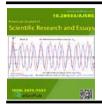
**Review Article** AJSRE (2019), 4:24



# **American Journal of Scientific Research and Essays** (ISSN:2475-7527)



# Finding Waldo: Two Routes for Processing Visual Search **Complex Scenic Images**

## **Andrew Hwang**

**Newark Academy** 

#### **ABSTRACT**

This paper reviews the cognitive mechanisms underlying visual search in complex visual images. Visual search is a type of task that we are constantly engaged in our everyday lives. Here, the current paper introduces the current literature that suggests that there are two different processing routes that support visual search in human observers: Bottom-up and Top-down processing. Bottom-up processing allows the visual system to process the sensory information from the visual image as it is, whereas top-down processing guides the visual system to efficiently allocate attention to a part of the image to facilitate the search process. The two processing routes via bottom-up and topdown processing work in concert during complex visual search, guiding our search behavior to be more efficient and adaptive. In closing remark, this review also briefly discusses how the current knowledge about the roles of bottom-up and top-down processing in visual search can contribute to development of computer vision and artificial intelligent.

Keywords: Two Routes, Processing Visual Search, Complex Website: https://escipub.com/ Scenic Images

# \*Correspondence to Author:

Andrew Hwang **Newark Academy** 

### How to cite this article:

Andrew Hwang, Finding Waldo: Two Routes for Processing Visual Search in Complex Scenic Images. American Journal of Scientific Research and Essays, 2019 4:24.



Imagine you are in front of the arrivals door at the John F. Kennedy International Airport, peacefully waiting for an old school friend who is visiting you for summer vacation. As the plane lands and the arrivals door opens, surroundings suddenly become crowded and filled with people walking in and out, looking for someone, greeting one another, and rushing into the baggage claim. As people crowd the greeting area in front of the arrivals gate, your eyes and brain become overloaded searching for him, trying to recall how he looked at the moment you last saw him and guessing how he would have been changed since then (probably almost 5 years now). Such activity, scanning environment to search for something important

is defined "visual search" in the fields of cognitive psychology and neuroscience. Visual search is a type of fundamental psychological and perceptual process that occurs in many different contexts of everyday life (e.g., when your mom picks out a carton of milk on a supermarket shelf, when a TSA officer tries to detect any weapons or dangerous objects from an x-ray image of a baggage, or when you are simply playing "Where's Waldo?"; see Figure 1 for visual illustrations). In fact, anything that is currently important to your goal (e.g., picking up your old friend at the airport) can be your "search target" (e.g., your friend in this airport example) among all "distractors" (e.g., all the other objects and people in the environment).

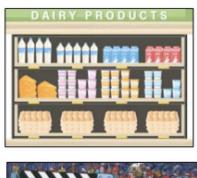






Figure 1. Some of the real-world examples of visual search, including "looking for a carton of milk in the grocery market", "detecting a gun from the security monitor", and "finding Waldo from the book".

Our ability to detect or locate a target amongst many distractors in various visual search tasks has been extensively studied over the past 40 years. Figure 2 illustrates a typical trial sequence of a visual search task that has been used in many research studies: a search target is

defined (e.g., a red letter "T") then a human observer scans a visual search array consisted of multiple items. The search array sometimes contains the target (i.e., present trials), and sometimes it does not (i.e., absent trials). Typically, the human observer is tasked with reporting whether or not the target item is located inside the visual array. While the human observer completes the task, some researchers

measures search accuracy [1]; some researchers measures response time (RT; [2]); and some others monitors and explores the patterns of eye movement and gazes of observers during search tasks [3]. These measures allow them to understand the nature of perceptual and cognitive processes underlying visual search.

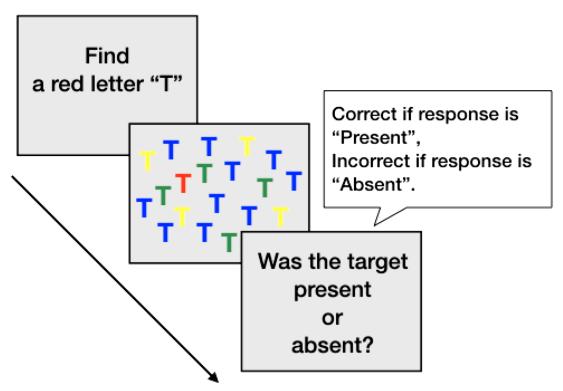


Figure 2. An example of an experimental sequence to test the ability to search for a visual target from a search array. In the experimental settings, a human participant is asked to find a target (defined by experimenters, for example a red letter "T" in this example). The search array that sometimes includes a target and sometimes not is presented; then the participant is to indicate whether the target was present or absent in the search array. Experimenters tend to measure how accurately and how fast the participant reported.

By using these behavioral measurements, much previous literature has shown that we are genuine experts at visual search from extremely complex natural scenes, still outperforming state-of-the-art computer vision systems [4]. How is this high performance achieved? In other words, how does our brain solve the Where's Waldo problem so easily and efficiently? There seem to be (at least) two different processors

that can guide and allocate our attention. These two processors not only support different attentional mechanisms, but work together to facilitate our visual search performance, possibly in more efficient and flexible manners than any types of computer vision or AI (Artificial Intelligent) can do. Because properly allocating your attention towards the location occupied by a target and suppressing attentional capture by

other distractors is a critical success factor for processors would determine the outcome of the your visual search [5], the functions of these two search.

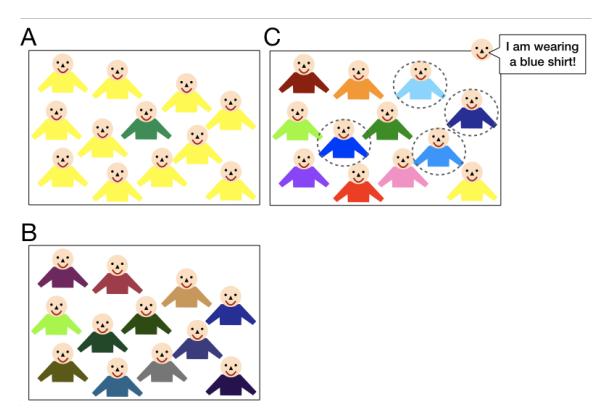


Figure 3. Visual illustrations of how bottom-up and topdown processing would work in the airport example. In panel A, the guy in a green shirt will attract your attention first, because it has a unique feature. In panel B, the guy in a neo green shirt will pop-out, because of its saliency. In panel C, your expectations about the color of the shirt that your friend is wearing will guide your attention so that you can only focus and search for blue shirts, filtering out other colors.

One of these search processors relies on bottom-up processing. Bottom-up processing involves a real-time, data-driven sensory analysis, starting with sensory information brought to the retina and proceeding to the visual cortex and other (higher) brain areas until representation of the object is formed in our minds [6]. Bottom-up processing plays roles in capturing and allocating our attention towards a target item that is unique, distinct, or salient. Let's consider the airport example again. If your old friend was the only one who was wearing a green shirt among others in yellow shirts (Figure 3A), you would be able to spot him easily and quickly (almost automatically and immediately): Your attention will be directed to the unique color by default, then you will recognize that he was

the one you were looking for. Alternatively, if he was wearing a shirt in a very bright, neon green color, saliency would lead bottom-up processing to attract your attention towards him, even if he was not the only unique one (Figure 3B). A list of such factors for guiding bottom-up processing in various visual search tasks that have been examined in previous research studies is found in Figure 4, along with relevant citations and visual illustrations. Relying on these principles underlying the bottom-up processing, recent mathematical and probabilistic models for AI and computer vision have successfully stimulated and even excelled the human performance in detecting a pop-out target in various visual search tasks [7].

Bottom-up factors	Example	References
Perceptual Saliency	+ + + + + + + + + + + + + + + + + + +	<ul> <li>Lamy, Leber, &amp; Egeth, Journal of Vision, 2002</li> <li>Thompson, Bichot, &amp; Sato, Journal of Neurophysiology, 2005</li> </ul>
Uniqueness (oddball)		<ul> <li>Chen &amp; Zelinsky, Vision Research, 2006</li> <li>Bravo &amp; Nakayama, Perception &amp; Psychophysics, 1992</li> </ul>
Importance (e.g., threat)	2	<ul> <li>Connor, Egeth, &amp; Yantis, Current Biology, 2004</li> <li>Wolfe, Current Biology, 2010</li> <li>Fabio &amp; Caprì, Heliyon, 2019</li> </ul>
Emotion		<ul> <li>Park &amp; Thayer, Frontiers in Psychology, 2014</li> <li>Calvo &amp; Marrero, Cognition and Emotion, 2009</li> </ul>

Figure 4. Visual illustrations of some bottom-up factors and relevant papers to read more about each of the examples.

The other search processor relies on top-down processing. Top-down processing utilizes what the brain already knows and expects about objects, scenes, concepts, contexts, and the entire environment in order to create a more efficient understanding based on this information. For example, your visual search can be guided by what you know about the world described in the search array. When you are asked to locate where a person is in Figure 5, it is more likely that you first look around the door and the bench on the ground and scan these spots more often and more carefully (with more frequent eye gazes to these spots, as well), rather than somewhere in a tree or the sky. Here, you are applying your general knowledge that people tend to sit on a bench or walk through doors on

the ground, but rarely fly! Conversely, when you are asked to locate a bird in the same image, then your attention would be most likely attracted to the sky, a tree or the bush, rather than the bench or the door, spending more time in scanning these spots (again, with more frequent eye gazes to these spots, as well) based on your reasonable expectation and prior knowledge about the birds. Sometimes, you can learn new information about the characteristics of a target item and utilize it to find the target faster. Coming back to the airport example above, if your friend told you that he is wearing a blue shirt, then you would be able to focus on people who are wearing blue shirts and skip scanning others who are not, in order to facilitate the search and terminate the process faster (Figure 5).

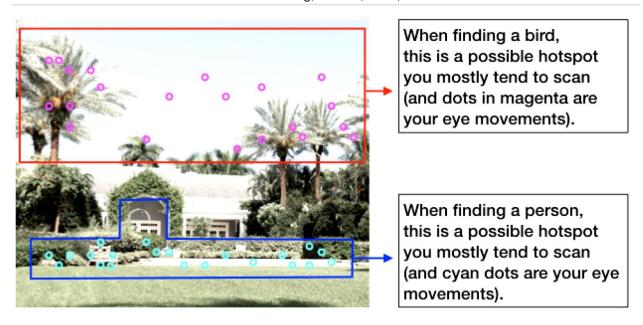


Figure 5. An example of visual search in a naturalistic scene. Our own knowledge and expectations about the world guide our eye movements (indicated by colored circles on this image) by reflecting our optimal strategy to effectively search for a target.

Top-down factors	References	
Expectation	Sobel, Gerrie, Poole, & Kane, Psychonomic Bulletin & Review	
	Summerfield & Egner, Trends in Cognitive Sciences, 2009	
Knowledge	Wolfe, Butcher, Lee, & Hyle, Journal of Experimental Psychology: Human Perception & Performance, 2003	
	Chun & Jiang, Psychological Science, 1999	
Current goals	<ul><li>Lookadoo, Yang, &amp; Merrill,</li><li>Awh, Belopolsky, &amp; Theeuwes</li></ul>	
Emotion	Mohanty & Sussman	
Past experience	<ul> <li>Hwang, Higgins, &amp; Pomplun</li> <li>Lee, Yang, Romero, &amp; Mumford</li> </ul>	
Motivation	<ul><li>Van Zoest&amp; Donk</li><li>Boot, Becic, &amp; Kramer</li></ul>	

Figure 6. Some examples of top-down factors and relevant papers to read more about each of the examples.

Other than the prior knowledge or expectation, there are various factors that can effectively guide the top-down processing in visual search, as summarized in Figure 6. Compared to any other Al systems, the human brain seems to be much better at utilizing and combining these factors for top-down processing even before most of the scene objects and details are recognized and consciously processed [4]. Thus, the top-down processing is the domain in which machine such as AI and computer vision systems cannot simulate human performance yet. How would the computer be able to utilize its past experience, maximize its motivation, rely on its long-term memory of prior knowledge and expectation about the world? Addressing this question will help scientists and engineers to fill in the gap between the human brain and machine in performing visual search in complex, natural scenes.

During the past 40 years or so, much research on visual search has been conducted and has given us a good grasp of the mechanisms of bottom-up processing and top-down processing that guide human observers to locate a target item in complex scenes (whether it is your old friend in the airport scene, a bird in a forest, or Waldo in your book). The remarkable search ability the human brain has shown is the result of attentional guidance mechanisms relying on the combination of bottom-up and top-down factors [8]. Given that computer vision and Al systems have been already achieved the level of ability that is comparable to the bottom-up processing in human [7], it seems to be the topdown processing such as semantic knowledge, prior information through past experiences and common sense about the external environment. and motivational or emotional factors that can make us outperform the machine in this important task. The field is just beginning to understand how bottom-up and top-down processing interact with each other in real time and how the interaction can be effectively implemented to improve search performance in the computer vision and AI systems and other

practical applications. Future challenges for the field will also include understanding how the principles of attentional guidance mechanism by bottom-up and bottom processing can be extended from two-dimensional images to immersive, dynamic, three-dimensional environments and how these can be utilized in practical applications that require outstanding visual search ability, such as surveillance and security cameras, selfdriving cars, and robots.

#### References

- 1. Wolfe, J. M. (1998). What Can 1 Million Trials Tell Us About Visual Search? *Psychological Science*, 9(1), 33-39. doi:10.1111/1467-9280.00006.
- 2. Schweizer, K. (1998). Visual search, reaction time, and cognitive ability. *Perceptual and Motor Skills*, 86(1), 79-84.
- 3. Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62,1457-1506.
- Wu, C. C., & Wolfe, J. M. (2019). Eye Movements in Medical Image Perception: A Selective Review of Past, Present and Future. *Vision*, 3:32
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1000-1014. doi:10.1037/0096-1523.20.5.1000.
- Sobel, K. V., Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. *Psychonomic Bulletin & Review, 14*(5), 840-845.
- 7. Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.
- 8. van Zoest W, Donk M. Bottom-up and top-down control in visual search. *Perception*. 2004; 33(8). 927-937.

