Salient Foreground Object Detection based on Sparse Reconstruction for Artificial Awareness

Jingyu Wang[1], Ke Zhang[1], Kurosh Madani[2], Christophe Sabourin[2] and Jing Zhang[3]

[1]*School of Astronautics, Northwestern Polytechnical University, Xi'an, China*

[2]*Signals Images & Intelligent Systems Laboratory (LISSI/EA3956), Université Paris-Est, Paris-Lieusaint, France*

[3]*School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China*

Keywords: Foreground Object Detection, Informative Saliency, Sparse Representation, Reconstruction Error, Artificial Awareness.

Abstract: Artificial awareness is an interesting way of realizing artificial intelligent perception for machines. Since the foreground object can provide more useful information for perception and informative description of the environment than background regions, the informative saliency characteristics of the foreground object can be treated as a important cue of the objectness property. Thus, a sparse reconstruction error based detection approach is proposed in this paper. To be specific, the overcomplete dictionary is trained by using the image features derived from randomly selected background images, while the reconstruction error is computed in several scales to obtain better detection performance. Experiments on popular image dataset are conducted by applying the proposed approach, while comparison tests by using a state of the art visual saliency detection method are demonstrated as well. The experimental results have shown that the proposed approach is able to detect the foreground object which is distinct for awareness, and has better performance in detecting the information salient foreground object for artificial awareness than the state of the art visual saliency method.

1 INTRODUCTION

Due to the perception importance and distinctive representation of visual information, it dominates the perceptual information acqusited from environment. Thus, visual object detection plays a vital role in the perception process of the surrounding environment in our lives. Since machines that with certain level of intelligence have been frequently depolyed in the dangerours or complex environment to accomplish complicate tasks instead of human beings more than ever before, the accuracy and efficiency of visual channel perception is extremely crucial and highly important. However, as image requires much more resource for higher level processing, it is difficult and practically impossible for artificial machines to exhaustively analyze all the image data.

As human perception is such a sophisticated and purely biological process, only some features of the phenomenal world have been tentatively modeled or even implemented in robotic systems (Fingelkurts, 2012). Alternatively, an interesting way of achieving human-like intelligent perception has been proposed as a lower level and preliminary stage of artificial consciousness, which is known as awareness (Ramík, 2013).

According to the discussion of (Reggia, 2013), the artificial conscious awareness or the information processing capabilities associated with the conscious mind would be an interesting way, even a door to more powerful and general artificial intelligence technology. However, very little work has been done to realize the awareness ability in machines. The difficulty is that current approaches always focus on the computational model of information processing, while the human awareness characteristic is hard to be simulated.

From the perspective of human visual awareness, it is obvious that we always intend to focus on the most informative region or object in an image in order to efficiently analyze what we have observed. This biological phenomenon is known as the visual saliency and has been well researched for years. Compared to the background regions, the foreground objects in an image contain more useful and unique informative cues in the perceptual process from the perspective of visual perception, which means that the foreground object is considered to be informative

Wang J., Zhang K., Madani K., Sabourin C. and Zhang J..
Salient Foreground Object Detection based on Sparse Reconstruction for Artificial Awareness.
DOI: 10.5220/0005571204300437
In *Proceedings of the 12th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2015), pages 430-437
ISBN: 978-989-758-123-6

salient. It is the perceptual awareness that makes the foreground objects more interesting and valuable, so that they can be treated as informative salient by human beings. Therefore, the detection of salient foreground object is a crucial and fundamental task in realizing the intelligent awareness for artificial machines.

From object detection point of view, foreground objects can be either salient or non-salient to human vision (see Figure 4 in Section 4.2). However, the foreground object has informative saliency features compared to the background region. Thus, novel approach that can detect the saliency property of foreground object in information level is required to mimic the human awareness characteristic. The rest part of this paper is organized as follows. Section 2 briefly introduces and discusses the state of the art of related works. Section 3 describes the proposed detection approach in detail. Section 4 demonstrates the experiment setup and gives results, while the discussion and comparison are presented afterwards. Section 5 summarizes the conclusions.

2 RELATED WORKS

Traditional visual saliency detection approaches have been well researched and can be generally illustrated into local and global schemes. Most of them are based on the centre-surround operator, contrast operator as well as some other saliency features. Since these features are mostly derived in pixel level from image, the intrinsic information of object such as objectness is rarely taken into account. As a result, the detected salient regions could not cover the expected objects in certain circumstances, especially when multiple objects exist or the objects are informative salient.

In the research work of (Wickens and Andre, 1990), the term of objectness is characterized as the visual representation that could be correlated with an object, thus an objectness based visual object shape detection approach is presented. The advantage of using objectness is that, it can be considered as a generic cue of object for further processing, which is the way more like the perceptual characteristic of our visual perception system. Notably, in (Alexe et al., 2010) and (Alexe et al., 2012) the objectness is used as a location prior to improve the object detection methods, the yielded results have shown that it outperforms many other approaches, including traditional saliency, interest point detector, semantic learning as well as the HOG detector, while good results can be achieved in both static images and videos. Thereafter, in the works of (Chang et al., 2011), (Spampinato et al., 2012) and (Cheng et al., 2014) the objectness property is used as the generic cue that to be combined with many other saliency characteristics to achieve a better performance in salient object detection, the experimental results of which have proved that objectness is an important property as well as an efficient way in the detection of objects and can be applied to many object-related scenarios. Therefore, it is worthy of researching the approach of detecting information salient foreground objects by measuring objectness and conduct it in an autonomous way. Moreover, inspired by the early research of (Olshausen and Field, 1997) which revealed the biological foundation of sparse coding, researches of (Mairal et al., 2008) and (Wright et al., 2009) have shown that the sparse representation is a powerful mathematical tool for representing and compressing high dimensional signal in computer vision, including natural image restoration, image denoising and human face recognition.

In (Ji et al., 2013), a foreground object extraction approach is proposed for analyzing the image of video surveillance, in which the background region is represented by the spatiotemporal spectrum in 3D DCT domain while the foreground object pixels are identified as an outlier of the sparse model of the spectrum. By updating the background dictionary of sparse model, the dissimilarity between background and foreground can be measured and the foreground object can be extracted. Experiment on video frames shows a good performance of the proposed approach, however, the images only contain simple foreground object and the objectness property is not taken into account. Meanwhile, (Sun et al., 2013) proposed an automatic foreground object detection approach, in which the robust SIFT trajectories are constructed in terms of the calculated feature point probability. By using a consensus foreground object template, object in the foreground of video can be detected. Despite that the experiment results derived from real videos have proved the effectiveness of proposed approach, the applied objects are in a close-up scene and are both informative salient and visually salient, which limits its application in real world.

Recently, (Biswas and Babu, 2014) proposed a foreground anomaly detection approach based on the sparse reconstruction error for surveillance, in which the applied enhanced local dictionary is computed based on the similarity of usual behavior with spatial neighbors in the image. The experiment results have shown better detection performance compared to the traditional approaches, which indicate that the error of sparse reconstruction can represent the objectness property of foreground objects that are informative salient and describe the perceptual informative dissimilarity between foreground and background.

As motivated before, a reconstruction error based salient foreground object detection approach is proposed in this paper. Different from other works, we propose to use the informative saliency instead of visual saliency. To be specific, the informative saliency is described as the objectness property and measured by the sparse reconstruction error. The foreground object with salient informative meaning is detected by calculating the reconstruction error of the feature matrix over an overcomplete background dictionary which describes the dissimilarity between object and background. Since the theoretical basis and derivation of sparse representation has been well studied, the detailed introduction of sparse coding is omitted while the illustrations of key components of our approach will be given in detail.

3 SALIENT FOREGROUND OBJECT DETECTION

3.1 Overview of Approach

In general, the proposed approach in this paper consists of two stages which are the learning of background dictionary and the sparse reconstruction error computation in different scales, respectively.

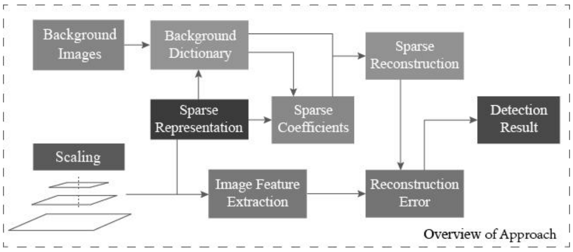


Figure 1: The overview of proposed detection approach, in which the blue arrow indicates the procedure of processing.

To be specific, foreground objects are considered to be much more informative salient with respect to the background region, as the foreground objects are more interesting and informative salient to human awareness than the background. The overview of the proposed approach is illustrated in Figure 1.

As demonstrated in Fig. 1, the visual image of the environment will be processed in different scales, the goal of which is to cover objects with different sizes. Notably, to simplify the question, the objects with ordinary and fixed sizes are considered in this paper. The dictionary is pre-learned by using a set of background images, while the Gabor features of input image are obtained.

Thereafter, the sparse coefficients are computed and used to generate the reconstruction feature vector. Finally, the errors of sparse reconstruction will be calculated between the original Gabor features and the reconstructed Gabor features, which indicate the informative saliency of local image patches in different scales. By assigning a threshold of reconstruction error, the patches with error value larger than the threshold are the potential locations of informative salient foreground regions.

The contribution of our work is the using of the reconstruction error, which is computed between the input and reconstructed image feature matrix. Since sparse decomposition is an optimal approximation, the reconstructed feature could be slightly different from the input feature vector, due to the dissimilarity of objectness between the foreground objects and background. Consequently, the sparse reconstruction error is applied as the representation of informative salient foreground object for awareness.

3.2 Sparse Reconstruction

3.2.1 Image Feature Extraction

Since the kernel of Gabor filters is believed to be a good model that similar to the receptive field profiles of cortical simple cells (Hubel and Wiesel, 1968), Gabor filter is used to capture the local feature of image in multiple frequencies (scales) and orientations due to its good performance of spatial localization and orientation selection. The two-dimensional Gabor function can therefore enhance the features of edge, peak and ridge and robust to illumination and posture to a certain extent.

Considering the statistic property of image, the kernel of Gabor function can be defined as (Liu and Wechsler, 2002)

$$\psi_{u,v}(x,y)=\frac{\|k_{u,v}\|^2}{\sigma^2}\exp\left(\frac{\|k_{u,v}\|^2(x^2+y^2)}{-2\sigma^2}\right)\cdot(\exp(ik_{u,v}\cdot\begin{pmatrix}x\\y\end{pmatrix})-\exp\left(-\frac{\sigma^2}{2}\right)) \quad (1)$$

where u and v represent the orientation and scale of the Gabor kernels, x and y are the coordinates of pixel location, $\|\cdot\|$ denotes the norm operator and σ determines the ratio of the Gaussian window width to wavelength. Particularly, the wave vector $k_{u,v}$ is defined as follows

$$k_{u,v}=k_v e^{i\phi_u} \quad (2)$$

where $k_v=k_{max}/f_s^v$ and $\phi_u=\pi u/8$, in which k_{max} is the maximum frequency and f_s is the spacing factor between kernels in the frequency domain. By using different values of u and v, a set of Gabor filters with different scales and orientations can be obtained.

Meanwhile, the Gabor features of an image are the convolution of the image with a set of Gabor filters in the filter bank which defined by Eq.(1). The formulation of the Gabor feature derived from the image $I(x,y)$ can be defined as

$$G_{u,v}(x,y)=I(x,y)*\psi_{u,v}(x,y) \quad (3)$$

where $G_{u,v}(x,y)$ is the Gabor feature of image $I(x,y)$ in orientation u and scale v, the * symbol represents the convolution operator.

As foreground objects in the environment are mostly regular in shape and contour, thus the scale parameters is set to be 3 so as to cover objects with different sizes in 3 scales, and the orientation is set to be 2 to obtain the Gabor features of vertical and horizontal axes.

3.2.2 Background Dictionary Learning

Considering the general problem model of sparse representation, the sparse representation of a column signal $x \in \Re^n$ with the corresponding overcomplete dictionary $D \in \Re^{n \times K}$, in which the parameter K indicates the number of dictionary atoms, can be described by the following sparse approximation problem as

$$\min_{\alpha} \|\alpha\|_0 \text{ subject to } \|x - D\alpha\|_2 \leq \varepsilon \quad (4)$$

where $\|\cdot\|_0$ is the l^0-norm which counts the nonzero entries of a vector, α is the sparse coefficient and ε is the error tolerance.

According to the research work of (Davis et al., 1997), the extract determination of the sparsest representation which defined in Eq.(4) has been known as a non-deterministic polynomial (NP) -hard problem. This means that the sparsest solution of Eq.(4) has no optimal result but trying all subsets of the entries for signal x which could be computational unavailable.

Nevertheless, researches have proved that if the sought solution x is sparse enough, the solution of the l^0-norm problem could be replaced by the approximated version of the l^1-norm as

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } \|x - D\alpha\|_2 \leq \varepsilon \quad (5)$$

where $\|\cdot\|_1$ is the l^1-norm. The similarity in finding sparse solution between using the l^1-norm and the l^0-norm has been supported by the work of (Donoho and Tsaig, 2008).

Current dictionary learning methods can be categorized into two kinds based on the discussion in (Rubinstein et al., 2010), which are the analytic approach and the learning-based approach. The first approach refers to the dictionaries which generated from the standard mathematical models, such as Fourier, Wavelet and Gabor, to name a few, which have no informative meaning correlate to the natural images.

On the other hand, the second approach uses machine learning based techniques to generate the dictionary from image examples. Therefore, the obtained dictionary could represent the examples in a close manner. Compared to the first approach which prespecifies the dictionary atoms, the second way is an adaptation process between the dictionary and examples from the machine learning perspective. Although the analytic dictionary is simple to be implemented, the learning-based dictionary has a better performance in image processing.

Considering the requirement of our work, the dictionary learned based on image examples is used to provide informative description of the background image. In this paper, we simply apply a frequently used dictionary learning algorithm that described in (Aharon, 2006) to generate the sparse atoms of the overcomplete dictionary.

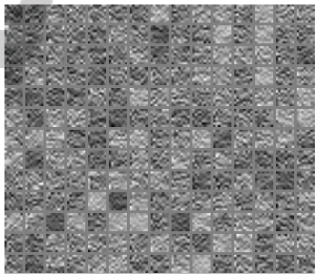


Figure 2: The learned background dictionary in gray scale.

Thus, the dictionary D is used to represent the image features of background regions. By using the dictionary, sparse coding is able to approximately represent the input features as a linear combination of sparse atoms. Particularly, the gray scale image of learned background dictionary is shown in Figure 2, in which the sequence of local image patch indicates the visualization representation of dictionary atom.

3.2.3 The Computation of Reconstruction Error

When the aforementioned background dictionary D has been learned, the objectness property of foreground object can be obtained by calculating the reconstruction error of input feature vector derived from a detection window over the learned dictionary. The underlying assumption of this approach is that, as the representation of a local image patch, each local feature vector contains the objectness property.

Meanwhile, objectness is characterized here as the dissimilarity between input feature vector and background dictionary. By using the obtained sparse representation coefficients α of a feature vector generated from the dictionary, the reconstructed feature vector can be restored by applying an inverse operation of sparse decomposition. However, since the reconstructed feature vector derived from sparse coding is the approximation of the original feature vector, a reconstruction error between these two vectors can be calculated to indicate the dissimilarity between the current local image patch and the background image. Thus, the objectness property of each detection window could be measured for foreground object detection.

Assume x_i, $i=1,\ldots,N$ is the corresponding feature vector for i^{th} local image patch, the sparse coefficient can be computed by coding each x_i over the learned dictionary D based on the l^1-minimization as

$$\min_{\alpha} \|\alpha\|_1 \text{ subject to } x = D\alpha \tag{6}$$

In order to obtain the sparse coefficient α, many decomposition approaches have been proposed so far and proved to be effective, such as Basis Pursuit (BP), Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP) and Least Absolute Shrinkage and Selection Operator (LASSO).

Considering the computational cost and the requirement of research goal, the LASSO algorithm (Tibshirani, 1996) is applied to compute the sparse coefficients α of the input feature vector. Thus, the reconstructed feature vector $\hat{x}$ can be calculated based on the sparse coefficients as

$$\hat{x} = D\alpha \tag{7}$$

Since $\hat{x}$ is the approximation solution result of x, the reconstruction error can be quantitatively given as

$$\varepsilon = \|x - \hat{x}\|_2^2 = \|x - D\alpha\|_2^2 \tag{8}$$

where $\|\cdot\|_2^2$ denotes the Euclidean distance.

Particularly, as the input image is processed in multiple scales to reveal the characteristics of object in different sizes, the input feature vector x_i of each scale will be evaluated differently as

$$\varepsilon_{fo} = \sum \left\{ \forall \varepsilon_i^{s_j} > \rho_{s_j} \right\}, \quad j = 1,2,3 \tag{9}$$

where ε_i denotes the reconstruction error of i^{th} local image patch in each scale and ε_{fo} represents the set consists of reconstruction errors ε_i^{Sj} that larger than the error threshold of ρ_{Sj} in S_j scale. Thus, the information salient object can be extracted by finding the detection window which indicated by ε_{fo}.

4 EXPERIMENTAL RESULTS

To validate the effectiveness of the proposed approach, natural images taken from real world including both outdoor and indoor environment are applied. The object images of clock, phone, police car, bus and tree are chosen to generate the experiment dataset. To compare the performance of proposed approach with the state of the art visual saliency detection approach, the method proposed by (Perazzi et al., 2012) is used to obtain the visual saliency detection results.

4.1 Experimental Setup

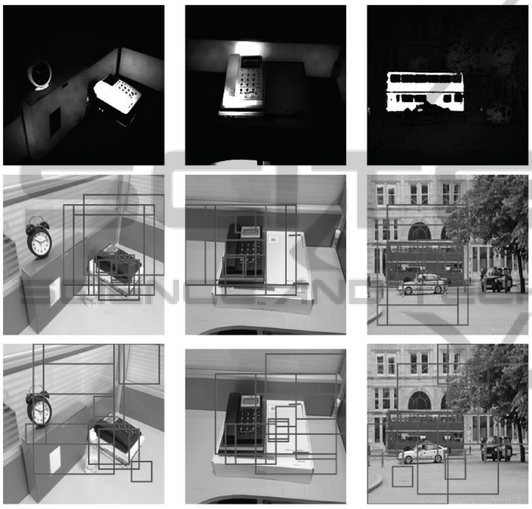
In general, the clock, phone and the white box underneath the phone are the expected foreground objects in the test images of indoor environment, while the police car, bus and tree are considered to be informative salient and are the expected objects in outdoor environment. To ensure the quality and resolution of the test images can represent the actual requirement of real world, the images of clock and phone are taken in a typical office room, while the images of police and bus are randomly selected from the Internet via *Google.fr*. There are 150 pictures which randomly chosen from internet with different colors and shapes for training the dictionary. The pictures rarely have foreground objects and are taken from ordinary environments which can be commonly seen in human world. The learning process is conducted on the laptop with Intel i7-3630QM cores of 2.4 GHz and 8 GB internal storage, 100 iterations are deployed as a compromise of time and computational cost.

Notably, other objects that are simultaneously appeared in the pictures which could be treated as interferences, while some of which are also visual salient to human visual perception.

4.2 Results and Discussion

In Figure 3, the visual saliency images derived from the approach of (Perazzi et al., 2012) are given as in the first row, while detection results of the information salient foreground objects by applying visual saliency method and proposed approach are shown in the second and third row, respectively.

The second images from Figure 3(a) and 3(c) of Figure 4 show that, visually salient objects could be detected while informative salient foreground object can not be located, such as the clock in Figure 3(a) and the car under the tree in Figure 3(c). Though this could has little influence to the further processing while the salient foreground object is not the expected object, such as the car under the tree in Figure 3(c), it still could lead to a failure in potential further processing, such as object classification.



(a) (b) (c)

Figure 3: The salient foreground objects detection results of (a) clock, (b) phone and (c) police car, bus and tree.

The images in the third row from both Figure 3(a) and 3(c) have shown that, all the salient foreground objects could be covered with at least one detection window. Particularly, both of the expected objects of clock and phone are detected by using objectness based approach as shown in the last image of Figure 3(a), and the detection windows in the last image of Figure 3(c) are more close to the expected police car compare to the visual saliency detection results in the third image of Figure 3(c). Particularly, the detection windows in Figure 3(c) can also cover the tree that in the foreground. These two examples have shown that the proposed method is able to detect the informative salient foreground objects successfully, when the expected objects are not visually salient.

Meanwhile, a set of test images that consists of a visually salient object of phone is given in Figure 3(b). Though the image from second row of Figure 3(b) shows that the result of using visual saliency detection approach is correct in detecting the phone, but the white box can not be fully covered by the detection window. However, better detection result that the entire box and phone can be located by the detection window by using the proposed as shown in the last image of Figure 3(b). Nevertheless, there are still some mismatched detection windows exist in the results obtained by using proposed approach, the explanation for this limitation is that only a small number (N=150) of background images are applied in our work to train the background dictionary. Thus, the dictionary is not well constructed based on the experimental data and not all the background images can be comprehensively represented by the learned dictionary. In fact, informative boundary between background and foreground is ambiguous and even subjectively different according to the differentials in visual perception system of different people.

(a) (b) (c)

Figure 4: The foreground objects detection results of three test images.

Furthermore, to compare the performances of both visual saliency method and proposed method, the detection results derived from the frequently used PASCAL VOC2007 dataset (Everingham et al., 2008) are shown in Figure 4 to show the importance of using the objectness property as the informative saliency feature in foreground object detection. In Figure 4, three example images from both indoor and outdoor environments have been given to demonstrate the differential detection results. To be specific, the original images, saliency maps and detection results are shown in the first, second and third row, respectively. It can be seen from the original images that informative salient foreground object with respect to the awareness characteristic in each test image can be illustrated as: two sheep in Figure 4(a), chairs and small sofas in Figure 4(b) and the computer with keyboard in Figure 4(c).

The visual saliency detection results in the second row of Figure 4 have shown that, the salient regions in Figure 4(a) represent the grass with green color behind the sheep and a small part (i.e. legs) of one sheep (i.e. left) while the majority of the two sheep have not been detected as salient objects; the most salient objects detected in Figure 4(b) are the door and ceiling of the room with dark color which are less interesting as they can be considered as backgrounds, another salient region represents the table which masked by the chairs and all the chairs have not been correctly detected. The middle image from Figure 4(c) shows that the blue part of computer screen has been detected as salient region, while the entire computer and the keyboard are the expected salient foreground objects. Therefore, the result images from the second row have shown that the visual saliency detection could not extract the expected foreground objects when the objects are not visually salient but informative salient.

The detection results by using proposed method of three test images are shown in the third row of Figure 4, in which the red windows with different size are detection windows used in different scales. It can be clearly seen from the result images that, despite there are a few mismatched windows that located in the background, such as the wall in Figure 4(b), the majority of all the detection windows can correctly include the expected foreground objects. Since the objects within the detection windows will be considered as the candidates of foreground object, windows which only cover a small part of the object will not affect the further classification process as long as the objects are covered by large windows.

5 CONCLUSIONS AND FUTURE WORK

In this paper, a novel foreground object detection approach for information salient foreground object is proposed based on the sparse reconstruction error. Regarding the generic characteristic of foreground object, the objectness property is characterized as informative salient. In order to detect the interesting foreground objects for artificial awareness, a sparse representation based method is initially presented to obtain the objectness feature of object different from other approaches. To be specific, the objectness of salient foreground object is obtained by calculating the dissimilarity between the object feature and the background dictionary based on the reconstruction error. Experiment results derived from the popular VOC2007 dataset show that, the proposed approach of using reconstruction error can correctly detect the informative salient foreground objects when visual saliency detection fails, which demonstrates the effectiveness of proposed approach.

The experimental results conducted on the real world images have shown that, the performance of proposed approach is quite competitive in detecting salient foreground object. Despite that mismatched detection window could exist in the background, more accurate results are considered to be possible when comprehensive dictionary learning process is applied. In general, the visual information awareness characteristic of salient foreground environmental object for machine can be obtained by applying the proposed approach in this paper, while the visual perception information can be achieved by applying state of the art classification approach to form the visual representation knowledge of environmental object for further higher level processing.

Considering the future work, more dictionary entries or different entries will be taken into account, while different sparse decomposition methods shall be researched. Moreover, the false-positive or false-negative recognition rates will also be investigated as well.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable suggestions and critical comments which have led to a much improved paper. The work accomplished in this paper is supported by the National Natural Science Foundation of China (Grant No. 61174204).

REFERENCES

Fingelkurts, A. A., Fingelkurts, A. A., & Neves, C. F. (2012). "Machine" consciousness and "artificial" thought: An operational architectonics model guided approach. *Brain research*, 1428, 80-92.

Ramík, D. M., Madani, K., & Sabourin, C. (2013). From visual patterns to semantic description: A cognitive approach using artificial curiosity as the foundation. *Pattern Recognition Letters*, 34(14), 1577-1588.

Reggia, J. A. (2013). The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks*, *44*, 112-131.

Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. *Human Factors: The Journal of the*

Human Factors and Ergonomics Society, *32*(1), 61-77.

Alexe, B., Deselaers, T., & Ferrari, V. (2010, June). What is an object?. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 73-80). IEEE.

Alexe, B., Deselaers, T., & Ferrari, V. (2012). Measuring the objectness of image windows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *34*(11), 2189-2202.

Chang, K. Y., Liu, T. L., Chen, H. T., & Lai, S. H. (2011, November). Fusing generic objectness and visual saliency for salient object detection. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 914-921). IEEE.

Spampinato, C., & Palazzo, S. (2012, November). Enhancing object detection performance by integrating motion objectness and perceptual organization. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 3640-3643). IEEE.

Cheng, M. M., Zhang, Z., Lin, W. Y., & Torr, P. (2014, June). BING: Binarized normed gradients for objectness estimation at 300fps. *In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on* (pp. 3286-3293). IEEE.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1?. *Vision research*, *37*(23), 3311-3325.

Mairal, J., Elad, M., & Sapiro, G. (2008). Sparse representation for color image restoration. *Image Processing, IEEE Transactions on*, *17*(1), 53-69.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *31*(2), 210-227.

Ji, Z. J., Wang, W. Q., & Lu, K. (2013). Extract foreground objects based on sparse model of spatiotemporal spectrum. *In Image Processing (ICIP), 2013 IEEE International Conference on* (pp. 3441-3445). IEEE.

Sun, S. W., Wang, Y. C. F., Huang, F., & Liao, H. Y. M. (2013). Moving foreground object detection via robust SIFT trajectories. *Journal of Visual Communication and Image Representation*, 24(3), 232-243.

Biswas, S., & Babu, R. V. (2014, October). Sparse representation based anomaly detection with enhanced local dictionaries. *In Image Processing (ICIP), 2014 IEEE International Conference on* (pp. 5532-5536). IEEE.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.

Liu, C., & Wechsler, H. (2002). Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *Image processing, IEEE Transactions on*, 11(4), 467-476.

Davis, G., Mallat, S., & Avellaneda, M. (1997). Adaptive greedy approximations. *Constructive approximation*, *13*(1), 57-98.

Donoho, D. L., & Tsaig, Y. (2008). Fast solution of l^1-norm minimization problems when the solution may be sparse. *Information Theory, IEEE Transactions on*, *54*(11), 4789-4812.

Rubinstein, R., Zibulevsky, M., & Elad, M. (2010). Double sparsity: Learning sparse dictionaries for sparse signal approximation. *Signal Processing, IEEE Transactions on*, *58*(3), 1553-1564.

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *Signal Processing, IEEE Transactions on*, 54(11), 4311-4322.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

Perazzi, F., Krahenbuhl, P., Pritch, Y., & Hornung, A. (2012, June). Saliency filters: Contrast based filtering for salient region detection. *In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 733-740). IEEE.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2008). The PASCAL Visual Object Classes Challenge 2007 (VOC 2007) Results (2007). In *URL http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.*