

Annual Report for Period:07/2009 - 06/2010**Submitted on:** 07/23/2010**Principal Investigator:** Wu, Yingnian .**Award ID:** 0707055**Organization:** U of Cal Los Angeles**Submitted By:**

Wu, Yingnian - Principal Investigator

Title:

From Information Scaling to Regimes of Statistical Models of Natural Image Patterns

Project Participants**Senior Personnel****Name:** Wu, Yingnian**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Ying Nian Wu worked on various projects on image modeling.

Name: Zhu, Song-Chun**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Song-Chun Zhu worked on statistical modeling and computing in vision.

Post-doc**Graduate Student****Name:** Si, Zhangzhang**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Zhangzhang Si worked under the supervision of PI and co-PI on extending the active basis model. He worked on learning mixed templates, shape script model and active plates model.

Name: Chen, Nicole**Worked for more than 160 Hours:** Yes**Contribution to Project:**

Nicole Chen worked under the supervision of PI on active basis model.

Undergraduate Student**Technician, Programmer****Other Participant****Research Experience for Undergraduates****Organizational Partners****Other Collaborators or Contacts**

Activities and Findings

Research and Education Activities:

The goal of our research is to develop statistical models and model-based algorithms for representing and recognizing visual patterns in natural images.

This is the third year of the proposed research. The following are our activities.

- (1) Active basis model. We continue our work on active basis model for object recognition. We have made the following progresses. (a) We further study the EM algorithm for fitting mixture of active basis models and show that it is possible to learn multiple templates from training images that contain objects of different categories or views or poses. (b) We extend the active basis model to represent three-dimensional templates.
- (2) Layered hierarchical extensions of active basis model. (a) We develop a shape script model, where the template is a composition of a small number of deformable elementary geometric shapes, such as line segments, angles and ellipsoids. These elementary shapes are allowed to change their geometric attributes, and each elementary shape is represented by an active basis model. (b) We develop an active plates model, which is a composition of partial templates (or plates) that are represented by active basis models.
- (3) Reproducibility page. We continue to develop the reproducibility pages for our papers. The following page <http://www.stat.ucla.edu/~ywu/AB/ActiveBasisMarkII.html> is the major one. We have further polished the matlab/C code, and we have added the source code and data sets for five more sets of experiments.
- (4) Supervision of graduate students. Nicole Chen, Zhangzhang Si and Tess Nesbitt received support from this grant and have been working under the supervision of the PI (Wu) and co-PI (Zhu).

Findings: (See PDF version submitted by PI at the end of the report)

Please see the attached 7-page pdf file.

Training and Development:

- (1) The research in this project represents the state of art in statistical modeling, learning and computing in computer vision. It provides opportunities for students to learn modern statistics while working on an important scientific problem.
- (2) The work requires considerable amount of programming. It provides opportunities to train students in programming in matlab and C. All the source code has been posted on the reproducibility pages that we have constructed.
- (3) Students receive training in conducting critical review of the current literature, and in communicating and presenting their work.
- (4) Students receive training in reproducible research. We ask students to construct reproducibility pages for the papers they submit or publish, where all the figures, plots, tables and numbers reported in the papers should be accompanied by the data, code and readme files that reproduce them.

Outreach Activities:

- (1) The PI (Wu) served as an associate editor for JASA applications and case studies.
- (2) The PI (Wu) served as an associate editor for Statistica Sinica.
- (3) The PI (Wu) was invited to give an Introductory Overview Lecture on 'Statistical modeling of visual patterns' in JSM 2010. The lecture is

scheduled on August 4, 2010.

(4) The PI (Wu) was invited to give a seminar in the Department of Statistics, Texas A and M University, in April 2010.

(5) The PI (Wu) served as referee for JASA, NIPS, etc.

(6) The co-PI (Zhu) participated in organizing Sino-USA Summer School in Vision, Learning and Pattern Recognition 2010 (Perception, Motion and Events) to be held July 18 - 26, 2010 in Xi'an, China.

Journal Publications

Wu, Y. N., Si, Z., Gong, H., and Zhu, S. C., "Learning active basis model for object detection and recognition", International Journal of Computer Vision, p. , vol. , (2010). on-line first publication, 10.1007/s11263-009-0287-0

Zhu, S. C., Shi, K., and Si, Z., "Learning explicit and implicit visual manifolds by information projection", Pattern Recognition Letter, p. 667, vol. 31, (2010). Published, 10.1016/j.patrec.2009.07.020

Wu, T. and Zhu, S. C., "A numeric study of the bottom-up and top-down inference processes in and-or graphs", International Journal of Computer Vision, p. , vol. , (2010). on-line first publication, 10.1007/s11263-010-0339-5

Chen, R. B., Chu, C. H., Lai, T. Y. and Wu, Y. N., "Stochastic matching pursuit for Bayesian variable selection", Statistics and Computing, p. , vol. , (2009). on-line first publication, 10.1007/s11222-009-9165-4

Books or Other One-time Publications

Si, Z. and Wu, Y. N., "Wavelet, active basis, and shape script --- a tour in the sparse land", (2010). conference workshop proceedings, Published

Collection: ACM SIGMM International Conference on Multimedia Information Retrieval, Special session on Statistical Modeling and Learning for Multimedia

Bibliography: NA

Hu, W. Z. and Zhu, S. C., "Learning a probabilistic model mixing 3D and 2D primitives for view invariant object recognition", (2010). conference proceedings, Published

Collection: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)

Bibliography: NA

Web/Internet Site

URL(s):

<http://www.stat.ucla.edu/~ywu/AB/ActiveBasisMarkII.html>

Description:

This is a comprehensive reproducibility webpage for our papers on active basis model. We have included the source code and data for reproducing 15 sets of experiments.

Other Specific Products

Contributions

Contributions within Discipline:

Our work on this project during the past year is focused on the active basis model and its extensions.

Contribution to statistics: One may consider the active basis model as a linear regression model, where the regressors are indexed by attributes such as locations, orientations and scales. In the active basis model, the regressors are allowed to locally vary their attributes. This is similar to the random effects model, but the randomness comes from the variations in the attributes of the regressors, in addition to the coefficients of the regressors. The regressors are to be selected from a large dictionary, similar to the variable selection problem in modern linear regression. The distributions of the coefficients and the residuals are highly non-Gaussian. So this model is a very interesting generalization of the linear regression model.

We also work on layered hierarchical extensions of the active basis model. One example is the shape script model, where an image is represented by a small number of deformable elementary geometric shapes, such as line segments, angles and ellipsoids. These elementary shapes are allowed to vary their overall geometric attributes and they are represented by the active basis models so that they are locally deformable. This leads to a layered sparse coding scheme that goes beyond wavelet sparse coding.

Contribution to vision: The active basis model is among a very small number of statistical models in computer vision that are built on image intensities, or the so-called generative models. The generative models are crucial for image representation and image understanding, but it is very difficult to construct generative models which are simple but flexible, and which can generate reasonably realistic visual patterns. The active basis model and its extensions are important advances in statistical modeling in computer vision.

Contributions to Other Disciplines:

Contribution to neuroscience. The active basis model and its layered hierarchical extensions are biologically plausible and cortex-like. They shed light on the primary visual cortex and beyond. They explain the role of the complex cells in the primary visual cortex in terms of local maximum pooling and, more important, the retrieval of the arg-max elements in the top-down pass. The shape script model sheds light on the role of neurons beyond the primary visual cortex.

Contribution to wavelets and harmonic analysis. The shape script model can be an important next step beyond wavelet sparse coding. In wavelet sparse coding, an image is represented by a small number of wavelet elements selected from a large dictionary. The geometric attributes of the wavelet elements, such as their locations, orientations and scales, can be further represented by deformable elementary geometric shapes or shape motifs in the shape script model. In general, pursuing further sparse coding of the attributes of the current sparse coding elements leads to a recursive sparse coding scheme that results in cortex-like layered hierarchical model.

Contributions to Human Resource Development:**Contributions to Resources for Research and Education:****Contributions Beyond Science and Engineering:****Conference Proceedings****Special Requirements**

Special reporting requirements: None

Change in Objectives or Scope: None

Animal, Human Subjects, Biohazards: None

Categories for which nothing is reported:

Organizational Partners

Any Product

Contributions: To Any Human Resource Development

Contributions: To Any Resources for Research and Education

Contributions: To Any Beyond Science and Engineering

Any Conference

Describe the major findings resulting from these activities

During the past year, our work has been focused on the active basis model and its layered hierarchical extensions. We have made interesting progresses. We shall first remind the reviewer the basic idea of the active basis model. Then we shall report our major findings.

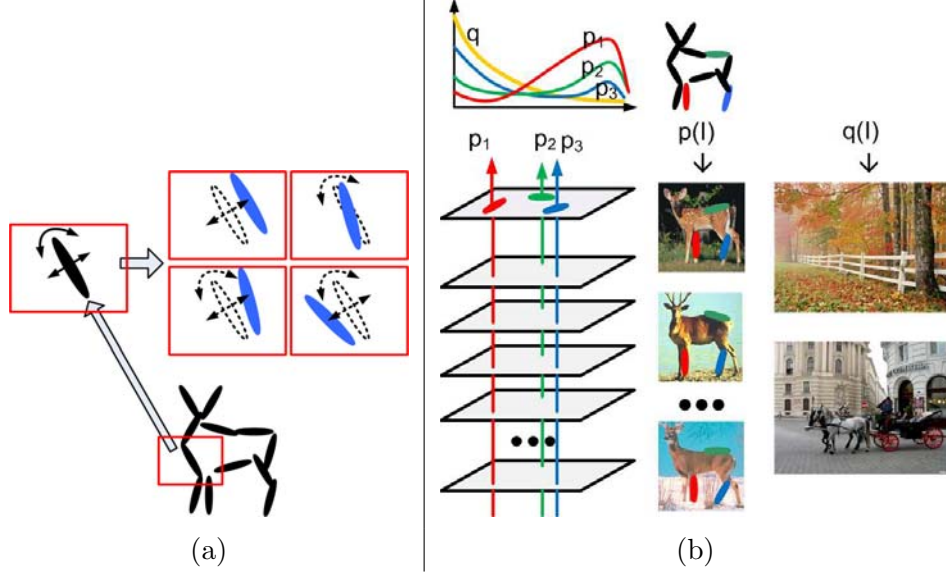


Figure 1: (a) An active basis template $\mathbf{B} = (B_{x_i, s, \alpha_i}, i = 1, \dots, n)$ of a deer, where each Gabor wavelet element B_{x_i, s, α_i} is illustrated by an elongated ellipsoid. An element B_{x_i, s, α_i} (black ellipsoid) can slightly shift its location and orientation and change to $B_{x_i + \Delta x_{m,i}, s, \alpha_i + \Delta \alpha_{m,i}}$ (blue ellipsoids) for coding image \mathbf{I}_m . (b) The elements of the active basis \mathbf{B} are shared by all the training images $\{\mathbf{I}_m, m = 1, \dots, M\}$ of deer, subject to local perturbations or activities $(\Delta x_{m,i}, \Delta \alpha_{m,i}, i = 1, \dots, n)$ that deform the active basis template \mathbf{B} . The elements are selected in the order of the Kullback-Leibler divergence between the foreground distribution p_i of the Gabor filter responses pooled from training images of deer, and the background distribution q pooled from the two natural images of rural and urban scenes.

The active basis model is based on the following linear regression model,

$$\mathbf{I}_m = \sum_{i=1}^n c_{m,i} B_{x_i + \Delta x_{m,i}, s, \alpha_i + \Delta \alpha_{m,i}} + U_m, \quad m = 1, \dots, M, \quad (1)$$

where $\{\mathbf{I}_m, m = 1, \dots, M\}$ is a set of training images, initially assumed to come from the same object category at the same scale and pose. Each \mathbf{I}_m is represented as a linear superposition of a set of basis elements that are perturbed versions of $(B_{x_i, s, \alpha_i}, i = 1, \dots, n)$, where each B_{x_i, s, α_i} is a localized, elongated and oriented Gabor wavelet element localized at location x_i , scale s , and orientation α_i . For each image \mathbf{I}_m , B_{x_i, s, α_i} is perturbed to $B_{x_i + \Delta x_{m,i}, s, \alpha_i + \Delta \alpha_{m,i}}$ to account for shape deformation.

$c_{m,i}$ is the coefficient of the perturbed basis element, and U_m is the unexplained residual image. The set of basis elements $\mathbf{B} = (B_{x_i,s,\alpha_i}, i = 1, \dots, n)$ are selected from a dictionary of Gabor wavelets. \mathbf{B} can be considered a deformable template.

For statistical modeling, we put non-Gaussian probability distributions on $c_{m,i}$ and U_m , and we put uniform distributions on the activities of the basis elements $(\Delta x_{m,i}, \Delta \alpha_{m,i}, i = 1, \dots, n)$, where the activities are assumed to be restricted within a local range. Figure (1) illustrates the active basis model and the scheme for learning the model by selecting the basis elements.

We can generalize the model to incorporate Gabor wavelets at multiple scales s . We can also fit the model to training images $\{\mathbf{I}_m\}$ where the objects may appear at different locations and scales in the training images.

1 Mixture of active basis models

The training images may be a mixture of different categories or poses, so we should learn multiple active basis models from the training images while separating them into different clusters. We can fit a mixture of active basis models by the EM algorithm, where the M-step learns different active basis templates for different clusters based on the E-step soft classification. This is an example of the so-called unsupervised learning.



Figure 2: Fitting mixture model by EM. In the templates, each selected Gabor wavelet element is depicted by a bar at the same location, orientation and scale as the element. Number of images: Cat-cattle-wolf-bear: 320. Horse: 188. Fashion: 57.

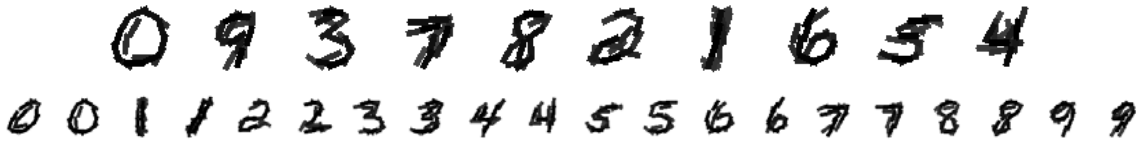


Figure 3: EM mixture. Top row is overall clustering: fitting a mixture model to 500 MNIST images, with the number of clusters set at 10. Bottom row is within digit clustering: fitting a mixture model to 100 or 200 MNIST images from each digit category. Number of clusters is set at 2.

We continue to work on the EM algorithm for mixture model, and our experiments suggest that it is possible to learn multiple active basis templates from training images in an unsupervised manner. Figures (2) and (3) display some examples of EM clustering, initialized by random clustering.

It is interesting to see that our method can separate the handwritten digits images of the 10 digits.

We are currently working on scaling up the experiments.

2 Comparing generative and discriminative approaches

The active basis model can be learned from roughly aligned images. Figure (4) displays the learned templates from various training sets.

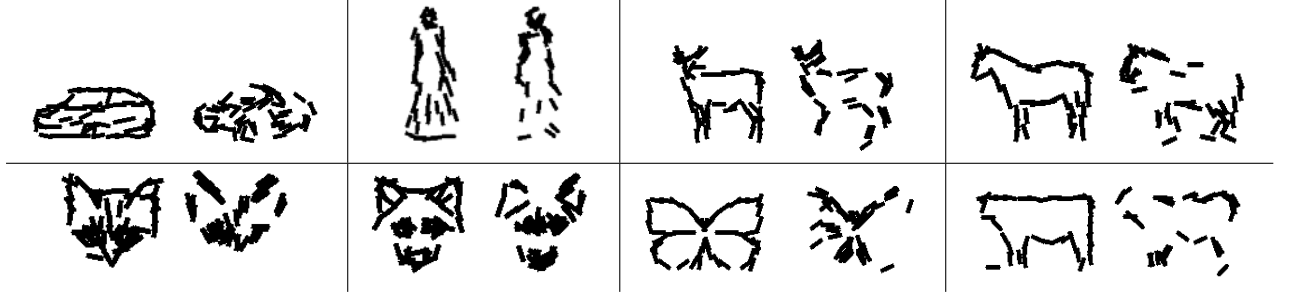


Figure 4: Active basis template (left) and adaboost template (right). Number of elements, number of positives, and number of negatives: Car: 60, 37, 1065. Fashion: 50, 15, 1147. Deer: 50, 9, 1138. Horse: 40, 280, 1511. Cat: 60, 89, 1493. Wolf: 60, 53, 1493. Butterfly: 50, 223, 1004. Cow: 40, 12, 1241.

One can also train the model discriminatively using the adaboost method with weak classifiers based on thresholding the local maxima of the filter responses, where the threshold is to be selected at each step of adaboost from a grid of 50 equally spaced values. Figure (4) displays the adaboost templates alongside the active basis templates. For each experiment, both templates are learned from the same positive training set with the same number of elements and under the same parameter setting.

The learning of active basis template does not require negative training images, except a one-dimensional marginal histogram pooled from two background images. It is therefore much faster than adaboost. The 1000+ negative image patches for training adaboost templates are randomly cropped from more than 200 large natural images at multiple resolutions. The adaboost learning is initialized from balanced weights, i.e., the total weights for positive images and negative images are both $1/2$.

In general, the active basis template is cleaner than the adaboost template. The active basis model is a generative model based on a linear additive structure. Adaboost is a discriminative method. In the adaboost method, once a basis element B_i is selected, the training examples are re-weighted to neutralize B_i . In contrast, in the active basis model, once a basis element B_i is selected, the residual images U_m is updated to $U_m \leftarrow U_m - c_{m,i}B_{x_i+\Delta x_{m,i},s,\alpha_i+\Delta\alpha_{m,i}}$, in order to neutralize B_i . We implemented an approximated version of this updating. In the generative learning of the active basis template, there is no re-weighting of the training examples, and therefore, there is no loss of effective training examples.

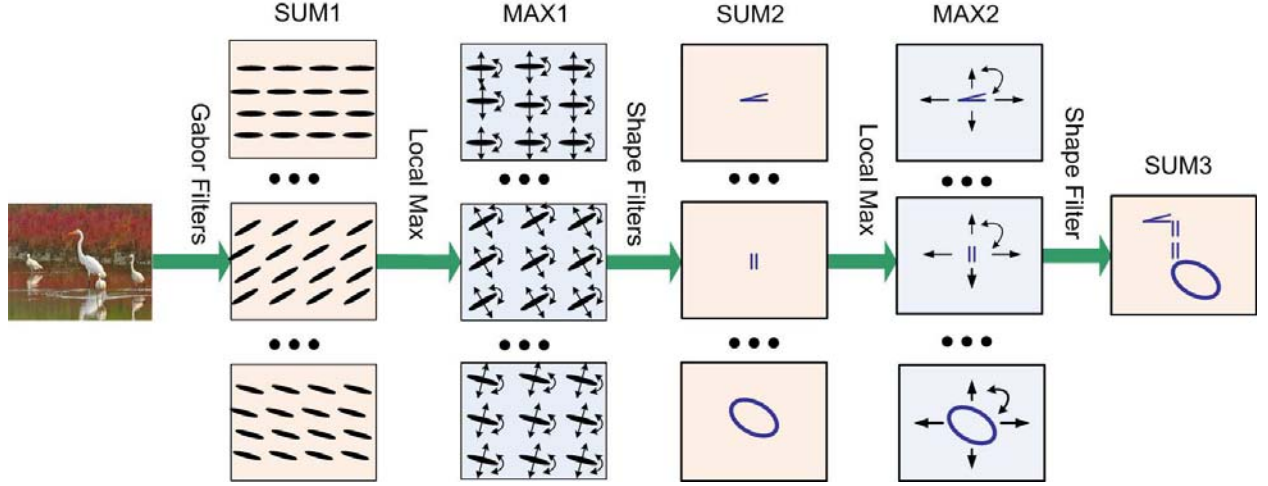


Figure 5: Recursive sum-max maps. The SUM2 maps score the matching of the shape motifs. The MAX2 maps are obtained by local maximum pooling of the SUM2 maps. The SUM3 map scores the matching of the shape script template.

We are currently working on a theoretical investigation of this issue. The active basis model is a very good example for revealing the difference between generative approach and discriminative approach, which is a fundamental problem in machine learning.

3 Shape script model

The wavelet representations take advantage of the fact that natural images or images of geometric shapes mostly contain edges at different scales. The question is: What is beyond these wavelet elements? In an analogy to language, if these elements are “letters,” then what are the “words” so that these “words” lead to even sparser representations?

We propose a shape script model as a highly sparse and symbolic representation of images based on elementary geometric shapes or shape motifs, such as line and curve segments, parallel bars, angles and corners, and ellipsoids. These shape motifs may change their overall geometric attributes to account for large scale shape deformation. Each shape motif can be represented by an active basis model, so that the shape motif can deform to fit the image data.

Recall that an active basis model can be written in the following form

$$\mathbf{I}_m = \sum_{i=1}^n c_{m,i} B_{x_{m,i}, s, \alpha_{m,i}} + U_m = C_m \mathbf{B}_m + U_m,$$

where \mathbf{B}_m is a deformed template that is composed of Gabor wavelet elements. The shape script model is a hierarchical recursion of the above model. It is a composition of K shape motifs which are themselves active basis models:

$$\mathbf{I}_m = \sum_{k=1}^K C_{m,k} \mathbf{B}_{x_{m,k}, s_{m,k}, \rho_{m,k}, \alpha_{m,k}}^{(t_k)} + U_m, \quad (2)$$

where t_k is the type of the k -th shape motif (e.g., ellipsoid, angle, parallel bars, etc.), which is endowed with hyper-parameters, such as the overall location x , scale s , aspect ratio ρ and orientation α . Similar to the active basis, we may allow perturbations of these hyper-parameters, and the perturbations can be quite large because the sizes of the shape motifs are much larger than the Gabor wavelets. Such perturbations cause global deformations of the shape motifs, so that the model is capable of representing large deformations and articulations of object shapes. In addition, on top of the hyper-parameters, we also allow perturbations of the location, scale, and orientation parameters of the Gabor elements that belong to each shape motif. This causes local deformations of shape motifs. So model (3) is a recursive compositional model.



Figure 6: Detecting the objects in the testing images using the designed shape script template.

We can illustrate shape script by a simple experiment of detecting egrets from testing images. We design the shape script template for egrets. The detection process is illustrated in Figure (5) where the shape script template consists of four shape motifs: one ellipsoid for the body, two parallel bars for the neck, and one angle for the beak.

In the current implementation, we assume the following model for each testing image \mathbf{I} :

$$\mathbf{I} = \sum_{k=1}^K C_k \mathbf{B}_{x+x_k+\Delta x_k, s_k+\Delta s_k, \rho_k+\Delta \rho_k, \alpha_k+\Delta \alpha_k}^{(t_k)} + U,$$

where $K = 4$ is the number of shape motifs, $t_k \in \{ \text{ellipsoid, parallel bars, angle} \}$ indexes the type of motif k , $(x_k, s_k, \rho_k, \alpha_k)$ are the location, scale, aspect ratio, and orientation of the k -th shape motif. In this experiment, we design the shape script template by giving the values of the parameters of the four shape motifs. Given the shape script, we will estimate x , the location of the object in the testing image \mathbf{I} , as well as the deformation of the template, i.e., $(\Delta x_k, \Delta s_k, \Delta \rho_k, \Delta \alpha_k, k = 1, \dots, K)$.

The log-likelihood of x is computed by

$$l(x) = \sum_{k=1}^K \max_{(\Delta x, \Delta s, \Delta \rho, \Delta \alpha)} [\mathbf{I}, \mathbf{B}_{x+x_k+\Delta x, s_k+\Delta s, \rho_k+\Delta \rho, \alpha_k+\Delta \alpha}^{(t_k)}], \quad (3)$$

where $[\mathbf{I}, \mathbf{B}_{x+x_k+\Delta x, s_k+\Delta s, \rho_k+\Delta \rho, \alpha_k+\Delta \alpha}^{(t_k)}]$ is the log-likelihood resulting from matching shape motif k .

Equation (3) can be implemented by a recursive structure of sum-max maps illustrated in Figure (5). In this recursive cortex-like structure, the SUM2 maps score the matching of the shape motifs. The MAX2 maps compute the local maxima of the SUM2 maps. The local maximization computation estimates the shape deformation $(\Delta x_k, \Delta s_k, \Delta \rho_k, \Delta \alpha_k, k = 1, \dots, K)$. The SUM3 map scores the overall matching of the shape script template. After detecting the overall location of the object, a top-down process then retrieves the arg-max of the local maximization computation in MAX2 and MAX1, so that we can deform the shape script template to sketch the image.

Figure (6) displays some examples of detecting egrets by the designed shape script template using the recursive sum-max maps. For each testing image, we superpose the deformed template obtained by the aforementioned top-down retrieval process, where the deformation involves changing the parameters of the shape motifs as well as the parameters of the Gabor wavelet elements of the shape motifs. Because we allow big changes in the parameters of the shape motifs, the shape script model can account for large deformations, articulations, and pose changes. The SUM3 scores of the detected objects are generally higher than those of natural background images.

We are currently working on designing a dictionary of shape motifs, and learning the shape script templates from training images by selecting the shape motifs from the dictionary.

The shape script model is a layered hierarchical extension of the active basis model. The SUM2 maps may correspond to the functions of neurons in V2 area of the visual cortex. The shape script model may well be the next step beyond wavelet representations.

The shape script model also reveals a general principle for layered hierarchical model for visual cortex. The basic idea is that after we obtain a sparse coding of the original data, we need to continue to pursue further sparse coding of the attributes of the sparse coding elements. This naturally leads to a higher layer of sparse coding. The residuals in coding the attributes correspond to the activities in the active basis model. This suggests that active basis model may be an inevitable building block in a hierarchical representation of natural images.

4 Active plates model

Besides the shape script model, we have also worked on another layered hierarchical extension of the active basis model. We call this model the active plates model. In this model, each template is a composition of a number of partial templates or plates. These partial templates or plates are allowed to shift their overall locations, scales and orientations to account for large shape deformation. Each partial template or plate is represented by an active basis model to account for more local shape deformation. The partial templates can be learned from training images. Currently we allow the partial templates to have some overlap with each other, so that this redundant representation is immune to occlusions.

Figure (7) illustrates the basic idea. The plot on the left displays an array of partial templates learned from the images of cat heads, where each learned partial template carries a log-likelihood score. The plot on the right displays some of the training images and their sketches by the active

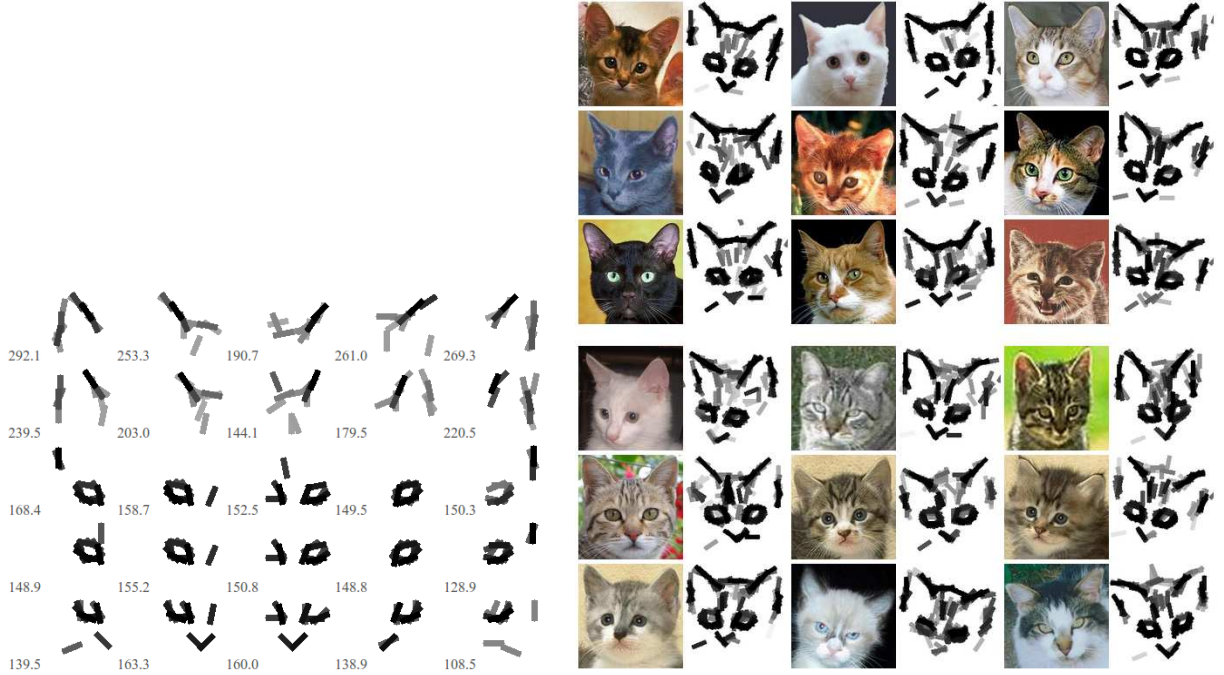


Figure 7: Active plates model. Left: the learned partial templates with their log-likelihood scores. Right: some training images and the matched templates.

plates model. It is clear that the active plates model is flexible enough to account for shape deformations in this example.

We are currently working on unsupervised learning of a dictionary of partial templates from non-aligned training images. The learned partial templates can then be composed into active plates models.

In addition to the above findings, the PI and collaborators have worked on a stochastic matching pursuit algorithm for variable selection in linear regression. The co-PI (Zhu) and a student (Hu, W.) have extended the active basis model to learning three-dimensional templates. The co-PI (Zhu) and another student (Wu, T.) have also studied the bottom-up and top-down inference processes, which are partially based on the active basis models.

5 Reproducibility

The code and data sets for reproducing the above findings can be downloaded at

<http://www.stat.ucla.edu/~ywu/AB/ActiveBasisMarkII.html>