Reasoning about Occlusions during Hypothesis Verification

Charlie Rothwell

INRIA, 2004, Route des Lucioles, Sophia Antipolis, 06902 CEDEX, France

Abstract. In this paper we study the limitations of current verification strategies in object recognition and suggest how they may be enhanced. On the whole *object topology* is exploited little during verification. In practice, understanding the connectivity relationships between features in the image, or on the object, can lead to significantly more accurate evaluations of recognition hypotheses. We study how topology reasoning allows us to hypothesize the presence of occlusions in the image. Analysis of these hypotheses provides information which turns out to be crucial to the quality of our overall verification results.

1 Introduction

In an object recognition system, the process which is most used to differentiate between identification hypotheses is the final stage, called *verification*. The typical processing path for a recognition system is to start off by extracting features from the image, then to do feature grouping and model selection (indexing), and penultimately a correspondence stage is used to pair together sets of model and image features. Finally, in mature systems a verification step is then included which ultimately determines the degree of correctness of the model-image correspondences. Of course not all recognition systems follow this path, and often we find that verification is not performed as a separate process to correspondence. Nevertheless, a fairly general list of recognition algorithms which record the variations in the methods contains the efforts of Ayache and Faugeras [1], Pollard, *et al.* [9], Grimson and Lozano-Pérez [6], Bolles and Horaud [2], Faugeras and Hébert [5], Thompson and Mundy [14], Huttenlocher and Ullman [7], Lamdan and Wolfson [8], Stein and Medioni [13], and Califano and Mohan [3].

In this article we study how verification based on understanding model and image *topology* leads to better verification results. We use the term topology to represent the connectivity relationships between features. The importance of the use of topology in vision is discussed more completely in [10], here it is sufficient to state that *none* of the processing techniques discussed in this paper would be possible without a suitable understanding of topology. Typically, verification is based entirely on *geometric* methods.

Locally, topology conveys whether two model features are adjacent (connected) and should be seen as such in an image. Globally it reveals whether a continuous chain of features is present between any pair of primitives, and so defines notions of global connectivity. Topology thus enables neighbourhood based inferences. The observance of a particular model feature in an image would most likely indicate that all features adjacent to it should also be visible. The failure to find the adjacent features in the image would thus indicate either of two things: perhaps an occlusion is present and so the features are hidden; or conversely that the hypothesis is wrong and should either be discarded or at least have its importance diminished. Discrimination between these two types of incident can only be achieved by simultaneous topological and occlusion analysis.

Commonly verification is done as follows: the final conclusion of a recognition algorithm is that a set of individual model features matches a set of individual image features (rather than just saying that the model matches the image). The precise set of correspondences implies a geometric mapping from the model to the image. Additionally, one would also recover a measure of the number of matching features as a percentage of the whole; this number would form the basis of a verification score on which a criterion for the acceptance of the hypothesis would be built. The actual score would be computed by projecting all of the model features (perhaps a set of line segments, or even edgels from an acquisition image) to the image, and then counting the number of projected features which find *image support*. The hypothesis is accepted if a certain proportion of the features are matched. A test for image support of a projected model feature could be whether the feature lies close to an observed image feature which has a similar orientation. A complete description of such a verification strategy can be found in [14]. Note especially that approach is entirely *geometric*.

In the rest of this article we build on the work of [12], but include the following considerations in order to improve the verification method of the previous paragraph:

- Negative evidence must be explained. Primarily we should be able to locate occlusion events which justify the lack of measurement of a scene-model match. Failure to find occlusion events reduces the likelihood that a hypothesis is correct.
- Unless there is positive evidence of occlusion, some notion of object topology must be preserved. Therefore two image features should not be marked as coming from adjacent features on an object's boundary unless they are either connected in the image, or unless there exists an occlusion event between them.
- Under generic viewing conditions there must be uniqueness of solution. A single feature cannot belong to more than one object.

This last point requires development. The nature of the results given by different recognition systems varies dramatically depending on the application. Often the recognition problem is posed as the task of finding a specific object in a scene, and then immediately terminating the processing [6]. This problem is significantly different from, and easier, than that of identifying all objects in a scene which might correspond to any of a number of objects in a model base (that is a complete evaluation of the scene). In this situation, a single mistake can lead to catastrophic failures in interpretation because any one decision influences all subsequent processing. Thus, we would be likely to find that accepting a particular hypothesis might in some way prevent the formation of another hypothesis (particularly if the former is incorrect). It is perhaps therefore wise to be conservative and to allow multiple interpretations so that no truly correct hypothesis is discarded. Certainly, in light of the results given in [12], where it was stated that a significant number of false positives are likely to be recovered in a scene, we should at first follow this line of thinking.

However, most applications might be expected to provide unique and accurate scene interpretations. We are therefore interested in moving towards the notion of single feature interpretations whilst still working in relatively unstructured and unknown environments. Unfortunately, the immaturity of most verification schemes makes this difficult. It is for this reason that we have taken another look at the competences of these schemes, and we see that combining a topological representation with our previous geometric measures provides an initial step towards improvement.

We now describe how topological reasoning can be used as an aid to verification. First we show how a system such as that described in [12] produces incorrect recognition results due to the frailty of geometric verification methods, and then we show how topology reasoning introduces opportunities for occlusion analysis leading to more robust recognition. A more complete version of this paper is available as [11].

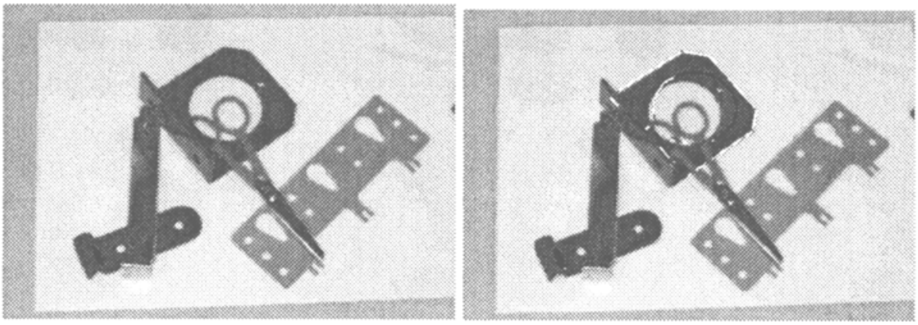

Fig. 1. *A model from the library having an outer boundary similar to that of the bracket visible in the image is hypothesized as being present in the scene.*

2 A sample recognition system

The recognition system we base our work on is called LEWIS, which uses invariant-indexing to form recognition hypotheses. A complete review of LEWIS can be found in [12], as well as examples of the system recognizing objects and performance statistics. Here we concentrate on examples failures of the system in identifying objects.

A characteristic example of recognition failure is shown in Fig. 1. After edgel detection and extraction of lines and conics, features are grouped together to form indexes based on plane projective invariants. Five lines around the boundary of the bracket just above centre-left in the image happens to have an index value which matches the invariants for a particular object which is distinct from the bracket in the model base.

Under verification the model for the incorrect hypothesis is projected to the image. Within LEWIS a model is represented as a set of edgel data recovered from an acquisition view of the isolated object, and a combination of the fitted lines, conics, and computed invariants for the edgel data. In essence, only the edgel data of the model are projected to the image. These are matched to image edgels using distance and orientation criteria: their orientations must differ by less than fifteen degrees and separations by less than five pixels. In the right-hand image of Fig. 1 the edges which found complete support are drawn in white, those which found matching image edgels within five pixels but whose orientation differed too much are drawn in grey (these may not be visible), and those failing both matching criteria are marked in black. In this example 55.0% of the model features were found to have complete image support. Any score over 50% is considered by LEWIS to be sufficiently high, and so the hypothesis is marked as accepted. Obviously there is an error as the wrong object has been identified. However, using geometric verification measures it cannot be ruled out.

A different failure is shown in Fig. 2. Here a false positive is created due to the presence of spurious scene features (linear texture). The problem arises because unconnected image features provide support for the model hypothesis over a large area. In this case 55.2% of the model edgels found matches in the scene even though only a few of them actually projected onto the features used for indexing.

3 Enhanced verification methods

Figures 1 and 2 highlight two areas where the insufficiency of current reasoning causes verification to fail. We cannot correct these errors using purely geometrical processes but rather must exploit integrated topological structure. Advances can be made rapidly

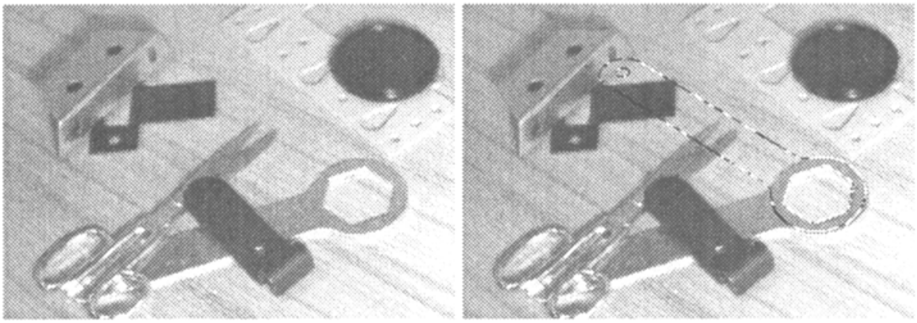

Fig. 2. *In this case an invariant configuration causes the estimation of the incorrect pose of an object. This is due to symmetry on part of the model.*

when we start to diffuse the local acceptance or rejection of a hypothesis around an object boundary. In short, in considering the acceptance of any one model element, we must analyse the outcome of the verification procedure for its neighbouring elements.

The two driving notions in verification thus become *topology* and consequently the understanding of *occlusion*. The topology of the image features which are assigned correspondences to model features must match the topology of the model features exactly. Exception can only be permitted due to occlusion, or when we are faced with the time-old problem of extracting reliable segmentations from images. We thus need to develop algorithms for analyzing the topology of features and for estimating the presence of occlusions. Now, in cluttered scenes we are very seldom able to recover image support for the entire boundary of the projection of a model hypothesized through indexing. Often this is because the objects are occluded in scenes, or perhaps they may suffer partial self-occlusion if they possess their own three-dimensional structure. The key issue is that the projection of the model into the image indicates where occlusions might arise (due to the loss of image support) and so our task becomes that of finding independent evidence for the occlusions. If we cannot, then it is likely that the original recognition hypothesis is incorrect and so we might do better by considering a different interpretation.

Occlusion events are typically marked by the presence of 'T' junctions in the image edgel structure. Generically, an object feature which undergoes occlusion will be cut and terminated by a locally straight transverse line segment. Thus when occlusions are hypothesized by the sudden loss of image support for the projection of the model we should look for 'T' junctions. Hypothesizing occlusions is roughly done as follows (see [11] for a more thorough treatment): find all of the projected model edgels which have image support, these form connected sub-sets. Then, at the boundaries of each of these sets (where image support is first totally lost), hypothesize an occlusion event.

We have evaluated occlusion hypotheses in two ways. The first uses the edgel data computed as the initial step of image processing. Initially we recover all of the junctions in the original edgel image which have an order greater or equal to three. Junctions of order three are those where three edgel chain curves meet at a single point. We declare the presence of a 'T' junction when the angle between any pair of the edges meeting at an order 3 junction are within twenty degrees of 180 degrees. The large tolerance reflects the fact that edgel contours are frequently displaced by significant amounts near junctions, and it also allows for occlusion by curved objects. Then, if the edgel-'T' junction lies sufficiently close to where the recognition hypothesis deduced that there should be an occlusion event, then we can add confidence to the original recognition hypothesis.

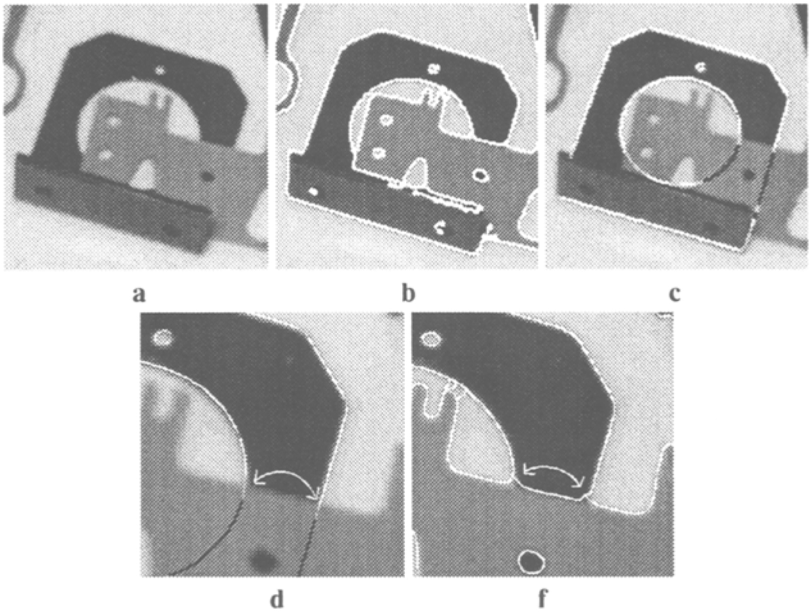

Fig. 3. *(a) shows part of an image which includes an object from the model base. In (b) are edgels from the edgel detector and (c) the (correct) projection of the model. This suggests occlusions at the points marked by the arrows in (d). Near to these points are triple junctions in the edgel description which have the forms of 'T's, as shown in (f).*

In practice this type of processing does not resolve all of the occlusion events we find in an image. This is because edgel detectors are notoriously poor at recovering meaningful connectivity at junctions. Consequently, it will seldom recover all of the 'T' junctions. We therefore resort to a second test which examines the overall structure of the image intensity data near to the locations of each hypothesized occlusion. This is done by parametric model fitting to junctions such as suggested by Deriche and Blaszka [4], though there are many other approaches to model fitting, some of which are mentioned in [11]. The type of parametric model fitted by [4] assumes that the surface is composed of a number of constant intensity plateaux which meet at the junction and are separated by straight edges. Each plateau represents an image region and smoothing is accounted for between the image regions through an approximate parametrization of Gaussian smoothing. The algorithm of [4] fits such a model over a specific window size at a given location in the image (where we suspect that there is an occlusion), and returns a number of different parameters which represent the interpretation of the intensity surface. The key measures which are returned are a fitting cost, the grey level values of the plateaux, and the angles at which the edges come into the junction. We can then estimate whether the junction is a real 'T' junction by looking at the angles between the edges and by making sure that the plateaux have sufficiently different grey levels.

Examples - edgel contour junctions

In Fig. 3 we show how data from the edgel detector hypothesizes 'T' junctions near to where occlusion events should be found on the strength of a specific recognition hypothesis. This is an example of positive support for an object hypothesis, with the measurement of the low-level junction description enhancing the confidence in a hypothesis. In

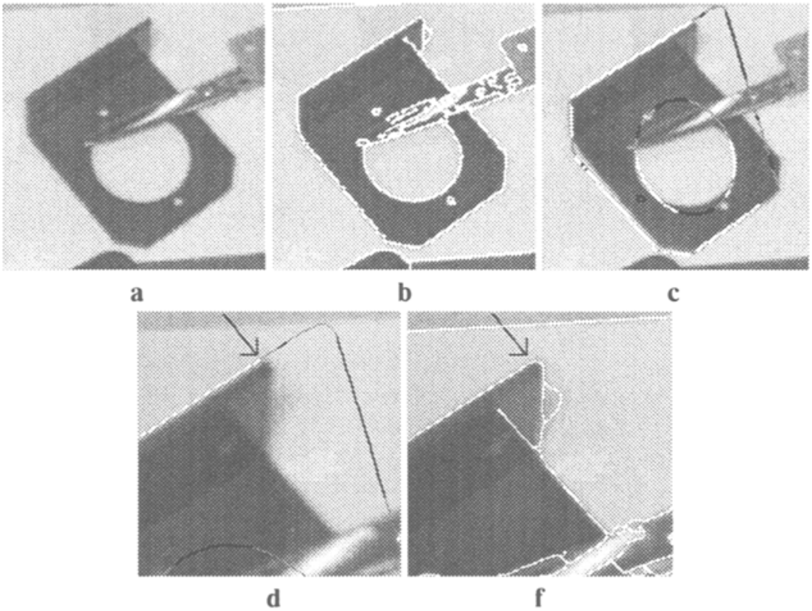

Fig. 4. *The edgels for sub-image (a) are shown in (b). LEWIS deduces the incorrect hypothesis (c). Near the white-grey transition of the projected outline in (d), we expect to find a junction in the edgel description (that is if the hypothesis were correct). However, as shown in (f) the failure of the edgel detector to record a junction in the right place suggests that the hypothesis is false.*

Fig. 4 we demonstrate an example of negative evidence which renders a hypothesis unlikely (or perhaps may even cause it to be rejected outright). The projected model curves are shown initially following a contour in the image, but at one stage (due to an incorrect model being hypothesized), the projected model curve parts quite clearly from the image curve. The edgel detector fails to find any junctions near the point of departure and so one can start to doubt whether the original hypothesis is correct.

This type of reasoning appears very attractive on the weight of the two examples given. However, most edgel detectors have difficulty recovering correct edgel contour connectivity near junctions. This means that failure to record the presence of a 'T' junction can be taken as only partial evidence against a hypothesis, and should not be used too strongly. Conversely, if we are able to find a 'T' junction where predicted by a recognition hypothesis then we can add considerable weight to the hypothesis. The progress we make in marking 'T' junctions is apparent. Henceforth, the projected edgel data from the hypothesis which originally failed to gain image support (and so was treated neutrally) can be treated as *positive support* if associated with a confirmed occlusion event.

Examples - parametric junction models

The alternative approach to evaluating 'T' junctions fits parametric junction models. In Figs 5 and 6 we show similar examples to those for the edgel junction reasoning, but this time 'T' junctions are evaluated using the Deriche-Blaszka operator. In the first case a pair of suitable 'T' junctions are recovered and shown superimposed in the figure, and in the second case no such fit could be found and so the hypothesis is marked as being unlikely. Thus we see that parametric model fitting can enhance the understanding of

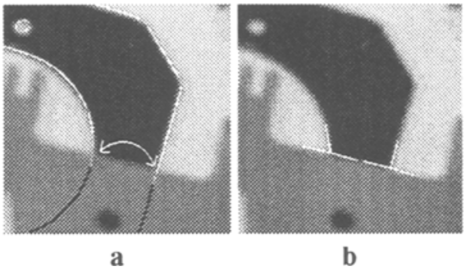

Fig. 5. *The output of the Deriche-Blaszka junction detector is shown in (b). Both junctions appear to be sufficiently close to 'T' junctions.*

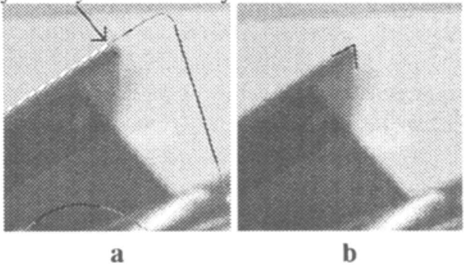

Fig. 6. *The Deriche-Blaszka filter fails to find a suitable 'T' junction near the potential occlusion event already discussed in Fig. 4. Instead it finds the 'L' junction shown in (b). Such a feature has no relation to an occlusion event, and so the original hypothesis is likely to be false.*

recognition hypotheses when directed by prior topological reasoning.

Again, these examples of the use of the parametric junction model approach do not describe the whole truth. Over repeated trials we have found that the Deriche-Blaszka model does not actually represent the image intensity surface correctly. In [11] we discuss to some extent why this is the case; overall more work is required in developing junction detectors which model the image intensity surface correctly.

In summary, we have so far demonstrated how occlusion events can be hypothesized by studying the model topology information contained in hypotheses produced by a typical recognition system. We have also seen that hypothesized occlusions can be evaluated in different ways. Two methods have been discussed: the first using bottom-up information recovered from the original output of an edgel filter; and the second derives top-down data resulting from the application of parametric junction model fitting. For certain cases the second method is more accurate and more reliable, but in general it relies on making incorrect assumptions about the shape of the intensity surface (which is seldom composed of smoothed constant-intensity plateaux). Nevertheless we can employ both methods with caution. Whenever either approach suggests the presence of a 'T' junction, we can be fairly confident that it is right. However, they both frequently reject junctions which do actually correspond to occlusion events.

3.1 How to update hypotheses

Now, whenever we find an occluded region which is terminated at one end by a verified 'T' junction, it can be marked as making a positive contribution to the hypothesis. This is to say that the score of *visible model edgels* used to compute the overall verification score should be incremented by the number of edgels in the occluded region. Consequently we can transform a hypothesis such as that in Fig. 7 from a 70.5% score to

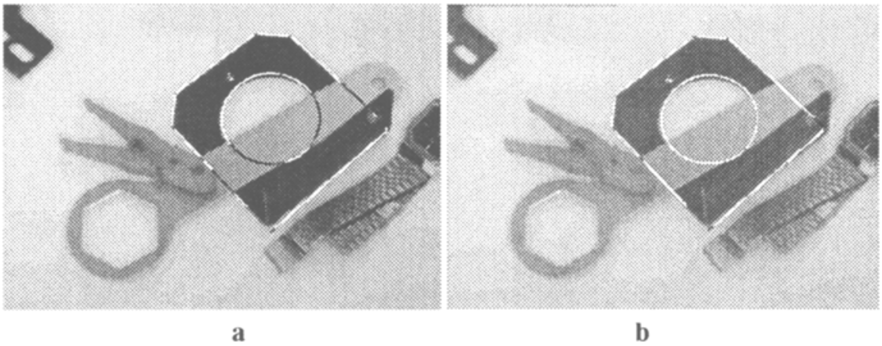

Fig. 7. *The original verification method produced only a* 70.5% *score for the hypothesis shown in (a). However, after the prediction and verification of the various 'T' junctions bounding the occluded parts of the hypothesized object, we increase the verification score to* 93.6%. *Edgels which were previously unmatched, but are now marked as positively occluded, are depicted in white in (b). The overall verification score is incremented by the number of edgels in the positively identified occluded regions.*

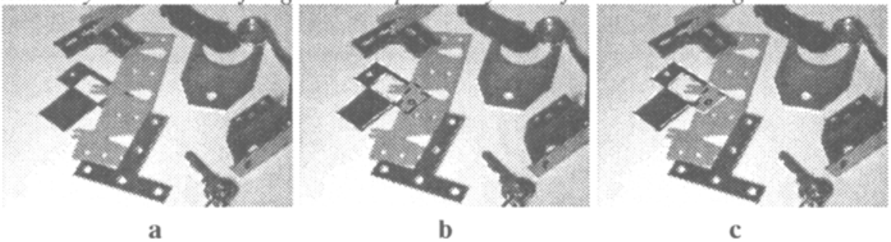

Fig. 8. *The original verification score for the object in (b) was* 70.7%. *After occlusion reasoning this rises to* 83.6% *which is a clear indication that the hypothesis is correct.*

a re-evaluated score of 93.6%, and hence have little doubt that the hypothesis is correct. Ideally we would hope that the new score would tend towards 100%, but there are always short sections of projected model curve which project near to image features with the wrong orientation (and so are not marked as being caused by occlusion). Taking these into account would make the hypothesis in Fig. 7 take on a score of little less than 100% (in fact 98.2% when we ignore small section of less than five pixels in length). The dominance of a good hypothesis such as this one significantly enhances our understanding of the scene. From a number of experiments we are able to conclude that a final matching score of over 90% leaves little doubt as to the identity of an object. Another example is in Fig. 8 where the score of an occluded object rises from 70.7% to 83.6% after occlusion reasoning (and 91.3% after removal of short unmatched chains).

3.2 Correctness of image topology

So far in this paper the effects we have been interested in have been dominated by model topology and cause-and-effect reasoning for connected parts of the model to be either occluded or visible. It is quite understandable that we can make reciprocal considerations with regard to image topology, or more properly between the consistency of both of the model and image descriptions. In an ideal world (where we would of course cease to be frustrated by problems in segmentation), the image topology should match the projection of the model topology exactly.

Even with our current segmentation abilities we can develop simple tests which

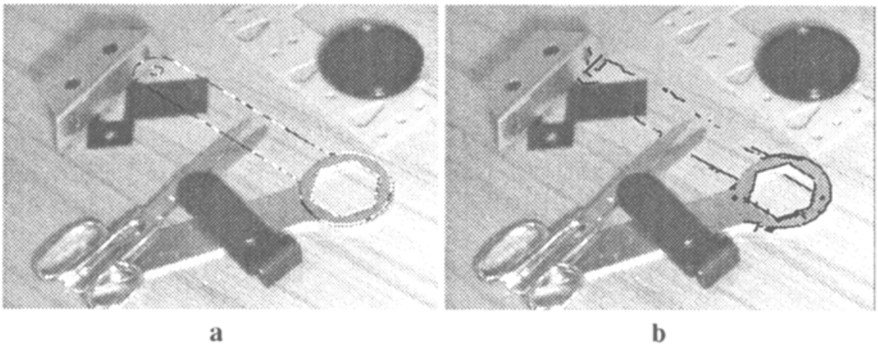

Fig. 9. *The poorly oriented hypothesis in Fig. 2 required support from unconnected texture features. These have an image topology which is inconsistent with that of the model. As there are no indications of occlusion events at the ends of each image curve, we infer that the correspondences are incorrect. Subsequently we can doubt the hypothesis.*

show some robustness to errors in the extraction of the low-level image description, and which indicate whether the grounds for believing a hypothesis should be reduced. For instance, we can start off by seeing which sets of image features have been given a correspondence with a single topological feature from a model. Due to the trivial connectivity constraints which exist on a lone model feature, one would also expect the corresponding image features to be connected. There are a number of different reasons for the image features to become distinct, though still to remain topologically linked. Perhaps the segmentation and fitting procedures separated the features at the geometric level by attributing each one to different geometric objects, or perhaps the edgel detector erroneously placed a junction between them to provide connectivity with other features. However, in both of these cases the image topology is consistent with a single model feature, and so we need not doubt the integrity of a particular hypothesis.

Conversely, we should mark cases in which connectivity has been lost. Of course there might be a perfectly reasonable explanation such as the presence of an occlusion event, but if not, we should add further doubt to the interpretation. Thus, our way of reasoning is again led back to the detection of occlusion events.

We show the success of this line of reasoning in Fig. 9. Model features have been projected into the image and have found sufficient image support along their lengths. However, the support has actually been provided by sets of unconnected image features. We thus test for the presence of occlusion events at the ends of the image features, and if they are not found, we mark the hypothesis as being unreasonable.

3.3 Uniqueness of description

By this stage of the proceedings the hypotheses have undergone a detailed level of topological analysis. Those which have been attributed near perfect scores are very likely to be correct, whilst those with poorer verification tallies may either be erroneous, or might just be suffering due to difficulties in segmentation or occlusion event detection.

We now return to the fact that a single feature in an image is caused by a single scene feature. Therefore, a consequence of recognition should be that the correspondence between model and image features is at most one-to-one. Should two hypotheses match a single image feature we can be sure that *at least one* of the hypotheses is incorrect, and so should try and eliminate the least likely. This process is risky should the confidence levels in the hypotheses be poorly defined, but as our abilities at verification improve,

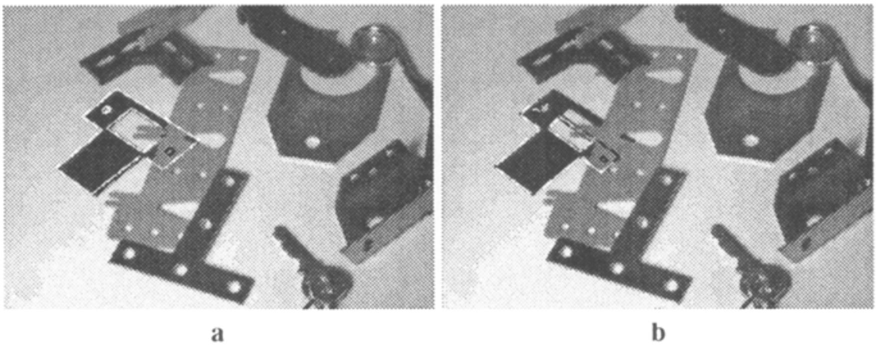

Fig. 10. *In Fig. 8 we were able to find two hypotheses which matched to common scene features. After all of the topological processing the hypothesis in (a) scored* 91.3% *and that in (b)* 68.2%. *Any score over* 90% *provides very strong confidence, and so we can eliminate any other hypotheses which match to the same image features.*

we can start to attribute error measures with reasonable accuracy.

Remembering that once occlusion analysis has been performed, we can mark hypotheses more clearly for acceptance or rejection. Thus we can proceed by accepting the single hypothesis which has gained the highest recognition score. Then, all of the image features which have been given a correspondence to any of the features in this model are marked as being explained. All other hypotheses which have correspondences with these image features are marked as being inconsistent, and rejected. We then take the next best hypothesis, and proceed by examining its image features. In short, this process ensures a uniqueness of description of the image features. An example of this type of reasoning is given in Fig. 10 where we are able to rule out the hypothesis in (b) because it shares scene features with the very highly scored hypothesis in (a).

4 Conclusions

In this article we have demonstrated how reasoning about model and image topology enhances our object recognition verification abilities. A typical recognition system such as that of [14] computes the final match score for any hypothesis by determining whether a set of independent model features finds support in an image. [12] demonstrated that such a geometric strategy does not produce conclusive recognition results. We have found that diffusing verification information around connected components of a model means that a lack of image support can actually be turned into *positive evidence.* This in turn means that verification thresholds can perhaps be raised from 50% up to somewhere in excess of a 90% level of image support.

The development of our verification algorithm involves reasoning about discrepancies between the model and image topologies. The differences are used to hypothesize where occlusion events should lie in the image. The presence of such events strengthens a recognition hypothesis, and the lack of one suggests that a hypothesis might be false. The detection of the junctions is done via both edgel detector output, and the Deriche-Blaszka feature detector [4]. Neither of these filters function perfectly, though when either hypothesizes the presence of a 'T' junction we can be relatively sure that it is correct.

Of course our results are not entirely complete. Whilst working with single images, we need to analyse the effects of other feature detectors. There are a large number of other filters which require evaluation in either of the domains of bottom-up or top-down processing. We also need to test the algorithms on a more varied range of objects of which a three-dimensional model base is the ultimate goal.

On a different level, we have only made use of the boundary information contained within the models. Such geometric primitives obviously provide very easy access to object descriptions. However, a full verification scheme should include analysis about surface properties such as texture, and also even colour. Certainly with the aid of top-down segmentation based on the recognition hypotheses one would be able to test out other object properties in conjunction with the more geometric aspects.

A more complete version of this paper is available through ftp as a technical report [11] from: <URL ftp://ftp.inria.fr/INRIA/publication/publi-ps-gz/RR/RR-2673.ps.gz>.

Acknowledgments

A number of people were involved with discussions during the original development of LEWIS and its re-implementation within the **TargetJr** software system initially developed at GE CRD (on which this paper is based). As such I would like to thank David Forsyth, Bill Hoffman, Joe Mundy, and Andrew Zisserman. Thierry Blaszka provided his implementation of the Deriche-Blaszka filter. I also acknowledge discussions with Theo Papadopoulo and Nour-Eddine Deriche, and Pippa Hook helped to ensure the correctness of the English in this article. This research was funded by a Human Capital and Mobility grant from the European Community.

References

[1] N. Ayache and O. Faugeras. HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects. *PAMI*, 8(1):44–54, 1986.

[2] R. Bolles and R. Horaud. 3DPO: A Three-dimensional Part Orientation System. *IJRR*, 5(3):3–26, 1986.

[3] A. Califano and R. Mohan. Systematic design of indexing strategies for object recognition. *Proc. CVPR*, p.709–710, 1993.

[4] R. Deriche and T. Blaszka. Recovering and characterizing image features using an efficient model based approach. *Proc. CVPR*, p.530–535, 1993.

[5] O. Faugeras and M. Hébert. The representation, recognition, and locating of 3d shapes from range data. *IJRR*, 5:27–52, 1986.

[6] W.E.L. Grimson and T. Lozano-Pérez. Localizing overlapping parts by searching the interpretation tree. *PAMI*, 9(4):469–482, 1987.

[7] D. Huttenlocher and S. Ullman. Recognizing Solid Objects by Alignment with an Image. *IJCV*, 5(2):195–212, 1990.

[8] Y. Lamdan and H. Wolfson. Geometric Hashing: A General and Efficient Model-Based Recognition Scheme. *Proc. ICCV*, p.238–249, 1988.

[9] S. Pollard, J. Porrill, J. Mayhew, and J. Frisby. Matching geometrical descriptions in three-space. *IVC*, 5(2):73–78, 1987.

[10] C. Rothwell, J. Mundy, and W. Hoffman. Representing objects using topology. In preparation, 1996.

[11] C. Rothwell. The importance of reasoning about occlusions during hypothesis verification in object recognition. TR 2673, INRIA, 1995.

[12] C. Rothwell. *Object recognition through invariant indexing*. Oxford University Press, 1995.

[13] F. Stein and G. Medioni. Structural Indexing: Efficient 3-D Object Recognition. *PAMI*, 14(2):125–145, 1992.

[14] D. Thompson and J. Mundy. Three-dimensional model matching from an unconstrained viewpoint. *Proc. ICRA*, p.208–220, 1987.