Cognitive psychology: human object recognition

For a very long time object recognition in human beings has been one of the most debated topics in computer vision and psychology. Within the past ten years, psychologists have shifted their approach from construction objects in three dimensions to the view-based approach, which encourages storage of object snapshots. The recent decade has also seen debates concentrated on the static domain of images and object recognition from databases that have many photos. However, in the past few years, cognitive philosophical results have indicated that the human beings use temporal information commonly involved in the visual input while recognizing objects. This paper discovers some approaches and experiments that researchers have made trying to understand how human object recognition mechanism operates.

Key words: Human object recognitionHuman Object Recognition

Introduction

This paper will discuss the issue of human object recognition as the cognitive phenomenon. Object recognition can be defined as the ability to see and perceive the physical properties of an object, such as texture and color, and manage to apply the semantic properties, which encompasses understanding of its use and how the objects relate to each other (Bosco et al., 1995). Several studies have tried to understand the processes that human brains undergo cognizing an object (Susan, 1995). It is worth noting that until now, there is no study that can give a substantive explanation of the mechanism that the brain uses in object recognition. However, studies have developed theoretical models that try to explain this mechanism. Nevertheless, the empirical data and substantive measurements of brain mechanisms in human beings are still rare. Recent studies have evidences arguing that the human brain is a parallel system having highly specialized organs, where different stages of processing information occur. For instance, the theory of neurology synchronization argues that cooperation between different compartments of the brain is achieved through their rhythmic synchronization of activities, thus resulting in emergence of dynamic network that last for a short while (Heinrich et al., 2000).

Object Recognition Theory in Human Beings

Failure to give a rational explanation to how the brain works implies that it is not easy to describe how human brains recognize objects (Kirkpatrick, 2000). Despite the fact that many researches have been conducted to explain the object recognition mechanism, there is not any study that has given substantive explanation of the mechanism involved. However, the available theories have intensively tried to give rational explanations. Experiments conducted to identify the way human brains recognize objects involve use of assumptions in these theories. Thus, it is important to understand one of them (Susan, 1995).

Recognition by component (RBC) theory is one of the principles of object recognition in human beings that tries to explain how human brains identify objects or images despite unpredicted changes in the image orientation. In addition, RBC tries to clarify how moderately degraded and occluded images give novel examples of images that can be easily recognized by the human visual system. Proponents of this theory argue that the human object recognition system extracts geometric ions (geons), which are used to identify images as well as objects (Kirkpatrick, 2000). The proponents explain geons as simple volumes such as cylinders, cubes, spheres, and wages. RBC argues the brain stores object representations as structural descriptions, which contains specifications of an object geons and its interrelations. For instance, structural description specifies that a cylinder is below a cube (Heinrich et al., 2000).

The visual system analyzes a perceived object and parses the image into the respective constituent geons. From there, the interrelation is determined including issues such as size and relative location. The interrelation created is then matched with the geons and compared with the structural description stored in the brain. Then, if a good match is identified, the brain will successfully identify the object. It is worth noting that RBC explanation of how objects are identified is analogous to speech recognition, where small phonemes sets are combined to arise with millions of different words through organizational rules (Bosco et al., 1995). Therefore, in RBC, a geon is the same as a phoneme, and the organization rules mean the same as the spatial interrelations. Proponents of RBC argue that 36 geons are used to produce millions of different or unique objects. It is, however, important to understand that researchers developed the RBC to deal with primal recognition of objects. Primal recognition does not involve high-level objective cognitive processes and is fast acting. High-level object recognition usually engages shading of color or texture in giving finer discrimination of objects. In addition, top-down processing occurs in cases where environmental cues must be used to discriminate particular instances of an image or object. For instance, it is easy to recognize a pencil when it is occluded by several other pencils, but it is challenging to identify a pencil in a pile of leaves (Kirkpatrick, 2000).

Many studies conducted in this field have involved this and several other theories for explaining unclear concepts. Despite the fact that studies have given different explanations behind the human object recognition mechanism, all of them indicate that there are several complex processes involved, and understanding every individual process might not be possible. Some experiments conducted in this field are discussed below (Bosco et al., 1995).

Susan and Roberta (1994) used several experiments to prove that human beings are capable of recognizing or identifying common objects, such as hammer and chisels, using the sense of touch alone. In one of the experiments, they used a sample population of 100 participants. They blindfolded te participants and asked them to identify several sets of common objects as accurately and quickly as they could (Heinrich et al., 2000). It was noted that most of these objects were identified quickly in 2 to 3 seconds. The results of this experiment surprised many people, since at the time the experiment was being conducted, several studies had substantively argued that human touch is not capable of conducting such kind of high-level processing (Susan, 1995). The researchers, therefore, started arguing that the people manually and actively explore similar multidimensional objects, and images might be very important in uncovering and eventually giving rational explanations of information-processing abilities. To give a further illustration of this argument, the researchers conducted another experiment (Gao, 1998).

In this experiment, they asked the participants to conduct a haptic match-to-sample activity. The participants were given standard objects, in addition to three comparison objects, in each trial. The comparison objects differed greately in such dimensions as texture, shape, and compliance, and the participants were asked to handle one dimension alone (Bosco et al., 1995). The participants had to select one comparison object that matched the dimension standard identified. As the experiment continued, the researcher produced several different multidimensional objects that were custom designed. They designed each set to be used specifically with a given property-matching instruction (John et al. 2000). As they conducted the experiment, instruction for all trials were selected equally often from properties such as hardness, texture, weight, thermal properties, envelope shape, volume, and exact shape. Furthermore, they video-recorded and particularly analyzed participant’s hand motion during the experiment. They noticed that manual exploration was always systematic, i. e. the participants performed stereotypical motions that they termed exploratory process (EPS) (Heinrich et al., 2000).

They decided to conduct particular EP in relation to specific instructions on property matching. As the participants attempted to execute the required task, they demonstrated several motions such as lateral motion EP, pressure EP, static contact, unsupported holding, and counter following (Olivia & Antonio, 2007). The lateral motion EP was a back-forth movement of the object surface transfer, which was typically conducted with instructions on texture matching. Pressure EP involved the application of physical pressure, torques, or forces on the object axis. This was usually used to test the hardness of the object. Simple contact or static contact was made between the object and skin and was associated with testing thermal properties (Bosco et al., 1995). Unsupported holding involved lifting of the object from the supporting surface. It was used for matching the weights. Molding the palm and fingers to the contours of the object or enclosure was applied to ascertain both the envelope shape and volume. Lastly, the edge or contour following was most often used in conjunction with both exact shape and envelope shape instruction (Lawson, 2012). Look at fig 1 below for more clarification.

In another experiment Biederman and Gerhardstein (1993) tried to taste the rationality of viewpoint-dependent and viewpoint invariant human object recognition theories in their article“ recognizing depth-rotated objects: Evidence and conditions for three dimensional viewpoint invariance”. The difference between these theories is their methods of prediction that concern how a person’s recognition system performs when the viewpoint is changed (Lawson, 2012). The researcher observed both view-point invariant and view-point dependent under several experimental conditions. While conducting this experiment, the researchers noted that the existence of patterns did not support one theory over the other one. Instead, it showed how the poronents of the theories interpreted the relative relevance and the validity of experimental condition to the ecology (Susan, 1995). Claims can be made over the wider class mechanism explanatory power only if the results are seen as having normal generalizability. For example, one must perform the following Gedanken experiment: it is necessary to obtain a set of differently colored objects. The person has to imagine that the objects are presented one at a time, and his responsibility is simply to identify each object. The performance in this task or the similar one will not vary with the viewpoint, size, mirror reflection, and any other transformations. Yet, as long as new colours that are similar to the objects are added into the set, one will change his consideration and start distinguishing them in " the real world" perspective. A person might realize that no colour is diagnostic or unique for the objects, and thus one might find the results that generalize the objects (Heinrich et al., 2000).

Bierderman and Gerhadstein (1993) argued in a similar way. They suggested that state of viewpoint-invariant recognition with one that uses geon structural description (GSD) give an exclusion state of an everyday object recognition in human beings. GSD theory makes one assumption that the approximate or relative object shapes are given by the configuration of three-dimensional parts. The innovation made by the GSD theory is the use of nonaccidental properties combination as a way of recovering parts. Since this combination is assumed to be viewpoint invariant, the recovered GSD showed strict viewpoint invariance (Lawson, 2012).

Another experiment of Bosco et al (1995) was conducted on the factors that limited object recognition enforced by factors related to low-level sensory and stimuli constraint. The researchers applied methods of trying to understand how human beings use sensory signals while recognizing 3 dimensional objects in signal-detection theory (Heinrich et al., 2000). The participants in this experiment recognized familiar objects by relating their visual images to some specific objects they remembered. Secondly, they noticed a novel exemplar of common or familiar objects using structural or functional criteria. Bosco et al. (1995) used four objects namely cylindeer, cone, wedge, and pyramid in their experiment, with every object shown in eight viewpoints. This implies that the researchers had 32 images of the four objects. In each attempt to identify the objects, the researchers randomly selected one object out of the 32 and presented in noise. In addition, the researchers chose a few familiar objects with an aim of reducing minimizing memory demands (Mary & Hanny, 2000). Furthermore, they chose objects that could be viewed to act as objects recognized in the real world from multiple viewpoints. The participants were also trained in all test views to rely on high-level or mid-level memory representations (Lawson, 2012).

Psychological studies have substantively proven that human beings actively use the temporal information, such as contiguity of images in trying to explain, how human beings behave recognizing the objects (Gupta, et al., 2012). This can be the reason why several studies have discovered recognition systems that use temporal contiguity to give an extensible representation of images and objects. The human object recognition system properly performs both synthetic and real-world data and demonstrates the robustness under different illuminations (Eldelman, 1997). In addition, many studies have questioned the thought that the position of the viewer determines the way an object is recognized. Proponents of view-based approach have argued in support of this statement. For instance, Heinrich et al (2000) conducted an experiment in which the participants had to confuse two different faces. This was done through combination of associated or related to coherent temporal sequence. The researchers used several faces that were vied in three dimensions laser-scanned heads images that were altered using laser technology. Each sequence was formed of five images that were morphed continuously from the original identity. Look at the figure below.

Most of the experiments on human object recognition mechanism were conducted using carefully planned procedures and variables. However, some few studies have used processes and variables that appear confounded. For instance, Bierderman and Gerhadstein (1993) conducted an experiment that tried to taste the rationality of viewpoint dependent and viewpoint invariant human object recognition theories. In the experiment, the researchers explained that the researchers used objects colored differently in red, blue, and green. These images were simultaneously presented to the participants each of them being asked to identify individual objects. This experiment was invariant over viewpoint, mirror reflection, size, and other transformations (Gupta, et al., 2012). However, a variable that seems to confuse is the argument that when similar colors are introduced in this experiment, it is difficult to recognize this object color solely. People might fail to recognize objects that appear in different but closely related colors if they have problems with identifying colors. In addition, the objects cannot be identified based on their colors alone, but other attributes, such as weight, mass, temperature, and shape, can be considered. Therefore, failure to identify objects due to their close colors appears to be confusing (Susan, 1995).

Alternatively, Bierdman and Gerhadstein (1993) could have used other variables that could have clearly seen physical differences between the objects. The experiment could appear as follows: one could identify objects, such as cubes, pyramids, star shaped woods, cones, and cylinders, that differ in color being green, white, and black. Using a small sample population one should identify these objects while maintaining a given distance. Then, it is necessary tointroduce other objects of different but closely related colors (American Psychology Association, 2009). For instance, objects, such as rectangular woods, small and large sized cylinders, pyramids of different sizes, can be introduced. In addition, they should be similar in appearance, texture, shape, and size. The participants should be asked to recognize all of the objects while maintaining the same distance. In this experiment, the object recognition system will be put in real test since it will be needed to recognize very small details that differentiate the objects (Michael, 2008). For instance, a participant needs to identify a cube from a rectangular object, based on the dimensions of the object. Similarly, a triangular object is different from a pyramid because it has a triangular base, while a pyramid has a rectangular or square one. These results will prove that one has to consider several object properties while trying to identify it precisely. Identification of objects solely depending on the color of the object depending on color can give an impaired impression of human object recognition mechanism (Glen et al. 2001).

In addition, it seems that most of researchers have pushed the fact that object recognition is a systematic process aside. This has been proven by many scientific studies, with most of them suggesting that the object recognition is a four or six staged process. Basing on neuropsychological evidence, Michael & Heinrich (2000) classified the process into four stages. At the first stage, human beings undergo a basic process where the object components, such as depth, form, and color, are identified. In the second stage, the components identified in the second stage are grouped based on factors such as similarity and distinctiveness (Michael, 2008). Additionally, segregating the image basing on the figure and shape takes place. Thirdly, the visual representation that the brain creates is matched with several structural descriptions stored in the memory. In the last stage, the brain applies semantic attributes to the visual representation, which provides meaning, and thus the object is recognized (Yao & Fei-Fei, 2000).

In conclusion, despite the fact that the process of human object recognition is complex, the studies that have been undertaken until now give a good representation of what actually happens. Although many theories have been developed to support the experiments done in this field, there is a need for future studies to decrease the theoretical approach and adopt practical or evidence based argument.