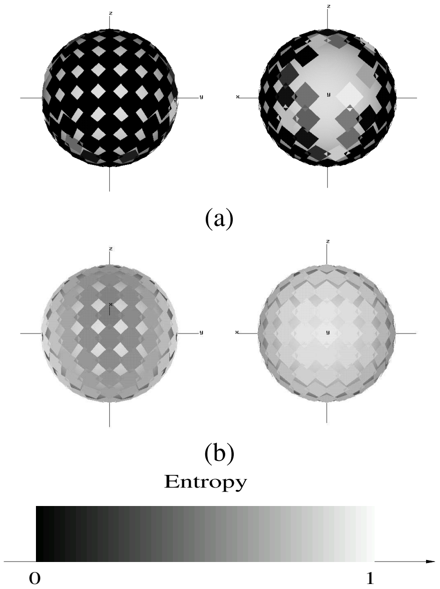


Figure 2: (a) Two Views of an Entropy Map, (b) Corresponding Smoothed Maps.

3.2 Using Entropy Maps to Plan Gaze

Two problems must be solved prior to planning gaze: 1) a particular map must be selected and 2) the pose of that map must be determined relative to the data acquired. As measurements are made on-line, the maximum a posteriori (MAP) solution corresponding to $p(O|\mathbf{x})$ is used to determine the most likely object hypothesis for the measured data $\mathbf{x}$. This estimate is subsequently used to select the entropy map to be used for planning the next best view. Of course the particular gaze planning strategy must be carefully structured to operate stably in these circumstances. More will be said about this shortly.

Pose can be estimated at minimal expense by retaining the location information along with the image measures acquired during training. For example, appearance-based methods can be used to index these measures using the data acquired on-line [15]. In fact, the implementation described in Section 2 already uses appearance-based techniques in the process of determining the likelihoods for the different object hypotheses. As such, the computational overhead of determining pose is minimal. Once the camera pose is established in the coordinates of the training viewsphere, it is straightforward to determine the relative transform taking the camera to the desired position within this frame (Figure 3, Steps 2–3). By applying this same transform to the current camera frame, the camera is positioned accordingly as shown in (Figure 3, Step 4).

The gaze planning strategy itself must be sufficiently robust to accommodate errors in pose determination and entropy map selection. Errors in the former are accommodated in part by the smoothing applied to the entropy map and a strategy that avoids placement in the vicinity of singularities and discontinuities. A partial solution to the selection problem is effected by choosing a next best view that minimizes the entropy on the most likely object hypothesis map. Over time the expectation is that confidence in an incorrectly chosen hypothesis will decrease as further evidence is uncovered.

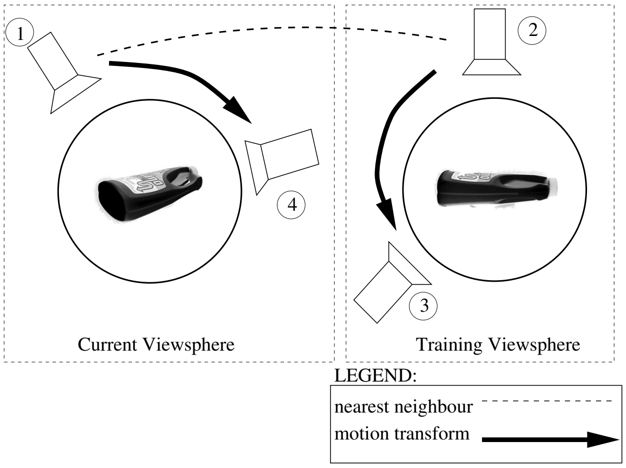


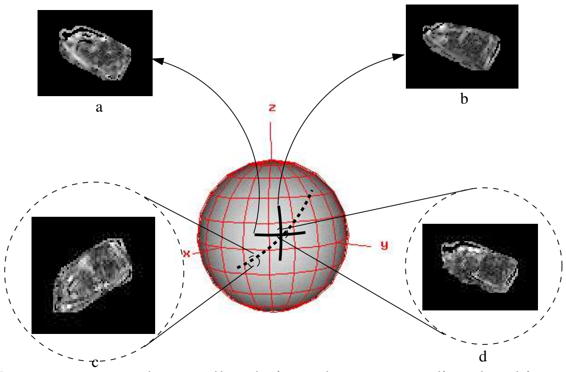
Figure 3: Navigation Strategy.

4 Experimental Results

The system described in this paper was applied to real, on-line recognition experiments, where the task of the mobile agent was to identify an object in a scene within a finite amount of steps. We will illustrate the system's ability to (a) build entropy maps off-line for each object in the database using optical flow images as inputs, (b) resolve recognition ambiguities resulting from tests with single flow images by accumulating evidence over time, and (c) plan the next gaze position and corresponding motion trajectory of a robot based on these maps during on-line experimentation. We will also show the system's superiority over random navigation techniques in terms of speed of convergence and recognition accuracy.

We begin with a description of our set-up. Off-line, training was performed on flow magnitude images of 15 household products (see Figure 4), by gathering images at equally spaced locations around a coarsely tessellated viewsphere. This was accomplished by placing each object on a rotary table within the working space of a gantry robot arm, at the origin of the viewsphere. The robot arm was used to move latitudinally and to gather the motion sequences on the sphere, while the rotary table served to

Figure 4: Database of Objects.



Here, one can see the tessellated viewsphere surrounding the object of interest (eg. a liquid drano bottle). The perpendicular darker lines illustrate the arc-like trajectories comprising the motion basis for training at a particular location on the viewsphere. Examples of the resulting flow images at that location can be found in (a) and (b). The dotted lines indicate an example of a trajectory taken during a navigation experiment. Examples of 2 flow images resulting from gathering images along the trajectory can be found in (c) and (d). Notice that the flow image in (d) resembles both (a) and (b) despite having been computed along a trajectory not trained on.

Figure 5: Motion basis and test trajectory along viewsphere. Liquid drano bottle example: (a)–(b) Flow images along basis used for training. (c)–(d) Flow images along trajectory.

move the object longitudinally. Specifically, at each position on the viewsphere, the robot arm moved along a horizontal and along a vertical arc at fixed distances from the object. A CCD camera, mounted on its end-effector (see Figure 1), gathered three images in sequence along each trajectory from which optical flow was computed (using a strategy as in [4]). This served to create a local basis for flow. The expectation was that other on-line motions could be inferred from this basis. Speed normalization was achieved by normalizing the optical flow magnitudes to lie between 0 and 255. Flow was used to localize the object of interest within the images.

Figure 5 shows an example of a training set and a typical test sequence (and illustrates their corresponding flow images) on the viewsphere about an object of interest. This figure illustrates the system's capability at generalizing motions based on a small training set, such that it can successfully identify objects based on novel motions (in this case along a diagonal arc direction). In this case, a 20 dimensional basis for flow, built using standard PCA techniques [18], was sufficient to represent the flow images.

Off-line, entropy maps are built from the flow images gathered. The hypothesis is that the structure of the maps lend themselves to on-line navigation experiments based on them. Figure 6 shows images of an object from the database and the corresponding smoothed entropy map taken from two different camera viewpoints. The system located the best viewpoint for identification of this object, one that was maximally far from the most ambiguous ones. The structure of the map illustrates a continuous degradation from good views to bad ones. Notice that the entropy maps match an intuitive notion of viewpoint ambiguity.

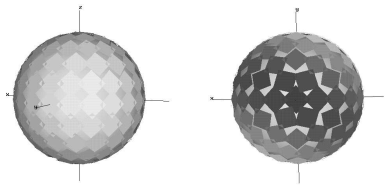
Next, the off-line entropy maps were used in a series of on-line recognition experiments that applied the gaze planning strategy to a real robot system. The test was to see whether the maps would guide the sensor to desirable locations for recognition, and to see whether convergence to the correct hypothesis would be more accurate than if one were to use a random sequential navigation strategy.

On-line, the system used the same physical set-up as in training. Starting from a random location, the gantry arm was moved along a local, curvilinear trajectory on the viewsphere according to the proposed navigation strategy. At each coordinate sampled along this path, a local flow measurement was made by sweeping the arm along two short curvilinear arcs (the cross shown in Figure 5). Recognition was then performed using the corresponding optical flow generated by this local motion. As a precautionary measure, an initialization procedure was performed to ensure that the system did not start out from a local minimum in the entropy map. Empirically, it was found that this lead to an improvement in the results.

The system iterated until the entropy reached an arbitrarily small convergence value (e.g. 0.01 was chosen). Five hundred such tests were performed on all the objects in the database in order to examine average performances. The percentage of correct MAP recognition results at convergence using the described navigation strategy can be found in Figure 7. Here, one can see that the system converges to the correct solution in most cases. In addition, convergence occurs in less than 3 iterations on average.

A similar experiment to the one above was performed,

(a)



(b)

Images of the toothpaste tube and the corresponding smoothed entropy map are seen at two locations. The system chose the right view as the most informative (seen with darker shading on the map), and the left view as a relatively bad one (lighter tiles). This corresponds to an intuitive notion of "good" and "bad" views.

Figure 6: (a) Images of a Toothpaste Tube, (b) Smoothed Entropy Maps.

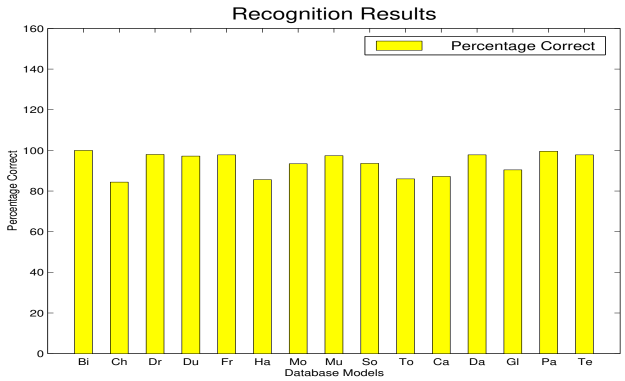
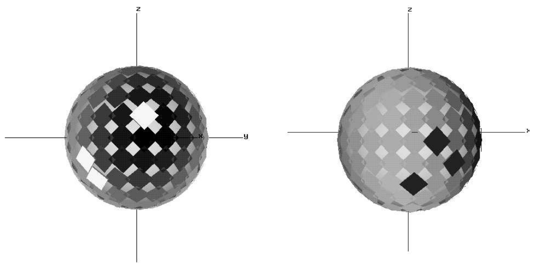


Figure 7: Recognition Results at Convergence.



(a) (b)

Above we can see the results of 2 navigation experiments: (a) using our strategy (in white), and (b) using a random one (in black) superimposed onto the entropy map of the bread roll. In (a), notice that the system converges to the correct solution in the desired neighborhood (seen darker tiles). In (b), the system converged to the wrong solution, having gathered data from highly ambiguous viewpoints (seen in lighter shades).

Figure 8: (a) Entropy Map Navigation, (b) Random Navigation.

this time using a random navigation strategy. The results indicate that both the proposed strategy and the random one performed quite well, in terms of recognition results and quick convergence. This is mostly due to the strength of the Bayesian chaining algorithm at eliminating false hypotheses quickly. However, it was found that navigating based on entropy maps outperformed the random approach, particularly in cases where the system started from ambiguous viewpoints.

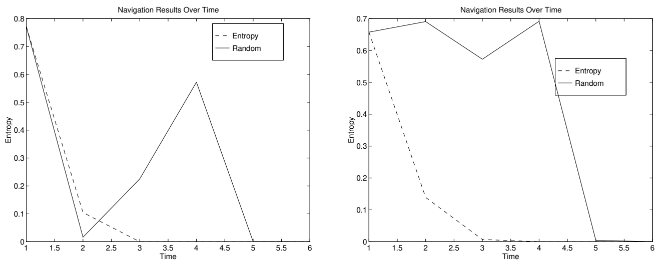
Figure 8 shows an example comparing the two strategies when initialized from the same high entropy location. Figure 8(a) illustrates the results of a navigation sequence that converges to the correct solution in three iterations, superimposed onto the entropy map of the object of interest. Notice that the system leads the sensor near to the entropy map minimum of the correct hypothesis Figure 8(b) shows the result of a random sequence, where convergence to the wrong model, reached in five iterations, was due to several images taken at locations where strong belief in the wrong model was present.

Figure 9 illustrates a comparison of navigation results (using both strategies) starting at high entropy locations in the cases of the duck and the bread roll (chala). One can see that the proposed strategy converges faster than the random strategy in both cases. Notice that, in Figure 9(a), the random strategy caused the sensor to move to a "bad" local minimum (low entropy, wrong model case) at iteration 2. A formal comparison of the performance of the two approaches when started from ambiguous viewpoints can be seen in Figure 10. Only the results where a difference exists are plotted. For the most part, our strategy wins over the random one.

Empirical evidence indicates the benefits of using an off-line entropy minimization strategy, over on-line methods, in leading the system towards the *global* entropy minimum. In cases where the sensor began with a high confidence in the wrong model, the entropy may increase with each step as it leaves the local minimum, before converging to a global low entropy state. On-line entropy minimization strategies converge to a *local* entropy minimum, even if it belongs to a false assertion.

5 Discussion and Conclusions

The experimental results would appear to confirm our central hypothesis that the optical flow patterns resulting from the motion of a mobile observer can be used for the purpose of object recognition. What makes this approach tractable is that a mobile observer can constrain its motion so that the patterns induced on the camera bear some re-



(a) (b)

Figure 9: Navigation Results Over Time for (a) duck, (b) bread roll.

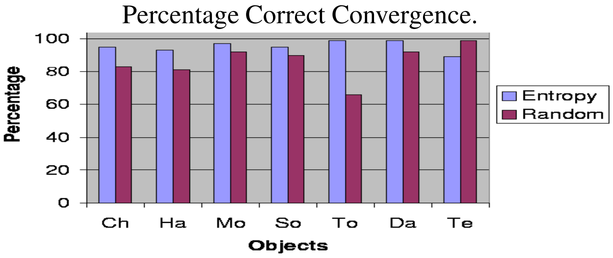


Figure 10: Comparison of Two Strategies.

semblance to those of the training set. The entropy-based gaze planning strategy further ensures that vantage points are chosen where the patterns best serve to discriminate one object from another. Finally, rather than make an instantaneous assertion as to the identity of the unknown object based on maximum likelihood, we instead accumulate evidence in the form of belief distributions over several measurements. The latter acts as a temporal filter by maximizing the probability of the object hypothesis that is most consistent with the data over time.

An interesting question, and a direction for future research, concerns the degree to which the basis obtained through training generalizes to a much broader range of motions. We suspect that feedback afforded by the gaze-planning strategy and the use of temporal filtering should permit a fairly broad range of generalization, but this will have to be verified experimentally. Further improvements to the algorithm, e.g., a better method for estimating pose, a navigation strategy that can encompass *several* competing hypotheses, and the exploitation of entropy measurements computed on-line, would also improve the robustness of the algorithm. Finally we note that this approach should also be adaptable to other sensory modalities which encode the structure of the local environment.

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, (2):283–310, 1989.

[2] G. Aviv. Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field. *PAMI*, II(5):477–489, May 1989.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[4] S. M. Benoit and F. P. Ferrie. Monocular optical flow for real-time vision systems. In *Proceedings of the 13th International Conference on Pattern Recognition*, pages 864–868, Vienna, Austria, 25–30 Aug. 1996.

[5] M. J. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parametrized models of image motion. *Int. Journal of Computer Vision*, 25(1):23–48, 1997. Also found in Xerox PARC, Techinical Report SPL-95-020.

[6] K. Bowyer and C. Dyer. Aspect graphs: An introduction and survey of recent results. In *Close Range Photogrammetry Meets Machine Vision*, volume 1395, pages 200–208. SPIE, 1990.

[7] W. Burgard, D. Fox, and S. Thrun. Active mobile robot localization. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, Nagoya, Japan, August 23-29 1997.

[8] F. G. Callari and F. P. Ferrie. Active recognition: Using uncertainty to reduce ambiguity. In *Proceedings of the 13th International Conference on Pattern Recognition*, pages 925–929, Vienna, Austria, 25–30 Aug. 1996. International Association for Pattern Recognition, IEEE-CS.

[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley & Sons, 1991.

[10] D. Eggert, K. Bowyer, C. Dyer, H. Christensen, and D. Goldgof. The scale space aspect graph. In *Proceedings, Conference on Computer Vision and Pattern Recognition*, pages 335–340, Champaign, Il., June 15-18 1992. IEEE.

[11] D. Fleet. *Measurement of Image Velocity*. Kluwer Academic Publishers, Norwell, MA, 1992.

[12] J. Gibson. *The Perception of the Visual World*. Houghton Mifflin Company, Boston, 1950.

[13] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[14] D. Kriegman and J. Ponce. Computing exact aspect graphs of curved objects: Solids of revolution. In *PROC. of IEEE Workshop on the Interpretation of 3-D Scenes*, pages 116–122, Austin, Texas, November 27-29 1989. IEEE.

[15] S. K. Nayar, H. Murase, and S. A. Nene. *Parametric Appearance Representation in Early Visual Learning*, chapter 6. Oxford University Press, February 1996.

[16] K. Prazdny. Egomotion and relative depthmap from optical flow. *Biological Cybernetics*, 36:87–102, 1980.

[17] P. Singh, A. amd Allen. Image -flow computation: An estimation-theoretic framework and a unified perspective. *CVGIP: Image Understanding*, 56:152–177, 1992.

[18] M. Turk and A. Pentland. Eigenfaces for recognition. *CogNeuro*, 3(1):71–96, 1991.

[19] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *PAMI*, II(5):490–498, May 1989.