Integrating Vision Processing and Natural Language Processing with a clinical application

Peter Hall
Department of Computer Science
P.O. Box 600
Victoria University of Wellington
Wellington
New Zealand
Email peter@comp.vuw.ac.nz

Paul Mc Kevitt
Department of Computer Science
211 Portobello Street
Univeristy of Sheffield
Sheffield, GB- S1 4DP
England, UK
Email: p.mckevitt@dcs.shef.ac.uk

Abstract

It is an interesting observation that humans are able to combine information from ostensibly disparate sources in order to better comprehend the world around them. For example, the sense information from the eyes and the ears are combined at every instant – tasks such as road crossing would be far more difficult without such integration. But such combinations are non–linear in nature, so examining each sense in isolation will not shed light on how they combine within the cognitive frame. We are interested in the way humans combine their visual sense with natural language. This is an important area that appears to be little researched to date, despite the large amount of literature in both Vision Processing and Natural Language Processing. In response to the non–linear nature of the combination we proceed by synthesising the phenomenon rather than analysing it. To advance our work we focus upon a particular application that not only makes such a combination feasible but has a practical use in the field of medicine.

1. Introduction

Students find that books with both text and pictures in them are easier to understand than books with either alone. It appears that there is some cognitive process in action that is able to combine the information from the two sources into an understanding that is greater than the sum of its parts to produce a synergy between language and vision. Such a combination is non–linear which makes understanding the phenomena by a detailed analysis of its parts an impossible undertaking. Although the integration of language and vision is important to an understanding of cognition and has many practical benefits [1], the subject is little researched. We can posit that humans use a unique target representation when combining information from different sources. What this representation looks like is a matter for debate in both Philosophy and Psychology [2,3].

Because language and vision combine in a non–linear way we seek to synthesise the phenomenon rather than analyse it. That is, we construct a model that mimics certain aspects of complex behaviour rather than deconstruct the real thing. Investigating complex systems by synthesis has more in common with methods used in Artificial Life (AL) than with Artificial Intelligence (AI) which is the traditional umbrella for both Natural Language Processing (NLP) and Vision Processing (VP). However, unlike the AL approach we do not attempt to synthesise complex behaviour as an emergent property given simple rules. Rather, we maximise our use of the large bodies of literature that exist in both VP and NLP so our "starting point" comprises VP and NLP modules – this is at a much higher level than is typical in AL. What makes the integration practically feasible is that we have a Knowledge Representation (KR) independent of both the NLP and VP modules. The KR contains expert knowledge about blood vessel anatomy.

The specific task is to build a three dimensional model of blood vessels in a patient's brain (*cerebral vasculature*) given x–ray pictures and associated medical reports. The x–ray pictures are perspective projections of the cerebral vasculature and are separated by an angle of about ninety degrees (*biplane angiograms*). The medical reports are prepared by an expert Radiologist while examining the angiograms. The three–dimensional model our system produces is a *reconstruction* that should account for all input angiograms and medical reports. As Hall *et al.*[4] report, the clinical motivation for this task is to locate any malformations within the three dimensional vascular bed.

This area is suitable to investigate the integration of NLP and VP for six reasons: (1) The medical reports are brief, well structured, and exhibit the characteristics of a sub–language [5]. They describe particular vasculature in angiograms and this information can be extracted from

0-8186-7174-2/95 $04.00 © 1995 IEEE

them; (2) The angiograms are complicated but methods exist for extracting relevant information from them; (3) The KR in our system models a collection of vasculature. Each entry in the collection is an anatomical model of a particular vasculature. This makes the KR a suitable target for both NLP and VP modules; (4) Reconstruction of vasculature from either angiograms or medical reports alone is inherently ambiguous. Integration helps disambiguate candidate reconstructions because additional information is provided; (5) The KR is incrementally manufactured by adding new reconstructions. So integration addresses the *knowledge elicitation* problem [6]; (6) The integration means our KR will hold anatomically correct models whose parts are assigned correct clinical terminology. This would not be possible from a single source of information.

2. Background

Reconstruction of a vascular system from biplane angiograms poses an under–determined problem that is soluble only if a priori information is given. Previous approaches have been expert systems that provide rules which are used during reconstruction [7,8,9,10,11]. Hall [12] gives an overview of such methods. However, none of these methods is linked to NLP. With regard to medical text processing, systems for analysis have been built [5,13,14,15]. However, none of these is linked to VP. There have been some efforts to integrate NLP and VP. These analyse images to produce text [16,17,18]. None of these act in the clinical domain and, importantly, none analyse both text and images to produce a single representation. We differ in that our efforts are devoted to combining analyses from different sources into a single representation. We begin by treating the NLP and VP modules as independent components that can communicate both directly with each other as well as via our KR. We present the integrating architecture below, but first we discuss the KR, VP, and NLP modules.

3. The knowledge representation

Our KR for a collection of vasculature is described in Hall and McGregor [19] and Hall [20]. With respect to the current discussion the important aspects of the KR are that: (1) it is no more than an information repository about a collection of vasculature, and (2) information can be added to it incrementally.

The collection of vasculature is represented by a single graph that we call a *proto–graph*. Nodes in the proto–graph correspond to places where vessels divide or rejoin (*furcations*) while arcs correspond to *vessel segments* that connect furcations. *Vessels* as recognised by clinicians [21] are paths, or possibly small subgraphs, in the protograph. The proto–graph is manufactured by adding new instances. These instances are graphs that represent the topology of particular vasculature. Vascular topology can be recovered from the proto–graph using a Boolean valued function that we call the *discrimination proposition*. This discriminates between arbitrary subgraphs of the proto–graph and those that were component to its manufacture. The proto–graph is labelled with geometric and statistical information. The geometrical information is used to fix the shape of the vasculature in three–dimensions while the statistical information records variances in both geometry and topology. The proto–graph is also labelled with correct clinical terminology. In short, the KR is a catalogue of vasculature and their variations. As such it has no specific purpose but can be used by a variety of applications.

All of the information in the KR is acquired from data sources such as angiograms or medical reports. For example, suppose a reconstruction has been produced from angiograms and medical reports. Expert clinicians will be consulted to confirm the reconstruction before it is added into the KR. The KR therefore reflects expert opinion as to the nature of vascular anatomy. In general the topology, the geometry, and the statistics of the KR will all be updated by adding new entries. In this way the performance of applications that use the KR, such as reconstruction, should improve over time.

4. The Vision Processor

By itself the vision processor accepts angiograms as input and produces a set of candidate reconstructions as output. Each candidate in the set accounts for the input angiograms. For the purposes of reconstruction it is helpful to represent each angiogram as a graph that we call an α–*graph*. In this graph the furcations correspond either to images of furcations in the vasculature or to *crossings*. A crossing is an artefact of projection created by the apparent intersection of arcs in three dimensions. Because arcs are curved they can appear to intersect themselves and such places are crossings also. The arcs in the graph that represents the angiograms represent fractions of a vessel segment (*slivers*) that connect furcations and crossings.

Reconstruction can now be described as finding a maximal, consistent set of mappings that carry nodes in the KR to nodes in every α–graph, and arcs in the KR to nodes and arcs in every α–graph. The reason for the latter mapping is that crossings in the α–graph arise from vessel segments in the KR. If the set of nodes and arcs discovered this way satisfy the discrimination proposition then the reconstruction has been "recognised" by the VP as being an individual entry in the catalogue. Otherwise several options exist: (1) the reconstruction is a subgraph of a catalogue entry; (2) the reconstruction comprises subgraphs from several catalogue entries; or (3) the reconstruction contains entirely new pieces. In such cases an expert clinician can be prompted to confirm or complete the reconstruction.

5. The Natural Language Processor

By itself the NLP module accepts a medical report as input and produces a set of candidate reconstructions as output. Each candidate in the set accounts for the input medical report.

The NLP module must construct a set of graph structures from its input text to match against the KR and so yield a final result. Initially the NLP module produces a graph, that we call the ρ–graph whose nodes correspond to features that are named in the given report. Typically these features are vessels and any associated malformations. In addition, it is typical for medical reports to describe how these features are connected to one another, and give weak geometric relations also. Importantly, the medical report locates unusual structures that are present in the vasculature. This helps when building a reconstruction including any malformations – which is the clinical motivation behind this work. All of this information is used in constructing a ρ–graph.

Reconstruction comprises finding a set of consistent, maximal mappings from the KR to the ρ–graph. In this case, groups of vessel segments in the KR must be matched against a single node in the ρ–graph (recall that in clinical terminology a vessel is a path or subgraph in the KR). The furcations in the KR map to arcs in the ρ–graph. Once a set of reconstructions have been made they can each be tested against the discrimination proposition in the same way as reconstructions from the VP.

6. An integrated system

The architecture of the integrated vision and language system is shown below in Figure 1. So far the KR, VP, and NLP modules have all been outlined. We now discuss the operation of the decision module and the purpose of the communication channels.

Suppose the VP and NLP modules each worked independently to produce a set of reconstructions. Each of these reconstructions should be regarded as an hypothesis. We can discover where the modules agree by intersection of hypotheses. The task of the decision module is to perform such an intersection – we expect that the intersection of sets will usually yield a set smaller than either of them. Thus we see that the benefit of combining NLP and VP is that they disambiguate one another. More generally we can consider table 1, below, where G_{vis} and G_{lang} are the hypotheses

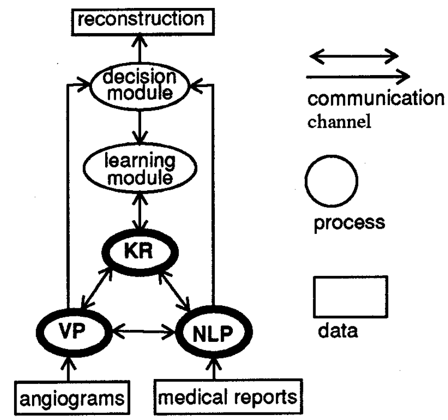

Figure 1. The integrated system architecture

from the VP and NLP modules respectively, ∅ is the empty set, and "match" denotes a non–empty set. Of the possibilities shown the most useful has the interpretation "full agreement" when ($G_{lang} \cap G_{vis}$) is a singleton set. In this case the result is optimal because there is a unique solution agreed upon by independent methods. The integration therefore provides a consistency check.

Table 1 Decision table for integration

G_{lang}«G_{vis}	G_{lang} / G_{vis}	G_{vis} / G_{lang}	Row interpretation
∅	∅	∅	no reconstruction
∅	∅	match	vision only
∅	match	∅	language only
∅	match	match	no agreement
match	∅	∅	full agreement
match	∅	match	vision dominant
match	match	∅	language dominant
match	match	match	inconclusive

We argue that if the NLP and VP are allowed to communicate directly then this analysis will not be altered. This is because we have added no new information into the overall system so that no further ambiguity resolution is possible. Rather, the system will become more efficient.

The learning module, who's task is to add new instances of individual vasculature to the proto–graph, has not been discussed; its operation is fully discussed elsewhere [19].

7. Conclusion

We have described how the power of NLP and VP might be harnessed together into a single system. The key to our integration is the use of a single target representation. Their combination is more powerful than either one in isolation because they mutually disambiguate and confirm one another. In this way we have a synergy between the component parts. The fact that the system will provide extra information to the KR, and will be more robust and more efficient are important practical outcomes. We conclude that work on integrating complex systems can play an important role in understanding the cognitive process in humans and therefore impact upon Artificial Intelligence, Cognitive Science, Psychology, and Philosophy.

8. References

[1] W. Wahlster, "*One word says more than a thousand pictures. On the automatic verbalization of the results of image sequence analysis.*" Bereich Nr. 25, Universitat des Saarlandes, FB 14 Informatik IV, Im Stadtwald 15, D–6600, Saarbrucken 11, Germany, EU, February 1988

[2] Z. Pylyshyn, "What the mind's eye tells the mind's brain: a critique of mental imagery," *Psychological Bulletin,* 80, 1973, pp 1–24.

[3] S.M. Kosslyn, and J.R. Pomerantz, "Imagery, propositions and the form of internal representations." *Cognitive Psychology,* 9, 1977, pp 52–76

[4] P.M. Hall, R. Feltham, R., and T. Fitzjohn, "Automated analysis of x–ray angiograms," *New Zealand Computer Society Conference* 1995 (forthcoming). Also in Technical Report CS–TR–94–14. Dept. Comp. Sci., Victoria University of Wellington, Wellington, New Zealand, 1994

[5] M. Schröder, "Knowledge based analysis of radiology reports using conceptual graphs," In *Proceedings of the Seventh Annual Workshop on Conceptual Graphs,* Heather Pfeiffer (Ed.), pp 213–222, New Mexico State University (NMSU), Las Cruces, New Mexico, USA, July 1992.

[6] D. Diaper, and Kidd (ed.), "*Knowledge elicitation, principles, techniques and applications,*" Chichester: Ellis Horwood. 1989

[7] D. Delaere C. Smets, P. Suetens, and G. Marchal, "A knowledge–based system for the angiographic reconstruction of blood vessels from two angiographic projections," In *Proc. of the North Sea Conference on Biomedical Engineering,* Nov. 1990.

[8] M. Garreau, J.L. Coatrieux, R. Collerec, and C. Chardenon, "A knowledge based approach for reconstruction and labelling of vascular networks from biplane angiographic projections," IEEE Transactions on Medical Imaging, 10(2), 1991, pp 122–131.

[9] S.T. Rake, and L.D.R. Smith, "The interpretation of x–ray angiograms using a blackboard control architecture," *Proc. Computer assisted radiology* 1987, pp 681–685.

[10] S.A. Stansfield, "ANGY: A rule base expert system for automatic segmentation of coronary vessels from digital subtracted angiograms," *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 8(2), 1986, pp188–199:

[11] P. Suetens, van Cleynenbreugel, J., Fierens, F., Smets, C., Oosterlink, A., and Wilms, G., "An expert system for blood vessel segmentation on subtraction angiogram, " *SPIE 767 Medical Imaging,* 1987, pp 454–459

[12] P.M. Hall, "*A review of procedures for reconstruction of the vasculature from angiograms,*" Technical Report CS–93–06, Dept. Comp. Sci., University of Sheffield, Sheffield, UK. 1993

[13] N.C. Sager, C. Friedman, and M.S. Lyman, "*Medical language processing: computer management of narrative data,*" Reading, MA: Addison–Wesley. 1987

[14] D.L. Ranum, D.L. "Knowledge based understanding of radiology text," In *12th Annual Symposium on computer applications in medical care,* 1988, Greenes, R.A. (Ed.), IEEE Computer Society Press, November.

[15] R.H. Baud, A.–M. Rassinoux and J.–R. Scherrer, "Knowledge representation of discharge summaries." In *Proc. of The Third Conference on Artificial Intelligence in Medicine* (AIME 1991), pp 173–182, Maastricht, June.

[16] G. Herzog, and G. Retz–Schmidt, "Das System SOCCER: Simultane Interpretation und naturalichsprachliche Beschreibung zeitveranderlicher Szenen," In *Sport und Informatik,* J. Perl (Ed.), 95–119, Schorndorf: Hofmann. 1990

[17] E. André, G. Herzog, and T. Rist, "On the simultaneous interpretation of real–world image sequences and their natural language description: the system SOCCER," In *Proc. of the 8th European Conference on Artificial Intelligence,* 449–454, Munich, Germany. 1988

[18] W. Maaß, P. Wizinski, and G. Herzog "*VITRA GUIDE: Multimodal route descriptions for computer assisted vehicle navigation,*" Bereich Nr. 93, Universitat des Saarlandes, FB 14 Informatik IV, Im Stadtwald 15, D–6600, Saarbrucken 11, Germany, EU, February 1993.

[19] P.M. Hall, and McGregor, J.J. "A graph based model of a collection of physical vasculature." *Proc. Digital Image Computinf Techniques and Applications* 1993, Sydney, Macquarie University, pp 414–421.

[20] P.M. Hall, "Simulating angiography," *Mathematical Modelling and Scientific Computing,* Vol 6, 1996 (forthcoming). Also in Technical Report CS–TR–94–15. Dept. Comp. Sci., Victoria University of Wellington, Wellington, New Zealand. 1994

[21] G. Salamon, and Y.P. Huang, "*A radiological anatomy of the brain,*" Springer-Verlag, Berlin. 1976