IPSJ Transactions on Computer Vision and Applications Vol.7 35–40 (Apr. 2015)

[DOI: 10.2197/ipsjtcva.7.35]

Express Paper

Spatial Visual Attention for Novelty Detection: A Space-based Saliency Model in 3D Using Spatial Memory

Nevrez Imamoglu[1,a)] Enrique Dorronzoro[2,3,b)] Masashi Sekine[1,3,c)]
Kahori Kita[3,d)] Wenwei Yu[3,e)]

Received: October 28, 2014, Accepted: January 22, 2015, Released: April 16, 2015

Abstract: Saliency maps as visual attention computational models can reveal novel regions within a scene (as in the human visual system), which can decrease the amount of data to be processed in task specific computer vision applications. Most of the saliency computation models do not take advantage of prior spatial memory by giving priority to spatial or object based features to obtain bottom-up or top-down saliency maps. In our previous experiments, we demonstrated that spatial memory regardless of object features can aid detection and tracking tasks with a mobile robot by using a 2D global environment memory of the robot and local Kinect data in 2D to compute the space-based saliency map. However, in complex scenes where 2D space-based saliency is not enough (i.e., subject lying on the bed), 3D scene analysis is necessary to extract novelty within the scene by using spatial memory. Therefore, in this work, to improve the detection of novelty in a known environment, we proposed a space-based spatial saliency with 3D local information by improving 2D space base saliency with height as prior information about the specific locations. Moreover, the algorithm can also be integrated with other bottom-up or top-down saliency computational models to improve the detection results. Experimental results demonstrate that high accuracy for novelty detection can be obtained, and computational time can be reduced for existing state of the art detection and tracking models with the proposed algorithm.

Keywords: space-based spatial saliency, saliency map with 3D data, top-down saliency, spatial memory, visual attention

1. Introduction

Visual attention (VA) mechanism of the human visual system (HVS) led our attention on the regions with higher significance or novelty [1], [2], [3], [4]. VA mechanism helps us gaze into specific regions within a scene instead of processing all of the visual information on the visual field by decreasing redundancy and encouraging our attention to shift [1], [2], [3], [4]. Inspired by the VA mechanism, many works [4], [5], [6], [7], [8] have aimed for developing computational models to obtain saliency maps (attention maps).

In our research [7], [8], we have been developing algorithms to integrate visual attention computational models to at-home biomonitoring mobile robot applications for people in need who have difficulties leaving their home. The mobile robot has to detect, track and recognize the activities of the subject [7], [8], [9], [10] during daily activities in an indoor environment. Obviously, the vision tasks of such a system are the most crucial parts that have to be robust to detect and track novel regions within a scene (e.g., subjects) to enable the system to perform activity recognition task.

Many studies on 2D and 3D data [11], [12], [13], [14], [15] have been proposed for detection and tracking applications by using saliency information to improve current conventional state-of-the-art algorithms. However, these applications do not consider the spatial memory of the observer during visual search. And, the saliency results of these models can include both novelty and feature based attention points irrelevant to the observer based on the spatial memory. As a result, they may not incorporate the known spatial association between sensory memory and the currently perceived scene on the task dependent novelty region extraction for the detection and tracking tasks. Moreover, the results of these approaches are still redundant without the use of spatial memory, which results in higher computational costs for the high level vision functions such as object detection or recognition models (i.e., Refs. [13] and [14]).

As stated by Chun and Wolfe [16], memory of prior perceptual observations affects the allocation of attention to the perceived scene [13], [17], [18]. And, saliency approaches with 3D information are generally using range sensors to take advantage of depth data [12], [15] or stereo vision from color cameras to calculate disparity maps [12]. Range sensor based approaches have a certain advantage over stereo vision for low light or dark environmental conditions. Therefore, in our previous study [8], we demonstrated that the spatial memory of the observer can be used to detect and track novel regions in the scene regardless of the

1 Graduate School of Engineering, Chiba University, Chiba 263–8522, Japan
2 Department of Electronic Technology, University of Seville, Spain
3 Center for Frontier Medical Engineering, Chiba University, Chiba 263–8522, Japan
a) nevrez.imamoglu@chiba-u.jp/nevrez@gmail.com
b) enriquedz@dte.us.es
c) sekine@office.chiba-u.jp
d) kkita@chiba-u.jp
e) yuwill@faculty.chiba-u.jp

low-level or high-level features of the objects by using Kinect and Lidar (Light Detection and Ranging) range sensors [8]. The algorithm in Ref. [8] compares local observation from the Kinect sensor with the environment memory of the robot as the global 2D map of the place, which is obtained by a Lidar sensor [8]. However, in the spatial attention model of Ref. [8], the results were 2D so it can not handle the cases with complex environments such as a person lying down on a bed. To handle these problems with the previous method [8], we proposed a 3D approach by integrating height heuristics for specific regions on the global map to enable 3D saliency computation based on the spatial memory of the environment with a mobile robot. Because, if 2D information is used, only the subject moving or standing on the floor could be detected as novelty, but the subject on or beside furniture objects (such as a bed, treadmill, table, etc.) could not be detected. Hence, height information of the objects of interest shall be integrated to achieve a 3D spatial memory.

Moreover, we integrated the proposed approach so as to assist one of the existing top-down saliency models [14] for person detection to demonstrate its advantage to existing state-of-the-art algorithms. In addition, the approach is proposed for known environments. But, in this paper, we focus on at-home monitoring for the basis of our research work in order to present one case to develop and test the model. Experimental results demonstrated that the proposed 3D space-based saliency can result in high detection accuracy of the novel regions in the scene and can decrease computation time for integrated approaches.

The paper is organized as follows; the proposed 3D spaced-based saliency model is expressed in Section 2, experimental results are given in Section 3, and finally, discussion and conclusion remarks are stated in Section 4 and Section 5 respectively.

2. Proposed Novelty Detection with Space-Based Spatial Saliency in 3D

Space-based attention not only consists of spatial features such as contrast or orientation, but also it is related to the position of object relative to a spatial reference in space [19]. Object based attention investigates the structure of the object to aid the recognition of whole object [19]. Moreover, memory, including spatial relations, is also an important factor during visual search [16], [20] since the memory of prior observations affects the designation of the attention on the perceived scene [16], [17], [18]. In summary, spatial memory is one of the keys to efficient visual search for detecting novelty, which can assist with reducing redundancy to increase performance of monitoring tasks. Therefore, by using the mobile robot platform, we developed an algorithm to obtain space-based spatial saliency maps using 3D data with height and size heuristics to find novelties such as subjects in our mobile robot based monitoring application. The flowchart of the proposed algorithm can be seen in **Fig. 1**.

As demonstrated in Fig. 1, robot pose and global map are the two necessary pieces of information crucial to the application, where local Kinect observation can be projected on to a global 2D map [8]. For this purpose, the Robot Operating System (ROS) [21] is used to handle robot navigation [22], [23], environment mapping [24], [25] as the spatial memory of the robot, and localization [26] of the robot by utilizing the Lidar range sensor [21], [22], [23], [24], [25], [26]. Then, with the global 2D map as the environment memory and the robot's pose obtained from the mobile robot platform and ROS, the space-based saliency can be generated for each point on the Kinect sensor data.

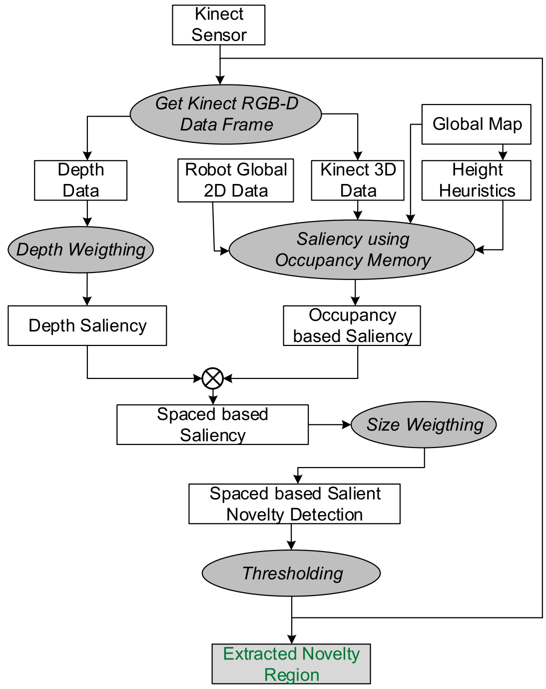

Fig. 1 Flowchart of the proposed space-based salient novelty detection algorithm.

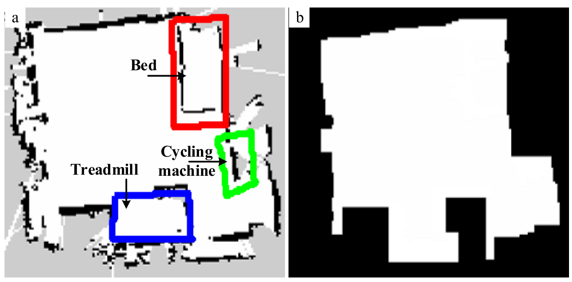

Fig. 2 (a) Location boundaries that are updated as candidate regions for novelties, (b) updated occupied memory map for the regions where novelties can not exist.

First, since Kinect data is 3D and the global map is 2D, some occupied regions of the global map have the possibility to include subjects such as on the bed, on the sofa, etc.. So the global map should support some height information to handle these possible false negative detection cases. Then, spatial saliency on these complex global positions (i.e., bed) can be obtained with the integration of 3D information by removing irrelevant points, which is not included in our previous application [8]. Also, the map is updated for irrelevant regions such that the possibility of seeing a subject does not exist similar to the table plane or unused regions of the environment [8]. As it can be seen in **Fig. 2**, the bed, cycling machine, and treadmill region is included as candidate regions for the search of novel regions (white areas of Fig. 2 (a)) with their maximum possible height without any subjects; whereas, tables or irrelevant regions are defined as occupied regions (black areas of Fig. 2 (b)) in the environment memory as

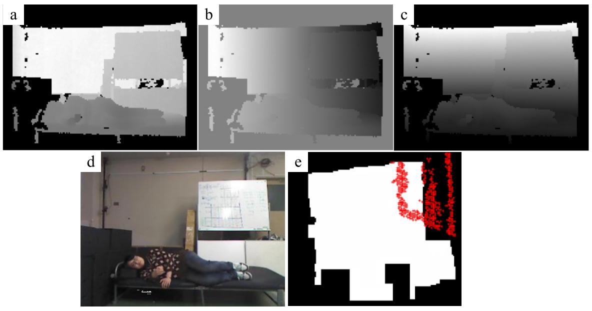

Fig. 3 (top-row) Kinect 3D local data; (a) depth, (b) horizontal, (c) vertical images respectively (bottom-row), (d) Color image of Kinect 3D observation, and (e) $K_G(x, y)$ transformed points are shown on the 2D map with red marks.

the occupancy map.

After prior adjustments, the system can start working by processing each Kinect 3D data. Initial steps of the proposed algorithm are similar to our earlier work [8]. To reduce the number of points to be processed, first, the data is filtered by removing the ceiling and floor from the 3D observation (**Fig. 3** (top-row)) and the 3D points are arranged into 5 cm grids. Then, all the local Kinect points on the 3D space are transformed to a global 2D map as described in Eq. (1) [8].

$$\begin{bmatrix} K_{Gx} \\ K_{Gy} \end{bmatrix} = \mathbf{A} \times \begin{bmatrix} K_{Lx} \\ K_{Ly} \end{bmatrix} + \begin{bmatrix} T_{Gx} \\ T_{Gy} \end{bmatrix} \tag{1}$$

$$\mathbf{A} = \begin{bmatrix} s_x \cos(\alpha) & -s_y \sin(\alpha) \\ s_x \sin(\alpha) & s_y \cos(\alpha) \end{bmatrix} \tag{2}$$

$$\begin{aligned} T_{Gx} &= R_{Gx} - d\cos(\theta) \\ T_{Gy} &= R_{Gy} - d\cos(\theta) \end{aligned} \tag{3}$$

where K_{Gx} and K_{Gy} are the transformed Kinect global x and y positions on the global 2D map, K_{Lx} (depth data in Fig. 3 (top-row (a))) and K_{Lx} (horizontal data in Fig. 3 (top-row (b))) are the local Kinect data on the relevant x and y axis. α is the Kinect pose on global map calculated by the robot pose and rotating table angle. $\mathbf{A}$ is the transformation matrix given in Eq. (2), in which s_x and s_y are the scaling coefficients on x and y axes. T_{Gx} and T_{Gy} are the translation values (Eq. (3)) due to the difference between the robot's center and the Kinect position on the robot. And, in Eq. (3), R_{Gx} and R_{Gy} are the robot's global position, d is the distance of the Kinect position from the robot's center, and θ is the robot's pose on the global map.

As demonstrated in Fig. 3, the next step is to create occupancy based saliency map by taking advantage of height heuristics for the defined regions, the global Kinect 3D points and occupied regions on the global map. So, the comparison of observations to the spatial memory will provide the saliency of the occupied points as described in **Table 1**, in which $S_o(K_{Gx}, K_{Gy})$ is the occupancy based saliency of the given $K_G(x, y, h)$ point with height h based on the environment knowledge (global 2D map) $\mathbf{O}^g_m$, and $MAXH_B$, $MAXH_T$ and $MAXH_C$ are the maximum height of the bed, treadmill and cycling machine for the given sample map, respectively. Data in Fig. 2 is used to remove them and extract the main novelty from the Kinect data. And, as in Ref. [8], $l_2 - norm$ of the vectors, expressed as each Kinect point to each global map points, were used to compare the perceived scene to

Table 1 Pseudo-code of occupancy based saliency for each Kinect global data for the given spatial memory (global map).

foreach $K_G(x, y, h)$
  if isOnBed $(K_G(x, y))$ && $K_G(h) \leqslant MAXH_B$
    $\mathbf{S}_o(K_{Gx}, K_{Gy}, h) = 0$
  elseif isOnTreadmill $(K_G(x, y))$ && $K_G(h) \leqslant MAXH_T$
    $\mathbf{S}_o(K_{Gx}, K_{Gy}, h) = 0$

  elseif isOnCycling $(K_G(x, y))$ && $K_G(h) \leqslant MAXH_C$
    $\mathbf{S}_o(K_{Gx}, K_{Gy}, h) = 0$
  else
    $\mathbf{S}_o(K_{Gx}, K_{Gy}, h) = \mathtt{min}(\, \mathtt{norm}_{l_2}(\, \mathbf{O}^g_m,\ K_G(x, y)\,)\,)$

Table 2 Pseudo-code of size weighting.

1) Create binary image **BW** from $\mathbf{S}_s$ as:
$$BW(K_{Gx}, K_{Gy}, h) = \begin{cases} 1 & if\ S_s(K_{Gx}, K_{Gy}, h) \\ 0 & otherwise \end{cases}$$

2) for each extracted region BW_i
  - calculate area as the number of pixels for $i^t h$ region R_i
  - if area is smaller than a defined threshold remove the region

3) Assign each area values to corresponding pixels of the region:
$A_i(K_{Gx}, K_{Gy}, h) = R_i$

4) 4) Then size weighting map S_A is normalized values as
$S_A(K_{Gx}, K_{Gy}, h) = \mathrm{N}(A(K_{Gx}, K_{Gy}, h)) + 1$

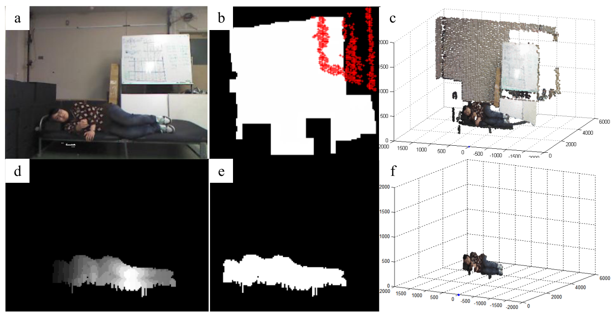

Fig. 4 (top-row) (a) Color image of Kinect 3D observation, (b) $K_G(x, y)$ transformed points are shown on the 2D map with red marks, (c) representation of the Kinect data in 3D (bottom-row), (d) space-based saliency maps with height information of the bed, (e) extracted novelty region from space-based saliency map, (f) space-based saliency region in 3D.

the spatial memory for detecting saliency values. As also stated in Ref. [8], to make the closer regions with higher saliency, depth weights (Eq. (4)) are used on the occupancy based saliency to obtain space-based saliency as in Eq. (5) since the distance of the objects to the observer may also affect attention during visual search.

$$\mathbf{S}_D = \frac{1}{\mathrm{N}(\mathbf{X}^{depth}_k) + 1} \tag{4}$$

$$\mathbf{S}_S(K_{Gx}, K_{Gy}, h) = \mathbf{S}_D(K_{Gx}, K_{Gy}, h) \times \mathbf{S}_o(K_{Gx}, K_{Gy}, h) \tag{5}$$

where S_D is the depth saliency map, which is used for weighting the occupancy based saliency map S_o, and N(.) is the normalization function to the {0-1} scale for local depth data $\mathbf{X}^{depth}_k$ of the Kinect sensor.

Then, size weighting (Eq. (6)) is applied to each salient region on the space-based saliency map S_S as in **Table 2** to obtain the final saliency map for novelty detection as demonstrated in **Fig. 4**.

In Table 2, the normalization function N(.) is the same as Eq. (4). Finally, saliency map obtained from top-down spatial relations can be updated with the size weighting values of SA as below:

$$\mathbf{S}_S^{'}(K_{Gx}, K_{Gy}, h) = \mathbf{S}_S(K_{Gx}, K_{Gy}, h) \times \mathbf{S}_A(K_{Gx}, K_{Gy}, h) \qquad (6)$$

In Fig. 4, the novelty region extracted by the proposed space-based saliency map is given both in 2D and 3D. As it can be seen, using the height prior information the bed is removed and only the subject is the novelty of the scene.

3. Experimental Results

An experiment in a room environment with continuous data was done, where the subject moves in the room randomly to test the detection performance of the algorithm. For this part, we used the layout in Fig. 2 (a) that includes a furniture bed, table, treadmill, and cycling machine. The aim of the test with this layout is to show that detection of the subject as novelty can be done efficiently to provide of at-home monitoring support while the subject is doing various daily activities such as standing, walking, sitting, lying on the bed, using a treadmill, and using a cycling machine. For this purpose, 6,065 frames were used as a test benchmark.

In the detection and tracking experiment with the given dataset, tracking is simply done by using the centroid of the extracted region and marked on the novelty image as a tracking point as in Ref. [8]. Then, if the tracking point is on the subject, we will count the result as a successful detection and tracking case. However, detection failure will occur if the subject body is not detected as a novel region. Tracking failure condition happens when the tracking point is not on the subject body. With these conditions, performance of the algorithm is tested on Dataset-2, and the results are given in **Table 3**.

In **Fig. 5**, various detection and tracking cases are given with the experimental layout (Fig. 2) and given dataset, in which the subject randomly moves inside the room to simulate daily life activity situations. The results show that detection and tracking can be very useful by pointing out the novelty in the scene with the spatial working memory regardless of space or object base features. Moreover, subjects as a novel part of the scene can be detected and tracked accurately in order to support at-home monitoring surveillance mobile robot systems to handle or improve the performance of the existing state of the art models with appropriate integration.

As it can be seen from Table 3, detection and tracking were done in 6,048 frames among the 6,065 frame of dataset, and only 17 cases failed at tracking. The subject body is mistaken for obstacles in spatial memory when the subject is too close to the real obstacle such as the wall, tables, etc. due to low map resolution and the obstacle map in Fig. 2 (b). This can be improved with

Table 3 Novelty detection results for subject monitoring.

	Detection and Tracking Results		
	Success	Detect fail	Track Fail
# frames	6,048	0	17
Accuracy	99.72%		

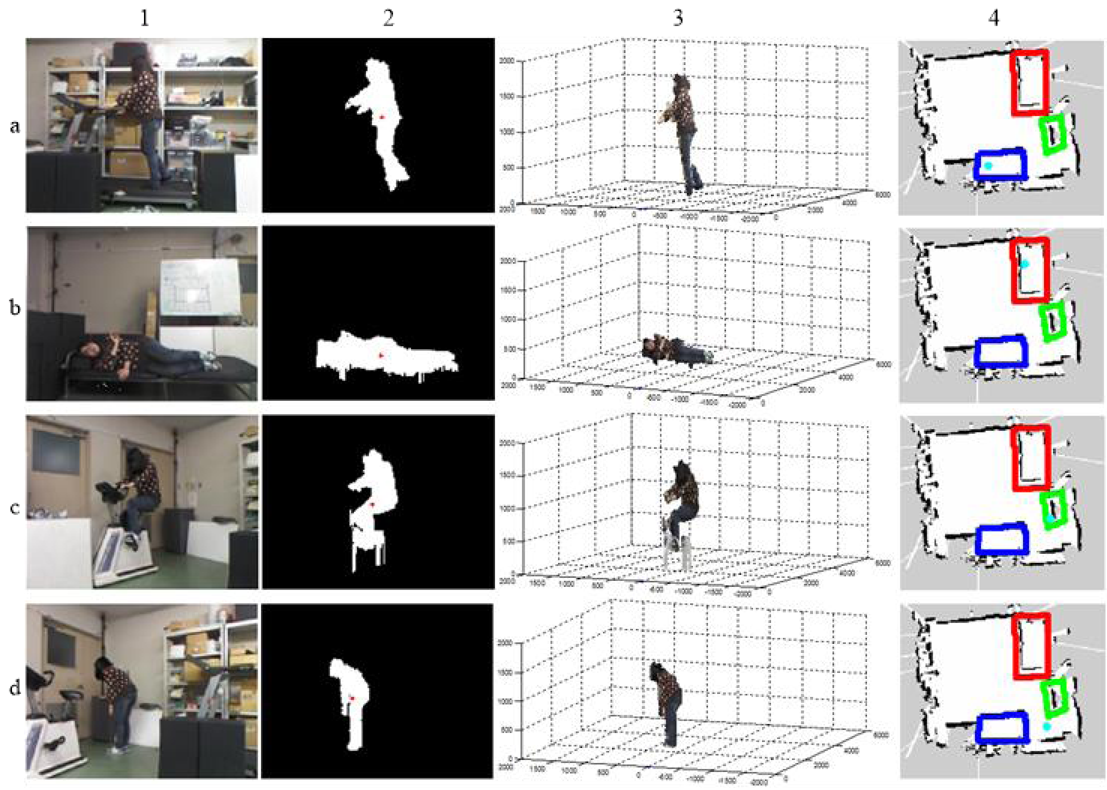

Fig. 5 1. a–d) sample color images, 2. a–d) extracted novelty region with tracked point is marked as the centroid of the region, 3. a–d) 3D representation of the novelty region extracted by the proposed space-based saliency using spatial memory, 4. a–d) global map with defined regions as in Fig. 2 and subject location on the map (marked cyan point).

higher resolution global map and more accurate obstacle map rather than a rough data. However, overall detection and tracking performance is quite reliable at around 99.72% accuracy.

3.1 Integration of Proposed Space-Based Saliency to an Existing Saliency Algorithm for Human Detection

In this part of the experimental analysis, we tried to integrate the proposed space-based novelty extraction algorithm to an existing top-down saliency based human detection model [14] to observe time and detection efficiency. The study in Ref. [14] uses color images to extract SIFT features and calculates the likelihood of the features based on the training data obtained by the object model, where the person is selected as the object of interest in our experiments to examine the top-down saliency information [14]. Integration is done in a way that the novelty region is extracted from a rectangular template to be used as an input image on the top-down saliency model of Ref. [14] rather than processing the whole image with high-level feature analysis. First, we selected a sequential set of data (100 frames) randomly from the whole dataset to compute the average time and to check the efficiency of the integration of the proposed novelty detection algorithm to Ref. [14]. The implementations of the proposed model and the study in Ref. [14] are built in Matlab®. So, the tests for the processing time of the algorithms are done using Matlab®. And, the desktop PC is used in which the CPU is Intel® Core(TM) i7-3930K 3.20 GHz and RAM is 16 GB DDR3 type.

In **Table 4**, test results for the processing time of Ref. [14] and the proposed integration approach are given. We used color images with 480x640 image resolution for testing the algorithm [14], and the proposed integration algorithm utilizes 120x160x3 3D data for generating space-based saliency map to combine with 480x640 color images to be processed by Ref. [14]. The saliency model in Ref. [14] takes 3.3179 seconds for the selected set of data; however, this processing time becomes 0.7223 seconds with the spatial novelty information by removing irrelevant regions for saliency calculation with high-level features such as SIFT. It can be seen that average computation time decreased by 78.23% after the integration of spatial saliency based novelty detection algorithm to Ref. [14] for top-down saliency on human category.

Table 4 Processing time of the dataset.

Average computation time (sec.) of saliency models			
# frames	TBS-H [14]	Improved TBS-H [14] by proposed spatial saliency	Efficiency
100	3.3179	0.7223	78.23%

A sample of the color image from the dataset, proposed saliency map result, extracted novelty region using the proposed model, and selected cropped color image region to be processed by the algorithm [14] are given in **Fig. 6** (a)–(d). The idea is to process only the significant part of the image with novelty (Fig. 6 (d)) for the top-down saliency model [14] instead of using the whole image. By this way, the salient points of the whole scene with no novelty can be removed. In Fig. 6 (e) and Fig. 6 (f), the marked color regions are the detected person regions by the saliency approaches for algorithm [14] only and an improved version with the help of proposed spatial saliency of this paper, respectively.

4. Discussion

It is obvious that spatial attention can greatly improve the visual search task to extract novelties from a scene. By this way, novelties can be detected and tracked easily within the scene for monitoring tasks such as human tracking and activity recognition. Moreover, this algorithm can reduce the overall processing time by removing irrelevant information. Therefore, as in the human visual system, pre-processing of the spatial environment with the aid of spatial memory regardless of features can improve the attention or visual search processes for task dependent applications such as at-home bio-monitoring, human detection, object detection, novelty tracking, and etc.

From the application's perspective, the proposed spatial

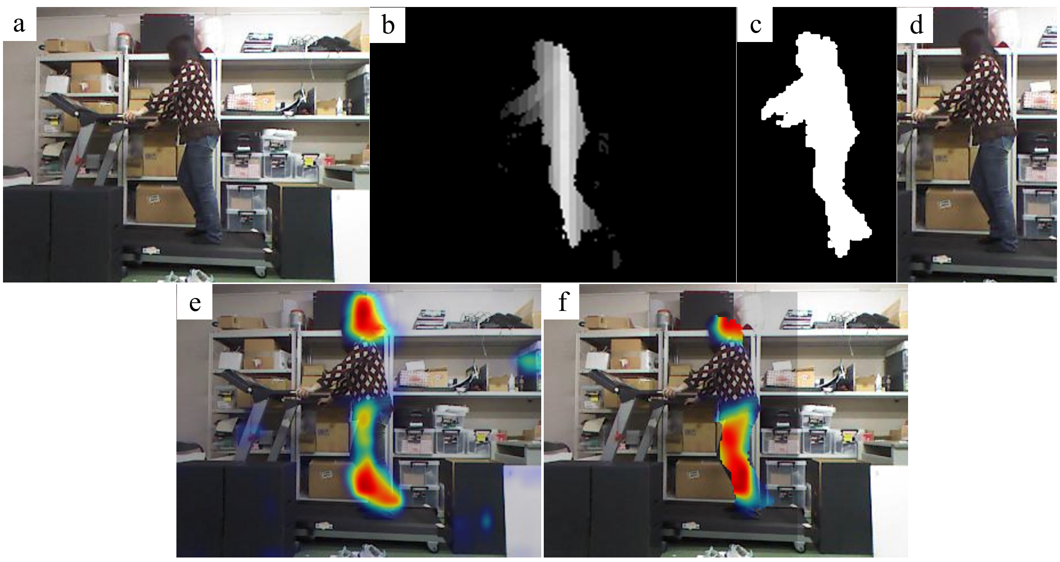

Fig. 6 (a) sample color image, (b) proposed saliency map demonstrated in 2D image template, (c) extracted novelty region, (d) selected cropped color image region based on the novelty region, (e) top-down saliency result for person features [14], (f) proposed saliency integration result.

saliency model is very effective at giving significant information from the scene by removing irrelevant data to reduce redundancy. Also, it provides a simple and fast model to extract novelty regions, which is crucial for real time applications. For example, in Table 4, the average processing time of the integrated system (proposed saliency and the study [14]) is given 0.7223 seconds, where most of this process' time (0.5875 seconds) was consumed by the top-down saliency map algorithm of Ref. [14]. On the other hand, the proposed spatial saliency in 3D takes only 0.1348 seconds out of the whole process. In addition, we improved our previous spatial attention algorithm [8], which can not separate a person lying on the bed or sofa from the furniture. However, current 3D approaches with prior height information of specific regions on the global map can remove irrelevant parts and identify novel subjects as demonstrated in Fig. 4 and Fig. 5 examples.

5. Conclusion

In this paper, a new approach for space-based spatial saliency for 3D room layouts is proposed by integrating height heuristics for specific regions on the global map to enable 3D saliency computation based on the spatial memory of the environment with a mobile robot. So, with new heuristics, the system can work in more complex environments such as a room with furniture (sofa, bed, treadmill, cycling machine, etc). By this way, at-home monitoring of persons living alone can be improved with a mobile robot using spatial attention for visual search. As future work, the processing speed can be increased more, and the use of a 3D global map can be tried for 3D spatial attention with higher flexibility. Also, the 5 cm grid resolution can be decreased to get smoother results for extracting the novelty region boundaries. Finally, the semantic mapping approach can be used together for changing the environment by updating the global map since the current solution assumes that the global map is static and does not change.

Acknowledgments The authors would like to thank Zhixuan Wei and Prof. Weidong Chen [7] from the Institute of Robotics and Intelligent Information Processing, Department of Automation, Shanghai Jiao Tong University, Shanghai, People's Republic of China for helpful discussion on algorithm and experimental analysis.

References

[1] Treisman, A. and Gelade, G.: A feature-integration theory of attention, *Cognit. Psychol.*, Vol.12, No.1, pp.97–136 (1980).

[2] Wolfe, J.: Guided search 2.0: A revised model of guided search, *Psychonomic Bull. Rev.*, Vol.1, No.2, pp.202–238 (1994).

[3] Logan, G.D.: The CODE theory of Visual Attention: An integration of space-based and object-based attention, *Psychological Review*, Vol.103, No.4, pp.603–649 (1996).

[4] Zhang, L. and Lin, W.: *Selective Visual Attention: Computational Models and Applications*, Wiley-IEEE Press (2013).

[5] Itti, L.: Models of bottom-up and top-down visual attention, Ph.D. Dissertation, Dept. Computat. Neur. Syst., California Inst. Technol, Pasadena (2000).

[6] Fang, Y., Chen, Z., Lin, W. and Lin, C.-W.: Saliency detection in the compressed domain for adaptive image retargeting, *IEEE Trans. Image Processing* (*IEEE TIP*), Vol.21, No.9, pp.3888–3901 (2012).

[7] Imamoglu, N., Wei, Z., Shi, H., Yoshida, Y., Nergui, M., Gonzalez, J., Gu, D., Chen, W., Nonami, K. and Yu, W.: An Improved Saliency for RGB-D Visual Tracking and Control Strategies for a Bio-monitoring Mobile Robot, *Communications in Computer and Information Science*, Vol.386, pp.1–12 (2013).

[8] Imamoglu, N., Dorronzoro, E., Sekine, M., Kita, K. and Yu, W.: Top-down spatial attention for visual search: Novelty detection-tracking using spatial memory with a mobile robot, *Advances in Image and Video Processing*, Vol.2, No.5, pp.40–58 (2014).

[9] Nergui, M., Imamoglu, N., Yoshida, Y. and Yu, W.: Human gait behaviour classification on lower body triangular joint features, *IASTED International conference on Signal and Image Processing* (*SIP*), pp.212–219 (2012).

[10] Nergui, M., Yoshida, Y., Imamoglu, N., Gonzalez, J., Otake, M. and Yu, W.: Human Activity Recognition Using Body Contour Parameters Extracted from Depth Images, *Journal of Medical Imaging and Health Informatics*, Vol.3, No.3, pp.455–461 (2013).

[11] Frintrop, S., Konigs, A., Hoeller, F. and Schulz, D.: A component based approach to visual person tracking from a mobile platform, *International Journal of Social Robotics*, Vol.2, No.1, pp.53–62 (2010).

[12] Wang, J., Perreira Da Silva, M., Le Callet, P. and Ricordel, V.: A computational model of stereoscopic 3D visual saliency, *IEEE Trans. Image Processing* (*IEEE TIP*), Vol.22, No.6, pp.2151–2165 (2013).

[13] Zhang, G., Yuan, Z. and Zheng, N.: Key object discovery and tracking based on context aware saliency, *International Journal of Advanced Robotic Systems*, Vol.10, No.15, pp.1–12 (2013).

[14] Yang, J. and Yang, M.-H.: Top-down visual saliency via joint CRF and dictionary learning, *IEEE International Conference on Computer Vision and Pattern recognition* (*CVPR*), pp.2296–2303 (2012).

[15] Garcia, G.M. and Frintrop, S.: A computational framework for attentional 3D object detection, *Proc. Annual Meeting of the Cognitive Science Society* (2013).

[16] Chun, M.M. and Wolfe, J.M.: *Visual Attention, Handbook of Sensation and Perception* (*Chapter 9*), Goldstein, E.B. (Ed.), pp.273–310, Blackwell Publishing (2005).

[17] Chun, M.M. and Jiang, Y.: Contextual cueing: Implicit learning and memory of visual context guides spatial attention, *Cognitive Psychology*, Vol.36, pp.28–71 (1998).

[18] Desimone, R. and Duncan, J.: Neural Mechanisms of selective visual attention *Annual Review of Neuroscience*, Vol.18, pp.193–222 (1995).

[19] Fink, G.R., Dolan, R.J., Halligan, P.W., Marshall, J.C. and Frith, C.D.: Space-based and object-based visual attention: shared and specific neural domains, *Brain*, Vol.120, pp.2013–2028 (1998).

[20] Oh, S.H. and Kim, M.S.: The role of spatial working memory in visual search efficiency, *Psychonomic Bulletin and Review*, Vol.11, No.2, pp.275–281 (2004).

[21] Robot Operating System (ROS): (online), available from ⟨http://wiki.ros.org/⟩

[22] ROSARIA: (online), available from ⟨http://wiki.ros.org/ROSARIA⟩

[23] ROS Costmap: (online), available from ⟨http://wiki.ros.org/costmap_2d⟩

[24] Grisetti, G., Stachniss, C. and Burgard, W.: Improved techniques for grid mapping with rao-blackwellized particle filters, *IEEE Trans. Robotics*, Vol.23, pp.34–46 (2007).

[25] ROS Gmapping: (online), available from ⟨http://wiki.ros.org/gmapping/⟩

[26] Thrun, S. and Burgard, W.: *Probabilistic robotics*, The MIT Press Cambridge (2005).

(Communicated by *Jong-Il Park*)