

Autonomous Exploration Using Rapid Perception of Low-Resolution Image Information

Vidya N. Murali and Stanley T. Birchfield
Electrical and Computer Engineering Department
Clemson University
Clemson, South Carolina 29634
Email: {vmurali, stb}@clemson.edu

Abstract

We present a technique for mobile robot exploration in unknown indoor environments using only a single forward-facing camera. Rather than processing all the data, the method intermittently examines only small 32×24 downsampled grayscale images. We show that for the task of indoor exploration the visual information is highly redundant, allowing successful navigation even using only a small fraction of the available data. The method keeps the robot centered in the corridor by estimating two state parameters: the orientation within the corridor, and the distance to the end of the corridor. The orientation is determined by combining the results of five complementary measures, while the estimated distance to the end combines the results of three complementary measures. These measures, which are predominantly information-theoretic, are analyzed independently, and the combined system is tested in several unknown corridor buildings exhibiting a wide variety of appearances, showing the sufficiency of low-resolution visual information for mobile robot exploration. Because the algorithm discards such a large percentage of the pixels both spatially and temporally, processing occurs at an average of 1000 frames per second, thus freeing the processor for other concurrent tasks.

1 Introduction

It has been known for some time that natural visual systems contain parallel pathways for recognition and guidance [46, 43, 36]. While *recognition* capabilities deteriorate severely under poor visual conditions such as low resolution or low illumination, *guidance* capabilities remain largely intact. Studying the behavior of human drivers under adverse conditions, Leibowitz *et al.* [21, 31, 6] proposed the “selective degradation hypothesis” to explain the fact that some visual abilities such as vehicle steering and speed control remain relatively easy despite significant loss in visual acuity and color vision. Their psychovisual experiments revealed that the subjective magnitude of *vection* (that is, the vivid sensation of self-motion induced by optical flow in the visual field) is unaffected by reductions of luminance of eight orders of magnitude and by refractive errors of up to 20 diopters. In contrast, virtually all aspects of focal vision (e.g., visual acuity, peak contrast sensitivity, and accommodation) deteriorate rapidly when light levels drop from daytime to night.

Motivated by these studies showing that high-resolution information is not needed to accomplish basic tasks such as navigation, we present a vision-based mobile robot system that uses only low-resolution (32×24) grayscale images from a single forward-facing camera to explore a previously unknown corridor environment. Our system also exploits temporal redundancy to process only a subset of image frames from the sequence, rather than processing the frames continuously. This reduction is motivated by psychological research into the limits of temporal resolution that show that subjects walking over even terrain sample the environment for only a fraction of the total travel time [38, 10, 32].

Falling within the general framework of minimalistic sensing [42, 30], there are several reasons for undertaking such a study. First, by restricting ourselves to such impoverished sensory data, it is possible to make

quantitative claims about how much information is needed to accomplish a given task. Secondly, the impoverished sensor necessarily limits the variety of algorithms that can be applied, thus providing focus and faster convergence to the research endeavor. Finally, the reduced amount of data available leads to greatly reduced processing times, which can be used either to facilitate the use of low-cost, low-power embedded processors, or to free up the processor to spend more cycles on higher-level focal vision tasks such as recognition.

We present solutions to two specific subproblems: estimating the robot’s orientation in the corridor, and estimating the distance to the end of the corridor. In both cases we propose a number of complementary measures, most of which are information-theoretic. After describing these individual measures, we present an integrated system that is able to autonomously explore an unknown indoor environment, recovering from difficult situations like corners, blank walls, and initial heading toward a wall. This approach extends our previous work [29], in which reactive behavior was demonstrated with a subset of measures, but without explicit estimation of orientation or distance to the end. All of this behavior is accomplished at a rate of 1000 Hz on a standard computer using only 0.02% of the pixels available from a standard 30 Hz color VGA (640×480) video camera, discarding 99.98% of the information.

2 Previous work

Researchers in computer vision, independently of the aforementioned psychophysical and psychovisual experiments, have also emphasized the importance of low-resolution vision. Torralba *et al.* [40], for example, have shown extensive results on a large database of 80 million images for the problems of non-parametric object and scene recognition. In one particularly noteworthy aspect of their work, results on person detection and localization using low-resolution images are comparable to those of the popular Viola-Jones detector [45]. Additional experiments conducted by Torralba and Sinha [41] have established lower bounds on image resolution needed for reliable discrimination between face and non-face patterns, indicating that the human visual system is surprisingly effective at detecting faces in low resolutions. Similarly, Hayashi and Hasegawa [12] have developed a method that achieves a face detection rate of 71% for as small as 6×6 face patterns extracted from larger images.

There is a connection between low-resolution vision and the notion of scale space [47, 22]. Scale space processing emphasizes extracting information from multiple scales and is the basis of popular feature detection algorithms such as SIFT [24] and SURF [4]. Koenderink [17] speaks of the “deep structure” within images, arguing that human visual system perceives images at several levels of resolution simultaneously. His distinction between deep and superficial structure is closely related to the distinction between focal (for recognition) and ambient (for guidance) vision [46].

Historically, low-resolution images have been used for various mobile robotic tasks because of the limitations of processing speeds. For example, in developing a tour-giving autonomous navigating robot, Horswill [15] used 64×48 images for navigation and 16×12 images for place recognition. Similarly, the ALVINN neural network controlled the autonomous CMU Navlab using just 30×32 images as input [33]. Robust obstacle avoidance was achieved by Lorigo *et al.* [23] using 64×48 images. In contrast to this historical work, our approach is driven not by hardware limitations but rather inspired by the limits of possibility, as in Torralba *et al.* [40, 41] and in Basu and Li [3], who argue that different resolutions should be used for different robotic tasks. Our work is unique in that we demonstrate autonomous navigation in unknown indoor environments using not only low-resolution images but also intermittent processing.

3 What resolution is needed?

To get a sense of the visible content in a typical corridor image, Fig. 1 shows an example image at successively downsampled resolutions. It can be seen that as the image is decreased in size from 640×480 to 32×24 , the corridor remains recognizable. However, at the resolution of 16×12 , a noticeable drop in recognizability occurs, in which it is difficult to discern that the image is of a corridor at all. This observation is confirmed by noting that the Fourier coefficients are dominated by the low-frequency terms.

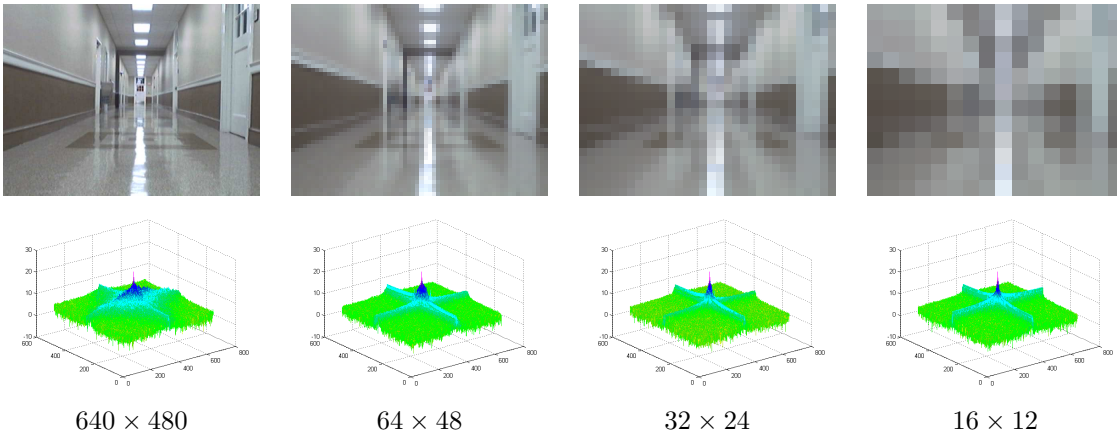


Figure 1: TOP: A typical corridor shown at different image resolutions. As the image resolution drops from 640×480 to 32×24 , the pattern of the corridor is still observable, but at 16×12 it is difficult to recognize the scene at all. BOTTOM: The corresponding Fourier coefficients (logarithmic display) show that the low frequency coefficients are more prominent.

To quantify these results, let $I : \Omega \rightarrow \mathcal{V}$ be a grayscale image, where $\mathbf{x} = (x, y) \in \Omega \subset \mathbb{R}^2$ are the coordinates of a pixel in the image plane, $v \in \mathcal{V} = \{0, \dots, 2^n - 1\}$ is a scalar intensity value, and n is the number of bits per pixel. If we assume that the pixel values in the image were drawn independently according to the probability mass function (PMF) $p(v)$, then we can say that

$$H(V; I) = \sum_{v \in \mathcal{V}} -p(v) \log p(v) \quad (1)$$

is a measure of the information content in the image, where $H(V; I)$ the entropy of a random variable V with PMF $p(v)$. Typically $p(v)$ is estimated by the normalized graylevel histogram of the image. Fig. 2 shows that for a variety of corridor environments, the entropy of an image does not change significantly as the image is downsized from 640×480 to 32×24 . In fact, at the latter resolution, there is only a 5% drop in entropy from the original image. However, as the resolution drops below 32×24 , the entropy drops sharply.

It should be emphasized that the entropy of the random variable associated with the graylevel values of the pixels in an image is not the only way to measure information content. As a result, the experiments above do not necessarily establish any validity to the claim that for a given task such a resolution is sufficient. However, we argue that there is a close connection between the coarse information needed for autonomous guidance (whether exploration or navigation) of a mobile robot and the more general problem of scene recognition. In both cases, the task is aimed at gleaning summary information from the entire image rather than discovering particular identities of objects in the scene. To this end, it is interesting to note that our results are similar to those of Torralba and colleagues. Their independent work on determining the spatial resolution limit for scene recognition [39, 40] has established through psychovisual experiments that 32×32 is sufficient for the identification of semantic categories of real world scenes. Our result of 32×24 is a close approximation governed by the desire to preserve the aspect ratio of the original image. As we shall see, these resolutions are corroborated by our own experiments on the specific task of autonomous exploration of a mobile robot.

4 Approach

Having considered the possibility of using low-resolution information, we now present a novel algorithm for estimating parameters necessary for autonomous exploration from a low-resolution 32×24 grayscale image. Although there are good theoretical reasons to smooth before downsampling, we adopt the extreme approach

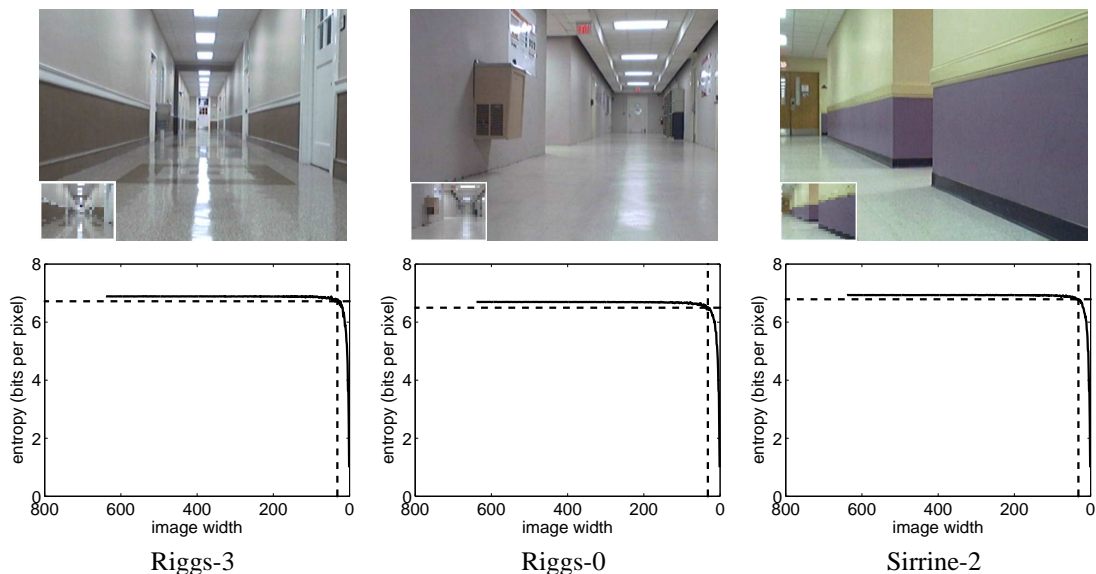


Figure 2: Plots of entropy versus image resolution for three different corridors. A loss of just 5% of information occurs as the image is downsized from 640×480 to 32×24 , but a sharp drop occurs at lower resolutions. For each image, the inset in the lower-left corner shows the 32×24 version (upsampled for display) in which the scene is clearly recognizable.

of simply downsampling the image without smoothing, both to test the limits of such an idea and to decrease significantly the computation time by avoiding most pixels entirely. (We did not observe much difference with or without smoothing.) Our system estimates two state parameters: the orientation of the robot in the corridor, and the distance to the end of the corridor. We now describe our solutions to these two problems.

4.1 Orientation in a corridor

In order for a mobile robot to autonomously maneuver through an indoor office environment, one obvious parameter that must be estimated is the robot’s orientation within the corridor, which is assumed to consist of straight parallel walls. We propose to combine five complementary ways of estimating this value from low-resolution images: the entropy of the image, the symmetry as measured by mutual information, aggregate phase, vanishing points using self-similarity of the image, and the median of the bright pixels. The goal is to learn a mapping $f : I \rightarrow \theta$, where I is the low-resolution image and θ is the orientation of the robot with respect the primary axis of the corridor.

4.1.1 Entropy

We have found empirically that, as a general rule, entropy is maximum when the camera is pointing down the corridor. The reason for this perhaps surprising result is that such an orientation causes scene surfaces from a variety of depths to be visible, yielding an increase of image information at this orientation. In contrast, when the robot is turned so that it faces one of the side walls, the range of visible depths is much smaller, and therefore the variety of pixel intensities usually decreases. A similar observation has been noted by other researchers in the context of using omnidirectional images [5, 11], but we show that the relationship between entropy and orientation holds even for standard camera geometries. In addition, we have found that the relationship is not significantly affected by whether the walls are textured.

We exploit this property by dividing the image into overlapping vertical slices and computing the graylevel

entropy of the image pixels in each slice. The horizontal coordinate yielding the maximum entropy is then an estimate of the orientation. More precisely, let us define a vertical slice of pixels centered at x as $\mathcal{C}_\omega(x) = \{(x', y') : \frac{\omega}{2} \leq |x - x'| < \frac{\omega}{2}\}$, where ω is the width of the slice. If $p(v; x)$ is the normalized histogram of pixel values in $\mathcal{C}_\omega(x)$, then the graylevel entropy of the slice is given by $H(V; I, x) = -\sum_{v \in \mathcal{V}} p(v; x) \log p(v; x)$. The orientation estimate is then given by $f_1(I) = \psi(\arg \max_x H(V; I, x))$, where the function ψ converts from pixels to degrees. With a flat image sensor and no lens distortion, the horizontal pixel coordinate is proportional to the tangent of the angle that the projection ray makes with the optical axis. Since the tangent function is approximately linear for angles less than 30 degrees, we approximate this transformation by applying a scalar factor: $\psi(x) = \alpha x$, where the factor α is determined empirically.

4.1.2 Symmetry by mutual information

Another property of corridors is that they tend to be symmetric about their primary axis. Various approaches to detecting and measuring symmetry have been proposed [48, 20, 2, 7]. However, in our problem domain it is important to measure the *amount* of symmetry rather than to simply detect axes of symmetry. One way to measure the amount of reflective symmetry about an axis is to compare the two regions on either side of the axis using mutual information. Mutual information is a measure of the amount of information that one random variable contains about another random variable, or equivalently, it is the reduction in the uncertainty of one random variable due to the knowledge of the other. Mutual information has emerged in recent years as an effective similarity measure for comparing images [16, 35, 13]. As with entropy, a column of pixels $\mathcal{C}(x)$ is considered for each horizontal coordinate x , where we have dropped the ω subscript for notational simplicity. The column is divided in half along its vertical center into two columns $\mathcal{C}_L(x)$ and $\mathcal{C}_R(x)$. The normalized graylevel histograms of these two regions are used as the two probability mass functions (PMFs), and the mutual information between the two functions is computed:

$$MI(x) = \sum_{v \in \mathcal{V}} \sum_{w \in \mathcal{V}} p(v, w; x) \log \frac{p(v, w; x)}{p_L(v; x)p_R(w; x)}, \quad (2)$$

where $p(v, w; x)$ is the joint PMF of the intensities in both sides, and $p_L(v; x)$ and $p_R(w; x)$ are the PMFs computed separately of the intensities of the two sides. As before, the orientation estimate is given by $f_2(I) = \psi(\arg \max_x MI(x))$.

4.1.3 Aggregate phase

A third property of corridors is that the dominant intensity edges tend to point down the length of the corridor. Therefore, near the center of the corridor, the phase angles of these edges on the left and right sides will balance each other, yielding a small sum when they are added together. We compute the gradient of the image using a Sobel operator and retain only the phase $\phi(x, y)$ of the gradient at each pixel. For each horizontal coordinate x we simply add the phase angle of all the pixels in the vertical slice: $AP(x) = \sum_{(x, y) \in \mathcal{C}(x)} \phi(x, y)$. The orientation estimate is given by $f_3(I) = \psi(\arg \min_x AP(x))$. Phase angles overlaid on several example images are shown in Figure 3.

4.1.4 Vanishing point using self-similarity

An additional property of corridors is the central vanishing point, which is nearly always present in the image when the robot is facing down the corridor. Our approach is based on the work of Kogan *et al.* [18], who developed a novel self-similarity based method for vanishing point estimation in man-made scenes. The key idea of their approach, based upon the work of Stentiford [37], is that a central vanishing point (meaning a vanishing point that is visible in the image) corresponds to the point around which the image is locally self similar under scaling changes. See Figure 4. While Kogan *et al.* [18] use 1D cross-sections of the image for similarity matching using affine transformation and cross correlation, we instead shift the downsampled image across the original image and calculate the mutual information between the two windows. The point at which

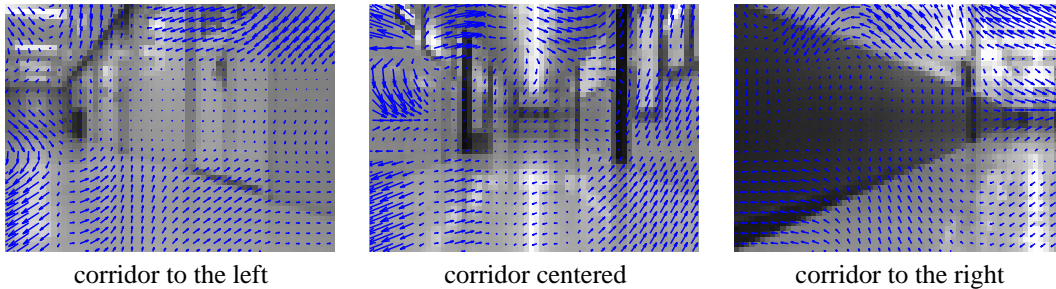


Figure 3: Gradient phase vectors overlaid on corridor images. From left to right: The center of the corridor is on the left side of the image, in the center of the image, and on the right side of the image. The phase vectors generally point toward the center of the corridor, so that in a vertical stripe near the center, the vectors balance each other.

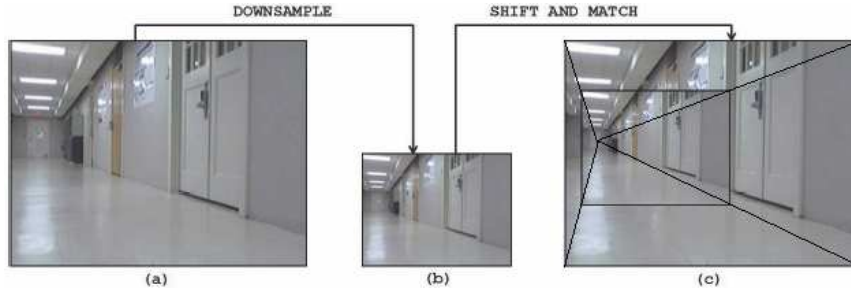


Figure 4: Vanishing point estimation from global self-similarity. (a) Original image of a corridor. (b) The image downsampled by a factor of two. (c) The downsampled image overlaid on the original image at the location of maximum self-similarity. The intersection of the lines connecting the corners of the two images yields the vanishing point. Normal resolution images are shown here only for display purposes; the actual algorithm uses low-resolution 32×24 images.

the mutual information between the two images is maximum yields a location for the downsampled image. The vanishing point is then found by intersecting the lines connecting the corners of the two images. Once we find the vanishing point, we discard the y coordinate, retaining only the x coordinate because our goal is to determine the robot's orientation within the corridor. This leads to $f_4(I) = \psi(\hat{x})$, where (\hat{x}, \hat{y}) is the intersection of the corner-connecting lines.

Our self-similarity approach has several advantages over existing techniques: It is simple, computationally efficient, and yields good results even for low-resolution images. Traditional techniques [26, 34] involve clustering detected lines, which performs poorly in low-resolution images because lines are not easily detected in such images. A more recent approach by Kong *et al.* [19] uses Gabor filters to yield texture estimates, and an adaptive voting scheme allows pixels to decide the confidence of an orientation rather than relying upon explicit line detection. Not only is their approach much more computationally intensive than ours, but with indoor low-resolution images the results are less accurate. See Figure 5 for some examples.

4.1.5 Median of bright pixels

The ceiling lights, which are usually symmetric with respect to the main corridor axis, provide another important cue. Due to the low resolution of the image, it is not possible to analyze the shape of the lights, as in [8]. Moreover, sometimes the lights are not in the center of the corridor but rather on the sides. A simple technique that overcomes these difficulties is to apply the k -means algorithm [25] to the graylevel values in the upper half



Figure 5: Comparison between our vanishing point estimation approach (green circle) using self-similarity and that of Kong *et al.* [19] (red plus). Our approach is more robust to the scenario of low texture information which is common in indoor scenes.

of the image, with $k = 2$. The median horizontal position of the brighter of the two regions is calculated, yielding an estimate of the center of the corridor. (The use of median as opposed to mean prevents the result from being affected by specular reflections on either wall.) We have found this approach to be not only simpler, but also more accurate and more generally applicable, than the shape-based technique in [8]. Note that ceiling lights provide an added advantage over vanishing points because they are affected by translation, thus enabling the robot to remain in the center of the corridor while also aligning its orientation with the walls. As with the previous measure, the horizontal coordinate is transformed to an angle by applying the same scalar factor. Therefore, $f_5(I) = \psi(\text{med}\{x : (x, y) \in \mathcal{R}_{\text{bright}}\})$, where $\mathcal{R}_{\text{bright}}$ is the set of bright pixels.

4.2 Distance to the end of the corridor

The second state parameter to be estimated is the distance to the end of the corridor. We assume a rectilinear structure, so that the end of the corridor is defined as the perpendicular, flat wall that one would encounter if continuing to travel along the corridor, either due to a dead end or a T- or L-junction. To solve this problem, we combine three complementary measures: Time-to-collision, Jeffrey divergence, and entropy.

4.2.1 Time-to-collision

Time-to-collision (TTC) is defined as the time it will take the center of projection of a camera to reach the opaque surface intersecting the optical axis, if the relative velocity between the camera and the surface remains constant. Traditional methods of computing TTC [1, 9] require computing the divergence of the estimated optical flow, which is not only computationally intensive but, more importantly, requires a significant amount of texture in the scene. To overcome these problems, Horn *et al.* [14] have recently described a *direct* method to determine the time-to-collision using image brightness derivatives (temporal and spatial) without any calibration, tracking, or optical flow estimation. The method computes the TTC using just two frames of a sequence, filtering the output using a median filter, to yield a reliable estimate as the camera approaches the object. This method is particularly applicable to our scenario in which the robot approaches a planar surface by translating in a direction parallel to the optical axis, a scenario for which the algorithm achieves an extremely simple formulation. Given two successive image frames $I^{(1)}$ and $I^{(2)}$ taken at different times, the TTC is computed as

$$\tau(I^{(1)}, I^{(2)}) = \frac{-\sum_{x,y} (G(x, y))^2}{\sum_{x,y} G(x, y) I_t(x, y)}, \quad (3)$$

where $G(x, y) = xI_x(x, y) + yI_y(x, y)$, I_x and I_y are the spatial derivatives of the image intensity function, and $I_t(x, y)$ is the temporal derivative. Normally, the summation would be computed over the desired planar object, but in our case we compute the sum over the entire image. Although the scene is not strictly planar when the robot is at the beginning of the corridor, we have found empirically that the TTC values are nevertheless higher at the beginning of the corridor, indicating that the method succeeds in estimate the TTC qualitatively even at larger distances. As the robot approaches the end of the corridor, the scene in the field of view becomes more planar, thereby increasing the accuracy of the estimated TTC. Since the formula for TTC yields a result in units of the time between image frames, to transform the TTC to an estimate of the distance to the end we multiply by the robot translational speed s divided by the camera frame rate f : $g_1(I^{(1)}, I^{(2)}) = (s/f) \cdot \tau(I^{(1)}, I^{(2)})$.

4.2.2 Jeffrey Divergence

As the robot approaches the end of the corridor, the pixel velocities increase, thereby causing the image to change more rapidly. As a result, another way to estimate the distance to the end is to measure the distance between two images. A convenient way to compare two images is to measure the Jeffrey divergence [44], which is a symmetric version of the Kullback-Leibler divergence:

$$J(p, q) = \sum_{v \in \mathcal{V}} \left(p(v) \log \left(\frac{p(v)}{q(v)} \right) + q(v) \log \left(\frac{q(v)}{p(v)} \right) \right), \quad (4)$$

where p and q are the graylevel histograms of the two successive images $I^{(1)}$ and $I^{(2)}$, respectively, and the summations are over the entire image. There is an inverse relationship between the divergence and the distance, so we transform this value to an estimate of the distance to the end by subtracting a scaled version from a constant to keep the result non-negative: $g_2(I^{(1)}, I^{(2)}) = \beta_2 - \alpha_2 J(p^{(1)}, q^{(2)})$, where β_2 is the offset.

4.2.3 Entropy

It is also true that, as the robot approaches the end of the corridor, the entropy of the image increases more rapidly. An alternate way to estimate the distance to the end, then, is to compute the difference in entropy between consecutive image frames, which also has an inverse relationship with distance: $g_3(I^{(1)}, I^{(2)}) = \alpha_3 / (H(V; I^{(1)}) - H(V; I^{(2)}))$, where α_3 is a scale factor. This difference in entropy is similar to the Jeffrey divergence, except that it eliminates the cross terms $p(v) \log q(v) - q(v) \log p(v)$, which have the most effect when the distributions are changing rapidly.

5 Experimental Results

To test the proposed approach, we analyze the accuracy of the various individual measures for the orientation and distance to the end. Then we describe the combined system and evaluate its performance exploring several unknown environments.

5.1 Orientation along the corridor

We collected data for 10 unique corridors in 5 different buildings on our campus, shown in Figure 6. Multiple corridors were used for the same building only when they were located on different floors and exhibited varying appearances. For every corridor, at equally spaced 15 feet intervals along the corridor we rotated the robot from -20 degrees to $+20$ degrees while collecting odometry data, laser readings, and images. The equipment used included an ActivMedia P3AT mobile robot, a SICK LMS-291 laser, and a forward-facing Logitech QuickCam Pro 4000 webcam. The robot was rotated at a speed of 2 degrees per second, and data was stored at the rate of 0.5 Hz, leading to densely sampled data approximately 1 degree apart. The laser provided depth readings in a 180-degree horizontal plane in increments of 1 degree, leading to 180 laser depth readings per sample time. The peak of these depth readings, after smoothing, was used to estimate the ground truth orientation, except for one corridor containing large specular surfaces (Lowry-0), where the heading was obtained from the odometry readings.

Figure 7 shows the orientation estimates of the different measures on four example images from different corridors. For each image, the overlaid vertical line indicates the estimate of the corridor centerline. There is in general wide agreement between the various measures, and their error with respect to ground truth is generally less than about 5 degrees. Overall, the median of the bright pixels yields the most accurate estimate, with the accuracy from entropy being only slightly degraded. The other three techniques also produce good results, but their accuracy is noticeably less. From the plots, it can also be seen that the median of the bright pixels yields a very sharp peak compared with the other measures. Due to space limitations, vanishing point is not shown in the figure, but its estimate is similar to the others except that it is computed directly.

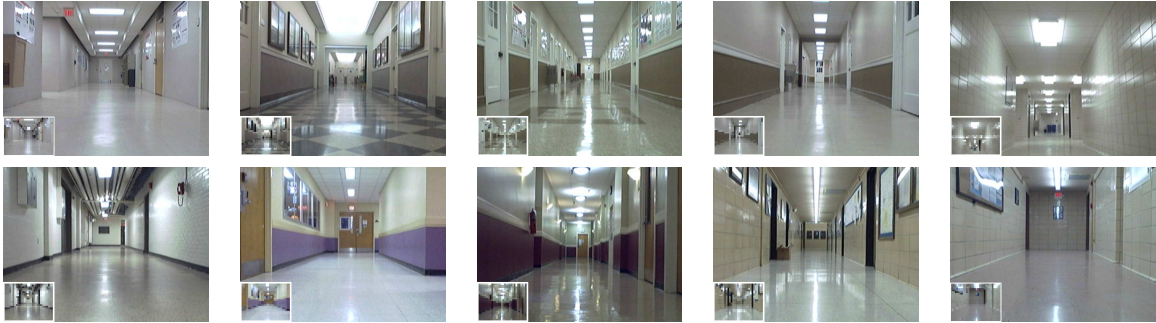


Figure 6: 10 distinct corridors in 5 different buildings. The inset shows the 32×24 downsampled version, upsampled to make it more visible. In lexicographic order, the corridors are Riggs-0, Riggs-1, Riggs-2, Riggs-3, Rhodes-4, Sirrine-0, Sirrine-2, Freeman-0, Lowry-0, Lowry-2, where the number indicates the floor.

To determine the broad applicability of the approach, Figure 8 shows the results of the five measures on the database. Half of the recorded environments (five corridors) were used for developing algorithm parameters, while the other half were used for testing. The error bars show the range $\pm 2\sigma$ capturing 95% of the data, where σ is the standard deviation of the error. As can be seen, the error for entropy is generally less than 10 degrees, the error for bright pixels is less than 5 degrees, and the error for the other techniques is less than approximately 15 degrees. For entropy and bright pixels, the error does not vary significantly across headings, while for the other three measures the error varies widely. Over the entire test database, the final weighted combination of all five measures yielded a root mean squared error of 7.2 degrees.

For driving a robot down the center of the corridor, the most important information is whether the orientation estimate is in the correct direction. In other words, the estimate should tell the robot to turn right when it is pointing to the left, and it should tell the robot to turn left when it is pointing to the right. Figure 9 shows the percentage of locations in which each of the five measures computes the correct sign for the orientation, with the center of the image defined as the zero heading. Again, outside a 5- to 10-degree range around the center of the corridor, both the entropy and bright light measures nearly always produce the correct direction.

5.2 Distance to the end

For the same 10 corridors mentioned earlier, we drove the robot along the corridor three times: down the center of the corridor, down the left side (1.5 feet from the center), and down the right side (1.5 feet from the center). While driving, the robot collected 640×480 images along with their corresponding 180-degree laser readings. The three measures for estimating the distance to the end were compared with ground truth, which was estimated from the central laser reading after median filtering.

Figure 10 shows the results of the three measures. After performing a linear fit to determine the two scale factors α_2 and α_3 , all three measures performed reasonably well at estimating the distance to the end, with time-to-collision being the most accurate. All measures performed more accurately as the robot approached the end of the corridor, until a distance of about 0.25 meters. Combined, the three measures yielded a root mean squared error of 0.49 meters over the entire test database.

5.3 Exploration in an unknown environment

The final system consisted of the mobile base equipped with a single forward-facing camera on the front. The only input to the system consisted of the 32×24 downsampled grayscale images from the 30 Hz camera. In our previous work we estimated orientation using only the median of bright pixels, and distance to the end of the corridor was largely determined by entropy [29, 27]. In this work we show that a linear combination (weighted average) of five (orientation) and three (distance to the end) complementary measures is more effective

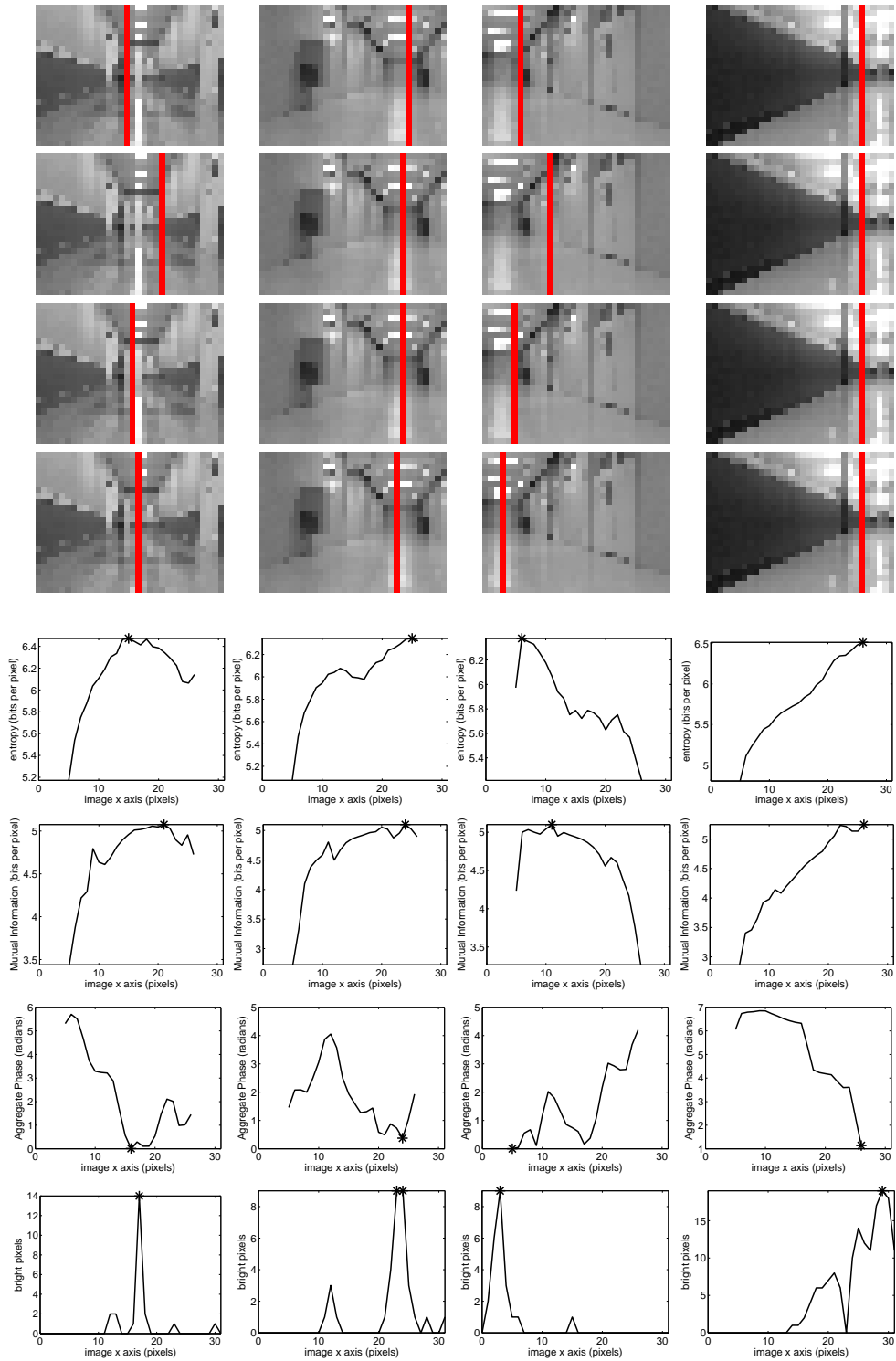


Figure 7: Estimating the center of the corridor using various cues. TOP ROWS: Four example tiny corridor images, with the pixel column corresponding to the center of the corridor overlaid for each of four different methods: entropy, symmetry, aggregate phase, and bright pixels. BOTTOM ROWS: Plots of the function computed.

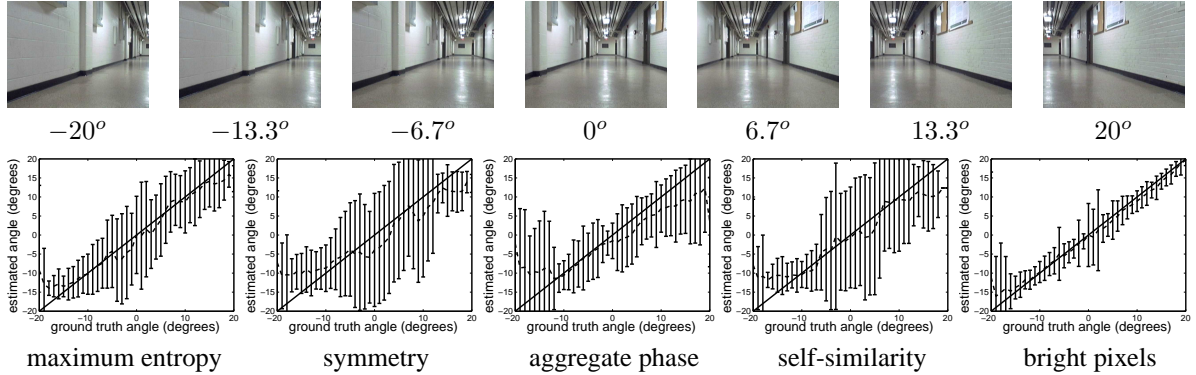


Figure 8: TOP: Sample images from the orientation experiment in one corridor. BOTTOM: Orientation estimate plots for the five measures on the database. Each plot shows, for each angle of the robot, the mean estimated angle across the database along with error bars indicating $\pm 2\sigma$, where σ is the standard deviation. The diagonal line is the ground truth. Note again the high accuracy of the bright pixels.

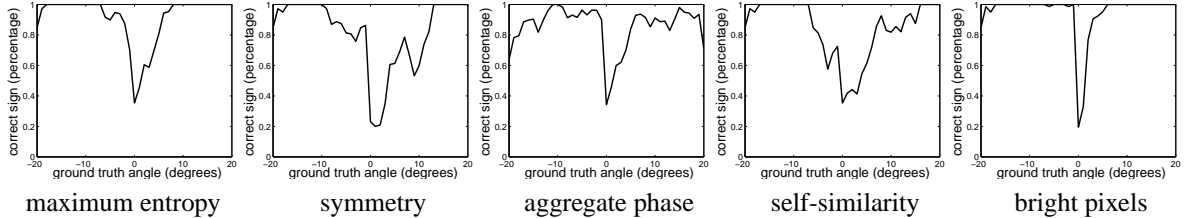


Figure 9: Plots of the orientation sign estimates for the five techniques. Each plot shows, for each angle of the robot, the percentage of images across the database in which the technique estimated the sign of the orientation correctly, with the center of the image corresponding to a heading of zero.

for achieving success in multiple environments. More sophisticated schemes for combining the estimates, like Kalman filtering [28], can produce even more reliable estimates but are omitted here due to space constraints.

The exploration capability of the robot was tested in several corridor environments. The robot did not have any knowledge of the corridors before the run, but some assumptions included rectilinear (Manhattan) corridor structure, opaque planar surfaces perpendicular to the direction of travel at the end of corridors, and the absence (or minimal effect) of sunlight shining through windows. (Although ceiling lights are not strictly speaking required by the algorithm, robustness is greatly increased by their presence; but whether they exist in the center or sides of the corridor is immaterial.) The goal of the robot was to autonomously explore the environment by repeatedly driving down the corridor and turning at the end. The turning direction at the end of a corridor was determined by attempting the two possibilities of turning right / turning left in an arbitrary order, using the entropy to distinguish between an open side corridor and a wall. In the event of a dead end, the robot made a 180-degree turn to return along the direction from which it came.

Results of the system showing successful end-to-end autonomous exploration in several different corridors are displayed in Figure 11. Only six of the ten corridors are shown to conserve space in the paper, but results for the other corridors were similar. The robot was able to stay in the center of the corridor, detect the end, stop, turn 90 degrees in the appropriate direction, and continue driving. The layout of the buildings, along with the path taken by the robot in each run, are shown in the figure. Note that the laser readings were used only to generate the plots, not to guide the robot. We also conducted experiments in which the robot started facing the wall, or started close to a wall; in both cases the robot corrected its orientation and position and continued exploring the environment. The longest successful exploration was a corridor in which the robot continuously ran for 45 minutes, navigating a total (overlapping) distance greater than 850 meters.

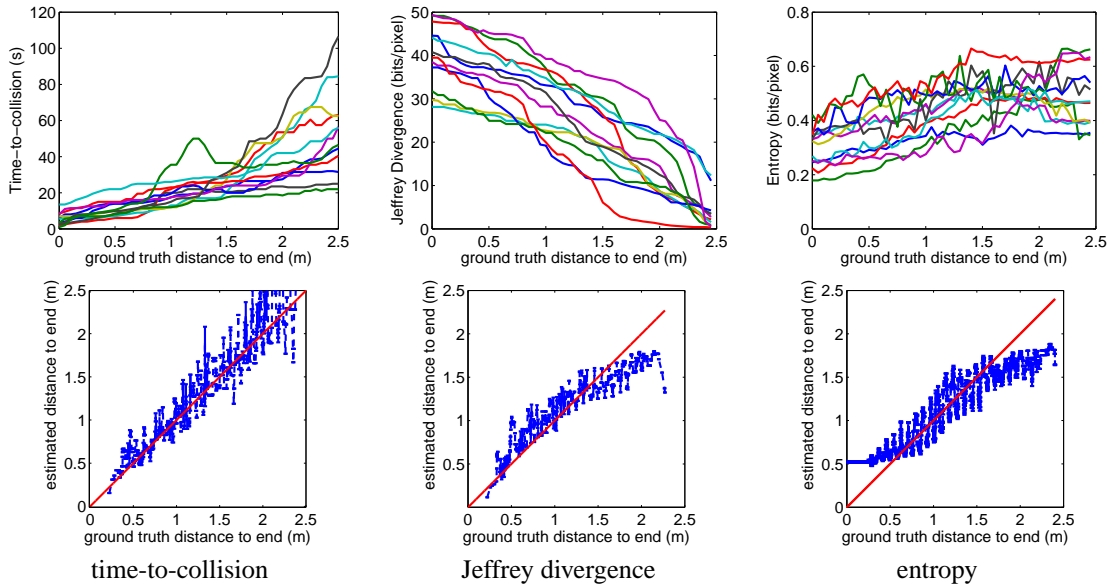


Figure 10: TOP: The estimated distance from the three measures plotted versus the distance from the robot to the end of the corridor. As the robot approaches the end, the entropy and time-to-collision decrease, while the Jeffrey divergence increases. Each line is a separate run of the robot along a different corridor / location within the corridor. BOTTOM: Estimated distance to the end of the corridor plotted against ground truth (obtained from the laser) for all three measures. Each vertical error bar indicates $\pm 2\sigma$, where σ is the standard deviation.

A few failure situations are also evident from the plots in the figure. In Riggs-0 and Lowry-0, the final corridor ends in a pair of transparent glass doors or reflective wall which confuse all the methods for determining the end of the corridor. In Riggs-1, the presence of a brightly textured vending machine in the final corridor causes erroneous estimates for the robot orientation. While the robot successfully explores other corridors containing vending machines, in this corridor the machine is located close to the turn so that it dominates the robot’s field of view immediately after the final left turn, giving the robot insufficient time to react. Images from these environments are shown in Figure 12. Overall, we have found that, when the assumptions are met, the robot is extremely reliable and can repeat runs with essentially 100% success. The primary failures we have observed occur because of glass doors or windows, large obstacles in the corridor (e.g., vending machines, rows of chairs), reflective surfaces on the walls, and narrow corridor openings. In addition to the ten corridors presented here, the robot was tested in a corridor in which one of the walls was lined with wall-to-ceiling windows. This was the only environment in which the robot was completely unsuccessful.

As the robot moves down the corridor, consecutive images usually differ from each other by only a small amount. To exploit this redundancy, we compare the entropy of consecutive images from the current frame t and previous frame $t - 1$, normalized by the first frame of the corridor (which presumably is near the maximum, since the entropy difference decreases as the robot travels down the corridor): $|H(V; I^{(t)}) - H(V; I^{(t-1)})| / H(V; I^{(1)})$. A histogram of these normalized entropy differences is shown in the top of Figure 13 for three different corridors, from which it is clear that most differences are small. As a result, our final system processes each image a minimal amount in order to determine whether the image needs to be fully processed by the five methods for orientation and three methods for distance-to-the-end. If the normalized entropy difference is less than 0.01 (10%), then no further processing is performed on the image. The results of this *rapid perception* are shown in the middle and bottom of the figure, from which it can be seen that nearly 80% of the images are only minimally processed, with the robot driving at 400 mm/s.

To perform the entire processing on a single image (five orientation plus three distance measures), the system takes 3.89 ms. However, the rapid perception module mentioned above quickly determines whether the current

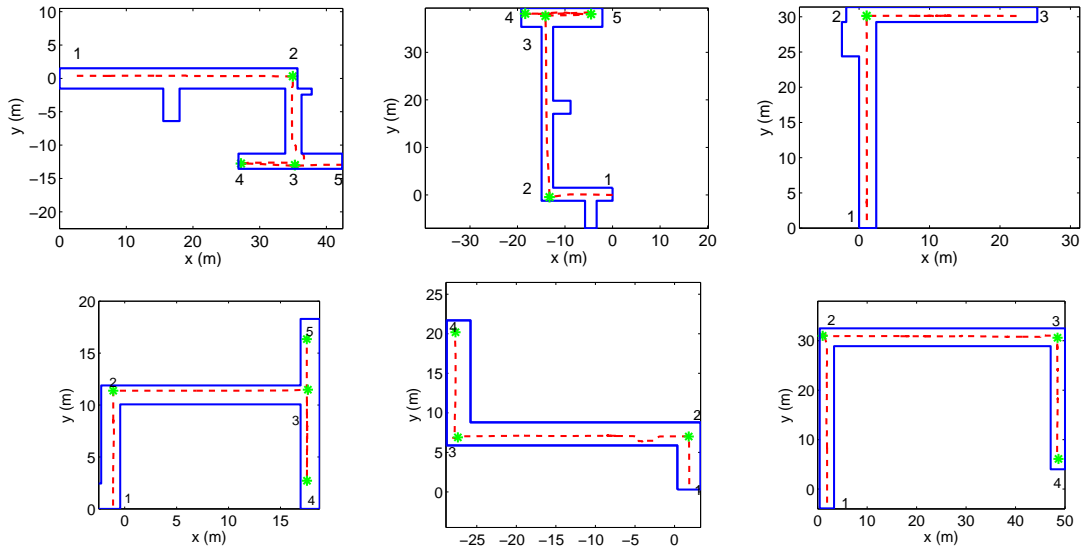


Figure 11: Exploration experiments for six different corridors, showing the path taken by the robot (red dashed, obtained from laser measurements) and the end-of-corridor detections (green asterisks) along different corridors (blue solid, manually measured). The numbering shows the temporal sequence of the robot’s locations from start to end. In lexicographic order, the corridors are Riggs-0, Riggs-3, Lowry-0, Freeman-0, Sirriner-0, and Sirriner-2.

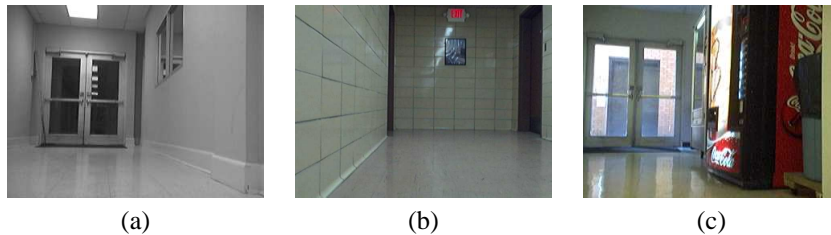


Figure 12: Problematic corridors: (a) reflective glass doors in Riggs-0, (b) reflective wall in Lowry-0, and (c) soda machine in Riggs-1.

image needs to be processed in its entirety. This module requires just 0.28 ms to make that decision, and 80% of the images are not processed further. Therefore, the time needed to process an image frame, on average, is $(0.8)(0.28) + (0.2)(3.89) = 1.002$ ms, which is equivalent to processing about 1000 frames per second. Stated another way, with a standard 30 Hz camera, the system consumes only 3% of the CPU, thus freeing the processor for other concurrent tasks that might be needed in a real system.

6 Conclusion

We have presented a low-resolution vision-based robot exploration system that can navigate down the center of a typical unknown indoor corridor, and turn at the end of the corridor. The exploratory behavior of a mobile robot is modeled by a set of visual percepts that work in conjunction to correct its path in an indoor environment based on different measures. Special emphasis is placed on using low-resolution images for computational efficiency, and on measures that capture information content that cannot be represented using traditional point features and methods. The resultant algorithm enables end-to-end navigation in indoor environments with self-directed decision making at corridor ends, without the use of any prior information or map.

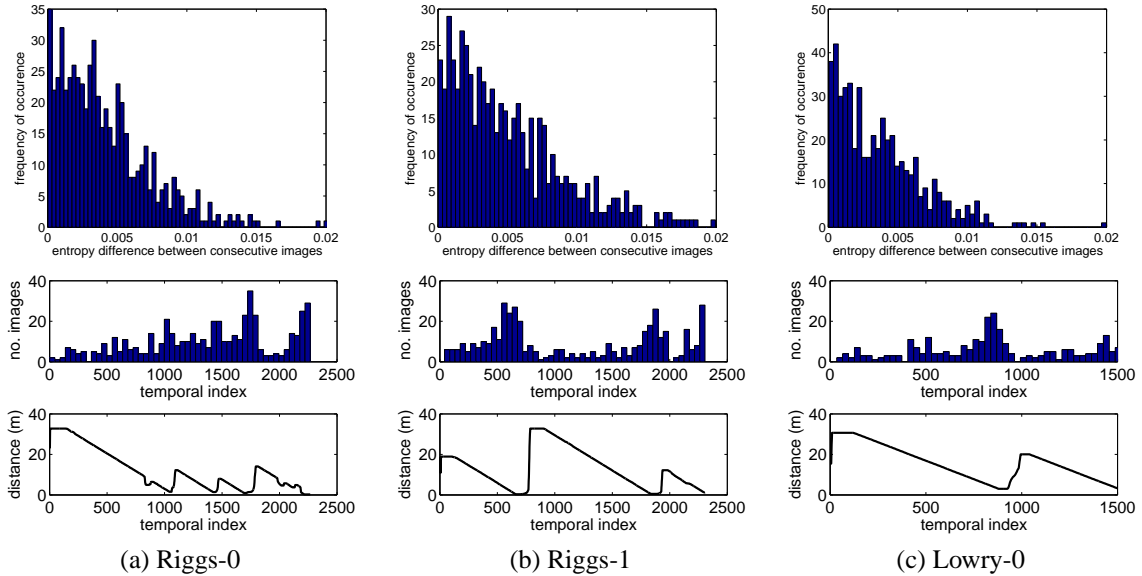


Figure 13: Rapid perception for three different corridors. TOP: Histogram of normalized entropy difference between consecutive images; most values are less than 0.01. MIDDLE: The number of images fully processed by the system versus the temporal index in the image sequence; most images are discarded after the rapid perception processing. Each bin captures 50 consecutive images, so that complete processing of all frames would be indicated by all bars reaching a height of 50. The percentage of frames fully processed, approximately 20%, is therefore the area under the curve divided by the area of the rectangle of height 50 and length equal to the number of frames in the sequence. BOTTOM: The distance from the robot to the end of the corridor versus the image index. By comparing the two plots for each corridor, note that when the robot nears the end of the corridor, the number of images fully processed increases. For all plots, the robot drove at 400 mm/s and was equipped with a 30 Hz camera.

The primary contribution of this work is the spatio-temporal compression of image information for computing navigation parameters of the robot, yielding high computational efficiency while maintaining robustness. With a 30 Hz VGA (640×480) camera, we discard 99.75% of pixels spatially (downsampling) and 80% of pixels temporally (rapid perception), leading to an algorithm that works on just 0.02% of the bytes available from the sensor. The advantage of such reduction is not only the tremendous computational efficiency which frees processor cycles to perform higher-level tasks such as recognition, but also the proof-of-concept regarding the amount of information needed for a particular task, in the spirit of minimalistic sensing.

There is much room for improvement in this line of work. First, the algorithm is challenged by specular or transparent surfaces, particularly when they dominate the field of view. However, this limitation does not seem to be intrinsic to the reduced amount of information available in the data, since a human viewer has no trouble interpreting the scene when watching the low-resolution video. Another improvement would be to combine the low-level exploration capabilities of such a robot with high-level recognition algorithms to provide a more detailed sense of the robot’s location within the environment by recognizing landmarks. Another open problem is to explore alternative ways of fusing measurements from multiple modules. Ultimately, we believe that minimalistic low-resolution sensing is a promising approach for low-level mobile robot tasks such as navigation and exploration.

7 Acknowledgments

The authors gratefully acknowledge the partial support of NSF grant IIS-1017007.

References

- [1] N. Ancona and T. Poggio. Optical flow from 1-D correlation: Application to a simple time-to-crash detector. *International Journal of Computer Vision*, 14(2):131–146, Mar. 1995.
- [2] M. J. Atallah. On symmetry detection. *IEEE Transactions on Computers*, 34(7):663–666, July 1985.
- [3] A. Basu and X. Li. A framework for variable-resolution vision. In *Proceedings of the International Conference on Computing and Information: Advances in Computing and Information*, pages 721–732, 1991.
- [4] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.
- [5] B. Bonev, M. Cazorla, and F. Escolano. Robot navigation behaviors based on omnidirectional vision and information theory. *Journal of Physical Agents*, 1(1):27–35, September 2007.
- [6] J. C. Brooks and D. A. Owens. Effects of luminance, blur, and tunnel vision on postural stability. *Journal of Vision*, 1(3):304, 2001.
- [7] M. Cho and K. Lee. Bilateral symmetry detection via symmetry-growing. In *Proceedings of the British Machine Vision Conference*, 2009.
- [8] K. Choi, S. Bae, Y. Lee, and C. Park. A lateral position and orientation estimating algorithm for the navigation of the vision-based wheeled mobile robot in a corridor. In *SICE 2003 Annual Conference*, volume 3, 2003.
- [9] D. Coombs, M. Herman, T.-H. Hong, and M. Nashman. Real-time obstacle avoidance using central flow divergence, and peripheral flow. *IEEE Transactions on Robotics and Automation*, 14(1):49–59, Feb. 1998.
- [10] J. Corlett. The role of vision in the planning and guidance of locomotion through the environment. In L. Proteau and D. Elliot, editors, *Advances in Psychology - Vision and Motor Control*, pages 375–397. Elsevier Science, 1992.
- [11] F. Escolano, B. Bonev, P. Suau, W. Aguilar, Y. Frauel, J. Saez, and M. Cazorla. Contextual visual localization: cascaded submap classification, optimized saliency detection, and fast view matching. In *IEEE International Conference on Intelligent Robots and Systems*, 2007.
- [12] S. Hayashi and O. Hasegawa. Detecting faces from low-resolution images. In *Proceedings of the 7th Asian Conference on Computer Vision*, pages 787–796, 2006.
- [13] H. Hirschmüller. Stereo processing by semi-global matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, Feb. 2008.
- [14] B. K. P. Horn, Y. Fang, and I. Masaki. Time to contact relative to a planar surface. *IEEE Intelligent Vehicles Symposium*, pages 68–74, June 2007.
- [15] I. D. Horswill. Polly: A vision-based artificial agent. In *Proceedings of the National Conference on Artificial Intelligence*, pages 824–829, 1993.
- [16] J. Kim, V. Kolmogorov, and R. Zabih. Visual correspondence using energy minimization and mutual information. In *Proceedings of the International Conference on Computer Vision*, 2003.
- [17] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984.
- [18] H. Kogan, R. Maurer, and R. Keshet. Vanishing points estimation by self-similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 755–761, 2009.
- [19] H. Kong, J.-Y. Audibert, and J. Ponce. General road detection from a single image. *IEEE Transactions on Image Processing*, 19(8):2211 – 2220, Aug. 2010.
- [20] P. Kovési. Symmetry and asymmetry from local phase. In *Tenth Australian Joint Conference on Artificial Intelligence*, pages 185–190, Dec. 1997.
- [21] H. W. Leibowitz, C. S. Rodemer, and J. Dichgans. The independence of dynamic spatial orientation from luminance and refractive error. *Perception & Psychophysics*, 25(2):75–79, Feb. 1979.
- [22] T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):224–270, 1994.
- [23] L. M. Lorigo, R. A. Brooks, and W. E. L. Grimson. Visually-guided obstacle avoidance in unstructured environments. In *IEEE Conference on Intelligent Robots and Systems*, volume 1, pages 373–379, Sept. 1997.
- [24] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [25] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.
- [26] G. McLean and D. Kotturi. Vanishing point detection by line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:1090–1095, 1995.

- [27] V. Murali. Autonomous navigation and mapping using monocular low-resolution grayscale vision. Master's thesis, Clemson University, Aug. 2008.
- [28] V. Murali. *Low-Resolution Vision for Autonomous Mobile Robots*. PhD thesis, Clemson University, Aug. 2011.
- [29] V. N. Murali and S. T. Birchfield. Autonomous navigation and mapping using monocular low-resolution grayscale vision. In *Workshop on Visual Localization for Mobile Platforms (in association with CVPR)*, June 2008.
- [30] J. M. O'Kane and S. M. LaValle. Almost-sensorless localization. In *Proc. IEEE International Conference on Robotics and Automation*, 2005.
- [31] D. A. Owens. Twilight vision and road safety. In J. Andre, D. A. Owens, and L. O. Harvey, Jr., editors, *Visual perception : The influence of H. W. Leibowitz*. Washington, DC: American Psychological Association, 2003.
- [32] A. E. Patla. Understanding the roles of vision in the control of human locomotion. *Gait and Posture*, 5(1):54–69, October 1997.
- [33] D. A. Pomerleau. Efficient training of artificial neural networks for autonomous navigation. *Neural Computation*, 3(1):88–97, 1991.
- [34] L. Quan and R. Mohr. Determining perspective structures using hierarchical Hough transform. *Pattern Recognition Letters*, 9(4):279 – 286, 1989.
- [35] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer. Image similarity using mutual information of regions. In *Proceedings of the 8th European Conference on Computer Vision*, pages 596–607, 2004.
- [36] G. E. Schneider. Two visual systems. *Science*, 163(3870):895–902, Feb. 1969.
- [37] F. Stentiford. Attention-based vanishing point detection. In *Proceedings of the IEEE International Conference on Image Processing*, pages 417–420, 2006.
- [38] J. A. Thomson. Is continuous visual monitoring necessary in visually guided locomotion? *Journal of Experimental Psychology: Human Perception and Performance*, 9(3):427–443, 1983.
- [39] A. Torralba. How many pixels make an image? *Visual Neuroscience*, 26(1):123–131, 2009.
- [40] A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1958 –1970, Nov. 2008.
- [41] A. Torralba and P. Sinha. Detecting faces in impoverished images. A.I. Memo 2001-28, MIT-AI, May 2001.
- [42] B. Tovar, L. Guilamo, and S. M. Lavalle. Gap navigation trees: Minimal representation for visibility-based tasks. In *Proceedings of the Workshop on the Algorithmic Foundations of Robotics*, pages 11–26, 2004.
- [43] C. B. Trevarthen. Two mechanisms of vision in primates. *Psychologische Forschung*, 31(4):299–337, 1968.
- [44] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1023–1029, Apr. 2000.
- [45] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [46] J. S. Werner and L. M. Chalupa. *The Visual Neurosciences*. The MIT Press, 2004.
- [47] A. Witkin. Scale space filtering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1983.
- [48] H. Zabrodsky, S. Peleg, and D. Avnir. Symmetry as a continuous feature. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12):1154–1166, Dec. 1995.