in: Proceedings of the ICANN 1996, C. v.d. Malsburg,
W. v. Seelen, J.C. Vorbrüggen, B. Sendhoff (eds.),
Springer Verlag, Berlin, Heidelberg, New York, 1996.

# Reconstruction from graphs labeled with responses of Gabor filters*

M. Pötzsch[1], T. Maurer[1], L. Wiskott[1]** and C. v.d. Malsburg[1,2]

[1] Ruhr-Universität Bochum, Institut für Neuroinformatik, 44780 Bochum, Germany
[2] University of Southern California, Dept. of Computer Science and Section for Neurobiology, Los Angeles, CA

**Abstract.** The work presented is part of a larger effort to build a general object recognition system. Objects as well as human faces are represented by graphs labeled with Gabor filter responses. We describe an optimal method to reconstruct images from such graphs. Two examples of how this can be used to analyze the object representation or to compensate for its deficiencies are presented. Since the reconstruction method is formulated generally for an arbitray set of linear filters, it can also be applied to data produced by other systems, artificial or biological.

## 1 Introduction

Our point of departure is an object recognition system [1, 2, 3], which uses Gabor filters as basic features. The region surrounding a given pixel in the image is represented by the responses of a set of Gabor filters of different frequencies and orientations, all centered at the same pixel position. This set of responses is called a *jet*. Objects are represented by graphs whose nodes are labeled by jets, and whose edges describe topographical relations (see Fig. 3a). An object is identified and located by Elastic Graph Matching (EGM), which is a simple algorithmic caricature of Dynamic Link Matching, a neural model based on synchrony coding of feature binding and rapid reversible synaptic plasticity [4]. The system has been successfully applied to face recognition and face segmentation [1, 3], as well as object recognition in complex scenes [5].

Here, we describe how an image can be reconstructed from a graph labeled with jets. In contrast to the method presented in [6], our reconstruction is optimal in the sense that all information is preserved, and it is general, because it is not restricted to rectangular graphs. In Section 2, we describe a reconstruction scheme for an arbitrary set of linear filters and explain certain aspects of it by the example of Gabor filters. In section 3, the reconstruction is applied to labeled graphs, and the results are discussed in the last section.

* Supported by the German Federal Ministry of Science and Technology
** currently at the Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, San Diego, CA

## 2 Reconstruction

A simple model for the cell responses in the visual cortex is the linear receptive field. Thus, the response $J_\nu$ of neuron $\nu$ can be written as

$$J_\nu = (\mathcal{T}I)_\nu = \int I(\mathbf{x})\, p_\nu(\mathbf{x}) d\mathbf{x}, \tag{1}$$

with the shape of its receptive field given by $p_\nu(\mathbf{x})$ and the image $I(\mathbf{x})$ as a stimulus. This operation can also be interpreted as transformation $\mathcal{T}$ of an image. Since the intergral computes a projection of the image onto filter $p_\nu$, $p_\nu$ is also called a *projection function*.

Furthermore, it has been shown that receptive field profiles of *simple cells* in the visual cortex V1 can be modeled by Gabor filters $\psi_{\mathbf{k},\mathbf{x}_0}$ as projection functions [7]:

$$\psi_{\mathbf{k},\mathbf{x}_0}(\mathbf{x}) = \frac{\mathbf{k}^2}{\sigma^2} \exp\left(-\frac{\mathbf{k}^2(\mathbf{x}_0 - \mathbf{x})^2}{2\sigma^2}\right) \exp\left(i\mathbf{k}(\mathbf{x}_0 - \mathbf{x})\right), \tag{2}$$

where $\mathbf{k}$ is the main frequency of the filter and $\mathbf{x}_0$ specifies the location of the filter. The filters have the shape of complex waves (third factor) restricted by a Gaussian envelope function (second factor). Thus, the complex–valued $\psi_\nu$ (with $\nu := (\mathbf{k}_\nu, \mathbf{x}_{0,\nu})$) is composed of an even (cosine-type) and an odd (sine-type) part. The first factor compensates for the frequency-dependent decrease of the power spectrum in natural images [8]. (There is an additional correction for the DC-value of the filters which is not shown here.)

Since the transformation $\mathcal{T}$ is linear, the optimal reconstruction of the image from the values $J_\nu$ is linear as well and given by a well-known concept of linear algebra:

$$I^{\mathcal{R}}(\mathbf{x}) = \mathcal{R}\mathbf{J} = \sum_\nu J_\nu b_\nu(\mathbf{x}), \tag{3}$$

where $\mathcal{R}$ symbolizes the reconstruction operation and $b_\nu(\mathbf{x})$ are appropriate basis functions.

If the projection functions were orthogonal and normalized ($\int \overline{p}_\nu(\mathbf{x})\, p_\rho(\mathbf{x}) d\mathbf{x} = \delta_{\nu\rho}$) the basis functions would simply be $b_\nu(\mathbf{x}) = \overline{p}_\nu(\mathbf{x})$. Since they are not orthogonal in case of Gabor functions, their affinity must be taken into consideration by using a particular linear combination of the projection functions:

$$b_\nu(\mathbf{x}) = \sum_\rho \left(P^{-1}\right)_{\nu\rho} \overline{p}_\rho(\mathbf{x}), \quad \text{with} \quad P_{\nu\rho} := \int \overline{p}_\nu(\mathbf{x})\, p_\rho(\mathbf{x}) d\mathbf{x}. \tag{4}$$

The matrix coefficients $P_{\nu\rho}$ are dot products of the projection functions. In case of Gabor filters, these products can be obtained analytically.

Gabor filters can also be used as basis functions ($b_\nu = \psi_\nu$) [9], in which case the values $\tilde{J}_\nu$ have to be computed by minimizing $\|I(\mathbf{x}) - \sum \tilde{J}_\nu \psi_\nu\|^2$. Since the $\tilde{J}_\nu$ are defined by the reconstruction formula, they differ from those computed by

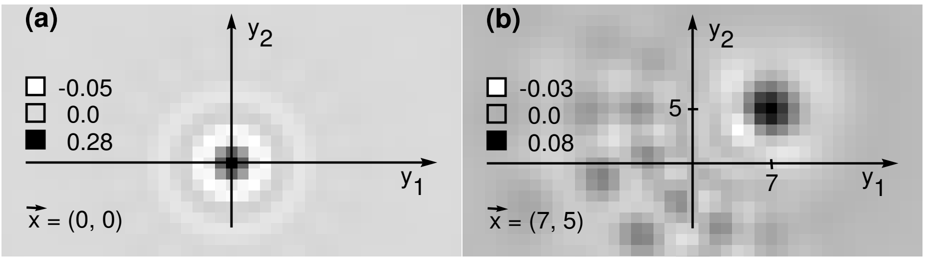


**Fig. 1.** The projector $\mathcal{P} = \mathcal{RT}$ shown in this figure was derived from a set of 32 Gabor filters as projection functions all centered at the same position $\mathbf{y}_0 = (0, 0)$. The figure shows how the original image contributes to the reconstructed image $I^{\mathcal{R}}$ at a certain point $\mathbf{x}$ ((0,0) in (a) and (7,5) in (b)). E.g., in (b), the contributions come mainly from the gray values around the corresponding point $\mathbf{y} = (7, 5)$ but also from the gray values at all other points within the extent of the Gabor filters. Formally speaking, the kernel $c(\mathbf{x}, \mathbf{y})$ of the projection process $I^{\mathcal{R}}(\mathbf{x}) = \mathcal{P}I(\mathbf{y}) = \int c(\mathbf{x}, \mathbf{y})I(\mathbf{y})d\mathbf{y}$ is shown: (a) $c((0, 0), \mathbf{y})$, (b) $c((7, 5), \mathbf{y})$. (The wave vector $\mathbf{k} = (k, \varphi)$ of the applied Gabor filters takes on four different frequency values $k$ and eight different orientation values $\varphi$.)

Eq. (1) and the image can be reconstructed directly by $I^{\mathcal{R}} = \sum \tilde{J}_\nu \psi_\nu$. However, in the sense of the findings in [7], this approach cannot serve as a model for the cell responses in sensory cortex.

The reconstruction defined by Eqs. (3) and (4) is perfect in the sense that the amount of information in the reconstructed image $I^{\mathcal{R}}(\mathbf{x})$ is equal to that of the transformed data $\mathbf{J}$. That also implies that the values $\mathbf{J}$ can be identically recalculated by transforming $I^{\mathcal{R}}$, i.e., the compound operation $\mathcal{TR}$ is the identity operator[1].

The compound operator $\mathcal{P} = \mathcal{RT}$ is called a *projector* and it projects images into the "space" of images which can be represented by the chosen set of projection functions $p_\nu$. $\mathcal{P}$ satisfies the condition for being a projector ($\mathcal{P}^2 = \mathcal{P}$), for $\mathcal{TR}$ equals the unity operator. Figure 1 shows such a projector for the set of Gabor filters all centered at the same positions.

The reconstruction formulas presented can only be applied to a linear independent set of projection functions, because otherwise the determinant of $P$ vanishes and $P$ cannot be inverted. In other words, the transformed data of a linear dependent set of $p_\nu$ is redundant. This means at least one $p_\mu$ can be

[1] The compound operation $\mathcal{TR}$ is the identity operator, for

$$(\mathcal{TR}\mathbf{J})_\mu = \int I^{\mathcal{R}}(\mathbf{x})\, p_\mu(\mathbf{x}) d\mathbf{x} = \sum_\nu \left( \int b_\nu(\mathbf{x})\, p_\mu(\mathbf{x}) d\mathbf{x} \right) \cdot J_\nu$$

$$= \sum_\nu \left( \sum_\rho (P^{-1})_{\nu\rho} \underbrace{\int \overline{p}_\rho(\mathbf{x})\, p_\mu(\mathbf{x}) d\mathbf{x}}_{= P_{\rho\mu}} \right) \cdot J_\nu = \sum_\nu \delta_{\nu\mu} J_\nu = J_\mu .$$

represented by a linear combination of (a subset of) the other projection functions and thus, the corresponding value(s) $J_\mu$ can be omitted for the purpose of reconstruction without loss of information.

## 3 Application to labeled graphs

The object recognition system described in [1, 2, 3] represents objects as well as human faces by graphs whose nodes are labeled by jets. The amount of information contained in such a representation can be visualized by reconstruction. Eq. (3) and (4) provide an optimal reconstruction for any arbitrary set of linear filter responses. However, since the coefficients most important for reconstructing a given image point are contained in the jet nearest to that point, and the nodes are sufficiently far apart, the reconstruction can be approximated by using the components of just that jet independent of others. This has the advantage that only a small set of basis functions has to be computed, which can be applied to all nodes (see Fig. 3b as an example). Each so called *local* reconstruction is restricted to a Voronoi area around its location, because no interaction between the basis functions of adjacent jets is taken into account. This approximated reconstruction differs little from the exact one, and can be computed much faster (in about 4 sec. –3.5 sec. to compute the basis functions and 0.5 sec. to apply them– compared to about 1 hour needed for the exact version on a Sparc 20).

Visualization of the information stored in a labeled graph may help expose deficiencies in the object representation. It may even help to compensate for these deficiencies. As an example, consider the nodes located near the outline of an object. Their jets are not only influenced by the object itself but also by the background, because of the spatial extent of the filters. A simple linear tranformation on the image, namely the multiplication with a 2-dimensional Heavyside-function properly located, could suppress the background locally. Let us denote this transformation by $\Theta$. In order to perform the corresponding transformation directly on a jet, one simply has to concatenate reconstruction, background suppression, and retransformation to form a single linear transformation $\mathcal{L} = \mathcal{T}\Theta\mathcal{R}$, which can then be applied to a jet (see Fig. 2c) [10, 11].

For the task of finding a face in an image without attempting to identify the person, a graph containing the knowledge about several other faces is used [3]. For this purpose, each node is labeled with a *bunch* of jets, each jet extracted from the image of a different person but at the same facial landmark (such as the tip of the nose or the left eye). Each bunch may, for example, contain jets from 100 different persons. Thus, it covers a wide variety of shapes for a single landmark. Such a graph is called a *bunch graph*.

During the process, in which the bunch graph is matched onto an image with an unknown face, the jet fitting best to the presented face is automatically selected in each bunch. A reconstruction of an image from these jets fitting best leads to a *phantom face*; a face created from transformed data taken from several different persons (see Fig. 3c). Phantom faces have already been introduced in [3], where they have been used for the determination of facial attributes. How-

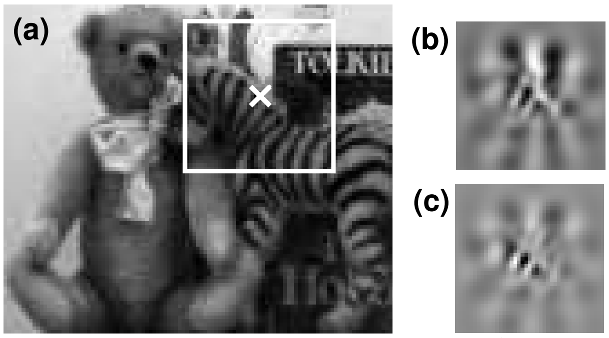


**Fig. 2.** Background suppression: **(a)** Scene of toys (a zebra in front of a bear and a book). **(b)** Reconstruction from the jet marked in (a). **(c)** Reconstruction of a modified version of the jet marked in (a). Knowing the outline of the zebra the considered jet can be linearly transformed to suppress the influence of the background.

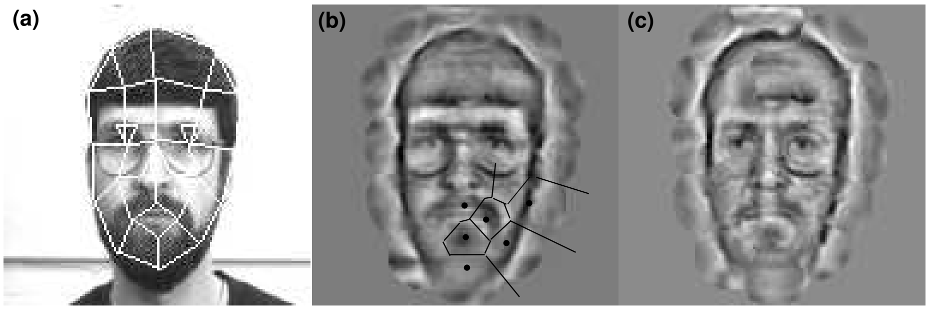


**Fig. 3.** Reconstruction from a labeled graph: **(a)** A face with a graph. **(b)** Reconstruction from the labeled graph taken from the image in (a). The reconstruction is achieved by a local reconstruction of each jet restricted to a Voronoi area around its location. Some of these areas are indicated by lines. **(c)** The phantom face: a reconstruction of the jets in a bunch graph which fitted best to the image (a). The bunch graph contained jets from about 100 different persons (not including the one shown in (a)). **(b,c)** As the DC-part of the chosen filters vanishes, the absolute grey levels are missing in the reconstructed images.

ever, they have not been created from jets themselves (by reconstructing them). Instead, patches of the corresponding original images have been used.

## 4 Discussion

Sensory data, e.g., visual or auditory, are preprocessed in the cortex as in many artificial systems. Usually the chosen filters applied for preprocessing are designed to simplify the perceptual task by transforming the original data appropriately. Visualization of the remaining amount of information after preprocessing is a great help in understanding and analyzing an artificial or even biological system. We have presented a reconstruction method (optimal in the sense that all information is preserved) for data which have been computed by a set of projection functions (or linear filters). The method has been applied to an arbitrary set of Gabor filters, and an approximating variant of it to labeled graphs. Two aspects of an object recognition system have been analyzed by visulisation: the background problem concerning the representation of objects, and the detection of faces. In the latter case, phantom faces have been created to show what information is used to find a face without regard to identity.

## References


1. M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R.P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.*, 42(3):300–311, 1993.
2. J. Buhmann, M. Lades, and C. v.d. Malsburg. Size and distortion invariant object recognition by hierarchical graph matching. In *Proc. of the IJCNN International Joint Conference on Neural Networks*, pages II 411–416, San Diego, June 1990. IEEE.
3. L. Wiskott, J.M. Fellous, N. Krüger, and C. v.d. Malsburg. Face recognition and gender determination. In *Proc. of the International Workshop on Automatic Face- and Gesture-Recognition, IWAFGR 95*, Zurich, June 1995.
4. C. v.d. Malsburg. The correlation theory of brain function. Internal report, 81-2, Max-Planck-Institut für Biophysikalische Chemie, Göttingen, FRG, 1981. Reprinted in E. Domany, J.L. van Hemmen, and K.Schulten, editors, *Models of Neural Networks II*, pages 95–119. Springer, Berlin, 1994.
5. L. Wiskott and C. v.d. Malsburg. A neural system for the recognition of partially occluded objects in cluttered scenes. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(4):935–948, 1993.
6. R.P. Würtz. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition*, volume 41 of *Reihe Physik*. Verlag Harri Deutsch, Frankfurt a. Main, 1995. PhD thesis.
7. J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.
8. D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Optical Society of America*, 4(12):2379–2394, 1987.
9. J.D. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36:1169–1179, 1988.
10. M. Pötzsch, N. Krüger, and C. v.d. Malsburg. Improving object recognition by transforming Gabor filter responses. *Network: Computation in Neural Systems*, 7(2), 1996.
11. M. Pötzsch. Die Behandlung der Wavelet-Transformation von Bildern in der Nähe von Objektkanten. Internal Report (IR-INI) 94-04, Institut für Neuroinformatik, Ruhr-Universität Bochum, 1994.


This article was processed using the LaTeX macro package with LLNCS style