The Cognitive Informatics Theory and Mathematical Models of Visual Information Processing in the Brain

Yingxu Wang, University of Calgary, Canada

ABSTRACT

It is recognized that the internal mechanisms for visual information processing are based on semantic inferences where visual information is represented and processed as visual semantic objects rather than direct images or episode pictures in the long-term memory. This paper presents a cognitive informatics theory of visual information and knowledge processing in the brain. A set of cognitive principles of visual perception is reviewed particularly the classic gestalt principles, the cognitive informatics principles, and the hypercolumn theory. A visual frame theory is developed to explain the visual information processing mechanisms of human vision, where the size of a unit visual frame is tested and calibrated based on vision experiments. The framework of human visual information processing is established in order to elaborate mechanisms of visual information processing and the compatibility of internal representations between visual and abstract information and knowledge in the brain. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: AI; The Basic Visual Frame; The Brain; Cognitive Informatics; Cognitive Models; Cognitive Principles; Computational Intelligence; Gestalt Principles; Hierarchical Abstraction Model; Hypercolumns; Visual Frame; Visual Information Processing; Visual Invariance

INTRODUCTION

It is recognized that, although over 90% information receptors of the brain are in the visual form, the internal processing mechanisms for the visual information are based on semantic or symbolic inferences rather than graphical reasoning (Hubel and Wiesel, 1959; Matlin, 1998; Payne and Wenger, 1998; Pinel, 1997; Westen, 1999; Wilson, 2001). In other words, the brain carries out thinking, reasoning, and inference on visual stimuli and image information in an abstract approach, and all visual information is represented and processed as visual semantic objects rather than direct images or episode pictures in long-term memory.

A fundamental question about the mechanisms of the brain is what the form of internal
representations of visual information is in long-term memory (Glickstein, 1988; Goldstein, 1999; Wang, 2009b; Wang and Wang, 2006). Early studies perceived that visual information is stored as pictures and the eyes work as cameras (Gray, 1994; Smith, 1993). Contemporary studies reveal that it may be true only in Sensory Buffer Memory (SBM) and Short-Term Memory (STM), but images retained and recognized in LTM are in the form of abstract visual semantics or symbolic concepts (Coaen et al., 1994; Hubel and Wiesel, 1959; Wang, 2009b). Therefore, the mechanisms of visual knowledge processing are based on abstract semantic analyses and syntheses.

This paper presents the cognitive informatics foundations of visual information processing in the brain and their applications in knowledge engineering and computational intelligence. In the remainder of this paper, fundamental principles of visual perceptions such as the gestalt principles, the cognitive informatics principles, and the hypercolumn theory, are described. The visual information processing mechanisms are explained by the visual frame theory and the calibration of the size of a unit visual frame. The framework of human visual information processing is developed to elaborate the fundamental mechanisms of visual information processing in the brain for visual knowledge representation and manipulation.

**COGNITIVE FOUNDATIONS OF VISUAL INFORMATION PROCESSING**

The mechanisms of visual information representation, processing, recognition, and comprehension, as well as their relationships to those of abstract information processing, are a set of fundamental questions in explaining the nature of human vision. This section presents the classic gestalt (holistic) principles and the cognitive informatics principles of visual information processing. Hubel and Wiesel’s hypercolumn theory for visual information processing in the visual cortex is introduced, which reveals the important mechanism of internal image information representation, interpretation, and processing.

**The Holistic Principles**

The classic gestalt principles of visual perception are developed in Germany based on experiments conducted in the 1920s and 1930s, where the term gestalt means an organized whole that is related to the philosophical doctrine of holism (Gray, 1994; Westen, 1999). The gestalt or holistic philosophy states that the whole is greater than the sum of its parts, which is inherited by modern system science.

In system algebra (Wang, 2008b), Wang creates a mathematical model of the holistic system principle that reveals the mechanism of abstract systems gains known as incremental union.

**Definition 1.** An incremental union of two sets of relations $R_1$ and $R_2$, denoted by $\bigoplus$, are a union of $R_1$ and $R_2$ plus a newly generated incremental set of relations $\Delta R_{12}$, i.e.:

$$R_1 \bigoplus R_2 \triangleq R_1 \cup R_2 \cup R_{12}$$

(1)

where $\Delta R_{12} \subseteq R_1 \cup R_2$ and $\Delta R_{12} = 2(\#C_1 \cdot \#C_2)$.

The incremental union operation on abstract systems is a new denotational mathematical structure, which provides a generic mathematical model for revealing the fusion principle and system gains during system unions and compositions.

Six gestalt principles for visual object and pattern perception are identified (Kanizsa, 1979) such as similarity, proximity, good continuation, simplicity, closure, and background contrast, as summarized in Table 1.
The classic gestalt principles reveal a set of important natural tendencies and fundamental mechanisms of human visual perceptions. However, they are inadequate in rigorousness in order to form a theory of visual information processing of the brain.

**The Cognitive Informatics Principles**

In cognitive informatics (Wang, 2002, 2003, 2007b, 2009a; Wang et al., 2009), a set of cognitive principles for visual information perceptions is identified as summarized in Table 2. The cognitive principles for visual perception may be considered as the aesthetic principles, which people intend to apply in perception and identification of perfect and coherent human figures, physical objects, and natural surroundings.

It is noteworthy that many of the visual perception principles in Table 2 have also identified in mathematics and philosophy. This is an interesting finding on the relationship between

---

**Table 1. The Gestalt Principles of Visual Perceptions**

<table>
<thead>
<tr>
<th>No.</th>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Similarity</td>
<td>The tendency to see resemble objects and patterns belong to a same group.</td>
</tr>
<tr>
<td>2</td>
<td>Proximity</td>
<td>The tendency to partite discrete image into groups.</td>
</tr>
<tr>
<td>3</td>
<td>Good continuation</td>
<td>The tendency to see intersected curves and images continued smoothly.</td>
</tr>
<tr>
<td>4</td>
<td>Simplicity</td>
<td>The tendency to see an image in the simplest way by analysis.</td>
</tr>
<tr>
<td>5</td>
<td>Closure</td>
<td>The tendency to see a completely enclosed border by ignoring gaps or cloaks.</td>
</tr>
<tr>
<td>6</td>
<td>Background contrast</td>
<td>The tendency to identify larger and dark objects in an image as the ground; and the smaller and brighter objects as the front-end figure.</td>
</tr>
</tbody>
</table>

**Table 2. The Cognitive Informatics Principles of Visual Perceptions**

<table>
<thead>
<tr>
<th>No.</th>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Association</td>
<td>The tendency to find links and relations among individual objects and images.</td>
</tr>
<tr>
<td>2</td>
<td>Symmetry</td>
<td>The tendency to identify a symmetry in images.</td>
</tr>
<tr>
<td>3</td>
<td>Perfection</td>
<td>The tendency to perceive a perfect image from the given partial information.</td>
</tr>
<tr>
<td>4</td>
<td>Abstraction</td>
<td>The tendency to use a semantic label to denote an image.</td>
</tr>
<tr>
<td>5</td>
<td>Categorization</td>
<td>The tendency to classify similar images into a group.</td>
</tr>
<tr>
<td>6</td>
<td>Analysis</td>
<td>The tendency to identify common meta-shapes or meta-figures in images.</td>
</tr>
<tr>
<td>7</td>
<td>Appreciation</td>
<td>The tendency to be sensitive on borders, intersections, changing points, or differences in images.</td>
</tr>
</tbody>
</table>
humanity, aesthetics, cognitive informatics, mathematics, and philosophy. Particularly, the perfection principle elicits an important human tendency to reconstruct the whole when only a portion of an image or picture can be seen.

The Hypercolumn Theory of the Visual Cortex

It is recognized there are 1.5 million axons that link each Lateral Geniculate Nucleus (LGN) cells to the visual cortex known as the striate cortex (Glickstein, 1988; Goldstein, 1999). Hubel and Wiesel, Nobel Prize laureates in Physiology and Medicine in 1981, discovered the special orientation selectivity of visual neurons for barlike stimuli with specific orientations (Hubel and Wiesel; 1959, 1979). Hubel and Wiesel revealed that the basic structure of vision cells known as the hypercolumns.

Definition 2. A hypercolumn (HC) is a structured visual processing module in the visual cortex, corresponding to a unit area of the retina, which is capable to process a basic unit of visual information in the three aspects of ocular dominance, location on the retina, and orientation of the stimuli.

Tovee reported that there are over 2,500 hypercolumns in the striate cortex (Tovee, 1996), which is equivalent to the size of a visual frame as determined in the experiment described in the next section on the visual frame theory.

Based on Definition 2 and the number of HCs in a visual frame, a mathematical model of HC is developed by the author in order to formalize the discovery of Hubel and Wiesel (1959).

Definition 3. The formal model of an HC is a 3-tuple, i.e.:

\[ HC \triangleq (E, P, O) = (\{L_e, R_e\}, \{X, Y\}, \{0', 5', 10', ..., 90', ..., 175', 180'\}) \]

where \( E \) represents the ocular dominance (the left eye (\( L_e \)) or the right eye (\( R_e \))), \( P \) the location on the retina determined by the coordinates (\( X, Y \)), and \( O \) the orientation of a stimulus in the scope of 0° through 180°, i.e.:

\[ E = \{L_e, R_e\} \]

\[ P = X \times Y = [0...50] \times [0...50]^T \]

\[ = \begin{bmatrix} (0,0) & \cdots & (0,50) \\ \vdots & \ddots & \vdots \\ (50,0) & \cdots & (50,50) \end{bmatrix} \]

\[ O = \{O_0, O_1, O_2, ..., O_{34}, O_{35}\} = \{0', 5', 10', ..., 175', 180'\} \]

For instance, the following HCs represent that the left eye detects a 45° bar at the coordinates (2, 49) in the retina, and the right eye detects a 135° bar at the coordinates (32, 6), respectively.

\[ HC_1 = (E_1, P_1, O_1) = (L_e, (2,49), 45'); \quad HC_2 = (E_2, P_2, O_2) = (R_e, (32,6), 135') \]

Complex shapes and images can be represented by multiple HCs corresponding to the given pattern of the stimuli.

Based on Hubel and Wiesel’s, as well as Tovee’s, experiments and Definitions 2 and 3, the following theorem and corollary can be derived.

Theorem 1. The symbolic representation mechanism of vision states that the basic unit
of vision is a barlike area modeled by the HC rather than a simple dot.

Proof: Hubel and Wiesel’s experiments and the HC layout determine that the semantics of internal image representation is an abstract symbolic structure as shown in Eq. 2, i.e., \( HC = (E, P, O) \).

Corollary 1. An image frame is represented by a set of \( 50 \times 50 \) HCs.

Proof: Directly based on Definitions 2 and 3, Corollary 1 can be proven.

The advantage of the HC mechanisms in visual image detection and representation is that it is reliable, fault-tolerant, and anti-noisy. Any noisy HC in an image frame may be easily identified and corrected in later-phase cognitive processing. According to the hypercolumn theory and Theorem 1, the most sensitive shapes of traffic signs to the brain would be in the forms of bars rather than circles.

THE VISUAL FRAME THEORY OF HUMAN VISIONS

Definition 4. A visual frame is an invariant area of eyes with a certain sensorial resolution in number of pixels or bits.

The following psychological experiment is designed to test the typical size and resolution of a unit visual frame of human vision.

The Experimental Test of Human Visual Frames

Experiment 1. The layout of the experimental test is illustrated in Figure 1, where a CRT is adopted with length 28.0cm, width 20.6cm, and its resolution is \( 1,024 \times 1,024 = 1Mb \) pixels. The minimum area of the visual frame of the eye is represented by \( A_v \), where the minimum pixels can be visible at the nearest distance, \( l_0 \approx 4.6cm \). The diameter of the visual frame is tested as \( d \approx 0.9cm \).

In the test, the basic area of the visual frame \( A_v \), the area of the CRT \( A_{CRT} \), and the resolution of the CRT \( R_{CRT} \) are obtained as follows:

\[
A_v = 2 \cdot \frac{1}{4} \pi d^2 = 0.5 \cdot 3.14 \cdot 0.9^2 = 1.3 \text{ [cm}^2]\]

\[
A_{CRT} = 28.0 \cdot 20.6 = 576.8 \text{ [cm}^2]\]

\[
R_{CRT} = 1,024 \cdot 1,024 = 1,048,576 \text{ [bit]}\]

(7)

On the basis of the above testing results, the resolution of a unit vision frame of humans, \( R_v \), can be calibrated with that of the CRT by proportional equivalency, i.e., \( A_v : R_v = A_{CRT} : R_{CRT} \). This leads to the finding of the unit

Figure 1. Calibration of the invariant size of a visual frame
The size of human visual frame in the following theorem.

**Theorem 2.** *The maximum resolution of a vision frame is a constant that is proportional to the number of visual sensation nerves, or the number of visible pixels in the visual frame, i.e.:

\[
R_{v0} = \frac{A_v \cdot R_{CRT}}{A_{CRT}} = \frac{1.3 \times 1,048,576}{576.8} = 2,363.3 \text{ [bit]}
\]

This result conforms well with the number of hypercolumns, \#HC \approx 2,500, according to Toovee (1996). Theorem 2 indicates that, although there are about 5 million cones and 120 million rods in the retina as the array of light receptors (Goldstein, 1999), the size of visual frame or the resulted image pixels of each eye is much smaller and it is invariant from the distance between the eyes and the visual object.

**Properties of Visual Frames**

Theorem 2 indicates that the size of a human visual frame is invariant in term of number of pixels within the visual field. According to Theorem 2, the closer the object, the higher the resolution in the visual frame; and vice versa. The maximum information that can be obtained by eyes without saccade is \(R_{v0} = 2,363 \text{ bits (pixels)}\), no matter how large is the external image.

Therefore, large-frame information must be scanned by multiple units of visual frames via saccades. This explains why no one may read more than three or four words in a line without move one’s eyes.

**Corollary 2.** *The visual resolution of eyes \(R_v\) is inversely proportional to the distance of the object \(l\), i.e.:

\[
R_v \leq R_{v0} | l \geq l_0 = 4.6 \text{ cm}
\]

According to Corollary 2, the visual resolution of the eyes is decreasing when the distance of the object in the visual frame is increasing.

When the maximum sample rate of vision \(s_v\) and the sample rate of each HC \(s_{HC}\) are known, i.e., \(s_v = 50 \text{ frame/s (Tuker, 1997)}\), and \(s_{HC} = 50\text{bit/s}\), the processing speed of visual information can be determined as follows.

**Corollary 3.** *The maximum rate of human visual information processing \(S_v\) is a product of the maximum sample rate \(s_v\) and the maximum resolution of the vision frame \(R_{v0}\), i.e.:

\[
S_v \triangleq s_v \cdot R_{v0} = 50 \text{ [frame/s]} \cdot 2,363 \text{ [bit/frame]} = 118,150 \text{ [bit/s]}
\]

Corollary 3 indicates that the maximum visual information transformation rate between sensory-buffer memory and short-term memory is equivalent to approximately 118.2kbps, which forms the upper bound of visual information processing of the human brain.

**THE FRAMEWORK OF HUMAN VISUAL INFORMATION PROCESSING SYSTEM**

The Framework of Visual Information Processing (FVIP) of the brain is shown in Figure 2, where three forms of memories, known as the Sensory Buffer Memory (SBM), Short-Term Memory (STM), and Long-Term Memory (LTM), are involved under the control of the Perception Engine (PE) of the brain (Wang and Wang, 2006; Wang, 2007a). The visual information is stored, retrieved, and manipulated in the three memories in different forms. SBM temporarily stores the analog visual information as a direct image of the external object, which is transferred into STM as an analog visual frame. Except the part of abstract or symbolic information, the visual information retained in LTM can be classified into three types, namely...
the basic image base, the semantic image base, and the episodic image base.

The FVIP model reveals that the major forms of visual information represented in LTM is non-analog or non-photonic. Instead, it is symbolic, semantic, and denotational, except a small part of the visual information in the semantic image base such as common and simple shapes and solid figures (Wang, 2009b), or the episodic image base such as image of family members, home, highly impressed scenes of events, highly familiar places, and very frequently used facilities or tools.

**Theorem 3.** Acquired visual information is represented in symbolic or semantic form in LTM.

**Proof:** This theorem can be proven by Theorem 1 and related experiments.

Theorem 3 is supported by many observations and psychological experiments. For instance, Reed and his colleagues reported that the mental images in LTM are in propositional codes (Reed, 1972; Reed et al., 1974) or in the form of semantics.

**Experiment 2.** Novel and fancy images that have never been seen in one’s experience may be perceived in subconscious dreams or during conscious imageries. This mental phenomenon indicates that the LTM does not retain photonic images from the real world, and images one sees during thinking is reconstructed in STM by the perceptual engine rather than retrieved by searching from LTM.

An advantage of the abstract internal representation of visual information is that a category of equivalent images may be treated by the same semantic images as shown in Figure 3. Therefore, highly efficient, flexible, and adaptive visual information processing and comprehension mechanisms may be naturally implemented (Wang, 2008a, 2009b).

It is noteworthy that the STM is the space for both analogy image processing acquired from SBM and internal image reconstruction retrieved from LTM. In other words, STM is the space of image information processing, coding, decoding, and reconstruction. Therefore, what the mind sees during visual information and knowledge processing are restructured images located in STM, particularly in the visual cortex.

**THE MECHANISMS OF VISUAL**

Figure 2. The Framework of Visual Information Processing (FVIP) in the brain

**SMB** – Sensory buffer memory; **STM** – Short-term memory; **LTM** – Long-term memory
INFORMATION PROCESSING

The FVIP model developed in preceding section explains the fundamental mechanisms of human vision, visual information processing, imagery, perception, image reconstruction, and pattern recognition. On the basis of FVIP, the following fundamental questions about human vision can be answered:

- Why human long-term memories of images are always blurred and vague?
- Why it is easy to compare two photos in STM or in both STM and SBM, but is not easy to compare those in LTM?
- How internal visual information is represented?

Table 3. The Role of Abstraction in Human Inferences

<table>
<thead>
<tr>
<th></th>
<th>Abstract (Concept)</th>
<th>Analog (Image)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal inference</td>
<td>√</td>
<td>×</td>
<td>Reasoning based on internal abstract concepts or images</td>
</tr>
<tr>
<td>External inference</td>
<td>√</td>
<td>√</td>
<td>Reasoning based on external abstract or visual objects</td>
</tr>
</tbody>
</table>

Figure 4. The semantic representation of concrete images

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bird</td>
<td>Camel</td>
<td>Elephant</td>
</tr>
<tr>
<td>Car</td>
<td>Hammer</td>
<td>Fork</td>
</tr>
</tbody>
</table>
**Corollary 4.** *The human tendency in visual information processing is to perform abstract or semantic visual inferences rather than direct diagram-based visual inference.*

The above corollary is supported by the following experiment.

**Experiment 3.** *The brain cannot carry out image inferences without looking at the real world images, e.g. the images and their relations in Figure 2 or Figure 3, because this cognitive process requires too large memory space beyond the capacity of STM in the brain for supporting complex inferences.*

Based on Experiment 3, the role of abstraction in human inference can be observed and contrasted in Table 3, where an internal analog inference cannot be carried out based only on analogy images. Table 3 provides another evidence to support Theorem 3 and the FVIP model. That is, the internal visual representation in LTM is abstract semantic objects rather than image objects.

For example, typical semantic objects of basic shapes and images can be represented as shown in Figure 4, where 6 semantic objects are given to represent 6 elementary images adopted from the Silhouette database at [http://www.lems.brown.edu/vision/software/216shapes.tar.gz](http://www.lems.brown.edu/vision/software/216shapes.tar.gz), where 216 visual objects have been created.

**CONCLUSION**

This paper has presented the cognitive informatics foundations of visual information and knowledge processing in the brain. A set of cognitive principles of visual perception, such as the Gestalt principles, the cognitive informatics principles, and the hypercolumn theory, has been elaborated. The visual information processing mechanisms of human vision have been explained by the invariant visual frame theory, where the size of a unit visual frame has been determined. The framework of human visual information processing has been developed to elaborate the mechanisms of human visual information processing. Based on it, the mechanisms of visual information processing and their compatibility with abstract information processing have been analyzed and contrasted.

It has been revealed that the visual information is represented in symbolic or semantic forms. The basic unit of vision has been identified as a barlike area known as Hubel and Wiesel’s hypercolumns, and an image frame is represented by a set of $50 \times 50$ hypercolumns. The size of a visual frame has been calibrated as about 2,363 pixels with the property of an invariant resolution inversely proportional to the distance of visual objects. One of the major findings reported in this paper has been that the human tendency in visual information processing is to perform abstract semantic visual inferences rather than direct diagram-based visual inferences. In other words, the internal representation of visual and abstract knowledge shares a unified and coherent form known as concept networks. According to the Hierarchical Abstraction Model (HAM), the internal representation of both visual and symbolic knowledge are in abstract forms, and only a higher-level abstract means is precise and adequate to express an object at a given level of abstraction in the HAM model.

**ACKNOWLEDGEMENT**

This work is partially sponsored by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors would like to thank the anonymous reviewers for their valuable suggestions and comments on this work.

**REFERENCES**


Yingxu Wang is professor of cognitive informatics and software engineering, Director of International Center for Cognitive Informatics (ICfCI), and Director of Theoretical and Empirical Software Engineering Research Center (TESERC) at the University of Calgary. He is a Fellow of WIF, a P.Eng of Canada, a Senior Member of IEEE and ACM, and a member of ISO/IEC JTC1 and the Canadian Advisory Committee (CAC) for ISO. He received a PhD in Software Engineering from The Nottingham Trent University, UK, in 1997, and a BSc in Electrical Engineering from Shanghai Tiedao University in 1983. He has industrial experience since 1972 and has been a full professor since 1994. He was a visiting professor in the Computing Laboratory at Oxford University in 1995, Dept. of Computer Science at Stanford University in 2008, and the Berkeley Initiative in Soft Computing (BISC) Lab at University of California, Berkeley in 2008, respectively. He is the founder and steering committee chair of the annual IEEE International Conference on Cognitive Informatics (ICCI). He is founding editor-in-chief of International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), founding editor-in-chief of International Journal of Software Science and Computational Intelligence (IJSSCI), associate editor of IEEE Trans on System, Man, and Cybernetics (A), and editor-in-chief of CRC Book Series in Software Engineering. He is the initiator of a number of cutting-edge research fields such as cognitive informatics, abstract intelligence, denotational mathematics, theoretical software engineering, and built-in tests. He has published over 105 peer reviewed journal papers, 190 peer reviewed full conference papers, and 12 books in cognitive informatics, software engineering, and computational intelligence. He is the recipient of dozens international awards on academic leadership, outstanding contribution, research achievement, best paper, and teaching in the last 30 years.