Applying Dynamic Link Matching to Object Recognition in Real World Images *

Wolfgang Konen,† Jan C. Vorbrüggen
Institut für Neuroinformatik, Ruhr-Universität Bochum, FRG

Abstract

We apply the dynamic link matching algorithm to object recognition in gray level images. The algorithm is able to map from one view of an object to different—e. g., translated, rotated, or mirror-reflected—views, being at the same time tolerant of small distortions. A sparse representation (10%) of the image data is used as a boundary condition for a self-organizing mechanism which performs the object match within a modest number of iterations ($\sim 10^2$). The mechanism can be derived from local neural dynamics [1].

Introduction. Invariant object or pattern recognition is considered to be one of the hardest problems in vision research. This is mainly due to the large variety of transformations which the image of an object can undergo while the object is still being perceived as the same by our visual system. Neural networks offer promising approaches to this problem since they can deal well with the variability of natural scenes as far as incomplete information or noise (e. g., associative memory [2, 3]) or distortions (e. g., in handwritten digits [4, 5]) are concerned. They cannot, however, easily perform geometric transformations of an object's image. Graph matching schemes [6, 7] offer the ability to match transformed (e. g., translated, rotated, or stretched) patterns, since the quality of the match is only evaluated in terms of *relative* and not of absolute arrangement of features within the patterns. Graph matching has been applied successfully to the recognition of facial images [8].

An algorithm for finding the best graph match may, however, be hindered by energy barriers between the initial estimate and the globally optimal solution. Consider for example the case of mirror reflection, where the whole graph has to be flipped in order to achieve the best match. In this paper, we want to overcome these limitations by connecting the graph nodes with dynamic links which code for the probability that two nodes are matching counterparts. This frees the system from specifying an initial estimate. An appropriate self-organizing mechanism then allows it to search in parallel among many possible transformations, even those (like mirror reflection) which cannot be reached from the identity mapping by successive infinitesimal transformations. The principle of the dynamic link matching (DLM) algorithm and a possible implementation by neural dynamics have been reported previously [1, 9] for the case of symmetry recognition in synthetic data. Here, we want to demonstrate its applicability to realistic gray level images.

Dynamic Link Matching (DLM). In order to allow for a flexible match between images, we will not work on the pixel level but use a sparse graph of nodes within each image. Each node carries local feature information obtained by preprocessing the image with localized filters (see below). For simplicity, graphs will be rectangular grids (see Fig. 1), but the algorithm could use arbitrary planar graphs as well.

In all object recognition tasks we have an image graph I and one or several model graphs M^i, $i = 1, \ldots, N$. All nodes $a \in I$ and $b^i \in M^i$ are connected by dynamic links $J^i_{ba} \in [0, 1]$, where $a, b, \ldots$ specify two-dimensional locations in a planar graph.

A high value of J^i_{ba} denotes a high confidence that nodes a and b^i are matching counterparts, a low value denotes low confidence. The effective connectivity $J^i_{ba}T^i_{ba}$ between two nodes a and b^i is the product of the dynamic link and the local feature similarity $T^i_{ba} \in [0, 1]$. Again, a high value of T^i_{ba} denotes strong similarity of the features attached to a and b^i, while a low value denotes dissimilarity.

Within this system architecture the task of object recognition can be formulated as a self-organizing mechanism selectively strengthening those links J^i which connect nodes in I with their corresponding nodes in M^i. This mechanism uses activations of subgraphs of nodes—termed *blobs* in this work—within each graph. In a truly neural system, the activation would be mediated by lateral connectivity within each graph layer [1]. For simplicity, we replace the lateral dynamics with an algorithm where nodes within a blob $B(a_c)$ are set to an active state. $B(a_c)$ is a square of $m \times m$ nodes with center at a_c. (If a part of the square crosses the border of the graph, it is truncated.)

*Supported by a grant from the German Federal Ministry for Science and Technology (413-5839-01 IN 101 B/9).

†Email address: wolfgang@neuroinformatik.ruhr-uni-bochum.de

The DLM algorithm. Initialize the dynamic links with $J^i_{ba} = T^i_{ba} / \sum_{a'} T^i_{ba'}$.

(i) Randomly choose a center a_c among the nodes of I and activate the nodes $a \in B(a_c)$. The effective connectivity from I to each of the model graphs M^i leads to an input activity at each node b^i:

$$I(b^i) = \sum_{a \in B(a_c)} J^i_{ba} T^i_{ba}. \tag{1}$$

(ii) Find the center position b^i_c at which the blob $B(b^i_c)$ will have its largest overlap with the input activity:

$$\sum_{b \in B(b^i_c)} I(b) = \max_{i, \beta^i \in M^i} \sum_{b \in B(\beta^i)} I(b) \tag{2}$$

and activate the nodes $b \in B(b^i_c)$.

(iii) Update the dynamic links between active nodes such that the total link strength converging on each cell b is kept constant:

$$J^i_{ba} = \frac{J^i_{ba} + \epsilon J^i_{ba} T^i_{ba}}{\sum_{a'} (J^i_{ba'} + \epsilon J^i_{ba'} T^i_{ba'})} \qquad \forall a \in B(a_c), b \in B(b^i_c). \tag{3}$$

(iv) Reset all active nodes and proceed with step (i).

The algorithm has three free parameters: the update parameter ϵ and the sizes m_I and m_M of the blobs in I and M^i, respectively.

Image Preprocessing. As feature input to the DLM algorithm we use a simple preprocessing stage based on localized filtering of camera images. Two requirements have to be met by the filters: (i) The absolute placement of the rectangular grid on the image should be irrelevant, i.e., the feature information should be insensitive against a shift of the grid by at most half the distance of neighboring nodes. This implies sufficient spatial extent of the filters. (ii) Since we want to match rotated and reflected images, the local features must be invariant against local rotation and reflection.

We extract approximately rotationally invariant features $c_k(a)$ by convolving the image $\mathcal{I}$ at node location $\vec{r}_a$ with a family of Gabor wavelets $g_{k\phi}$ [8]:

$$c_k(a) = \sum_\phi |\, (g_{k\phi} * \mathcal{I})(\vec{r}_a) \,|, \quad g_{k\phi}(\vec{r}) = 2\pi G(\vec{r}, \sigma/k)(e^{i\vec{k}\vec{x}} - e^{-\sigma^2/2}) \tag{4}$$

with normalized Gaussians $G(\vec{r}, s)$ of width s and $\sigma = 2\pi$. The sum runs over 8 different orientations ϕ and we use 7 frequency levels with half-octave spacings starting from the lowest frequency $k = 3\pi/32$. We note in passing that we also tested DoG (Laplace) filters as an alternative to the Gabor wavelets and found qualitatively the same results.

Coding the image information in terms of the coefficients $c_k(a)$ instead of the original 128×128 pixels means a large information reduction due to the sparseness of the grid (down to 11% and 2.7% for a 16×16- or 8×8-grid, resp., using 7 frequency levels). From this sparse representation we compute the local feature similarity of nodes a and b:

$$T_{ba} = \exp\Big[-\alpha \sum_k (c_k(a) - c_k(b))^2 \Big]. \tag{5}$$

Note that the DLM algorithm is independent of the specific kind of preprocessing used: Any kind of feature can be incorporated once a suitable similarity function T_{ba} has been defined.

Object recognition. We have tested the DLM algorithm with different recognition tasks:

Single object recognition: Given two different views of the same object (Fig. 1), the task is to find the best match from the graph of Fig. 1A onto Fig. 1B. After 100 iterations the dynamic links have established the correct transformation as can be seen from the smoothed maximum link map shown in Fig. 1C. Minor imperfections of the map are mainly due to the coarseness of the grids and to the use of filters with low spatial resolution. Note that node pixels in grid 1A in general do not have an exact correspondence among the node pixels of grid 1B.

Object discrimination and patch identification: The DLM algorithm is also able to discriminate an object I among a number of model choices $M^i, i = 1, 2, 3$ (Fig. 2), or to locate different patches

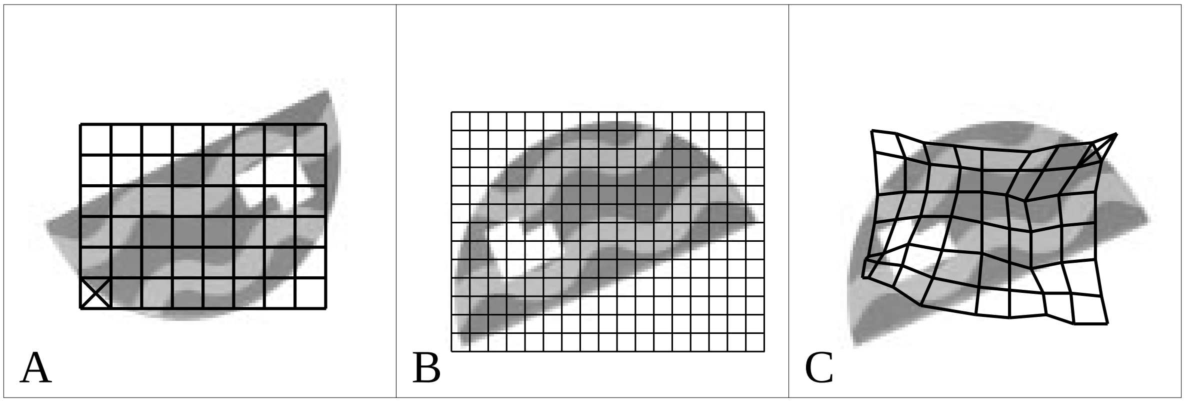

Figure 1: **Single object recognition.** The task is to find an estimate of the transformation (here: a 180°-rotation) which matches image **A** onto image **B**. Model graph M and image graph I are shown overlayed in **A** and **B**, resp. After 100 iterations of the DLM algorithm the dynamic links attached to M have grown into the correct regions of image **B**. This result is demonstrated in **C** as the *maximum link map* which is defined as follows: For each node $b \in M$ we mark the location of node $a \in I$ whose link J_{ba} is largest among all links converging on b. For better visualization the resulting map of locations is smoothed with a window of 3×3 nodes (the nearest neighbors of b) giving the map shown in **C**. One corresponding corner of graph and map is marked by an '×' showing that the map is rotated correctly. Parameters are $\epsilon = 1.0, m_M = 4, m_I = 7, \alpha = 0.8$.

M^i of an object (e. g., the head region of the elephant in Fig. 3) within a different image of this object.

Conclusion. The DLM algorithm has sufficient flexibility to deal with the noisy data of gray level images. Various object matching tasks requiring different invariance transformations have been demonstrated. The advantage of the system is that the *same* algorithm can be used for searching in parallel among different image transformations, as well as solving different recognition tasks. Currently, objects are represented by planar graphs, limiting the system to map only views of an 3D-object from similar viewpoints. Object discrimination has been investigated only for a small number (3) of objects; it remains to be shown that the algorithm can also cope with a larger number of objects. Of course, this algorithm is only a simple sketch of a more mature system, because it uses only coarse image features on fixed rectangular grids which are by no means adapted to the object information. For a more realistic neural system the grids should be replaced by arbitrary planar graphs of salient nodes with activation spreading along the edges of these graphs. Further work will go into this direction.

Acknowledgements. We would like to thank C. v. d. Malsburg, L. Wiskott and R. Würtz for helpful discussions and K. Behrmann and A. Weitzenfeldt for valuable support in programming.

References

[1] W. Konen and C.v.d. Malsburg. Learning symmetries from single examples in the dynamic link architecture. *To appear in: Neural Computation*, 1993.

[2] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558, 1982.

[3] T. Kohonen. *Associative Memory*. Springer, Berlin, 1977.

[4] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989.

[5] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition. *Biological Cybernetics*, 36:193, 1980.

[6] C.v.d. Malsburg. Pattern recognition by labeled graph matching. *Neural Networks*, 1:141–148, 1988.

[7] E. Bienenstock and R. Doursat. Elastic matching and pattern recognition in neural networks. In L. Personnaz and G. Dreyfus, editors, *Neural Networks: From Models to Applications*. IDSET, Paris, 1989.

[8] 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. Comp.*, March 1993.

[9] W. Konen and C.v.d. Malsburg. Unsupervised symmetry detection: A network which learns from single examples. In I. Aleksander, editor, *Proceedings of the International Conference on Artificial Neural Networks*, pages 121–125. North-Holland, Amsterdam, 1992.

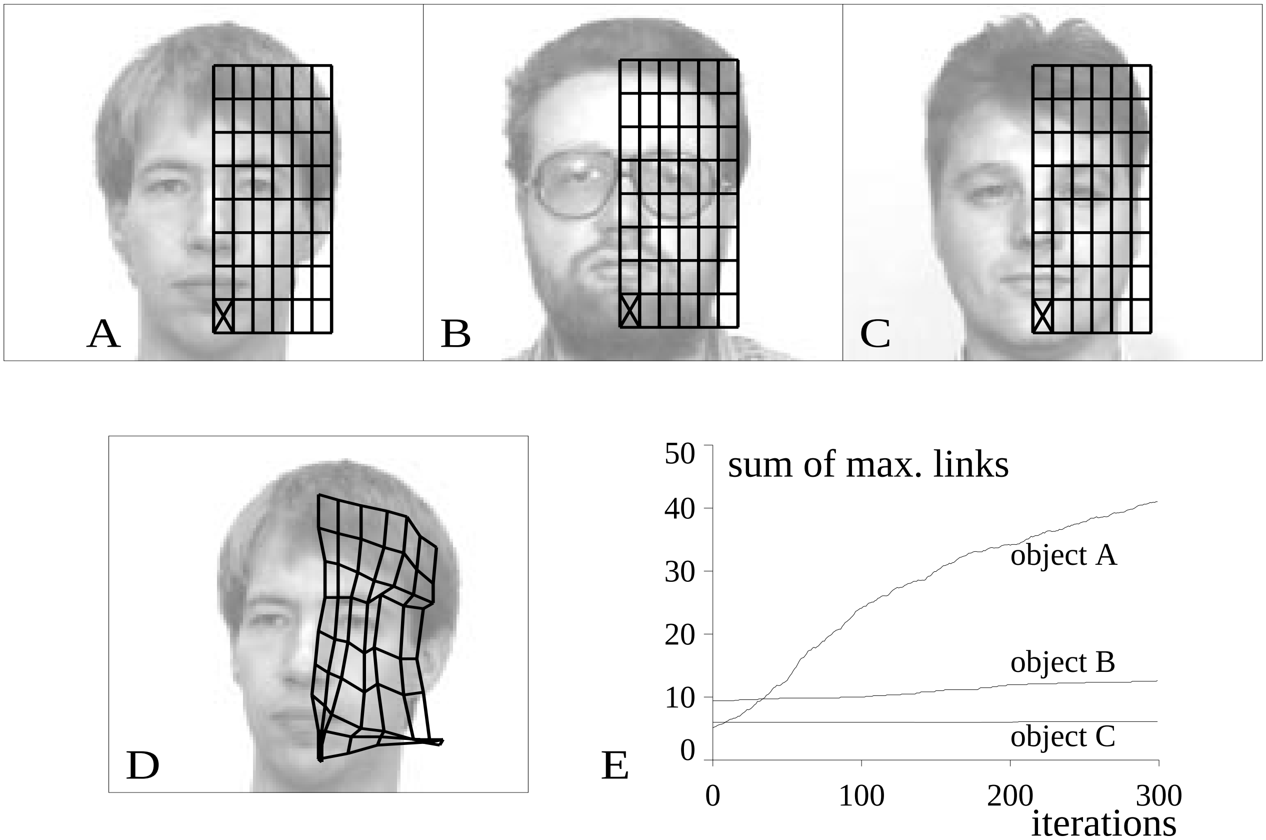

Figure 2: **Object discrimination.** The DLM algorithm allows fast discrimination to which stored model M^i (**A, B, C**) a different image **D** of object **A** belongs. **E**: The best matching M^i can be easily read off from the sum of the maximal links converging on objects **A, B** and **C**, resp. Note that initially object A has the lowest sum, i. e., the discrimination could not have been obtained from the similarity function T^i_{ba} alone. Parameters are $\alpha = \epsilon = 1.0, m_I = m_M = 4$ for the first 100 iterations and $m_I = m_M = 3$ thereafter.

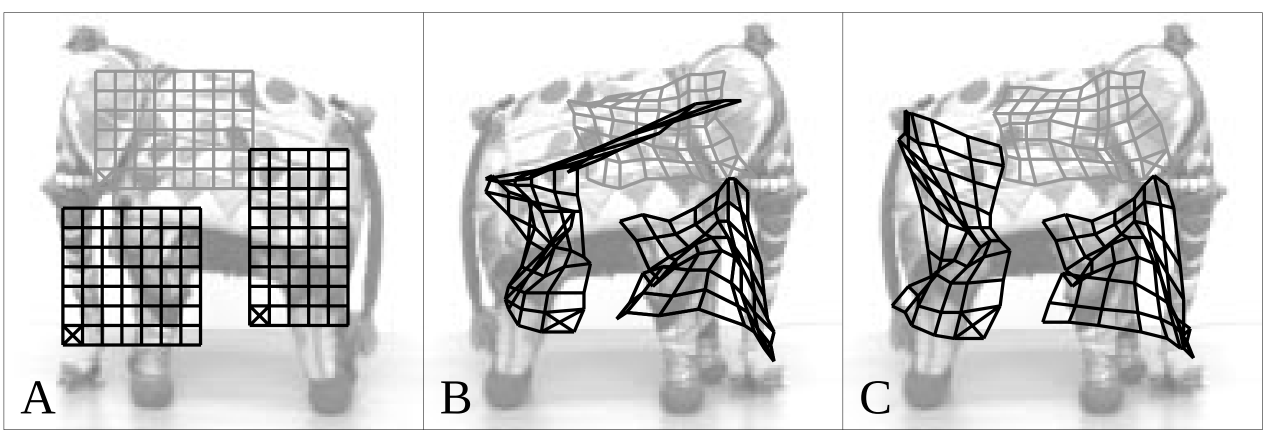

Figure 3: **Patch identification.** This figure demonstrates the ability of the system to correctly locate and orientate the three patches M^i defined by the grids in **A** within a mirror-reflected image I shown in **B**. Note that both images additionally differ in a slight rotation in depth of the elephant. Image I was covered by a 16×16-grid (not shown). **B** and **C** show the maximum link maps for the different patches obtained after 300 and 600 iterations, resp. Parameters are $m_I = m_M = 5, \alpha = 0.5$ and a varying update parameter ϵ, starting with 0.1 and doubled every 150 iterations.