**Military Technical College**
**Kobry El-Kobba**
**Cairo, Egypt**

**12-th International Conference**
**on**
**Aerospace Sciences &**
**Aviation Technology**

# A SYSTEM FOR FINDING AND SEGMENTING A HAND IN A PARTIALLY CLUTTERED SCENE

R.M. Farouk* and M. Ali Gomaa**

## ABSTRACT

In this paper we discuss a system for finding and segmenting a human hand, which is holding an object. This is a difficult computer vision problem, because of the many degrees of freedom, the flexibility, and the clutter problem of the hand. In order to find the correspondences we have used a lateral similarity function. In the lateral excitation we have not only compared different features at one point but also the features at that point and its neighbors ( for example : Gobar feature , and skin color feature ). The system is an extension of the Elastic Graph Matching (EGM) algorithm [1]. EGM has been shown to be successful in numerous object vision tasks, its best performance being in face recognition.

## KEY WORDS

Hand detection, Object detection, Object recognition, Lateral excitation,
Elastic Graph Matching, Gabor Wavelets.

---

* Department of mathematics , Faculty of science , Zagazig , University , Egypt .
* * Department of mathematics , Faculty of science for Girls , Al Azhar , University , Cairo , Egypt .

## INTRODUCTION

Recognizing non-rigid objects like the human hand is a very hard computer vision problem. Different methods have been presented in the literature, whose main drawbacks are low robustness or high computational load in the analysis of cluttered scenes [2]. In order to achieve immersive human-computer interaction, human body parts like the hand could be considered as a natural input which motivates the research of tracking, analyzing, and recognizing human body movements.
As an application example, the gestures could be used to represent some commanding inputs such as pointing, rotating, starting, stopping, etc. One of the important characteristics of human body parts is the skin color. Many good algorithms are based on the skin color to detect the hand in the scene [3], [4].
We have used the skin color feature in our algorithm, and we have found that the lighting and background (as in many other algorithms) play a big role in the response of the feature. The human hand has received attention from many researchers in the last decade. Many of the researchers have concentrated in hand tracking [6], fingerprint identification [7], hand sign recognition [7], and hand segmentation as in [8]. In our system we have looked for the characteristics of the hand and tried to find the correspondences of these characteristics by Labeled Graph Matching [1].
The rest of this paper is organized as follows. In section 2 we discuss the object representation with Gabor wavelets and describe the different features that we have used in our graph matching algorithm. In section 3 we explain
the generation of the algorithm described in section 2 with the EGM algorithm.
In section 4 we discuss our results and finally the advantages and disadvantages of our system.

## 2 OBJECT REPRESENTATION

Let $f(x)$ be the distribution of the input image in the color space HSI (Hue,Saturation, and Intensity), ([9], [5]). We have represented the components of the input image separately: for the intensity component I (x) , we have convolved it with a family of Gabor kernels $\psi_k(x)$ as in the formula:

$$(WI)(k, x_0) = \int_R \psi_k(x_0 - x) I(x) d^2 x \qquad (1)$$

the operator W symbolizing the convolution with a family of Gabor wavelet kernels which take the form of a plane wave restricted by a Gaussian envelope function as given by the formula

$$\psi_k(x) = \frac{k^2}{\sigma^2} \exp\left(\frac{-k^2 x^2}{2\sigma^2}\right) [\exp(ikx) - \exp\left(\frac{-\sigma^2}{2}\right)] \qquad (2)$$

which are parameterized by two-dimensional vector $k$ . The first term in the square brackets determines the oscillatory part of the kernel. The second term compensates for the DC-value of the kernel, to avoid unwanted dependence of the filter response on the absolute intensity of the image. For sufficiently high values of σ the effect of

the DC-term becomes negligible. The complex valued $\psi_k$ combine an even (cosine-type) and odd (sine-type) part. The filter response of $\psi_k$ in frequency space is given by

$$(F\psi_k)(k_0) = \exp\left(-\frac{\sigma^2(k_0-k)^2}{2k^2}\right) - \exp\left(-\frac{\sigma^2(k_0+k)^2}{2k^2}\right) \qquad (3)$$

where F denotes the Fourier transform. The first Gaussian centered at the characteristic frequency k provides a band pass filter. The second exponential removes the DC-component of $\psi_k$ . Equation (3) does not normalize the energy picked up by the kernel in the convolution. Consequently, this energy will be proportional to | $k^2$ |. D. Field [10] noted that the power spectrum of natural images decreases like $1/|k^2|$ . The energy resulting components of our image transform should therefore be roughly independent of $|k^2|$ .The $\psi_k$ form a family that is self-similar under the application of the group of translation, rotation, and scalings . The family is also known as “Gabor wavelets”. The wavelets are parameterized by the wave vector k , which controls the width of the Gaussian window and the wavelength and orientation of the oscillatory part. The parameter σ determines the ratio of window width to wavelength, i.e., the number of oscillations under the envelope function. The Gabor wavelets seem to be a good approximation to the sensitivity profiles of neurons found in visual cortex of higher vertebrates [11].
In order to construct a group of numerical values of Eq. (1), we employed a discrete lattice of k values

$$k_{\nu\mu} = \begin{pmatrix} k_\nu \cos(\phi_\mu) \\ k_\nu \sin(\phi_\mu) \end{pmatrix} \quad , k_\nu = \frac{k_{\max}}{f^\nu} \quad , \phi_\mu = \frac{\mu\pi}{D} \qquad (4)$$

where $\nu \in \{0, 1, ...,L-1\}$ , L is the number of scales and $\mu \in \{0, 1, 2, ...,D-1\}$, D is the number of orientations. In our algorithm we have used L = 3, D = 6, $k_{\max} = \frac{\pi}{4.0}$, $f^\nu = 0.5$ and $\sigma = \pi$ $k_{\max} = \frac{\pi}{4.0}$, $f^\nu = 0.5$ and $\sigma = \pi$ ,The wavelet transform as in Eq. (1) produces a complex numbers of oscillating phase. This leads to local similarity optima when matching stored wavelet features to the transform of an image. We call the vector of the stored wavelet features that taken from the intensity component I(x) at a position $x_0$ Gabor jet and we can write it in the form

$$(WI)(k, x_0) = J(k, x_0)\exp^{i\phi(k,x_0)} \qquad (5)$$

where $J$ is the amplitude, which is slowly varying with position, and phase $\phi$ varying with the spatial frequency given by the characteristic wave vector $k$. Gabor wavelets were chosen for their technical properties and biological relevance. Since they are DC-free, they provide robustness against varying illumination in the image. Robustness against varying contrast can be obtained by normalizing the jets. The limited localization in space and frequency yields a certain amount of robustness against translation, distortion, rotation, and scaling.

Only the phase changes drastically with translation, and can there to can be used for estimating displacement. [12], [13] provided more information about wavelets and its properties. For two jets, J extracted at one point in an image, and $J'$ at another point in the same image or a another image. There are two similarity functions [14]: the first uses only the absolute value in Eq. (5), and it is given by the formula

$$A_{abs}(J,J') = \frac{J.J'}{\|J\|\|J'\|} \tag{6}$$

The second similarity function uses the phase and amplitude and defined by the formula

$$A_{phase}(J,J') = \frac{J.J'\cos(\phi_1 - \phi_2)}{\|J\|\|J'\|} \tag{7}$$

The second feature that we have used is a local color average in HSI (for transform from RGB color space to the HSI color space [9], [15]. Two image regions are compared by their color averages. We consider regions of the 8-neighbors to compare color regions with the formula:

$$A_{col}(J,J') = \frac{\langle \underline{\Gamma}(J), \underline{\Gamma}(J') \rangle}{\|\underline{\Gamma}(J)\|.\|\underline{\Gamma}(J')\|} \tag{8}$$

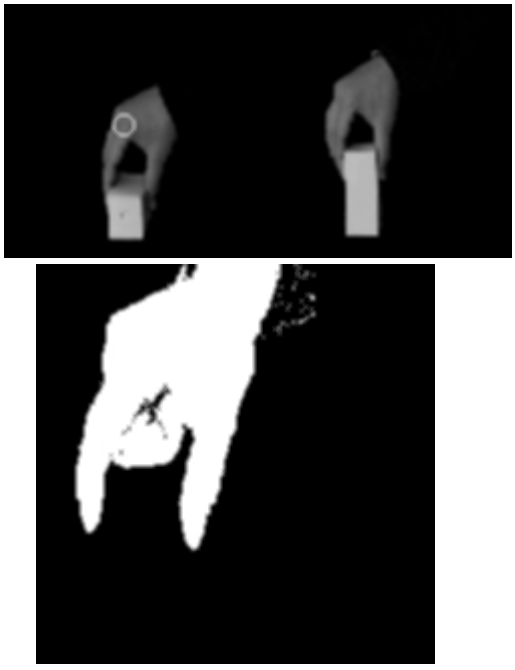

Fig. 1. The upper left is input image with a circle about the region that we try to find its correspondence in the second image uper right, a segmented skin color region that we have got according to Eq. (9), and Eq. (10), respectively.

Where

$\underline{\Gamma}(\Omega) = (\tau_H \Omega_H, \tau_S \Omega_S, \tau_I \Omega_I)^{\mathrm{T}}$

is the 8-neighbors averaged color. The result of this feature is displayed in Fig. 2 . The third feature that we have used is a skin color similarity image comparing each pixel in the original image with a prototype skin color $(H_0, S_0)$ as in the equation:

$$B_{skin} = \chi\left(1 - \sqrt{\left(\frac{H - H_0}{\tau_H}\right)^2 + \left(\frac{S_H - S_0}{\tau_S}\right)^2}\right) \quad (9)$$

where $\tau_H = 1$ , and $\tau_S = 0.2$ are the normalization factors, and $\chi(x)$ is a threshold function defined at time x > 0 by the equation:

$$\chi(x) = \begin{cases} 1 & If \quad x \geq 0; \\ 0 & If \quad x < 0. \end{cases} \quad (10)$$

we have extracted the gabor jet $k$ at each pixel from the resulting image B skin. To compare two jets $k$ extracted at one point in an image and $k'$ at another point in the same image or a another image, we have used Eq.(7).

### 2.1 Jet of Multiple Features

We have extracted the three different features at each pixel in the database. In order to compare two jets with the multiple features. We have used the weighted sum of the different similarities features as in:

$$A_{total}(J, J') = \sum_n w_n A_n(J, J'), \quad \sum_n w_n = 1 \quad (11)$$

where $n$ is an index over the multiple features, and $w_n$ are normalized weighting factors. Experiments as in Fig. 1, have shown that the skin color segmented feature plays the major role in finding the corresponding region.

## 3 GENERALIZATION OF ALGORITHM

### 3.1 Elastic Graph Matching

A visual pattern can be represented via a graph containing nodes labeled with local features and links encoding the topological relationship between the features [14]. Based on this representation, the problem of pattern recognition can be formulated as the one-to-one correspondence between the nodes of an input graph and one from the database. A good correspondence is one that respects the topological

relationships between the nodes, and finds high similarity between the labels of the correspondence nodes. In the neural implementation of EGM, two separate neural layers represent the input and the stored graphs. Graph nodes and graph links are represented by neurons and excitatory connection between the neurons, respectively. Node labels are represented by receptive field profiles of the neurons. The correspondence between the two graphs is found through a dynamical process where at the final state each neuron in the input layer is mapped to its corresponding neuron in the stored model layer [12].

### 3.2 Lateral Excitation

The traditional measure of similarity A between two Gabor jets $J$, and $J'$ is the cosine of the angle between the two jets is given by Eq. (6), and Eq. (7). This measure is useful as it provides robustness to variations in degree of contrast.
The optimization of similarity between node labels $V = \{J_1, J_2, ..., J_n\}$ and $V' = \{J'_1, J'_2, ..., J'_n\}$ of two graphs, G, and $G'$, respectively, requires the definition of a cost function $A_{graph}$. A natural choice is the sum of all pairwise jets similarities:

$$A_{graph}((G,G') = \frac{1}{n}\sum_{i=1}^{n} A_{abs}(J_i, J'(i)) \qquad (12)$$

The previous algorithmic implementations of graph matching for object recognition have used this cost function. An alternative cost function would be to use a count of nodes with a similarity beyond a threshold. While the relative performance of alternative cost function including the latter scheme have not been documented, the latter method seems to be inferior in that it requires defining a threshold for the similarity values to be accepted as high. In our new version of EGM, the similarity between two jets is computed in our algorithm according to Eq. (11). However, we augmented the graph similarity function with an element that emphasize the topological coherence of the match. The new graph similarity function $A_{lat}$ involves the enhancement of each pairwise similarity value $\overline{A}_{total}$ by its neighboring similarity values as in the following formula.

$$A_{lat}(G,G') = \frac{1}{n}\sum_{i=1}^{n} \overline{A}_{total}(J_i, J'_c(i)) \qquad (13)$$

$$\overline{A}_{total}(J_i, J'_c(i)) = A_{total}(J_i, J'_c(i)) + A_{total}(J_i, J'_c(i))\sum_k A_{total}(J_k, J'_{c(k)}) \qquad (14)$$

where $k$ is the index of neighbors of $J_i$ in the graph topology, and $c(k)$ is the index of the jet matched with $J_k$ (for biological analog of this function [16]). The second term in Eq. (14) is the excitation received by $A_{total} = (J_i, J'_c(i))$. This is consistent with the physiological findings in the visual cortex. It has been shown that "... the level of excitation induced by activating the horizontal inputs depends on the level of depolarization of the target cell: the more depolarized the cell is, the larger the excitatory postsynaptic potential, as the result of voltage-dependent sodium conductance. Thus one can think of the effect of horizontal connections as being

state dependent and influenced by the level of activation of other inputs converging onto the cell" ([17]). The motivation behind the similarity measure with lateral excitation was to fortify the algorithm against noise and clutter.

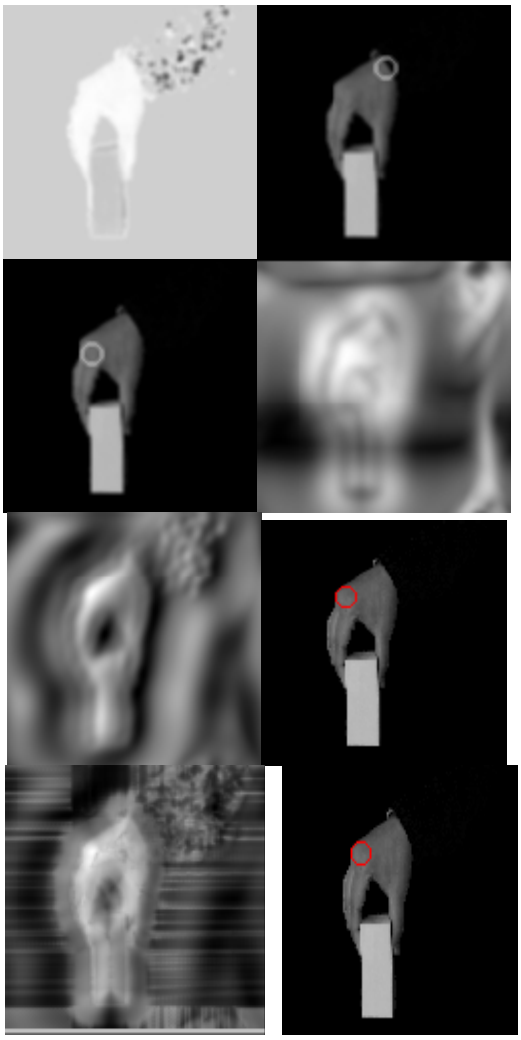

Fig. 2. The first column from top to bottom shows the color similarity distribution computed from Eq. (8), the Gabor jet similarity as defined in Eq. (6), the similarity of the segmented skin image Eq. (7), total similarity of different features computed with the form in Eq. (11), respectively. The second column marked the highest similarity to the circled region in the input image in Fig. 1.

## 4 EXPERIMENTS AND RESULTS

We have tested our system on a database of pictures taken by a Sony firewire camera with a constant background, different lighting degrees, and complex background (Fig. 3). In our test we have found that the system is sensitive against background and the illumination conditions as in many different algorithms.
Another condition that we have tried to control is the pose of the grip of the hand. Due to the flexibility of the hand, and different characteristics of each pose of the

hand, we have found that the system is tolerant with regard to slight pose changes. The use of the lateral similarity in Eqs. (13) and (14) is practically important, because the second term in Eq. (14) is the excitation received by $A_{total} = (J_i, J'_c(i))$ . As it can be seen, the amount of excitation directly depends on the value of $A_{total} = (J_i, J'_c(i))$ . The matching process declined when the pose difference became greater. In our experiment we have found that the matching depends upon the background, lighting, and the pose of the grip. Some of our results have been shown in Fig. 4. As it can be seen, the correspondence in the first example (the first row in Fig. 4) is perfectly matched because we have tried to control the pose as good as possible, and then we have succeeded to segment the hand perfectly.

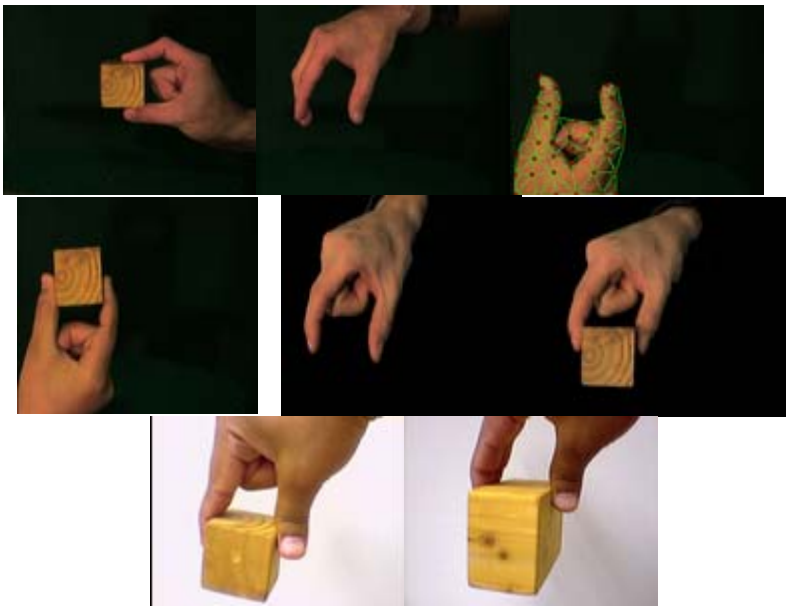

Fig. 3. A sample of the objects that we have used to test our system

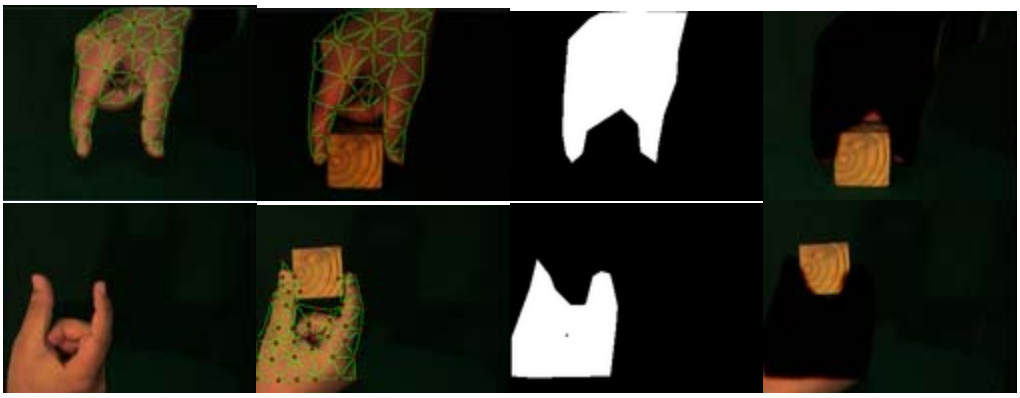

Fig. 4. In the first row from left to right: The grip of the hand with the model graph, the model graph after the matching process, segmented part of the hand, and the separated part of the object after separating it from the hand. The second row shows another example with a different grip.

## 5 DISCUSSION

The finding of non-rigid parts of the human body in our opinion is not a trivialcomputer vision tasks. The main aim from our system was how to find the handwhich is partially occluded by another objects, and then separating the objectfrom the hand by segmenting the hand. We have found that the task is not trivialbecause the hand has no fixed characteristics, but we have tried to extract features at the most important parts of the hand like finger tips, finger joints and other important positions. The system for the constant background, and the slightly pose gives a good results. The slighter the pose is, the better results are. The Elastic Graph Matching(EGM) [1] was introduced over a decade ago as a biologically inspired method for pattern recognition. Our results demonstrated that despite its age, EGM remains a competitive algorithm, even in comparison with the modern analytically developed statistical information processing methods. EGM with lateral excitations(similarity) in Eq. (13) reveals the following characteristic. Given equal similarity values across all the nodes in the graph for a given match, the nodes at the graph boundary receive less excitation from their neighbors than the nodes in the center because they have fewer neighbors. In other words, given random similarity values, the nodes at object boundary are weighted less than those in the center, and thus their contribution to the total graph similarity is reduced. Of course this argument only holds on average. In cases where a boundary node is surrounded by very high similarity values, or where a center node is surrounded by low similarity nodes, this will no longer be the case. However, this subtle weighting scheme can be important in situations where the objects are to be detected in clutter. The jets located at or near the boundary of the object are affected by the structure in the background, hence leading to low similarity with the boundary jets in a model graph. Lateral excitation in effect weights down the contribution from the boundary jets and improves the robustness of graph matching. In the future work, we will discussions the complexity of our system and compare the result with the existing systems.
The effect of lighting is only sensitive for the skin color features, but our system is an integrated system from different cues as it is clear in above sections.

## ACKNOWLEDGMENTS

The authors would like to thank the developers of the FLAVOR software environment, which served as the platform for this research.The first author also would like to thank Prof.Dr. C. von der Malsburg for his friendly discussions.

## REFERENCES

[1] Christoph von der Malsburg. Pattern recoggnition by labeled graph matching. Neural Networks, Vol.1:pp.141–148, 1988.

[2] Christoph von der Malsburg Laurenz Wiskott. A neural system for the recognition of partially occluded objects in cluttered scenes. Int.J.of Pattern Recognition and Artificial Intelligence, Vol.7(No.4):pp.935–948, 1993.

[3]Thomas S. Huang Ying Wu. Nonstationary color tracking for vision-based human computer interaction. IEEE Transactions on Neural Networks, Vol.13(No.4), 2002.

[4] Jochen Triesch. Vision-Based Robotic Gesture Recognition. PhD thesis, Berichte aus der Informatik, Shaker Verlag Aachen, 1999.

[5] Seiji Igi Akira Utsumi, Nobuji Tetsutani. Hand detection and tracking using pixel value distribution model for multiple-camera-based gesture interactions. Proceedings of the IEEE Workshop on Knowledge Media Networking, 2002.

[6] Anil K. Jain Sharath Pankanti, Salil Prabhakar. On the individuality of fingerprints. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages pp.805–812, Hawaii, December 11-13 2001.

[7] A. C. Downton and H. Drouet. Image analysis for model-based sign language coding. in Progress in image analysis and processing II: Proc. of the 6th International Conference on Image Analysis and Processing, pages pp.79–89, 1991.

[8] Y. Cui and J. Weng. Hand segmentation using learning-based prediction and verification for hand-sign recognition. In in Proc. IEEE Conf. Computer Vision and Pattern Recog., pages pp.88–93, june 1996.

[9] Richard E. Woods Rafael G. Gonzalez. Digital Image Processing. Addison-Wesley Publishing Company, 1992.

[10] D.Field. Relation between the statistics of natural images and the response properties of cortical cells. J.Opt.Soc.Amer.A, Vol.4(No.12):pp.2379–2394, 1987.

[11] J. Jones and L. Palmer. An evaluation of two-dimenional gabor filter model of simple receptive fields in cat straite cortex. J. Neurophysiol, pages pp.1233–1258, 1987.

[12] Rolf P. W¨urtz. Multilayer dynamic link networks for establishing image point correspondences and visual object recognition. PhD thesis, volume 53 of Reihe Physik,Verlag Harri Deutsch,Thun,Frankfort a. main, 1995.

[13] Gerald Kaiser. Friendly Guide to Wavelets. Birkh¨auser, 1994.

[14] L. Wiskott. Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis. PhD thesis, volume 53 of Reihe Physik,Verlag Harri Deutsch,Thun,Frankfort a. main, 1995.

[15] B. Prados-Su`arez J.Chamorro-Martinez, D. S`anchez. A fuzzy color image segmentation applied to robot vision. 1998.

[16] C. Gilbert. Horizontal integration and cortical dynamics. Neuron, Vol.9(No.1-13):pp.121–128, 1992.

[17] Gillbert J.Hirsch. Synaptic physiology of horzontal connections in the cat's visual cortex. J.of Neuroscience, Vol.11:pp.1800–1809, 1991.