Fundamental Research 5 (2025) 354–359

Contents lists available at ScienceDirect

Fundamental Research

journal homepage: http://www.keaipublishing.com/en/journals/fundamental-research/

Article

Towards reliable object representation via sparse directional patches and spatial center cues

Muwei Jian[a,*], Hui Yu[b,*]

[a] School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China
[b] School of Creative Technologies, University of Portsmouth, Portsmouth 200021, UK

ARTICLE INFO

Article history:
Received 21 February 2023
Received in revised form 31 July 2023
Accepted 3 August 2023
Available online 10 August 2023

Keywords:
Multiscale analysis
Image patches
Visual perception
Shearlet transform
Object representation

ABSTRACT

In the process of image understanding, the human visual system (HVS) performs multiscale analysis on various objects. HVS primarily focuses on marginally conspicuous image patches located within or around distinct objects rather than scanning the image pixels point by point. Inspired by the HVS mechanism, in this paper, we aimed to describe and exploit multiscale decomposition-based patch detection models for automatic visual feature representation and object localization in images. Our investigation into mimicking and modeling the HVS to capture conspicuous sparse patches and their spatial distribution clues makes a profound contribution to the automatic comprehension and characterization of images by machines. This study demonstrates that the sparse patch-based visual representation with spatial center cues is intrinsically tolerant to object positioning and understanding beyond object variations in spatial position, multiresolution, and chrominance, which has significant implications for many vision-based automatic object grabbing and perception applications, such as robotics, human–machine interaction, and unmanned aerial vehicles (UAVs).

1. Introduction

Our daily lives involve constant perception of various objects in the primary visual cortex [1]. When observing real-world visual data (e.g., images or videos), the human visual system (HVS) is generally focused on visually striking and attention-grabbing patches or regions within the image for object perception and manipulation. In fact, HVS perceives and comprehends objects not by progressively scanning pixel by pixel, but by capturing and parsing image patches or regions for visual analysis [2]. Therefore, patchwise-based image analysis, which is in line with the HVS perception mechanism, has received tremendous attention for diverse object-representation modeling and has practical applications in various fields such as image indexing [5,12], salient object detection [6,51], image reconstruction [14], and medical image understanding [15,52]. Patchwise-based image analysis has sparked interest among scholars in various disciplines, such as cognitive science, visual psychology, and computer vision.

Cognitive research has revealed that image patches can be implicitly clustered and arranged together in the HVS for visual object modeling and perceiving [47–50]. Meanwhile, as multiscale analysis (e.g., wavelet transform) can represent images at multiple resolutions, it has been employed to detect salient patches with indexing using patch clustering for content-based image retrieval [4,5]. By integrating orientation clues of informative and directional patches in the discriminant color channel, Jian et al. [6] exploited a visual wavelet-based patch-attention-aware mechanism to imitate the HVS for salient object detection. In [7], Liu et al. proposed a pixel-by-pixel contextual attention network for saliency detection, selectively focusing on the contextual location of information at each pixel. Later, Zhao et al. [8] designed a pyramidal feature attention network to augment higher-order contextual features and lower-order spatial structure features to make various feature maps of convolutional neural networks in the saliency-detection task. Recent cognitive research has indicated that the HVS perceives color and direction stimuli in the cerebral visual cortex collectively and concurrently for object perception and image understanding [9]. In view of image patch exemplars, Varma and Zisserman proposed a texton-based algorithm for material classification [10]. Additionally, a hierarchical patch analysis approach was developed for facial component detection and spatial localization by considering regions such as the eyes, nose and mouth in facial images [11]. According to the similarity of image patches, an unsupervised embedding strategy for deep neural networks was introduced to represent image characteristics [12]. Afterward, Hu et al. [13] proposed an efficient patch-based hierarchical scheme for image matching. Through division of the original input into diverse image patches, a multiscale patch log likelihood approach was designed for image restoration [14]. For patchwise-based medical image analysis, Song et al. [15] de-

* Corresponding authors.
E-mail addresses: jianmuwei@ouc.edu.cn (M. Jian), hui.yu@port.ac.uk (H. Yu).

https://doi.org/10.1016/j.fmre.2023.08.001

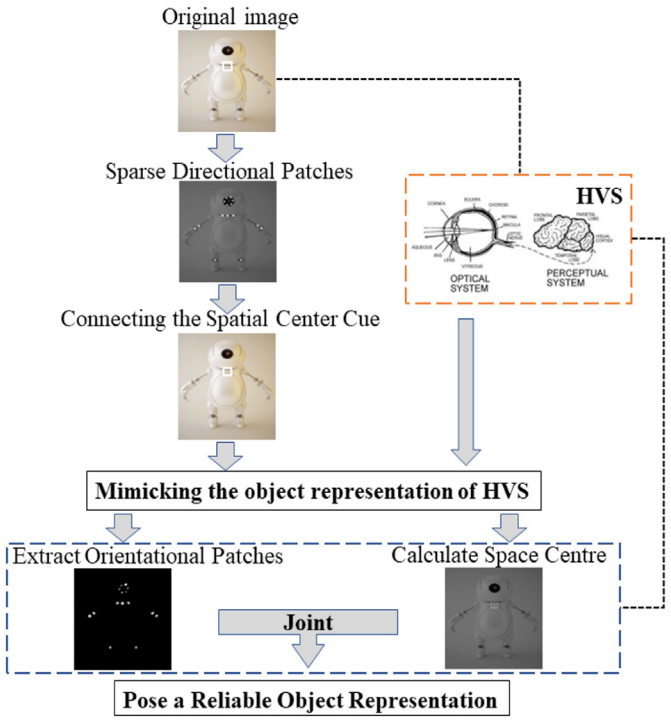

Fig. 1. The main framework of the proposed method.

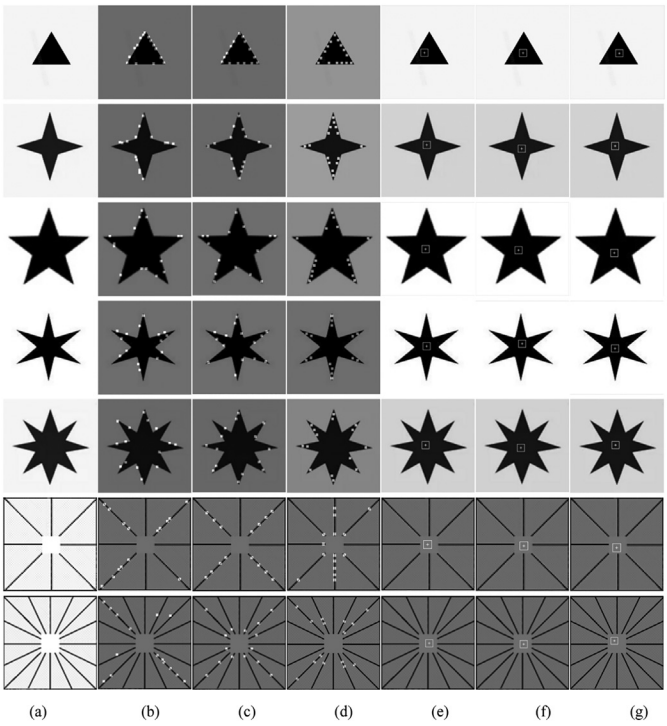

Fig. 2. Comparison results of multiscale decomposition-based patch detection and object spatial localization. (a) Input images, from top to bottom: triangle, four-pointed star, five-pointed star, six-pointed star, eight-pointed star, intersecting grid with 8 directions and mosaic background noises, and intersecting grid with 16 directions and mosaic background noises; (b) salient patches detected by DWT; (c) directional patches detected by DWFT; (d) directional patches detected by DST; (e) estimated spatial center by DWT; (f) estimated spatial center by DWFT; and (g) estimated spatial center by DST.

veloped a patch adaptation approximation framework for lung tissue classification. Later, an improved U-Net model based on image patches was introduced for microscopy imaging systems [16].

From another perspective of visual understanding, visual neuroscience research reveals that the observation of objects by HVS is a multiscale sensing process involving zoom-in and zoom-out patterns [17]. In recent decades, multiscale analysis techniques have become a research focus in the disciplines of signal processing and computer vision [18,19] and have also been successfully applied in image denoising [20], image compression [21], multiresolution image representation [22], fault diagnosis [23], etc. To reveal how the working mode functions for visual apprehension and in light of the universality of multiscale analysis of the inherent physiological mechanism in the HVS, we introduce and evaluate the emblematic frameworks of multiscale decomposition-based image representation, namely, the classic Discrete Wavelet Transform (DWT), Discrete Wavelet Frame Transform (DWFT) and Discrete Shearlet Transform (DST). These frameworks are used for sparse patch localization and spatial center expression in object modeling and visual understanding. The main flowchart of this method is shown in Fig. 1. These typical multiscale decomposition-based frameworks are instrumental in imitating and modeling the zooming-in and zooming-out transactions of the HVS for visual object representation.

2. Materials and methods

2.1. Spatial localization of saliency

Based on psychological investigation, the visual intensity of cortical responses depends on salient patch variances of contrast [24,25,29], and spatial visual stimuli (e.g., the spatial center of the object) are jointly influenced and attract human attention during object localization and perception [30,31]. Therefore, the spatial distribution of these sparsely salient patches undoubtedly provides an inherent and trustworthy cue to correlate with object modeling and capturing. Thus, we take the centroid of the salient patches as a spatial center clue for object representation to forecast the object's spatial center cue as follows:

$$Center(\bar{x},\bar{y}) = \left(\frac{\sum_{i=1}^{K} x_i DP_i(x_i,y_i)}{\sum_{i=1}^{K} DP_i(x_i,y_i)}, \frac{\sum_{i=1}^{K} y_i DP_i(x_i,y_i)}{\sum_{i=1}^{K} DP_i(x_i,y_i)} \right) \quad (1)$$

where DP_i (i = 1, 2, …, K) represents the K extracted patches (e.g., K = 15).

The sample results of estimating the spatial center position of diverse objects are illustrated in Fig. 2e, which demonstrates the significance of the geometric center of an object in capturing visual attraction for reliability prognosticates.

The HVS is sensitive to the contrast of luminance of images [24,25]. The primate perception system unconsciously tends to target and perceive salient objects in the visual world according to the exceedingly multiscale sparse visual clues of pixel change and contrast in brightness [26–28]. As a characteristic multiscale analysis scheme, the coefficients of wavelet decomposition can reflect image intensity variation and position changes [18,19]. To represent the multiscale pixel variations in brightness, we propose an efficient wavelet-based sparse salient point extraction procedure for image representation, which is briefly described as follows. For any image pixel $f(n)$, with regard to the compactly supported DWT (e.g., db4 with the wavelet regularity p equal to 4), an arbitrary wavelet coefficient $W_{2^j} f(n)$ can be calculated with $2^{-j}p$ signal points on the scale 2^j. The wavelet coefficients can be further analyzed on a finer scale of 2^{j+1}. Regarding the identical image pixel decomposed on the scale 2^j, the children coefficients $C(W_{2^j} f(n))$ of the upper-scale coefficient $W_{2^j} f(n)$ can be linked as follows [27,28]:

$$C(W_{2^j} f(n)) = \left\{ W_{2^{j+1}} f(k), 2n \le k \le 2n + 2p - 1 \right\} \quad (2)$$

where $0 \le n \le 2^j N$, and N is the length of the input. Specifically, these children's coefficients $C(W_{2^j} f(n))$ also identify the alteration of the same input image pixel, and the most salient point is the one with the largest absolute value of the wavelet coefficient. Consequently, the salient points in the image can be selected by calculating the magnitude of the DWT coefficients across various scales [27,28]:

$$saliency = \sum_{k=1}^{-j} |C^{(k)}(W_{2^j} f(n))|, 0 \le n \le 2^j N, -\log_2 N \le j \le -1 \quad (3)$$

Nevertheless, the maximum values of the wavelet coefficients among individual multiscales contain different variation ranges of values for different scopes of variety. In general, the values in the largest wavelet coefficient set of the upper scale embody larger magnitudes than those in the current scale 2^{j+1}. To calculate the saliency values more accurately, the largest magnitude across all the various decomposed multiscale can be comprehensively modified with normalization [5,6]:

$$saliency = \sum_{k=1}^{-j} |w_k C^{(k)}(W_{2^j} f(n))|, 0 \le n \le 2^j N, -\log_2 N \le j \le -1 \quad (4)$$

where $W_{2^j} f(n)$ denotes the adapted weighting value at the corresponding multiscale k.

Considering the limited capacity of the retina in the human visual system, a wavelet-based salient patch detection approach was designed to enhance image feature representation in machine learning [5,6]. Postulate an image patch (e.g., quadrangular or hexagon) with the spatial center coordinate (x, y) of an object. Taking the wavelet coefficients within the patch as a unit, the computational mechanism is implemented on the three different high-frequency sub-bands (i.e., horizontal, vertical, and diagonal subimages) in terms of Eq. 4. Then, with the integration of the saliency values in the multiscale subimages, the holistic saliency values of distinct patches can be computed and sorted by their magnitude coefficients of DWT, as shown in Fig. 2b.

2.2. Orientational visual cues

Cognitive neuroscience research has revealed that orientation selectivity is a ubiquitous processing mode in the primate cortical area V1, so directional visual stimuli are innately conducive to image perception and object understanding [32,33]. Visual cortical mechanisms indicate that orientational visual cues can shape the representation of the image scene and the modeling and perceiving of diverse conspicuous objects, which is consistent with cognitive discoveries [39–42]. However, this observation is the research of neuroscience, which has not yet been technically realized and proven. To our knowledge, this work is the first of its kind to design an orientational visual cue procedure to imitate image representation.

Spatial connectivity generally responds to orientation selectivity synchronously in the HVS [34], which expresses the spatial representation of directional visual cue links and constitutes one of the fundamental and irreplaceable elements for sensing objects and visual perception [35]. In the primary visual cortex, the low-level visual property of the object center cue forms a stable and reliable perceptual stimulus and sensory feedback for HVS to detect and comprehend individual objects in various complex environments [36].

2.3. Three-directional saliency

To imitate the human visual perception process for manipulating directional visual characteristics and spatial representation of objects, a discrete wavelet frame transform (DWFT)-based method is exploited to forecast the orientation selectivity associated with the object spatial center simultaneously. Relative to DFT, the discrete wavelet frame transform (DWFT) is also a multiscale analysis approach [37,38]. The gigantic strength of DWFT is that it is translation-invariant for feature representation while avoiding downsampling through an overcomplete wavelet transform, which is particularly sufficient to characterize and reflect multiscale directional visual cues. In the following subsection, we concisely describe the specific DWFT-based directional patch detection procedure.

Each scale of the DWFT engenders three directional subbands: the LH^{DWFT} sub-band reflects the horizontal orientation selectivity of high frequency of the original image; HL^{DWFT} manifests vertical orientation variations of high frequency; and HH^{DWFT} responds to diagonal orientation variances of high frequency of the initial input. Then, concerning each distinct column in the $LH^{DWFT}(x,\ y)$ subimage, the vertical direction map is available in the following mathematical expression [36]:

$$D_{ir_V}(x, y) = \frac{1}{2W} \left| LH^{DWFT}(x, y) \right| \left\{ \sum_{r=-W}^{W} \left| LH^{DWFT}(x, y + r) \right| \right\} \quad (5)$$

where W represents the window size of the successive coefficients (e.g., set as the same value as the wavelet regularity empirically).

Using the same computing strategy as Eq. 5, the horizontal direction and the diagonal direction maps can be acquired from their respective orientation selectivity. Finally, the multiscale and triple directional maps are directly merged into a unified and complete integrated orientation map as follows:

$$D_{ir}(x, y) = \frac{1}{3} \left[D_{ir_V}(x, y) + D_{ir_H}(x, y) + D_{ir_D}(x, y) \right] \quad (6)$$

where $D_{ir_V}(x, y)$, $D_{ir_H}(x, y)$ and $D_{ir_D}(x, y)$ denote the triple distinct direction maps computed by Eq. 5 similarly.

At this point, the directional sparse patches based on DFWT can be easily detected from the monolithic integrated orientation map $D_{ir}(x, y)$, as illustrated in Fig. 2c.

To estimate the spatial position clue, the sparsely extracted directional patches are implemented to calculate the centroid of the object by employing a similar computing procedure according to Eq. 1. Some typical examples of the calculated spatial position center of objects with different geometric configurations are illustrated in Fig. 2f.

2.4. Multiple-directional saliency

Visual representation based on sparse patches with spatial center clues essentially locates and understands the changes beyond the spatial position, multiresolution and chromaticity of objects (Fig. 2). Within the visual receptive field, the HVS has the extraordinary ability to detect and encode extremely sparse patches with orientation discontinuities to represent salient objects [39,40]. However, visual stimulation in more directions is more prone to benefiting the formation of a complete and versatile representation of the object shape and structure; therefore, with the result, it is also potentially instrumental in perceiving and recognizing objects [41,42]. To simplify the modeling process, we explored three directional and multiple directional multiscale decompositions to simulate the object description.

More directional visual stimuli are often employed to describe and characterize objects, stemming from the fact that they facilitate the formation of a complete representation of the object's shape and structure. Based on this visual cortical mechanism, we excogitate a discrete shearlet transform (DST)-based directional sparse patch detection method, which furnishes more orientation properties (e.g., multiscale DST with 12 or 16 directional features) and optimally sparse representation for the effective characterization of objects.

Unlike the traditional DWT and DWFT, which engender triple multiscale directional decompositions (namely, horizontal, vertical, and diagonal), the DST employs shearing to enhance orientation selectivity, providing a remarkable capacity to capture edges, boundaries, and other anisotropic features from multidirectional perspectives [43–45]. Taking advantage of DST, an input image can be decomposed into multiscale subimages with multiple directional high-frequency information. Similar to the computing framework of directional patch detection based on

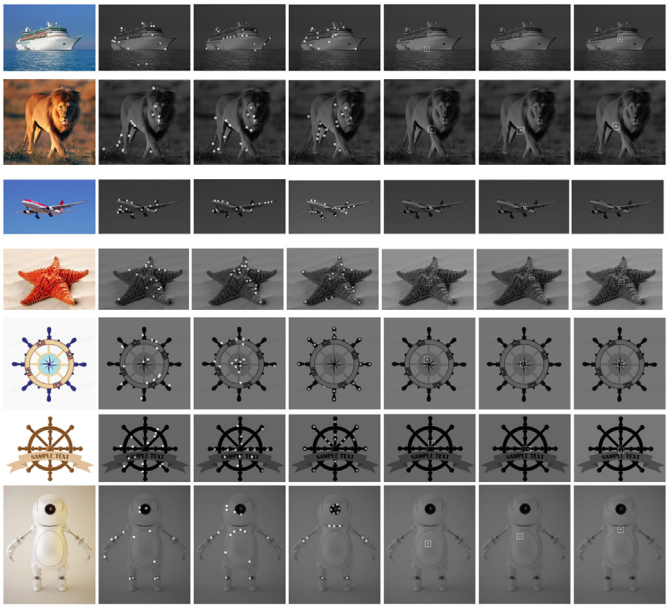

Fig. 3. Experimental results of real-world images produced by multiscale decomposition-based patches detection and object spatial localization. (a) Input image; (b) salient patches detected by DWT; (c) directional patches detected by DWFT; (d) directional patches detected by DST; (e) the estimated spatial center by DWT; (f) the estimated spatial center by DWFT; and (g) the estimated spatial center by DST.

DFWT, the corresponding direction map can be obtained computationally based on Eq. 5 regarding the individual high-frequency subimage of DST. Thereafter, the composite orientation map derived from the directional multiscale high-frequency subimages is computed straightforwardly in accordance with Eq. 6. Then, the directional patches based on DST can be extracted conveniently in terms of the composite orientation map. Some experimental results of directional patches and the spatial position center produced by DST are respectively shown in Fig. 2d and 2g.

3. Results and discussion

Images in the real world with symmetry of geometric structures, richness of directional characteristics and mixture of background textures are collected for a comparative and comprehensive study. For the sake of implementing fair comparison, the number of directional patches K is set as 15, and a three-level multiscale decomposition using the orthogonal db4 wavelet with a compact support of the wavelet regularity $p = 4$ is adopted. For DST, 16 directivities were selected and applied for multiple directional representation of the input image.

3.1. Qualitative analysis

A number of typical visual comparisons with the individual multiscale decomposition-based approaches are displayed in Fig. 3. As shown in Fig. 3, by conducting simulated experiments, those distinct salient and directional image patches on the object constitute an extremely sparse orientation discontinuity visual stimulus to reflect the modeling and perceiving of diverse conspicuous objects, which is consistent with cognitive discoveries [39–42].

Moreover, in contrast to the DWT- and DWFT-based patch extraction models, the devised DST-based method tends to focus on scattered patches fixed at more directional positions as visual spatial cues, which indicates higher directional selectivity, as illustrated in Fig. 4. With multidirectional decomposition, the DST-based patch detection framework

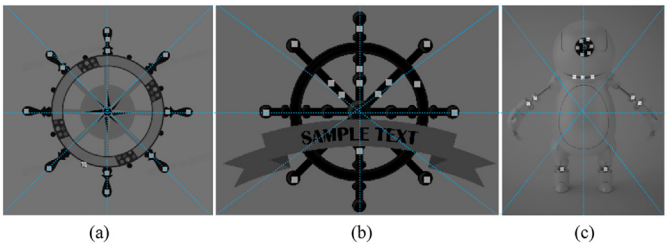

Fig. 4. Illustration of symmetry in multiple aliquot orientations of the extracted image patch-based DST. (a) and (b) compass images with directional diversity; (c) robots with geometric symmetry.

exhibits perfect spatial directional positioning capability as well as symmetry in geometric shape for object representation, such as compass images with directional diversity (Fig. 4a, b) and robots with geometric symmetry (Fig. 4c). Through visual comparison of the results obtained by the DWT-, DWFT-, and DST-based patch extraction schemes, DST outperforms the DWT and DWFT in terms of the symmetrical positioning capacity of the spatial direction for object characterization.

3.2. Quantitative analysis

For quantitative measurements, the accuracy of the mean square error (MSE) in Euclidean space is used as an objective assessment indicator to statistically evaluate the spatial center clues estimated by different methods based on multiscale analysis. We adopt MSE in Euclidean space to calculate the error between the predicted center point and the ground truth, and the center of the current sample is evaluated as correctly predicted when the error is less than a threshold λ (e.g., $\lambda = 5$). We test the individual method on the commonly used MSRA image database [46] with a variety of 300 images randomly selected for comparative evaluation.

Table 1 provides the average accuracy of the estimated spatial centers of the individual multiscale decomposition-based approach according to the MSRA database. No image-preprocessing operations, such as background noise suppression, texture elimination or foreground highlighting, are performed in advance to enhance the performance index. The main purpose of this study is to verify the cognitive mechanism rather than break the performance indicator or achieve a new record. From Table 1, it can be observed that those multiscale decomposition-based frameworks can predict the spatial center cues of diverse objects. The spatial position clue of objects, as a cognitive mechanism promulgated, is reliable for representing and understanding objects. Meanwhile, since the DST has a better capacity of directional characteristics, the developed DST-based algorithm is superior to the DWT- and DWFT-based models in line with the quantitative comparison.

In addition, it is worth noting that visual noise interference in the HVS is a ubiquitous phenomenon. For example, in the first and second images in Fig. 3b, two patches were located in the background of ocean surface texture, while a single patch landed on the image background of grass. Although background effects can be easily eliminated with simple image preprocessing, such as background suppression or blurring operations, the predicted spatial centers of objects still yield satisfactory results, as displayed in Fig. 3d. This universal phenomenon can be explained by the fact that the HVS not only concentrated on the conspicuous foreground objects in the wide receptive field of the visual cortex

Table 1
The accuracy of the predicted spatial centers of typical multiscale decomposition-based methods in terms of the MSRA database.

Methods	DWT[6]	DWFT[36]	The devised DST
Accuracy	87.08	90.31	91.59

but also processed these intensely salient visual stimuli in the image background simultaneously. The evidence confirms that the HVS also possesses the reliability to resist disturbance and is immune to faults, even in the presence of a small amount of scattered noise (e.g., the first two images in Fig. 3b), and the prediction of the object center remains unaffected.

Meanwhile, as indicated in the pioneering research [3,5,41], the sparse patches in the image can easily be clustered, programmed and classified into various categories subconsciously in the HVS for depicting different objects and separating the foreground object from the image background. This foundation will effectively facilitate the visual representation of multitarget objects in an image. Furthermore, the combination of direction selectivity and spatial connectivity provides the HVS with a remarkable advantage in perceiving and modeling objects, as revealed in previous investigations [6,30,35]. These discoveries verify that the HVS has an inherent perception capacity to sense a wide variety of objects in a visual stimulus processing mode of sparse direction selectivity associated with spatial connectivity jointly, which also provides insight into guiding the design of visual perception and explainable deep neural networks proceeding from the perspective of cognitive and instinctive mechanisms.

4. Conclusion and future work

Object perception and representation are fundamental tasks confronted by the HVS when perceiving the visual world. A biological visual mechanism of the HVS is that it can recognize and identify objects in the scene at multiple scales, both in holistic characteristics and meticulous details. From a cognitive perspective, this research simulates how the HVS processes sparsely directional visual cues of distinct objects based on multiresolution analysis. Extensive evaluation has shown that the combination of directional patches considering orientation selectivity connecting the spatial center cue in synchronization indeed carries a sparse approximation and a reliable spatial representation of various objects.

The comparative experiment indicates that multiscale sparse image patches together with spatial location clues play a basic and critical role during visual understanding. This finding demonstrates that the sparse visual directional patches and spatial center characteristics are consistent with the integrated disposal pattern of visual stimuli, which corroborates that direction selectivity joining the spatial cue constitutes an essential and connected visual stimulation in the HVS for perceiving and encoding objects. This mechanism is also correlated with resistance in the alteration of spatial localization and immune to the unanticipated uncertainty of diverse image chromaticity and multiple resolutions.

In the future, we will explore how the computable multiscale decomposition-based image patch representation can assist visual understanding by designing reliable and explainable deep learning architectures that can drive the promotion of object representation-related computer vision tasks, including object detection, semantic segmentation, saliency detection, and object recognition.

Code availability

Source data will be provided with this paper.

Declaration of competing interest

The authors declare that they have no conflicts of interest in this work.

Data availability

The raw data required to reproduce these findings can be available from the first author upon reasonable request.

Acknowledgments

This work was supported by National Natural Science Foundation of China (61976123, 61601427); Taishan Young Scholars Program of Shandong Province; Key Development Program for Basic Research of Shandong Province (ZR2020ZD44) and Royal Society - K. C. Wong International Fellowship (NIF\R1\180909).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fmre.2023.08.001.

References

[1] H. Yin, A. Varava, D. Kragic, Modeling, learning, perception, and control methods for deformable object manipulation, Sci. Robot. 6 (54) (2021) eabd8803.
[2] Y. Fang, W. Lin, B.S. Lee, et al., Bottom-up saliency detection model based on human visual sensitivity and amplitude spectrum, IEEE Trans. Multimedia 14 (1) (2011) 187–198.
[3] Y. Sugita, Grouping of image fragments in primary visual cortex, Nature 401 (6750) (1999) 269–272.
[4] M. Jian, J. Dong, R. Jiang, et al., Wavelet-based salient regions and their spatial distribution for image retrieval, in: In: Proc. IEEE Int. Conf. Multimedia Expo., 2007, pp. 2194–2197.
[5] M. Jian, J. Dong, J. Ma, Image retrieval using wavelet-based salient regions, Imaging Sci. J. 59 (4) (2011) 219–231.
[6] M. Jian, K.M. Lam, J. Dong, et al., Visual-patch-attention-aware saliency detection, IEEE Trans. Cybern. 45 (8) (2015) 1575–1586.
[7] N. Liu, J. Han, M.H. Yang, Picanet: Learning pixel-wise contextual attention for saliency detection, in: In: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 3089–3098.
[8] T. Zhao, X. Wu, Pyramid feature attention network for saliency detection, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 3085–3094.
[9] A.K. Garg, P. Li, M.S. Rashid, et al., Color and orientation are jointly coded and spatially organized in primate primary visual cortex, Science 364 (6447) (2019) 1275–1279.
[10] M. Varma, A. Zisserman, A statistical approach to material classification using image patch exemplars, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2009) 2032–2047.
[11] M. Jian, K.M. Lam, J. Dong, Facial-feature detection and localization based on a hierarchical scheme, Inf. Sci. 262 (2014) 1–14.
[12] D. Danon, H. Averbuch-Elor, O. Fried, et al., Unsupervised natural image patch learning, Comput. Vis. Media 5 (2019) 229–237.
[13] S.M. Hu, F.L. Zhang, M. Wang, et al., PatchNet: A patch-based image representation for interactive library-driven image editing, ACM Trans. Graph. 32 (6) (2013) 1–12.
[14] V. Papyan, M. Elad, Multi-scale patch-based image restoration, IEEE Trans. Image Process. 25 (1) (2016) 249–261.
[15] Y. Song, W. Cai, Y. Zhou, et al., Feature-based image patch approximation for lung tissue classification, IEEE Trans. Med. Imaging 32 (4) (2013) 797–808.
[16] A.C. Li, S. Vyas, Y.H. Lin, et al., Patch-Based U-Net model for isotropic quantitative differential phase contrast imaging, IEEE Trans. Med. Imaging 40 (11) (2021) 3229–3237.
[17] S.C. Nercessian, K.A. Panetta, S.S. Again, Non-linear direct multi-scale image enhancement based on the luminance and contrast masking characteristics of the human visual system, IEEE Trans. Image Process. 22 (9) (2013) 3549–3561.
[18] S.G. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
[19] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inf. Theory 36 (5) (1990) 961–1005.
[20] S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, IEEE Trans. Image Process. 9 (9) (2000) 1532–1546.
[21] B.E. Usevitch, A tutorial on modern lossy wavelet image compression: Foundations of JPEG 2000, IEEE Signal. Proc. Mag. 18 (5) (2001) 22–35.
[22] M.N. Do, M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
[23] S. Lee, C.D. Yoo, T. Kalker, Reversible image watermarking based on integer–to-integer wavelet transform, IEEE Trans. Inf. Forensics Secur. 2 (3) (2007) 321–330.
[24] A.F. Rossi, C.D. Rittenhouse, M.A. Paradiso, The representation of brightness in primary visual cortex, Science 273 (5278) (1996) 1104–1107.
[25] M. Carandini, D. Ferster, A tonic hyperpolarization underlying contrast adaptation in cat visual cortex, Science 276 (5314) (1997) 949–952.
[26] J.J. Todd, R. Marois, Capacity limit of visual short-term memory in human posterior parietal cortex, Nature 428 (6984) (2004) 751–754.
[27] N. Sebe, Q. Tian, E. Loupias, et al., Color indexing using wavelet-based salient points, in: Proc. IEEE Workshop Content-Based Access Image Video Libr., 2000, pp. 15–19.
[28] Q. Tian, N. Sebe, M.S. Lew, et al., Image retrieval using wavelet-based salient points, J. Electron. Imaging 10 (4) (2001) 835–849.

[29] R. Margolin, A. Tal, L. Zelnik-Manor, What makes a patch distinct? in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2013, pp. 1139–1146.
[30] G. Nyman, P. Laurinen, Reconstruction of spatial information in the human visual system, Nature 297 (5864) (1982) 324–325.
[31] Y. Xiao, Y. Wang, D.J. Felleman, A spatially organized representation of colour in macaque cortical area V2, Nature 421 (6922) (2003) 535–539.
[32] C.D. Gilbert, T.N. Wiesel, The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat, Vision Res. 30 (11) (1990) 1689–1701.
[33] N.J. Priebe, Mechanisms of orientation selectivity in the primary visual cortex, Annu. Rev. Vis. Sci. 2 (2016) 85–107.
[34] E.N. Johnson, M.J. Hawken, R. Shapley, The orientation selectivity of color-responsive neurons in macaque V1, J. Neurosci. 28 (32) (2008) 8096–8106.
[35] L.F. Rossi, K.D. Harris, M. Carandini, Spatial connectivity matches direction selectivity in visual cortex, Nature 588 (7839) (2020) 648–652.
[36] M. Jian, J. Wang, H. Yu, et al., Visual saliency detection by integrating spatial position prior of object with background cues, Expert Syst. Appl. 168 (2021) 114219.
[37] M. Unser, Texture classification and segmentation using wavelet frames, IEEE Trans. Image. Process. 4 (11) (1995) 1549–1560.
[38] M. Unser, N. Chenouard, D. Van De Ville, Steerable pyramids and tight wavelet frames in L_2(\BBR^d), IEEE Trans. Image. Process. 20 (10) (2011) 2705–2721.
[39] J.I. Nelson, B.J. Frost, Orientation-selective inhibition from beyond the classic visual receptive field, Brain Res. 139 (2) (1978) 359–365.
[40] A.M. Slllito, K.L. Grieve, H.E. Jones, et al., Visual cortical mechanisms detecting focal orientation discontinuities, Nature 378 (6556) (1995) 492–496.
[41] J.P. Gottlieb, M. Kusunoki, M.E. Goldberg, The representation of visual salience in monkey parietal cortex, Nature 391 (6666) (1998) 481–484.
[42] Z. Kourtzi, N. Kanwisher, Representation of perceived object shape by the human lateral occipital complex, Science 293 (5534) (2001) 1506–1509.
[43] G. Kutyniok, W.Q. Lim, Compactly supported shearlets are optimally sparse, J. Approx. Theory 163 (11) (2011) 1564–1589.
[44] G. Kutyniok, W.Q. Lim, R. Reisenhofer, ShearLab 3D: Faithful digital shearlet transforms based on compactly supported shearlets, ACM Trans. Math. Softw. 42 (1) (2016) 1–42.
[45] T.A. Bubba, D. Labate, G. Zanghirati, et al., Shearlet-based regularized reconstruction in region-of-interest computed tomography, Math. Model. Nat. Phenom. 13 (4) (2018) 34.
[46] T. Liu, Z. Yuan, J. Sun, et al., Learning to detect a salient object, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 33, 2010, pp. 353–367.
[47] K. Nakayama, S. Shimojo, G.H. Silve rman, Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects, Perception 18 (1) (1989) 55–68.
[48] S. Shimojo, K. Nakayama, Amodal representation of occluded surfaces: Role of invisible stimuli in apparent motion correspondence, Perception 19 (3) (1990) 285–299.
[49] K. Nakayama, S. Shimojo, Experiencing and perceiving visual surfaces, Science 257 (5075) (1992) 1357–1363.
[50] K. Nakayama, Binocular visual surface perception, Proc. Natl. Acad. Sci. U.S.A. 93 (2) (1996) 634–639.
[51] Y. Cong, C. Gu, T. Zhang, et al., Underwater robot sensing technology: A survey, Fund. Res. 1 (3) (2021) 337–345.
[52] J. Wang, Y. Wang, X. Tao, et al., PCA-U-Net based breast cancer nest segmentation from microarray hyperspectral images, Fund. Res. 1 (5) (2021) 631–640.

Author profile

Muwei Jian received the PhD degree from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, in 2014. Currently, Dr. Jian is a distinguished professor and Ph.D supervisor at the School of Computer Science and Technology, Shandong University of Finance and Economics. Dr. Jian has been awarded the Royal Society – K. C. Wong International Fellow under Newton International Fellowship (NIF). His current research interests include human face recognition, image and video processing, machine learning and computer vision. He serves as an Associate Editor of *IET Computers & Digital Techniques* and the *Journal of Image and Graphics*.

Hui Yu (Senior Member, IEEE) is professor with the University of Portsmouth and an Industrial Fellow of the Royal Academy of Engineering, UK. He worked at the University of Glasgow and Queen's University Belfast before joining the University of Portsmouth in 2012. His research interests include vision, creative computing and AI with applications to 4D facial and affective analysis, human-machine interaction, VR/AR, intelligent vehicles and image clustering. He serves as an Associate Editor of the *IEEE Transactions on Human-Machine Systems* and the *IEEE Transactions on Computational Social Systems journal.*