Learning the Topology of Object Views

Jan Wieghardt[1], Rolf P. Würtz[2], and Christoph von der Malsburg[2,3]

[1] SIEMENS AG, CT SE 1, Otto-Hahn-Ring 6, D-81730 München, Germany
jan.wieghardt@mchp.siemens.de
[2] Institut für Neuroinformatik, Ruhr-Universität Bochum
D-44780 Bochum, Germany
rolf.wuertz@neuroinformatik.ruhr-uni-bochum.de
[3] Laboratory for Computational and Biological Vision
University of Southern California, Los Angeles, USA

Abstract. A visual representation of an object must meet at least three basic requirements. First, it must allow identification of the object in the presence of slight but unpredictable changes in its visual appearance. Second, it must account for larger changes in appearance due to variations in the object's fundamental degrees of freedom, such as, e.g., changes in pose. And last, any object representation must be derivable from visual input alone, i.e., it must be learnable.
We here construct such a representation by deriving transformations between the different views of a given object, so that they can be parameterized in terms of the object's physical degrees of freedom. Our method allows to automatically derive the appearance representations of an object in conjunction with their linear deformation model from example images. These are subsequently used to provide linear charts to the entire appearance manifold of a three-dimensional object. In contrast to approaches aiming at mere dimensionality reduction the local linear charts to the object's appearance manifold are estimated on a strictly local basis avoiding any reference to a metric embedding space to all views. A real understanding of the object's appearance in terms of its physical degrees of freedom is this way learned from single views alone.

Keywords

Object recognition, pose estimation, view sphere, correspondence maps, learning, topology

1 Introduction

An object representation useful for computer vision should meet at least three basic requirements. First, it must allow identification of an object in the presence of slight but unpredictable changes in its visual appearance. This a fact widely acknowledged and a problem frequently tackled. One successful way of dealing with this invariance problem is provided by the *elastic graph matching* approach [6]. Second, it must be able to yield an interpretation of larger changes in appearance, which are due to variations in the object's fundamental degrees of freedom, such as, e.g., changes in pose. And last, it is important that the object representation be learnable, i.e., it should be derived from visual input alone, without interaction of a user.

In this paper, we construct such a representation by deriving transformations between the different visual appearances of a given object, so that they can be parameterized in terms of the object's physical degrees of freedom. Our method allows to automatically derive the appearance representations of an object in conjunction with a model for their linear deformation from example images. These are subsequently used to provide linear charts to the entire appearance manifold of a three-dimensional object. A topology on the space of all views of an object is created, without reference to an embedding metric space. Finally, it is shown that the retrieved topology is indeed ideally suited to relate the single views of an object to parameterizations of the object's physical degrees of freedom. An understanding of the object's appearance in terms of its physical degrees of freedom is thus learned from nothing but the visual input.

2 View representation by labeled graphs

A single view of an object is here represented by *labeled graph* such that the similarity between different views can be assessed via *elastic graph matching* [6]. In order to extract the labeled graph from a given image, the location of the object in the image must first be determined.

2.1 Segmentation

As fully-fledged *data driven segmentation* is beyond the scope of this paper some simplifying assumption about the nature of the image data are made. It is assumed that images contain only one object, that the background is fairly homogeneous, and that the object is relatively close to the center of the image.

If these requirements are fulfilled (and they are in the image data we used in this paper) the background can be determined using the spin segmentation approach described in [4]. From the thus determined object region a regular grid-graph can be extracted as shown in figure 1.

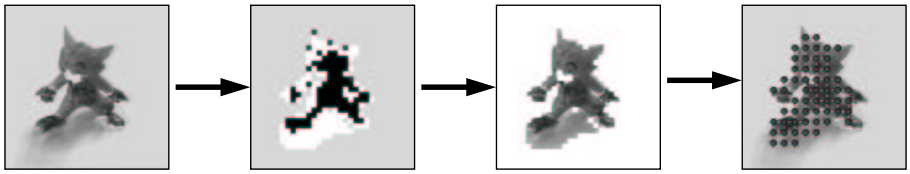


Fig. 1. Single View Models: In order to describe an object, a suitable representation of a single view is found by first splitting the image into a background and a foreground region. To this end the spin segmentation method [12, 4] is used, in connection with the knowledge that only one object is presented approximately at the center of the image and that there is a nearly homogeneous background. After the segmentation step the resulting object region is represented by a jet-labeled graph.

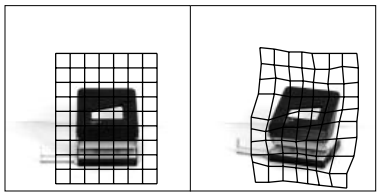


Fig. 2. Elastic Graph Matching: The labeled graph (*left*) is used to determine the correspondences with a novel but similar view (*right*). The figure is copied with permission from [6].

2.2 Model graphs

Each graph node is now labeled with a *Gabor-jet*, which locally describes the image content. Given an image of a novel but similar view of the same object *elastic graph matching* [6] can be be used to determine the corresponding points in the new image for each node of the graph (see figure 2).

Elastic graph matching thus yields a set of x_i and y_i coordinates in the new image for each node i in the graph as well as the the similarity s_i between Gabor-jet of node i and the Gabor-jet at the corresponding location $(x_i, y_i)^T$ in the new image. The overall similarity $s_{min} = \min_i(s_i)$ can then be used as a similarity measure between views.

3 Aspects

Equipped with this powerful method of comparing views, one can try to group views, according to their similarities, into larger entities which we will call *aspects*. Given a model graph g and a set of views U, the *aspect* a is defined as the subset of views u_i of U, whose similarity s_{min} to g is above a given threshold t, i.e.

$$a(g, U) = \{u_i \in U | s_{min}(g, u_i) > t\}. \tag{1}$$

In this context we call g the *representational graph* of the *aspect* a. This representational graph can be used to establish a meaningful topology on the views of the aspect a.

3.1 Local linear descriptions

Let g be a representational graph and a its associated aspect, the correspondences between g and all views in a are then also given, as a was derived from the training set U by thresholding on the graph similarities, which in turn were established by elastic graph matching. The correspondence mapping can now be associated with a point $(x_1, y_1, \ldots, x_N, y_N)^T$ in Euclidean space, where $(x_l, y_l)^T$ is the mapped position of the l-th node of the representational graph. It must be noted that for a given aspect the dimensionality of the mapping space is fixed and given by twice the number of nodes N of the representational graph, but between different aspects associated with different representational graphs the dimensionality can vary.

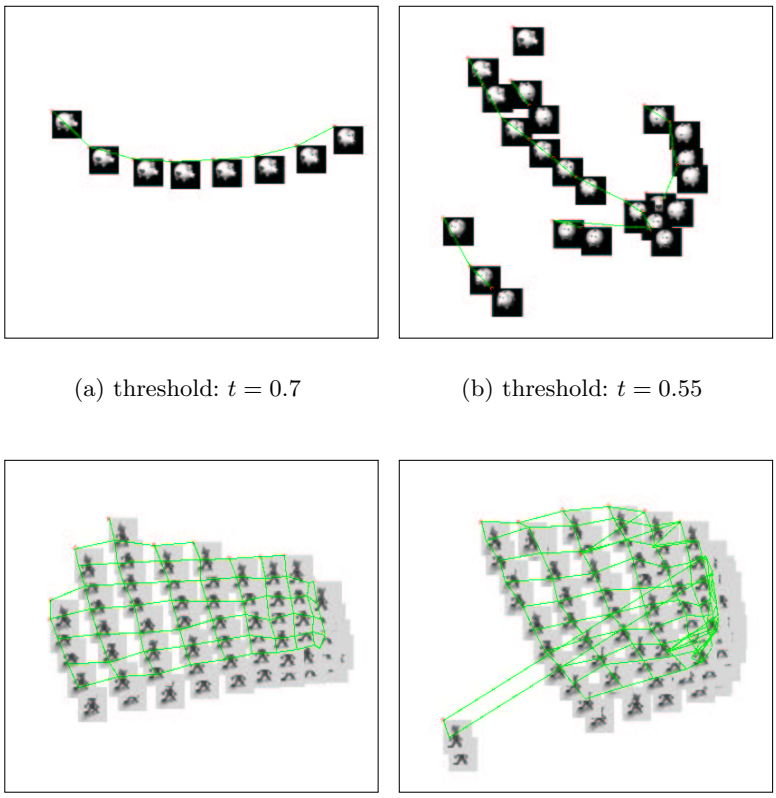
To get rid of all the variation in this coordinate vector stemming from simple translation, the correspondence mappings are normalized by subtracting the graphs center of gravity, i.e.

$$\boldsymbol{X}^T = \left(x_1 - \frac{1}{N}\sum_n x_n, y_1 - \frac{1}{N}\sum_n y_n, \ldots, y_N - \frac{1}{N}\sum_n y_n \right). \tag{2}$$

To see whether the variations between the coordinate vectors of different views within the same aspect can be understood in terms of variations in some underlying low-dimensional parameter space *principal component analysis* is performed on the set of all the $\boldsymbol{X}_i$ of all views view i in the aspect a. The resulting eigenvector matrix P and the mean $\boldsymbol{M}$ can now be used to introduce a topological structure on the aspect a. For a given dimensionality K of the aspect, i.e. the number of degrees of freedom along which the observer-object-relations varied to create those views, the first K eigenvectors $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_K$ form an orthogonal basis, which spans an affine subspace of the correspondence mapping space. A correspondence mapping $\boldsymbol{X}$ can then be expressed in this coordinate system by the vector

$$\boldsymbol{\xi} = \left((\boldsymbol{X} - \boldsymbol{M})^T \boldsymbol{P}_1, \ldots, (\boldsymbol{X} - \boldsymbol{M})^T \boldsymbol{P}_K \right)^T. \tag{3}$$

As each view within the aspect is associated with a correspondence mapping, these coordinates impose a topology upon the views of the aspect. Examples of the such induced topology are shown in figure 3 for $K = 2$ and different values of t. The quality of the established topology depends of course critically on a suitable choice for t. Luckily, the similarity as established by *elastic graph matching* is a fairly good estimate of view-similarity independent of the concrete object at hand, as previously shown in [5, 1], such that $t = 0.7$ yielded good results in all following experiments.



(a) threshold: $t = 0.7$ (b) threshold: $t = 0.55$

(c) threshold: $t = 0.7$ (d) threshold: $t = 0.55$

Fig. 3. Examples of aspect representations: Shown are four groups of views. Each is displayed in its local coordinate system spanned by the first two principal components. Views that are known from ground truth to correspond to neighboring samples on the view sphere are connected by a line. For each group the value of the threshold used listed below the diagram. The data sets (a) and (b) have one degree of freedom, whereas (c) and (d) have two.

3.2 Estimating local dimensionality

In order to estimate the aspect dimensionality K automatically, a modified version of the *Scree*-test is used. Given the eigenvalues of the principal components of a distribution, it is assumed that they can be separated into two classes, one belonging to variations due to changes in some underlying variables and the second being created by noise. The first group, consisting of the largest eigenvalues, contributes most to the overall variation. The second group, generated by noise, usually shows a characteristic flat and monotonous decline, which is approximately exponential [3].

To detect this transition the ratio between the eigenvalue p_i and its immediate predecessor p_{i-1} is calculated. The minimum of those ratios $\frac{p_i}{p_{i-1}}$ is supposed to mark the transition. So the first K principal components are assumed to be due to variations in the object's degrees of freedom if

$$K = \arg\min_i \left(\frac{p_i}{p_{i-1}} \right) - 1. \tag{4}$$

Using this method, a correct estimation of the number of underlying degrees of freedom was possible for 97% of the aspects.

4 View manifold

It was demonstrated how a set of views can be decomposed into aspects, which in turn can be represented in terms of simple linear models via *principal component analysis*. In order to arrive at a complete parameterization of the view manifold it is again assumed that a training set U of views is given. It is also assumed that a set of representational graphs and the aspects they define were extracted from U such that all views in U are part of at least one aspect and that pairs of aspects which are neighbors in appearance space have a set of views in common. The set of views common to two aspects a and b is denoted O_{ab}. Because these sets O contain views which are represented in more than one aspect, they can be exploited to establish relations between the aspects.

4.1 Defining aspect distances

Let a and b be two K-dimensional aspects with representational graphs g^a and g^b, mean correspondence mappings $\boldsymbol{M}^a$ and $\boldsymbol{M}^b$, and eigenvector-matrices P^a and P^b, respectively. Then each view which is contained in both a and b can be represented by a point in a K-dimensional Euclidean space for each of the two aspect representations (see section 3.1). The point coordinates are derived from the correspondence mappings $\boldsymbol{X}^a$ and $\boldsymbol{X}^b$ from the representational graphs g^a and g^b associated with the view. They yield the following coordinates in the affine subspaces of each aspect.

$$\boldsymbol{\xi}^a = \left((\boldsymbol{X}^a - \boldsymbol{M}^a)^T \boldsymbol{P}^a_1, \ldots, (\boldsymbol{X}^a - \boldsymbol{M}^a)^T \boldsymbol{P}^a_K \right)^T, \tag{5}$$

$$\boldsymbol{\xi}^b = \left(\left(\boldsymbol{X}^b - \boldsymbol{M}^b\right)^T \boldsymbol{P}^b_1, \ldots, \left(\boldsymbol{X}^b - \boldsymbol{M}^b\right)^T \boldsymbol{P}^b_K \right)^T. \tag{6}$$

It must be noted that the dimensionalities of $\boldsymbol{\xi}^a$ and $\boldsymbol{\xi}^b$ are all equal to K, because K is the number of the object's degrees of freedom and must be constant over the whole view manifold. On the other hand $\boldsymbol{X}^a$ and $\boldsymbol{X}^b$ may have completely different dimensionalities, because they are determined by the number of nodes of g^a and g^b and are therefore of a fairly arbitrary nature.

If a whole set of common views O_{ab} is given one can estimate the centers of gravity $\boldsymbol{o}^{\,a}$,$\boldsymbol{o}^{\,b}$ in the two aspect coordinate systems.

$$\boldsymbol{o}^{\,a} = \frac{1}{N} \sum_{n \in O_{ab}} \boldsymbol{\xi}_n^{\,a}, \tag{7}$$

$$\boldsymbol{o}^{\,b} = \frac{1}{N} \sum_{n \in O_{ab}} \boldsymbol{\xi}_n^{\,b}, \tag{8}$$

where N is the number of views in the overlap. One can now define the distance Δ_{ab} between the origins of the two aspects a and b by

$$\Delta_{ab} = \|\boldsymbol{o}^{\,a}\| + \|\boldsymbol{o}^{\,b}\|. \tag{9}$$

Given the distances between all overlapping aspects, the distance between non-overlapping aspects can be defined to be the shortest connecting path of overlapping aspects. If, e.g., aspects a and c are not overlapping, but the distances Δ_{ab} and Δ_{bc} to the aspect b are already defined, Δ_{ac} is defined by $\Delta_{ac} = \Delta_{ab} + \Delta_{bc}$.

If these rules lead to different definitions of the distance between a pair of aspects the distance is taken to be the shortest of those distances. In this way all violations of the *triangle inequality* are eliminated and a metric is imposed upon the aspects.

In case the set of aspects is composed of two or more completely unconnected subsets, these are treated separately in all following considerations and are said to constitute different *object hypotheses*.

It is important to note that this somewhat awkward distance measure has unique qualities rendering it fundamentally different from all other previously proposed methods. The distances are measured solely *inside* the view manifold, no reference is made to an embedding space. This distinguishes it from other methods which retrieve perceptual manifolds in an unsupervised fashion, such as classical neural network approaches or more recent techniques, e.g., [9, 11]. Even most techniques which establish the topology of views in a supervised fashion such as [7, 2, 10, 8] require an embedding space in which distances can be measured. This point is so important, because a space embedding all views of an object requires finding a view representation suitable for the whole view sphere of an object beforehand. This in turn raises difficult questions regarding the *missing data problem* [8], caused for example by self-occlusion, or it requires normalization procedures in order to render the view representation comparable even across large changes in viewpoint [7, 2].

4.2 Global parameterization

So far, a local topology has been established between all views within an aspect, and distances between the single aspects have been defined. What is missing for complete knowledge about the manifold is a global parameterization of all views. To achieve this *metric multi-dimensional scaling* is applied to the distance matrix Δ of the aspects. Given a distance matrix between a number of objects,

in our case the aspects, metric multi-dimensional scaling yields a representation of these objects as points in an L-dimensional Euclidean space, such that the distances between those points approximate the original distance matrix. Very similar to principal component analysis each dimension in this Euclidean space is associated with an eigenvalue, such that the appropriate dimensionality L can be estimated by the previously introduced *Scree*-test.

This way a point $\boldsymbol{c}^{\,a}$ in an L-dimensional Euclidean space $\mathcal{G}$ is assigned to each aspect a. But as this only defines coordinates in the global parameterization space $\mathcal{G}$ for the origin of the local aspect parameterization, one has to take one more step, namely to estimate a mapping from the local parameterization of each aspect to the global one.

4.3 Aligning the aspects

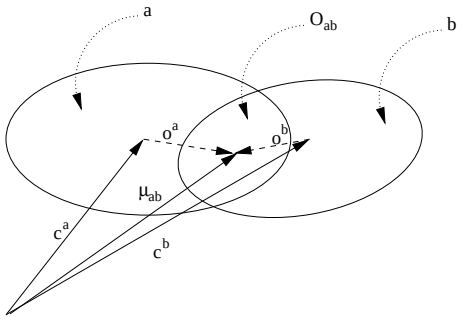


Fig. 4. Neighboring Aspects: Shown is a schematic drawing of two neighboring aspects a and b. The coordinates of the origins of the aspect representation are given by $\boldsymbol{c}^{\,a}$ and $\boldsymbol{c}^{\,b}$, respectively. The set of views in the overlap of a and b is denoted O_{ab}. The center of gravity of O_{ab} in the local coordinate systems of a and b is given by $\boldsymbol{o}^{\,a}$ and $\boldsymbol{o}^{\,b}$.

In order to retrieve mappings from the local parameterization of each aspect into the global parameter space the already established global coordinates $\boldsymbol{c}^{\,a}$ of the aspect origins are used to determine the rough position of views and the overlap sets O_{ab} are again exploited to yield the correct orientation of the different aspect mappings.

Let $\boldsymbol{c}^{\,a}$ and $\boldsymbol{c}^{\,b}$ be the global coordinates of two overlapping aspects a and b, and let $\boldsymbol{o}^{\,a}$ and $\boldsymbol{o}^{\,b}$ be the center of gravity of the overlap set O_{ab} in local coordinates of a and b respectively. In this situation the global coordinates $\boldsymbol{\mu}^{\,ab}$ of the center of gravity of O_{ab} is estimated by

$$\boldsymbol{\mu}^{\,ab} = \boldsymbol{c}^{\,a} + \frac{\|\boldsymbol{o}^{\,a}\|}{\|\boldsymbol{o}^{\,a}\| + \|\boldsymbol{o}^{\,b}\|} \left(\boldsymbol{c}^{\,b} - \boldsymbol{c}^{\,a}\right). \tag{10}$$

In doing so it was assumed that $\boldsymbol{o}^{\,a}$ and $\boldsymbol{o}^{\,b}$ are mapped onto the same point $\boldsymbol{\mu}^{\,ab}$ and moreover that the mapped origins $\boldsymbol{c}^{\,a}$, $\boldsymbol{c}^{\,b}$, and $\boldsymbol{\mu}^{\,ab}$ lie on a straight line (see figure 4).

Assuming now that the mapping between the aspect a and the global parameter space can be approximated linearly, a matrix $\mathbf{A}_a$ can be estimated from

$$\boldsymbol{\mu}^{\,ab} - \boldsymbol{c}^{\,a} = \mathbf{A}_a \boldsymbol{o}^{\,a}. \tag{11}$$

$\mathbf{A}_a$ is a $L \times K$ matrix, where L is the dimensionality of the global parameterization space $\mathcal{G}$ and K is the the dimensionality of the local parameterization, or in other words the dimensionality of the view manifold. One such equation can be obtained for each overlap of a and another aspect. Consequently, K overlap sets are sufficient to determine $\mathbf{A}_a$. If more are available a solution optimal in a least-square sense can be found by standard techniques. If fewer are available, the aspect representation cannot be incorporated into the global parameterization and must be discarded. This, however, is rarely the case as the required number of overlapping aspects is just the number of the physical degrees of freedom of the object.

So, if a new view, represented by an image, is to acquire coordinates $\boldsymbol{\zeta}$ in terms of the global parameterization, all representational graphs g^a are matched on the image. The one with the highest similarity is then used to assign to the new view local coordinates $\boldsymbol{\xi}^{\,a_{max}}$ in terms of the aspect representation. These are then transformed into the global coordinate system by

$$\boldsymbol{\zeta} = \mathbf{A}_{a_{max}} \boldsymbol{\xi}^{\,a_{max}} + \boldsymbol{c}^{\,a_{max}}. \tag{12}$$

4.4 Summary

All the previously described processing stages can now be integrated into one algorithm which derives a complete object representation from single views.

If views of an object are presented to the system one by one in no particular order, the algorithm proceeds the following way. First a view is associated with an aspects, describing the transformation properties of the object by a local linear model. These linear charts to the actual view manifold are then integrated into one global embedding parameter space by determining their topology from the mutual overlap.

So, for each view of an object which is presented to the system, the following steps are taken:

1. **Initialization:** If this is the first view and no aspect representation has so far been created, a graph is extracted from the view, after prior segmentation. This graph is then the representational graph g^a of a new aspect a.
2. **Aspects:** All representational graphs g^a associated with local aspect representations a are matched onto the view.
   (a) If the similarity between the new view and a representational graph g^a exceeds the threshold value t, i.e. $s_{min} > t$, the view is assumed to be

well represented in the associated aspect a. The retrieved correspondence mapping $\boldsymbol{X}^a$ between graph and view is stored and the linear aspect model is updated by recalculating the principal components of the correspondence mappings (see section 3.1).

(b) In case a view is represented well by two aspects a and b the view belongs to the overlap set O_{ab} of those two aspects, and they are said to be neighboring.

(c) If the view cannot be represented in any of the current aspects, a graph is extracted from the view, after prior segmentation. This graph is then the representational graph g^a of a new aspect a.

3. **Object Hypotheses:** Aspects that are neighboring or can be connected by a succession of neighboring aspects are grouped together to form an object hypothesis. An object hypothesis is thus a coherent patch, or atlas, of an object's view manifold for which neighborhood relations between the linear models as well as distances between the aspects in the sense of section 4.1 are defined.
4. **Global Parameter Spaces:** A global parameter space $\mathcal{G}$ is now assigned to each object hypothesis. The mapping between the origins of the linear aspect representations and the new global coordinate system is retrieved via multi-dimensional scaling (see section 4.2).
5. **Mappings from Global to Local Representations:** The transformations $\mathbf{A}_a$ from the local linear parameterizations of all aspects a into the embedding global parameter space $\mathcal{G}$ are retrieved by once again exploiting the overlap between the aspects (see section 4.3).

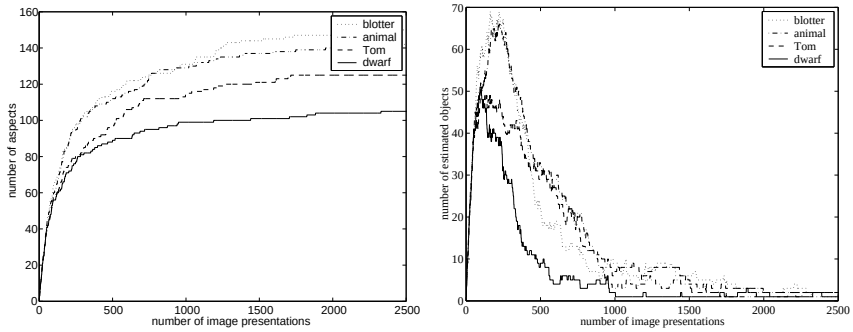
In this fashion the object representation can be refined as more and more views become available. The process has no well defined point of termination, because it cannot be guaranteed, that all degrees of freedom of one object are covered for any given representation. But when only views are presented to the system, which are already well represented, the object representation should slowly converge towards a stable solution.

5 Experiments

The algorithm was tested separately on image databases of four different objects, two plastic figures, an ink blotter, and an animal (the plastic figures were real objects whereas the other two objects were generated by the raytracer POVRAY). Of each object 2500 views were recorded by varying longitudinal viewing-angle in steps of 3.6° from 0° to 360° and the latitudinal viewing-angle in steps of 3.6° from 0° to 90°. The views were presented to the system by selecting a view at random in each iteration step. Multiple presentations of the same view were not prohibited.

5.1 Results

Figure 5 shows the evolution of the number of aspects and object hypothesizes as a function of views presented for those sets of views reflecting two degrees of



(a) Number of aspects as function of iterations

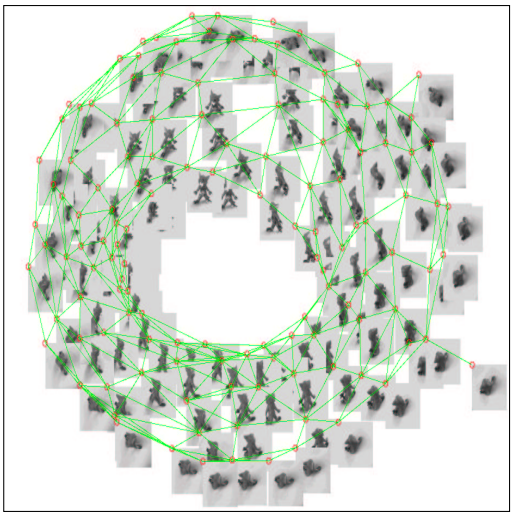
(b) Number of estimated objects as function of iterations

Fig. 5. Evolution of representations: For four example objects with training images containing two degrees of freedom the change in the number of aspects and the number of estimated objects are shown as more and more images are presented. After 1000 to 1500 presentations most aspects seem to have a corresponding representation. At about the same time the representation of single aspects are fused to form a small set of integrated object representations. After about 2000 presentations all but one or two aspects are fused to from a complete object representation.

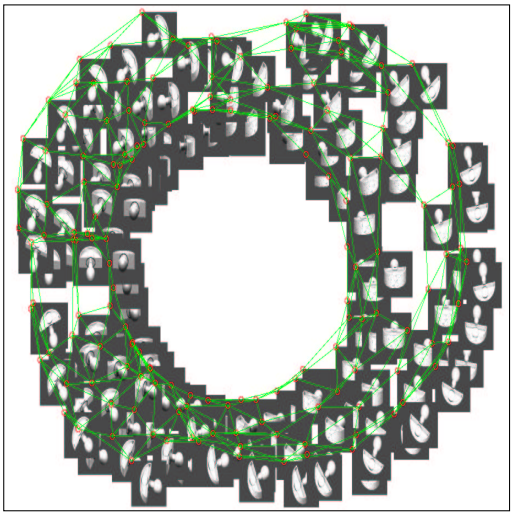
freedom. After $1000 - 1500$ presentations the final number of aspects is nearly reached, all remaining views are already represented by the existing aspects. After another 500 presentations most aspects are integrated into into one coherent object representation.

For two of the four objects (see figures 6(a),6(b)) the derived topology and embedding into a three-dimensional space the nature of the underlying transformation is well captured. Views which are located in close proximity on the view sphere are associated with similar coordinates in the embedding space and vice versa.

The view topology, derived for the *dwarf* object (figure 7(a)), is less intuitive. Here the view topology, as for example given by the relative viewing angle, is not well established. The reason is to be found in the asymmetric properties of the similarity function used. To determine whether a certain view is part of a given aspect, the representational graph is matched onto the view and all subsequent evaluations are solely based on the resulting similarity. But to yield a high similarity it suffices if the representational graph is similar to a *part* of the view. The top view of the dwarf, which shows only the hat, does match the hat of the side view of the dwarf very well and the views are thus assumed to belong to the same object aspect. This could be trivially avoided by employing a symmetric similarity function (for an example see [13]), which requires the similarity between the top of the dwarf to the side of the dwarf to be the same as vice versa. Such a similarity function was not used as it would require the solution



(a)



(b)

Fig. 6. Two examples where the topology of the view sphere was captured very well by the algorithm. (a) shows real pictures of a plastic figure of the cat Tom, (b) generated graphics of an ink blotter.

of the correspondence problem in two directions and thus double the necessary calculation time, which was already fairly long. But it must be emphasized that the poor result for the dwarf is due to a detail in the implementation rather than due to the general approach.

More general problems are exhibited in figure 7(b). In this case the topology of views is not represented well. Views which, although similar, are not neighbors in physical terms, are represented as being part of the same aspect of the object. This in turn yields a topology which cannot be realized as a manifold. This is so because the assumed virtual transition between the physically unconnected views locally introduces a new degree of freedom, which is not present at other points of the view sphere. In the examples presented this effect might be avoided by employing a more sensitive similarity measure or by resorting to images of higher resolution. But this problem cannot be avoided in general, because objects are thinkable that actually have identical front and rear views. In such cases the similarity of single views is insufficient to determine their topology and it shows a point where additional information must be used to construct a reliable object representation.

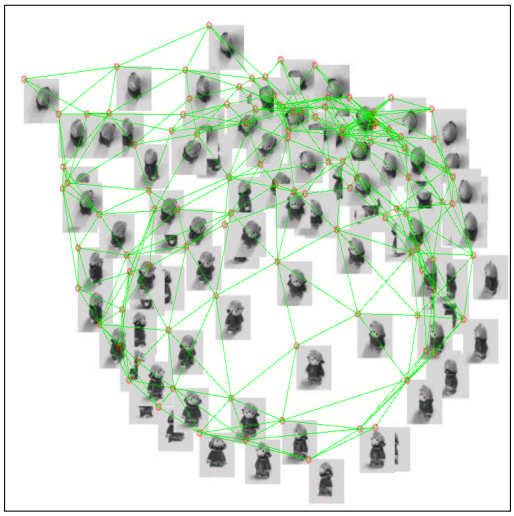
6 Conclusion

It was demonstrated that it is feasible to derive an appearance-based object representation from nothing but single pictures. The approach taken went all the way from the single views of an object to aspects to a completely integrated object description.

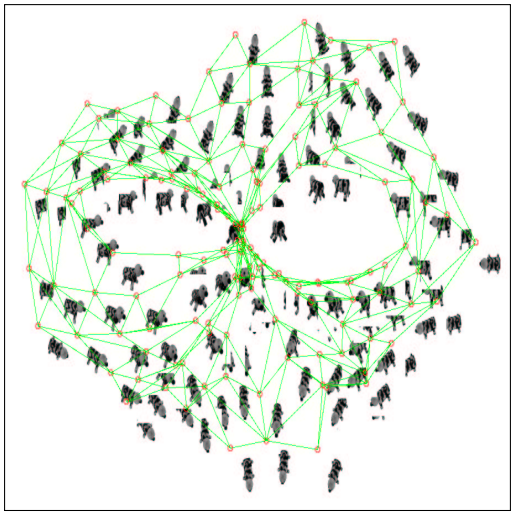
At each stage of integration the entities of representation (model graphs, aspects, view manifolds) provide new knowledge about an object. The description of single views in terms of representational graphs enables the system to recognize known views in the presence of slight variations by elastic graph matching without any further knowledge about the overall structure. The representation of aspects in terms of representational graphs and linear transformation properties helps to determine the local degrees of freedom of an object. Thus creating a foundation for understanding the varying appearance of one object in terms of variations in lower-dimensional parameter spaces.

The global representation, finally, integrates the local topologies and parameterizations into one global topology-preserving parameterization. This way all the different view of an object can be related to each other, forming an ideal basis to describe an object's appearance in terms of its physical parameters.

Starting from a simple view representation and establishing local topologies by aspect representations before integration in terms of a global parameterization also eliminated the need for an embedding space for all views.In the exemplary case of a representation by graphs creating an embedding space would imply that all views must be dealt with in terms of a single graph. Such a graph would unavoidably face the problem that some of its nodes disappear from the field of view while the object is rotating. The task of handling those invisible parts which are nonetheless still part of the representation was termed the *missing*



(a)



(b)

Fig. 7. Two examples where the algorithm yields results different from the view sphere. In (a), real views of a plastic dwarf are arranged poorly, because the system is confused about the very similar top views. In (b), generated graphics of an animal lead to a singularity, where the front and back views are very similar although distant in the view sphere topology.

data problem in [8]. By utilizing completely independent representations for each aspect of the object we here avoided the *missing data problem*. We also kept the representational space low-dimensional at all times, because invisible parts of the object were simply not represented. Differentiating between the local number of degrees of freedom and the dimensionality of the embedding parameter space also prevented the spread of local errors in topology and allowed to properly parameterize non-Euclidean manifolds.

These results shed some light on the discussion of how new objects can be learned. They provide evidence that single view information can be enough to learn object properties, but additional information, such as continuous motion or even manipulation is sometimes required to resolve ambiguities. The algorithm also provides a basis for learning facts about the three dimensional physical space from visual experience.

References

1. Mark Becker, Efthimia Kefalea, Eric Maël, Christoph von der Malsburg, Mike Pagel, Jochen Triesch, Jan C. Vorbrüggen, Rolf P. Würtz, and Stefan Zadel. . *Autonomous Robots*, 6(2):203–221, 1999.
2. D. Beymer and T. Poggio. Image representations for visual learning. *Science*, 272:1905 – 1909, June 1996.
3. R.B. Cattel. The Scree test for the number factors. *Multivar. Behav. Res.*, 1:245 – 276, 1966.
4. Christian Eckes and Jan C. Vorbrüggen. . In *Proceedings WCNN96*, pages 868–875, San Diego, CA, USA, 16–18 September, 1996. INNS Press & Lawrence Erlbaum Ass.
5. E. Kefalea. . In *Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (IECON 98)*, Aachen, Germany, 1998.
6. 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 Transactions on Computers*, 42:300–311, 1993.
7. Hiroshi Murase and Shree K. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.
8. Kazunori Okada. *Analysis, Synthesis and Recognition of Human Faces with Pose Variations*. PhD thesis, Computer Science Department, University of Southern California, 2001.
9. J. B. Tenenbaum. Mapping a manifold of perceptual observations. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing*, volume 10, pages 682–688. MIT Press, 1998.
10. J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12(6):1247–1284, June 2000.
11. J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319 – 2323, December 2000.
12. Jan C. Vorbrüggen. , volume 47 of . , Thun, Frankfurt am Main, 1995.
13. J. Wieghardt and C. von der Malsburg. Pose-independent object representation by 2-d views. In *IEEE International Workshop on Biologically Motivated Computer Vision, May 15-17, Seoul*, 2000.