Bianca Falcidieno
Michela Spagnuolo
Yannis Avrithis
Ioannis Kompatsiaris
Paul Buitelaar (Eds.)

LNCS 4816

Semantic Multimedia

Second International Conference on Semantic and Digital Media Technologies, SAMT 2007
Genoa, Italy, December 2007, Proceedings

Springer

Bianca Falcidieno Michela Spagnuolo
Yannis Avrithis Ioannis Kompatsiaris
Paul Buitelaar (Eds.)

Semantic Multimedia

Second International Conference on Semantic and Digital Media Technologies, SAMT 2007
Genoa, Italy, December 5-7, 2007
Proceedings

Springer

Volume Editors

Bianca Falcidieno
Michela Spagnuolo
Consiglio Nazionale delle Ricerche
Istituto per la Matematica Applicata e Tecnologie Informatiche
Via De Marini 6, 16149 Genova, Italy
E-mail: {bianca.falcidieno,michela.spagnuolo}@ge.imati.cnr.it

Yannis Avrithis
National Technical University of Athens
School of Electrical and Computer Engineering
Image, Video and Multimedia Systems Laboratory
9 Iroon Polytechniou Str., 157 80 Athens, Greece
E-mail: iavr@image.ntua.gr

Ioannis Kompatsiaris
Informatics and Telematics Institute
1st Km Thermi-Panorama Road, Thessaloniki, 57001 Greece
E-mail: ikom@iti.gr

Paul Buitelaar
DFKI GmbH
Language Technology Lab
Campus D3 2, Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany
E-mail: paulb@dfki.de

Library of Congress Control Number: 2007940049

CR Subject Classification (1998): H.5.1, H.4, H.5, I.2.10, I.4, I.7, C.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-540-77033-X Springer Berlin Heidelberg New York
ISBN-13	978-3-540-77033-6 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12197521 06/3180 5 4 3 2 1 0

Preface

The conference on Semantics And digital Media Technologies (SAMT) covers a wide scope of subjects that contribute, from different perspectives, to narrow the large disparity between the low-level descriptors of multimedia content and the richness and subjectivity of semantics in user queries and human interpretations of content: The Semantic Gap. The second international conference SAMT 2007 took place in Italy, in the beautiful town of Genova, and continued to gather interest and high-quality papers from across Europe and beyond, bringing together forums, projects, institutions and individuals investigating the integration of knowledge, semantics and low-level multimedia processing, including new emerging media and application areas.

In response to the call for papers, 55 papers were submitted. After a thorough review process, only 16 contributions were accepted as full papers with oral presentation, in addition, 20 contributions were selected as short papers. We would like to thank all the members of the Technical Program Committee, the authors of submitted papers, and the additional reviewers for their efforts in setting the quality of this volume. The conference program also includes two invited keynote talks from Steffen Staab and Remco Veltkamp and we are very grateful to them for their insightful presentations. This volume includes a special section with three awarded short papers from the K-Space PhD Workshop that took place in October 2007 in Berlin.

We acknowledge and would like to thank the entire Organizing Committee for the excellent contribution to the coordination of the various events: the programme of SAMT 2007 included three workshops, one tutorial, three special sessions, project and demo session, the SAMT 2007 industry day, and two invited talks from European Commission representatives, Albert Gauthier of the INFSO.E2 Knowledge and Content Technologies, and Luis Rodríguez-Roselló of the INFSO.D2 Networked Media Systems.

The SAMT 2007 conference was organized by IMATI-CNR with the sponsorship of the European FP6 Networks of Excellence AIM@SHAPE and K-Space and of the Department of Information and Communication Technology of the CNR of Italy. Moreover, SAMT 2007 ran in cooperation with the European Commission and the SALERO project. Finally, we would like to thank Marinella Pescaglia and Sandra Burlando for their invaluable administrative support and the whole staff of the Shape Modelling Group of IMATI-CNR that contributed to making the conference happen.

December 2007

Bianca Falcidieno
Michela Spagnuolo
Yannis Avrithis
Ioannis Kompatsiaris
Paul Buitelaar

Organization

SAMT 2007 was organized by the department of Genova of the Institute of Applied Mathematics and Information Technology of the National Research Council of Italy (IMATI-CNR).

Conference Committees

General Chairs	Bianca Falcidieno *IMATI-CNR, Genova, Italy* Yannis Avrithis *National Technical Univ. of Athens, Greece*
Program Chairs	Michela Spagnuolo *IMATI-CNR, Genova, Italy* Ioannis Kompatsiaris *Informatics and Telematics Institute, Greece* Paul Buitelaar *DFKI, Germany*
Special Sessions	Werner Haas *Joanneum Research, Austria*
Tutorials	Noel O'Connor *Dublin City University, Ireland*
Workshops	Catherine Houstis *Informatics and Telematics Institute, Greece*
Industry Day	Franca Giannini and Chiara Catalano *IMATI-CNR, Genova, Italy*
Posters and Demos	Riccardo Albertoni and Michela Mortara *IMATI-CNR, Genova, Italy*

Program Committee

Bruno Bachimont
Wolf-Tilo Balke
Jenny Benois-Pineau
Jesús Bescós
Nozha Boujemaa
Patrick Bouthemy
Pablo Castells
Andrea Cavallaro
Stavros Christodoulakis
Fabio Ciravegna
Thierry Declerck
Anastasios Delopoulos
Edward Delp
Martin Dzbor
Touradj Ebrahimi
Jerome Euzenat
Borko Furht
Moncef Gabbouj
Christophe García
William I. Grosky
Werner Haas
Christian Halaschek-Wiener

Siegfried Handschuh
Alan Hanjalic
Lynda Hardman
Andreas Hotho
Jane Hunter
Antoine Isaac
Ebroul Izquierdo
Franciska de Jong
Joemon Jose
Aggelos K. Katsaggelos
Hyoung Joong Kim
Joachim Koehler
Stefanos Kollias
Paul Lewis
Petros Maragos
Ferran Marques
José Martínez
Adrian Matellanes
Bernard Merialdo
Frank Nack
Jacco van Ossenbruggen
Jeff Pan
Thrasyvoulos Pappas
Ioannis Pitas
Dietrich Paulus
Eric Pauwels
Dennis Quan
Keith van Rijsbergen
Andrew Salway
Mark Sandler
Simone Santini
Guus Schreiber
Timothy Shih
Sergej Sizov
Alan Smeaton
John R. Smith
Giorgos Stamou
Fred S. Stentiford
Michael Strintzis
Rudi Studer
Vojtech Svatek
Murat Tekalp
Raphael Troncy
Paulo Villegas
Gerhard Widmer
Li-Qun Xu

Additional Reviewers

Samer Abdallah
Riccardo Albertoni
Alexia Briassouli
Stamatia Dasiopoulou
Alberto Machì
Joao Magalhaes
Vassilis Mezaris
Francesco Robbiano
Massimo Romanelli

Sponsoring Institutions

Institute of Applied Mathematics and Information Technology of the CNR, Italy
Department of Information and Communication Technology of the CNR, Italy
FP6 Network of Excellence AIM@SHAPE (http://www.aimatshape.net/)
FP6 Network of Excellence K-Space (http://www.k-space.eu/)

Lecture Notes in Computer Science

Sublibrary 3: Information Systems and Application, incl. Internet/Web and HCI

For information about Vols. 1–4469
please contact your bookseller or Springer

Vol. 4871: M. Cavazza, S. Donikian (Eds.), Virtual Storytelling. XIII, 219 pages. 2007.

Vol. 4857: J.M. Ware, G.E. Taylor (Eds.), Web and Wireless Geographical Information Systems. XI, 293 pages. 2007.

Vol. 4853: F. Fonseca, M.A. Rodríguez, S. Levashkin (Eds.), GeoSpatial Semantics. X, 289 pages. 2007.

Vol. 4836: H. Ichikawa, W.-D. Cho, I. Satoh, H.Y. Youn (Eds.), Ubiquitous Computing Systems. XIII, 307 pages. 2007.

Vol. 4832: M. Weske, M.-S. Hacid, C. Godart (Eds.), Web Information Systems Engineering – WISE 2007 Workshops. XV, 518 pages. 2007.

Vol. 4831: B. Benatallah, F. Casati, D. Georgakopoulos, C. Bartolini, W. Sadiq, C. Godart (Eds.), Web Information Systems Engineering – WISE 2007. XVI, 675 pages. 2007.

Vol. 4825: K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, P. Cudré-Mauroux (Eds.), The Semantic Web. XXVII, 973 pages. 2007.

Vol. 4816: B. Falcidieno, M. Spagnuolo, Y. Avrithis, I. Kompatsiaris, P. Buitelaar (Eds.), Semantic Multimedia. XII, 306 pages. 2007.

Vol. 4813: I. Oakley, S. Brewster (Eds.), Haptic and Audio Interaction Design. XIV, 145 pages. 2007.

Vol. 4806: R. Meersman, Z. Tari, P. Herrero (Eds.), On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops, Part II. XXXIV, 611 pages. 2007.

Vol. 4805: R. Meersman, Z. Tari, P. Herrero (Eds.), On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops, Part I. XXXIV, 757 pages. 2007.

Vol. 4804: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS, Part II. XXIX, 683 pages. 2007.

Vol. 4803: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS, Part I. XXIX, 1173 pages. 2007.

Vol. 4802: J.-L. Hainaut, E.A. Rundensteiner, M. Kirchberg, M. Bertolotto, M. Brochhausen, Y.-P.P. Chen, S.S.-S. Cherfi, M. Doerr, H. Han, S. Hartmann, J. Parsons, G. Poels, C. Rolland, J. Trujillo, E. Yu, E. Zimányie (Eds.), Advances in Conceptual Modeling – Foundations and Applications. XIX, 420 pages. 2007.

Vol. 4801: C. Parent, K.-D. Schewe, V.C. Storey, B. Thalheim (Eds.), Conceptual Modeling - ER 2007. XVI, 616 pages. 2007.

Vol. 4797: M. Arenas, M.I. Schwartzbach (Eds.), Database Programming Languages. VIII, 261 pages. 2007.

Vol. 4796: M. Lew, N. Sebe, T.S. Huang, E.M. Bakker (Eds.), Human–Computer Interaction. X, 157 pages. 2007.

Vol. 4794: B. Schiele, A.K. Dey, H. Gellersen, B. de Ruyter, M. Tscheligi, R. Wichert, E. Aarts, A. Buchmann (Eds.), Ambient Intelligence. XV, 375 pages. 2007.

Vol. 4777: S. Bhalla (Ed.), Databases in Networked Information Systems. X, 329 pages. 2007.

Vol. 4761: R. Obermaisser, Y. Nah, P. Puschner, F.J. Rammig (Eds.), Software Technologies for Embedded and Ubiquitous Systems. XIV, 563 pages. 2007.

Vol. 4747: S. Džeroski, J. Struyf (Eds.), Knowledge Discovery in Inductive Databases. X, 301 pages. 2007.

Vol. 4740: L. Ma, M. Rauterberg, R. Nakatsu (Eds.), Entertainment Computing – ICEC 2007. XXX, 480 pages. 2007.

Vol. 4730: C. Peters, P. Clough, F.C. Gey, J. Karlgren, B. Magnini, D.W. Oard, M. de Rijke, M. Stempfhuber (Eds.), Evaluation of Multilingual and Multi-modal Information Retrieval. XXIV, 998 pages. 2007.

Vol. 4723: M. R. Berthold, J. Shawe-Taylor, N. Lavrač (Eds.), Advances in Intelligent Data Analysis VII. XIV, 380 pages. 2007.

Vol. 4721: W. Jonker, M. Petković (Eds.), Secure Data Management. X, 213 pages. 2007.

Vol. 4718: J. Hightower, B. Schiele, T. Strang (Eds.), Location- and Context-Awareness. X, 297 pages. 2007.

Vol. 4717: J. Krumm, G.D. Abowd, A. Seneviratne, T. Strang (Eds.), UbiComp 2007: Ubiquitous Computing. XIX, 520 pages. 2007.

Vol. 4715: J.M. Haake, S.F. Ochoa, A. Cechich (Eds.), Groupware: Design, Implementation, and Use. XIII, 355 pages. 2007.

Vol. 4714: G. Alonso, P. Dadam, M. Rosemann (Eds.), Business Process Management. XIII, 418 pages. 2007.

Vol. 4704: D. Barbosa, A. Bonifati, Z. Bellahsène, E. Hunt, R. Unland (Eds.), Database and XML Technologies. X, 141 pages. 2007.

Vol. 4690: Y. Ioannidis, B. Novikov, B. Rachev (Eds.), Advances in Databases and Information Systems. XIII, 377 pages. 2007.

Vol. 4675: L. Kovács, N. Fuhr, C. Meghini (Eds.), Research and Advanced Technology for Digital Libraries. XVII, 585 pages. 2007.

Vol. 4674: Y. Luo (Ed.), Cooperative Design, Visualization, and Engineering. XIII, 431 pages. 2007.

Vol. 4663: C. Baranauskas, P. Palanque, J. Abascal, S.D.J. Barbosa (Eds.), Human-Computer Interaction – INTERACT 2007, Part II. XXXIII, 735 pages. 2007.

Vol. 4662: C. Baranauskas, P. Palanque, J. Abascal, S.D.J. Barbosa (Eds.), Human-Computer Interaction – INTERACT 2007, Part I. XXXIII, 637 pages. 2007.

Vol. 4658: T. Enokido, L. Barolli, M. Takizawa (Eds.), Network-Based Information Systems. XIII, 544 pages. 2007.

Vol. 4656: M.A. Wimmer, J. Scholl, Å. Grönlund (Eds.), Electronic Government. XIV, 450 pages. 2007.

Vol. 4655: G. Psaila, R. Wagner (Eds.), E-Commerce and Web Technologies. VII, 229 pages. 2007.

Vol. 4654: I.-Y. Song, J. Eder, T.M. Nguyen (Eds.), Data Warehousing and Knowledge Discovery. XVI, 482 pages. 2007.

Vol. 4653: R. Wagner, N. Revell, G. Pernul (Eds.), Database and Expert Systems Applications. XXII, 907 pages. 2007.

Vol. 4636: G. Antoniou, U. Aßmann, C. Baroglio, S. Decker, N. Henze, P.-L. Patranjan, R. Tolksdorf (Eds.), Reasoning Web. IX, 345 pages. 2007.

Vol. 4611: J. Indulska, J. Ma, L.T. Yang, T. Ungerer, J. Cao (Eds.), Ubiquitous Intelligence and Computing. XXIII, 1257 pages. 2007.

Vol. 4607: L. Baresi, P. Fraternali, G.-J. Houben (Eds.), Web Engineering. XVI, 576 pages. 2007.

Vol. 4606: A. Pras, M. van Sinderen (Eds.), Dependable and Adaptable Networks and Services. XIV, 149 pages. 2007.

Vol. 4605: D. Papadias, D. Zhang, G. Kollios (Eds.), Advances in Spatial and Temporal Databases. X, 479 pages. 2007.

Vol. 4602: S. Barker, G.-J. Ahn (Eds.), Data and Applications Security XXI. X, 291 pages. 2007.

Vol. 4601: S. Spaccapietra, P. Atzeni, F. Fages, M.-S. Hacid, M. Kifer, J. Mylopoulos, B. Pernici, P. Shvaiko, J. Trujillo, I. Zaihrayeu (Eds.), Journal on Data Semantics IX. XV, 197 pages. 2007.

Vol. 4592: Z. Kedad, N. Lammari, E. Métais, F. Meziane, Y. Rezgui (Eds.), Natural Language Processing and Information Systems. XIV, 442 pages. 2007.

Vol. 4587: R. Cooper, J. Kennedy (Eds.), Data Management. XIII, 259 pages. 2007.

Vol. 4577: N. Sebe, Y. Liu, Y.-t. Zhuang, T.S. Huang (Eds.), Multimedia Content Analysis and Mining. XIII, 513 pages. 2007.

Vol. 4568: T. Ishida, S. R. Fussell, P. T. J. M. Vossen (Eds.), Intercultural Collaboration. XIII, 395 pages. 2007.

Vol. 4566: M.J. Dainoff (Ed.), Ergonomics and Health Aspects of Work with Computers. XVIII, 390 pages. 2007.

Vol. 4564: D. Schuler (Ed.), Online Communities and Social Computing. XVII, 520 pages. 2007.

Vol. 4563: R. Shumaker (Ed.), Virtual Reality. XXII, 762 pages. 2007.

Vol. 4561: V.G. Duffy (Ed.), Digital Human Modeling. XXIII, 1068 pages. 2007.

Vol. 4560: N. Aykin (Ed.), Usability and Internationalization, Part II. XVIII, 576 pages. 2007.

Vol. 4559: N. Aykin (Ed.), Usability and Internationalization, Part I. XVIII, 661 pages. 2007.

Vol. 4558: M.J. Smith, G. Salvendy (Eds.), Human Interface and the Management of Information, Part II. XXIII, 1162 pages. 2007.

Vol. 4557: M.J. Smith, G. Salvendy (Eds.), Human Interface and the Management of Information, Part I. XXII, 1030 pages. 2007.

Vol. 4541: T. Okadome, T. Yamazaki, M. Makhtari (Eds.), Pervasive Computing for Quality of Life Enhancement. IX, 248 pages. 2007.

Vol. 4537: K.C.-C. Chang, W. Wang, L. Chen, C.A. Ellis, C.-H. Hsu, A.C. Tsoi, H. Wang (Eds.), Advances in Web and Network Technologies, and Information Management. XXIII, 707 pages. 2007.

Vol. 4531: J. Indulska, K. Raymond (Eds.), Distributed Applications and Interoperable Systems. XI, 337 pages. 2007.

Vol. 4526: M. Malek, M. Reitenspieß, A. van Moorsel (Eds.), Service Availability. X, 155 pages. 2007.

Vol. 4524: M. Marchiori, J.Z. Pan, C.d.S. Marie (Eds.), Web Reasoning and Rule Systems. XI, 382 pages. 2007.

Vol. 4519: E. Franconi, M. Kifer, W. May (Eds.), The Semantic Web: Research and Applications. XVIII, 830 pages. 2007.

Vol. 4518: N. Fuhr, M. Lalmas, A. Trotman (Eds.), Comparative Evaluation of XML Information Retrieval Systems. XII, 554 pages. 2007.

Vol. 4508: M.-Y. Kao, X.-Y. Li (Eds.), Algorithmic Aspects in Information and Management. VIII, 428 pages. 2007.

Vol. 4506: D. Zeng, I. Gotham, K. Komatsu, C. Lynch, M. Thurmond, D. Madigan, B. Lober, J. Kvach, H. Chen (Eds.), Intelligence and Security Informatics: Biosurveillance. XI, 234 pages. 2007.

Vol. 4505: G. Dong, X. Lin, W. Wang, Y. Yang, J.X. Yu (Eds.), Advances in Data and Web Management. XXII, 896 pages. 2007.

Vol. 4504: J. Huang, R. Kowalczyk, Z. Maamar, D. Martin, I. Müller, S. Stoutenburg, K.P. Sycara (Eds.), Service-Oriented Computing: Agents, Semantics, and Engineering. X, 175 pages. 2007.

Vol. 4500: N.A. Streitz, A.D. Kameas, I. Mavrommati (Eds.), The Disappearing Computer. XVIII, 304 pages. 2007.

Vol. 4495: J. Krogstie, A. Opdahl, G. Sindre (Eds.), Advanced Information Systems Engineering. XVI, 606 pages. 2007.

Vol. 4480: A. LaMarca, M. Langheinrich, K.N. Truong (Eds.), Pervasive Computing. XIII, 369 pages. 2007.

Vol. 4473: D. Draheim, G. Weber (Eds.), Trends in Enterprise Application Architecture. X, 355 pages. 2007.

Vol. 4471: P. Cesar, K. Chorianopoulos, J.F. Jensen (Eds.), Interactive TV: A Shared Experience. XIII, 236 pages. 2007.

Table of Contents

Classification and Annotation of Multidimensional Content

Content Adaptation

MX: the IEEE Standard for Interactive Music

Semantic Multimedia Annotation II

Short Papers

Improving the Accuracy of Global Feature Fusion Based Image Categorisation*

Ville Viitaniemi and Jorma Laaksonen

Laboratory of Computer and Information Science, Helsinki University of Technology,
P.O.Box 5400, FIN-02015 TKK, Finland
{ville.viitaniemi,jorma.laaksonen}@tkk.fi

Abstract. In this paper we consider the task of categorising images of the Corel collection into semantic classes. In our earlier work, we demonstrated that state-of-the-art accuracy of supervised categorising of these images could be improved significantly by fusion of a large number of global image features. In this work, we preserve the general framework, but improve the components of the system: we modify the set of image features to include interest point histogram features, perform elementary feature classification with support vector machines (SVM) instead of self-organising map (SOM) based classifiers, and fuse the classification results with either an additive, multiplicative or SVM-based technique. As the main result of this paper, we are able to achieve a significant improvement of image categorisation accuracy by applying these generic state-of-the-art image content analysis techniques.

1 Introduction

In this paper we consider categorisation of images into semantic classes. Image categorisation is a task closely related to the more general problem of image content understanding. The capabilities of image content analysis techniques can be demonstrated by applying them to the image categorisation task. In our earlier work [17] we have demonstrated image annotation and categorisation performance that compares favourable to methods presented in literature by using a system architecture that adaptively fuses a large set of global image features. One of the benchmark task concerns categorisation of images of the Corel collection, approached earlier by Andrews et al.[1], Chen and Zwang [3] and Qi and Han [16] with SVM-based multiple-instance learning methods.

In this work, we consider the same image categorisation task, taking our earlier PicSOM system as a baseline. We retain the overall feature fusion based system architecture since feature fusion has nowadays proven to be an effective approach to solving large-scale problems, such as analysis of the content

* Supported by the Academy of Finland in the projects *Neural methods in information retrieval based on automatic content analysis and relevance feedback* and *Finnish Centre of Excellence in Adaptive Informatics Research.* Special thanks to Xiaojun Qi and Yutao Han for helping with the experimental setup.

B. Falcidieno et al. (Eds.): SAMT 2007, LNCS 4816, pp. 1–14, 2007.

of large video collections. However, we refine the constituent components inside the system. We modify the set of visual features that are fused by including interest point SIFT feature [13] histograms. Such features have recently become popular and demonstrated a good performance in a variety of image content analysis tasks [20]. Based on the individual features in the feature set, the image categories are detected with support vector machine (SVM) classifiers in the improved system, instead of using self-organising map (SOM) based PicSOM classifiers. For fusing the detector outputs, we experiment with additive, multiplicative and SVM-based fusion mechanisms. With the improved system, we perform experiments that replicate the experimental setup of [16]. In addition to demonstrating a significant improvement of the overall image categorisation accuracy, we are able to experimentally compare the performance of different feature sets and fusion mechanisms.

The rest of the paper is organised as follows. Section 2 introduces the image categorisation problem and details the Corel categories benchmark task. In Sections 3 and 4 we describe the baseline system and the improvements, respectively. In Section 5 the performed experiments are described and their results reported. In Section 6 we draw conclusions from the results.

2 Image Categorisation

In the image categorisation task each image is assigned to exactly one of a list of possible categories on the basis of visual content of the image. We regard the task as a supervised learning problem, where a number of training images together with their ground truth categorisation is used to learn the connection between the images' visual properties and the category labels. The quality of the learned model is tested by predicting the categories of a previously unseen set of test images solely based on their visual contents and comparing the predictions with a manually specified ground truth. Another type of image content analysis problem—image annotation—is closely related to image categorisation. In that problem, however, an image may be assigned to any number of categories simultaneously. Sometimes the term annotation is used in literature also for categorisation problems.

As an experimental benchmark categorisation task we use the Corel categories task that was first defined by Chen and Wang [3] for 20 image categories. The image data consists of images from 20 Corel stock photograph CDs. Each of the CDs contains 100 images from a distinct topic and forms a target category for the task. A label is chosen to describe each of the categories. Table 1 shows the labels of the 20 categories. Some of the categories are very concrete, some are more abstract. The task was later extended to 60 categories by Qi and Han [16] who introduced 40 more Corel CDs. We use all the 60 categories in our experiments. The task also defines an ordering of the image categories and a corresponding succession of incremental subsets of categories: the set with M categories includes the first M categories of the ordering.

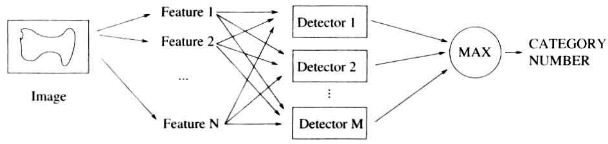

Fig. 1. The general architecture of our image categorisation system

As we want to compare our categorisation performance with that of the earlier efforts, we use the same metric for measuring the categorisation performance: the average accuracy. This is simply the fraction of correctly categorised test images over their total number. For measuring annotation quality, more sophisticated measures [18] are usually used, but the equal number of images in each category warrants the use of this simple average accuracy measure in this categorisation task.

Table 1. Labels of the first 20 categories of the Corel categories task ([3,16])

African people and villages	Beach	Historical buildings	Buses
Dinosaurs	Elephants	Flowers	Horses
Mountains and glaciers	Food	Dogs	Lizards
Fashion	Sunsets	Cars	Waterfalls
Antiques	Battle ships	Skiing	Desert

3 Baseline System

As a baseline system, we use our PicSOM image annotation system that was demonstrated to perform well in the Corel categories task in our earlier work [17]. Figure 1 shows the general architecture of the system. To categorise a test set image, a number of elementary visual features is extracted from it. The extracted set of feature vectors is fed parallel to several detectors, one for detecting each of the categories. The output of each detector is converted to the estimated posterior probability of the category. The predicted category of the test image is selected to be the category whose detector outputs the highest probability.

The rest of this section details the components of the baseline system: the used visual features (Sec. 3.1) and the detector modules (Sec. 3.2).

3.1 Visual Features

As a basis for the classification, a set of global visual features is extracted from the images. The extracted ten elementary features, listed in Table 2, are encoded as feature vectors with dimensionalities given in the rightmost column of the table. The first four rows correspond to features that more or less closely resemble the ColorLayout, DominantColor, EdgeHistogram and ScalableColor features of the MPEG-7 standard [8]. The column "Tiling" of the table shows that some of the features are calculated truly globally, such as the global colour histogram

Table 2. The elementary visual features extracted from the images

Feature	Tiling	Dim.
DCT coefficients of average colour in rectangular grid	global	12
CIE L*a*b* colour of two dominant colour clusters	global	6
MPEG-7 EdgeHistogram descriptor	4×4	80
Haar transform of quantised HSV colour histogram	global	256
Average CIE L*a*b* colour	5	15
Three central moments of CIE L*a*b* colour distribution	5	45
Co-occurence matrix of four Sobel edge directions	5	80
Magnitude of the 16×16 FFT of Sobel edge image	global	128
Histogram of four Sobel edge directions	5	20
Histogram of relative brightness of neighboring pixels	5	40

feature of the fourth row, others, such as the edge histogram feature of the third row, encode some spatial information by using a fixed image grid. The features calculated for five tiles employ a tiling mask where the image area is divided into four tiles by the two diagonals of the image, on top of which a circular center tile is overlaid.

3.2 Classifiers

The category detection is achieved by training a separate detector for each category in our PicSOM image content analysis framework. The framework classifies images by adaptively fusing information given by several different elementary low-level image features. The framework is readily described elsewhere (e.g. [12]).

In the PicSOM image content analysis framework, the input to the image classifier consists of three sets of images: training images annotated with a keyword (positive examples), training images not annotated with the keyword (negative examples), and test images. The task of the classifier is to associate a score to each test image so that the score reflects simultaneously the image's similarity to the positive examples and dissimilarity from the negative ones.

To obtain a detector for an image category, a set of visual features is first extracted from the example images. The PicSOM framework is then used to automatically generate representations of the features that are adaptive to both 1) the context of the totality of images in the image collection, and 2) the context of the present task, expressed in terms of sets of positive and negative example images. To this end, the components of the elementary feature vectors are divided into several overlapping subsets, feature spaces. As the number of all possible feature combinations is overwhelming, we must content ourselves with a rather arbitrarily chosen subset. In the baseline system, we consider all the 10 elementary feature vectors individually, almost all pairs of them, and some heuristically chosen triplets and quadruplets, resulting in 98 feature spaces in total.

Each of the feature spaces is quantised using a TS-SOM [11], a tree-structured variant of the self-organising map [10]. For the experiments reported here, we use TS-SOMs with three stacked levels, the bottom levels measuring 64×64 map units. For each two-dimensional quantised TS-SOM representation, positive and negative impulses are placed at the best-matching-unit (BMU) projections of the positive and negative example images on the TS-SOM surface. The impulses are then normalised and low-pass filtered. This associates a partial, feature-dependent classifier score to each map unit of the corresponding feature SOM. A total classifier score for a test set image is obtained by projecting the image to each of the SOM-quantised feature spaces and summing the partial classifier scores from the corresponding SOM's BMUs. In this procedure, the low-pass filtered summation of positive and negative examples effectively emphasises features that perform well in separating the positive and negative example images of that particular classification task.

To facilitate the comparison between outputs of detectors for different image categories, each output is converted to a probability estimate. We have observed that a simple logistic sigmoid model

$$p_i(w|s_w(i)) = \frac{1}{1 + e^{-\theta_1^w s_w(i) - \theta_0^w}} \tag{1}$$

suits this purpose well. Here $s_w(i)$ is the classifier score and and $p_i(w|s_w(i))$ the probability of image i belonging to category w. The model parameters $\theta_{\{0,1\}}^w$ are estimated separately for each category w by regarding all the training images as independent samples and maximising the likelihood

$$L_w = \prod_i \frac{I_w(i) + (1 - I_w(i))\, e^{-\theta_1^w s_w(i) - \theta_0^w}}{1 + e^{-\theta_1^w s_w(i) - \theta_0^w}} \tag{2}$$

with respect to $\theta_{\{0,1\}}^w$. Here $I_w(i)$ is the binary indicator variable for the training image i to belong to the category w. The product is taken over all the training images. The maximisation is performed numerically using the classical Newton-Raphson method.

4 Improved System

The improvements of the image categorisation system are made to the components of baseline system, the overall system architecture of Figure 1 is not affected. Both the set of visual features (Sec. 4.1) and the category detectors (Sec. 4.2) are improved.

4.1 Improved Visual Features

In the improved system, the set of visual features is extended with histograms of interest point features. The interest points are detected using a Harris-Laplace detector [14]. The SIFT feature [13], based on local gradient orientation, is calculated for each interest point. Histograms of interest point features have proven to be efficient in image content analysis and are gaining popularity in various image analysis tasks [5,15].