

Fast Sum of Absolute Differences Visual Landmark Detector

Craig Watman*, David Austin*[†], Nick Barnes[†], Gary Overett* and Simon Thompson[‡]

*Robotic Systems Laboratory

Department of Systems Engineering, RSISE, Australian National University

Canberra, ACT 0200 Australia

Email: d.austin@computer.org

[†]Autonomous Systems and Sensing Technologies Program, National ICT Australia

Locked Bag 8001 Canberra, ACT 2601 Australia

[‡]Digital Human Research Center

National Institute of Advanced Industrial Science and Technology

Aomi, Koto-ku Japan

Abstract— This paper presents various optimisations that can be applied to the Sum of Absolute Differences (SAD) correlation algorithm for automated landmark detection. This has applications in mobile robotic navigation and mapping. We show how some assumptions about the environment and the generic form of strong landmarks selected by the SAD correlation algorithm have led to the development of an algorithm to enable near real time selection of strong landmarks from visual information.

The landmarks that have been selected from a series of frames using our optimisations are shown to be stable through the image sequence, demonstrating the scale invariance of the landmarks that are selected by the SAD correlation algorithm.

I. INTRODUCTION

The use of vision for robotic navigation can dramatically decrease the cost of production of a mobile robot by replacing more expensive sensors with relatively inexpensive stereo cameras.

The identification of natural landmarks in real time using visual information is a key problem that needs to be solved before inexpensive autonomous robots can be widely used. A landmark is a visual feature that is relatively unique within an image and is stable over time, usually from different distances and angles.

In this paper we present optimisations of the Sum of Absolute Differences (SAD) algorithm to enable the near real time identification of strong visual landmarks gathered from visual information in the environment, this will enable dynamic construction of environment maps and navigation through the environment using these maps. Our current work is restricted to indoor environments as outdoor environments pose greater problems and complexity.

We propose to use the SAD correlation algorithm to identify landmarks in near real time. Our current results enable the selection of ten natural landmarks from an indoor environment at speeds of up to 0.17 seconds per frame, with good repeated selection of landmarks over frames indicating that the landmarks are strong within the environment and that they can be consistently selected from different perspectives.

These results indicate that the SAD correlation algorithm can offer a potential alternative to the SIFT method proposed by Lowe [5] for landmark selection. The SIFT method is used to extract features that are scale, translation, rotation and to some extent illumination invariant. The aim of this method is to select a set of features that characterise an image as stated by Se [6], we argue that the aim in feature extraction is to select stable features from the world that can be re-detected easily, rather than select features that are invariant to scale, rotation and translation in image space.

A feature of the algorithm presented in this paper is that it enforces spatial separation of the features that are selected. This is critical in using visual features for navigation, as accurate triangulation requires features that are spatially separated in the 3D world.

II. SUM OF ABSOLUTE DIFFERENCES FOR LANDMARK SELECTION

Static landmarks can be selected from visual information based on the uniqueness of a landmark template in the surrounding local area. We use an adaption of the SAD method as used by Bianco and Zelinsky [1], based on “The Valley Method” proposed by Mori et al. [2], and a slight adaption of the method proposed by Thompson [3].

The SAD method generates a metric of local uniqueness by calculating the distortion from a template image within a search window centered on the template image. For each pixel in the search window, the SAD is calculated by computing the correlation value using

$$SAD = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |I_{i,j} - T_{i,j}| \quad (1)$$

where the template T of M by N pixels is correlated with a surrounding image I for the same size drawn from within the search window space. We use a template size of 16 by 16 and a search window area centered on the template of 32 by 32.

This provides a 16 by 16 distortion matrix from which to compute a metric of local uniqueness. The distortion matrix will always have a zero distortion value in the center where the template is matched to itself as can be seen in both images in Fig. 1. The upper image in Fig. 1 shows the distortion matrix that is obtained from a random noise image, the lower image shows the distortion matrix from a strong landmark from an indoor area. The values in the distortion matrix for the strong landmark are generally smaller than the values obtained from the noisy landmark.

The lower image in Fig. 1 demonstrates the valleys that are usually associated with strong landmarks when calculated using the SAD correlation algorithm. We use this feature to generate a simple distortion metric based on the ratio between the global minimum in the center, and local minimum, selected from the rest of the distortion matrix. Given that the global minimum is always zero, the distortion metric is simply the local minimum.

We have not used the Normalized Cross Correlation (NCC) [7] algorithm here as we are trying to identify areas within the same image scene that exhibit local differences. The NCC algorithm uses the mean pixel value of an image to normalize lighting effects. The NCC is traditionally used to reduce the effect of differing lighting in two images, however as we are searching within a single image the light conditions are not altered, further more the use of the NCC algorithm in this context tends to exaggerate all differences in pixel values [3].

The local minimum is found by scanning the distortion matrix for the lowest value excluding the center distortion value, however given generic form of strong landmarks we propose an optimisation called the “Lazy Spiral” that reduces the number of computations required, we describe is optimisation in detail later in this paper.

In order to select highly distinct regions in the image, we are seeking a distortion matrix that exhibits a high local minimum. A high local minimum indicated that there is high distortion between the template image and the surrounding image area. High distortion indicates that the template area contains image features that are not present in the area immediately surrounding it, making the template area a good candidate for a landmark.

We select the top ten non-overlapping landmarks based on the SAD values that they produce. We have chosen to use non-overlapping landmarks because we are applying a method that seeks regions of interest not points of interest. Thus it is easy to see that pixels in one region produce similar SAD values as the adjacent pixels, as can be seen by the valleys exhibiting lower values near the global minimum. This implies a strong correlation between the template image and the immediately surrounding images. Based on this we propose an optimisation that uses a template stepping approach to reduce the computational load.

III. OPTIMISATION

The calculation of the distortion matrices for every possible template in an images of 640 by 480 pixels using an x

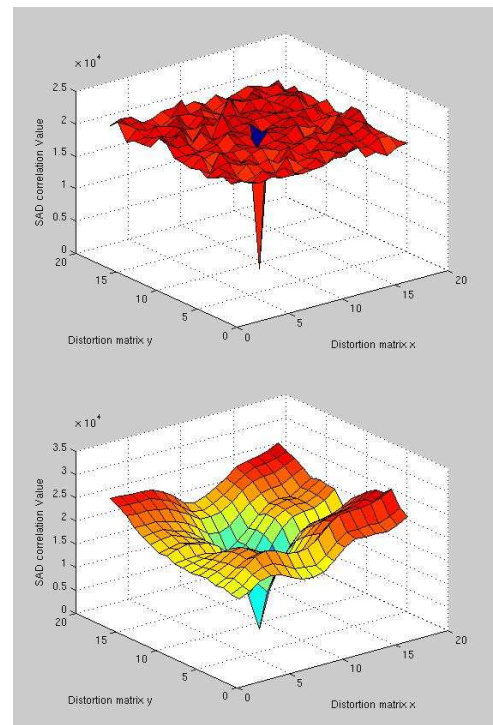


Fig. 1. Strong and Weak Distortion Matrices

to y scanning pattern takes 123.58 seconds on an AMD Athlon™XP 2100+ processor. We have implemented a series of optimisations that enable the processing of a 640 by 480 image in 0.17 seconds.

We have achieved this through four optimisations, template stepping, uniform area reduction, adaptive thresholding and lazy spiral calculation of the SAD. We have also experimented with a fifth optimisation that assumes that the distortion matrices take on the normal valley appearance as seen in Fig. 1, we have named this the normal form calculation of the SAD.

A. Template Stepping

The template stepping method assumes that the template image in a local region is relatively similar to the template images centered on the surrounding pixels. This assumption enables us to perform a single pass over the image calculating the SAD values at every i th pixel. Now that we have calculated SAD values over the image, we select the upper ten non-overlapping unique regions based on the SAD metric and perform a second series of SAD calculations within a $2i$ by $2i$ search window centered on the regions identified by the first pass. The second pass does not use template stepping and is a refinement step to select the best possible SAD from within the $2i$ by $2i$ region. In essence we are using the first pass to obtain ten “ball park” estimations of locally unique regions and then refining these by performing the full SAD correlation algorithm locally.

Through experimentation with different template step sizes we have selected a size of three. This value was chosen

because it provides good performance with little or no alteration in the final landmark positions. The graph in Fig. 2 shows the processing time for a range of template sizes using the lazy spiral calculation and adaptive thresholding method outlined later in this paper.

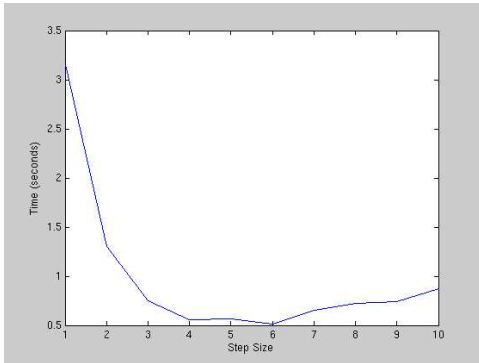


Fig. 2. Template Step selection

The template stepping method used alone has given speeds of 14.28 seconds, with little alteration in the selected landmarks from the exhaustive SAD calculation.

B. Uniform Area Reduction

We use a simple method to identify pixels that are in uniform areas of the image. These areas of the image include regions on a wall, carpet, or doors and they are characterised by little variation in color with neighboring pixels.

As there is little variation in color on uniform regions these regions do not yield landmarks that are locally unique, so the fast identification and the removal of these areas from the SAD processing can significantly reduce the processing time for indoor images. Natural outdoor images do not generally exhibit many uniform areas and this method will not yield major efficiencies in these image types. The method we have chosen to identify uniform areas is simple and computationally inexpensive, we search around the circumference of the template centered on each pixel and compare the color of each circumference pixel to the color of the center pixel. We set a threshold of 100 color units for the difference between the color of the circumference pixel and the center pixel, if the threshold is not exceeded then the pixel designated as a uniform area pixel and is not processed.

The threshold is aggressive requiring a color change of approximately 40 percent for at least one pixel on the template circumference and has led to a reduction of the number of points that are processed to less than 25,000 in most cases from the initial 272,384 points. The top image in Fig. 3 shows an original image frame from our indoor test image sequence and the lower image shows the points that have been selected to be processed using the uniform area reduction method.

C. Adaptive Thesholding

Early recognition of weak landmarks is achieved through the use of a threshold on the value of the local minimum. This

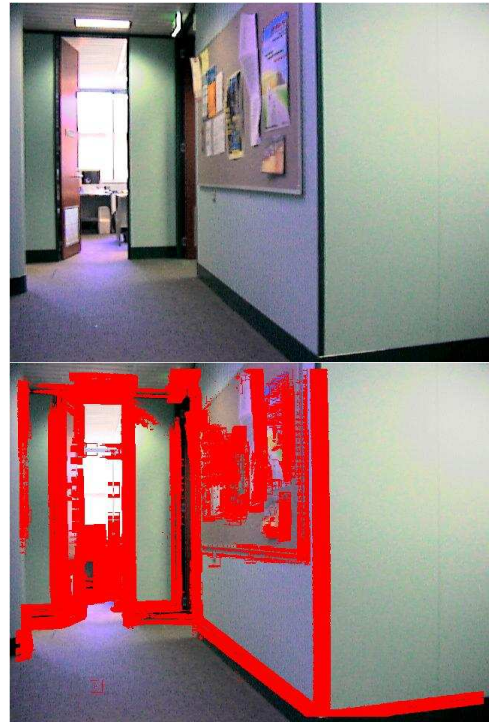


Fig. 3. Point Reduction

is implemented by first setting the threshold to zero, after the first ten landmarks have been selected the threshold is set at the minimum SAD value from the landmark set. As further SAD values are calculated if the SAD value falls below the threshold we can stop calculating the SAD within the search window as the local minimum is already below the acceptable value to be included in the landmark set.

This method quickly reduces the number of calculations that need to be performed to assess each landmark template for local uniqueness. By recognising early when a landmark is weak we can halt further calculations for that template. This has led to a reduction in the average processing time to 4.45 seconds per image with template stepping.

D. Lazy Spiral Calculation

The lazy spiral SAD calculation recognises that given the general form of the distortion matrices the local minimum is most likely to be found in the pixels surrounding the global minimum. Coupling this knowledge with the adaptive thresholding method we search outwards from the global minimum, and we are more likely to locate the local minimum earlier thus enabling recognition of poor landmarks with less computations.

Searching for the local minimum is most efficient when the pixels are assessed in an increasing spiral extending from the global minimum, as illustrated in Fig. 4, with the black dot in the center representing the global minimum. Each pixel on the spirals is assessed starting from the inner most spiral and moving outwards.

This spiral technique combined with the adaptive threshold

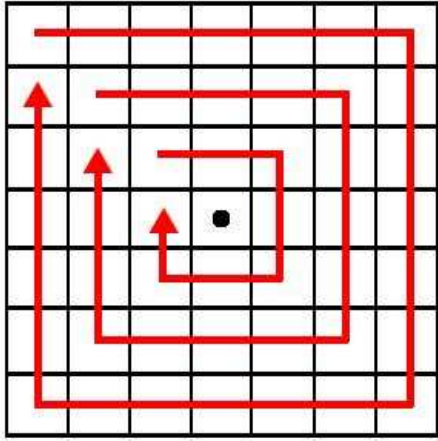


Fig. 4. Lazy Spiral Calculation

yields significant reductions in the processing time by ensuring that only strong landmarks have the full number of SAD calculations carried out. This method enables the processing of an image in 0.97 seconds.

E. Normal Form Calculation

The lazy spiral SAD calculation method relies on the fact that strong landmarks exhibit strong valleys through the middle of the distortion matrix as seen in Fig. 1. This leads to the local minimums almost invariably being located in a small area surrounding the global minimum.

We can incorporate this prior knowledge by restricting the lazy spiral to the immediately surrounding pixels when searching for the local minimum. This leads to significantly less calculations and achieves an average processing time of 0.17 seconds per image. This method is equivalent to using a search window size of template size plus two and is based on the interest operators proposed by Moravec[4].

The inclusion of the adaptive threshold in combination with the normal form calculation provides a novel refinement of currently existing landmark selection algorithms resulting in lower processing times while still providing a good set of strong landmarks.

IV. RESULTS

A. Speed

The results below summarise the speeds of the various combinations of optimisations, all numbers are in seconds.

The two bottom rows in Table I show the results where stepping and uniform area reduction have been used together, generally this is not good practice as the template stepping may step over a pixel that has been tagged for processing by the uniform area method. The pixel that is stepped over may be of importance and as can be seen from the results some landmarks do not feature in the methods that use uniform area reduction. It is encouraging to note that in most cases the

top four landmarks are included in the landmark set for all variations of optimisations.

TABLE I
SPEED COMPARISON

Step Size	Uniform Area	Adaptive Threshold	Search Types		
			X Y Scan	Lazy Spiral	Normal Form
1	N	N	123.58	115.89	4.57
1	N	Y	32.97	5.01	1.19
1	Y	N	9.30	8.58	0.42
1	Y	Y	4.97	0.85	0.26
3	N	N	14.28	13.23	0.52
3	N	Y	4.45	0.97	0.17
3	Y	N	1.24	1.16	0.07
3	Y	Y	0.82	0.36	0.05

It is clear from inspection of this table that the template stepping method is faster in all cases.

Within the variations of stepping and non-stepping algorithms, variations of adaptive thresholding and uniform area identification yield similar patterns of speed, with the uniform area identification and adaptive threshold being the quickest (see Table I).

B. Output

The figures below show the results over four image frames taken from a sequence five frames apart. The landmarks that have been selected by six different combinations of parameters have been plotted onto the image frame with a table below each figure showing the mapping of each point to each combination. The table is ordered with the strongest landmarks at the top progressing down to the weakest. The header of each table shows the combinations of template step size, uniform area reduction and the SAD calculation type used, all combinations use the adaptive threshold. When plotting these points a tolerance of ± 5 pixels has been used.

Fig. 5 contains double size images of the first ten landmarks that have been selected. By comparing the template image to the surrounding area in the full image, we can see that the template that has been selected is locally unique.



Fig. 5. Frame 1

TABLE II
FRAME 1 LANDMARKS

Step Size	1	1	1	3	3	3
Uniform	No	Yes	No	No	Yes	Yes
Search Type	LS	LS	NF	LS	LS	NF
1	1	1	1	1	1	1
2	2	2	2	2	3	3
3	3	3	3	3	2	2
4	4	4	4	4	4	4
5	5	7	5	5	7	7
6	6	8	6	6	8	8
7	7	11	7	7	15	15
8	8	12	8	8	6	6
9	9	13	9	9	16	16
10	10	14	10	10	17	17

The remaining three frames are shown to provide visual evidence of the repeated selection of landmarks from the indoor environment.

Below each frame the correlation image for the strongest landmark found in the frame is shown.

The strongest landmark is characterised by a white dot and cross centered on the location of the landmark template and shows that the template is strong compared to the immediately surrounding image and also exhibits the generic form of a strong landmark, as evidenced by the white cross.

There is no other noticeable white dots in the images indicating that the landmark template is also globally strong as well as locally strong.

It should also be noted that the correlation values range over 10,000 and this range has been compressed into 256 values for display reasons.

This image sequence has been taken as the robot is rounding a corner and then progressing down a corridor. We can see from the landmarks selected that the feature detector repeatedly selects the same regions when they are visible even under scale and rotation of the landmark regions relative to the cameras.

TABLE III
FRAME 5 LANDMARKS

Step Size	1	1	1	3	3	3
Uniform	No	Yes	No	No	Yes	Yes
Search Type	LS	LS	NF	LS	LS	NF
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	13	13
6	6	6	7	6	5	5
7	7	8	6	7	6	6
8	8	10	8	8	8	8
9	9	11	9	11	11	11
10	10	12	10	12	14	14

The results above show consistent landmark selection results over the frames, with six primary points being successfully selected in all images where those features are visible, with

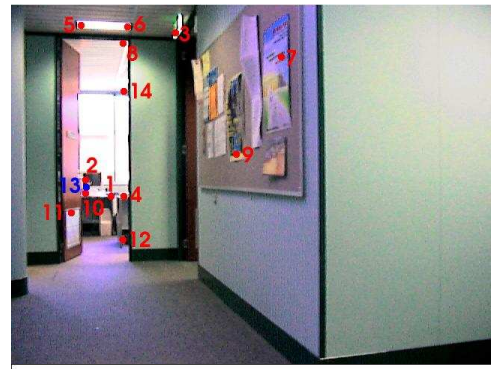


Fig. 6. Frame 5



Fig. 7. Frame 10

one feature on the light, two on the desk in the background, two on the posters to the right and one on the exit sign present in the upper right side of most images.

Although this experiment is not a landmark tracking

TABLE IV
FRAME 10 LANDMARKS

Step Size	1	1	1	3	3	3
Uniform	No	Yes	No	No	Yes	Yes
Search Type	LS	LS	NF	LS	LS	NF
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	5	5
5	5	5	5	5	10	10
6	6	7	6	6	11	11
7	7	11	7	8	12	12
8	8	12	8	7	7	7
9	9	13	9	9	14	16
10	10	10	10	10	15	14



Fig. 8. Frame 15

TABLE V
FRAME 15 LANDMARKS

Step Size	1	1	1	3	3	3
Uniform	No	Yes	No	No	Yes	Yes
Search Type	LS	LS	NF	LS	LS	NF
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	7	6	6	7	7
7	7	8	7	7	6	6
8	8	6	8	9	8	11
9	9	9	11	8	9	8
10	10	10	9	10	12	9

experiment, the fact that we consistently select the same objects and patterns in the environment from differing perspectives shows that the SAD correlation algorithm can be used to identify strong landmarks that are scale invariant over time.

V. CONCLUSION

Given the above results using the optimised SAD with template stepping, the lazy spiral calculation, no uniform area reduction and adaptive thresholding, we can extract natural indoor features over a series of frames in 0.17 seconds consistently.

The optimisation that uses template stepping, lazy spiral calculation, uniform area reduction and the adaptive thresholding produces a good set of features in 0.05 seconds per frame. While some of the features selected are not stable over time, the majority are. For real time navigation purposes this will provide a good set of landmarks to track.

Further efficiency can be achieved by scaling the input images though this tends to slightly reduce the quality of the feature set, however, the features still tend to be stable over time. Additionally our implementation has not been optimised for MMX however it is expected that this will further decrease the processing time.

ACKNOWLEDGMENT

National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence Program.

REFERENCES

- [1] G. Bianco and A. Zelinsky, *Biologically inspired visual landmark learning and navigation for mobile robots*, Proceedings of the 1999 IEEE/RSJ International Conference on Intelligent Robotic Systems (IROS '99), vol. 2, pp. 671-676, 1999.
- [2] T. Mori et al., *Trackable attention point generation based on classification of correlation value distribution*, JSME Annual Conference on Robotics and Mechatronics (ROBOMECH '95), pp. 1076-1079.
- [3] S. Thompson, *A Multi-Level Spatial Memory for Vision-Based Mobile Robot Localisation*, PhD Thesis, Australian National University, 2002.
- [4] H. Moravec, *Towards automatic visual obstacle avoidance*, Proceedings of the 5th International Joint Conference on Artificial Intelligence, Vol Vision-1, pp. 584.
- [5] D. Lowe, *Object Recognition from Local Scale-Invariant Features*, Proceedings of the 7th International Conference on Computer Vision (ICCV '99), pp. 1150-1157, 1999.
- [6] S. Se, D. Lowe, and J. Little, *Vision-based Mobile Robot Localization and Mapping using Scale-Invariant Features*, Proceedings of the IEEE International Conference on Robotics and Automation 2001 (ICRA '01), 2001.
- [7] J. Lewis, *Fast normalized cross-correlation*, In Vision Interface, 1995.