# COPYRIGHT NOTICE

# A System for Person-Independent Hand Posture Recognition against Complex Backgrounds

Jochen Triesch and Christoph von der Malsburg

**Abstract**—A computer vision system for person-independent recognition of hand postures against complex backgrounds is presented. The system is based on Elastic Graph Matching (EGM), which was extended to allow for combinations of different feature types at the graph nodes.

**Index Terms**—Computer vision, human computer interaction, human robot interaction, hand posture recognition, gesture recognition, object recognition, segmentation, complex backgrounds, elastic graph matching, Gabor wavelets.

## 1 Introduction

Gestures are a very powerful means of communication among humans. In fact, gesturing is so deeply rooted in our communication that people often continue gesturing when speaking on the telephone. Consequently, a significant amount of research has recently addressed building gesture-based human-computer and human-robot interfaces [15], [2], [8], [16], [17]. The recognition of hand postures is an important ingredient for building gesture-based interfaces. Although there have been attempts to recognize gestures and sign language without recognizing the hand posture it is clear that these approaches are very limited. For example, in American Sign Language, the same arm movement can have very different meanings depending on the posture of the gesturing hand.

A number of requirements make hand posture recognition particularly challenging. First, if a system is intended to be person independent, it must cope with geometric distortions due to different hand anatomy or different performance of gestures by different persons. Second, the system must be able to deal with complex, cluttered backgrounds, making segmentation of the gesturing hand difficult. Most work in the field tries to circumvent the problem by either using colored markers, or by requiring the background to be static, or by requiring the hand to be the only skin-colored object in the scene [20], [3], [7], [6], [12], [9], [14], [13]. In contrast to this, our goal was a system that works in relatively unconstrained environments, where segmentation based on primitive cues is not always possible. Along similar lines, it should not be necessary for subjects to take off rings, etc., before interacting with the system but they should be able to just "come as they are." Other important requirements that, however, were not at the focus of this work, are real-time performance, naturalness of gestures, and robustness to varying lighting conditions.

We chose Elastic Graph Matching (EGM), a neurally inspired architecture for view-based object recognition, as our starting point because 1) it has an inherent ability to handle geometric distortions, 2) it does not require a perfectly segmented input image, and 3) it elegantly represents variances in object appearance with the *bunch graph* method. In order to make EGM more robust in the presence of complex backgrounds, we have extended it to handle multiple feature types. Our experiments demonstrate the advantage of using multiple feature types over the classical approach. The remainder of the paper is organized as follows: In Section 2, we will review EGM. Section 3 describes the generalization of EGM to multiple feature types. In Section 4, we present our experiments. Finally, Section 5 gives a discussion.

- *J. Triesch is with the Department of Cognitive Science, University of California at San Diego. E-mail: triesch@ucsd.edu.*
- *C. von der Malsburg is with the Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany and Laboratory for Computational and Biological Vision, Hedco-Neuroscience Building 228, University of Southern California, Los Angeles, California 90089-2520. E-mail: malsburg@neuroinformatik.ruhr-uni-bochum.de.*

*Manuscript received 9 Dec. 1999; revised 17 Nov. 2000; accepted 4 Mar. 2001.*
*Recommended for acceptance by R. Sharma.*
*For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111060.*

## 2 Elastic Graph Matching

Elastic Graph Matching (EGM) is a neurally inspired object recognition architecture [11], which has already been successfully applied to object recognition, face finding and recognition, and the analysis of cluttered scenes. Although being motivated by a theory of neural information processing, EGM is similar to other elastic matching approaches [5]. In EGM, views of objects are represented as labeled graphs with an underlying two-dimensional topology. The nodes of a graph are labeled with a local image description. Edges of a graph are labeled with a distance vector. Elastic matching of a model graph to an image means to search for a set of node positions such that 1) the local image description attached to each node matches the image region around the position where the node is placed and 2) the graph is not distorted too much.

**Gabor Jets.** As a local image description, a *Gabor jet* is usually used. This is a vector of responses of Gabor wavelets:

$$\psi_{\mathbf{k}}(\mathbf{x}) = \frac{\mathbf{k}^2}{\sigma^2} \exp\left(-\frac{\mathbf{k}^2\mathbf{x}^2}{2\sigma^2}\right)\left[\exp(i\mathbf{k}\mathbf{x}) - \exp\left(\frac{-\sigma^2}{2}\right)\right]. \quad (1)$$

These wavelets represent a plane wave with wave vector $\mathbf{k}$ restricted by a Gaussian envelope function of width $\sigma$; $\mathbf{x}$ denotes the two-dimensional image location. The responses of several such filters with different size and orientation, parameterized by different $\mathbf{k}$, constitute a jet:

$$\mathbf{k}_{\nu\mu} = k_\nu \begin{pmatrix} \cos\phi_\mu \\ \sin\phi_\mu \end{pmatrix} \quad \text{with} \quad k_\nu = k_{max}/f^\nu,\ \phi_\mu = \frac{\mu\pi}{D}. \quad (2)$$

Here, the index $\nu \in \{0, \ldots, L-1\}$ labels different spatial scales or frequencies and $\mu \in \{0, \ldots, D-1\}$ labels different orientations. Hence, $L$ is the number of frequency levels used and $D$ is the number of orientations. The value $f$ is the spacing factor between kernels in the frequency domain and $k_{max}$ is the maximum wave number. In this paper, we use $L = 3$, $D = 8$, $f = 1/\sqrt{2}$, $k_{max} = 1.7$, and $\sigma = 2.5$. A jet is a complex vector composed of the $L \times D$ complex filter responses $c_j$, where the index $j$ is a double index running over different orientations and scales. The $c_j$ are represented as absolute value $a_j$ and phase $\phi_j$: $c_j = a_j e^{i\phi_j}$.

Two similarity functions have been proposed for comparing Gabor jets [11], [21]:

$$\begin{aligned} S_{\text{abs}}(J, J') &= \frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}, \\ S_{\text{pha}}(J, J') &= \frac{1}{2}\left(1 + \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j)}{\sqrt{\sum_j a_j^2 \sum_j a_j'^2}}\right). \end{aligned} \quad (3)$$

$S_{\text{abs}}$ uses only magnitudes of the complex filter responses. It has the form of a normalized scalar product. $S_{\text{pha}}$ uses the magnitude and phase of the complex filter responses. Both functions yield similarity values between zero and one. While $S_{\text{abs}}(J, J')$ slowly changes when $J'$ is moved across the image, $S_{\text{pha}}(J, J')$ varies very rapidly because the phases of filter responses change significantly on a spatial scale corresponding to the wave-vector $\mathbf{k}$ of strongly responding kernels.

**Bunch Graphs.** The *bunch graph* idea [22] was invented to model variability in object appearance. The natural variability in the jets of corresponding points in several images of the same object or a class of objects (e.g., several left eyes of different persons) is captured by labeling each node of a graph with a set or

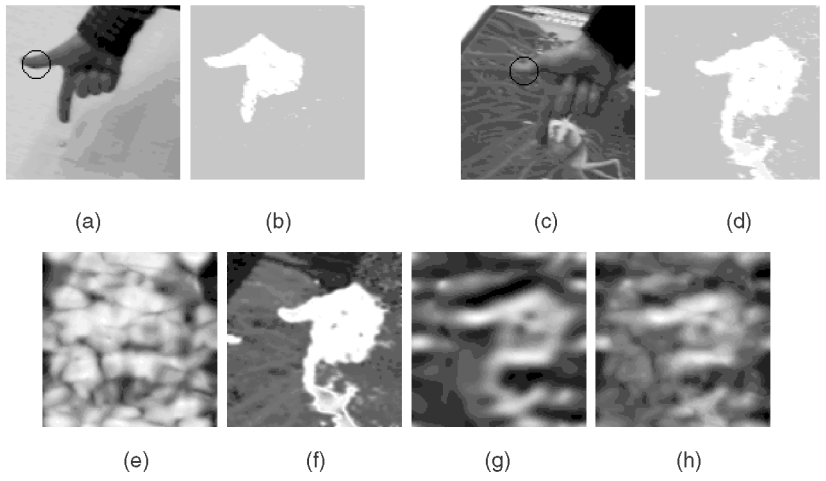


Fig. 1. Establishing correspondences between image points using different feature types. The task is to find the position in the target image (c) that corresponds to the marked position in the source image (a) ((b), (d) are the corresponding skin color segmentations). (e)-(h) are similarity landscapes obtained when comparing local image descriptions extracted at each position in (c) with the local image description taken at the marked position in (a). Light areas represent high similarity. (e) Only Gabor Jets, (f) only color average, (g) only colorGabor Jets, and (h) Compound Jets comprising a linear combination of all three. The circle in (c) corresponds to the maximum in (h).

"bunch" of jets (short: *bunch jet*) extracted from corresponding points in different example images rather than only with a single jet. Previously, we have shown that this method can also be used to elegantly model complex backgrounds [19]. For the matching process, we need a similarity function comparing the set of jets attached to each node of the bunch graph with local image information. The similarity of a bunch jet $B^n$ comprising $K$ jets at a node $n$ to a single jet $J(\mathbf{x})$ taken at a point $\mathbf{x}$ in an image is defined to be the maximum of the similarities of the $K$ individual jets:

$$S^{\mathrm{B}}_{\mathrm{abs}}(B^n, J(\mathbf{x})) = \max_k \{S_{\mathrm{abs}}(J^n(k), J(\mathbf{x}))\ , k = 1, \ldots, K\}\,. \quad (4)$$

$S^{\mathrm{B}}_{\mathrm{pha}}$ is defined analogously. When a graph $G$ with $N$ nodes is matched to an image $I$ at the node positions $\mathbf{x}_n$, its total similarity is given by the average of the node similarities:

$$S^{\mathrm{G}}_{\mathrm{abs}}(G, I) = \frac{1}{N}\sum_{n=1}^{N} S^{B}_{\mathrm{abs}}(B^n, J(\mathbf{x}_n)). \quad (5)$$

Again, we define $S^{\mathrm{G}}_{\mathrm{pha}}$ in an analogous manner.

## 3 EGM with Multiple Feature Types

While earlier versions of EGM have only worked with a single shape or texture feature type such as Gabor or Mallat filters, we have recently extended it for handling multiple feature types [18]. There are several ways of extending EGM for multiple feature types. First, the feature types used may differ from node to node. For example, a model graph describing a face could employ edge features at the outline of the face and texture features at the inside. Second, all nodes may be labeled with a combination of different feature types which is identical for all nodes. Third, nodes may be labeled with combinations of feature types which may differ from node to node. Here, we focus on the second scheme.

**Compound Jets.** In order to allow for multiple feature types at a node, we introduce the notion of a *compound jet*. In a compound jet, several local image descriptions are simply concatenated. When similarities between two compound jets $\mathcal{J}$ and $\mathcal{J}'$ are computed, first the similarities of their corresponding constituents are considered. These are are then combined by computing a weighted average:

$$\mathcal{S}(\mathcal{J}, \mathcal{J}') = \sum_{\mathcal{F}} w_{\mathcal{F}}\ S_{\mathcal{F}}(J_{\mathcal{F}}, J'_{\mathcal{F}})\,, \quad \sum_{\mathcal{F}} w_{\mathcal{F}} = 1\,. \quad (6)$$

The $S_{\mathcal{F}}$ are similarities between jets of a particular feature type $\mathcal{F}$.

**Compound Bunch Graphs.** We can extend the bunch graph idea to graphs with compound jets as well. *Compound bunch graphs* are constructed in the following manner:

- For every node $n$ of a particular graph of a view of an object, the jets of the same feature type $\mathcal{F}$ extracted from $K$ different training images are integrated into a bunch jet $B^n_{\mathcal{F}} = \{J^n_{\mathcal{F}}(1), \ldots, J^n_{\mathcal{F}}(K)\}$. With these bunch jets as node labels, one bunch graph is created for every feature type whose geometry is averaged from the constituent graphs.
- The bunch jets $B^n_{\mathcal{F}}$ of different feature types but from the same node $n$ of the same view are concatenated to a *compound bunch jet* $\mathcal{B}^n$. With these compound bunch jets as node labels, a compound bunch graph is created whose geometry is identical to that of the constituent bunch graphs of different feature types.

For evaluating the similarity $\mathcal{S}^{\mathrm{B}}(\mathcal{B}^n, \mathcal{J}(\mathbf{x}))$ between a compound bunch jet $\mathcal{B}^n$ and a compound jet $\mathcal{J}(\mathbf{x})$ extracted at a particular location $\mathbf{x}$ a weighted average of contributions $S^{\mathrm{B}}_{\mathcal{F}}$ stemming from different feature types is used:

$$\mathcal{S}^{\mathrm{B}}(\mathcal{B}^n, \mathcal{J}(\mathbf{x})) = \sum_{\mathcal{F}} w_{\mathcal{F}}\ S^{\mathrm{B}}_{\mathcal{F}}(B^n_{\mathcal{F}},\ J_{\mathcal{F}}(\mathbf{x}))\,, \quad \sum_{\mathcal{F}} w_{\mathcal{F}} = 1\,. \quad (7)$$

The functions $S^{\mathrm{B}}_{\mathcal{F}}$ compare a bunch jet $B^n_{\mathcal{F}}$ to a simple jet $J_{\mathcal{F}}$ of the same feature type $\mathcal{F}$ in analogy to (4):

$$S^{\mathrm{B}}_{\mathcal{F}}(B^n_{\mathcal{F}}, J_{\mathcal{F}}(\mathbf{x})) = \max_k \{S_{\mathcal{F}}(J^n_{\mathcal{F}}(k),\ J_{\mathcal{F}}(\mathbf{x})), \quad k = 1, \ldots, K\}\,. \quad (8)$$

The similarity functions $S_{\mathcal{F}}$ directly compare two jets of a particular feature type $\mathcal{F}$. The similarity of a compound bunch graph $\mathcal{G}$ with node positions $\mathbf{x}_n$ to an image $I$ is defined as the average of the similarities of all $N$ nodes, in analogy to the previous section:

$$\mathcal{S}^{\mathrm{G}}(\mathcal{G}, I) = \frac{1}{N}\sum_{n} \mathcal{S}^{\mathrm{B}}(\mathcal{B}^n, \mathcal{J}(\mathbf{x}_n))\,. \quad (9)$$

## 4 Experiments

### 4.1 Additional Feature Types

In addition to conventional Gabor jets we chose to include two color feature types in our representation of hand postures. Fig. 1 shows how the different feature types compare for finding image point correspondences. We represent colors in $HS$ (hue,

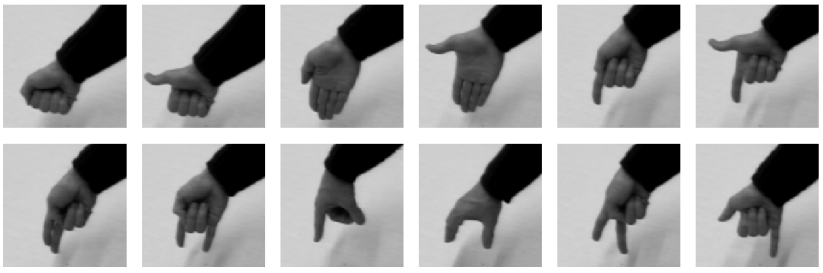

Fig. 2. The 12 hand postures to be recognized.

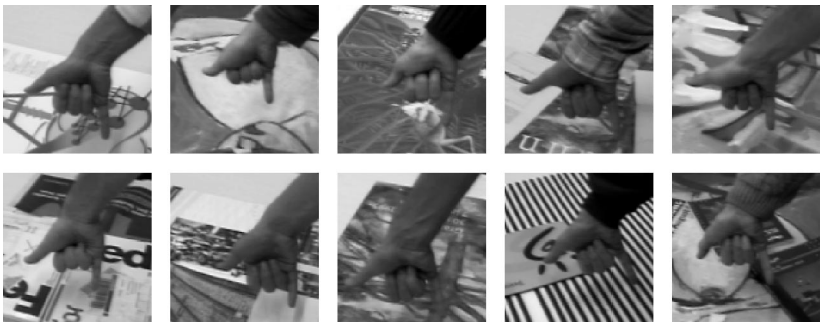

Fig. 3. Posture 12 performed by 10 out of 19 subjects against complex backgrounds.

saturation) space. Preliminary tests showed that, for our data set, skin color tones fall in a region centered at $H_0 = 0.118$ and $S_0 = 0.333$, whose elongation in the $S$ direction is about 5 times bigger than in the $H$ direction (all subjects were Caucasian). We account for this by introducing scaling factors $\alpha_H = 1$, $\alpha_S = 1/5$.

**Color Average.** We consider the color averages of $3 \times 3$ pixel regions. The similarity function is defined by:

$$S_{\text{col}}(J, J') = \frac{\alpha_H^2 HH' + \alpha_S^2 SS'}{\sqrt{\alpha_H^2 H^2 + \alpha_S^2 S^2}\sqrt{\alpha_H^2 H'^2 + \alpha_S^2 S'^2}}. \tag{10}$$

**ColorGabor Jet.** A *colorGabor jet* is similar to a Gabor jet, but convolutions are performed on images expressing each pixel's similarity to skin color. The skin color similarity compares each pixel to the prototypical skin color $(H_0, S_0)$:

$$\begin{aligned} S_{\text{skin}} &= \mathcal{R}_t\left(1 - \sqrt{\alpha_H^2 (H - H_0)^2 + \alpha_S^2 (S - S_0)^2}\right), \\ \mathcal{R}_t(x) &= \begin{cases} 0 & : \quad x \le t \\ x & : \quad x > t \end{cases}. \end{aligned} \tag{11}$$

$\mathcal{R}_t(x)$ is a thresholded ramp function parameterized by a constant $t = 0.784$. The colorGabor jets strongly respond to edge - and bar-like structures in the skin color segmentation. As the fingertips are usually closer to skin color than the background, it is advantageous to consider the phases of the filter responses in the similarity function. Only the phase information can distinguish between, e.g., a skin-colored bar on not-skin-colored background from the opposite. Hence, we use $S_{\text{pha}}$ from (3) for computing similarities between colorGabor jets during the graph matching.

### 4.2 Image Database and Training

Our database consists of more than 1,000 color images of $128^2$ pixels of 12 hand postures (Fig. 2) performed by 19 persons against simple and complex backgrounds (Fig. 3) with varying amount of skin color. The images of three subjects signing against uniform light and dark backgrounds formed the training set, giving six training images per posture, the remaining images formed the test set. For the images in the training set, we constructed graphs of 15 nodes. All 15 nodes were manually placed at anatomically significant points. The number of training images is quite small, but because preliminary experiments were very encouraging and, due to the amount of manual work involved in creating the model graphs, we chose not to add more images to the training set. Jets of the three different feature types were extracted at the node positions of every training image. Then, we created one model graph for every feature type for every training image of every posture giving a total of $3 \times 6 \times 12 = 216$ model graphs. These graphs were combined into 12 compound bunch graphs (one for every posture) as described in Section 3.

### 4.3 Matching Procedure and Recognition

For recognition, the 12 compound bunch graphs are sequentially matched to an input image as described below. The graph obtaining the highest similarity represents the classification result. The matching procedure operates in four steps:

1. Coarse positioning of the graph: The image is scanned in coarse steps of five pixels in x and y direction without graph distortion.
2. Rotation in plane: The graph may rotate around its center of gravity by up to $15^\circ$ in the image plane (seven orientations). Additionally, it may shift its position by up to six pixels (two steps of three pixels) in x and y direction.
3. Rescaling of the graph: The graph is allowed to grow by up to 20 percent or shrink by up to 10 percent without relative changes of the edge lengths (five scales). Additionally, it is allowed to shift its position by up to four pixels (two steps of two pixels) in the x and y direction.
4. Local diffusion of single nodes: All nodes may shift their positions by one pixel to compensate for residual geometric distortions.

During all matching steps, similarities between Gabor jets are computed using $S_{\text{abs}}$, while $S_{\text{pha}}$ is used for the colorGabor jets (3). An example of a successful match is given in Fig. 4. The nodes usually find their proper positions during the matching, even if the background is very complex or contains large regions of skin color.

### 4.4 Results

Cross runs on a test set of 604 images taken against uniform light or dark background and 338 images against complex backgrounds

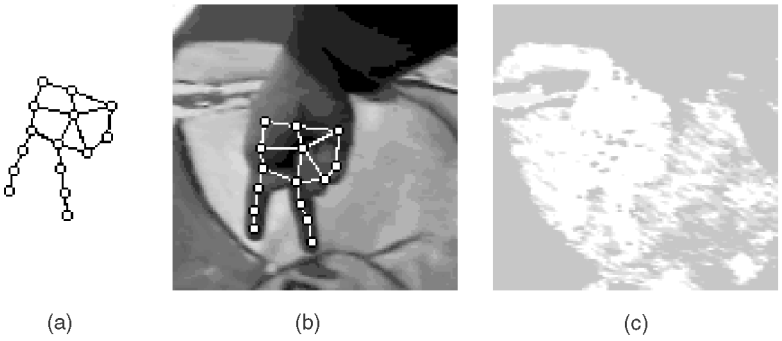


Fig. 4. Example of a model graph being matched onto an input image. (a) Original graph. (b) Matched graph. (c) Skin color segmentation.

were performed. The results are summarized in Table 1. The weighting factors $w_{\mathcal{F}}$ between different feature types were systematically varied in an exhaustive manner considering all relative weightings of the types 1:0:0, 4:1:0, 3:2:0, 3:1:1, 2:2:1, 2:1:1, and 1:1:1, and system performance on the test set was measured. The best result was obtained for $w_{\text{Gabor}} = w_{\text{ColorGabor}} = 25\%$, $w_{\text{Color}} = 50\%$. Since the weighting of cues only introduces two free parameters, we decided not to use a separate validation set. It turns out that a proper combination of the three feature types outperforms any of them alone. For example, the error for the best combination found is less than half as big as that for using Gabor features alone. Performance does not depend critically on the precise weighting between the features if all are being used (Fig. 5). The recognition rates vary smoothly with the weighting and there is a large plateau of weightings yielding recognition rates higher than any of the feature types alone could account for. If only Gabor and colorGabor features but not color features are used, performance is still very good, suggesting that the color features are not essential for the system's performance.

An analysis of the system's errors reveals that strong geometric distortions due to either differences in performance of a posture, different hand anatomy, strong rotation in depth, or combinations of these account for practically all errors. One might suggest that greater flexibility of the graph during matching could alleviate this problem. However, allowing for more diffusion in the last matching step does not have the desired effect. The reason seems to be that the geometric distortions are not of a statistical nature, which would be well-modeled by a diffusion process, but instead show high correlations. This is due to the hand's kinematics, which only allow for certain correlated movements of nodes. A more refined modeling of likely graph distortions seems like a good avenue for future research.

## 5 Discussion

We have presented a system for person-independent recognition of hand postures against complex backgrounds. Our experiments demonstrate that Elastic Graph Matching with multiple feature types is a powerful architecture for view-based object recognition. The good performance, despite the small number of training images suggests, that the graph representation has good generalization properties. The introduction of additional feature types gave significant improvements over just using Gabor jets. In another experiment [18], the advantages of using multiple feature types were even more pronounced: For the analysis of cluttered scenes of occluding objects against complex backgrounds, the recognition rate climbed from 10.0 percent of correctly analyzed scenes to 62.9 percent due to the use of compound jets instead of only Gabor jets. There, however, the success was more dependent on the precise weighting between the feature types.

TABLE 1
Percentages of Correct Recognition

| weighting | simple background | complex background |
|---|---|---|
| only Gabor | 82.6% | 70.4% |
| only Color | 39.7% | 34.6% |
| only ColorGabor | 88.2% | 76.3% |
| best Mixture | 92.9% | 85.8% |

We integrated the system into a gesture interface for an anthropomorphic autonomous robot with an active vision system [1]. The demonstration system operates in a simple pick and place scenario. A human operator points to objects on a table in front of the robot with a particular hand posture, which signals the robot how to pick up the object pointed to, e.g., from the side or from above. The robot fixates the hand, recognizes its posture, and picks up the object in the requested manner. It places the object at location indicated by a second pointing gesture. The system has proven its robustness in numerous demonstrations.[1]

In contrast to almost all systems for hand posture recognition described in the literature, our system performs fairly well in situations where there is no easy way of segmenting the hand from the background. The only other system in the literature attempting to face the difficult challenge of complex backgrounds that we are aware of was presented by Cui and Weng [4]. It recognizes 28 different hand gestures in front of complex backgrounds. It reaches 93.1 percent correct recognition rate, but does not seem to be person independent. Furthermore, it relies on a separate segmentation stage taking 58.3 seconds per image on a SGI INDIGO 2 workstation. It is our conviction that treating segmentation and recognition as separate processing stages is a fundamental misconception and that systems attempting to do so are severely limited. In many situations, the necessary information for correctly grouping an image's pixels into segments corresponding to the scene's objects simply is not there at the level of primitive features.

For future work, we would like to highlight two promising avenues. First, the current system combines collections of relatively low-level features (Gabor jets, local color averages, and colorGabor jets) directly to high-level object representations in the form of a labeled graph. In contrast, object recognition in the primate brain seems to involve computing features at a hierarchy of different complexity levels. A similar hierarchy could be elegantly introduced into our system by using hierarchical graphs as object descriptions. For example one might consider lower-level graphs describing individual fingers, which are themselves the nodes of a higher-order graph describing the whole hand. In this way, correlated geometric distortions could be handled more effectively. Second, for the construction of the model graphs, the proper node positions in a large number of images have to be edited by hand. Trying to overcome this limitation would be a very worthwhile endeavor [10], the ultimate goal being that the system constructs its

1. Video at: http://www.neuroinformatik.ruhr-uni-bocum.de/VDM/research/robotics/contents.html.

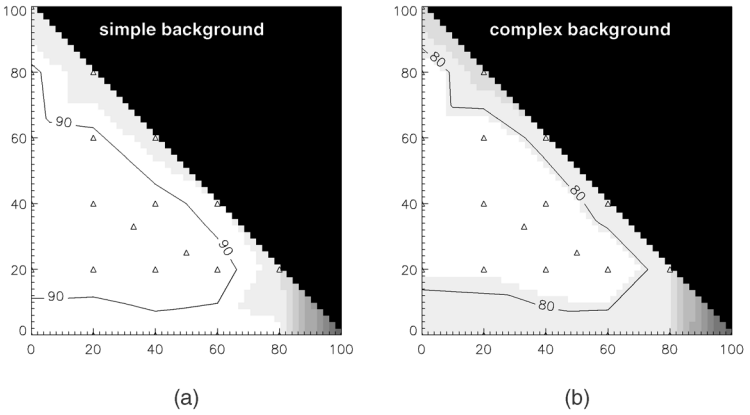


Fig. 5. Dependence of recognition performance on the relative weighting of feature types (left: uniform backgrounds, right: complex backgrounds). Actual sample points are indicated by triangles. On the x-Axis the weight of the color features $w_{col}$ is plotted. On the y-Axis the weight of the Gabor features $w_{gab}$ is plotted. The weight of the colorGabor features is given by $w_{cg} = 100\% - w_{col} - w_{gab}$ so that the weights add up to 100 percent. The gray level indicates recognition performance with light indicating high recognition rates and dark low recognition rates. For both uniform and complex backgrounds, there is a big plateau where recognition is very good. All sample points within the contour lines are above 90 percent (left) or 80 percent (right). Compare with Table 1.

own object representation on the basis of training images, for which only the class of the hand posture is known.

## ACKNOWLEDGMENTS

This work was in part supported by a grant from the German Federal Ministry for Science and Technology (01 IN 504 E9). The first author is currently supported by NIH/PHS research grant P41 RR09283. The authors would like to thank the developers of the FLAVOR software environment, which served as the platform for this research. They also thank three anonymous reviewers for useful comments on an earlier draft.

## REFERENCES

[1] M. Becker, E. Kefalea, E. Maäl, C. von der Malsburg, M. Pagel, J. Triesch, J.C. Vorbrüggen, R.P. Würtz, and S. Zadel, "GripSee: A Gesture-Controlled Robot for Object Perception and Manipulation" *Autonomous Robots*, vol. 6, no. 2, pp. 203-221, 1999.

[2] *Proc. Int'l Workshop Automatic Face—and Gesture—Recognition,* M. Bichsel, ed., June 1995.

[3] Y. Cui and J.J. Weng, "Learning-Based Hand Sign Recognition," *Proc. Int'l Workshop Automatic Face—and Gesture Recognition,* pp. 201-206, June 1995.

[4] Y. Cui and J.J. Weng, "Hand Sign Recognition from Intensity Image Sequences with Complex Backgrounds," *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition,* Oct. 1996.

[5] M.A. Fischler and R.A. Elschlager, "The Representation and Matching of Pictorial Structures," *IEEE Trans. Computers,* vol. 22, no. 1, pp. 67-92, 1973.

[6] W.T. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition," *Proc. Int'l Workshop Automatic Face—and Gesture Recognition,* pp. 296-301, June 1995.

[7] E. Hunter, J. Schlenzig, and R. Jain, "Posture Estimation in Reduced-Model Gesture Input Systems," *Proc. Int'l Workshop Automatic Face—and Gesture Recognition,* pp. 290-295, June 1995.

[8] *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition Killington,* M.E. Kavanaugh, ed., Oct. 1996.

[9] M. Kohler, "Technical Details and Ergonomical Aspects of Gesture Recognition Applied in Intelligent Home Environments," Technical Report 638, Informatik VII, Univ. of Dortmund, Germany, Jan. 1997.

[10] N. Krüger, "Visual Learning with A Priori Constraints," PhD thesis, Technische Fakultät, Univ. Bielefeld, Germany, 1998.

[11] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Computers,* vol. 42, pp. 300-311, 1993.

[12] C. Maggioni, "Gesturecomputer—New Ways of Operating a Computer," *Proc. Int'l Workshop Automatic Face—and Gesture Recognition,* 1995.

[13] B.-W. Min, H.-S. Yoon, J. Soh, T. Ohashi, and T. Ejima, "Visual Recognition of Static Dynamic Gesture: Gesture-Driven Editing System," *J. Visual Languages and Computing,* vol. 10, pp. 291-309, 1999.

[14] C. Nölker and H. Ritter, "Detection of Fingertips in Human Hand Movement Sequences," *Gesture and Sign Language in Human-Computer Interaction,* I. Wachsmuth and M. Fröhlich, eds., pp. 209-218, 1997.

[15] V.I. Pavlovic, R. Sharma, and T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, 1997.

[16] *Proc. Third Int'l Conf. Automatic Face and Gesture Recognition,* P. Storms, ed., Apr. 1998.

[17] *Proc. Fourth Int'l Conf. Automatic Face and Gesture Recognition,* F.M. Titsworth, ed., Mar. 2000.

[18] J. Triesch and C. Eckes, "Object Recognition with Multiple Feature Types," *Proc. Int'l Conf. Artificial Neural Networks, (ICANN '98),* 1998.

[19] J. Triesch and C. von der Malsburg, "Robust Classification of Hand Postures Against Complex Backgrounds" *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition,* Oct. 1996.

[20] C. Uras and A. Verri, "Hand Gesture Recognition from Edge Maps," *Proc. Int'l Workshop Automatic Face—and Gesture Recognition,* pp. 116-121, June 1995.

[21] L. Wiskott, "Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis," *Reihe Physik,* vol. 53, PhD thesis, Verlag Harri Deutsch, Thun, Frankfurt, Main, Germany, 1995.

[22] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.