/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

ELSEVIER

Journal of Neuroscience Methods 00 (2001) 000–000

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

Efficient evaluation of serial sections by iterative Gabor matching

Peter König [a,*], Christoph Kayser [a], Vincent Bonin [a], Rolf P. Würtz [a,b]

[a] *Institute of Neuroinformatics, ETH/University Zürich, Winterthurerstraße 190, 8057 Zürich, Switzerland*
[b] *Institute of Neuroinformatics, Ruhr-University Bochum, 44780 Bochum, Germany*

Received 5 March 2001; received in revised form 11 June 2001; accepted 13 July 2001

Abstract

Evaluation of electron microscopic images of serial sections is a time-consuming process requiring a high level of expertise. Here we present an algorithm to ease and accelerate this process. It is a modification of an algorithm successfully used in computer vision for object recognition. However, rather than recognising individual structures, we estimate the spatial mapping of a whole section onto the consecutive one. This mapping is used to transfer labelled information of the very first section, e.g. a classification by a human expert of different visible structures, onto structures visible in the next section. We investigate its performance on an artificially constructed benchmark as well as on real electron microscopic samples taken in primary visual cortex and demonstrate its potential for dramatically facilitating the evaluation process of serial sections. © 2001 Published by Elsevier Science B.V.

Keywords: Electron microscopy; Wavelet description; 3D Reconstruction; Morphology

1. Introduction

Vision research is a highly active and fertile area investigating, among others, higher order processes such as object recognition and pattern classification. Humans excel at such tasks and no artificial system achieves a performance that comes even close to that of biological systems. A notable example is the evaluation of serial sections: A three-dimensional (3D) sample is cut into thin two-dimensional slices for microscopic evaluation and the original 3D shape reconstructed. This widely used technique is of fundamental importance to elucidate the anatomy of the nervous system (for example, Anderson et al., 1999; Hack et al., 1999; Rockland and Knutson, 2000; Gibbins et al., 2000; Fukuda and Kosaka, 2000). For this reconstruction, structures visible in each slice are matched to their counterparts in neighbouring sections. Indeed, consecutive sections look very similar and the matching process itself is not difficult as such. Nevertheless, no general solution to this problem is available and, as a consequence, it is the task of the scientist to match structures in hundreds and thousands of serial sections.

We address this problem by modifying a method used for object recognition. In order to recognise an object visible in an image, parts of the image must be compared with a stored representation of known objects. This is only possible after the 'correspondence problem' has been solved: In order to compare the image features correctly, pairs of image points that are projections of the same physical point have to be identified. This process is also called 'matching' and yields a mapping from points in the stored image to points in the image to be analysed. A particularly successful example of such a recognition system is the dynamic link matching approach (von der Malsburg, 1988; Lades et al., 1993; Würtz, 1997). This algorithm creates correspondence mappings from each model stored in a database to the image to be analysed on the basis of feature similarity and topological consistency. Using these mappings, the similarities between corresponding features are added up to produce a similarity measure between each model and the image. Then, the model with the highest similarity is assumed to show the object present in the image. In the recognition process, the mappings between models and image only serve the purpose of comparing the correct features, and are of no interest themselves. In the current context, automatic recognition of structures in electron

* Corresponding author. Tel.: +41-1-635-3060; Fax: +41-1-635-3053.
E-mail address: peterk@ini.phys.ethz.ch (P. König).

0165-0270/01/$ - see front matter © 2001 Published by Elsevier Science B.V.
PII: S0165-0270(01)00439-3

UNCORRECTED PROOF

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

microscopic images is a very difficult problem: The variability of structures is very large and it is far from clear which structures are relevant to the research question at hand. Therefore, we leave the task to label important structures, e.g. dendrites, axons, spines etc., in the first slice to the scientist and use only the matching process to transfer this information to a subsequent section. This process can be iterated following visible structures over several sections, while newly appearing structures have to be labelled at least once in the section they appear the first time. Thus, the algorithm holds the promise to considerably reduce the amount of work necessary to evaluate subsequent sections of a series.

2. Methods and results

2.1. Electron microscopy

We have evaluated electron microscopic photographs taken for a different study, kindly provided by J.C. Anderson and K.A.C. Martin. The investigated tissue was taken from cat primary visual cortex (area 17, layer 3) and processed according to standard protocols (Anderson et al., 1994). Briefly stated, neurons were labelled intracellularly with horseradish peroxidase (HRP) using an in vivo preparation. At the end of the experiment, the animal was sacrificed and perfused transcardially (1% paraformaldehyde and 2.5% glutaraldehyde in 0.1 M phosphate buffer). The cerebral cortex was blocked and serially sectioned using an Oxford vibratome in the coronal plane at a thickness of 80 μm. Further treatment of the sections for light microscopy included incubation, nickel/cobalt intensification, treatment with 1% osmium tetroxide in 0.1 M phosphate buffer, dehydration with alcohols (the 70% step containing 1% uranyl acetate) and propylene oxide and flat embedding in Durcupan ACM resin (Hanker et al., 1977; Adams, 1981; Anderson et al., 1994). After light microscopic evaluation, selected portions were re-embedded and serially sectioned at 70 nm thickness. These ultrathin sections were collected on Pioloform-coated single slot copper grids and photographed at 21 000 magnification. For the present project, glossy prints have been digitised at a resolution of 6 nm per pixel. Thus, the distance between two consecutive images corresponds to a spatial distance of about 12 image pixels.

2.2. Image representation

The mathematical analysis of the images is based on a representation by Gabor wavelets (von der Malsburg, 1988; Lades et al., 1993; Würtz, 1997). As a first step, convenient representations of the images are computed. Images are roughly aligned manually and portions of 512×512 pixel size are cropped. Features at different spatial scales and orientations are extracted by convolution of the images with a set of two Gabor kernels. These kernels are planar sinusoidal waves whose amplitude is modulated by a Gaussian envelope. The envelope has a width that is inversely proportional to the spatial frequency. Thus, depending on the ratio of wavelength to the width of the Gaussian envelope, a varying number of ripples is obtained. The parameters used in this project led to wavelets with two visible maxima. For computational efficiency, convolution is performed in frequency space. One of the functions defining the kernel is symmetric with respect to the Gaussian envelope (cosine phase), whereas the other is anti-symmetric (sine phase). The convolution results for both kernels are squared, summed, and subjected to a square root. Thus, at each location, the result is a vector of positive numbers, one for each combination of orientation and spatial frequency. The values can be interpreted as the degree to which given Gabor functions are contained in the image at each location. As sine and cosine contributions are summed, the result is not sensitive to the position (phase) of the planar wave within the Gaussian envelope. This Gabor wavelet decomposition allows joint localisation of features in the spatial and the frequency domain.

Each Gabor kernel is characterised by its spatial frequency, orientation and width. In the work presented here, we use three different spatial frequencies (with wavelengths of 64, 32, and 16 pixels, respectively), and eight different orientations (spaced by 22.5°). The information contained in the Gabor representation can be visualised by reconstruction of the image. Fig. 1A shows the sampling grids used at low spatial frequencies (large squares), intermediate (small squares) and high spatial frequencies (dots). High spatial frequency information has to be sampled at a higher density compared with lower spatial frequencies. Support points where the kernel used for convolution extends beyond the image boundary are excluded. For the matching process, a twofold oversampling, as compared with the minimum density set by the sampling theorem, turned out to be useful (Fig. 1A). Fig. 1B shows the reconstruction from these three spatial frequencies.

2.3. Matching features

Based on the Gabor representation, the matching process constructs a mapping between the two images. We name one of the images the 'model', and the other image the 'target'. We search for the location in the target image corresponding to a particular point in the model image.

The algorithm proceeds in several steps, using information from higher and higher spatial frequencies.

UNCORRECTED PROOF

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

Firstly, model and target representations are restricted to the lowest spatial frequency sampled on a single coarse grid covering the entire image (Fig. 1C). We compute the multidimensional equivalent of a cross-correlation function between the Gabor coefficients from the model and the target image: The coefficients for different orientations of the model are multiplied pointwise with the respective coefficients of the target image. All products are then summed and normalised by the product of the lengths of both feature vectors.

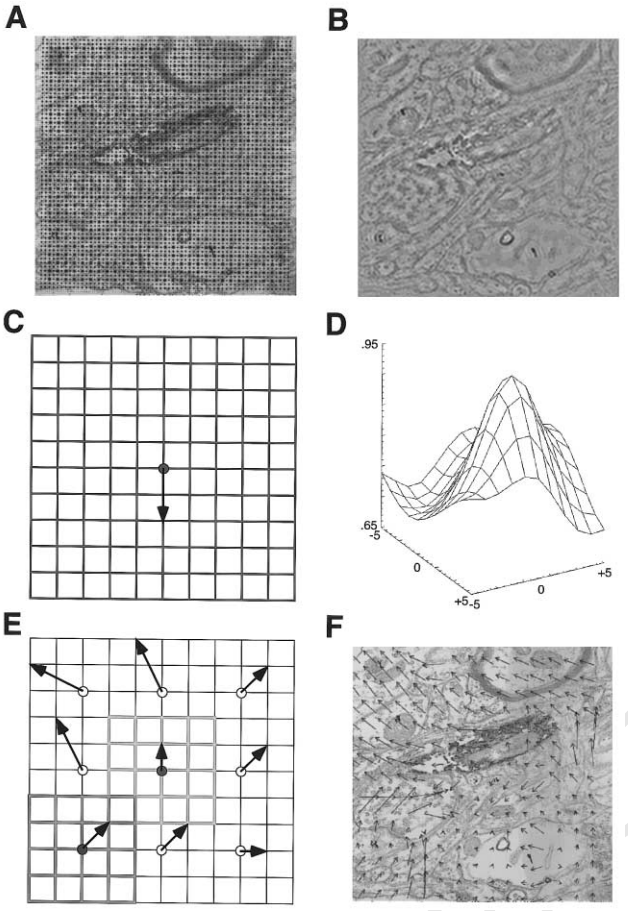

Fig. 1. Wavelet representation of electron microscopic images and the matching process. (A) A section of 512×512 pixels cut out of an image taken in layer 3 of cat primary visual cortex (area 17). Sampling points used for wavelets of low (+), intermediate (○) and high (•) spatial frequencies. (B) Reconstruction of the image with the full set of Gabor wavelets. (C) Amplitudes of the lowest spatial frequency located on a coarse grid in the model image are shifted (arrow) to achieve optimal fit with the respective grid of the target image. (D) The correlation of model and target grid is shown for different shifts. A shift by 2 units in the vertical direction and no shift in the horizontal direction results in the best fit. Note that the global maximum is well defined and local minima occur only at the boundary of the parameter range investigated. (E) Subgrids, which may partly overlap, are shifted (arrows) to improve the local fit between model and target. Note that possible shifts are limited by the spatial resolution of the grid. (F) The mapping resulting from the iterative procedure. Each arrow indicates the mapping of a point in the model to the target image.

Table 1

Iteration	Spatial frequency	Number of grids	Grid size	Range of shift
1	1	1	13	11
2	1	9	9	11
3	2	25	15	11
4	3	100	19	21

This normalisation allows comparing structures independent of the total contrast. For example, in Fig. 1A the right half of the image has considerably lower contrast than the left, which also shows in the reconstruction. For the matching, however, this difference in contrast level is largely ignored. Since all vector components are positive, this measure can not be smaller than 0 and reaches a maximum of 1 if the vectors are identical up to a common factor. The grid in the model is shifted with respect to the target grid and the process is repeated. Due to the restriction to (and the appropriate choice of) the lowest spatial frequency, the cross-correlation function obtained is smooth and exhibits a single maximum (Fig. 1D). In contrast to this, computing a cross-correlation of the brightness patterns of the images directly would result in rapidly oscillating correlation functions and multiple local maxima. The maximum at $(0, -2)$ in Fig. 1D indicates that the low spatial frequencies in the two images match best if the model is shifted downwards by two grid units. The resulting mapping is a translation in space.

In the next step, the mapping is improved by individually matching a set of smaller overlapping subgrids (Fig. 1E). As in the first step already described, cross-correlation functions for the lowest spatial frequency are calculated for these smaller grids. The shift corresponding to the optimal match is assigned to the centre point of each subgrid (coloured dots in Fig. 1E). For all other points, the mapping is estimated by bilinear interpolation.

This process is iterated using increasingly higher spatial frequencies and increasingly localised subgrids. The set of parameters (spatial frequencies, subgrid sizes and allowable shifts) employed in the present work is given in Table 1. We have not formally investigated the impact of parameters on the performance of our algorithm. However, a few rules of thumb can be easily identified. Firstly, the number of samples for the computation of the cross-correlation should be kept relatively constant over all frequencies. Denser sampling compensates for smaller grids. Secondly, the number of samples must also account for the maximum additional shift allowed for each iteration. The larger the area to be searched, the larger should the sample set be. The amount of information available for a particular

UNCORRECTED PROOF

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

matching step must be large enough to compensate for the range of shifts considered. These rules have been applied in the first step using a single large grid. Then, the mapping is refined gradually, considering more detailed spatial information while searching more localised regions. The resulting mapping can be conveniently displayed as a vector field (Fig. 1F), where the arrows denote the displacement caused by the mapping. Fig. 1 illustrates why this is a reasonable procedure for matching. At the lowest frequency, only low-pass filtered versions of the structures are available, so that finer differences between images can be ignored for a global alignment. Once this global alignment is correct, more details are used for comparison, resulting in finer local readjustments.

2.4. Implementation

The algorithm has been implemented in a fourth-generation programming language (IDL 5.2; Research Systems Inc.) and is running on a Macintosh G3 (233 MHz) under OS 9.0. With a standard set of parameters (see Table 1) corresponding to three spatial frequencies and eight orientations, the matching of two images takes around 180 s for the Gabor decomposition and 100 s for the matching process itself. Note that for the analysis of series of sections, the decomposition of a single image is used twice: first as a target, then as a model. Thus, under these circumstances, the effective time necessary for the transform is cut in half. Please note that these computations do not require user interactions; thus, they can conveniently run in batch mode, processing many hundreds of images overnight. Furthermore, there is ample room for software optimisation, which has not yet been attempted.

2.5. Testing against a constructed benchmark

In order to quantify the performance of the algorithm, we have generated a pair of electron microscopic images with a known mapping of corresponding points. The first section employed is shown in Fig. 3A (left). This image was then distorted by stretching the left and right border region (25%) on the horizontal axis by 20% and compressing the centre half of the image horizontally by 20%. Furthermore, the lower 40% of the image have been compressed vertically by 20%; an equally sized part above stretched vertically by 20%. The upper 20% of the image has not been modified on the vertical axis. The artificially obtained image has electron microscopic features, but shows distortions of 60.5 pixels on average. These distortions are of the same order of magnitude as those encountered in serial sections. However, the solution for the optimal mapping between the model and the target is known, which makes it suitable for a benchmark. The first iteration does not reduce the average deviation, which is not surprising since no global shift was introduced in the distortion. However, the following iterations quickly reduce the deviations (step 2, 53.4 pixels; step 3, 26.0 pixels; step 4, 12.7 pixels). Indeed, in the central half of the image, where enough information is available to match the subgrids, the final step reduces the average error to 5.3 pixels. This is less than one-tenth of the initial average distortion of the target image. Note that the performance of the algorithm is lower bounded by the spatial resolution of the finest grid used. This resolution is in the present case 4 pixels, and thus only little below the actual performance achieved. The quality of the mapping in the central half of the image region is demonstrated in Fig. 2A. Points 1–10 are mapped fairly precisely. At points 11 and 12, downward deviations of the mapped points in the target image are seen. This is an instance of the general problem that less information is available near the image borders. This problem can be addressed by increasing the spatial localisation in the peripheral region or by using larger prints. In this work, however, we do not consider this problem and assume that all struc-

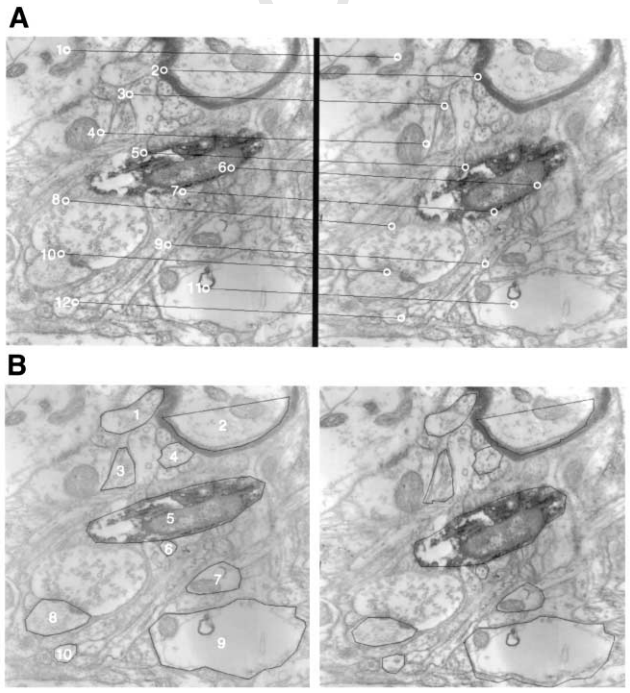

Fig. 2. A morphed image as a benchmark. (A) An original electron microscopic image (left) is distorted by horizontal and vertical stretching and squeezing (for details, see text). Corresponding points as determined by the algorithm are connected by black lines. The numbering refers to a further discussion in the text. (B) Anatomical structures of interest are labelled manually on the left. These include myelinated and unmyelinated axons, dendrites and an HRP stained neuron. The algorithm transfers this labelling to the distorted image. This mapping is based on the matching of features and does not use any knowledge of the type of distortions introduced. Please note that most structures are correctly located in the target image.

UNCORRECTED COPY

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

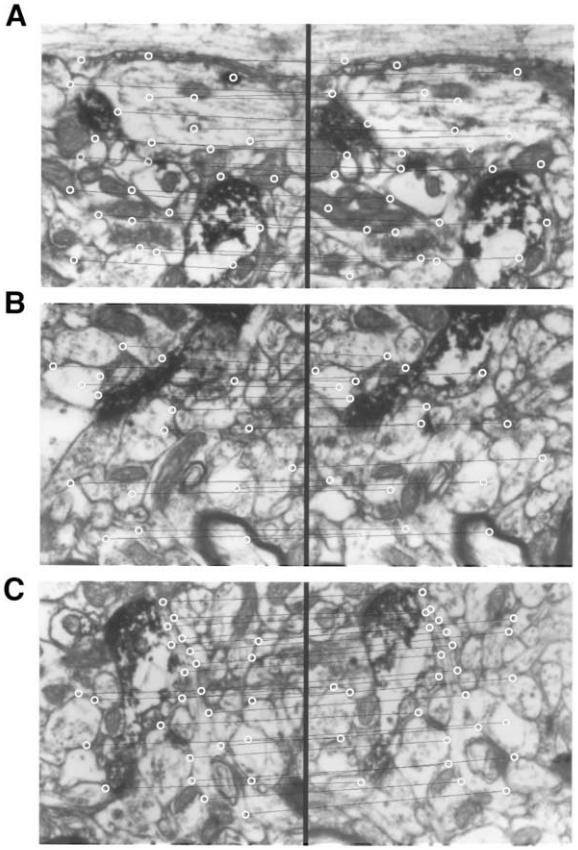

Fig. 3. Matching of individual points in neighbouring sections. (A–C) Points characterising anatomical structures of interest are manually marked in one section (left, red circles) and automatically mapped to the consecutive section (right, white circles). Black lines connect corresponding pairs. For the purpose of illustration, some points in the middle of homogeneous regions have been selected too (see text for details).

tures of interest are located well outside the border region.

In Fig. 2B, the mapping of some structures of interest is investigated. A subset of the visible axons, dendrites and spines are marked by polygons in the model image (left). The vertices are then mapped to the target image (right) and the polygons reconstructed. Automatically generated delineation of unmyelinated axons (1, 3, 4, 6, and 7), the myelinated axon (2) and the HRP labelled neuron (5) precisely match the visible structures. The two axons (8 and 10) and the dendrite (9) show small deviations on the side facing the border of the image. Thus, all structures lying in the central region are mapped to the distorted image rather satisfactorily.

2.6. A real-world test

In this section, we investigate the mapping process between real neighbouring sections (Fig. 3). These sections are also taken from primary visual cortex and show a variety of structures. Labelled corresponding points are marked by red circles and connected by black lines. Scrutinising the mapping reveals that points at higher contrast contours are mapped with a precision matching the results with the artificial benchmark. Two issues are worth noting. Firstly, in some instances points within large uniform structures, e.g. upper middle in Fig. 3A, are placed within the matching homogeneous area but far from a place a human would select. This is possible, as currently in the algorithm no direct neighbourhood interaction is implemented. Thus, the shift assigned to a point at the centre of a subgrid at a given resolution is independent of the shift of other subgrids. Because different structures (axons, dendrites and subcellular structures) may run in different directions, this is a desirable feature. In case a subgrid encompasses large homogeneous regions, however, the energy landscape is flat and no information is available where to map the respective point. This problem may be addressed by placing such points in a constant geometrical relation to neighbouring points. However, in practice, the interest to label the interior of large homogenous regions is limited, and therefore we decided not to implement this feature in the current version of the algorithm. Secondly, in some cases topological discontinuities appear. In Fig. 3B, a merging mitochondrion can be seen in the left half. This leads to features that have no counterpart in the following section. This is an instance of newly appearing or disappearing objects. In these cases, comparable with the first labelled section, the human expert must give additional information. Aside from these two issues, the examples demonstrate that the mapping of the overwhelming majority of points marking structures of interest is satisfactory.

As an even stricter test, we have included images that contain severe artefacts. Please note that this test is carried out with the identical parameter set as the aforementioned experiments. Prominent features in the model image are a large myelinated axon, an axon labelled with HRP, many smaller axons, some containing mitochondria and a dendrite cut obliquely (Fig. 3). In addition, many vesicles and other subcellular compartments are visible. These structures are a source of noise for the present purpose, because they are too small to be visible in neighbouring sections. In Fig. 4, a selection of structures is marked by smooth polygons in the model image (top): a myelinated large axon (1), a small axon (2), a dendrite with mitochondrion (3), an HRP-labelled neuron (4), a small axon (5), and a large dendrite (6). Structures that are not completely visible on the image are labelled as open polygons. The mapping to the subsequent section is shown below (middle). It can be seen that most structures are correctly mapped from model to target, with one exception. At the bottom of the image, one of the corner points of the polygon marking the dendrite is dragged towards the

UNCORRECTED PROOF

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

image boundary (Fig. 3). This can be readily understood, since the mapping near the boundary is less reliable than within the interior of the image. Nevertheless, a satisfactory mapping for more than 80% of labelled points is achieved. This performance can easily be improved by staying clear off artefacts and the image boundary.

These structures are now mapped—without correction of the aberrant corner points—to the following section (Fig. 3, bottom). This section contains a severe artefact near the centre (white arrows). It can be seen that the algorithm degrades gracefully and maps most points that are not directly affected by the artefact correctly. The algorithm cannot correct the aberrant corner point near the bottom. To optimise performance, the few obvious errors should be corrected before the structures are mapped to the following section. In this way, it is possible to map labelled structures not only to the subsequent section, but to efficiently evaluate a sequence of electron microscopic images.

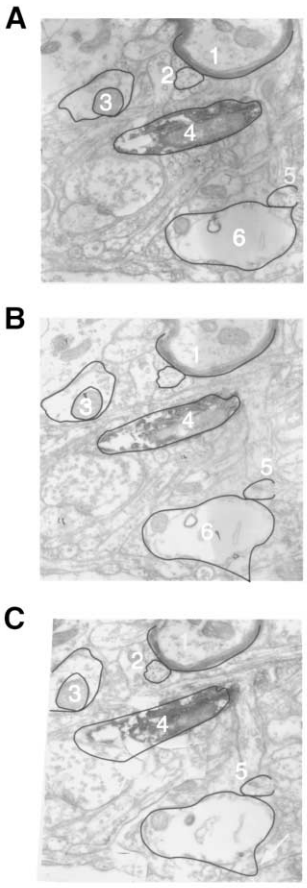

Fig. 4. Matching of a sequence of serial sections. (A) Anatomical structures of interest are manually marked by polygons in one section. Numbers refer to the different structures of interest. See text for details. (B) The algorithm maps all vertices of the polygons to the consecutive sections. Please note that most polygons are mapped to mark the corresponding structures in the target image reasonably well. (C) Automatic mapping of the mapped structures to the subsequent section, which contains some artefacts. The white arrows indicate vertices where the algorithm results in deviations from the optimal solution (see text for further details).

3. Discussion

In this work, we have applied an algorithm, originally proposed for object recognition, to compute a mapping between a model and a target image taken out of a sequence of serial electron microscopic sections. This allows the transfer of labelled information, i.e. which structure is an axon, a dendrite, a spine or a soma, from one image to the other. As the fraction of correctly mapped points is greater than 80% (see Figs. 2 and 3), the evaluation of the second section is greatly facilitated. Given that the total overhead in data handling is small, speed-up by a factor of 4 may be reached.

Nevertheless, the algorithm in its current form has several shortcomings. We now discuss several options available for improvements.

Firstly, in the iterative process, subgrids of the model image are matched to the best fitting position in the target image. This is done irrespective of how much information, i.e. structure in the local region of the image, is available to do so. Thus, for a given grid size and spatial frequency, solutions for optimal matching of noisy or homogeneous regions may be ambiguous. Under these circumstances, it is possible that the quality of some parts of the transformation degrades with the number of iterations. As a remedy, a threshold on the local contrast (i.e. local structure of the image) and on the quality of the match may be introduced. The mapping of points where not enough information is available could be left to interpolation between neighbouring points. This is a special form of neighbourhood interactions that could ensure smooth mappings. The correctly mapped neighbouring points could restrict in particular aberrant points.

Secondly, the current algorithm only depends on the squared sums of the convolution with the pair of Gabor kernels, and deliberately neglected the phase information. This is useful in order to obtain a smooth and well-behaved similarity landscape as shown in Fig. 1D, and to avoid multiple maxima that would cause ambiguous mappings. However, as a consequence, the algorithm cannot discriminate light/dark from dark/light edges. This explains why sometimes the mapping of a point located on a high contrast border is reasonably good, but not exact. A viable compromise is to establish a mapping using amplitude information only, and to exploit phase information in a final correction step to precisely align structure boundaries.

UNCORRECTED PROOF

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

Thirdly, the algorithm as presented here has been investigated with parameters (e.g. relative width of Gabor kernels, number of orientations, spatial frequencies, grid sizes, etc.) that seemed reasonable to us, but no systematic optimisation of these parameters has been performed. Indeed, it has been emphasised that the selection of features to track is critical for the performance of an algorithm (Shi and Tomasi, 1994). Thus, in an application of the algorithm presented here, the first step should be a test of a number of parameter sets, and to investigate for which set the reconstructed images contain the information necessary to label the relevant structures. For example, using very low spatial frequencies makes the matching process sensitive to cutting or photographic artefacts. Similarly, using very high spatial frequencies would induce noise in the computed mapping since they correspond to individual vesicles, which have no correspondences in consecutive sections. Noise will deteriorate the quality of the matching process. Then, it is dependent on the scientific question investigated that structures are marked. Thus, the optimal strategy is to carefully adapt the range of spatial frequencies to the type of images to be processed and only include the relevant features.

Fourthly, in its present implementation, the evaluation of an image takes a few minutes. However, an advantage of the presented algorithm is that the computationally intensive part is not dependent on any user-specified data. Thus, batch processing of a stack of images over night, storage of computed mappings and boosted evaluation of images from serial sections seems an efficient approach. Furthermore, the present implementation concentrates on complete documentation for investigating the properties and the performance of the algorithm, and is not tuned for speed as a production system would be. Furthermore, the standard parameter set uses oversampling, which is very memory and CPU intensive. Implementation in programming languages as C/C++ offers room for improvement. Last but not least, using state of the art computer hardware running a 1.5 GHz processor should give an additional factor of 6 in speed. Combined, it seems feasible to achieve times for computation of a mapping between subsequent images of a serial section as a few seconds. This is further supported by the performance of the original algorithm in on-line object recognition. Thus, the proposed algorithm is suitable for boosting image analysis on-line and transparent for the user.

Although we address in the current work the special problem of tracing cellular structures in electron microscopic images, the proposed algorithm may well be compared with other algorithms used in computer vision. Tracking two-dimensional structures in serial sections is quite comparable with tracking two-dimensional structures over time (Lucas and Kanade, 1981). Matching stereo images raises the correspondence problem too, although the image variations have only one degree of freedom, giving an important constraint (Marr and Poggio, 1976; Jones and Malik, 1992). Compared with these approaches, we address the problem of selection of scale by a hierarchical approach. Recent work on a selection of natural images has demonstrated the high performance of such an approach (Wu et al., 2000).

The process of object definition by a human expert and the tracking of features may be reversed: Motion corresponds to the Gestalt criterion of 'common fate' and is a very powerful cue to segment images (Shi and Malik, 1998; Wiskott, 1999). The result of tracking features can be used to define objects in the first place. In combination with other work on segmentation of natural images (Shi and Malik, 1997; Wiskott and van der Malsburg, 1993), this could then lead to a fully automated tracking algorithm.

However, it has to be noted that the images used in the present study are very different from standard natural scenes. They are more difficult, because they do not contain colour information, appear noisy and contain useful information only in a limited range of spatial frequencies. They may be easier to match; because occlusions of one object by another does not occur and the spatial distance between neighbouring sections is small compared with the spatial extent of the image, the corresponding velocities are limited to a small range. A systematic comparison between different algorithms, which differ in their scope, and tests with the same set of images are promising topics for future research.

The algorithm presented here creates a mapping between two reasonably similar images. It does not recognise anything on its own; rather, it transfers a classification, created by some other program or human expert, to another image. As a consequence it cannot recognise qualitative changes such as the bifurcation of an axon and will only match visual features on a best effort basis. This is a big advantage as the algorithm does not depend on the kind of data supplied and it can operate on a wide variety of image types by adaptation of few parameters to the structures that are typically expected.

UNCORRECTED PROOF

Acknowledgements

The authors are pleased to thank J.C. Anderson and K.A.C. Martin for supplying high quality electron microscopic serial images, fruitful discussions on the relevant aspects in electron microscopic images and helpful comments on an earlier version of this manuscript. This work was supported by SNF grant 3100-051059 to P.K. and SPP Bioinformatics, and the Körber Prize awarded in the year 2000 to C. von der Malsburg at the Institute

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

of Neuroinformatics in Bochum and R.J. Douglas at the Institute of Neuroinformatics in Zürich. The authors will be happy to provide on request the software described in this paper.

Appendix A. Gabor wavelets

The Gabor transform of an image $l(\vec{x})$ is defined as the absolute value of the convolution with a set of complex Gabor functions:

$$(WI)(\vec{k},\vec{x}) = \left|(\psi_{\vec{k}}^{*}I)(\vec{x})\right| \quad (1)$$

with

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

A.1. Sampling

The two-dimensional frequency vector $\vec{k}$ codes for a scale I and a an orientation d, and is defined as follows.

$$\vec{k}_{ld} = k_{\min}\left(\frac{k_{\max}}{k_{\min}}\right)^{1/L}\begin{bmatrix}\cos(d/D\cdot 180°)\\ \sin(d/D\cdot 180°)\end{bmatrix} \quad (3)$$

Spatial sampling is adapted to I in such a way that one wavelength of a wave with frequency $\vec{k}_{ld}$ contains four sampling points. This is an oversampling of 2.

A.2. Feature vectors

If I is kept constant, the transformation Eq. (1) assigns a feature vector of positive numbers to each image point $\vec{x}$. These feature vectors are the basis of the registration procedure, and their similarity is measured by their normed scalar product:

$$S_{\mathrm{loc}}(\vec{h}^I\vec{h}^M) = \frac{\sum_{d=0}^{D-1}(h_d^M)(h_d^I)}{\|\vec{h}^M\|\,\|\vec{h}^I\|} \quad (4)$$

A.3. Multidimensional template matching

Standard template matching must be slightly modified for matching feature vectors rather than scalars. Multidimensional template matching (MTM) consists of finding the displacement $\vec{y}$ for a 'data grid' $\vec{f}(\vec{x})$ and a 'template grid' $\vec{t}(\vec{x})$ such that the average similarity of the single feature vectors becomes maximal inside a defined range for $\vec{y}$.

$$S(\vec{f},\vec{t})(\vec{y}) = \frac{1}{N^2}\sum_{\vec{x}} S_{\mathrm{loc}}[\vec{f}(\vec{x}),\vec{t}(\vec{x}-\vec{y})] \quad (5)$$

where N^2 is the number of grid points in the template (which is generally square).

A.4. Coarse-fine schedule

A mapping is a function that maps coordinates of one image onto coordinates of the second image. Mappings are calculated coarse to fine. During the process, the following parameters can be varied:

1. the frequency $\vec{k}$, which is connected with the grid resolution;
2. the template size N, which is connected with the number of possible templates;
3. the range of the shift.

All these are, in principle, independent of each other, therefore they must be combined to a 'schedule' S_i, a list of values of these parameters. A useful schedule is shown in Table 1.

Given two images I_1 and I_2, and a schedule S_i, the matching process is as follows:

Calculate Gabor transforms $W_1(l,d,\vec{x})$ and $W_2(l,d,\vec{x})$ of both images.
Initialise $i = 0$.
Initialise mapping to identity $M_0(\vec{x}) = \vec{x}$ for all $\vec{x}$.
While $i < n_{\mathrm{schedules}}$ do:
  Set parameters according to S_i
  For each location $\vec{x}$ on the given grid in I_1 do:
  Define the template as the feature vectors in W_1.
  Define the data as the feature vectors in W_2 with centre in $M_1(\vec{x})$ and a size as specified by the range of shift in S_i.
  Define $M_{i+1}(\vec{x})$ as the result of multidimensional template matching of data and template.

This yields a series of mappings defined on grids, which depend on the schedule. The final mapping $M_{n_{\mathrm{schedules}}}$ is then interpolated to the full image resolution.

References

Adams JC. Heavy metal intensification of DAB based HRP reaction product. J Histochem Cytochem 1981;29:755.
Anderson JC, Douglas RJ, Martin KAC, Nelson JC. Synaptic output of physiologically identified spiny stellate neurons in cat visual cortex. J Comp Neurol 1994;341:16–24.
Anderson JC, Binzegger T, Kahana O, Martin KAC, Segev I. Dendritic asymmetry cannot account for directional responses of neurons in visual cortex. Nat Neurosci 1999;2:820–4.
Fukuda T, Kosaka T. Gap junctions linking the dendritic network of GABAergic interneurons in the hippocampus. J Neurosci 2000;20:1519–28.
Gibbins IL, Jobling P, Messenger JP, Teo EH, Morris JL. Neuronal morphology and the synaptic organisation of sympathetic ganglia. J Auton Nerv Syst 2000;81:104–9.
Hack I, Peichl L, Brandstatter JH. An alternative pathway for rod signals in the rodent retina: rod photoreceptors, cone bipolar cells, and the localization of glutamate receptors. Proc Natl Acad Sci USA 1999;96:14130–5.

UNCORRECTED PROOF

/typeset2:/sco3/jobs1/ELSEVIER/nsm/week.29/Pnsm3076.0011Wed Aug 15 15:55:36 2001 Page Wed Au

Hanker JS, Yates PE, Metz CB, Rustioni A. A new specific sensitive and non carcinogenic agent for the demonstration of horseradish peroxidase. Histochem J 1977;9:789–92.

Jones D, Malik J. Computational framework for determining stereo correspondence from a set of linear spatial filters. Image Vision Comput 1992;10:699–708.

Lades M, Vorbrüggen JC, Buhmann J, Lange J, von der Malsburg C, Würtz RP, Konen W. Distortion invariant object recognition in the dynamic link architecture. IEEE Trans Comp 1993;42:300–11.

Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. Proceedings of the 7th International Joint Conference on Artificial Intelligence, 1981, pp. 121–130.

Marr D, Poggio T. Cooperative computation of stereo disparity. Science 1976;194:283–7.

Rockland KS, Knutson T. Feedback connections from area MT of the squirrel monkey to areas V1 and V2. J Comp Neurol 2000;425:345–68.

Shi J, Tomasi C. Good features to track. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, 1994, pp. 593–600.

Shi J, Malik J. Normalized cuts and image segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997, pp. 731–737.

Shi J, Malik J. Motion segmentation and tracking using normaliyed cuts. International Conference on Computer Vision, 1998, pp. 1154–1160.

von der Malsburg C. Pattern recognition by labeled graph matching. Neural Networks 1988;1:141–8.

Wiskott L, von der Malsburg C. A neural system for the recognition of partially occluded objects in cluttered scenes. Int J Pattern Recog Artif Intell 1993;7:935–48.

Wiskott L. Segmentation from motion: combining Gabor- and mallat-wavelets to overcome the aperture and correspondence problems. Pattern Recog 1999;32:1751–66.

Wu Y, Kanade T, Li C-C, Cohn J. Image registration using wavelet-based motion model. Int J Comput Vision 2000;38:129–52.

Würtz RP. Object recognition under translations, deformations, and changes in background. IEEE Trans Pattern Anal Mach Intell 1997;19:769–75.

UNCORRECTED PROOF