DE GRUYTER
OPEN

JAISCR, 2016, Vol. 6, No. 3, pp. 137– 154
10.1515/jaiscr-2016-0011

A BIO-INSPIRED INTEGRATION METHOD FOR OBJECT SEMANTIC REPRESENTATION

Hui Wei

*Laboratory of Cognitive Modeling and Algorithms, Department of Computer Science,
Fudan University, Handan Road No.220, 200433 Shanghai, China*

Abstract

We have two motivations. Firstly, semantic gap is a tough problem puzzling almost all sub-fields of Artificial Intelligence. We think semantic gap is the conflict between the abstractness of high-level symbolic definition and the details, diversities of low-level stimulus. Secondly, in object recognition, a pre-defined prototype of object is crucial and indispensable for bi-directional perception processing. On the one hand this prototype was learned from perceptional experience, and on the other hand it should be able to guide future downward processing. Human can do this very well, so physiological mechanism is simulated here. We utilize a mechanism of classical and non-classical receptive field (nCRF) to design a hierarchical model and form a multi-layer prototype of an object. This also is a realistic definition of concept, and a representation of denoting semantic. We regard this model as the most fundamental infrastructure that can ground semantics. Here a AND-OR tree is constructed to record prototypes of a concept, in which either raw data at low-level or symbol at high-level is feasible, and explicit production rules are also available. For the sake of pixel processing, knowledge should be represented in a data form; for the sake of scene reasoning, knowledge should be represented in a symbolic form. The physiological mechanism happens to be the bridge that can join them together seamlessly. This provides a possibility for finding a solution to semantic gap problem, and prevents discontinuity in low-order structures.

Keywords: bio-inspired method, object representation, prototype

1 Introduction

1.1 How to ground visual concept semantic?

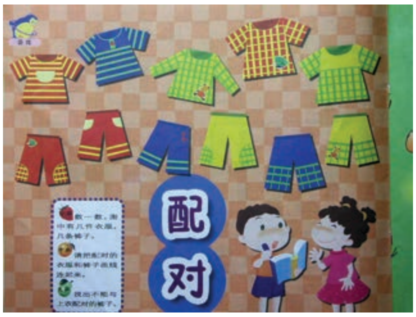


Figure 1. The problem solving task here is to match each T-shirt with its appropriate trousers

Figure 1 shows an example of problem solving carried out by kindergarten children. The children were asked, according to the given picture, to match a T-shirt with trousers. To accomplish this task, the children needed to have learned some basic concepts or knowledge about clothes. If this problem is presented to a computer in the form of a picture, to obtain an answer requires image understanding and knowledge-based reasoning. Realizing this in an artificial intelligence system, however, is quite difficult. It is not difficult for us to define some concepts and rules in a knowledge base, such as in a logical expression we define:

$$(\exists\, x)(\exists\, y)\ \mathit{Is\_Tshirt(x)} \wedge \mathit{Is\_trousers(y)}$$
$$\wedge\ \mathit{Same\_size(x,\ y)} \wedge \mathit{Same\_color(x,\ y)}$$
$$\wedge\ \mathit{Same\_pattern(x,\ y)} \rightarrow \mathit{Match(x,y)}.$$

But how can these concepts and rules correlate with their occurrences in a picture? If this connection can't be built, then the computer can't know what is *x* and where is *x*. So-called high level knowledge defined like this does not settle the problem of operability, i.e. how to apply these concepts on a picture is still unclear. Consequently, these symbolic rules will not be triggered successfully. Almost all applications of computer vision or image understanding concerned with knowledge are seriously suffering from semantic gap problem. In artificial intelligence there is a very famous and conventional hypothesis, which is the discontinuity in low-order structure. In the past history of AI, this hypothesis was a kind of compromise between high computational demand and simple computer. We think this is the reason of leading semantic gap. The reason man does not suffer from semantic problem is because human brain has rich representation layers and rich process layers. We need to find a solution that can connect abstract symbolic concept with pixel-level manipulations, on a condition of this solution submitting to formal paradigm.

In many applications on image retrieval, in advance, all training pictures were manually labeled what there are in a picture [1, 2]. And these symbols are regarded as the semantic of picture. This practice is very popular. For example, a word "car" was tagged to a picture, but this label did not define any detail about this concept. So this method is too simple and too coarse to solve semantic problem. Because (a) the words and syntax used for labeling is highly task-depended, and many semantic details are neglected; (b) these labels can neither be used as standards to include any other positive case, nor exclude any other negative case; (c) using these labels cannot differentiate an object from its environment. In fact, a so-called semantic label is only a brief index. It is far away from semantic definition.

This solution should have some kind of connectionist infrastructure.

1.2 Two aspects of this solution

Now there are two common senses in computer vision, saying that (a) an animal's vision system is much cleverer than a machine one, and (b) object recognition needs a help from higher level knowledge. To (a), the advantage of a biological vision system's that it has a systematic architecture with rich knowledge and rich levels of representing and processing. And some similar strategy has been applied to represent image semantic [3]. To (b), what is knowledge needs to be clarified on the base of a biological representation paradigm. These two aspects outline the main parts of our solution.

1.3 How to fulfill a more detailed definition of an object?

Many concept-representation concerned studies have been done. Some of them were facing image retrieval, thus the bag-of-words algorithm were applied[4; 5]. Another more complicated method is a structural decomposition model. In this model the shape of an object is described in terms of relatively a few generic components which are joined by spatial relationships [6, 7, 8, 9, 10, 11, 12, 13]. Region has been proved to be a kind of effective element to describe image semantic [14]. For example, due to region can provide much more extensive information than pixel, region-based representation can facilitate some advanced processing such as segmentation [15, 16] and tampering detection [17]. Similar to region, patch-based descriptors were also popular in semantic definition [18, 19]. And patch can also represent rich information in an expanded area, so using it can also implement some semantic-concerned task, such as image inpainting [20] and image synthesis [21]. For visual concept application, the representation form [22], topological relations among regions [23], template of an object [24] is very important. All above works were a good start on concept acquisition, explicit representation and top-down effect of visual concept. We also want to contribute on these themes.

We think a prototype representation of an object is one of a crucial types of knowledge for object recognition, especially when neither object nor its background is highly specific or severely restricted. That famous example, recognizing a spotted dog from an environment of swing tree shadow, is a typical instance of needing prototype. Theory of prototype can be found in the chapters of perception and concept acquisition of Cognitive Psychology [8, 25]. A physiological neuroscience study [26] shows that semantic is processed in the left inferior prefrontal cortex. This hints that a procedure of integration is very necessary, because only higher cortex can collect and process information from extensive area. Here the form of prototype or the materialization of this kind of knowledge is important for the computer vision (CV), because on the one hand it was the result of perceptual experience, and on the other hand it should be able to guide the future practice of downward processing. We can conclude that the prototype of an object is the final destination of learning and also is the source of bi-directional visual processing. So, the formalization schema of prototype representation is the key point. But how to realize this prototype definition is always worth studying. For the sake of defining the prototype of a type of object, two things are necessary. The first is representing the parts of an object, and the second is describing their topographical relationships. This schema should be compatible with upward raw-data input as well as downward pixel-organizing instruction. This has always been the basic goal of CV, but its priority has always been delayed.

The psychological experiments on mental image proof that the prototype of an object resembles the original, and experiments on memory proof that the prototype is somewhat abstract and declarable. From a perspective of AI, these are two conflict requirements: the former is pixel-suited and the latter is symbol-suited. Who can satisfy them simultaneously? The answer is neural vision system, a highly optimal system after a long time of natural selection. The biological vision is made up of many processing loops. The middle layers, from ganglion cell (GC) to V1, are the intersection of bottom-up data and top-down concept. We think these middle layers are very important for producing semantic and grounding semantic, or here are the key locations of semantic emerging. This paper focuses on a kind of non-classical receptive field (nCRF) mechanism, and uses it to form a prototype representation. Perhaps that is the basic tool for brain to bridge low-level stimulus and high-level semantics.

Here we use Figure 2 to highlight our fundamental motivation and technical solution. The second section of this paper is about neural mechanism of nCRF. The third section is a bio-inspired model designing as an infrastructure of forming representation. The forth section is about a nCRF-based Delaunary triangulation strategy to obtain many small normal triangles. The fifth section is about how to combine these candidates to some expanded polygons and similarity comparison between two polygons. The sixth section describes an explicit concept representation through a tree structure and also discusses how product rules can reflect direct connections from high-level knowledge to low-level pixel manipulations. The last two sections are about how this bio-inspired infrastructure might ground semantic.

2 Non-classical receptive field mechanism

2.1 Bio-inspired design is a good option

Why brain can protect against the problem of semantic gap should be concerned with its physiological structure. The biological system is worth simulating because it had been evolved for hundreds of thousands of years. We believe that its structure and function had been tested thoroughly and should have been highly optimized. So a bio-inspired principle should be much more rational for algorithm design. Now let's see a fundamental mechanism.

2.2 The neural mechanism of nCRF

Ganglion cells (GC) are the most important cells in retina. They locate at the rear path of information transmitting in retina. Since 1960s, many researchers found there was a large region outside the classical receptive field (CRF). In this region, light spot stimuli cannot directly cause a reaction of the cells, while they can modulate the reaction caused by the CRF. And this modulation can be facilitory, inhibitive or disinhibitive [27, 28], and this expanded receptive field is called as non-classical receptive field (nCRF). Neurophysiologic researches [29, 30] show a very complex formation of nCRF constructed by receptor cells (RC), horizontal cells (HC), bipolar cells (BC) and amacrine cells (AC), and also by outer and inner plexiform layer. Activities in the region can inhibit the antagonistic effect and compensate the loss of low spatial frequency

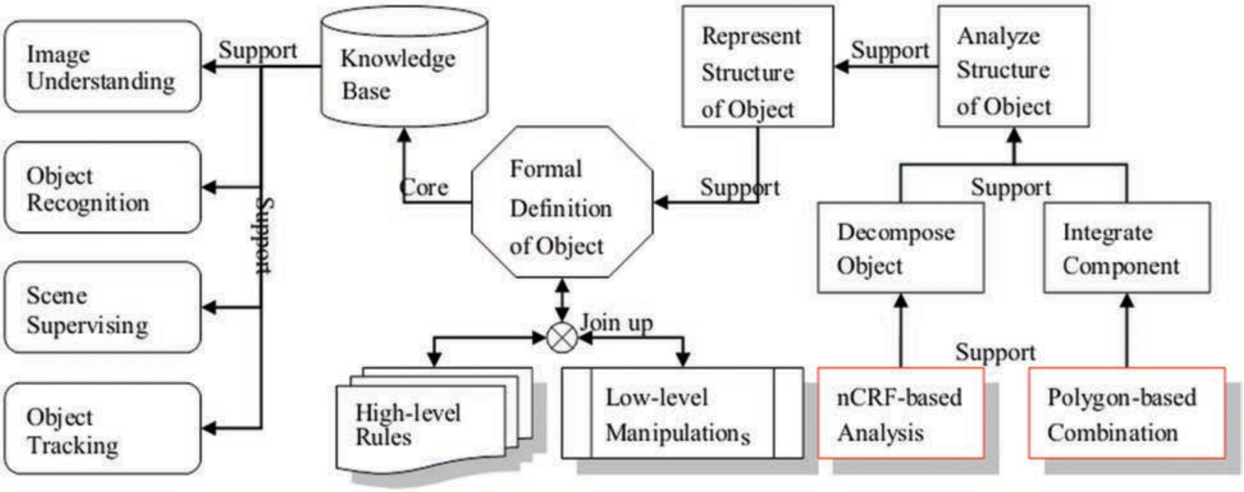


Figure 2. A whole logic framework of implementing formal representation of prototype

caused by the CRF center-periphery antagonism to some extent. nCRF plays an important role in representing contour [31], shape [32], curvature [33, 34]. nCRF can compensate loss of low spatial frequency to some extent by adding the output from extra surround area of CRF. Through its nCRF, a GC expands its information-receiving scope several times as CRF; undoubtedly this neural basis makes GC able to integrate image features in a large scale. Moreover, we think it plays a significant role in separating figures out of background.

2.3 nCRF can self-adapt its size so as to optimize its representational role

From the point of view of an image processing and understanding, GC and its nCRF mechanism are of great significance in a feature detection, and every GC plays a role of feature descriptor. What is surprising is that each GC can adjust its size of nCRF dynamically in order to make the characteristics occurring in its nCRF monotonous. So, a GC can reduce its size of nCRF to represent fine detail occurring in a local area, and also can expand its size to represent a big block with unitary feature. The GC and its nCRF are self-adaptable, localized, with regular shape, autonomous, and parallel. These attributes make it to be an ideal candidate of general descriptor. In traditional image processing, we note that GC was ussually used to extract boundary, filter noise or enhance image. But these are absolutely not the main functions of it, but somewhat wasting its talents.

3 A representation schema basing on nCRF mechanism

3.1 Size changeable nCRF can record patch, which can be a sub- component of an object

The research of the neurophysiology has shown, that according to different brightness, color or velocity of stimuli, the size of receptive field can be changed dynamically. This self-adaptability satisfies following two cases. In a dark environment, GCs will enlarge the size of receptive fields by means of reducing the spatial resolution, and accept much light through spatial summation. While distinguishing some fine details, the receptive fields will turn smaller so as to improve the spatial resolution. Each GC can implement this automatically by a local neural circuit. Besides the CRF, there are many rings. We call them sub-regions, and they are made up of nCRF. The maximum size of nCRF is about 3-6 times than the size of CRF. Figure 3(a) is a model of nCRF with multiple sub-regions, and (b) shows several initial RF being covered on an image, and (c) shows they were resized through increasing or decreasing sub-regions according to the stimuli they confronting with.

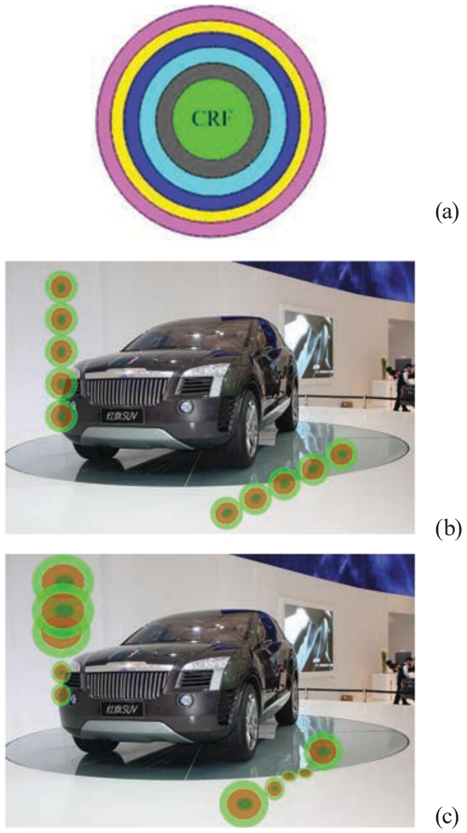
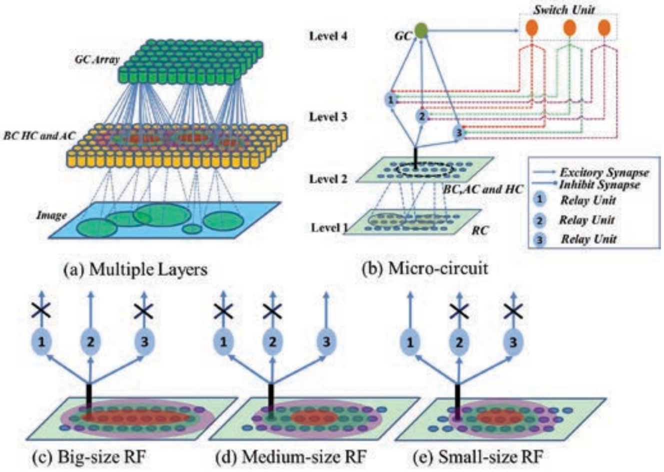


Figure 3. A model of RF and how it changes its size

Then, basing on the structure of neural circuit, we can design a schema to represent image. An image is actually the combination of many fine details and some homogeneous blocks. If we have a matrix of RFs and assign small size RF to record detail and assign big size RF to record block, then we may get an approximation of the original image (Figures 3(b) and (c)). This motivation is very important, because it reveals the basic principle of GC's working. GC always makes its role of specialized representation. For simulating this schema, we construct a matrix of computational units and let them change their nCRF sizes dynamically to fit the situation they confront. If a unit happens to cover a fine detail, then it will shrink its RF so as to record the detail more accurately. If it happens to cover a piece of homogeneous texture, then it will expand its RF to represent a unitary block. The fore-mentioned multiple sub-regions are for the sake of size changing. If expanding is necessary, then one or more sub-regions will be appended, and if shrinking is necessary, then one or more sub-regions will be deleted. This mechanism guarantees the feasibility of this self-adaptation. Now we can see that the function of GC matrix is to earn a just enough representation.



(a) Multiple Layers
(b) Micro-circuit
(c) Big-size RF
(d) Medium-size RF
(e) Small-size RF

Figure 4. A multi-layer neural architecture and how a GC adjusts its RF dynamically. (a) A multiple layers architecture. For clearness only several GCs and their RF are drawn. A RF has three parts: positive center, negative surround, and positive extra-surround. (b) This is a neural circuit of RF adjusting dynamically, which is a small functional unit of (a). Dynamic adjustment of RF due to neuron can change its destination of projecting output according to changing stimulus. This can be realized by three relay neurons and three switch neurons. A switch neuron imposes its backward control on three relay neurons, and selectively permits only one relay outputting its signal upwards to GC. This makes a relay neuron may have a chance to join one of three different rings of a RF. In (c)-(d), with the different switch turning on, the same neuron may exclusively participate in forming one of rings of a RF. Then a size-changing RF comes into being. (c) is a big one, (d) is a middle one, and (e) is a small one.

This GC-based image representation is more compact than pixel-based bitmap, because what a RF can represent is usually bigger than a pixel, so it is more efficient. Figure 4 is a hierarchical computational model of a GC and its inferior RF. At the highest level, a GCs array will turn a pixel-bitmap into a block-group. The dimension of block-group will decrease greatly than that of pixel-bitmap. A block-grained representation is more meaningful than pixel-wise representation, and must ease the emergence of semantic.

An ideal self-adaptability of GC is it adjusting its RF to a proper size coinciding with the scale of main component of stimuli occurring in its RF. This is done by a sequence of operations that append or withdraw sub-regions to or from one of three parts of RF. This causes that a GC can summary the attributes of stimuli in different area. Figure 5 is an algorithm to realize this. A more detailed implementation of this model can be seen in [35] and [36].

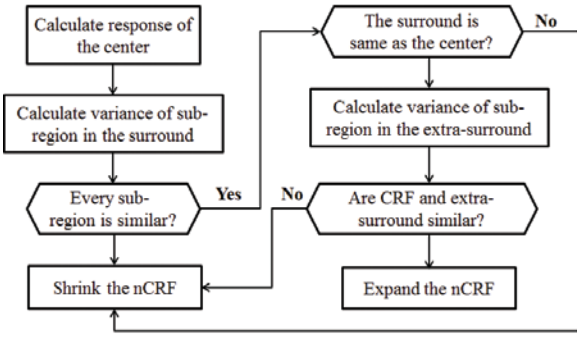

Calculate response of the center
Calculate variance of sub-region in the surround
Every sub-region is similar?
Yes
Shrink the nCRF
The surround is same as the center?
No
Calculate variance of sub-region in the extra-surround
No
Are CRF and extra-surround similar?
Expand the nCRF

Figure 5. The Dynamic Adjustment of RF

4 4 Prototype emerging from integration

In image understanding, a visual concept is activated if and only if thousands of pixels are arranged properly. So semantic is the result of integrating pixels in terms of some statistical pattern of distribution. The algorithm in previous section can facilitate the emergence of this kind of pattern, and one or several stable patterns are right to be prototypes that define a class of object.

4.1 RF mechanism archives a higher efficiency of representation

Here we also use salient object as learning sample [37]. The left of Figure 6 is an original picture, and the right is the result of GC-array running on the picture. The red circles denote the final sizes of RF after dynamic adjustment. For that bird, its back and wing possess same color or similar textures, so we represent them only by a dozen of big size RFs instead of many unorganized pixels. And its eye and beak possess tiny details, so we represent them by some small size RFs. So, a highly efficient representation is achieved. A fact that can't be ignored is that circle has regular shape and well defined algebraic formula, so it is easy for parameterization, and consequently it is easy to form a symbolic representation. And a parameterized representation also does not prevent original image from being rebuilt accurately.

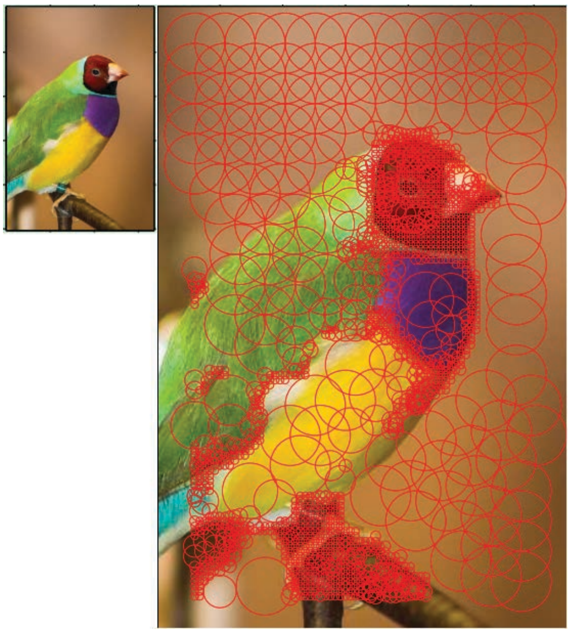
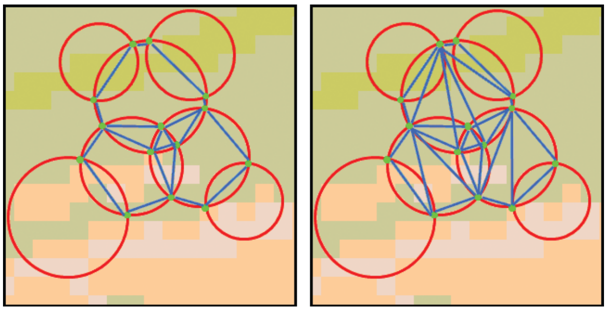


Figure 6. Size-changeable RFs bring an efficient representation

4.2 Set of regular blocks: a compact representation of object

Many existing algorithms in computer vision use the pixel-grid as the underlying representation. The pixel-grid, however, is not a natural representation of visual scenes. A good representation schema would be more natural, and presumably more efficient, to work with perceptually meaningful entities obtained from a low-level grouping process. Superpixels [38] represent a restricted form of region segmentation. Turbopixels [39] represent an image with a lattice-like structure of compact regions by dilating seeds so as to adapt to local image structure. The superpixel algorithm should partition an image into regions that are approximately uniform in size as: shape (compactness), minimizing region under segmentation, provided that superpixel size is comparable to the size of the smallest target region. Turbopixels achieve this by designing a geometric flow that dilates an initial set of uniformly distributed seeds, where each seed corresponds to one of superpixels. So, it can also be considered as a compact image representation, each of them should represent a simply connected set of pixels. Both of them provided two representation methods in order to achieve computational efficiency, representational efficiency, perceptual meaningfulness and near-completeness [40]. Our algorithm based on nCRF can also accomplish the same goals through using Inscribed polygon and Delaunay Triangulation from nCRF result. The mechanism to generate them is illustrated in Figure 7.



Every red circle denotes a nCRF and green dots denote intersections of circles. Both Inscribed Polygon (the left) and Delaunay Triangulation (the right) are derived from intersection points on neighboring circles.

Figure 7. A superpixel-like effect realized by nCRF-based algorithm

In order to testify the performance between our algorithm and superpixels, we run programs of superpixels, Inscribed Polygon and Delaunay Triangulation schemas on two different image databases. One is CityplaceBerkeley image database (481×321), the other is Microsoft Research image database (640×480). One of results is shown in Figure 8. It can be seen that the effects in (c) and (d) are similar to (a), which indicates that a complete coverage by Inscribed Polygons and Delaunay Triangulations can be calculated from RFs. But there still has a very important difference between algorithms through RFs and superpixels. Superpixel has an irregular shape, and still in a form of pixel-set instead of a form of brief vectors. This makes it difficult to be represented, recorded and operated algebraically. Due to the irregularity of shape, either the boundaries or the vertexes of a superpixel are dot-matrix data but not vectors. If the vectorization is required, it must degrade its time and storage consuming further. While the RF has a completely regular shape, circle or triangular, which can be easily represented, recorded and operated by a symbolic or algebraic means.

4.3 Run-time comparison

In order to compare the efficiencies of our nCRF-based Delaunary triangulation algorithm and Superpixel algorithm, we randomly selected 100 pictures with size of 481×321 pixels from Berkley Image Database. The test computer is with Intel Pentium Dual CPU E2200, 2.20GHz, 2G RAM. Figure 9 shows that our speed of producing small patches is much faster than Superpixel algorithm.

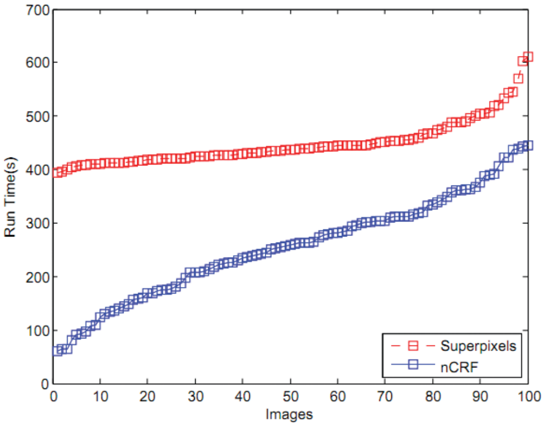


Figure 9. Speed comparison between nCRF-based algorithm and SuperPixel algorithm

5 5 Representing the parts of an object through polygons

5.1 Combining delaunay triangles into a polygon

In Figure 8(d), Delaunay triangulations provide a good start for further processing. So many small triangles can be regarded as components of an object. We can combine them into some polygons, and use these polygons as modules to construct an object. Because polygons are good at topographical and geometrical invariance, so using them to represent object offers the most stability in defining prototype.

During the learning period, we can provide some typical samples, such as clean cows without background disturbance, to computer, and ask it to form a prototype for this concept. Once cow images were uploaded to aforementioned nCRF-based neural computational system, we can obtain so many small triangles which covered a sample completely. Thus the first step of prototype-learning is to combine them appropriately. What we hope is that those combinations can reflect the structural characteristics possessed by a type of objects. A direct solution of combining triangles is to apply polygon, which

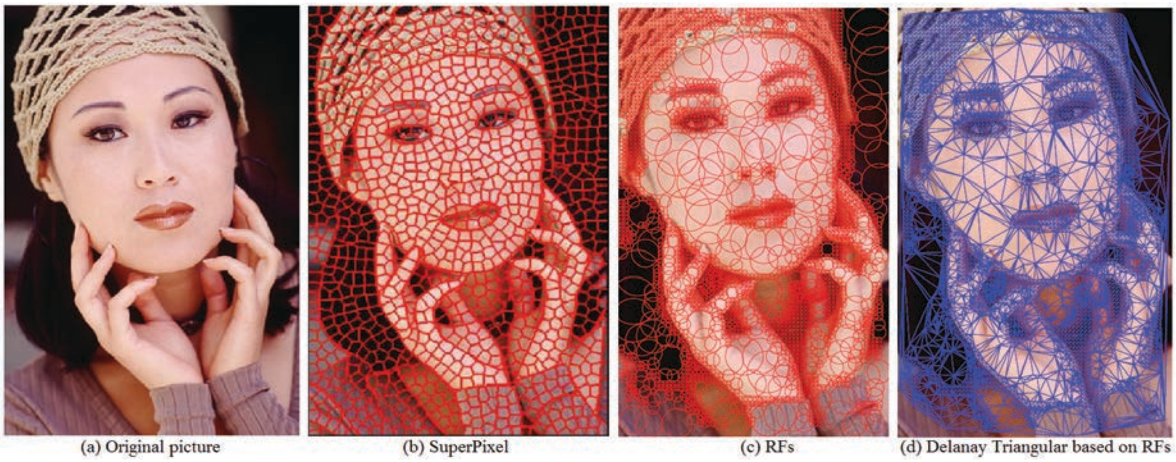


Figure 8. The partition results on an Example from Berkeley Image Database. (a) is original picture. (b) is superpixel result. (c) is RFs coverage on the original image and (d) is the result by Delaunay Triangulations.

is flexible enough to integrate triangles, and polygon's representation in analytic geometry is simple and compact.

In order to obtain a larger polygon from the result of nCRF algorithm, we design a clockwise spiral coordinate to order those triangles in 2D space. Then we can expand polygon by appending triangle one-by-one. Figure 10 demonstrates this expanding process and an algorithm of producing a polygon through binding some neighboring triangles. Here the principle of searching triangles is keeping the search closest to the periphery of growing-up polygon.

The nCRF-based mechanism provides a basic infrastructure to enable and facilitate patch-grained manipulations on geometrical level greatly. Searching algorithm perhaps is inefficient in pixel-grained space, but it is feasible in block-grained space.

At the top of Figure 11, there are at lease hundreds of small triangles, we want them to be allocated to different enlarged polygons, and at same time we hope these polygons happen to be the structural parts of an object. Tab. 1 is an algorithm to draw boundaries of polygons on the Delaunay triangulations. The down of Figure 11 is one of results of this algorithm. And Figure 12 includes two outputs after executing polygon-production algorithm.

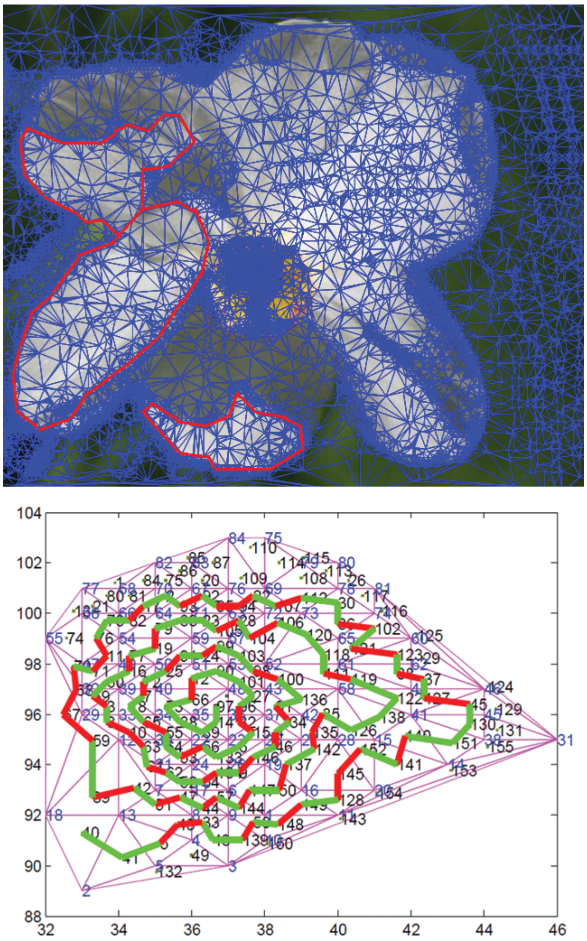
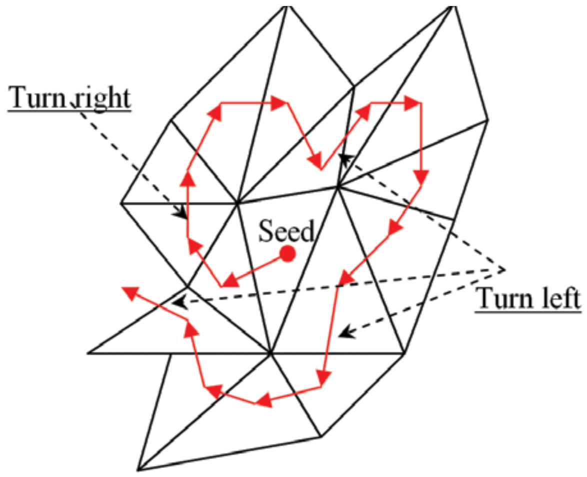


Figure 11. A Delaunay triangulations of a flower has many small triangles needing to be fitted by some polygons



The red arrows give directions of expanding.

Algorithm of forming a polygon along a clockwise spiral direction:
1. Clockwise ordering three vertexes of all triangles;
2. Randomly select a triangle as seed, and append it into polygon;
3. Randomly select an edge of seed, and expanding polygon to that triangle sharing this edge;
4. If both of right* and left triangles of the newly expanded triangle had been expanded, or there is no new triangle, then stop this algorithm and output polygon;
5. If the right neighbor triangle of this newly appended triangle has not been expanded yet, then expand it in next iteration, else expand the left one in next iteration;
6. Continues until the iteration limit is reached.

* The right side of an entrance of a triangle is the edge that decided by the upper vertex of being crossed edge and the third vertex. The order of vertexes was defined in advance in step 1.

Figure 10. Speed comparison between nCRF-based algorithm and SuperPixel algorithm

Table 1. Algorithm for producing multiple polygons on Delaunay triangulations

```
   Input data: int Small_Tri[id][x1][y1][x2][y2][x3][y3]; //All Delaunay triangles with their vertexes coordinates.

   Output data: List Polygon[id]; // A list dimension recording all polygons and their children triangles.

   Temp data: Boolean Neighboring[id][id]; // A matrix of all neighborhoods between triangles.

1. If the array of Small_Tri[id][x1][y1][x2][y2][x3][y3] is EMPTY Then End;

2. Clustering all vertexes by K-means algorithm, store all class center into center[K];

3. index=0;

4. For each center[k] do

   Finding Small_Tri[k] [x1][y1][x2][y2][x3][y3] whose center of gravity is closest to center[k];

   Setting Small_Tri[k] [x1][y1][x2][y2][x3][y3] as the seed of growing Polygon[index];

   Calling "Algorithm of forming a polygon along a clockwise spiral direction"; // Here Neighboring[id][id] are needed.

   According to the order of triangles being expanded, storing them into a list named Polygon[index];
   index=index+1;

5. For each Polygon[i]

   For each element in Polygon[i]

   Marking Small_Tri[element][x1][y1][x2][y2][x3][y3] by BEEN-DELETED;

6. For each Small_Tri[id][x1][y1][x2][y2][x3][y3]

   If all three neighboring triangles of Small_Tri[id][x1][y1][x2][y2][x3][y3] were marked by BEEN-DELETED Then Deleting all three vertexes of Small_Tri[id][x1][y1][x2][y2][x3][y3];

7. Iterate since 1.
```

Once this algorithm ends, some enlarged polygons come into being. The triangle is the simplest shape with edges, and this causes that searching from one triangle to its neighbors has only two possible choices. And while K is limited, K-means clustering algorithm can guarantee the seeds will not diverge far away from the local centers of topological components of an object. And what we focus on polygon, this further reduces the occasional disturbance of object's texture.

5.2 Matching between polygons

The core of inductive learning of prototype is to find the common components from different instances of a type of object. This is a hard work to do on pixel-level, or on small triangle-level. But it can be done easily on polygon-level. Once we divide an object into several larger polygons, then we can use famous shape context algorithm [41] to decide which two polygons corresponds well. Figure 13 shows two polygon-groups produced by previous algorithms, and they represent two cows respectively. Between two contours, a dozen of corresponding points can be established by a shape context algorithm. Usually, deciding initial points to

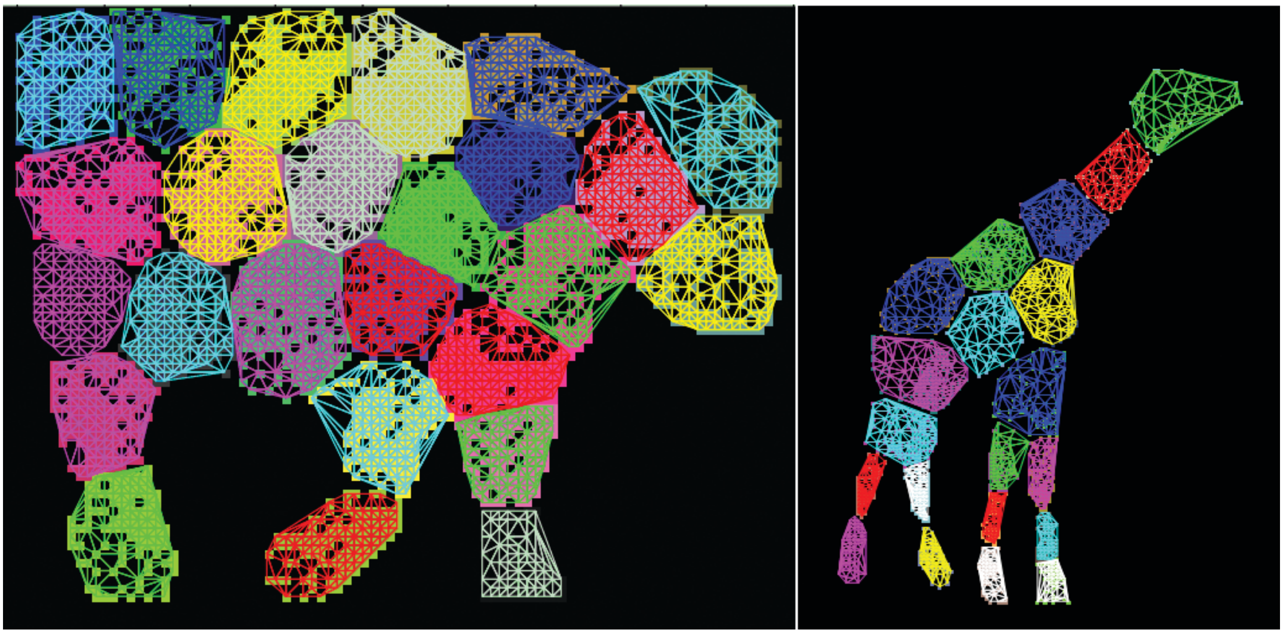


Figure 12. Some polygons have been produced on the base of many small triangles

match is a problem to shape context algorithm, because some points not belonging to contour might been selected with high possibility. This will reduce the accuracy of matching. But in a current situation, all polygons are represented algebraically, because they are defined by vertexes and all edges are vectors. Therefore on these edges it is easy to choose initial points. This algorithm is insensitive to position, size and pose, so we can concentrate on shape similarity.

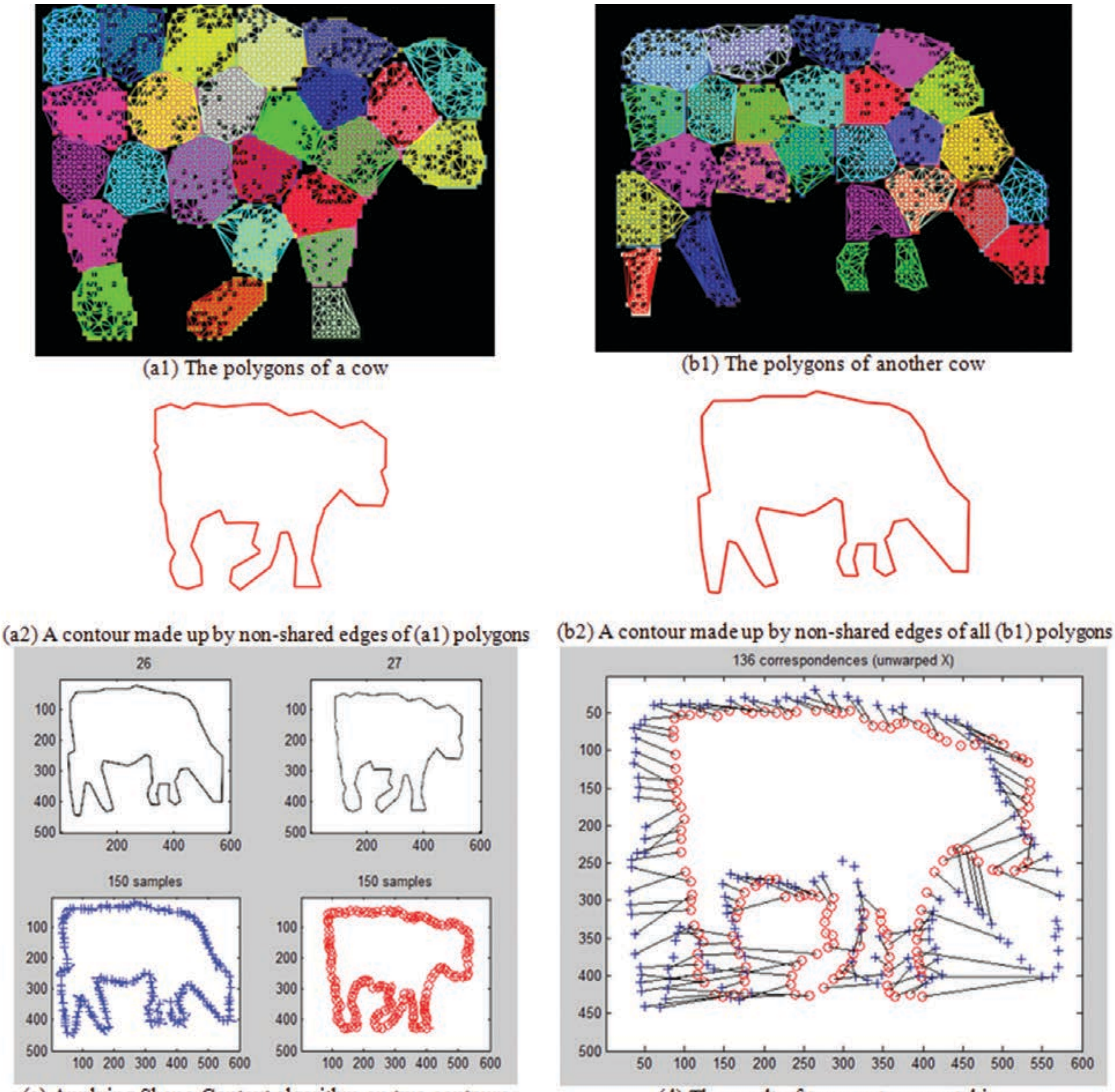
Another strategy that helps us to establish correspondences between polygons is minimum spanning tree algorithm. Firstly, all polygons were named. Secondly, according to their neighboring relationships, a connected graph was formed for each set of polygons in an image. Thirdly, selecting a node as tree root and starting Prim minimum spanning tree algorithm, after that we got a spanning tree growing up from a selected polygon. Figure 14 is the result of this process. We show there a two cow pictures which were made up by many named polygons. We selected three pairs of spanning trees taking root in A15-C15, A8-C7 and A23-C20 respectively. It is obvious that every pairs are very similar, because they really reflect the topological structure of an object. Therefore, applying this strategy we can greatly improve the matching accuracy between different cases.

After this step, we find out which polygons are corresponding in different samples, and they perhaps are the similar parts of a kind of object. Once correspondence relationships among parts or components are established, then an inductive learning procedure can be used to discover those inherent and persistent relationships. Thus, a prototype or some kind of formal semantic description of this type of an object can be defined by these relationships.

6 Representing a prototype by root-tree

6.1 A multi-layer concept-defining tree

In the previous Section, an image representation schema was developed. Basing on it, we can use the combination of multiple polygons to represent an object, and Figure 15 is an example for representing COW. Firstly, several cases of cow, with different appearances, had been watched, and each of them is an instance of the concept "cow". Secondly, nCRF-based algorithm was applied, and each instance was partitioned by polygons, and topological correlations of these polygons are important. Thirdly, a AND-OR tree was built to record combinational relations at polygon level and at component level. In Figure 15 from root to leaves, they respectively describe: a cow might have multiple instances, and every instance can be divided into several components, and every component includes



(a1) The polygons of a cow

(b1) The polygons of another cow

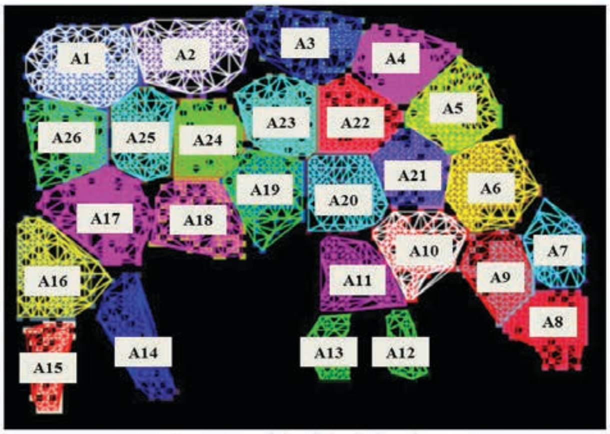
(a2) A contour made up by non-shared edges of (a1) polygons

(b2) A contour made up by non-shared edges of all (b1) polygons

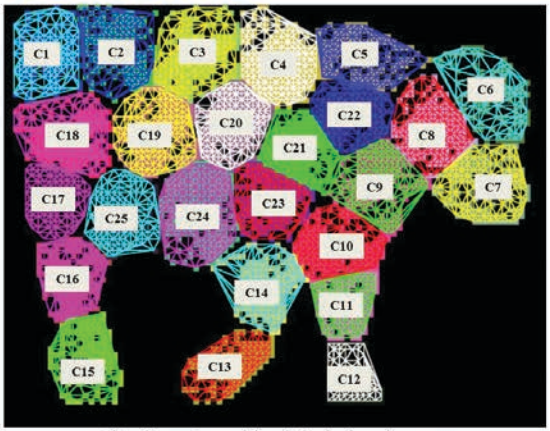
(c) Applying Shape-Context algorithm on two contours

(d) The result of two contours matching

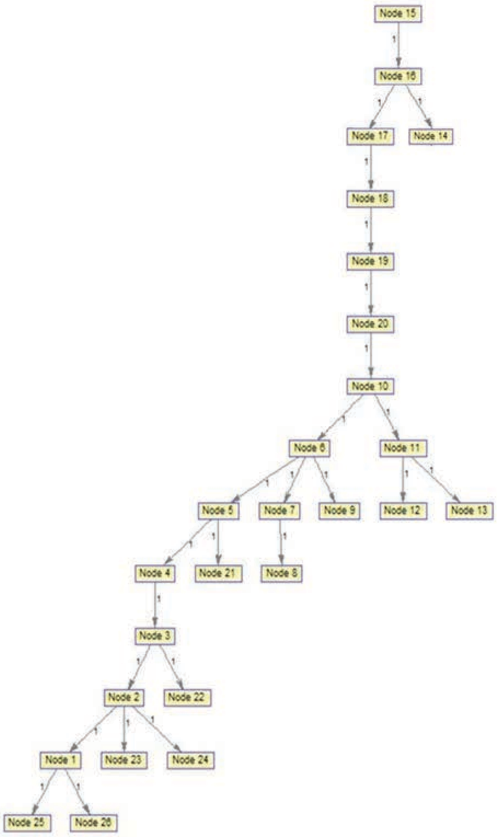
Figure 13. Shape context algorithm can be applied to establish the correspondence between different polygons



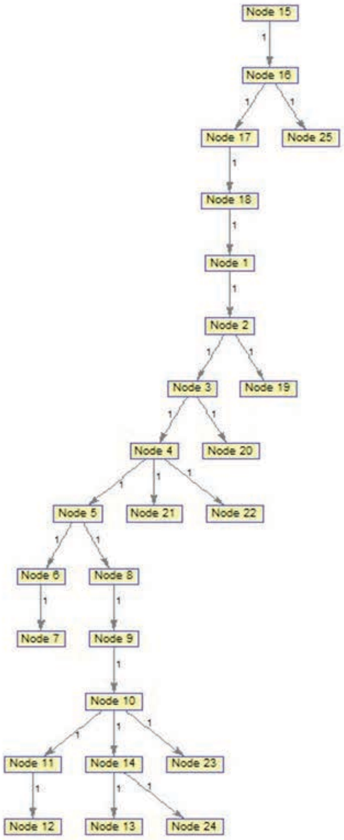
(a) Cow-1 and its labeled polygons



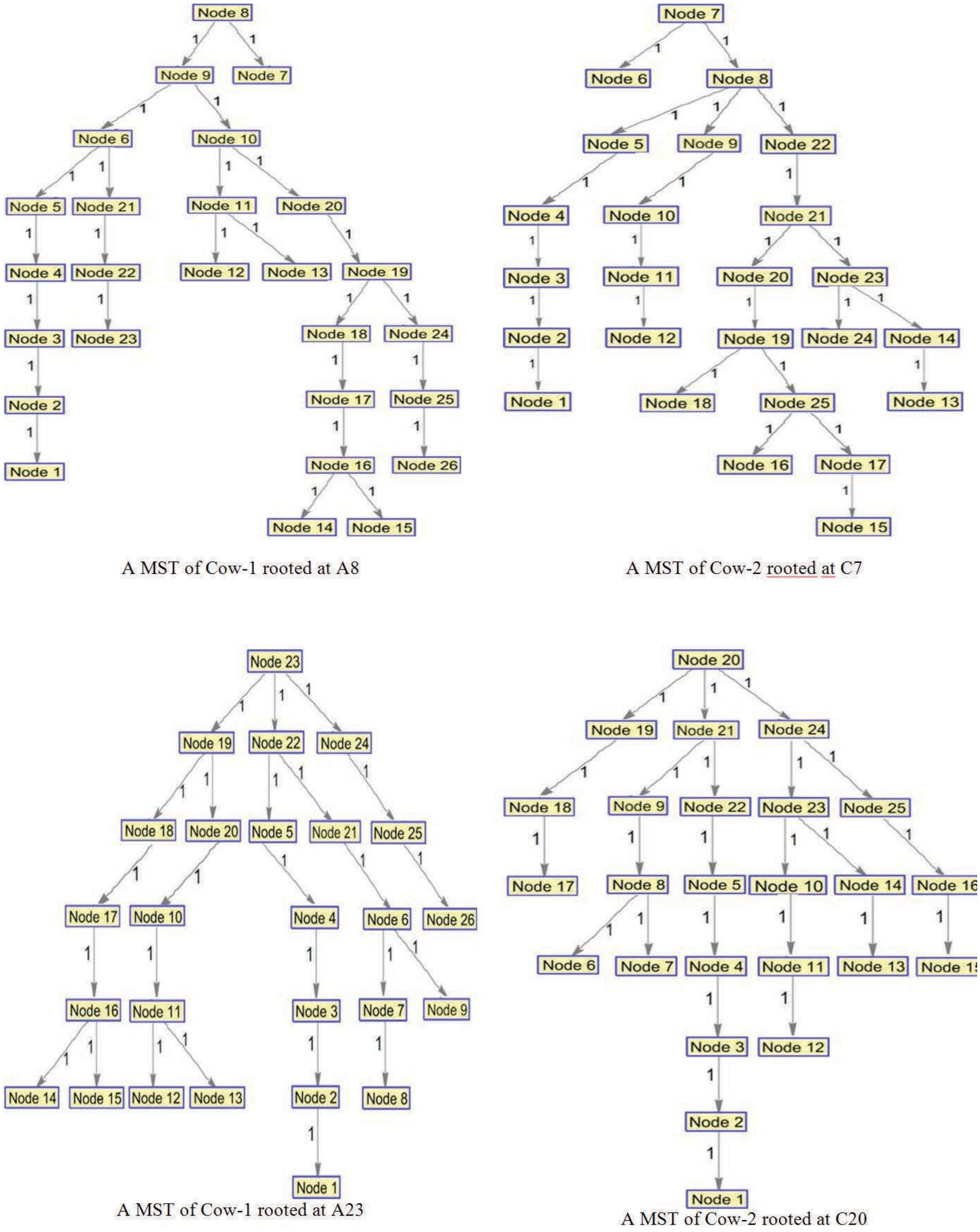
(b) Cow-2 and its labeled polygons



A MST of Cow-1 rooted at A15



A MST of Cow-2 rooted at C15



A MST of Cow-1 rooted at A8

A MST of Cow-2 rooted at C7

A MST of Cow-1 rooted at A23

A MST of Cow-2 rooted at C20

Figure 14. Spanning tree of connected graph can help establishing correspondence between polygons multiple blocks, and every block is made up of several polygons, and every polygon has its shape, and each shape can be defined by its vertexes, and any vertex can be obtained by several neighboring RFs. Using a tree-like data structure has been proved to be a practicable choice [42] to organize regions. All these calculations can be executed by algebraic equations of circle, and searching strategy is also workable [43].

The data at the bottom of Figure 13 are some vertexes of polygons, and all coordinates can be relocated according to a new datum point. Each row defines a shape, and no shape here is required to be absolutely fixed. Further more, these shapes can be zoomed in or out, or be rotated easily because they are defined by vectors. Using these tables we can rebuild several prototypes of a concept, and when a new sample occurring, we can compare it with those prototypes and identify which class the new sample belongs to. The concept tree is not absolute too, because the occurrences of Cow are different one to another. Fuzzy inference or probabilistic inference, such as Bayesian reasoning [44], or shape context algorithm [48] is a good tool to use this representation.

6.2 Direct manipulation that done to low-level pixels can be derived from this multi-layer tree

Now let's go back to the rules we formalized in Section 1. We obtain a concept tree about cow, and at the root of this tree it is symbolic, and at the leaf ends it contains many pixel-concerned coordinates. Moving from root to leaves, the concept definition turns gradually from abstract label to detailed object. This provides us an opportunity to join high level symbols and low-level BMP operations (such as searching pixel) together. Now let's define some production rules to show this practical method again. For explaining what is a cow:

```
IF Is_Cow(x) THEN Has_leg1(x) ∧ Has_leg2(x) ∧
Has_leg3(x) ∧ Has_leg4(x) ∧. . . ∧ Has_body(x);
IF Has_leg1(x) THEN Contain_area_like(x, Block1);
IF Verify(Block1) THEN Search_pixel_in(Set(Block1));
```

For deciding which pixels combine a cow:

```
IF RF_distribution_similar(x, Block5) THEN
Is_leg3(x);
IF Verify(RF_distribution_similar(y)) THEN
Decide_a_region_defined_by(y);
IF Verify(Similar(x, y)) THEN
Search_pixel_in_x_according_to_y(x, y);
```

The predicates like *Search_pixel_in( )* and *Contain_area_like( )* are totally operable at pixel level. Up to now, we show a feasible method, basing on an elaborate multi-level representation, which can ground semantic and apply them in image understanding.

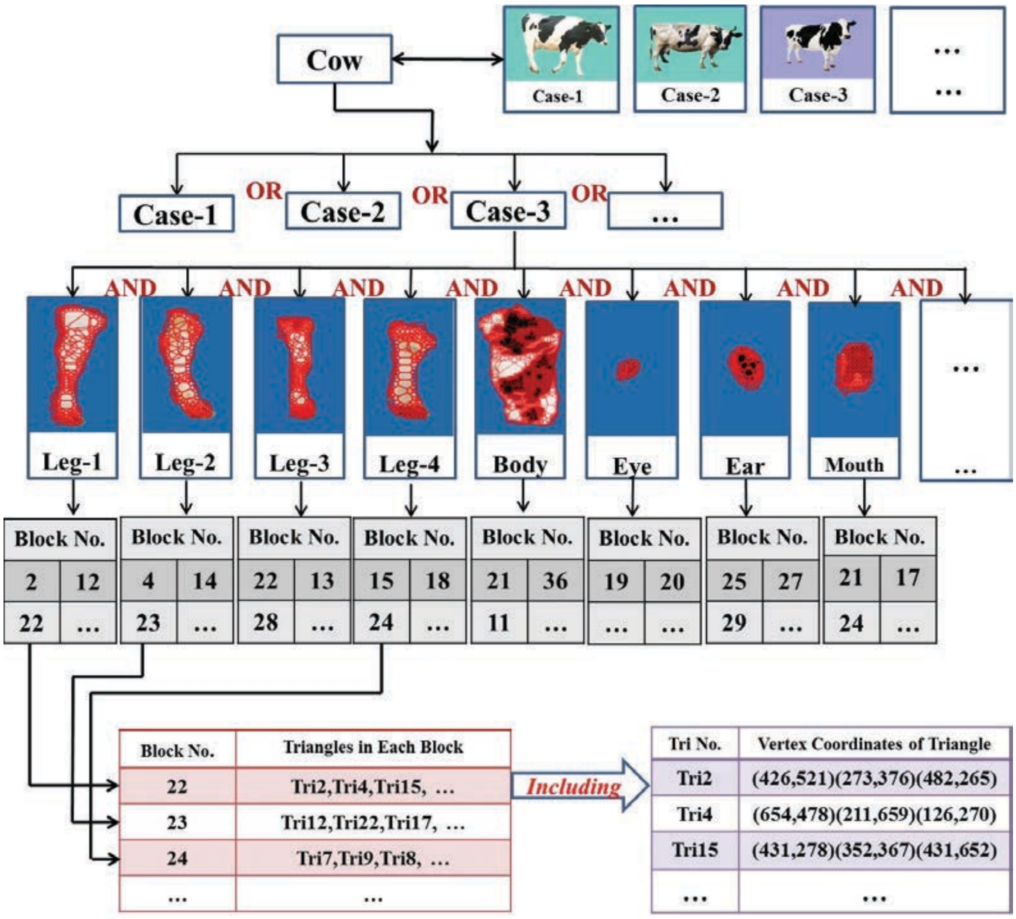
6.3 Why this tree can work?

Now a new problem arose. We know that tree, as a kind of data structure, is not new, and it was used extensively. For example, some knowledge for animals or plants classification is defined by tree, and some searching spaces are also defined by tree. What, then, makes tree work in grounding semantic this time? The most important reason is that GC-array and their RFs provide several gradual or transitional representation layers to ease the span between symbols and pixels. Another reason is that no information, whatever level or abstractness it is, is neglected, and on the contrary a realistic unit is assigned to represent it. That is to say we prevent our model from the famous and conventional hypothesis of discontinuity in low-order structures.

In Cognitive Psychology, cognitive modeling asks three key steps: (1) the stimulus must be translated into an internal representation, (2) the representation is manipulated by cognitive processes to derive new internal representations, and (3) these are in turn retranslated back into action [45]. Our methods of applying multi-layer representation and keeping intermediate links go along the same way, so a dense definition of semantic can be reached.

7 A semantic-grounding neural infrastructure

Perhaps we may suspect how so many rules can work efficiently. An alternative implementation way is doing this by a neural network. We know that production rules, including fuzzy rules and uncertain rules, can be realized equivalently by a neural network.



Leg-1 Block No.		Leg-2 Block No.		Leg-3 Block No.		Leg-4 Block No.		Body Block No.		Eye Block No.		Ear Block No.		Mouth Block No.	
2	12	4	14	22	13	15	18	21	36	19	20	25	27	21	17
22	...	23	...	28	...	24	...	11	...	...	...	29	...	24	...

Block No.	Triangles in Each Block
22	Tri2,Tri4,Tri15, ...
23	Tri12,Tri22,Tri17, ...
24	Tri7,Tri9,Tri8, ...
...	...

Tri No.	Vertex Coordinates of Triangle
Tri2	(426,521)(273,376)(482,265)
Tri4	(654,478)(211,659)(126,270)
Tri15	(431,278)(352,367)(431,652)
...	...

An important motivation of this paper is to construct an infrastructure for semantic-grounding. We think GC array provide a powerful base for this goal. Figure 16 is a neural structure [46] which can achieve semantic representation. Layer 2 is GC array, and each of them has a nCRF on photoreceptor layer. These two layers had described in fore-sections. And in layer 3, we design some feature recording units to produce semantic. We divided units in layer 2 and 3 into many small groups (three examples were drawn in each layer), and let them match one-by-one. The famous Bidirectional Associative Memory (BAM) algorithm was applied between each pair of groups. The BAM is unconditionally stable, and it is a heteroassociative network and indeed capable of error correction. Then two groups in a pair can feed each other upwards or downwards. The number of stable states in a BAM network is limited, but there are so many pairs, and the combinational number of states of different BAM networks is very huge. If we define units in layer 3 by some symbols and define units in layer 2 by image features, then layer 2 acts as explainer to ground semantic. When some more complicated network is built in layer 3, then a more flexible representation is possible [47]. This network fulfills the decomposition and integration of semantics with a fine span.

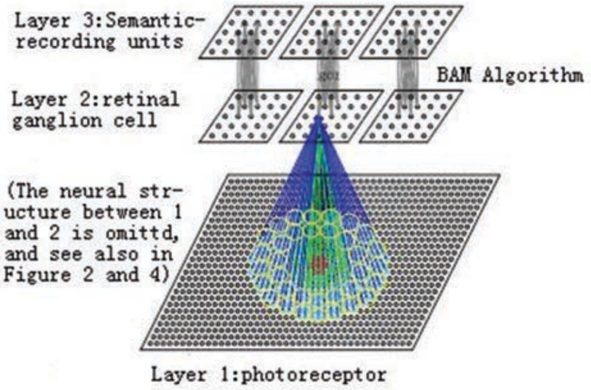

Layer 3:Semantic-recording units
BAM Algorithm
Layer 2:retinal ganglion cell
(The neural structure between 1 and 2 is omittd, and see also in Figure 2 and 4)
Layer 1:photoreceptor

Figure 16. A Multi-layer Structure for Semantic-grounding

8 Conclusion

When eye receiving an image as input, the retina in it will decide which information is significant and needs to be transmitted to central brain. A good heuristic rule is retaining variance and discarding redundancy. This task turns to be very difficult when scene keeps changing, and processing should be done in real time and the result should be in accordance with diverse upcoming tasks. This means a general and no task-specific schema should be sought. In this paper, we simulated a biological mechanism, using many receptive fields, for image representation, and based on it several important image processing tasks, such as segmentation and integration can be improved through some predefined production rules. Once a neuron-bounded representation is formed, the semantic-grounding turns to be practicable.

Semantic is crucial for computer vision (CV) and natural language understanding (NLU). In CV, possessing semantic means a program knows what a pile of pixels is; and in NLU, means a program knows how to apply rich relationships between concepts flexibly. A hierarchical structure can provide dense, continuous representations and rich linkages between them, from high-level concepts to low-level instances. This will greatly benefit knowledge applying in CV and NLU.

Acknowledgment

This work was supported by the 973 Program (Project No. 2010CB327900), the NSFC project (Project No. 61375122).

References

[1] G. Carneiro, A.B. Chan, P.J. Moreno, and placeN. Vasconcelos, Supervised learning of semantic classes for image annotation and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence (2007) 394-410.

[2] J.H. Su, C.L. Chou, C.Y. Lin, and V.S. Tseng, Effective Semantic Annotation by Image-to-Concept Distribution Model. Multimedia, IEEE Transactions on 13 (2011) 530-538.

[3] J. Fan, Y. Gao, H. Luo, and R. Jain, Mining multilevel image semantics via hierarchical classification. Multimedia, IEEE Transactions on 10 (2008) 167-187.

[4] J.R.R. Uijlings, A.W.M. Smeulders, and R.J.H. Scha, Real-time visual concept classification. Multimedia, IEEE Transactions on 12 (2010) 665-681.

[5] Y.G. Jiang, J. Yang, C.W. Ngo, and A.G. Hauptmann, Representations of keypoint-based semantic concept detection: A comprehensive study. Multimedia, IEEE Transactions on 12 (2010) 42-53.

[6] S. Edelman, Computational theories of object recognition. Trends in Cognitive Sciences 1 (1997) 296-304.

[7] B.J. Stankiewicz, and J.E. Hummel, MetriCat: A representation for basic and subordinate-level classification, Lawrence Erlbaum, 1996, pp. 254.

[8] I. Biederman, Recognition-by-components: A theory of human image understanding. Psychological review 94 (1987) 115.

[9] E.I. Le Dong, A Topology Preserving Approach for Image Classification. (2007).

[10] K. Engel, and K.D. Toennies, Hierarchical vibrations for part-based recognition of complex objects. Pattern Recognition 43 (2010) 2681-2691.

[11] F. Chen, H. Yu, and R. Hu, Simultaneous variational image segmentation and object recognition via shape sparse representation, IEEE, pp. 3057-3060.

[12] H. Liu, W. Liu, and L.J. Latecki, Convex shape decomposition. (2010).

[13] Y. Li, and J. Feng, Sparse representation shape model, IEEE, pp. 2733-2736.

[14] J. Vogel, and B. Schiele, Semantic modeling of natural scenes for content-based image retrieval. International Journal of Computer Vision 72 (2007) 133-157.

[15] B. Peng, L. Zhang, D. Zhang, and J. Yang, Image segmentation by iterated region merging with localized graph cuts. Pattern Recognition 44 (10-11), 25272538

[16] C. Huang, Q. Liu, and S. Yu, Regions of interest extraction from color image based on visual saliency. The Journal of Supercomputing1-14.

[17] K.S. Kim, M.J. Lee, J.W. Lee, T.W. Oh, and H.Y. Lee, Region-based tampering detection and recovery using homogeneity analysis in quality-sensitive imaging. Computer Vision and Image Understanding 115 (2011) 1308-1323.

[18] R. Farrahi Moghaddam, and M. Cheriet, Beyond pixels and regions: A non local patch means (NLPM) method for content-level restoration, enhancement, and reconstruction of degraded document images. Pattern Recognition 44 (2), 363-374

[19] C. Xiao, M. Liu, Y. Nie, and Z. Dong, Fast Exact Nearest Patch Matching for Patch-based Image Editing and Processing. IEEE Transactions on Visualization and Computer Graphics 17 (2011) 1122-1134.

[20] Z. Xu, and J. Sun, Image inpainting by patch propagation using patch sparsity. Image Processing, IEEE Transactions on 19 (2010) 1153-1165.

[21] C.W. Fang, and J.J.J. Lien, Rapid image completion system using multiresolution patch-based directional and nondirectional approaches. Image Processing, IEEE Transactions on 18 (2009) 2769-2779.

[22] H. Shvaytser, Learnable and nonlearnable visual concepts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 12 (1990) 459-466.

[23] P. Mylonas, E. Spyrou, Y. Avrithis, and S. Kollias, Using visual context and region semantics for high-level concept detection. Multimedia, IEEE Transactions on 11 (2009) 229-243.

[24] J.W. Hsieh, and W.E.L. Grimson, Spatial template extraction for image retrieval by region matching. Image Processing, IEEE Transactions on 12 (2003) 1404-1415.

[25] D. Marr, and H.K. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society of London. Series B. Biological Sciences 200 (1978) 269.

[26] J.T. Devlin, P.M. Matthews, and M.F.S. Rushworth, Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. Journal of Cognitive Neuroscience 15 (2003) 71-84.

[27] J. Allman, F. Miezin, and E. McGuinness, Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annual Review of Neuroscience 8 (1985) 407-430.

[28] D.J. Heeger, Normalization of cell responses in cat striate cortex. Visual Neuroscience 9 (1992) 181-197.

[29] X.L.I. YANG, F. GAO, and S.M. WU, Modulation of horizontal cell function by GABAA and GABAC receptors in dark-and light-adapted tiger salamander retina. Visual neuroscience 16 (1999) 967-979.

[30] Q.F.L. Chaoyi, Mathematical simulation of disinhibitory properties of concentric receptive field [J]. Acta Biophysica Sinica 11 (1995) 214-220.

[31] D. Fitzpatrick, Seeing beyond the receptive field in primary visual cortex. Current Opinion in Neurobiology 10 (2000) 438-443.

[32] A.M. Sillito, K.L. Grieve, H.E. Jones, J. Cudeiro, and J. Davis, Visual cortical mechanisms detecting focal orientation discontinuities. Nature 378 (1995) 492-496.

[33] H.R. Wilson, and R. Humanski, Spatial frequency adaptation and contrast gain control. Vision Research 33 (1993) 1133-1149.

[34] G. Krieger, and C. Zetzsche, Nonlinear image operators for the evaluation of local intrinsic dimensionality. IEEE Transactions On Image Processing 5 (1996), 1026-1042.

[35] Hui Wei, X-M Wang, L.L.Lai, A Compact Image Representation Model Based on Both nCRF and Reverse Control Mechanisms. IEEE Transactions on Neural Network and Learning Systems, Vol.23 (1), (2012), 150-162.

[36] Hui Wei, X-M Wang, A Neural Circuit Model for nCRF's Dynamic Adjustment and its Application on Image Representation, 2011 International Joint Conference on Neural Networks, San Jose, California , July 31 - August 5, (2011), 421-429.

[37] J. Fan, Y. Gao, H. Luo, and R. Jain, Mining multilevel image semantics via hierarchical classification. Multimedia, IEEE Transactions on 10 (2008) 167-187.

[38] X.F. Ren, and J. Malik, Learning a classification model for segmentation. Ninth IEEE International Conference on Computer Vision, Vols I and II, Proceedings (2003) 10-17.

[39] A. Levinshtein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, and K. Siddiqi, TurboPixels: Fast Superpixels Using Geometric Flows. IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 2290-2297.

[40] G. Mori, Guiding model search using segmentation, Computer Vision, 2005. ICCV 2005, (2005), pp. 1417-1423 Vol. 2.

[41] S. Belongie, J. Malik, and J. Puzicha, Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence (2002) 509-522.

[42] P. Arbelez, M. Maire, C. Fowlkes, and J. Malik, Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence (2010) 898-916.

[43] S. GAP, Visual-Concept Search Solved?. Computer 43 (2010) 76-78

[44] J. Luo, A.E. Savakis, and A. Singhal, A Bayesian network-based framework for semantic image understanding. Pattern Recognition 38 (2005) 919-934.

[45] S.J. Russell, P. Norvig, J.F. Canny, J.M. Malik, and D.D. Edwards, Artificial intelligence: a modern approach, Prentice hall Englewood Cliffs, NJ, (1995).

[46] X. Wang, and H. Wei, An Integration Model Based on Non-classical Receptive Fields, Springer, (2009), pp. 451-459.

[47] H. Wei, A Homogenous Associative Memory Neural Network Based on Structure Learning and Iterative Self-Mapping. Journal of Software Vol.13 (3) (2002) 438-446.

[48] Serge Belongie, Jitendra Malik, Jan Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.24, (2002), 509-522

Hui Wei received the Ph.D. degree at the department of computer science at Beihang University in 1998. From 1998 to 2000, he was a postdoctoral fellow at the Department of Computer Science and the Institute of Artificial Intelligence at Zhejiang University. Since November 2000, he has joined the Department of Computer Science and Technology at Fudan University. His research interests include artificial intelligence and cognitive science.