

Visual Attention-based Robot Navigation using Information Sampling

Niall Winters

José Santos-Victor

Computer Vision and Robotics Group,
Department of Computer Science,
University of Dublin, Trinity College,
Dublin 2 - Ireland.
Niall.Winters@cs.tcd.ie

Instituto de Sistemas e Robótica,
Instituto Superior Técnico,
Av. Rovisco Pais, 1,
1049-001 Lisboa - Portugal.
jasv@isr.ist.utl.pt

Abstract

This paper is concerned with focusing the attention of an autonomous mobile robot so it can fully utilize its computation resources in visually navigating through an indoor environment. Our sole input data are omnidirectional images obtained from a catadioptric panoramic camera. We present a statistical method termed Information Sampling for selecting the most interesting data from this set. These data could be a single pixel or a number scattered throughout an image. The main problem addressed is how to determine which image data points represent highly attentive regions of the environment. Once determined, we show how to rank the data from most to least significant and how, using only the best data, Information Sampling can be applied to determine the qualitative position of the robot in a topological context. Results are detailed in the paper.

1 Introduction

We address the question of how to obtain the most informative data when an autonomous mobile robot is visually guiding itself through a simple indoor environment. The aim here to identify and focus upon *highly distinctive* regions within the environment, which the mobile robot should memorize for future reference. In focusing upon attentive regions the robot makes maximum use of its limited computational resources.

Traditionally, research into solving the robot navigation problem concentrated almost exclusively on geometric solutions. While many worthwhile experiments were undertaken, real world solutions were of-

ten hampered by time constraints. In an effort to overcome this drawback, the past few years has seen a new approach emerge which seeks to turn the navigation problem from the domain of reconstruction to that of pattern classification or, as it has become known an *appearance-based* solution.

1.1 Related Work

Perhaps one of the earliest examples of the use of appearance-based methods for navigation was developed by Hong [6] et al. The goal was for a robot to home to a given destination. Hancock and Judd [4] developed “Ratbot”, where localization was achieved by simply matching vertical bars from runtime images to those acquired a priori. A View Sequence of images for navigation was proposed by Matsumoto et al. [13]. Zheng [22] presented a system which moved along a given route under human guidance and autonomously memorized a side-view of that route. These data were then used as a basis for route recognition and image matching was performed in a coarse-to-fine manner using dynamic programming. Ishiguro [7] used the frequency domain of the Fourier Transform as an image-based memory of the environment. Unfortunately, if repetitive structures need to be distinguished, this method failed. Maeda et al. [12] used the parametric eigenspace approach to image matching [14]. They noted that if one takes a single image from a standard camera, multiple matches could be obtained in environments where similar images appear a distance apart. Thus, robot position cannot be reliably determined. Their proposed solution to this problem is to employ an active vision strategy and take another image close to the current one by moving the robot, or the camera, and projecting this image into

the eigenspace. The disadvantage here is that when building the eigenspace multiple images have to be taken at each location, thus increasing the complexity of the approach.

In order to compress large amounts of image data, appearance-based systems are usually (but as can be seen from the related work, not exclusively) built using Principal Component Analysis. Construction of such a system involves computing the eigenvectors (sometimes called eigenimages) of an a priori set of images. The variance of this set is captured by its first few eigenvectors. This low dimensional subspace [14], also known as an eigenspace, forms an orthonormal basis into which each image from the a priori set is projected. Once this eigenspace has been built, real time recognition of an unknown image is achieved by projecting it into the eigenspace and using a simple distance measure to find its closest match to the previously projected points.

It is clear that previous research undertaken in this area represented a “global approach” to the localization problem, in that entire images were used for matching. No attempt was made to extract *attentive regions* from the input images. In the field of object recognition, several authors have noted problems with global matching approaches [15, 2]. Instead of projecting the entire image they proposed dividing each image into a number of smaller windows which they termed eigen windows. Eigenspace analysis was then applied to each window. Their basic idea was that even if a number of the windows were occluded, the remaining ones would contain enough information to perform the recognition task. A major problem with this approach is that it requires storage of a very large number of image windows. For example, if one had an a priori set of 1000 images of size 256×256 pixels, and each window was 8×8 pixels in size, then one would require 1,024 windows to represent an image or 1,024,000 to represent the entire a priori set. Clearly the chances of one window, acquired at runtime being matched to a number of images from the a priori set is high. This could be due, for example, to having many ambiguous regions within an image. Thus, it is highly desirable that only the most *effective* windows are selected from each acquired image, and that only these chosen windows be matched to the a priori set.

As a solution to this problem, Ohba and Ikeuchi propose using three criteria to eliminate the redundant windows, namely: detectability, uniqueness and reliability. Colin de Verdière and Crowley reformulate the problem as a question of whether to use the set of eigen windows selected by a particular interest opera-

tor or to use those windows selected from a predefined grid.

Interest points are local features where the signal changes two-dimensionally. Schmid and Mohr [18] used a Harris detector [5] to determine where to compute local grayvalue invariants in addressing the problem of image retrieval from a large database. Jugesur and Dudek [8] use a symmetry based context free attention operator from [16] to detect windows that contain a number of interest points above a certain threshold. Unfortunately, for such an approach to be effective, the images are required to be highly textured. Our a priori images do not exhibit such a property, as they were acquired along a corridor where the environment consists of plain white walls and brown doors. Additionally, with interest operators stability is hard to guarantee when changes in illumination occur.

As an alternative to using interest operators, one may use windows chosen from a pre-defined grid [2]. The first stage of this approach involves projecting *all* of the eigen windows into the eigenspace. Since an image will contain a number of windows, an image is represented in the eigenspace as a surface and a set of image by a set of surfaces. Naturally, on projection of an eigen window many matches will occur. Thus, suppression of redundant windows is required. This is usually achieved by noting that a search for the closest point in the eigenspace produces too many matches. Alternatively, suppression can occur at the training stage by noting that a redundant window is one that is projected many times. Nevertheless, this approach still requires enough space and computational power to store and search all of the eigen windows.

1.2 Our Approach

Our approach to this problem is somewhat different from those outlined above and is based on a method by Rendas and Perrone [17]. We noticed that the above approaches do not make use of the inherent information available from the a priori set of images. In the case of [15, 2] all the eigen windows from entire images are first collected, from which the interesting data must be found. In the case of [18, 8] their approaches require extraction of features from images *before* being able to proceed and solve their problem.

In contrast, our approach termed *Information Sampling* [21] selects the most interesting data from a set of images without using eigenspace analysis or needing to apply interest operators. Theoretically, it can be applied on a *pixel-by-pixel basis to any type of image*, as outlined in Section 3.1. In this paper, for com-

putational reasons, we use windows instead of pixels, extracted from omnidirectional images. Essentially, we can reconstruct an image using only the data selected by *Information Sampling* and then minimize the error associated with this reconstruction. We can rank the selected data and choose how much of it we wish to utilize for matching. *Information Sampling* selects quality image data as distinct from other methods which rely on a quantity of data. It is only after the ranking stage that we employ a local eigenspace approach to perform the qualitative recognition of a robot’s position within an indoor environment, i.e. its position on a topological map.

The outline of this paper is as follows: Section 2 briefly presents the Topological Navigation methodology. Section 3 presents the *Information Sampling* method and underlying statistics. In Section 4, we show how to use a local eigenspace approach for robot navigation while in Section 5, we present our experimental results and discuss the complexity of our method. Finally, in Section 6 we conclude and give the future directions of our research.

2 Topological Navigation

Topological Navigation is the mode used to travel long distances in the environment and does not require accurate control of the robot position along a path. Typically, it corresponds to commands such as “follow the corridor” or “turn left”. To be able to adopt this navigation strategy we need a suitable representation, since the traditional metric mapping approach is extremely inadequate. We represent the environment as a *Topological Map*, described by a graph structure. *Nodes* in the graph correspond to recognizable scene sites (landmarks) where specific actions may be elicited, such as entering a door, turning left, etc. *Links* connecting nodes in the topological map correspond to regions where some environmental structure can be used to control the robot. It is interesting to note that since the robot only navigates between nodes, global errors do not accumulate [1].

Topological Maps have proved beneficial in many research works [10, 20]. Previously, some researchers [9, 11] chose to detect nodes in the environment by the use of sonar readings. Unfortunately, using this approach one tends to have to deal with situations where nodes are relatively indistinguishable due to the nature of the obtained data. In our approach, nodes are represented by *Attention Windows* (see Section 3.2) and links by sequences of such windows. We use Attention Windows (selected from omnidirectional



Figure 1: Left: the omnidirectional camera. Right: the camera mounted on the mobile robot.

images), projected into a local appearance space (see Section 4) as an implicit topological representation of the environment. Progression towards the goal is obtained by comparing the most interesting attention window from the current view with attention windows acquired *a priori*.

Topological maps are a parsimonious representation of the environment that can be easily scaled, simply by connecting various subgraphs. Often the Topological Map can be associated with some qualitative structural elements of the environment, that can be used to control locally the robot trajectory.

3 The Information Sampling Method

As previously noted, our approach requires the use of a priori image data. We wish to make it clear that our method is *independent* of image type. For the experiments outlined in this paper images acquired from an omnidirectional camera with a spherical mirror, built in-house at the Instituto de Sistemas e Robótica, Lisboa were used. This camera was mounted on a Labmate mobile platform and images were captured as it traversed through a corridor environment. The system is shown in Figure 1. Once the images were captured, we determined which regions contained the most interesting information by applying *Information Sampling*. As a first step in explaining this process, Section 3.1 outlines the procedure for reconstructing an image, given only a small amount of data.

3.1 Image Reconstruction

We assume that the images captured by the robot’s camera can be modeled as a random vector I , characterized by a Gaussian distribution with mean \bar{I} and covariance Σ_I :

$$I \sim \mathcal{N}(\bar{I}, \Sigma_I) = p(I)$$

Usually, one can take an ensemble of images of the environment $[I_1 \dots I_m]$, which can be utilized for computing \bar{I} and Σ_I , so that $p(I)$ can be computed *a priori*. When the robot is navigating, we assume that the observations, d , consist of a selection of (noisy) image pixels (or sub-regions), rather than the entire image. Accordingly, the observation model can be expressed as:

$$d = SI + \eta \quad (1)$$

where d stands for the observed data and the measurement noise η is assumed to follow a Gaussian distribution with zero mean and covariance Σ_n . We further assume that I and η are independent. The selection matrix, S , is composed of a series of ones and zeros, the ones corresponding to the data points extracted from an image. We select a number of pixels to test by moving the set of ones in the selection matrix.

Having prior knowledge of I , in the form of a statistical distribution, $p(I)$, the problem now consists of estimating the (entire) image based on a partial (noisy) observations of a few pixels, d . This problem can be formulated as a *Maximum a Posteriori* estimation of I . The posterior probability can be determined from Bayes rule as:

$$p(I|d) = \frac{p(d|I)p(I)}{p(d)} \quad (2)$$

where $p(d|I)$ is the likelihood of a pixel (or set of pixels) given a known image, I ; the prior distribution is denoted by $p(I)$ and is assumed to have been learnt *a priori*. With this information we calculate the maximum a posteriori estimate of an image, \hat{I}_{MAP} as follows:

$$\hat{I}_{MAP} = \arg \max_I p(I|d)$$

$$\hat{I}_{MAP} = (\Sigma_I^{-1} + S^T \Sigma_n^{-1} S)^{-1} (\Sigma_I^{-1} \bar{I} + S^T \Sigma_n^{-1} d) \quad (3)$$

Thus, \hat{I}_{MAP} is the reconstructed image obtained using the pixel (or set of pixels), d . Notice that by combining the prior image distribution with the statistical observation model, we can estimate the entire image based on the observation of a limited number of pixels.

3.2 Choosing the Best Data: Attention Windows

Once we have reconstructed an image using the selected data, we can compute the error associated with this reconstruction. The error covariance matrix, Σ_{error} is given by:

$$\Sigma_{error} = \text{Cov}(I - \hat{I}_{MAP}) = (\Sigma_I^{-1} + S^T \Sigma_n^{-1} S)^{-1} \quad (4)$$

Of course, the quality of the estimate, and the “size” of Σ_{error} depend not only on the observation noise, η but also on the observed image pixels, as described by the selection matrix S . Equation (4) quantifies the quality of an estimate obtained from using a particular set of image pixels. In theory, we can evaluate the *information content* of any individual image pixel or combinations of pixels, simply by selecting an appropriate selection matrix, S , and determining the associated Σ_{error} .

The problem could be formulated as an experiment design process, in which we look for the optimal selection matrix S^* that minimizes (in some sense) the error covariance matrix. If we take the determinant of Σ_{error} as an indication of the “size” of the error, the optimal selection of image pixels would be given by:

$$S^* = \arg \min_S \{ \det((\Sigma_I^{-1} + S^T \Sigma_n^{-1} S)^{-1}) \} \quad (5)$$

In practice, to avoid computing the inverse we define the following equivalent optimization problem in terms of a modified uncertainty metric, U :

$$U = -\log \{ \det((\Sigma_I^{-1} + S^T \Sigma_n^{-1} S)) \} \\ S^* = \arg \min_S U \quad (6)$$

So far, we have described *Information Sampling* as a process for (i) reconstructing an entire image from the observation of a few (noisy) pixels and (ii) determining the *most interesting* image pixels, S^* , in the sense that they convey the most information about the image set.

Unfortunately, determining S^* is computationally impractical since we would have to compute Σ_{error} for all possible combinations of pixels scattered throughout the image. Instead, we partition the image into non-overlapping square windows of $(l \times l)$ pixels. We term these regions *Attention Windows*, denoted by $\mathbf{w} = [\mathbf{w}_1 \dots \mathbf{w}_n]$.

By using equation (6), we can rank *Attention Windows* or combinations of such windows, in terms of their information content. Again, as searching all possible combinations of windows within the image to minimize equation (6) would be computationally intensive, so instead we use two sub-optimal (greedy) algorithms. These algorithms are described in Section 5.1.

Notice that the information criterion is based on the entire set of images and not, as with other methods, on an image-by-image basis. For instance, a highly textured image region would only be selected if it varied significantly from one image to the next.

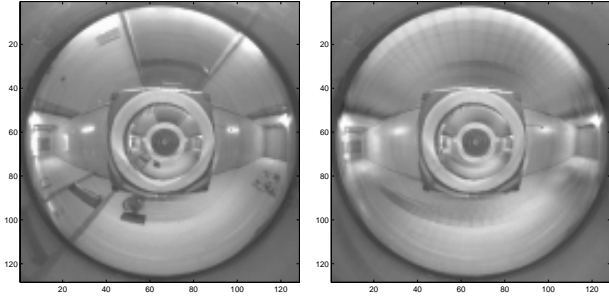


Figure 2: The input (left) and retrieved (right) omnidirectional images are very similar.

4 Local Appearance Space

We wish to use only the best Attention Windows, $\mathbf{w} = [\mathbf{w}_1, \dots, \mathbf{w}_n]$ obtained from *Information Sampling* to visually guide a mobile robot through a simple indoor environment. While conceivable this could be achieved by matching the reconstructed image, \hat{I}_{MAP} to the set of omnidirectional images, this would be computationally expensive.

Instead, to determine the robot position, this matching can be achieved in real time by reducing the amount of data using Principal Component Analysis (PCA). When using PCA, it is usual for entire omnidirectional images to be utilized in building the eigenspace. They are also used in the projection, i.e. a global approach. This was the approach taken in our previous work [3, 19, 20]. Figure 2 shows an image acquired by the robot on a run down a corridor and the reconstruction of the closest image from the a priori set. As one can see, the correct qualitative position of the robot within its environment was calculated.

We wish to improve upon this method by projecting *only* the most interesting information obtained by *Information Sampling* into a local appearance space. In this way we significantly reduce the number of projected windows, thus immediately reducing the level of possible ambiguity. Additionally we reduce even further the amount of data used for matching. The local appearance space has an orthonormal basis of eigenvectors of size $(l^2 \times 1)$, where l is the length of the side of a square Attention Window.

Following the standard PCA approach, and using Singular Value Decomposition (SVD) we can determine the eigenvectors (sometimes called eigen windows), \mathbf{e}_j , and eigenvalues, λ_j , of the covariance matrix Σ_{local} of the windows selected from the set of omnidirectional images, \mathbf{I}^w . We denote the selected windows from each image, $\mathbf{w} = [\mathbf{w}_1 \dots \mathbf{w}_m]$. We denote

α_j as the vector of co-efficients obtained by projecting each window from \mathbf{I}^w into the local eigenspace. We can reconstruct an entire unknown *image*, \hat{A}_{MAP} by replacing d in equation (3) by w_k and reconstruct its associated *window*, \hat{g} from PCA as follows:

$$\hat{g} = [e_1 \dots e_q] \alpha_j \quad (7)$$

5 Experimental Results

Information Sampling has been tested on images acquired by a mobile robot in a simple indoor office environment. Processing was carried out off-line on a Celeron 333MHz PC using Matlab. Our goal was to show that by using *Information Sampling* to focus upon attentive regions within the environment, effective navigation is possible. In the near term on-line experimentation shall be undertaken.

The a priori set of ninety omnidirectional images were obtained every 50 cm and ordered according to the direction of motion of a Labmate mobile platform. Each image was acquired at a resolution of 768×576 pixels using a Tekram acquisition board. Once each image was acquired it was filtered and subsampled to an image resolution of 16×16 pixels. The reason for such a small image size relates to the complexity of determining the error covariance matrix, Σ_{error} in equation (4). If one wished to use image sizes of 128×128 pixels, the computation of Σ_{error} would require the calculation of a matrix of $16,384 \times 16,384$ elements in size, a prohibitively large computation.

Each Attention Window was chosen to be 4×4 pixels in size, thus giving 16 non-overlapping Attention Windows per omnidirectional image. For image reconstruction using equation (3), these windows were of adequate size. As described in Section 4, we use a local appearance space to determine the qualitative position of a mobile robot within its environment. Additional positioning experiments were undertaken using eigen windows of 32×32 pixels in size, extracted from omnidirectional images of 128×128 pixels in size. In order to locate the 32×32 pixel sized Attention Windows, we used the ratio: $16:256 = 1,024:16,384$.

5.1 Ranking the Attention Windows

We tested two algorithms to rank the Attention Windows: Combinatorial Search and Simple Search.

1. **Combinatorial Search** We first search for the best *Attention Window*. Then, the search for the next best window is made keeping the first

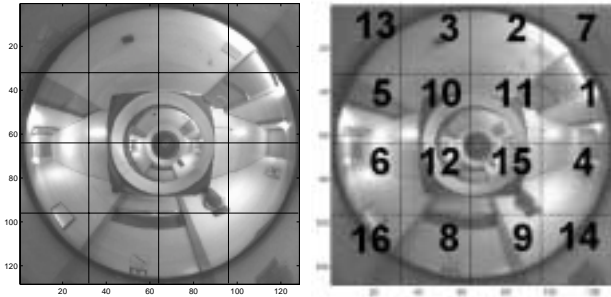


Figure 3: Left: The 16 non-overlapping Attention Windows. Right: Those windows ranked according to the amount of information they contain using Simple Search.

window fixed, thus locating the best *pair* of windows. As the method continues it determines the best triplet of windows, etc. If we denote n as the number of windows within an image, this method requires the evaluation of equation (6), $n!$ times. The method automatically groups the Attention Windows into a single window, a pair of windows, a triplet of windows etc.

2. Simple Search This is a quicker search algorithm. We rank the best information windows *independently*. In this case, equation (6) has only to be evaluated n times. As distinct from the combinatorial method, if we wish to group the best (single, pair, triplet etc. of) windows we must do it manually based on the initial ranking.

These methods were chosen given the complexity of searching for the optimal solution. Figure 3 (left) shows the Attention Windows available for selection and (right) these Attention Windows, *individually ranked* from the most (number 1) to the least informative (number 16) using simple search.

As can be seen from any of the omnidirectional images in this paper, the robot is in the centre of each image. Any information window which contains the robot is not a discriminating one and so it follows that such a window should have a lower ranking. As shown in Figure 3 (right), this proves to be the case: the four Attention Windows which contain the robot are ranked from numbers ten to fifteen. Additionally, the four windows at the periphery of the image also have a low ranking, since they only contain a portion of the omnidirectional image itself.

It should be noted that the corridor in which the a priori set of images were acquired has a number of offices on one side (the top half of the omnidirectional

images) and only a single door and notice-board on the other (the bottom half of the omnidirectional images). Thus, as the robot travels down the corridor more information change occurs in the top half of the omnidirectional images. Again, this is borne out by the window ranking, where the three highest ranking Attention Windows are all in the top half of the omnidirectional image.

Figure 4 shows the graphs of the information windows ranked using simple search (top) and combinatorial search (bottom). In both cases, the x -axis corresponds to the window ranking, from first to sixteenth and the y -axis corresponds to the uncertainty metric, U calculated using equation (6). The numbers along the graph line correspond to the 16 non-overlapping Attention Windows per omnidirectional image. For example, using simple search the left graph tells us that the eighth Attention Window exhibits the lowest uncertainty value and so is individually ranked in first position, while the third window, having a higher uncertainty value, is individually ranked in second position etc.

Using combinatorial search the right graph tells us that the eighth window is ranked in first position. This window is then fixed and the best *pair* of windows, in this case the eighth plus the third, is found. Thus, the third window contains the next best amount of information and is ranked in second position. Using combinatorial search the next best window added at each stage matches the window rank chosen by simple search.

Combinatorial search continues until all windows have been combined. As can be seen in Figure 4 (bottom) each *combination of Attention Windows* exhibits a lower uncertainty measure than the previous one, i.e. the uncertainty decreases per n best windows. Intuitively, this makes sense as the more information available, the better the image reconstruction should be. However, the payoff for using many Attention Windows is not significant as can be seen from the small drop in uncertainty. This result is also borne out by Figure 7, as detailed in Section 5.2. Clearly, the fact that the highest ranking Attention Window is not only the most interesting, but is the most interesting by a significant factor, is the reason why we need use only it for topological position estimation and additionally, for reconstruction. Figure 4 (top) shows how the uncertainty increases when each lower ranking window is used alone.

In terms of computation time, simple search took an average of 6.2 seconds to rank the Attention Windows while combinatorial search took an average of

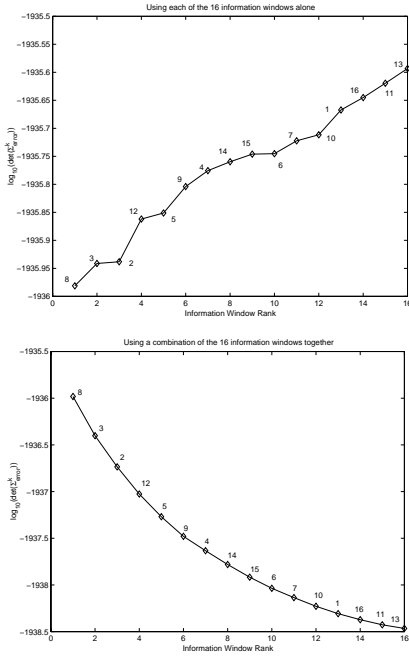


Figure 4: Graphs of the information contained in each eigen window versus the window rank when using simple search (top) and combinatorial search (bottom). The numbers along the graph line are the windows numbers.

63.9 seconds to determine the same information. The trade-off is accuracy versus computational power.

5.2 Window Results

We can divide our results into two main categories: Position Estimation using Eigen Windows and Reconstruction using Attention Windows.

1. Position Estimation using Eigen Windows

Having previously selected the best Attention Window using *Information Sampling*, we built a local appearance space using only the data selected with this window from each a priori image. This further compressed our data to only approximately one thousandth of the original 128×128 image data. Successful position estimation has been achieved using windows as small as 4×4 pixels in size.

We projected *only* the selected best Attention Window from each image into the eigenspace. This is an improvement on previous approaches, where all windows first had to be projected. Thus, we were able to immediately reduce the ambiguity associated with projection. The images in

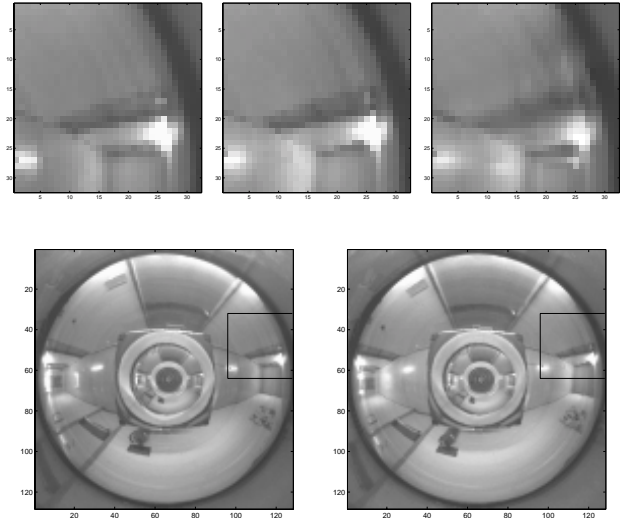


Figure 5: Top: Close-up of the 32×32 Attention Windows: unknown (left), closest (middle) and reconstructed (right). Bottom: The position of the unknown and closest images in their respective omnidirectional images.

Figure 5 show the results obtained using windows of 32×32 pixels in size.

The top row, from left to right shows the most interesting Attention Window from an unknown image, its closest match from the a priori set of omnidirectional images and its reconstruction using equation (7). The bottom row, shows the Attention Window in the unknown 128×128 image (left) and its closest match from the a priori set obtained by projecting only the most interesting Attention Window (right). We note here that we could in principle, given enough computing power, use equation (3) to reconstruct a 128×128 image using only the most interesting window.

Figure 6 shows the distance between images acquired at run time and those acquired a priori. The global minimum is the correct estimate of the robot's current topological position. Local minima correspond to similar images to the current one some distance away from the robot's current position. Figure 6 (top) shows the graph obtained using 16×16 images, while Figure 6 (bottom) that obtained using the most interesting 4×4 image window. While different local minima are obtained, it can clearly be seen the global minimum using both methods is maintained. Thus, we have shown that effective nav-

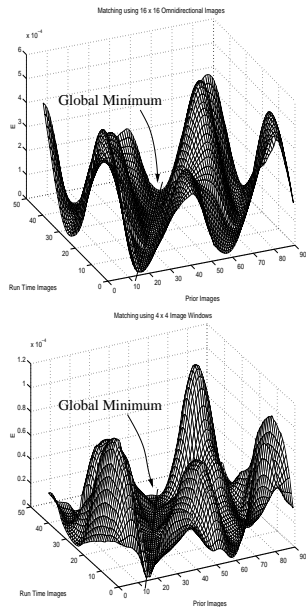


Figure 6: A 3D plot of image acquired at run time versus those acquired a priori using (top) 16×16 images and (bottom) 4×4 Attention Windows, respectively.

igation can be undertaken using only the most informative 16 pixels from each image.

2. Reconstruction using Attention Windows

The reconstruction results obtained using Attention Windows of size 4×4 pixels, selected from omnidirectional images of 16×16 pixels in size and of using 4×4 windows from images of 32×32 pixels in size, respectively are shown in Figure 7. The top row of Figure 7 (Top left) shows a 16×16 omnidirectional image from the a priori set, Figure 7 (Top middle) its reconstruction using only the *most informative* 4×4 Attention Window and 7 (Top right) its reconstruction using all of the Attention Windows. Reconstruction was achieved using equation (3). The bottom row of Figure 7 shows the same results obtained using 8×8 Attention Windows selected from 32×32 omnidirectional images. As can be seen from the images, a very good reconstruction is obtained using *only* the best Attention Window. This is an indication of the power of *Information Sampling*.

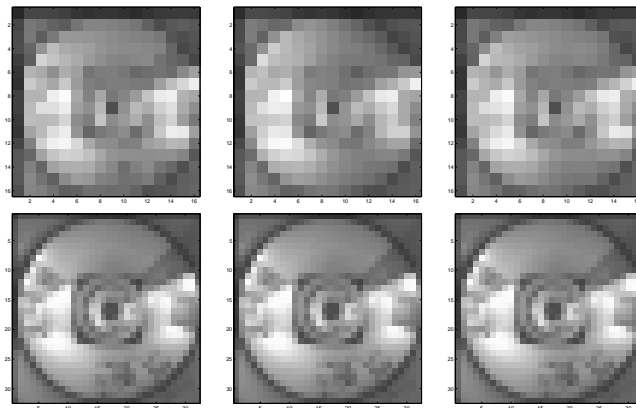


Figure 7: Top: (Left) A 16×16 omnidirectional image acquired at runtime, (Middle) its reconstruction using the *most informative* Attention Window and (Right) its reconstruction using all of the Attention Windows. Each Attention Window is 4×4 pixels in size. Bottom: Results obtained using 32×32 omnidirectional images and 8×8 Attention Windows.

6 Conclusions and Future Work

This paper presented a statistical method termed *Information Sampling* to aid in the visual guidance of a mobile robot through an indoor environment. This was achieved by focusing attention on the most interesting data from a set of images. These could be a single pixel or groups of pixels and for computational reasons, *Information Sampling* selected the most interesting data as *Attention Windows*. We showed how to use only these windows to determine the qualitative position of a mobile robot within its environment. This position estimation technique was an improvement over existing approaches since we only needed to project the best data, thus reducing ambiguity. Additionally, we detailed two methods for ranking the Attention Windows. By focusing using the highest ranking Attention Window, the robot makes the best use of its computational resources.

Our future work will be directed towards undertaking on-line experiments and extending the implementation of *Information Sampling*. The possibility of rebuilding the local appearance space in real time, as a robot explores new regions of its environment shall be investigated.

Acknowledgements

This work was partly funded by the European Union RTD - Future and Emerging Technologies Project Number: IST-1999-29017, Omniviews.

References

- [1] R. Brooks. Visual map making for a mobile robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'95)*, 1985.
- [2] V. Colin de Verdière and J. L. Crowley. Visual recognition using local appearance. In *5th European Conference on Computer Vision, (ECCV 1998)*, pages 640–654, Freiburg, Germany, June 1998.
- [3] J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omni-directional camera. *IEEE Transactions on Robotics and Automation*, 16(6):890–898, December 2000.
- [4] T. Hancock and S. Judd. Ratbot: Robot navigation using simple visual algorithms. In *IEEE Regional Conference on Control Systems*, pages 181–184, August 1993.
- [5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.
- [6] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Riseman. Image-based homing. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'91)*, pages 620–625, 1991.
- [7] H. Ishiguro and S. Tsuji. Image-based memory of environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'96)*, pages 634–639, 1996.
- [8] D. Jugessur and G. Dudek. Local appearance for robust recognition. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 834–839, Hilton Head Island, SC, USA, June 2000.
- [9] D. Kortenkamp and T. Weymouth. Topological mapping for mobile robots using a combination of sonar and visual sensing. In *AAAI'94*, July 1994.
- [10] Jana Kosecka. Visually guided navigation. In *Proc. 4th Int. Symp. on Intelligent Robotic Systems (SIRS'96)*, Lisbon, Portugal, July 1996.
- [11] B. J. Kuipers and Y. Byun. A robot exploration and mapping strategy on a semantic hierarchy of spatial representations. *Journal of Robotics and Autonomous Systems*, 8:47–63, 1991.
- [12] S. Maeda, Y. Kuno, and Y. Shirai. Active navigation vision based on eigenspace analysis. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS'97)*, pages 1018–1023, 1997.
- [13] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue. Exploration and navigation in corridor environment based on omni-view sequence. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'00)*, pages 1505–1510, 2000.
- [14] H. Murase and S. K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, January 1995.
- [15] K. Ohba and K. Ikeuchi. Detectibility, uniqueness and reliability of eigen windows for stable verification of partially occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):1043–1048, September 1997.
- [16] D. Reifeld, H. Wolfson, and Y. Yeshurun. Context free attention operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14(2):119–130, March 1995.
- [17] M. Rendas and M. Perrone. Using field subspaces for on-line survey guidance. In *Proceedings of Oceans 2000*, Providence, RI, USA, June 2000.
- [18] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, May 1997.
- [19] N. Winters, J. Gaspar, and J. Santos-Victor. Omni-directional vision for robot navigation. In *Proceedings of the 1st International IEEE Workshop on Omni-directional Vision at CVPR 2000*, pages 21–28, Hilton Head Island, SC, USA, June 2000.
- [20] N. Winters and J. Santos-Victor. Omni-directional visual navigation. In *Proceedings of*

the 7th International Symposium on Intelligent Robotics Systems, pages 109–118, Coimbra, Portugal, July 1999.

- [21] N. Winters and J. Santos-Victor. Information sampling for optimal image data selection. In *submitted to the 9th International Symposium on Intelligent Robotics Systems*, Toulouse, France, July 2001.
- [22] J. Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, 9(1):55–76, January 1992.