

Monocular Estimation of Spatial Displacement

Helio Perroni Filho (Doctorate Course, 1st year)^{†1}

Intelligent Robotics Laboratory, OHYA's group

Abstract— This article investigates how cues in monocular vision can be explored to estimate a robot's position relative to a set of visible landmarks. It presents a model for reasoning about the problem, a data set recorded to test the model, provisional solutions for the problems of selecting and comparing landmarks on images, and a set of statistical results computed from those elements. The results point towards the feasibility of visual estimation of landmark-relative displacement from monocular images.

Keywords: Image Processing, Machine Learning, Navigation

1. Introduction

Landmark-based navigation offers a solution to the problem of unrecoverable drift error accumulation, which is found in intrinsic localization methods such as odometry [1]. However it is not without its own open questions, such as how landmarks should be selected, how can they be identified, and how the robot's position relative to them can be computed. This last problem is specially difficult for monocular vision systems, where depth information is not immediately available. On the other hand, given the current ubiquity of monocular cameras, development of effective navigation methods for them could open the doors for a range of new applications.

This article investigates how cues in monocular vision can be exploited to estimate a robot's position relative to a set of visible landmarks. It starts by constructing a model for reasoning about variations in the perception of objects due to physical displacements. An hypothesis is formulated that it is possible to estimate physical displacements from measurements of such variations. A data set was collected to test the hypothesis: its features, and the procedure used for recording it, are reported. Next the practical problem of how to identify and compare landmarks in images of the data set is approached, and a provisional solution is devised. Statistics on the data set are calculated and analyzed in view of the proposed displacement model, and shown to be largely in agreement with the model's predictions. The article concludes with a discussion on directions for further research.

2. Displacement Estimation Model

Consider a robot equipped with a single front-mounted camera, wandering about an environment populated with stationary objects (or *landmarks*) by alternating among three *modes* of movement: straight (forward / backward), turning (clockwise / counter-clockwise) and sideways (left / right). As the robot's position changes relative to landmarks in its field of view, their images change in predictable ways:

- When the robot moves forward in a straight line, landmarks tend to appear bigger as they draw closer, and to drift towards the borders of the image as they are passed by (see Figure 1);
- When the robot turns in place, landmarks seem to move

horizontally in the opposite direction, but otherwise their images remain largely stable (see Figure 2);

- When the robot slides laterally, again landmarks appear to move in the opposite direction – but there is also a marked *motion parallax* effect, as landmarks closer to the robot seem to cross over the line of sight of those standing farther away (see Figure 3).

These observations suggest a model for relating *apparent* landmark movement in a series of pictures to the *actual* displacement of the robot picturing them.

Let $L = \{l_1, l_2, \dots, l_n\}$ be a set of uniquely identifiable landmarks. Given a landmark l_i , a *sighting* of l_i at time t_k is represented as the 4-tuple $g_{i,k} = (w_{i,k}, h_{i,k}, u_{i,k}, v_{i,k})$ containing two pairs of properties – the width ($w_{i,k}$) and height ($h_{i,k}$) of the image representation, and its horizontal ($u_{i,k}$) and vertical ($v_{i,k}$) positions in the whole picture. Figure 4 illustrates the concept.

Given a landmark l_i and two sightings $g_{i,a} = (w_{i,a}, h_{i,a}, u_{i,a}, v_{i,a})$ and $g_{i,b} = (w_{i,b}, h_{i,b}, u_{i,b}, v_{i,b})$ recorded at times t_a and t_b , the *absolute change* between the two sightings is represented as the 4-tuple $c_{i,a,b} = (\Delta w_{i,a,b}, \Delta h_{i,a,b}, \Delta u_{i,a,b}, \Delta v_{i,a,b})$, where $\Delta w_{i,a,b}$ and $\Delta h_{i,a,b}$ represent changes in width and height, $\Delta u_{i,a,b}$ and $\Delta v_{i,a,b}$ changes in position. Those values are calculated as below:

$$\Delta w_{i,a,b} = \frac{w_{i,b} - w_{i,a}}{\min(w_{i,a}, w_{i,b})} \quad (1)$$

$$\Delta h_{i,a,b} = \frac{h_{i,b} - h_{i,a}}{\min(h_{i,a}, h_{i,b})} \quad (2)$$

$$\Delta u_{i,a,b} = u_{i,b} - u_{i,a} \quad (3)$$

$$\Delta v_{i,a,b} = v_{i,b} - v_{i,a} \quad (4)$$

The formulas above have two important properties:

- 1) The value's sign is indicative of the *direction* of change – right/upward movements and size increments are positive, and left/downward movements and size decrements are negative;
- 2) The significance of a change in size must be taken relative to the original size of a sighting (e.g. an absolute increment that would double the area of a small sighting might be unremarkable for a larger one), so for width and height the *rate* of change is calculated instead of the simple difference.

^{†1} Graduate School of Systems and Information Engineering,
University of Tsukuba

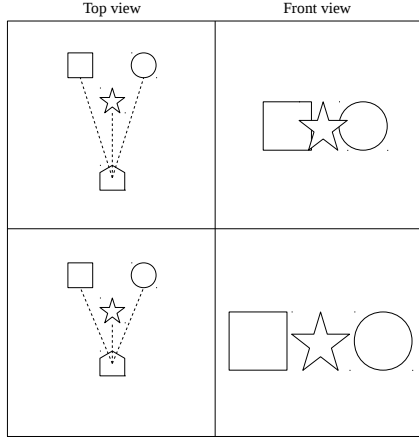


Fig. 1 When the robot moves forward in a straight line, landmarks tend to appear bigger as they draw closer, and to drift towards the borders of the image as they are passed by.

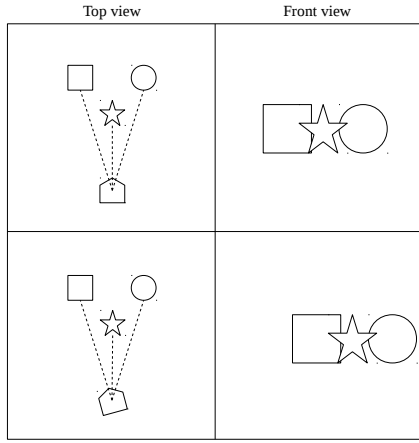


Fig. 2 When the robot turns in place, landmarks seem to move horizontally in the opposite direction.

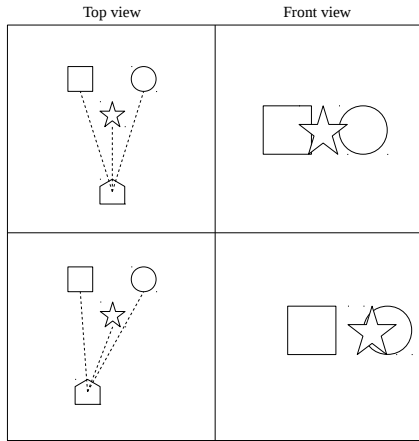


Fig. 3 When the robot slides laterally, landmarks appear to move horizontally, and due to motion parallax, closer landmarks may cross over the line of sight of those standing farther away.

Changes can also be measured relative to a pair of landmarks. Given two landmarks l_i and l_j , let $g_{i,a} = (w_{i,a}, h_{i,a}, u_{i,a}, v_{i,a})$ and $g_{j,a} = (w_{j,a}, h_{j,a}, u_{j,a}, v_{j,a})$ be sightings of l_i and l_j captured simultaneously at time t_a , and $g_{i,b} = (w_{i,b}, h_{i,b}, u_{i,b}, v_{i,b})$

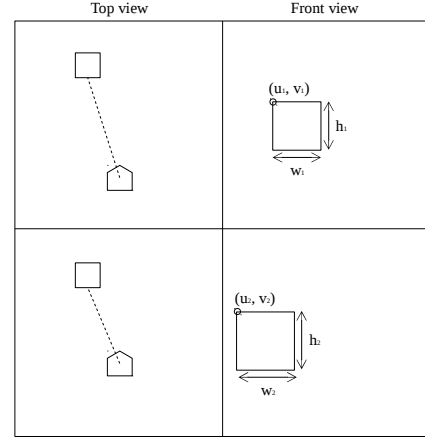


Fig. 4 Two separate sightings of the same landmark, with accompanying properties in the proposed model of displacement estimation. The width and height are not the landmark's actual physical dimensions, but merely those of its picture representations. Likewise, position coordinates are given relative to the upper-left corner of the whole image. Logically, for any given landmark, two sightings recorded from different spatial locations will yield different property values.

and $g_{j,b} = (w_{j,b}, h_{j,b}, u_{j,b}, v_{j,b})$ at time t_b . The *relative change* between l_i and l_j from time t_a to t_b is represented as the 4-tuple $r_{i,j,a,b} = (\Delta w_{i,j,a,b}, \Delta h_{i,j,a,b}, \Delta u_{i,j,a,b}, \Delta v_{i,j,a,b})$, whose values are calculated as below:

$$\Delta w_{i,j,a,b} = \frac{w_{i,b} - w_{j,b}}{w_{i,a} - w_{j,a}} \quad (5)$$

$$\Delta h_{i,j,a,b} = \frac{h_{i,b} - h_{j,b}}{h_{i,a} - h_{j,a}} \quad (6)$$

$$\Delta u_{i,j,a,b} = \frac{u_{i,b} - u_{j,b}}{u_{i,a} - u_{j,a}} \quad (7)$$

$$\Delta v_{i,j,a,b} = \frac{v_{i,b} - v_{j,b}}{v_{i,a} - v_{j,a}} \quad (8)$$

Compared to absolute changes, relative changes have the following properties:

- 1) The value's sign indicates whether there was an *inversion* in the relation between sightings relative to the measured property – e.g. if $u_{i,a} - u_{j,a} < 0$ but $u_{i,b} - u_{j,b} > 0$, it means that landmark l_i appeared to the left of l_j at time t_a , but was to the right of it at time t_b , and this inversion in relative position will result in $\Delta u_{i,j,a,b} < 0$ (conversely, if l_i remained to the left of l_j at t_b , then it would be that $\Delta u_{i,j,a,b} > 0$);
- 2) A value of $|\Delta \cdot| > 1.0$ indicates the disparity between sightings increased, whereas $|\Delta \cdot| < 1.0$ indicates the opposite – e.g. if $|u_{i,a} - u_{j,a}| < |u_{i,b} - u_{j,b}|$ it means the horizontal distance between sightings increased from t_a to t_b , and therefore $|\Delta u_{i,j,a,b}| > 1.0$ (conversely, had $|u_{i,a} - u_{j,a}| > |u_{i,b} - u_{j,b}|$ then it would be that $|\Delta u_{i,j,a,b}| < 1.0$).

Given a subset L'_k of landmarks visible at time t_k , a *scene* is defined as a sighting set $s_k = \{g_{i,k} | l_i \in L'_k\}$ recorded by the robot

at time t_k , while standing at a physical position $p_k = (x_k, y_k, \theta_k)$ relative to a universal reference point. Given two scenes s_a and s_b , the *difference* between them is defined as the pair $d_{a,b} = (C_{a,b}, R_{a,b})$ such that:

$$C_{a,b} = \{c_{i,a,b} \mid l_i \in L'_a \wedge l_i \in L'_b\} \quad (9)$$

$$R_{a,b} = \{r_{i,j,a,b} \mid l_i \in L'_a \wedge l_i \in L'_b, l_j \in L'_b\} \quad (10)$$

Finally, a *displacement* between two physical positions p_a and p_b is defined as $o_{a,b} = (x_b - x_a, y_b - y_a, \theta_b - \theta_a)$.

Given the model defined above, the *monocular displacement estimation hypothesis* states that a function Ψ exists such that:

$$o_{a,b} \approx \Psi(s_a, s_b) \mid L'_a \cap L'_b \neq \emptyset \quad (11)$$

That is, it is possible to design a function that given two scenes, approximates the displacement between the positions where they were recorded, provided they share any landmarks.

3. Data Set

The first step towards evaluating the monocular displacement estimation hypothesis is collecting a data base of images and associated positions. For that purpose an indoors environment was selected, and several objects scattered around it to serve as landmarks. See Figure 5 for a picture of this environment.



Fig. 5 Environment for collecting the data base of test images and associated positions. A collection of objects thought to be easily recognizable were scattered around to serve as landmarks.

In order to collect the images, a stock web camera was front-mounted to an Yamabiko robot. The robot was then placed at a distance of 2m from the front-most landmark, close and at a parallel angle to the right wall of the environment, and left to autonomously perform the following steps:

- 1) Turn 60° to the left;
- 2) Take a picture, record its position, then turn 10° to the right;
- 3) Repeat the previous step until collecting a total of 13 data points;
- 4) Turn left 60° (so it will again stand parallel to the side walls);

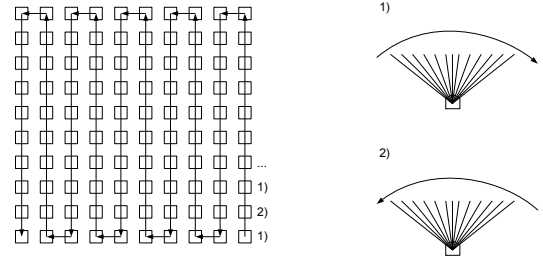


Fig. 6 Path traveled by the Yamabiko robot throughout the test environment. At each of the 100 stops 13 pictures were taken, amounting to a total of 1300 recorded pictures.

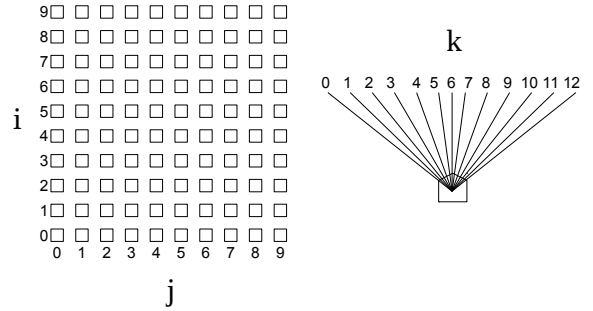


Fig. 7 Naming scheme for pictures in the test data base. Pictures at each stop are labeled with the forward and lateral distances from that stop to the origin at the rear-left, in decimeters; and also the distance from the leftmost orientation to the robot's orientation at the time, in tens of degrees.

- 5) Advance 10cm in a straight line;
- 6) Turn 60° to the right;
- 7) Take a picture, record its position, then turn 10° to the left;
- 8) Again, repeat the previous step until a total of 13 data points are collected;
- 9) Turn right 60°;
- 10) Advance another 10cm;
- 11) Repeat the steps above until advancing a total of 90cm;
- 12) Turn 90° to the left, advance 10cm, then turn 90° to the right;
- 13) Repeat steps 1-4, then move back 10cm;
- 14) Repeat steps 6-9, then move back another 10cm;
- 15) Repeat the last two steps until receding 90cm;
- 16) Repeat the pattern of movement above until moving 90cm to the left.

This path was designed so as to minimize the impact of random errors in the robot's odometer, making it so they partially cancel each other out. See Figure 6 for a schematic representa-

tion. The recorded image data base amounts to 1300 images of 640 X 480 pixels in size, taken at regular intervals of 10cm X 10cm X 10°.

Pictures were named in the form `still- i - j - k .png`, where i and j are respectively the forward and lateral distances (in decimeters) between the environment's origin point (at the rear-left end of the picture-taking area) and the robot's location at the time the picture was taken, and k the distance between the leftmost turning point and the robot's orientation at the time (in tens of degrees). See Figure 7 for an illustration.

4. Landmark Extraction and Comparison

Automatic selection of landmarks in sensory streams remains an open problem for robot navigation. Different approaches have been proposed, from employing an acquired global model of the environment to locate deviating local features [1], to statistical methods that attempt to maximize the utility of the extracted set [2], and even biological-inspired attention-based procedures [3]. For the purposes of evaluating the displacement model, a simpler strategy was devised: select the most "colorful" objects in each image as landmarks, where "colorful" is defined as the value of the Saturation channel of the HSL color space. Given an input image I , the procedure is performed as follows:

- 1) Calculate S_I , the HSL Saturation channel of I normalized across the range [0, 255];
- 2) Apply Otsu's method [4] to S_I , creating the binary image B_I ;
- 3) Apply an edge detection filter to S_I , creating the binary edge map E_I ;
- 4) Calculate the edge-subtracted binary image $G_I = B_I - E_I$;
- 5) Apply a morphological erosion operation to G_I , producing the foreground mask F_I ;
- 6) Apply a morphological dilation operation to F_I followed by the inversion of all bit values, producing the background mask R_I ;
- 7) Combine background and foreground masks in a gray-scale mask $M_I = 255 * F_I + 64 * R_I$;
- 8) Perform marker-based watershed segmentation [5] of I by M_I , producing the segment map P_I ;
- 9) Apply border-following segmentation [6] to P_I , extracting the list C_I of landmark contours;
- 10) Copy the pixels from I within each contour in C_I to new images, producing the list of landmark images L_I .

While it is certainly not without its misses, empirical tests with the data set presented in the previous section showed the procedure can identify a sufficient number of landmarks with high consistency, successfully recovering its outlines at different distances and camera orientations.

A procedure for measuring landmark similarity is also required, so that different object images recovered from distinct pictures can be recognized as pertaining to the same landmark. Again a simple procedure was devised:

- 1) Scale both images to dimensions equal to the average width and height between them;
- 2) For each of the (R, G, B) color channels plus the HSL Lightness channel, take the sum of the absolute pixel-wise difference between the scaled images;
- 3) Sum the channel-wise differences and divide by the scaled area.

One last problem remains: given two scenes $s_a = \{s_{a,1}, s_{a,2}, \dots, s_{a,m}\}$ and $s_b = \{s_{b,1}, s_{b,2}, \dots, s_{b,n}\}$ with possibly $m \neq n$, how to pair up the sightings in s_a and s_b that correspond to the same landmarks, and how to find which (if any) sightings in one scene have no pair in the other? Just selecting the pairs with highest similarity could cause one landmark from s_a to be paired with several landmarks from s_b or vice-versa – an obvious physical impossibility, as landmarks cannot be at two places at the same time.

The chosen solution was to construct an $m \times n$ matrix $S_{a,b}$ of pair-wise similarity measurements between s_a and s_b , then apply the Kuhn-Munkres assignment algorithm [7] to the matrix. In order to avoid pairings displaying incongruent changes (e.g. size reduction in straight forward motion, or rightward movement after a turn to the right) to be selected by the algorithm, any such cases are assigned the worst similarity value allowed by the computing environment, and discarded if nevertheless they get selected. The remaining set is used to compute the difference $d_{a,b}$ between the two scenes.

5. Statistical Analysis

If there is any truth to the monocular displacement estimation hypothesis, then differences between scenes must fulfill the following requirements:

- 1) Movement by different modes (e.g. straight v. sideways) must produce distinguishable variation patterns in change attributes;
- 2) Movement by different *amounts* on the same mode (e.g. straight 10cm v. 20cm) must produce variations in change attributes of similar patterns but different intensities.

In other words, in the very least it must be possible, by analyzing the differences between two scenes separated by a limited amount of single-mode movement, to identify what was the mode of movement, and estimate its amount. This can be verified by systematically comparing all pictures in the test data base, plotting their differences so as to allow quantitative comparison of movements falling in the conditions above. Some considerations on how to split and plot the data are in order, however.

Each change has four attributes (width, height, horizontal position and vertical position) which are expected to vary distinctly across differences; it's only logical, then, that a separate plot be drawn for each attribute. Moreover, the initial picture position of a landmark sighting may influence how different movements will affect future sightings – e.g. for straight forward movements, landmarks at the left and right sides tend to drift towards

the respective borders, while those at the center only drift down – therefore it makes sense to split results across nine vertical *origin bands*. Figure 8 illustrates this concept.



Fig. 8 The nine *origin bands* across which statistical results are split. Each band corresponds to a plot of the differences between sightings found at that band and their matches in another picture.

Obviously, for each movement mode, separate difference sets must be computed. Not all possible picture pairs need be compared for each mode – only those taken at locations that can be traversed on that mode alone. Moreover, results will clearly be symmetrical between directions (e.g. straight forward and backward directions will show the same results with inverted signs), so only one direction needs to be reported for each mode. Grouping results by pair-wise distance and computing the average allows band-wise results to be presented in a two-dimensional plot (distance X average change) instead of three-dimensional (distance X change instance X change value). Care must be taken, however, that as distance between locations increase, the number of valid image pairs differing by that amount decreases, and so statistics become less reliable. Figure 9 illustrates the problem.

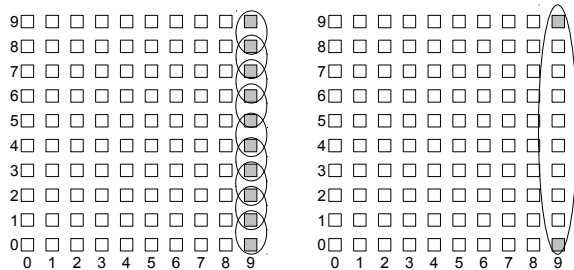


Fig. 9 Progressive depletion of test cases with distance. On column 9 of the data base, if only frontward orientation is considered, there are nine image pairs traversable by straight movement alone with a distance of 10cm (i.e. the pairs (0, 1), (1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7), (7, 8) and (8, 9)), but only one (the pair (0, 9)) with a distance of 90cm.

Figure 10 shows the result plots for absolute differences between straight forward movements, constructed in the manner discussed above. For each of the four change attributes – width, height, horizontal position (U) and vertical position (V) – a set

of nine plots was drawn, corresponding to the nine origin bands in the field of view (each band is identified by the horizontal position range of sightings that fall within it).

In each plot, the horizontal axis corresponds to the distance between pairs of pictures (in decimeters). Two curves are present at each plot: the first one (drawn in black), measured by the vertical scale on the left, shows the average change for each given distance (in pixels). The second (drawn in gray), measured by the scale on the right, shows how many sighting pairs were found at each given distance. This is an important information to add perspective to the results, since (as was noted before) as the number of available samples decrease, results become less reliable.

Looking at the plots, it's visible that for most bands, width and height increase steadily as the distance between locations increase. Also, as landmarks get close to the borders, the chance they won't be seen after the next forward movement quickly increases – so it's no surprise that results at the outer borders tend to be more erratic.

For the horizontal position (U) plot, sightings from right-side bands tend to change rightward, while left-side band sightings largely change leftward, and sightings at the central band display very little horizontal movement. The vertical position plot (V), in contrast, only shows upward movement on average – likely an artifact of the way samples are split across vertical bands, causing downward movement instances to be offset by upward ones.

Figure 11 shows the result plots for absolute differences between sideways rightward movements, structured exactly like the ones for straight forward movements. In this case most bands show negligible change in width, height and vertical position, while horizontal position decreases steadily as sightings shift towards the left border.

Figure 12 shows the result plots for absolute differences between turning clockwise movements. This set differs from the previous ones in that the horizontal axis of the plots measures distances in tens of degrees instead of decimeters, but otherwise the structure is the same. It also displays very similar results to the sideways case, with marked horizontal distance decreases between movements whereas other properties change little or too erratically to be considered.

Looking at the results above, it seems reasonable that straight movements could be estimated by changes in size and horizontal position of sightings, while turning and sideways movements could be estimated by horizontal position changes alone. The fact that the width and height of sightings hardly change between turning and sideways movements could also be used to distinguish them from straight movements. But to distinguish between sideways and turning movement, absolute differences alone don't seem to be enough. It is required, therefore, to look into relative difference plots to tell them apart.

Figure 14 shows the result plots for relative differences between turning rightward movements, and Figure 13, for relative differences between sideways rightward movements. In contrast

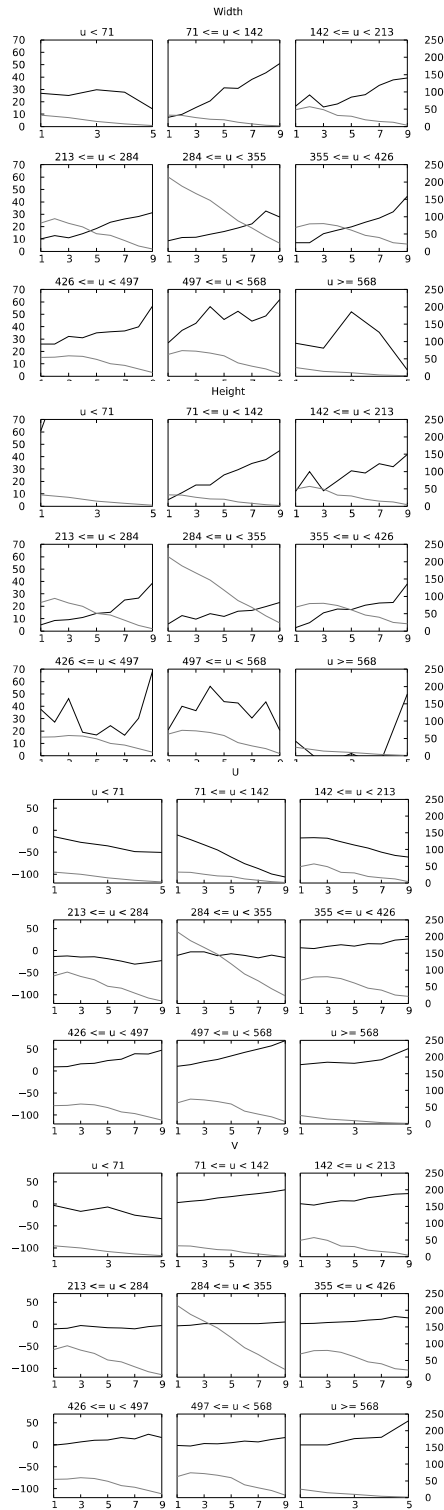


Fig. 10 Result plots for absolute differences between straight forward movements. For each of the four change attributes – width, height, horizontal position (U) and vertical position (V) – a set of nine plots is shown, corresponding to the nine origin bands in the field of view. Each band plot is identified by its horizontal position range. In each plot, the horizontal axis corresponds to the distance between pairs of pictures (in decimeters). Two curves are present at each plot: the first one (drawn in black) shows the average change for each given distance (in pixels). The second (drawn in gray) shows how many sighting pairs were found at each given distance.

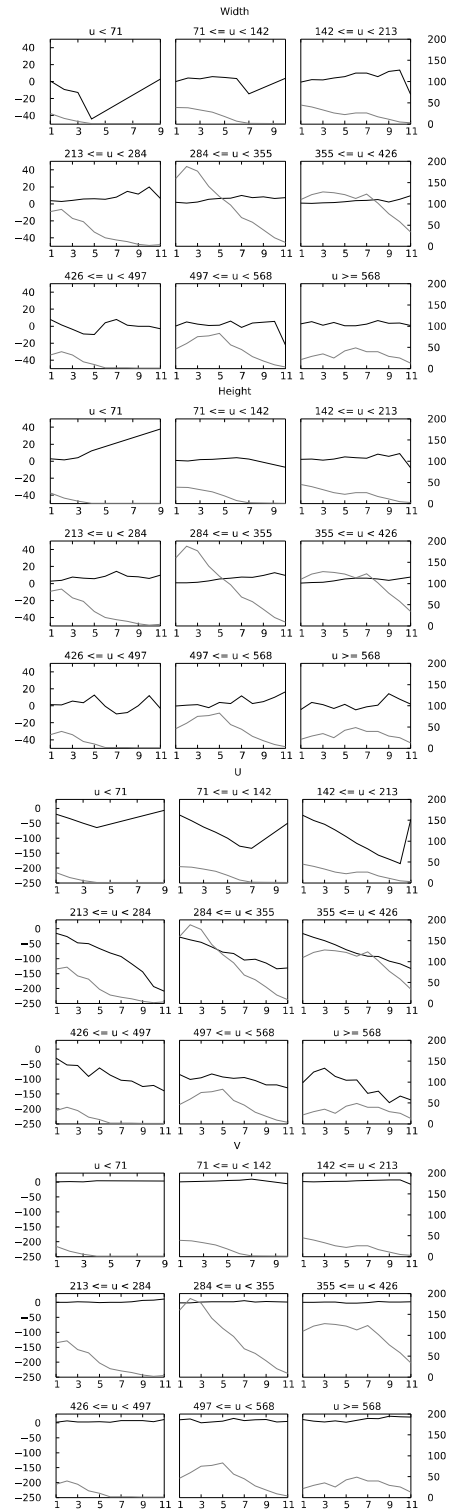


Fig. 11 Result plots for absolute differences between side-ways rightward movements. For each of the four change attributes – width, height, horizontal position (U) and vertical position (V) – a set of nine plots is shown, corresponding to the nine origin bands in the field of view. Each band plot is identified by its horizontal position range. In each plot, the horizontal axis corresponds to the distance between pairs of pictures (in decimeters). Two curves are present at each plot: the first one (drawn in black) shows the average change for each given distance (in pixels). The second (drawn in gray) shows how many sighting pairs were found at each given distance.

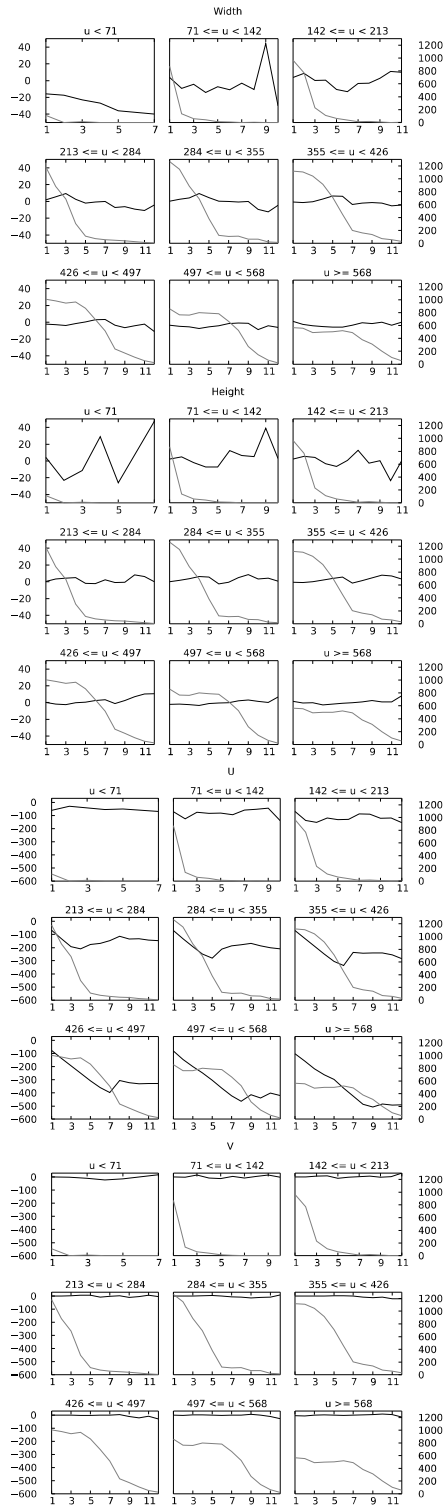


Fig. 12 Result plots for absolute differences between turning clockwise movements. For each of the four change attributes – width, height, horizontal position (U) and vertical position (V) – a set of nine plots is shown, corresponding to the nine origin bands in the field of view. Each band plot is identified by its horizontal position range. In each plot, the horizontal axis corresponds to the distance between pairs of pictures (in tens of degrees). Two curves are present at each plot: the first one (drawn in black) shows the average change for each given distance (in pixels). The second (drawn in gray) shows how many sighting pairs were found at each given distance.

to absolute plots, these are not split by bands: since the very principle of relative plots is that sightings at different image regions (and hence, bands) are compared to each other, it would make little sense to enforce this division. Otherwise their structure and measurement scales are the same as those used for absolute difference plots.

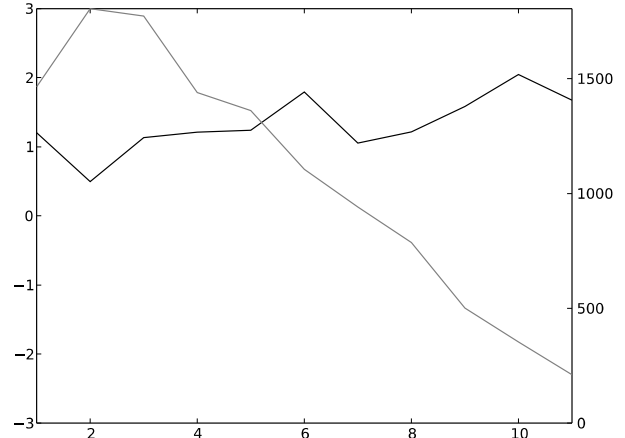


Fig. 13 Horizontal position (U) change plot for relative differences between sideways rightward movements. The horizontal axis corresponds to the distance between pairs of pictures (in tens of degrees). Two curves are present at each plot: the first one (drawn in black) shows the average change for each given distance (in pixels). The second (drawn in gray) shows how many sighting pairs were found at each given distance.

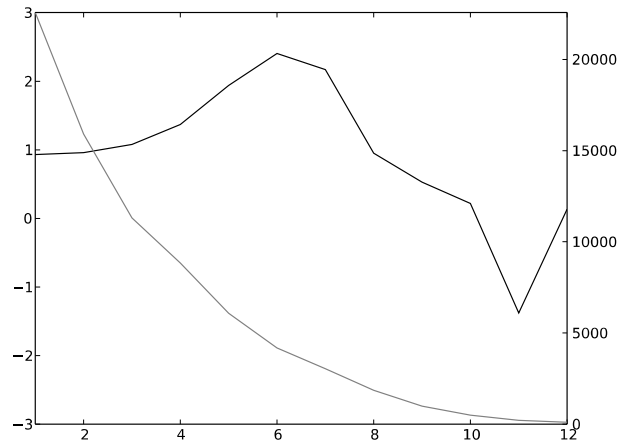


Fig. 14 Horizontal position (U) change plot for relative differences between turning clockwise movements. The horizontal axis corresponds to the distance between pairs of pictures (in tens of degrees). Two curves are present at each plot: the first one (drawn in black) shows the average change for each given distance (in pixels). The second (drawn in gray) shows how many sighting pairs were found at each given distance.

Comparing those plots, we can see that the relative change in horizontal positions increases steeply for turning movements (that is, until the depletion of test cases causes it to reverse course), whereas for sideways movements it alternates between increasing and decreasing cycles. This is in agreement with earlier observations on sighting changes for these two movement modes, and allows them to be told apart.

6. Conclusion

As a robot with a front-mounted camera wanders around an environment with recognizable stationary objects (or *landmarks*), their pictured images change in predictable ways. The *monocular displacement estimation hypothesis* proposes that those changes can be exploited as cues to estimate the direction and extent of physical displacement of the robot between successive camera takes. In order to verify this hypothesis, a test image data set was collected, and statistics compiled on it in accordance to the proposed visual displacement model. All data files and source code used to compute these results are available at [8].

Initial tests point towards a confirmation of the monocular displacement estimation hypothesis – i.e. it might in fact be possible to estimate a robot's physical displacement by analyzing apparent changes in position and size of landmarks captured by a front-mounted monocular camera. All single-mode movements correlate particularly well with horizontal position changes on average; straight movements, additionally, also correlate well to changes in width and height, which are negligible in the other modes. This fact makes it possible to tell straight movements apart; but telling apart turning and sideways movements requires that relative differences to horizontal position changes be also used.

One problem that was not taken into consideration is how to make sense of mixed-mode movements, i.e. those that combine straight, turning and / or sideways movements. In this case the problem is not determining what movement mode was employed, but to what extent (if any) each mode contributed to the movement. Further tests will be required to determine the effects of different mode mixes to change attributes, and consequently how to estimate physical movements from such changes.

The evaluation of supervised learning techniques as a way to construct an approximation of Ψ also arises as a logical next step. Preliminary experiments with VG-RAM networks [9] have shown promise in modeling the effects of mixed-mode movements to sighting changes, however estimation accuracy is limited, particularly at longer distances. The employment of more sophisticated landmark selection techniques and recording of a larger, more varied image data base will undoubtedly play an important role in such developments.

参考文献

- [1] S. Marsland, U. Nehmzow, and T. Duckett, "Learning to select distinctive landmarks for mobile robot navigation," *Robotics and Autonomous Systems*, vol. 37, no. 4, pp. 241–260, 2001.
- [2] S. Thrun, "Bayesian landmark learning for mobile robot localization," *Machine Learning*, vol. 33, no. 1, pp. 41–76, 1998.
- [3] A. Chella, I. Macaluso, and L. Riano, "Attention-based landmark selection in autonomous robotics," in *WAPCV*, ser. Lecture Notes in Computer Science, L. Paletta and E. Rome, Eds., vol. 4840. Springer, 2007, pp. 447–462. [Online]. Available: <http://dblp.uni-trier.de/db/conf/wapcv/wapcv2007.html#ChellaMR07>
- [4] N. Otsu, "A threshold selection method from gray-level histograms," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 9, no. 1, pp. 62–66, Jan 1979.

- [5] F. Meyer, "Color image segmentation," in *Proc. of the 4th Conference on Image Processing and its Applications*, 1992, pp. 302–306.
- [6] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985. [Online]. Available: <http://dblp.uni-trier.de/db/journals/cvgip/cvgip30.html#SuzukiA85>
- [7] H. W. Kuhn, "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1–2, pp. 83–97, March 1955.
- [8] H. P. Filho, "3rd Yamabiko Symposium 2013 Submission Files," 2014. [Online]. Available: <https://github.com/xperroni/Yamabiko/tree/master/2013-3>
- [9] H. P. Filho, "Weightless neural networks for depth estimation," in *2nd Yamabiko Symposium 2013*.