Improved Vision-based Robot Navigation Using a SDM and Sliding Window Search

Mateus Mendes∗†, A. Paulo Coimbra∗, and Manuel M. Crisóstomo∗

Abstract—Robust and fast vision-based robot navigation is a long sought goal, which requires comparing the robot’s current view with a database of visual memories. The technique described in the present paper uses a Sparse Distributed Memory (SDM) to store paths described by sequences of images, and a sliding window to narrow the search space for real-time operation. The use of the sliding window greatly reduces the processing time and the number of prediction errors. The use of a short-term memory confers on the robot the ability to still solve the kidnapped robot problem. The Sparse Distributed Memory is a kind of associative memory suitable to work with high-dimensional binary vectors, thus being appropriate to store long sequences of images.

Index Terms—Robot Navigation, View-based Navigation, Sliding Window, SDM, Sparse Distributed Memory

1 Introduction

Among all the techniques that may be used for robot localisation, vision-based approaches are probably the most cherished, for they are biologically inspired. Humans strongly depend on visual memories: approximately 80% of the sensorial information processed by an average person is visual [1]. Additionally, the sensors required for vision-based robot navigation are inexpensive. An average quality video camera may be enough for the simpler applications.

The main drawback of vision-based approaches is that the processing power needed is huge. Every single image is usually described by several hundreds or thousands of pixels, and every path that the robot learns is described by tens, hundreds or even thousands of images, depending on the path length and frame rate used. That makes the technique less appealing, because real-time operation may be compromised for large databases, due to increasing needs of processing power and error probability. As the robot learns more and more paths, the number of images it has to store grows continually. The memory requirements and the processing time increase in proportion to the database size. It should be noted, however, that modern evidence indicates that the human brain functions in a similar manner: it is a huge amount of memory, used to store sequences of events that will lead future analysis and actions [2, 3].

The images alone are a means for instantaneous localisation. View-based navigation is almost always based on the same idea: during a learning stage the robot learns a sequence of views and motor commands that, if followed with minimum drift, will lead it to a target location. The robot is later able to follow the learnt path by following the sequence of commands, possibly correcting the small drifts that may occur.

In previous work the authors presented a system to navigate a robot using images stored into a Sparse Distributed Memory (SDM) [4]. The SDM is a kind of associative memory based on the properties of high-dimensional boolean spaces, and thus suitable to work with large binary vectors such as images [3]. The method was efficient even under difficult conditions [5], but the processing requirements were very demanding. The present paper, which is an extended version of [6], describes an improvement to the system, in which a search sliding window truncates the search space and thus considerably reduces the time and processing requirements, as well as the number of robot localisation errors.

Section 2 reviews some popular navigation techniques and explains navigation based on view sequences in more detail. Section 3 briefly describes how the SDM works. In Section 4 the experimental platform used is described. Section 5 explains the problems encountered with the original navigation method, and how the application of a sliding window contributes to solve many of them. Section 6 shows and discusses the results obtained, and Section 8 draws some conclusions.

2 Robot Navigation

There are many different approaches for robot navigation. Vision-based approaches have been extensively used, although other techniques may be better for some types of common environments.
2.1 Popular techniques

Besides vision-based techniques, many different approaches have been tried to localise and navigate robots in a safe and robust way. Some of those approaches work only in structured environments, since they are based on the recognition of artificial landmarks, beacons, indoor/outdoor GPS or similar strategies [7]. Those strategies greatly improve the accuracy of the system. They trim the complexity of the environment and narrow the robot search space. The problem of robot localisation becomes much simpler, for the robot only has to look for selected signals and cut off all the remaining data that is unnecessary. The disadvantages, however, are obvious: such approaches are suitable only for structured environments. They are not a general solution to the problem of robot localisation.

More generic strategies that work in unstructured environments include mapping and localisation using laser range finders, as well as sonars or cameras for vision-based approaches. In this case the robot is equipped and programmed in such a way that it is expected to be able to succeed in localising itself in a wide range of environments, and those environments do not have to be intentionally structured in order for the robot to succeed. It should be able to map the environment, regardless of its characteristics, and later use that map to localise itself autonomously.

Popular sensors used for robot localisation are sonars, infrared and laser range finders. Sonars are cheap but in general offer poor precision. Infrared sensors offer better precision, but they hardly work on clear daylight. Lasers are expensive and may also have problems working in daylight. On the other hand, an average quality video-camera is nowadays cheap and can work in a wide range of environments with different illumination levels [8].

2.2 Navigation using view sequences

There are two popular approaches for vision-based navigation: one that uses plain images [9], the other that uses omnidirectional images [10]. Omnidirectional images offer a 360° view, which is richer than a plain front or rear view. However, that richness comes at the cost of even additional processing power requirements. Some authors have also proposed techniques to speed up processing and/or reduce memory needs. Matsumoto [11] used images as small as 32×32 pixels. Ishiguro replaced the images by their Fourier transforms [12]. Winters compressed the images using Principal Component Analysis [13].

In the present work, the approach followed to navigate the robot is based on using visual memories stored into a Sparse Distributed Memory, as described in [4]. It requires a supervised learning stage, in which the robot is manually guided. While being guided, the robot memorises a sequence of views automatically. It stores a sequence of views for each path. Images that are very similar to previously stored images are discarded, because they would, with high probability, not add any relevant information to the known information. They might even disturb important information already stored into the SDM, as explained further in Section 3.

While running autonomously, the robot performs automatic image-based localisation and obstacle detection. Localisation is estimated based on the similarity of two views: one stored during the supervised learning stage and another grabbed in real-time. To minimise possible drifts to the left or to the right, the robot tries to find matching areas between those two images and calculates the horizontal distance between them in order to infer how far it is from the correct path, being able to reduce the drift iteratively over time. There is no need to process the vertical distance, since the camera is fixed and vertical shifts are not expected. The technique is described in more detail in [4].

3 Sparse Distributed Memories

The Sparse Distributed Memory is an associative memory model proposed by Kanerva in the 1980s [3]. It is suitable to work with high-dimensional binary vectors. In the proposed approach, an image is regarded as a high-dimensional vector, and the SDM is used simultaneously as a sophisticated storage and retrieval mechanism and a pattern recognition tool.

3.1 Previous use of the SDM for robot navigation

The concept of the SDM is very attractive for robot navigation. It confers on the robot the ability to learn and follow paths the same way humans do. Some researchers have already explored the idea to some extent. Rao and Fuentes [14] simulated the use of a SDM to store position information from optical sensors, associated with motor controls, during a learning stage. Using that information, the robot was later able to follow the same paths. The authors presented only simulation results. Watanabe et al. [15] used a SDM for the task of scene recognition in a factory environment, where the robots had to move autonomously from one place to another. The SDM, however, was just used as an auxiliary method of scene recognition. It was not used for robot navigation purposes.

3.2 The original model

The underlying idea behind the SDM is the mapping of a huge binary memory onto a smaller set of physical locations, called hard locations. That way it is possible to mimic the existence of a much larger space, taking ad-
must be able to succeed in a wide range of difficult situations which are typical in robot navigation, such as partial occlusion, illumination changes and memory overflow. A number of experiments have been carried out to prove that assumption, and the results are described in [8].

Another interesting characteristic of the SDM model is that the same set of vectors can be used simultaneously to store the addresses and the data, as long as any given datum $\zeta$ is always stored at address $\zeta$. A practical consequence of this is that one of the arrays can be discarded, cutting the memory size down to about one half its original size. Such a memory, comprising only one array, where datum $\zeta$ is stored at address $\zeta$, is called auto-associative.

3.3 The models used

The original SDM model has been subject to various improvements and alternative implementations. In the present work, four variations have been studied: the arithmetic mode, the bitwise mode using the natural binary code, the bitwise mode using an optimised code, and the bitwise mode using a sum-code. Those models are described in [5].

All the models used are auto-associative and use the Randomised Reallocation (RR) algorithm [16]. Using the RR, the system starts with an empty memory and allocates new hard locations when there is a new datum which cannot be stored into enough existing locations. The new locations are placed randomly in the neighbourhood of the new datum address.

3.3.1 Bitwise SDM

The bitwise implementation is very similar to the original model. The difference is that it stores only one bit per input vector bit, thus dropping the bit counters, as shown in Figure 2. Writing in such a model consists in just replacing the old datum. The advantages are that the capacity of storing data is improved, and reading and writing is much faster. The model was inspired by Furber et al.’s approach [17].

3.3.2 Use of an optimised code

As described in [5], the Hamming distance between two binary numbers is not proportional to the arithmetic distance. For example, the Hamming distances $d_2(0111_2, 1111_2) = h_2(1110_2, 1111_2) = 1$. That happens because the Hamming distance does not take into account the positional values of the bits. However, the sensorial data is encoded using the natural binary code, which takes into account the positional values of the bits. Using arithmetic distances, $d_1(0111_2, 1111_2) = 8$ and $d_2(1110_2, 1111_2) = 3$. The use of an optimised code changes the situation, making the Hamming distance a better approximation of the arithmetic distance.

Figure 1: One model of a SDM, using bit counters. The address bits highlighted are different in the address vector, compared to the corresponding bits in the input address.

![Image of SDM model](Image 66x657 to 274x770)
1. Hence, different criteria are used to encode the input information and to process it inside the SDM according to Kanerva’s original model. That difference causes a loss of performance of the system, and to overcome the problem other memory models were implemented. The first alternative encodes the data using an optimised code. In that optimised code some bytes are sorted, in order to minimise the effect of using different criteria to encode the input data and to process it inside the SDM.

3.3.3 Use of a sum-code

In another model, the data is encoded using a sum-code of 9 graylevels. In that code, each binary number is mapped into the range \{00000000, 00000001, 00000011, ..., 11111111\}. That way the Hamming distance between any two binary numbers is proportional to the arithmetic distance.

3.3.4 Arithmetic SDM

In the arithmetic implementation, the bits are grouped as byte integers, as shown in Figure 3. Addressing is done using an arithmetic distance, instead of the Hamming distance. Learning is achieved updating each byte value using the equation:

$$h_k^t = h_k^{t-1} + \alpha \cdot (x^k - h_k^{t-1}), \quad \alpha \in \mathbb{R} \land 0 \leq \alpha \leq 1$$ (1)

In the equation, \(h_k^t\) is the \(k^{th}\) number of the hard location, at time \(t\), \(x^k\) is the corresponding number in the input vector \(x\) and \(\alpha\) is the learning rate. In the present implementation \(\alpha\) was set to 1, enforcing one shot learning.

4 Experimental platform

The robot used was a Surveyor\(^1\) SRV-1, a small robot with tank-style treads and differential drive via two precision DC gearmotors (Figure 4). Among other features, it has a built-in digital video camera and a 802.15.4 radio communication module. The robot was controlled in real time from a laptop with a 1.8 GHz processor and 1 Gb RAM. The overall software architecture is as shown in Figure 5. It contains three basic modules:

1. The SDM, where the information is stored.
2. The Focus (following Kanerva’s terminology), where the navigation algorithms are run.
3. An operational layer, responsible for interfacing the hardware and some tasks such as motor control, collision avoidance and image equalisation.

Navigation is based on vision, and has two modes: supervised learning, in which the robot is manually guided and captures images to store for future reference; and autonomous running, in which it uses previous knowledge to navigate autonomously, following any sequence previously learnt. The vectors stored in the SDM consist of arrays of bytes, as summarised in Equation 2:

$$x_i = \langle \text{im}_i, \text{seq}_{id}, i, \text{timestamp}, \text{motion} \rangle$$ (2)

In the vector \(x_i\), \(\text{im}_i\) is the image \(i\), in PGM (Portable Gray Map) format and 80×64 resolution. In PGM images, every pixel is represented by an 8-bit integer. 0 is black, 255 is white. \(\text{seq}_{id}\) is an auto-incremented,

4-byte integer, unique for each sequence. It is used to identify which sequence the vector belongs to. \( i \) is an auto-incremented, 4-byte integer, unique for every vector in the sequence, used to quickly identify every image in the sequence. \( \text{timestamp} \) is a 4-byte integer, storing Unix timestamp. It is not being used so far for navigation purposes. \( \text{motion} \) is a single character, identifying the type of movement the robot performed after the image was grabbed. The image alone uses 5120 bytes. The overhead information comprises 13 additional bytes. Hence, the input vector contains 5133 bytes.

5 Use of a search window

The use of a search window, which truncates the search space, greatly improves the speed and performance of the method.

5.1 The problems

There are two weaknesses of the view-based navigation approach described: i) processing time required to store and retrieve one image and ii) prediction errors, when the memory outputs a wrong image and motion command.

As for the processing time, it is proportional to the number of images stored in the memory. Each new image has to be compared to all the hard locations that exist in the memory. That may be a problem for real time operation, specially if a single processor is used.

As for the second problem, it is due primarily to the existence of noise in the images, which is impossible to avoid. When following a path, it is normal that the robot makes some wrong predictions. It is difficult to count the exact number of errors, but in this case we define the concept of “Momentary Localisation Error” (MLE). A MLE occurs when the system retrieves image \( \text{im}_i \), after having retrieved \( \text{im}_j \), for \( i, j > 0 \). That is a reasonable assumption, since, under normal circumstances, the robot is not expected to get back in the sequence. If at some point of a path the prediction is \( \text{im}_i \), and after that it is \( \text{im}_{i-j} \), then it means that at least one of the predictions was wrong. Those MLEs are not to worry when the robot is performing the same movement in both the correct and the wrongly retrieved image. That is often the case, since there are only 4 possible motions (forward, backward, turn left and turn right). But a prediction error could compromise the robot’s ability to complete a path if the correct motion and the motion associated with the retrieved image are different.

5.2 Distribution of the Momentary Localisation Errors

Table 1 shows the number of MLEs measured while following a typical path, described by a sequence of 130 images. The first row of the table indicates the distance of the image predicted by the memory to the last predicted image. The first column is the operation mode.

As the table shows, most of the MLEs occur with adjacent images: the distance between the expected image and the retrieved image is 1. More than 60% of the MLEs are between adjacent images, regardless of the memory operation mode. In the bitwise mode the MLEs are more distributed in the range of distances [1-5] than in the other modes. That makes sense, considering that the bitwise mode is, in general, the weakest of all [5]. In the example path no MLEs were detected at distances greater than 5 images, and that is also a normal behaviour of the system. Figure 6 shows an histogram of the distribution of MLEs.

5.3 The use of a sliding window

The use of a sliding window helps improving both the processing time and the number of MLEs. It works like the use of a kind of context, in which the topic is narrowed to a given subject. In the case of the SDM, that is equivalent to segmenting the search space.

L. Jaeckel proposed a method of segmenting the space by way of using only a limited set of coordinates, instead of all the binary vector, to determine the set of active locations [18]. The method implemented in the present work has some similarities to Jaeckel’s approach. The idea is narrowing the search field to a number of images before and after the last predicted image, as illustrated in Figure 7. For example, if the robot is following path A and the last image retrieved is image \( i \), in the next prediction it is expected to be still following path A and retrieve either image \( i \) or image \( i+1 \). Since the length of the step used in the autonomous run is 1/16th of that used during the learning stage, it will see image \( i \) for some time and that is no prediction error. The use of a sliding window of width \( 2 \times j + 1 \) consists in narrowing
Table 1: Distribution of the MLEs according to the operation mode, without search window, in a typical path described by 130 images.

<table>
<thead>
<tr>
<th></th>
<th>-5</th>
<th>-4</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Bitwise</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>Optimised code</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td>Sum-code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 6: Distribution of the MLEs, in the four operation modes, without search window, in a typical path described by 130 images.

Figure 7: Example of a sliding window of width 3. For image \( i \), search is first performed in the interval \([i-1, i+1]\).

6 Experiments and results

As Table 1 shows, more than 60% of the MLEs occur between adjacent images (distance -1). The other MLEs appear at absolute distances of 2, 3, 4 or 5 images. Although those errors account for less than 40% of the total, they are still undesirable.

In order to assess the performance of the system using a sliding window, the navigation algorithm was updated to narrow the search to the same sequence and a window of three images, in the interval \([im_{i-1}, im_{i+1}]\)—i.e., for each image, perform the search by comparing just with the last seen image, the image prior to that one in the sequence and the next expected image in the sequence.

Table 2 shows the results obtained when following the already presented example path, using a search window of width 3. One interesting conclusion is that the search window cut more MLEs than those counted out of its range, except for the arithmetic mode. That is explained by the fact that some MLEs may actually be the reason of other MLEs. For example, a MLE that causes a wrong motion of the robot may cause drifts and additional MLEs in the future. The improvements are of 50% or more, except for the arithmetic mode.

Figure 8 illustrates the data shown in Table 2, related to the number of MLEs counted with and without using the search window. The histogram clearly shows the impact of the method, specially in the bitwise modes.

Figure 9 illustrates the differences in processing time, as shown in Table 2. It is clear that there is an improvement of about 93% in the processing time. That makes sense, considering that the memory is loaded with a sequence of 130 images. The use of the search window makes the
algorithm skip all but three images, and those three images represent only 2.31% of the whole sequence. Since most of the time necessary to make a prediction is actually spent comparing images, an improvement of 93% is coherent with the theory.

7 Discussion

As shown in Section 6, the use of a search window greatly improves the performance of the system. In the example path it reduced the number of momentary localisation errors up to 67%, and the processing time up to 95%. That improvements are possible at the cost of truncating the search space. Under normal circumstances, truncating the search space should pose no problem to the robot. However, the solution loses generality, because it is strongly based on the robot’s short memory: the algorithm is based on the assumption that the robot is always close to its last position. Nonetheless, it may happen that the robot slips while moving, is manually moved by a human to another location, etc. That is commonly known as the “kidnapped robot” problem.

To achieve robust navigation, a robot must not rely strictly on a search window, otherwise it will not solve the kidnapped robot problem. Using a SDM that problem may be easily overcome. A general solution to the problem is to use an algorithm that, for each new image:

1. Search within the sliding window. If the search retrieves one or more images within the SDM access radius (as explained in Section 3.2), then assume that the prediction is correct.
2. If the search within the sliding window does not retrieve at least one image within the SDM access radius, then perform a global search in the SDM and use the best prediction.

The algorithm as described still takes advantage of the sliding window under normal circumstances, and is able to solve the kidnapped robot problem.

8 Conclusions

Robot navigation based on visual memories is a long sought goal. However, it requires heavy processing due to the amount of information that has to be processed in real time. The approach followed in the present work is vision-based robot navigation using images stored into a Sparse Distributed Memory. The speed of the process can be largely improved with the use of a search window. The search window truncates the search space, reducing significantly the processing time as well as the number of prediction errors, thus improving the real time performance operation of the robot.

Acknowledgements

This work is supported in part by grant SFRH/BD/44006/2008 from Fundação para a Ciência e Tecnologia, Portugal.

References

Table 2: MLEs and processing time without using search window and with search window of size 3, in a typical path described by 130 images.

<table>
<thead>
<tr>
<th></th>
<th>Arithmetic mode</th>
<th>Bitwise mode</th>
<th>Optimised code</th>
<th>Sum-code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MLE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without search window</td>
<td>11</td>
<td>53</td>
<td>55</td>
<td>14</td>
</tr>
<tr>
<td>With search window</td>
<td>7</td>
<td>19</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Improvement</td>
<td>36%</td>
<td>64%</td>
<td>67%</td>
<td>50%</td>
</tr>
<tr>
<td><strong>Time (µs)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without search window</td>
<td>1 511.38</td>
<td>14 567.16</td>
<td>16 269.21</td>
<td>116 846.54</td>
</tr>
<tr>
<td>With search window</td>
<td>1 011.14</td>
<td>988.44</td>
<td>1 000.95</td>
<td>5 539.48</td>
</tr>
<tr>
<td>Improvement</td>
<td>93%</td>
<td>93%</td>
<td>94%</td>
<td>95%</td>
</tr>
</tbody>
</table>


(Advance online publication: 12 November 2011)