

Emergence of ontological relations from visual data with Self-Organizing Maps

Jorma Laaksonen and Ville Viitaniemi

*Helsinki University of Technology
Adaptive Informatics Research Centre
FI-02015 TKK, FINLAND

jorma.laaksonen@tkk.fi ville.viitaniemi@tkk.fi

Abstract

In this paper we examine how Self-Organizing Maps (SOMs) can be used in detecting and describing emergent ontological relations between semantic objects and object classes in a visual database. The ontological relations we have studied include *co-existence*, *taxonomies of visual and semantic similarity* and *spatial relationships*. The used database contains 2618 images, each of which belongs to one or more of ten predefined semantic classes.

1 Introduction

Data-driven learning of semantic properties and ontological relations has for long been a goal in AI research. If machine learning were practicable, rule-based generation and presentation of *a priori* top-down knowledge could be complemented by information obtainable from bottom-up data processing and abstraction.

Techniques of traditional statistical pattern recognition and soft-computing, including neural networks and fuzzy logic, have been used for data-driven learning, but they have two severe limitations. First, they need training data, and supervised classifiers additionally *a priori* knowledge of correct classifications of the data. Second, statistical classifiers and other non-structural recognizers mostly operate on very low abstraction level data and there will inevitably exist the *semantic gap* between their output and the required input for syntactic and other structural processing stages.

In this paper, we will study a specific image collection of 2618 images. The images are accompanied with keyword-type annotations which specify a subset of ten available keywords for each image. In addition, there exists information on the location of the objects in the images in the form of bounding boxes. The image database has been prepared for evaluation of object recognition and detection techniques, but it lends itself also for the purpose of studying the structure and ontological properties of the contained images.

The experiments to be described have been per-

formed by using the PicSOM image analysis, classification and content-based information retrieval (CBIR) framework (Laaksonen et al., 2001, 2002). In the PicSOM system, we have recently added functionality for processing and analyzing also images which have hierarchical segmentations (Viitaniemi and Laaksonen, 2006b). As its name suggests, PicSOM is based on using Self-Organizing Maps (SOMs) (Kohonen, 2001) in organizing and indexing objects of a database. We believe that SOMs – due to their un-supervised training – can be a beneficial tool for facilitating data-driven emergence of ontological relations.

The rest of this paper is organized as follows. First, in Section 2, a short literature review is presented. Then, we give overviews of the Self-Organizing Map, the PicSOM CBIR system and the low-level visual features used in it in Section 3. Next, in Section 4, we describe the image data set used in the study. Section 5 presents and demonstrates the techniques we have developed for using SOMs in discovering and analyzing ontological relations in an image database. Conclusions are drawn and future directions discussed in Section 6.

2 Related work

Ontologies can be used to boost the performance of information retrieval systems (e.g. Bodner and Song (1996)). This applies to both large external general purpose ontologies, such as the WordNet (Fellbaum, 1998) with over one hundred thousand words, as well as ontologies mined automatically from the data it-

self. In the visual domain, ontologies have been used e.g. in aiding automatic image annotation (Jin et al., 2005; Srikanth et al., 2005), keyword sense disambiguation (Benitez and Chang, 2002), video retrieval (Hoogs et al., 2003; Hollink and Worring, 2005), and video annotation (Bertini et al., 2005).

For the ontologies to be usable in connection with visual data, they need to include visual components. In the literature, this has been achieved mostly by two means. First is linking pre-existing ontologies, such as the WordNet (Hoogs et al., 2003; Hollink and Worring, 2005; Jaimes and Smith, 2003) or domain ontologies, with visual concepts. The visual concepts are chosen so that they can be detected automatically from image or video documents. “Sky”, “water” and “man-made” are examples of such concepts. In (Bertini et al., 2005) the concepts obtained by clustering visual features of manually chosen representative video clips of soccer game highlights are linked directly to the soccer game domain ontology nodes as specializations. Also in other approaches the linking involves manual labor. An advantage of this method is the ability to use pre-existing ontologies that are possibly very large. However, if the ontology is large and there are only a few visual concepts, the coupling between specific ontology concepts and visual properties may remain weak.

Another approach for building visual ontologies is to hierarchically cluster a visual training data set (e.g. Khan and Wang (2002)). Such clustering methods (e.g. Frakes and Baeza-Yates (1992)) have been used earlier for ontology discovery in textual and numerical data (Bodner and Song, 1996; Clerkin et al., 2001). An alternative to clustering the training data itself is to calculate a similarity measure between the ontology concepts on basis of the training data, and cluster directly the concepts based on the measure. In the clustering approach, labels of the ontology nodes are pre-specified by labels of the training data, and the relations between the nodes are learned. The number of nodes varies from a few to few dozens. (Khan and Wang, 2002) performs the clustering on basis of weighted cosine distance measure between the region composition vectors of images. In (Koskela and Smeaton, 2006) the distances between concepts describing videos are measured by the Jeffrey’s divergence between their distributions in quantized feature spaces. In this work, we will consider measuring the distance between concepts by the performance of a visual classifier attempting to separate the concepts.

Describing spatial relationships in images has attracted research attention for decades (e.g. Winston (1975)). Qualitative descriptions based on logical for-

malisms have often been used. More recently, fuzzy descriptions of spatial relationships (Miyajima and Ralescu, 1994; Bloch, 2005) have become popular in the field. Spatial data mining from large databases (e.g. Koperski et al. (1996)) is a relatively new development. (Smith and Bridges, 2002; Lan et al., 2005) exemplify data mining based on fuzzy spatial relationships.

In this paper, we will use Self-Organizing Maps for the discovery and visualization of ontological relations—including spatial relationships—using visual data. The SOM has been used for ontology discovery and visualization earlier by Elliman and Pulido (2001, 2002) with textual and categorical data.

3 SOMs and PicSOM

3.1 Self-Organizing Map

The Self-Organizing Map (SOM) is an unsupervised, self-organizing neural algorithm widely used to visualize and interpret large high-dimensional data sets. The SOM defines an elastic net of points that are fitted to the distribution of the data in the input space. It can thus be used to visualize multi-dimensional data, usually on a two-dimensional grid.

The SOM consists of a two-dimensional lattice of neurons or map units. A model vector $\mathbf{m}_i \in \mathbb{R}^d$ is associated with each map unit i . The map attempts to represent all the available observations $\mathbf{x} \in \mathbb{R}^d$ with optimal accuracy by using the map units as a restricted set of models. During the training phase, the models become ordered on the grid so that similar models are close to and dissimilar models far from each other.

When training a SOM, the fitting of the model vectors is carried out by a sequential regression process, where $t = 0, 1, 2, \dots, t_{max} - 1$ is the step index: For each input sample $\mathbf{x}(t)$, first the index $c(\mathbf{x})$ of the best-matching unit (BMU) or the “winner” model $\mathbf{m}_{c(\mathbf{x})}(t)$ is identified by the condition

$$\forall i : \|\mathbf{x}(t) - \mathbf{m}_{c(\mathbf{x})}(t)\| \leq \|\mathbf{x}(t) - \mathbf{m}_i(t)\| . \quad (1)$$

The distance metric used here is usually the Euclidean one. After finding the BMU, the vectors of map units constituting a *neighborhood* centered around the node $c(\mathbf{x})$ are updated as

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h(t; c(\mathbf{x}), i)(\mathbf{x}(t) - \mathbf{m}_i(t)) . \quad (2)$$

Here $h(t; c(\mathbf{x}), i)$ is the neighborhood function, a decreasing function of the distance between the i th and

$c(\mathbf{x})$ th nodes on the map grid. This regression is iterated over the available samples and the value of $h(t; c(\mathbf{x}), i)$ is let decrease in time to guarantee convergence of the model vectors \mathbf{m}_i . Large values of the neighborhood function $h(t; c(\mathbf{x}), i)$ are used in the beginning of the training for initializing the network, and small values on later iterations are needed for fine-tuning. After the training, any vector in the feature space can be quantized to a two-dimensional index by its BMU on the SOM.

3.2 PicSOM

In the PicSOM¹ image analysis framework, target objects are ranked according to their similarity with a given set of positive example objects, simultaneously combined with the dissimilarity with a set of negative example objects. The objects are correlated in terms of a large set of visual features of statistical nature. For this purpose training and test set images are pre-processed in the same manner: the images are first automatically segmented and a large body of statistical features is extracted from both the segments and whole images. Several different segmentations of the same images can be used in parallel. In the current experiments we consider only visual features, but the framework has been used to index also e.g. videos and multimedia messages (Koskela et al., 2005).

After feature extraction, a SOM is trained in an unsupervised manner to quantize each of the formed feature spaces. The quantization forms representations of the feature spaces, where points on the SOM surfaces correspond to images and image segments. Due to the topology preserving property of SOM mapping, the classification in each of the individual feature spaces can be performed by evaluating the distance of representation of an object on the SOM grid to the representations of positive and negative example objects.

The positive and negative examples are marked on the SOM surfaces with small impulse values of corresponding sign. As the SOM maps similar objects in nearby map units, we are motivated to spatially spread these sparse values fields by a low-pass filter, i.e. to convolve them with a smoothing kernel. The size and shape of the convolution kernel is selected in a suitable way in order to produce a smooth value map. In the resulting map, each location is effectively assigned a relevance value according to the number of positive objects mapped to the nearby units.

A single combined similarity measure is formed by summing the contributions of the individual fea-

ture spaces. In the classification task where the target objects are images, the segment-wise similarities are finally combined within an image by summing the contributions of all the segments in the image.

3.3 Low-level visual features

The PicSOM system implements a number of methods for extracting different statistical visual features from images and image segments. These features include a set of MPEG-7 content descriptors (ISO/IEC, 2002; Laaksonen et al., 2002) and additionally some non-standard descriptors for color, shape and texture.

3.3.1 Color

Of the used MPEG-7 descriptors Color Layout, Dominant Color and Scalable Color describe the color content in image segments. In addition to the MPEG-7 color descriptors, both the average color in the CIE L*a*b* color space (CIE, 1976) and three first central moments of the color distribution are used as color features.

3.3.2 Shape

Besides the MPEG-7 Region Shape, the shape features include two non-standard descriptors. The first consists of the set of the Fourier descriptors for the region contour. Fourier descriptors are derived from the following expansion of the region contour:

$$z(s) = \sum_{n=-\infty}^{\infty} z_n e^{\frac{2\pi i n s}{L}}. \quad (3)$$

Here the Cartesian coordinates of the contour are represented by the real and the imaginary parts of the complex function $z(s)$, parameterized by the arc length s . The resulting feature vector includes a fixed number of low-order expansion coefficients z_n . The coefficients are then normalized against affine image transformations. In addition, the high-order coefficients are quadratically emphasized.

The second non-standard shape descriptor is formed from the Zernike moments (Khotanzad and Hong, 1990) of the region shape. The Zernike polynomials are a set of polar polynomials that are orthogonal in the unit disk. The Zernike moments A_{nm} are given by the expansion coefficients when the polar presentation of the region shape is represented in the basis of Zernike polynomials:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y I_{xy} V_{nm}(\rho_{xy}, \theta_{xy}), \quad (4)$$

¹<http://www.cis.hut.fi/picsom>



Figure 1: A sample image with keyword *dog* and its bounding box

where $n - |m|$ is even. Here n is the order of the moment, m the index of repetition, x, y are the rectangular image coordinates, and ρ_{xy}, θ_{xy} the corresponding polar coordinates. I_{xy} is the binary representation of the region shape and V_{nm} is the Zernike polynomial:

$$V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{im\theta} \quad (5)$$

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)! \rho^{n-2s}}{s! (\frac{n+|m|}{2} - s)! (\frac{n-|m|}{2} - s)!}. \quad (6)$$

The feature vector includes coefficients A_{nm} up to a selected order. The feature is normalized against translation and scaling by fitting the region inside the unit disk. Rotation invariance is achieved by taking the absolute values of the coefficients.

3.3.3 Texture

We have used MPEG-7's Edge Histogram descriptor to describe the statistical texture in image segments. For non-standard description of a region's texture the YIQ color space Y-values of the region pixels are compared with the values of their 8-neighbors. The feature vector describes the statistics of the resulting distribution.

4 VOC image database

The Visual Object Classes (VOC) 2006 image database was created and annotated with keywords and bounding box information for the purpose of a challenge evaluation of object recognition and detection techniques². It contains 2618 PNG images whose sizes are typically 500×375 pixels in either portrait or landscape orientation.

²<http://www.pascal-network.org/challenges/VOC/>

Figure 1 displays one example from the collection. It has been given the annotation *dogFrontal* and the associated bounding box. In our experiments, however, we have removed all pose information from the annotations and therefore used only the keyword *dog* for that image.

After discarding the pose information, there were ten keyword classes left in the database. The classes and the numbers of images in each of them are listed in Table 1. One may note that the total number of keywords is well above the number of images, which is an indication of the fact that many of the images have more than one keyword associated.

Table 1: The keyword classes in the VOC2006 database and their counts

<i>bicycle</i>	270
<i>bus</i>	174
<i>car</i>	553
<i>cat</i>	386
<i>cow</i>	206
<i>dog</i>	365
<i>horse</i>	247
<i>motorbike</i>	235
<i>person</i>	666
<i>sheep</i>	251

5 Ontological relations on SOMs

In this section we study three different types of information present in the VOC image data that can successfully be visualized and analyzed on a SOM.

5.1 Co-existence of classes

First we note that the ten keywords defined for the images in the collection form a ten-dimensional binary space and that each image is mapped to one point of that space. The ordering of the classes in such a vectorial presentation is meaningless, but the simultaneous existence of more than one keyword in one image is of special interest here.

For ten keywords, the *binary keyword vector* can have $2^{10} = 1024$ different combinations or values. The distribution of vectors in this ten-dimensional space will thus be difficult to visualize without means for low-dimensional projections. We have used SOM for this purpose as follows.

First, we have added small amount of white Gaussian noise in the binary vectors. This ensures that

there will not exist two exactly equal data vectors even though two images may have the same associated keywords. As a further consequence, those keyword combinations which are most common in the image collection will occupy largest areas on the SOM to be created.

Second, we trained a SOM and studied for each of the ten keyword classes how images that had been given that keyword are mapped on the keyword SOM. Figure 2 shows the SOM areas where four image classes are mapped. It can be seen that the class *person* intersects notably with both *dog* and *horse*, but not at all with *sheep* in this database.

After inspecting the distributions of all the ten keyword classes, one can depict a *co-existence graph* for describing the ontological relations of the classes with respect to the likeliness of simultaneous existence in a photograph. The areas on the SOM surface can be used to guide the placement of the nodes in the graph, as can be seen done in Figure 3. It is obvious that in this collection the *person* class is very central as almost all the other classes exhibit co-existence with it. Furthermore, all the different vehicles form a tightly interconnected subgraph.

5.2 Visual and semantic taxonomy

With our second experiment we show that the visual similarities between the classes can be used in data-driven creation of a taxonomy of object classes. For that purpose, we first calculated pairwise misclassifi-

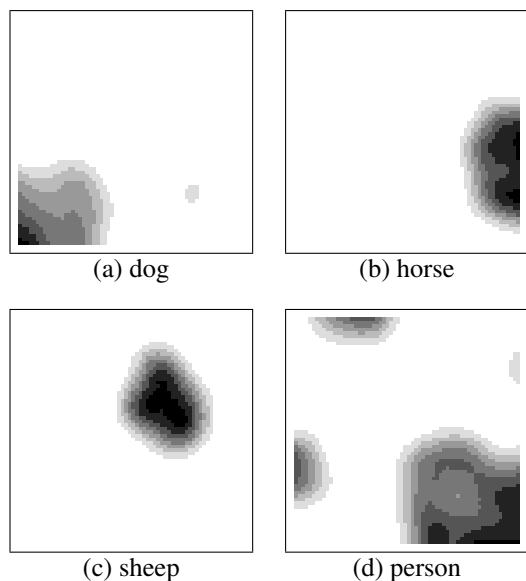


Figure 2: Areas occupied by four different image classes on keyword SOM

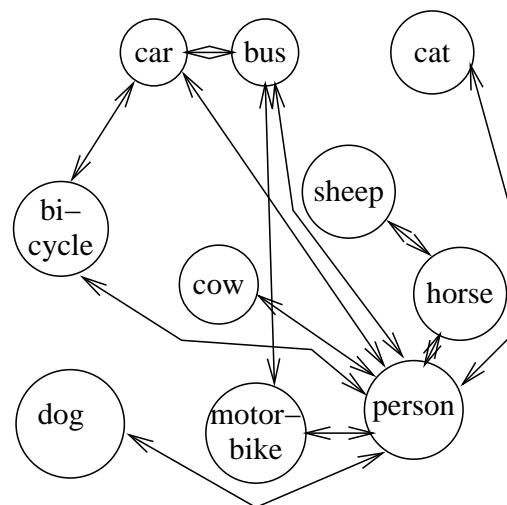


Figure 3: Co-existence graph of the VOC image collection, created to match corresponding class areas on the keyword SOM

cation rates between all class pairs. That is, we assessed the similarity of two classes by studying the difficulty of separating them. For achieving optimal classification we used a k -nearest neighbor classifier with optimization of the k value and sequential forward selection (SFS) of the used visual features.

When the two classes which were the most difficult to separate were found, we created a virtual joint class of those two and re-calculated all pairwise misclassification rates. This process produced the unbalanced binary tree of the ten classes shown in Figure 4. We argue that even though this *visual class taxonomy* has been created purely from low-level visual feature data, it to some extent coincides with the corresponding *semantic class taxonomy* one might come to by mental contemplation.

In the VOC data, there exist class pairs (*motorbike, bicycle*), (*cat,dog*) and (*cow,sheep*) which tend to be misclassified and are therefore combined in the early stages of the taxonomy tree formation process of Figure 4. Some of the later merges are, however, not as straightforward to interpret. (One may actually state that such an iterative merging process which only relies on local information is bound to be chaotic.)

The same principle can be used also to create a one-dimensional, non-hierarchical ordering of the classes where mutually similar ones reside near each other. In order to achieve this, we first changed for each class the cardinal numbers that showed the misclassification rate between that class and each other class to ordinal numbers which indicated the order of the other classes in decreasing misclassification rate.

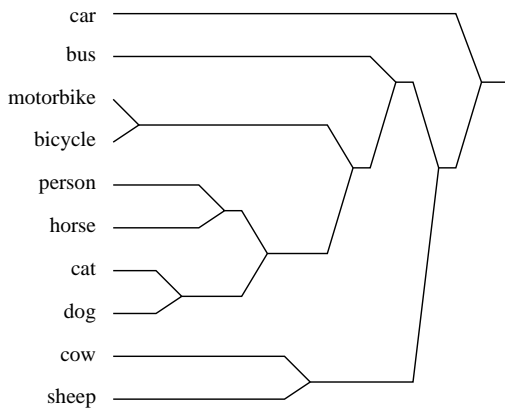


Figure 4: Taxonomy tree of visual similarity of the classes

Then we re-ordered the rows and columns of the resulting matrix so that the smallest numbers were forced to reside along the diagonal and the first subdiagonals whereas the largest numbers were moved to furthest away from the diagonal. The resulting ordering and *ordinal similarity matrix* is shown in Figure 5. It can be seen that there again exist obvious pairs (*motorbike, bicycle*), (*cat, dog*) and (*cow, sheep*), but they are now aligned on a linear continuum between the two opposites, *bus* and *sheep*.

The creation of the two similarity representations for the ten object class were based on classifications which used potentially all the available low-level features. Therefore, there does not exist a single SOM map, where the ordering of the object classes would match that of the class taxonomy tree or the ordered similarity matrix. Some features, however, may come quite close to that ordering. For an illustration, see Figure 6 where the VOC image dataset has been ordered on a 16×16 -sized SOM created from the MPEG-7 EdgeHistogram feature. One can see how the antidiagonal order-

<i>bus</i>	0	3	2	4	1	5	6	7	9	8
<i>car</i>	3	0	2	4	1	5	6	7	9	8
<i>motorbike</i>	4	5	0	1	2	3	6	7	9	8
<i>bicycle</i>	7	6	1	0	2	4	5	3	8	9
<i>person</i>	7	6	1	3	0	2	5	4	9	8
<i>horse</i>	9	8	3	6	1	0	3	2	5	7
<i>cat</i>	8	9	5	4	3	2	0	1	5	7
<i>dog</i>	8	9	7	6	3	2	1	0	5	4
<i>cow</i>	8	8	7	6	5	2	4	3	0	1
<i>sheep</i>	9	8	5	7	4	3	6	2	1	0

Figure 5: Ordered similarity matrix of the object classes

ing of the classes *car–bus–motorbike–bicycle–horse–dog/cat–sheep–cow* quite well matches those of Figures 4 and 5.



Figure 6: Ordering of object classes on MPEG-7 EdgeHistogram SOM

5.3 Spatial relationships

With our final experiment we want to show how the SOM can be used in presenting spatial relationships between objects in an image. We will here restrict our attention to images where there exists one *horse* and one *person*. The database contains a total of 84 such images. From the associated annotations we can extract the bounding boxes of those two objects and combine them in eight-dimensional feature vectors.

Figure 7 displays a 16×16 -sized SOM created from the eight-dimensional spatial location data. One may coarsely identify four regions on the map. First, on the left of the map there are images where the person is standing to the right of the horse. On the bottom left the situation is otherwise similar, but the spatial ordering of the objects is reversed. On the right of the map there are portrait images of riding and in the middle parts similar landscape images.

We argue that the *spatial relationship SOM* has been able to discover relations between the horses and persons which we would associate to “riding” and “standing”. This is an example of data-driven emergence by a SOM which orders objects topographically on the basis of their low-level statistical features. Only the linguistic concepts “riding” and “standing” have in this case been added by human inspection of the outcome of the self-organization.

In this example we have used only data from images where *horse* and *person* have been the two objects whose spatial relationships have been of interest.

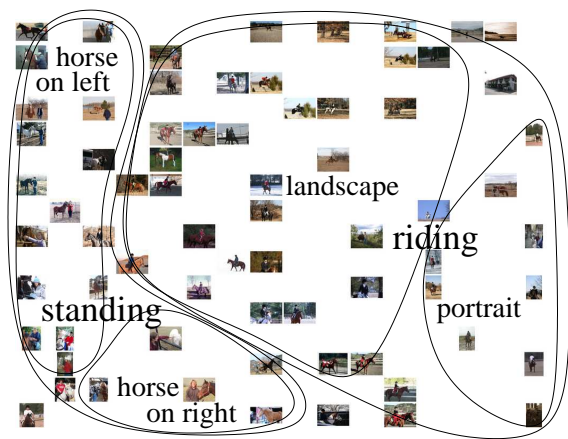


Figure 7: Spatial relationship SOM created from images of a horse and a person

The approach is, however, independent of the classes, so a single SOM could be created from data of all object pairs. Such a map will inevitably describe the relationships on a more general level, for example, the relationship of *person* “riding” *horse* would be replaced by a less context-specific expression *person* “above” *horse*.

6 Conclusions and discussion

In this paper we have shown that the Self-Organizing Map can successfully be applied in diverse data analysis tasks in which it is able to extract and display emergent ontological relations between semantic objects and object classes. The studied ontological relations included (i) simultaneous existence of objects from two or more object classes in one image, (ii) taxonomy of visual (and arguably to some extent also semantic) similarity, and (iii) spatial relationships between different object types in one image.

The experiments were performed within the PicSOM framework for image analysis and content-based information retrieval. The PicSOM system inherently uses SOMs in indexing and scoring images by their similarity to other images known to be relevant in a particular query instance. Consequently, the co-existence and spatial relationship SOMs, which were for the first time introduced in this paper, can easily be integrated in that framework.

In another related study (Viitaniemi and Laaksonen, 2006a) we have already questioned whether knowledge of a semantic taxonomy of object classes would be beneficial for detection of objects. This is an important issue because semantically related object types often appear in visually similar *contexts*.

Therefore hierarchical detection of such objects may to some level rely also on the context information, but on the final stages the visual appearance of the sole objects itself must be the only discriminating factor as the context will be equally probable for all the candidate object classes.

In the future we will turn our attention to more quantitative analysis and assessment of the benefits obtainable from the techniques presented in this paper. In the forthcoming experiments we will study how the emergent ontological relations can be first extracted from available annotated training data and then be used to facilitate efficient retrieval from unannotated test data.

Acknowledgments

This work has been 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*.

References

- Ana B. Benitez and Shih-Fu Chang. Semantic knowledge construction from annotated image collections. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME 2002)*, volume 2, pages 205–208, Lausanne, Switzerland, August 2002.
- Marco Bertini, Alberto Del Bimbo, and Carlo Torriani. Automatic video annotation using ontologies extended with visual information. In *ACM Multimedia*, pages 395–398, 2005.
- Isabelle Bloch. Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23:89–110, February 2005.
- R. Bodner and F. Song. Knowledge-based approaches to query expansion in information retrieval. In G. McGalla, editor, *Advances in Artificial Intelligence*, pages 146–158. Springer, New York, 1996.
- CIE. Supplement No. 2 to CIE publication No. 15 Colorimetry (E-1.3.1) 1971: Official recommendations on uniform color spaces, color-difference equations, and metric color terms, 1976.
- P. Clerkin, P. Cunningham, and C. Hayes. Ontology discovery for the semantic web using hierarchical clustering. In *Online proceedings of Semantic Web Mining Workshop at*

- ECML/PKDD-2001*, Freiburg, Germany, 2001. <http://semwebmine2001.aifb.uni-karlsruhe.de/>.
- D. Elliman and J. R. G. Pulido. Visualizing ontology component through Self-Organizing Maps. In *Proceedings of the International Conference on Information Visualisation (IV2002)*, pages 434–438, London, July 2002.
- Dave Elliman and J. Rafael Pulido. Automatic derivation of on-line document ontologies, 2001.
- Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, 1998.
- William B. Frakes and Ricardo Baeza-Yates, editors. *Information Retrieval, Data Structures and Algorithms*. Prentice Hall, New Jersey, USA, 1992.
- Laura Hollink and Marcel Worring. Building a visual ontology for video retrieval. In *ACM Multimedia*, pages 479–482, 2005.
- Anthony Hoogs, Jens Rittscher, Gees Stein, and John Schmiederer. Video content annotation using visual analysis and a large semantic knowledgebase. In *CVPR (2)*, pages 327–334, 2003.
- ISO/IEC. Information technology - Multimedia content description interface - Part 3: Visual, 2002. 15938-3:2002(E).
- Alejandro Jaimes and John R. Smith. Semi-automatic, data-driven construction of multimedia ontologies. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, Baltimore, July 2003.
- Yohan Jin, Latifur Khan, Lei Wang, and Mamoun Awad. Image annotations by combining multiple evidence & WordNet. In *ACM Multimedia*, pages 706–715, 2005.
- Latifur Khan and Lei Wang. Automatic ontology derivation using clustering for image classification. In *Multimedia Information Systems*, pages 56–65, Tempe, AZ, USA, October 2002.
- A. Khotanzad and Y. H. Hong. Invariant image recognition by Zernike moments. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 12(5): 489–497, 1990.
- Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer-Verlag, Berlin, third edition, 2001.
- Krzysztof Koperski, Junas Adhikary, and Jiawei Han. Spatial data mining: progress and challenges. In *Proceedings of the Workshop on Research Issues on Data Mining and Knowledge Discovery*, Montreal, Canada, 1996.
- Markus Koskela and Alan F. Smeaton. Clustering-based analysis of semantic concept models for video shots. In *International Conference on Multimedia & Expo (ICME 2006)*, Toronto, Canada, July 2006. To appear.
- Markus Koskela, Jorma Laaksonen, Mats Sjöberg, and Hannes Muurinen. PicSOM experiments in TRECVID 2005. In *Proceedings of the TRECVID 2005 Workshop*, pages 262–270, Gaithersburg, MD, USA, November 2005.
- Jorma Laaksonen, Markus Koskela, Sami Laakso, and Erkki Oja. Self-organizing maps as a relevance feedback technique in content-based image retrieval. *Pattern Analysis & Applications*, 4(2+3): 140–152, June 2001.
- Jorma Laaksonen, Markus Koskela, and Erkki Oja. PicSOM—Self-organizing image retrieval with MPEG-7 content descriptions. *IEEE Transactions on Neural Networks, Special Issue on Intelligent Multimedia Processing*, 13(4):841–853, July 2002.
- Rongqin Lan, Wenzhong Shi, Xiaomei Yang, and Guangyuan Lin. Mining fuzzy spatial configuration rules: methods and applications. In *IPRS Workshop on Service and Application of Spatial Data Infrastructure*, pages 319–324, 2005.
- Koji Miyajima and Anca Ralescu. Spatial organization in 2D images. In *Proceedings of the Third IEEE Confence on Fuzzy Systems*, volume 1, pages 100–105, 1994.
- George Brannon Smith and Susan M. Bridges. Fuzzy spatial data mining. In *Proceedings of the 2002 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS2002)*, pages 184–189, 2002.
- M. Srikanth, J. Varner, M. Bowden, and D. Moldovan. Exploiting ontologies for automatic image annotation. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 552–558, 2005.
- Ville Viitaniemi and Jorma Laaksonen. Techniques for still image scene classification and object detection. In *Proceedings of 16th International Conference on Artificial Neural Networks (ICANN 2006)*, September 2006a. To appear.
- Ville Viitaniemi and Jorma Laaksonen. *Focusing Keywords to Automatically Extracted Image Segments Using Self-Organising Maps*, volume 210 of *Studies in Fuzziness and Soft Computing*. Springer Verlag, 2006b. ISBN 3-540-38232-1. To appear.
- P. H. Winston. Learning structural descriptions from examples. In P. H. Winston, editor, *The Psychology of Computer Vision*. McGraw-Hill, 1975.