# Targets Detection Using Multiple Foveas

Petrucio R. T. Medeiros[*§], Rafael B. Gomes[†] and Luiz M. G. Gonçalves[*]
[*] Department of Computer Engineering
[†] Department of Computer Science and Applied Mathematics
Federal University of Rio Grande do Norte
Emails: petrucio.medeiros.082@ufrn.edu.br, rafaelbg@dimap.ufrn.br and lmarcos@dca.ufrn.br

***Abstract*—Target detection enables running a robotic task. However, their limited resources make large amount of data processing harder. Image foveation is an approach that can reduce processing demand by reducing the amount of data to be processed. However, as an important visual stimulli can be attenuated by this reduction, some strategy should be applied in order to keep/recover awareness of it. This work compares gradient descent (potential field), maximum likelihood, multilateration, trilateration, and barycentric coordinates to solve this problem in a multiple mobile foveas context. Our results demonstrate that the proposed methodology detects the target converging with an average euclidian distance of 51 pixels from the target's center position.**

## I. Introduction

Robots perform rescue of disaster victims [1], allow complex surgical procedures with greater precision, flexibility, and control than conventional techniques [2], and help the hospital disinfection with ultraviolet rays to eliminate viruses and bacteria [3], among other applications. Regardless of the task, robots capture external information through sensors and use their actuators to interact with the environment.

These systems have limited resources and operate in dynamic environments, that is, the robot must be able to find some target given in the task as well as react to some adverse event during its operation. Nonetheless, this requires runtime processing and this is a challenging task due to the volume of data captured. The operating environment of a robot contains a vast volume of visual information. If not all information can be processed simultaneously, then the robot needs to select which part of the information is essential to perform the task.

Visual attention is able to select relevant information [4] to conduct the robotic task (in top-down attention) and highlight perceptible stimuli that can make the robot inoperative (bottom-up attention). In this work, we implement the visual attention process using the foveation technique, which is characterized by keeping some small region of the image at high resolution, and reducing the resolution according to the approximation to the image periphery [5].

The foveation technique uses the movement of the fovea to select the relevant information in the image and this movement takes time. There are contexts where time is crucial, for example, during a robotic search for survivors in the chaotic scenario of a natural disaster. In this context, the robot needs to keep visual attention on the rescue task (locate victims) while dodging the wreckage of the disaster that could eventually render it inoperative [6]. This problem can be overcome using the multifoveation technique, which uses multiple mobile foveas in the image, allowing visual attention to be maintained at different points in the scene at the same time.

There are several strategies to positioning the foveas in the image [7]–[14]. These strategies produce good results when submitted to the appropriate contexts, but a few works use the fovea distribution for target detection. Considering the fovea distribution in the example case mentioned above, the robot establishes foveas in the task of detecting survivors and these foveas help detect any adverse event that puts the robot at risk making both tasks possible simultaneously.

In this work, we propose a new approach to target detection using multiple moving foveas, which means that the foveas may be positioned in different places of the image from one frame to the next one. Our basic idea is to have estimates for these foveas of being the targets in evidence at a given instant, and from these dynamic and fixed foveas to estimate the final targets' positions in the image. Due to the difficulty in the mathematical modeling and implementation, solving this problem remains a challenge. Here we use mathematical strategies adapted to the context of computer vision, which consider the distribution of the foveas to estimate the localization of the target in the image. The mathematical strategies adopted here are the gradient descent (potential field), maximum likelihood, multilateration, trilateration, and barycentric coordinates.

In the remaining of this article, Section II introduces the techniques for reduction of data based on multiresolution. Related works using these concepts are described in detail on Section III. Section IV has explanation about the proposed target detection algorithm using multifoveation, which is the core and main contributions of this work. Section V verifies the usefulness of our methods including experimental analysis. Finally, we discuss the main contributions and trace directions on Section VI.

## II. Basic concepts

In the context of the evolutionary paradigm, our vision system has gone through several transformations to reach a more complex state. Through these transformations, our vision established a non-uniform distribution of cones and rods on retina. Cones, which allow a more accurated vision, are more concentrated on fovea, a small region of the retina. Based on this concept, Leonard Uhr and Charles Vossler [15] propose

[§]This work is a result of the Ph.D. thesis defended in September 2020.

the first method of multi-resolution vision based on the fovea. Multiresolution refers to a set of algorithms that decompose a signal into different resolution levels [16]. Based on this theory, several methods were proposed, such as: Gaussian pyramid, Laplacian pyramid, log-polar, and use of wavelets, among others.

In general, these methods transform the image from the cartesian domain into another domain, where the image will be processed and analyzed according to the chosen method. Each method contains intrinsic properties that can facilitate signal manipulation depending on the situation and need. Among these methods, the Multiresolution Multi-Feature (MRMF), proposed by Gonçalves et al. [17], selects parts of the image and resizes them equally to smaller sizes (see Figure 1). This method stands out for reducing the processing time in extracting information but maintains the static structure, such as the fixed centered fovea positioning. Feature extraction in MRMF is performed in the multiresolution domain. Each MRMF level is stored in memory and each sub-image has its features extracted. Note that each multiresolution level represents a portion of the image.

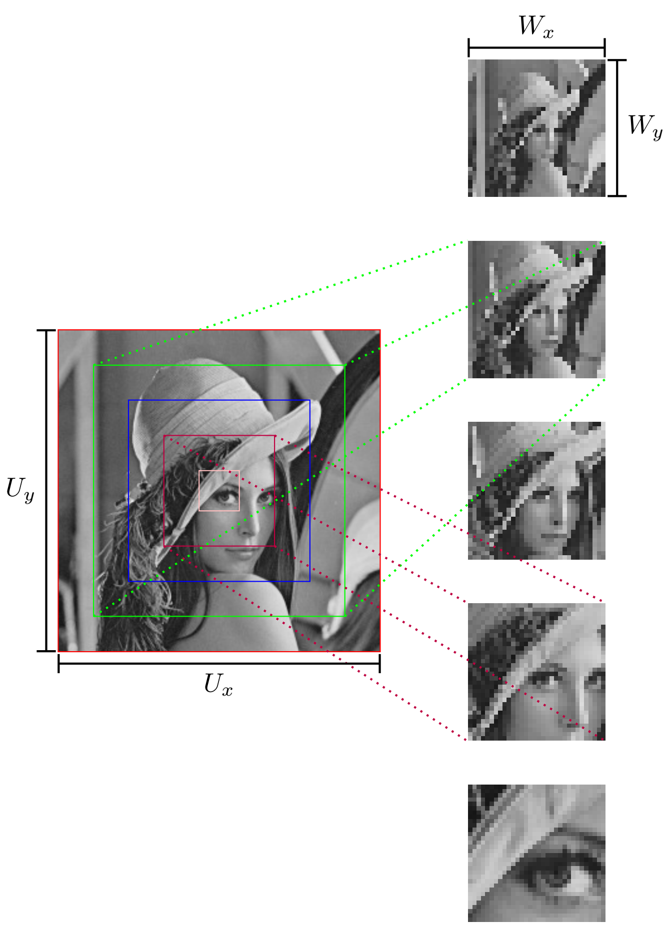



Fig. 1. Construction of levels with the MRMF.

Gomes et al. [13], [18]–[20] formalized the mathematics of the MRMF method and contributed by moving the structure along the image, known as Multiresolution with Moving Fovea (MMF). The objective of this approach is to discard features that are not essential for the task and allow execution in real time. In this multiresolution approach, the image is mapped to a set of $k$ levels, with constant size $W$, with indexes from $0$ to $m$, where $m$ is the level of the fovea, as seen in Figure 1. Let $I$ be an image of size $U = (U_x, U_y)$, and for each level $k$, it is delimited a portion of size $S = (S_x, S_y)$ of $I$, which will be mapped to the multi-resolution domain. It is defined that $S_0 = U$ and $S_m = W$, whereas the intermediate levels are obtained by interpolation, according to Equation 1.

$$S_k = \frac{mU + Wk - kU}{m} \tag{1}$$

Gomes et al. [13] defined that the motion of the fovea can be controlled by using a fovea vector $F$ inside the image domain (see Figure 2). Therefore, it is defined that the vector $F$, originating from the center of the image $I$, is between $(W - U)/2$ and $(U - W)/2$. Consequently, $F = (0, 0)$ when the fovea is positioned in the center of the image $I$. Equation 2 indicates the starting position of each region, in the space domain, which must be transformed.

$$\delta_k = \frac{k(U - W + 2F)}{2m} \tag{2}$$

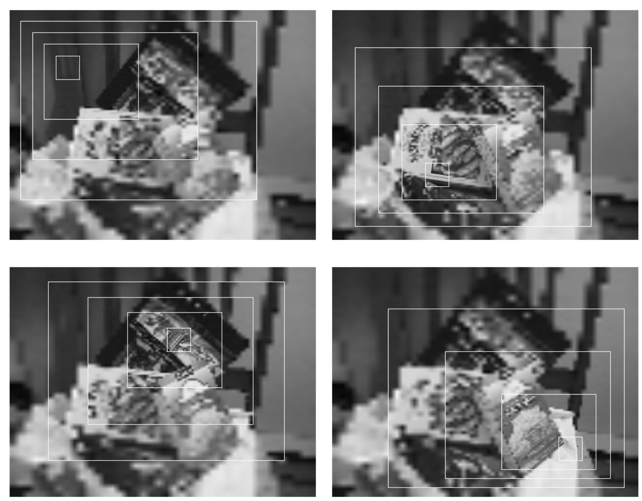


Fig. 2. Foveated image in different positions using the MMF method.

Medeiros et al. [21], [22] proposed the multifoveation by replication of the MMF structure, but that results in high computational costs since the structures are applied independently, leading to redundant processing of the image's portions. They solved this problem by removing the redundant processing between foveas (see Figure 3). A similar approach has also being applied to 3D point clouds for recognition purposes [23].

## III. Related Works

While analyzing the literature, we noticed that multi-foveation is not limited to computer vision. Dario et al. [10] proposed the construction of a tactile system with pyroelectric and piezoelectric sensors using multifoveation to reduce

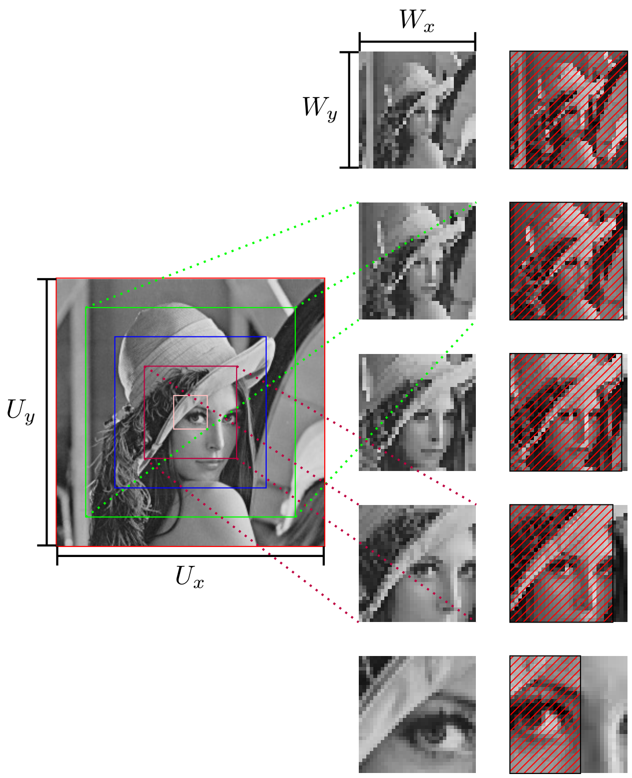



Fig. 3. Removing redundancies using MMMF with the second fovea shifted 20 pixels to the right. The first column with dimension $(W_x, W_y)$ refers to the first fovea and the second column refers to the second fovea added with the redundancy removed using MMMF method. The hatched region corresponds to redundant information between the two foveas.

resources. However, they designed a static tool, limiting the potential of the technique.

Multifoveation seeks to provide greater visual attention in a complex and dynamic context. Based on that, Lim et al. [24] developed a target tracking using multiple foveas with the log-polar method. The authors divided the image into regions, and features of the central parts determined which target will be tracked. Ude et al. [25] built a visual attention system for humanoid robots, which uses multiple foveas to generate a salience map and thus drive the saccadic movements of the robot's visual perception.

Tracking with movement of the robot's physical structure is limited due to the time it takes to move the physical apparatus from one target to another. Given this situation, Camacho et al. [26], [27] implemented an algorithm in Field Programmable Gate Array (FPGA) based on a gaussian pyramid that supports multiple foveas. This algorithm performs subdivision on each side of the fovea to build the levels of the structure. They used the last levels of the fovea to detect target movement.

Security, videoconferencing, and traffic supervision applications require quality visual information. In this context, Rodríguez et al. [28] proposed to compress static information and send resolutions at different baud rates using multifoveation. Basu et al. [7]–[9] researched lossy compression techniques with multiple foveas to solve problems related to video conferencing context. Wei & Li [29] also used multifoveation to control camera position and retrieve camera movement and zoom. Sankaran et al. [30] used frequency filtering to delimit visible information and target detecting in videos. Pioppo et al. [12] noted that the previous method discarded high-resolution information and proposed reusing it to produce better quality compressed videos.

Returning to the robotics context, Nicholas et Al. [31]–[37] studied a psychophysical model of visual perception and proposed the Information-gathering Partially Observable Markov Decision Process (I-POMDP) based on a salience map. The I-POMDP uses sequential searches to find the target and accumulates the acquired information. The salience map is widely used in the literature. Itti et al. [5], [11] constructed a salience map by combining features extracted from multiple Gaussian pyramids. Cavallaro et al. [38] proposed to create an efficient method to calculate the sensitivity functions, correspondence map and cutoff frequencies for a given overall sensitivity value using salience map and multifoveation. Salience maps are combined to identify possible target positions in the study by Xu et al. [39]. They used multifoveation and matching between SIFT descriptors, extracted from the highest resolution level, to indicate the position of the target.

There are few references in the literature that use the position of multiple foveas in the detection of visual stimuli. Soos et al. [40]–[43] used neighboring foveas to create an algorithm for detecting moving land objects from images captured by unmanned aerial vehicles. They implemented a field of local potential with multiple foveas. However, they did not consider the weighting of multi-resolution levels. Thus, target detection at other multi-resolution levels is compromised.

## IV. Methodology

A robotic task may contain one or multiple targets. A set of features stored in the robot's memory represent these targets. The feature is a portion of the image that contains relevant information. It is encoded into a series of numbers by a descriptor.

During an operation, the robot matches the descriptors of the target, stored in its memory, with those acquired by sensors. Matching descriptors results in false (outliers) or true (inliers) matches, which can be obtained by mask of the Random Sample Consensus (RANSAC). We combine these values into a inliers rate (see Equation 3).

$$p = \frac{Inliers}{Outliers + Inliers} \tag{3}$$

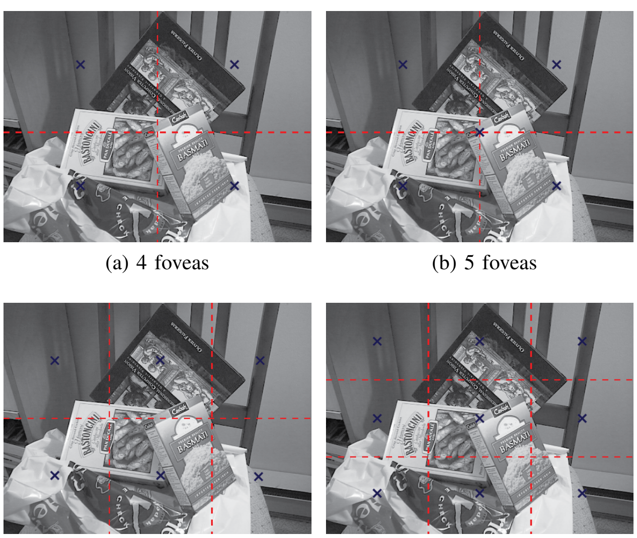
We use MMMF to add foveas in the images acquired by sensors. These foveas have $m$ levels, each one extracts features and has a inliers rate associated. There is no guarantee that the target is at all levels. Therefore, we propose a level weighting approach, where the level weights are given by the proportion of transformed pixels, defined by Equation 4. In this equation, $R_i$ represents the number of pixels of level $i$. Thus, the region with the highest number of pixels in the cartesian domain will contribute less to target detection compared to the levels that have the lowest number of pixels in the cartesian domain.

$$w_i = \frac{R_{m-i}}{\sum_{k=0}^{m} R_k}, \text{ where } i \in [0, m] \quad (4)$$

The linear combination of weighted levels and inlier rates provides a fovea detection function (see Equation 5). It returns a maximum value when the fovea is in the center of the target and this value decreases when we increase the distance between the fovea and the target. This behavior is similar to the results obtained in the study of photoreceptors developed by Osterberg [44].

$$f = \sum_{k=0}^{m} w_k p_k \quad (5)$$

We obtain a detection surface through successive placements of the fovea by the image. This procedure requires a huge computational cost, making its use in robotic applications unfeasible. Thus, instead of repositioning the fovea by all pixels in the image, we could (1) use spacing around the x and y axes, but this procedure also requires a lot of processing or (2) keep foveas fixed at strategic points in the image. We use the second possibility and we subdivided the image into 4, 5, 6, and 9 regions as shown in Figure 4. For each subregion we add a fovea (see Figure 5).

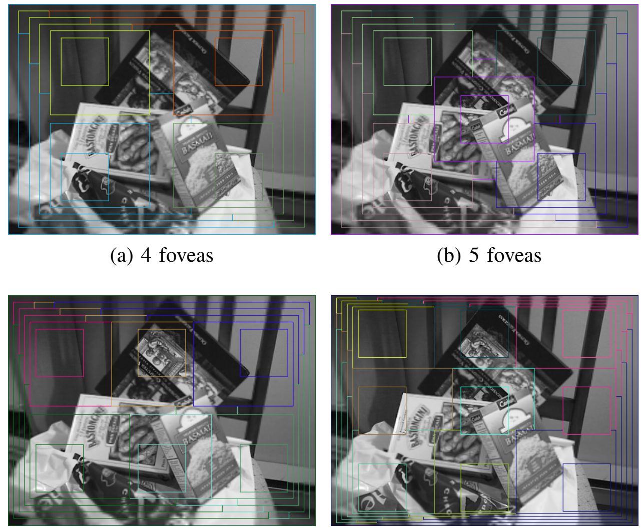


(a) 4 foveas (b) 5 foveas

(c) 6 foveas (d) 9 foveas

Fig. 4. The red dashed lines represent the subdivisions of regions and the positions marked with a 'x' blue are the central positions of each subregion, where the foveas will be added.

Each fovea of these settings has a target detection value obtained by Equation 5. We propose to use these information in methods based on gradient descent, maximum likelihood estimator, multilateration, trilateration, and weighted barycentric coordinates to estimate the target position.



(a) 4 foveas (b) 5 foveas

(c) 6 foveas (d) 9 foveas

Fig. 5. Foveas distributed in subregions.

### *A. Region Reduction and Intersection of Local Gradients*

The local gradient is calculated by the difference between the central fovea and the neighboring foveas. The region reduction method iteratively reduces the search region according to the direction of the local potential field. The local gradient intersection method indicates an estimate of the target's position according to the intersection of two local gradients. These methods make use of 3 arrangements: (1) using the 8 positions around the fovea, (2) using the 4 positions (up, down, left and right) and (3) using the 4 positions (northeast, southeast, southwest and northwest).

### *B. Maximum Likelihood Estimator (MLE)*

The detection surface discussed above converges to a multivariate normal distribution, where the mean represents the target position. We use MLE and conclude that given several foveas distributed across the image, it is possible to estimate the target's position using the average of their positions weighted by their detection rate.

### *C. Trilateration and Multilateration*

In these methods, we interpreted the detection of the fovea as a circumference that increases when the detection value decreases. In other words, there is an inversely proportional relationship between the radius of the circle and the detection value. These methods estimate the target position using the intersection between the circumferences and, for that, they need information about the distances between the foveas.

### *D. Weighted Barycentric Coordinates*

This method implements an interpolator that uses three foveas to estimate the target position. We selected three foveas according to their proximity and maximum detection value together. We weighted their positions using their target detection values.

## V. Results

We used the Toy video from Visual Tracker Benchmark [1] to validate our theory. This video contains a sequence of 271 frames under scale variation, fast movement and rotation of the tracking target. The target was extracted from the first frame of the video.

We consider foveas with 4 levels, dimension $W = (80, 80)$ and SURF feature extraction. However, we noticed the presence of oscillations, caused by false positives, which impair target detection. We solve this problem by defining a high-pass filter on the amount of matches using a threshold equal to 15. Using these information, we run an algorithm that applies the methodology defined in section IV. The methods IV-A to IV-D return an estimated position $(\hat{x}, \hat{y})$ of the center of the target and this estimated position can be evaluated using the ground truth available from the dataset.

However, ground truth provides a rectangular region that outlines the target. Therefore, we define Equation 6, where $(x_c, y_c)$ is the center of the region, $x$ and $y$ are positions of the upper left corner of the rectangle, and $sx$ and $sy$ are the dimensions of the $x$ and $y$ coordinates, respectively. Moreover, we calculate the error using the euclidean distance between the estimated positions and the center position of the ground truth region.

$$(x_c, y_c) = \left( \frac{x + sx}{2}, \frac{y + sy}{2} \right) \tag{6}$$

The results of methods based on gradient descent are presented in Table I. Our results demonstrate that the mean error of these methods are greater than 76 pixels and there is a scattering of the errors generating a high standard deviation. This information indicates that the methods may not converge to the target position.

TABLE I
Table With Mean Error (Err) and Standard Deviation (SD) With 8 Foveas (8 Fs), 4 Foveas in Up, Down, Left and Right (4 Fs1) and 4 Foveas in Northeast, Southeast, Southwest and Northwest (4 Fs2).

| Method | 8 Fs | | 4 Fs1 | | 4 Fs2 | |
|---|---|---|---|---|---|---|
| | Err | SD | Err | SD | Err | SD |
| Region Reduction | 82,29 | 37,97 | 80,37 | 36,37 | 100,48 | 45,67 |
| Intersection Gradients | 85,77 | 52,18 | - | - | 76,50 | 40,17 |

Table II shows that the mean error of estimated position by multilateration, trilateration, and barycentric coordinates are, on mean, $1,5$ times less than the MLE error. This happens because the statistical mean varies a lot depending on the data. In future works, we intend to implement this same approach using the median. Furthermore, we can observe that the standard deviations of the multilateration errors for the 4 settings show little variation due to the use of all fixed foveas to estimate the target position.

[1]Dataset available on the link: http://cvlab.hanyang.ac.kr/tracker_benchmark/index.html

## VI. Conclusion

Multifoveation can keep several targets in evidence at the same time, and this benefits several robotic applications. However, we realized that this technique is underutilized in literature. Therefore, we proposed here an approach that uses dynamic and fixed foveas to estimate the target's position in the image.

Basically, we shifted a mobile fovea by the image and found a detection surface, where the peak corresponds to the target's center. This surface has local minimums which spoils otimization routines. We solved it using a high-pass filter on the number of matches.

In this study, we implemented a region reduction and intersection of local gradient algorithms. The first converges to the target's position but consumes a lot of time and computational resources, whereas the second does not converge and requires more information. These algorithms do not use fixed foveas.

We tried to explore the surface properties using the MLE algorithm, but the accuracy of this method is proportional to the amount of fovea close to the target. Posteriorly, we realized that the algorithms were not exploring the geometry of the foveas and proposed trilateration, multilateration and barycentric coordinates. These last algorithms are efficient in target detection.

We conclude that multifoveation helps robots in the target detection task. However, some issues involving (1) multiple targets and (2) target absence were not discussed in this study and are essential problems to be studied in the context of visual detection in Robotics. As such, they can be matter of further research.

## Acknowledgment

The authors would like to thank Brazilian Sponsoring Agency for Higher Education (CAPES) for the grants of Luiz M. G. Gonçalves (88887.091733/2014-01) and Petrucio R. T. Medeiros (88882.145745/2017-01) and Brazilian Agency for Research Sponsoring (CNPq) for the grant of Luiz M. G. Gonçalves under number 8546600060032961.

## References

[1] G. Zhang, L. Bin, Z. Li, W. Cong, H. Zhang, S. Hong, H. Weijian, and Z. Tao, "Development of robotic spreader for earthquake rescue," in *2014 IEEE International Symposium on Safety, Security, and Rescue Robotics*, 10 2014.

[2] B. Peters, P. Armijo, C. Krause, S. Choudhury, and D. Oleynikov, "Review of emerging surgical robotic technology," *Surgical endoscopy*, vol. 32, no. 4, pp. 1636–1655, 2018.

[3] M. Diab-El Schahawi, W. Zingg, M. Vos, H. Humphreys, L. Lopez-Cerero, A. Fueszl, J. R. Zahar, and E. Presterl, "Ultraviolet disinfection robots to improve hospital cleaning: Real promise or just a gimmick?" *Annual Review of Psychology*, vol. 10, no. 33, pp. 2047–2994, 2021.

[4] H. Pashler, J. C. Johnston, and E. Ruthruff, "Attention and performance," *Annual Review of Psychology*, vol. 52, no. 1, pp. 629–651, 2001.

[5] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, Oct 2004.

TABLE II
TABLE WITH MEAN ERROR (ERR) AND STANDARD DEVIATION (SD) FOR FOVEAS' SETTINGS (FS). THE RED AND BLUE COLORS REPRESENT THE WORST AND BEST VALUES, RESPECTIVELY.

| Method | 4 Fs | | 5 Fs | | 6 Fs | | 9 Fs | |
|---|---|---|---|---|---|---|---|---|
| | Err | SD | Err | SD | Err | SD | Err | SD |
| MLE | 89,62 | 69,4 | 72,73 | 62,17 | 82,95 | 60,98 | 68,51 | 59,12 |
| Multilateration | 51,94 | 26,17 | 52,05 | 26,21 | 51,5 | 26,51 | 51,71 | 26,33 |
| Trilateration | 49,56 | 26,64 | 58,11 | 30,86 | 51,94 | 33,58 | 53,16 | 37,24 |
| Barycentric Coordinates | 50,48 | 28,43 | 54,29 | 31,01 | 52,26 | 35,82 | 54,95 | 38,24 |


[6] A. Quan, C. Herrmann, and H. Soliman, "Project vulture: A prototype for using drones in search and rescue operations," in *15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. Santorini Island, Greece: IEEE, May 2019, pp. 619–624.

[7] A. Basu, A. Sullivan, and K. J. Wiebe, "Variable resolution teleconferencing," in *Proceedings of IEEE Systems Man and Cybernetics Conference - SMC*. Le Touquet, France: IEEE, Oct. 1993.

[8] A. Basu and K. J. Wiebe, "Videoconferencing using spatially varying sensing with multiple and moving foveae," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 2 - Conference B: Computer Vision and Image Processing*. Jerusalem, Israel: IEEE, oct 1994, pp. 30–34.

[9] ——, "Enhancing videoconferencing using spatially varying sensing," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 28, pp. 137–148, mar 1998.

[10] P. Dario, M. Bergamasco, A. Fiorillo, and R. Leonardo, "Geometrical optimization criteria for the design of tactile sensing patterns," in *IEEE International Conference on Robotics And Automation*. IEEE, Apr. 1986, pp. 1268–1273.

[11] N. Dhavale and L. Itti, "Saliency-based multi-foveated mpeg compression," in *Proceedings of Seventh International Conference on Signal Processing and its Applications*. IEEE, Jul. 2003, pp. 229–232.

[12] G. Pioppo, R. Ansari, A. A. Khokhar, and G. Masera, "Low-complexity video compression combining adaptive multifoveation and reuse of high-resolution information," in *Proceedings of the International Conference on Image Processing, (ICIP)*, October 2006, pp. 3153–3156.

[13] R. B. Gomes, R. Q. Gardiman, L. E. C. Leite, B. M. Carvalho, and L. M. G. Goncalves, "Towards real time data reduction and feature abstraction for robotics vision," *Robot Vision*, pp. 345–362, mar 2010.

[14] X. C. Benjamim, R. B. Gomes, A. F. Burlamaqui, and L. M. G. Gonçalves, "Visual identification of medicine boxes using features matching," in *2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS) Proceedings*, July 2012, pp. 43–47.

[15] L. Uhr, "Layered "recognition cone" networks that preprocess, classify, and describe," *IEEE Transactions on Computers*, vol. C-21, no. 7, pp. 758–768, July 1972.

[16] L. P. Kobbelt, *Handbook of Computer Aided Geometric Design*. Elsevier Academic Press, 2002.

[17] L. M. G. Gonçalves, "A robotic control system for integration of multimodal sensory information (in portuguese)," Ph.D. dissertation, Federal University of Rio de Janeiro, 1999.

[18] R. B. Gomes, L. M. G. Gonçalves, and B. M. Carvalho, "Real time vision for robotics using a moving fovea approach with multi resolution," in *IEEE International Conference on Robotics and Automation*. Pasadena, CA, USA: IEEE, May 2008, pp. 19–23.

[19] R. B. Gomes, B. M. de Carvalho, and L. M. G. Gonçalves, "Visual attention guided features selection with foveated images," *Neurocomputing*, vol. 120, pp. 34 – 44, 2013, special issue on Image Feature Detection and Description.

[20] R. Gomes, "Feature selection guided by visual attention in images with fovea (in portuguese)," Ph.D. dissertation, Federal University of Rio Grande do Norte, 2013.

[21] P. R. T. Medeiros, "Multifoveation in multiresolution with mobile foveas (in portuguese)," Master's thesis, Federal University of Rio Grande do Norte, 2016.

[22] P. R. T. Medeiros, R. B. Gomes, E. W. G. Clua, and L. Gonçalves, "Dynamic multifoveated structure for real-time vision tasks in robotic systems: A tool for removing redundancy in multifoveated image processing," *Journal of Real-Time Image Processing*, vol. 17, no. 5, 2019.

[23] F. F. Oliveira, A. A. A. Souza, M. A. C. Fernandes, R. B. Gomes, and L. M. G. Goncalves, "Efficient 3d objects recognition using multifoveated point clouds," *Sensors*, vol. 18, no. 7, 2018.

[24] F. L. Lim, A. W. West, and S. Venkatesh, "Tracking in a space variant active vision system," in *Proceedings of the 13th International Conference on Pattern Recognition*, vol. 1. Vienna: IEEE, Aug. 1996, pp. 745–749.

[25] A. Ude, V. Wyart, L. Lin, and G. Cheng, "Distributed visual attention on a humanoid robot," in *5th IEEE-RAS International Conference on Humanoid Robots*, 2005, pp. 381–386.

[26] P. Camacho, F. Arrebola, and F. Sandoval, "Adaptive multifovea sensors for mobiles tracking," in *IEEE International Conference on Electronics, Circuits and Systems*. Lisboa: IEEE, Sep. 1998, pp. 449–452.

[27] ——, "Multiresolution sensors with adaptive structure," in *IECON 98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No.98CH36200*. Aachen, Germany: IEEE, Sep. 1998, p. 1230–1235.

[28] J. A. Rodríguez, C. Urdiales, A. Bandera, and F. Sandoval, "Nonuniform video coding by means of multifoveal geometries," *International Journal of Imaging Systems and Technology*, vol. 12, pp. 27–34, Jan. 2002.

[29] J. Wei and Z. Li, "On active camera control and camera motion recovery with foveate wavelet transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 896–903, Aug. 2001.

[30] S. Sankara, R. Ansari, and A. Khokhar, "Adaptive multifoveation for low-complexity video compression with a stationary camera perspective," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5685, 2005, pp. 5685–5697.

[31] N. J. Butko, L. Zhang, G. W. Cottrell, and J. R. Movellan, "Visual saliency model for robot cameras," in *IEEE International Conference on Robotics and Automation*. Pasadena, CA, USA: IEEE, May 2008, pp. 2398–2403.

[32] N. J. Butko and J. R. Movellan, "I-pomdp: An infomax model of eye movement," in *IEEE International Conference on Development and Learning (ICDL)*. Monterey, CA, USA: IEEE, Aug. 2008, pp. 669–672.

[33] ——, "Optimal scanning for faster object detection," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. Miami, FL, USA: IEEE, Jun. 2009, pp. 2751–2758.

[34] ——, "Infomax control of eye movements," *IEEE Transactions on Autonomous Mental Development*, vol. 2, pp. 91–107, Jun. 2010.

[35] ——, "Learning to look," in *IEEE International Conference on Development and Learning (ICDL)*. Ann Arbor, MI, USA: IEEE, Aug. 2010, pp. 70–75.

[36] N. J. Butko, "Active perception," Ph.D. dissertation, University of California, San Diego, 2010.

[37] W. A. Talbott, H. C. Huang, and J. Movellan, "Infomax models of oculomotor control," in *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, 2012, pp. 1–6.

[38] A. Cavallaro, D. Hands, and T. Popkin, "Multi-foveation filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, vol. 1, 2009, pp. 669–672.

[39] T. Xu, T. Zhang, K. Kühnlezn, and M. Buss, "Attentional object detection with an active multi-focal vision system," *International Journal of Humanoid Robotics*, vol. 7, no. 2, pp. 223–243, jan 2010.

[40] G. B. Soos and C. Rekeczky, "Elastic grid based analysis of motion field for object-motion detection in airborne video flows," in *2007 IEEE International Symposium on Circuits and Systems*, 2007, pp. 617–620.

[41] B. G. Soos, V. Szabo, and C. Rekeczky, "Elastic grid-based multi-fovea algorithm for real-time object-motion detection in airborne surveillance," in *Cellular Nanoscale Sensory Wave Computing*, C. Baatar, W. Porod, and T. Roska, Eds. Boston: Springer, 2009, pp. 181–213.
[42] B. G. Soos, "Multi-fovea architecture and algorithms based on cellular many-core processor arrays," Ph.D. dissertation, Faculty of Information Technology, 2010.
[43] A. Zarándy, C. Rekeczky, P. Földesy, R. C. Galán, G. L. Cembrano, B. G. Soos, A. R. Vázquez, and T. Roska, "Viscube: A multi-layer vision chip," in *Focal-Plane Sensor-Processor Chips*, A. Zarándy, Ed. New York: Springer-Verlag, 2011, pp. 181–208.
[44] G. Osterberg, "Topography of the layer of rods and cones in the human retina," *Acta ophthalmologica: Supplementum*, vol. 6, p. 1–102, 1935.