

Image Mapping and Visual Attention on a Sensory Ego-Sphere

Katherine Achim Fleming, Richard Alan Peters II, Robert E. Bodenheimer

Department of Electrical Engineering and Computer Science

Vanderbilt University

Nashville, TN

katherine.achim@vanderbilt.edu, rap2@vuse.vanderbilt.edu, bobbyb@vuse.vanderbilt.edu

Abstract— The Sensory Ego-Sphere (SES) is a short-term memory for a robot in the form of an egocentric, tessellated, spherical, sensory-motor map of the robot’s locale. This paper reports on: (1) the mapping to the SES of images with much higher resolution than the SES itself, and (2) the processing of visual attention on the SES. Described is a procedure to store spatially-overlapping imagery in the SES database and to generate and update a spherical composite of its visual contents. The composite image serves both as a visual map of the locale and as a representation of the local contents of the underlying full-resolution imagery. Visual attention enables fast alignment of overlapping images without warping or position optimization, since an attentional point (AP) on the composite typically corresponds to one on each of the collocated regions in the images. Such alignment speeds analysis of the multiple images of the area. Compositing and attention were performed two ways and compared: (1) APs were computed directly on the composite and not on the full resolution images until the time of retrieval. (2) The attentional operator was applied to all incoming imagery. Then the locations and values of the APs on the composite were computed as a function of those points within the corresponding regions of the collocated images. It was found that although the second method was slower, it produced consistent and, thereby, more useful APs.

I. INTRODUCTION

This paper reports methods for mapping images from a surrounding environment onto a robot’s Sensory Ego-Sphere (SES), and employing attentional methods to the result. The SES serves as a short term memory for a robot and as an interface to higher-level cognition, provided either by algorithms on the robot itself or by people teleoperating the robot [1]. Mapping images onto the SES is an important problem since the images can then be quickly recalled if needed, and applying attentional processes to these images is important in that it indicates where in the environment the robot or teleoperators may need to devote cognitive resources. Since the SES is a spatio-temporally organized database that records both sensory and motor information, image data, if mapped appropriately onto the SES, can provide resources for attention and cognition that supplement sensory-motor information.

The SES is a virtual tessellated sphere centered at the robot’s base frame, and its orientation remains fixed with respect to the world. As a robot moves its heading changes on the SES, and data can be moved from node to node based on the robot’s movements and how the data is accessed. The SES thus simplifies the organization, storage, and retrieval of egocentric information.

The tessellation used in the SES is geodesic and partitions space into a set of hexagonal or pentagonal cones that emanate from the frame. We use a tessellation frequency of $N=14$, which partitions space into 1963 regions such that the angle between nodes is on the order of $4.5^\circ - 5^\circ$. Each vertex in the tessellation (one per cone) corresponds to a node, or index, into a database where sensory and motor data related to the region is stored. For example, an object that has been visually identified in the environment is projected onto the sphere at azimuth and elevation angles that correspond to its location with respect to the SES frame. A label that identifies the object and other relevant information is stored into the database. The vertex on the sphere closest to an object’s projection becomes the registration node, or the location where the information is stored in the database, as illustrated in Fig. 1.

By virtue of its structure the SES has other capabilities including the fusion of multimodal, multiresolution sensory information that emanates from a single localized source [5] and mobile robot navigation [6]. It also provides a display of the robot’s locale-specific knowledge that can be of use to supervisors or remote operators of the robot as well as persons who interact with the robot.

A. Imagery on the SES

The SES has been used as a repository for defined objects, landmarks or sensory events (such as a sudden sound) that the robot has recognized and localized. Raw imagery was not stored even though it was captured periodically for various control processes. It can be advantageous to store such imagery temporarily for subsequent processing or comparison. An $N=14$ tessellation is too coarse for point-wise storage of high-resolution data such as imagery. However, blocks of high-resolution data (e.g.: a whole image) can be stored at each node. This paper describes two methods of doing so. Both make use of visual attention processing to simplify the matching of overlapping images.

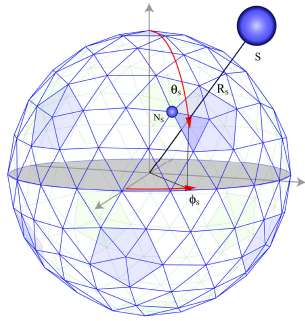


Fig. 1. Projection of an object onto the SES

If, while capturing images, the robot or its camera head rotates, the set of images will form a panorama of the area covered by the field of view (FoV). By storing the images on the SES, a short-term visual memory of the scene is retained. Each image is stored in the database at the node closest to the camera's optical axis at the time of capture. (The time is stored with the image.) Multiple images may be stored at the same node to cause a visual history of the location to be compiled. Or, if only the last view is needed a new image can overwrite an older one. By default, the SES stores the full image in the database. The FoV of ISAC's cameras are approximately 55° horizontal and 45° vertical. The $\sim 4.5^\circ$ spacing of the SES nodes results in considerable overlap between nodally adjacent images; in a fully imaged SES approximately 30 different images will overlap on any single node. Such redundancy may or may not be useful depending on the use of the imagery. If no (or minimal) overlap is required, the image size could be cropped to store only a 5° - 10° window centered on the optical axis.¹

It has been our experience that a graphical display of the contents of the SES can be quite useful to a person who is controlling or interacting with the robot. Previously, various objects, sounds, surfaces, or the locations of end-effector collisions with them, etc. have been displayed as icons on a computer graphic (CG) sphere. This has been extended by displaying a mosaic of imagery on the CG sphere. Such a display permits a person to guide the robot to objects or to avoid obstacles that the robot has not perceived. The mosaic is an approximate visual map of the locale. It is a conformal map of foveal images taken from the full resolution versions at the various nodes. These thumbnails are not warped or otherwise processed to present a smooth panorama since (we have observed) the rough set can convey in real-time or near real-time information useful to an operator. On ISAC, this results in a position error within $\pm 2.25^\circ$ in both directions. There are two formats for the mosaic: purely spherical and a spherical projection onto a plane. Both methods index into the database where the images are stored.

Current algorithms for image compositing and mosaic creation can create seamless panoramas [7], [8]. But these

¹ Such a window is commonly called a *fovea* in the computer vision literature. We will also refer to them as thumbnails even though they are cropped but not resized.

require both feature matching and image warping. The latter, especially, requires optimization and is computationally expensive. Some work on real-time image mosaicing has been done recently by Lu et al. [9]. They state, however, that, "Our method is efficient for small local distortions. For large frame-in-frame appearance transformation, our method can not guarantee the convergence, even with multiple iterations." The imagery projected onto the SES could exhibit small local distortions given a stationary robot and a relatively static environment. However, motion of the robot or objects in the environment often causes older imagery to differ significantly from recently acquired images. A presumption of small distortions cannot be made.

B. Visual Attention

The SES is most useful in a control system comprising parallel independent computational modules. If images are being captured, say for optical flow computation, and are subsequently stored on the SES rather than discarded, other modules can search them independently for information. For example, object recognizers can peruse the imagery for characteristic features. If one of them recognizes an object, its presence can then be marked on the SES. A module that detects change in the environment can compare incoming images to those already on the SES and flag differences. To maximize the usefulness to such computational modules, a set of images distributed across the SES should be (at least) loosely aligned with respect to their content. To minimize the computational complexity of alignment, easily discernible features should be used. We use a simple visual attention network to select common features in overlapping images.

Visual attention is a process that locates features in an image that have high information content so that limited computational resources can be directed toward them. Cave, [10] writes that attention "only allows a small part of the incoming sensory information to reach short-term memory and visual awareness, allowing us to break down the problem of scene understanding into rapid series of computationally less demanding, localized visual analysis problems." Cave assumes that information content is measured by visual *saliency*, a scalar measure of the conspicuousness or relevance of a pixel location. A saliency operator is defined in terms of those features that are important to a particular task.

For the purpose of image alignment, the choice of features is not as important as is the consistency of the attention algorithm. It should find the same features in separate images of the same area. If, however, the attentional algorithm is selected to find meaningful features, the benefit is two-fold: not only can the redundant images be aligned, but also features of interest can be located and tagged. We use Cave's *Feature-Gate* (FG) algorithm [10] because of its high flexibility in terms of feature selection. (The choice of visual feature sets can be varied on-the-fly and is independent of the algorithm structure.) With the appropriate features, FG has been demonstrated to mimic human visual attention [11] as well as to be programmable to favor specific objects [12]. FG operates in parallel on a set of feature maps extracted from an image. It returns a list of pixel locations, rank-ordered by saliency.

We tested the attention algorithm for two purposes: to find salient locations in the image mosaic and to select locations in images from the input stream that enable fast alignment with respect to features. We found that the discontinuities in the (unaligned) image mosaic generated spurious attention points. Attention applied to the input stream, however, was successful in aligning overlapping images.

II. IMPLEMENTATION

A. Visual Attention

In the FG model of human visual attention, each location in the visual scene has a vector of basic features, such as orientation or color, as well as an attentional gate that regulates the flow of information from the location to the output. The gated flow depends on that location’s features and the features of surrounding locations. The visual scene is partitioned into neighborhoods; a “winning” location in each neighborhood is passed to the next level. This proceeds iteratively until there is only one location remaining, the output of the model. FG contains two subsystems to handle bottom-up and top-down attentional mechanisms. The bottom-up process identifies the most salient location in the scene based on feature-value differences, and is independent of the task. The top-down process is task-related, a function of a *target* feature vector. We did not use the top-down procedure because we were not looking for a single location with a specific feature set. Instead, we wanted to find a set of the most discernible locations in each image.

Saliency is computed as the sum of the Euclidean distances between the feature vector, $\mathbf{v}(x,y)$, at pixel (x,y) and the feature vectors of its eight neighbors, $\mathbf{v}(x_k, y_k)$,

$$s(x, y) = \sum_{k=1}^8 \|\mathbf{v}(x, y) - \mathbf{v}(x_k, y_k)\|$$

The saliency image is partitioned into 2×2 blocks and the largest value in the block is passed to the next level. FG presumes that the greater the distance between the feature vector of a pixel and those of its neighbors, the more conspicuous that pixel is. Top-down control would have the effect of further limiting the number of attentional points.

B. Populating the SES

This work involved the mapping to the SES of 320×240 (col \times row) color images taken by the active pan-tilt camera-head of ISAC. A sequence of 519 images was generated using one camera by capturing an image at each of the pan-tilt angles that corresponded to an SES node. The entire image was stored in the database at the node. The images were not preprocessed and no particular objects were identified. Since ISAC’s cameras do not rotate through 360 degrees, they do not map the entire SES. A connected subset of the SES within the area of $+20$ to -60 degrees in tilt and $+80$ to -80 degrees in pan was populated. This range was chosen because the $\pm 80^\circ$ pan range is consistent with the human field of view [13]. The result was a complete mapping onto the SES of the forward visual scene.

B. Mosaic Construction

To create the mosaic, a foveal window at the optical center of each image was extracted and posted on the SES at the corresponding node location. Fig. 2 illustrates this procedure. The exact size of each window depends on the specific node. Although similar, not all the edges on a geodesic dome have the same length. For more precise results, the distances between each node and its 4 closest neighbors (top, down, left, and right) were calculated in degrees and converted to distance in pixels. An appropriately-sized fovea was then extracted from the center of the image and was posted on the SES at the node corresponding to its pan-tilt angle pair. Fig. 3 shows a spherical representation of all the foveal images posted on the Sensory Ego-Sphere with respect to ISAC’s camera frame.

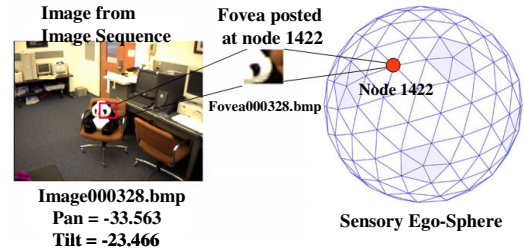


Fig. 2. Posting a fovea onto the SES

A mosaic image of the scene was reconstructed from all foveal images posted on the SES. A node map that associates each pixel in the mosaic image with a node on the SES was also generated. A flat mosaic image is illustrated in Fig. 4. Note that if the mosaic is used only as a visualization tool, it need not be constructed on the robot. The robot can transmit the thumbnails to a remote workstation where the mosaic can be generated.

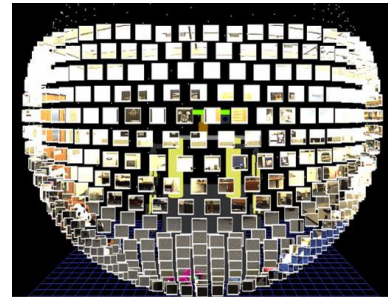


Fig. 3. Thumbnails posted on a robot’s SES

C. Attentional processing on individual images

The problem of attention arises once the SES is populated with dense information. Two procedures were tested. One was to apply FG directly to the SES image mosaic (*the mosaic method*). This was relatively fast but it found false discontinuities. Moreover, it cannot align overlapping images. The other approach (*the whole image method*) was to detect salient points within the individual images as they were captured and to combine them with the imagery that is already stored in the database. This simultaneously leverages the information in the overlapping images and estimates their overlap.

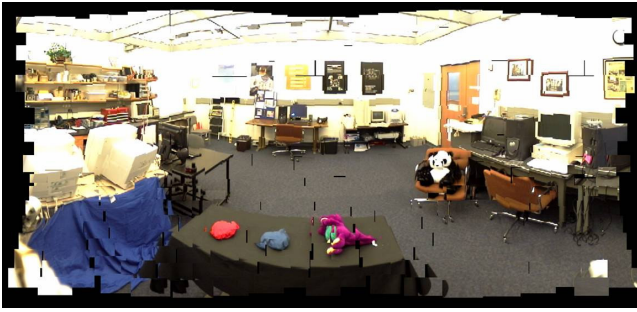


Fig. 4. Mosaic image constructed from SES thumbnails.

Attentional processing was performed using FG on each image in the sequence. The images were first blurred with a constant filter to reduce the sensitivity of the algorithm to insignificant changes. We implemented FG, with 12 feature-maps: the 9 Frei-Chen features [14], except for 3x3 average value, R, G, & B color intensities, and pixel intensity. The 12 most salient locations in each image were recorded as attentional points (AP) at the node corresponding to the optical center of the image. The locations were stored together with their saliency values and the pan and tilt angles of the image. 12 points were chosen because it was found that this usually resulted in a uniform distribution of APs throughout the image.

In Section I-A, we stated that as many as 30 images can overlap on a single SES node. The overlap implies that APs from different images will often refer to the same location in space. We estimated the visual importance of the various directions by computing one salience value for each node of the SES. This was done by combining the saliencies of all APs in the regions of the overlapping images that intersected the foveal window at the node. It was presumed that an AP that occurs in more than one image is more important (and should, therefore, have a higher value) than an AP found in only one image. The absolute SES coordinates of each AP was computed by converting its distance in pixels from the image center to pan and tilt angles. To those were added the angular location of the SES node. Due to localization errors, APs from the same feature can be mapped to adjacent nodes. For each node with a least 15 associated APs, the median pan and tilt angles of the APs were calculated. All APs within a 2° radius from the median were then mapped to the same node on the SES. If more than one node fell within the circle, then all the APs were mapped to the node with the most APs.

TABLE I
ALL ATTENTIONAL POINTS THAT MAP TO NODE 1422

Img CtrID	Saliency	Row	Col	ID	New pan	New tilt
1302	3528.456	197	146	1421	-38.769	-26.631
1626	4406/089	47	212	1421	-37.660	-26.918
1624	3865.287	41	140	1421	-39.610	-25.835
1421	3819.206	137	161	1421	-38.602	-26.537
1682	4790.870	26	236	1421	-37.308	-27.323
1340	3567.101	173	134	1421	-39.200	-26.030
1424	4096.694	131	233	1421	-36.692	-27.320
1679	4030.104	17	116	1421	-39.962	-25.698
1501	4254.137	98	236	1421	-36.789	-27.576
1303	4170.348	197	173	1421	-38.141	-26.680
1733	4671.133	5	266	1421	-37.252	-27.576

A 2° radius was selected because it represents one-fourth of the average fovea and was compact enough to isolate point clusters. An example of this is illustrated in Table I, which shows all original images (imgCtrID column) with an AP that maps to node 1421 (ID column) on the SES as well as each attentional point's calculated pan and tilt angles.

To determine the saliency of a node, the set of numerical saliency values from each AP posted at the node was summed. Fig. 5 shows the top 12 most salient locations (APs) in the scene (maxima over all the nodes).

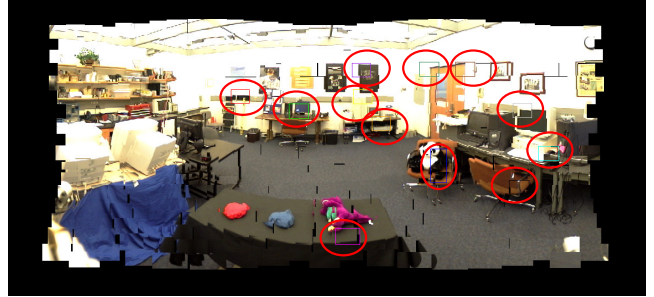


Fig. 5. Top 12 APs in scene by activation summation.

D. Attentional Processing on the Mosaic Image

We also tested the direct application of FG to the mosaic image. This process is much faster than the other because only one foveal image per node is processed rather than the entire set of complete images. The FG algorithm was modified to include a node map of the mosaic image – an array containing the node ID of each pixel location in the image. Thus FG returned not only the salient locations but also their node numbers. The results for the same image set used in the first approach are shown in Fig.6.

IV. RESULTS

A. Activation Summation

Fig. 7 is a graph of activation threshold versus number of nodes. It shows the number of nodes above threshold for values ranging from the minimum to the maximum total activation values per node. 672 nodes had attentional locations. Several thresholds were chosen and the percentage of nodes with activation above threshold level was computed. The first three columns of Table II indicate the activation levels necessary for a node to be a significant attentional location on the entire SES. For example, to be in the top 10% of attentional locations, a node would have to have a summed activation value of at least 100000.

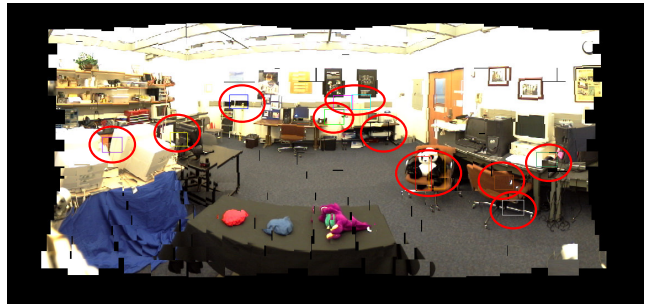


Fig.6. Top 12 APs by processing the mosaic image.

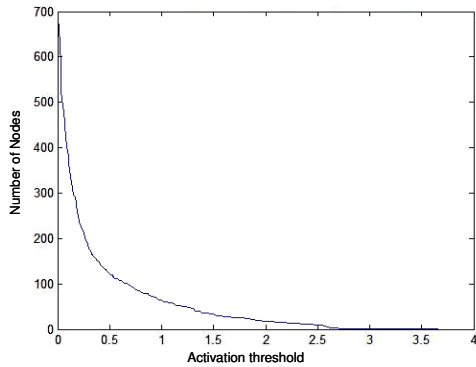


Fig. 7. Number of nodes above specific activation thresholds.

Another way to determine the attentional importance of a node with respect to the entire SES is to calculate the percentage of individual APs with above-threshold activation that map to the node. There were a total of 6228 APs on the SES. The calculations were performed for several thresholds. For example, if the nodes with activation values in the top 10% are chosen (threshold of 100000), the percentage of individual APs that map to one of them is 41%. In other words, 41% of individual attentional locations map to the top 10% node locations on the SES. The percentage calculations for different thresholds are listed in the last column of Table II.

TABLE II

Activation Threshold	Nodes above Thresh. (NAT)	Percentage of NAT	% Attentional pts. at NAT
27000	201	30%	77%
45000	134	20%	65.3%
100000	64	10%	41%

Another measure of the attentional importance of SES nodes is the percentage of attentional locations in the top N nodes. This is similar to the percentage comparison above, except that a fixed number of nodes is chosen. Regardless of the number of APs in a scene, only a fixed number of them can and should be attended to. For example, 19% of individual attentional locations were found to map to the top 20 node locations on the SES. In other words, the 20 most salient locations on the sphere represent 19% of all individual attentional locations. Table III shows the number of attentional locations for several values of N.

TABLE III

N	% attention points in top N node locations
20	19%
30	25.8%
50	36.2%

Note that all of the images underlying a given fovea are effectively aligned by their common attention points. Table IV shows that this tends to increase the saliency of a node. APs are first identified and posted at the appropriate nodes. The clustering procedure described on the previous page ensures that the points corresponding to a single image feature are mapped to a single node, thereby aligning the images on the foveal window of that node.

B. Whole Images versus Mosaic Image Attention

Attentional points found through the whole image approach were compared to attentional points found over the foveal mosaic image. This was done by processing the mosaic as a single image (e.g., Fig. 4) with FG to find the N nodes with highest activation. When attentional processing is performed on full-size individual images (as in section II-C), some attentional locations get mapped to nodes that do not correspond to an image piece posted in table tblSES. This occurs in images taken at nodes lying near the edges of the visual scene. These locations are not represented in the reconstructed visual scene image. Therefore, the top N locations that correspond to a node in the mosaic image were found. The attentional locations found through summation of the activation values in the complete set of images were then compared to those found in mosaic image directly (Table IV).

Objects such as the panda, Barney doll, trash can, left side shelves, and chair had features that were considered salient by FG. These objects were detected in both the summed activation image and the mosaic image. Features with definite edges and corners, such as the black frames on the front wall and the black wall-strip (marked with an asterisk in figure 41) were also detected in both images.

TABLE IV.
MATCHING ATTENTIONAL NODES BETWEEN INDIVIDUAL IMAGE ACTIVATION SUMMING AND RECONSTRUCTED IMAGE

N	Number / % matching nodes in top N locations
12	5 / 42%
20	8 / 40%
30	13 / 43%
50	21 / 42%
100	59 / 59%

The whole image method is better-suited to attention processing on the SES given the discontinuities in the mosaic image. The mosaic method makes no use of the redundancy inherent in the multiple overlapping images. Moreover, updating the saliency distribution on the SES as new images are captured is straightforward if the whole image method is implemented. Images can be processed as they are made available and the overall salient-point locations of the SES can be updated for the affected nodes. The mosaic method, even though it contains less information requires more processing since the entire mosaic must be processed to find attentional points with the new fovea in place.

Experiments were performed to test the robustness of the whole image method for attentional processing. A subset of the original visual scene was selected and image sequences of that scene under different illumination levels were generated. The number of matching nodes between sequences with differing illumination can be found in table V. The low light and low spotlight illumination levels were very different from the high and medium light levels. This accounts for the low percentage of matching nodes. However, the percentage of matching nodes between the high light and medium light levels were high, which indicates that the system will behave similarly when faced with moderately differing light levels.

TABLE V
MATCHING NODES PER ILLUMINATION LEVEL

N	High vs. Medium	High vs. Low	High vs Low Spot	Medium vs. Low	Low vs. Low Spot
12	11/92%	6/50%	3/25%	7/58%	5/42%
20	16/80%	10/50%	8/40%	11/55%	11/55%
30	25/83%	19/63%	13/43%	22/73%	17/34%
50	46/92%	39/78%	26/52%	42/84%	28/56%
100	87/87%	76/76%	58/58%	75/75%	60/60%

V. CONCLUSIONS

This paper has presented procedures to map high-resolution imagery on the SES and to perform visual attention processing on the SES. A sequence of images that covered a visual scene was mapped to the SES. A foveal window was extracted from the center of each image to create a mosaic representation of the scene on the SES.

Two approaches to visual attention on the SES were examined. Both methods used the FG model of visual attention. The first method performed attentional processing on individual full-size images and mapped each attentional location to the nearest SES node. A cumulative salience value was computed for each node by summing the salience values of the attention points at each node. This approach results in reasonably consistent and stable attentional points since they are combined from a number of different images of the same angular region of space. An attentional point that appears in several adjacent images is deemed to be more salient than one found in a single image. Larger activation values are assigned to SES nodes that contain APs visible in several images. The other approach was to process the thumbnail mosaic image with the attentional algorithm. The results of this were problematic, containing spurious APs and missing others.

To test the first method further, scene images were taken under different illumination levels. For reasonable illumination variations, the majority of attentional locations persisted. In future work, this method will be implemented on a humanoid robot and used in real time. The bulk of the processing time is due to the Frei-Chen filters used to create feature maps of the images. When the number of orientation filters is reduced from 8 to 4, the processing time is approximately 2-3 seconds per image. This time could be shortened by revising and optimizing feature detection and the particular implementation of FeatureGate. Note that the memory requirements for storing imagery on SES are not expensive: approximately 120MB for a partially populated SES (as is used on the stationary robot ISAC), and 4.5GB for a completely populated SES.

Updating the salience of the SES as new images become available can be done with relative ease by processing the new images separately and by combining their APs with those

already present. The activation at each node could be weighed by the age of each attentional point, giving more weight to newer points.

The work presented in this paper can be used to map a robot's environment in a meaningful way. It can also be used to direct the robot's attention to the potentially salient areas of that environment by combining attention information taken from sensors. The attention information is taken from images in this work, but this can be extended to other sensory modalities such as infrared and sound.

REFERENCES

- [1] Peters, R.A. II, Hambuchen, K.A., Kawamura, K., Wilkes, D.M. "The Sensory Ego-Sphere as a Short-Term memory for Humanoids." *Proceedings of the IEEE-RAS Conference on Humanoid Robots*, 2001, pp 451-60.
- [2] Ambrose, R.O. , H. Aldridge, R.S. Askew, R.R. Burrigge, W. Bluethmann, M. Diftler, C. Lovchik, D. Magruder, and F. Rehnmark, "Robonaut: Nasa's space humanoid," *IEEE Intelligent Systems*, vol. 15, pp. 57-63, July 2000.
- [3] Kawamura, K., R.A. Peters II, D.M. Wilkes, W.A. Alford, and T.E. Rogers, "Isac: foundations in human-humanoid interaction," *IEEE Intelligent-Systems*, vol. 15, no. 4, pp. 38-45, July 2000.
- [4] Edmondson, A.C. *A Fuller Explanation: the Synergetic Geometry of R. Buckminster Fuller*. Boston: Birkhauser Verlag, January 1987.
- [5] Hambuchen, K.A., *Multi-modal attention and event binding in humanoid robots using a sensory ego-sphere*, Ph.D. dissertation, Vanderbilt University, 2201 West End Ave., Nashville, TN 37235, May 2004.
- [6] Kawamura, K., A.B. Koku, D.M. Wilkes, R.A. Peters II, and A. Sekmen, "Toward egocentric navigation," *International Journal of Robotics and Automation*, vol. 17, no. 4, pp. 135-145, Oct. 2002.
- [7] Coorg S , Teller, S. "Spherical Mosaics with Quaternions and Dense Correlation," *International Journal of Computer Vision*, v.37 n.3, p.259-273, June 2000.
- [8] Debevec, P., Taylor, C., Malik, J. "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach." *Proceedings of SIGGRAPH 96*, pp 11-20, 1996.
- [9] Lu, L., Dai, X., Hager, G. "Real-Time Video Mosaicing with Adaptive Parametric Warping." Technical Report, Computational Interaction and Robotics Lab, Johns Hopkins University, 2003. <http://www.cs.jhu.edu/%7Eelelu/publication/realtimevideomosaicing-TR.pdf>
- [10] Cave, K.R. The FeatureGate model of visual selection. *Psychological Research*, 62, 182-194 (1999).
- [11] Driscoll, J.A., Peters, R.A., Cave, K.R. A Visual Attention Network for a Humanoid Robot. *Proc. 1998 IEEE/RSJ Int'l Conf. Intell. Robotic System. (IROS'98)*
- [12] Hambuchen, K.A. "Multiple alternative hypothesis testing with a visual attention network." M.S. Thesis, Vanderbilt University, 1999.
- [13] Weeks, Arthur J. Jr. *Fundamentals of Electronic Image Processing*. SPIE/IEEE Series on Image Science & Engineering, 1996.
- [14] Frei, W. and C. C. Chen, "Fast boundary detection: A generalization and a new algorithm", *IEEE Trans. Computers*, vol. 26, pp. 988-998, 1977.
- [15] Itti, L., Visual Attention, **In: The Handbook of Brain Theory and Neural Networks, 2nd Ed.**, (M. A. Arbib Ed.), pp. 1196-1201, MIT Press, Jan 2003.
- [16] Peters, R.A. II, Hambuchen, K.A., Bodenheimer, R.E. "The Sensory Ego-Sphere: A Mediating Interface Between Sensors and Cognition." Submitted to *IEEE Transactions on Systems, Man, and Cybernetics*, September, 2006, unpublished.
- [17] Shum H.Y. and Szeliski, R. "Systems and experiment paper: Construction of panoramic mosaics with global and local alignment." *IJCV*, 36(2):101-130, February 2000.