

AAAI-94 Workshop on Integration of Natural Language and Vision Processing

Twelfth National Conference on Artificial Intelligence (AAAI-94)
Washington State Convention and Trade Center
Seattle, Washington, USA

Tuesday/Wednesday, August 2nd/3rd, 1994

Chair:

Paul Mc Kevitt

Department of Computer Science
University of Sheffield, England, EU
E-mail: p.mckevitt@dcs.shef.ac.uk

Workshop Committee:

Prof. Mike Brady (Oxford, England)
Prof. Jerry Feldman (ICSI, Berkeley, USA)
Prof. John Frisby (Sheffield, England)
Prof. Frank Harary (CRL, New Mexico, USA)
Dr. Eduard Hovy (USC ISI, Los Angeles, USA)
Dr. Mark Maybury (MITRE, Cambridge, USA)
Dr. Ryuichi Oka (RWC P, Tsukuba, Japan)
Prof. Derek Partridge (Exeter, England)
Dr. Terry Regier (ICSI, Berkeley, USA)
Prof. Roger Schank (ILS, Illinois, USA)
Prof. Noel Sharkey (Sheffield, England)
Dr. Oliviero Stock (IRST, Italy)
Prof. Dr. Wolfgang Wahlster (DFKI, Germany)
Prof. Yorick Wilks (Sheffield, England)

Table of contents

Preface	iii
Programme	v
Domain-independent rules relating captions and pictures Neil C. Rowe	1
An image retrieval system that accepts natural language Hiromasa NAKATANI & Yukihiro ITOH	7
Integrating text and graphical input to a knowledge base Raman Rajagopalan	14
Photo understanding using visual constraints generated from accompanying text Rohini K. Srihari	22
Audiovisual recognition of speech units: a tentative functional model compatible with psychological data Jordi Robert-Ribes, Michel Piquemal, Jean-Luc Schwartz & Pierre Escudier	30
The spoken image system: on the visual interpretation of verbal scene descriptions Sean Ó Nualláin, Benoit Farley & Arnold G. Smith	36
Behavioural descriptions from image sequences Hilary Buxton & Richard Howarth	40
Visions for language Paul Mc Kevitt	47
Language animation A. Narayanan, L. Ford, D. Manuel, D. Tallis & M. Yazdani	58
Assembly plan generation by integrating pictorial and textual information in an assembly illustration Shoujie He, Norihiro Abe & Tadahiro Kitahashi	66
Multimedia presentation of interpreted visual data Elisabeth André, Gerd Herzog & Thomas Rist	74
Talking pictures: an empirical study into the usefulness of natural language output in a graphical interface Carla Huls, Edwin Bos & Alice Dijkstra	83
From verbal and gestural input to 3-D visual feedback Yuri A. TIJERINO, Tsutomu MIYASATO & Fumio KISHINO	91

An integration of natural language and vision processing towards an agent-based future TV system	
Yeun-Bae Kim, Masahiro Shibata & Masaki Hayashi	99
An AI module for reference based on perception	
John Moulton & Lawrence D. Roberts	107
Instruction use by a vision-based mobile robot	
Tomohiro Shibata, M. Inaba & H. Inoue	115
A cognitive approach to an interlingua representation of spatial descriptions	
Irina Rezero-Sans & Jun-ichi Tsujii	122
Describing spatial relations in weather reports through prepositions	
Stephan Kerpedjiev	131
Prepositional semantics in the WIP system	
Patrick Olivier & Jun-ichi Tsujii	139
Spatial layout identification and incremental descriptions	
Klaus-Peter Gapp & Wolfgang Maaß	145
Axiomatic support for event perception	
Jeffrey Mark Siskind	153
Towards a perceptually-based semantics of action verbs	
Anil S. Chakravarthy	161
Association of words and images	
Ian Radford	165
Conveying spatial information using vision and natural language	
Alicia Abella & John R. Kender	169
Distinctive feature analysis in intelligent Chinese OCR	
Cheng-ming Guo, Jun-ping Gong, Ying Xia, Xing-gong Chang & Hong-yan Jing ...	173
Visual information in word recognition: word shape or letter identities?	
Stuart J. Soltysiak	177
A study based on several types of co-operation between modalities and binding through synchrony	
Jean-Claude Martin	181
Exploiting lexical information for visual processing	
William B. Dolan	185
Psycholinguistic study of the interaction between language and vision	
M. Spivey-Knowlton, J. Sedivy, K. Eberhard & M. Tanenhaus	189
Links between language and vision	
Peter Forte	193

Preface

There has been a recent move towards considering the integration of perception sources in Artificial Intelligence (AI) (see Dennett, 1991 and Mc Kevitt, 1994). This workshop will focus on research involved in the integration of Natural Language Processing (NLP) and Vision Processing (VP).

Although there has been much progress in developing theories, models and systems in the areas of NLP and VP there has been little progress on integrating these two subareas of Artificial Intelligence (AI). It is not clear why there has not already been much activity in integrating these two areas. Is it because of the long-time reductionist trend in science up until the recent emphasis on chaos theory, nonlinear systems, and emergent behaviour? Or, is it because the people who have tended to work on NLP tend to be in other Departments, or of a different ilk, to those who have worked on VP?

We believe it is high time to bring together NLP and VP. Already we have advertised a call for papers for a special volume of the Journal of AI Review to focus on their integration and we have had a tremendous response. There will be three special issues focussing on theory, applications and intelligent multimedia for the integration of NLP and VP.

The workshop is of particular interest at this time because research in NLP and VP has advanced to the stage that they can each benefit from integrated approaches. Also, such integration is important as people in NLP and VP can gain insight from each others' work.

References

- Dennett, Daniel (1991) *Consciousness explained*. Harmondsworth: Penguin.
Mc Kevitt, Paul (1994) (Guest Editor) Integration of Natural Language and Vision Processing. In *Special Volume (Vol. 8) of AI Review Journal*, Dordrecht: Kluwer (forthcoming).

Workshop topics

The workshop will focus on the following themes:

- Multimedia retrieval
- Multimedia presentation
- Multimedia document processing
- Spatial relations
- Speech, gesture and gaze
- Multimedia interfaces
- Theory
- Reference

Reviewing

All papers were reviewed by leading researchers in the field including the Workshop Committee and authors. The following additional reviewers are to be thanked:

Gregory Anderson, Colin Beardon, Deborah Dahl, Andrew Fano, Lindsey Ford, Peter Hall, Kazuo Hiraki, Stuart Jackson, Henner Kollnig, Takio KURITA, Wolfgang Nagel, John Neary, Bernd Neumann, Itsuki Noda, Ryuichi Oka, Karl Schäfer, Roger Schank, Amanda Sharkey, Tony Simons, Oliviero Stock

I would like to point out that this workshop was, on the whole, organised electronically. The call for papers, programme checks, liaisons with Workshops Chair, AAAI, and the Workshop Committee, virtually all submissions of papers, reviewing, acceptances, and receipt of camera ready copies were conducted electronically. *UNIX uudecode, uncompress, gzip* as well as *postscript* proved useful here. Such is the world today! :-)

Hope

Finally, I would like to thank Donald Perlis, the AAAI-94 Workshops Chair, the Workshop Committee, the authors and Annette Eldredge, Mike Hamilton and Sally McLaughlin at AAAI for cooperating wonderfully on organisation. I hope that our visions for language will come true in the years to come! :)

Slan Leat,
Paul Mc Kevitt
Department of Computer Science
University of Sheffield
England, EU
May 1994.

Programme

Tuesday, August 2nd, 1994

INTRODUCTION I:

- 8.45 'Introduction'
Paul Mc Kevitt

MULTIMEDIA RETRIEVAL:

(Chair: Neil C. Rowe)

- 9.00 'Domain-independent rules relating captions and pictures'
Neil C. Rowe
Computer Science, U.S. Naval Postgraduate School, Monterey CA, USA
- 9.30 'An image retrieval system that accepts natural language'
Hiromasa NAKATANI and Yukihiro ITOH
Department of Information and Knowledge Engineering,
Shizuoka University, Hamamatsu, Japan

10.00 **Break**

MULTIMEDIA DOCUMENT PROCESSING:

(Chair: Rohini Srihari)

- 10.30 'Integrating text and graphical input to a knowledge base'
Raman Rajagopalan
Dept. of Computer Sciences, University of Texas at Austin, USA
- 11.00 'Photo understanding using visual constraints generated from accompanying text'
Rohini Srihari
Center of Excellence for Document Analysis and Recognition (CEDAR),
SUNY Buffalo, NY, USA

11.30 **Discussion**

SPEECH, GESTURE AND GAZE:

(Chair: Jordi Robert-Ribes)

- 12.00 'Audiovisual recognition of speech units: a tentative functional model compatible with psychological data'
Jordi Robert-Ribes, Michel Piquemal, Jean-Luc Schwartz & Pierre Escudier
Institut de la Communication Parlee (ICP)
Grenoble, France, EU

12.30 **Discussion**

12.45 **LUNCH**

SITE DESCRIPTION (VIDEO):

(Chair: Hilary Buxton)

- 2.00 ‘The spoken image system: on the visual interpretation of verbal scene descriptions’
Sean O Nuallain, Benoit Farley & Arnold G. Smith
Dublin City University, Dublin, Ireland, EU &
NRC, Ottawa, Canada

THEORY:

(Chair: Sean O Nuallain)

- 2.20 ‘Behavioural descriptions from image sequences’
Hilary Buxton and Richard Howarth
School of Cognitive and Computing Sciences, University of Sussex &
Department of Computing Science, QMW, University of London
- 2.50 ‘Visions for language’
Paul Mc Kevitt
Department of Computer Science, University of Sheffield, England, EU

3.15 **Discussion**

3.30 **Break**

- 4.00 ‘Language animation’
A. Narayanan, L. Ford, D. Manuel, D. Tallis, and M. Yazdani
Department of Computer Science, University of Exeter, England, EU

4.30 **Discussion**

MULTIMEDIA PRESENTATION:

(Chair: Gerd Herzog)

- 4.45 ‘Assembly plan generation by integrating pictorial and textual information in an assembly illustration’
Shoujie He, Norihiro Abe and Tadahiro Kitahashi
Dept of Information Systems and Computer Science,
National Univ. of Singapore, Singapore,
Faculty of Computer Science and Systems Engineering,
Kyushu Institute of Technology, Iizuka-shi, Japan &
The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan
- 5.15 ‘Multimedia presentation of interpreted visual data’
Elisabeth Andre, Gerd Herzog & Thomas Rist
DFKI & Universität des Saarlandes, Saarbrücken, Germany, EU

5.45 **Discussion**

6.00 **Oiche Mhaith**

Wednesday, August 3rd, 1994

INTRODUCTION:

8.45 'Introduction'
Paul Mc Kevitt

MULTIMEDIA INTERFACES:

(Chair: Yuri A. TIJERINO)

9.00 'Talking pictures: an empirical study into the usefulness of natural language output in a graphical interface'

Carla Huls, Edwin Bos & Alice Dijkstra
NICI, Nijmegen University, Nijmegen, The Netherlands &
Unit of Experimental and Theoretical Psychology, Leiden University,
The Netherlands

9.30 'From verbal and gestural input to 3-D visual feedback'

Yuri A. TIJERINO, Tsutomu MIYASATO & Fumio KISHINO
ATR Communication Systems Research Laboratories, Kyoto, Japan

10.00 **Break**

10.30 'An integration of natural language and vision processing towards an agent-based future TV system'

Yeun-Bae Kim, Masahiro Shibata & Masaki Hayashi
NHK (Japan Broadcasting Corporation)
Science & Technical Research Laboratories, Tokyo, Japan

11.00 **Discussion**

REFERENCE:

(Chair: Lawrence D. Roberts)

11.30 'An AI module for reference based on perception'

John Moulton, Hartwick College, Oneonta, N.Y. USA &
Lawrence D. Roberts, SUNY, Binghamton, N.Y. USA

12.00 'Instruction use by a vision-based mobile robot'

Tomohiro Shibata, M. Inaba, & H. Inoue
Department of Mechano Informatics, The University of Tokyo, Japan

12.30 **Discussion**

12.45 **LUNCH**

SPATIAL RELATIONS I:

(Chair: Stephan Kerpedjiev)

2.00 ‘A cognitive approach to an interlingua representation of spatial descriptions’
Irina Reyero-Sans & Jun-ichi Tsujii

Centre for Computational Linguistics, UMIST, Manchester, England, EU

2.30 ‘Describing spatial relations in weather reports through prepositions’
Stephan Kerpedjiev,

NOAA/ERL/Forecast Systems Laboratory, Boulder, Colorado, USA

3.00 **Discussion**

3.30 **Break**

SPATIAL RELATIONS II:

(Chair: Jeffrey Mark Siskind)

4.00 ‘Prepositional semantics in the WIP system’

Patrick Olivier & Jun-ichi Tsujii

Centre for Intelligent Systems

University of Wales at Aberystwyth, Penglais, Wales, EU &

Centre for Computational Linguistics, UMIST, Manchester, England, EU

4.30 ‘Spatial layout identification and incremental descriptions’

Klaus-Peter Gapp & Wolfgang Maaß

Cognitive Science Program, Saarbrücken, Germany, EU

5.00 ‘Axiomatic support for event perception’

Jeffrey Mark Siskind

Department of Computer Science, University of Toronto, Canada

5.30 **Discussion**

6.00 **Oíche Mhath**