

PROBABILISTIC COMBINATION OF SPATIAL CONTEXT WITH VISUAL AND CO-OCCURRENCE INFORMATION FOR SEMANTIC IMAGE ANALYSIS

Georgios Th. Papadopoulos^{1,2}, Vasileios Mezaris², Ioannis Kompatsiaris² and Michael G. Strintzis^{1,2}

¹Electrical & Computer Eng. Dep., Aristotle Univ. of Thessaloniki, Greece

²Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece

ABSTRACT

In this paper, a probabilistic approach to combining spatial context with visual and co-occurrence information for semantic image analysis is presented. Overall, the examined image is segmented and subsequently an initial classification of the resulting image regions to semantic concepts is performed based solely on visual information. Then, a Genetic Algorithm (GA) is introduced for deciding on the optimal semantic image interpretation, realizing image analysis as a global optimization problem. The fundamental novelty of this work is that the GA incorporates in its evolutionary procedure a set of Bayesian Networks (BNs), which probabilistically learn the impact of the available spatial, visual and co-occurrence information on the final outcome for every possible pair of semantic concepts. Experimental results on two publicly available datasets demonstrate the efficiency of the proposed approach.

Index Terms— Spatial context, bayesian network, genetic algorithm, semantic image analysis

1. INTRODUCTION

The widespread use of multimedia capturing devices with high storage capabilities and the continuously growing network access availability have resulted in an enormous increase of the total amount of image content that is exchanged among individuals or is made available over the internet. This has raised the need for techniques facilitating common image manipulation tasks like indexing, search and retrieval. Among the solutions that have received particular attention are semantic image analysis approaches [1], targeting the detection and recognition of the real-world objects that are depicted in an image. Despite the good recognition performance that has been reported for domain specific applications, this task has proven to be rather challenging in less constrained environments. The latter is mainly due to the ambiguity that is inherent in the visual medium. For overcoming this limitation, the use of context has been proposed.

Spatial context in particular is of increased importance in semantic image analysis. The latter models the spatial configuration of the objects and facilitates in discriminating between objects that exhibit similar visual characteristics. In [2], Yuan et al. employ simple grid-structure graphical models to characterize the spatial dependencies between the objects depicted in the image. Additionally, a Conditional Random Field (CRF)-based approach is presented in [3] that incorporates both co-occurrence as well as spatial contextual information. Wang et al. [4] propose a probabilistic approach for integrating feature distribution and spatial context models for image

region annotation. Moreover, individual spatial context techniques are comparatively evaluated with several different combinations of classifiers and low-level features in [5]. Although a series of spatial context techniques have already been presented, little work has been carried out towards the direction of examining under which circumstances spatial context should be used, i.e. identifying for which objects spatial context can facilitate their discrimination and subsequently adjusting its impact on their detection against the visual and the objects' co-occurrence information. Additionally, most of the existing approaches consider spatial context to have equal importance for all objects.

In this paper, a probabilistic approach to combining spatial context with visual and co-occurrence information for semantic image analysis is presented. Initially, the examined image is segmented and for every pair of regions a corresponding set of fuzzy directional spatial relations are estimated. Subsequently, an initial association of the computed image regions with a set of predefined high-level semantic concepts is performed using only visual features. Then, a Genetic Algorithm (GA) is introduced for estimating a globally optimal region-concept assignment. The fundamental novelty of this work is that the GA makes use of a set of Bayesian Networks (BNs) for probabilistically acquiring and utilizing complex contextual information. The BNs are provided with an appropriate network structure, which enables them to identify concept pairs for which spatial context can reinforce their discrimination. Consequently, they probabilistically adjust the weight of spatial context against the visual and co-occurrence information during the detection of every possible pair of semantic concepts.

The paper is organized as follows: Section 2 discusses the visual information processing. Section 3 focuses on the proposed approach for combining spatial context with visual and co-occurrence information for semantic image analysis. Experimental results are presented in Section 4 and conclusions are drawn in Section 5.

2. VISUAL INFORMATION PROCESSING

In order to perform the initial region-concept association, the examined image has to be segmented to regions and suitable low-level descriptions have to be extracted for every resulting segment. Under the proposed approach, the segmentation algorithm of [6] is used and the created spatial regions, which are likely to represent meaningful semantic objects, are denoted by s_n , $n \in [1, N]$.

For every image segment s_n , a corresponding region feature vector v_n is computed as follows: A set of keypoints are estimated for every region, using a point-of-interest detector as well as a pre-determined image grid, and a SIFT descriptor vector (with 128 elements) is extracted at each keypoint. Then, following the 'Bag of Words' (BoW) methodology [7] a 300-dimensional feature vector

The work presented in this paper was supported by the European Commission under contracts FP6-045547 VIDi-Video, FP7-214306 JUMAS and FP7-248984 GLOCAL.

v_n is created for region s_n based on its original SIFT descriptor vectors. In parallel to visual feature extraction, a set of fuzzy directional spatial relations are estimated for every ordered pair of image regions (s_n, s_m) , $n \neq m$. The set of directional relations utilized in this work, denoted by $R = \{r_\gamma, \gamma \in [1, \Gamma]\}$, comprises the following relations: Above, Right, Below, Left, Below-Right, Below-Left, Above-Right and Above-Left. Relation r_γ estimated for the region pair (s_n, s_m) is denoted by $r_\gamma(s_n, s_m) \in [0, 1]$. A detailed description of their extraction procedure can be found in [8].

Using only the visual features, an initial region-concept association is performed using Support Vector Machines (SVMs). In particular, an individual SVM is introduced for every defined concept c_k , $k \in [1, K]$, to detect the corresponding instances, and is trained under the 'one-against-all' approach. Each SVM receives as input the region feature vector v_n and estimates for every segment a posterior probability $h_{nk} \equiv P(c_k|v_n)$, which denotes the degree to which concept c_k is assigned to region s_n . This probability is calculated as follows: $h_{nk} = \frac{1}{1+e^{-\eta \cdot z_{nk}}}$, where z_{nk} is the distance of the particular input feature vector v_n from the corresponding SVM's separating hyperplane and η is a slope parameter set experimentally.

3. EXPLOITATION OF CONTEXTUAL INFORMATION

3.1. Genetic Algorithm

GAs have been extensively used in a wide variety of optimization problems, where they have been shown to outperform other traditional methods. Building on the authors' previous work [8], a GA is employed on top of the initial region-concept association results for deciding on the optimal semantic image interpretation by treating image analysis as a global optimization problem.

In this work, the GA employs an initial population of randomly generated chromosomes. Every chromosome T represents a possible solution, i.e. each gene assigns one of the defined concepts c_k to an image region s_n ; this assignment is denoted g_{nk} and therefore $T = \{g_{nk}, n \in [1, N]\}$. After the population initialization, new generations are iteratively produced, where each new generation comes from the current one after the application of evolutionary operators like selection, crossover and mutation, until the optimal solution is reached. The GA makes use of an appropriate fitness function for denoting the plausibility of every possible image interpretation, which has the form:

$$f(T) = \frac{\sum_{n,m} V(g_{nk}, g_{ml})}{N(N-1)}, \quad (1)$$

where $V(g_{nk}, g_{ml}) \in [0, 1]$ indicates the degree to which the g_{nk} , g_{ml} region-concept mappings are consistent with respect to the acquired contextual and other (e.g. visual) information and $N(N-1)$ denotes the number of permutations of the N image regions taken 2 at a time (i.e. the number of ordered region pairs that are present in the image and which contribute to the summation in the numerator).

The output of the GA is a final region-concept association which corresponds to the solution with the highest fitness value. The main issues related to the use of the GA in the presented semantic image analysis framework are: i) the contextual information acquisition procedure, and ii) the definition of function $V(g_{nk}, g_{ml})$ that exploits this contextual and other information for evaluating the consistency of the region-concept mappings. In this work, a probabilistic approach is followed for efficiently combining the spatial context with the visual and co-occurrence information for every possible pair of concepts, as opposed to the method of [8], where only a global weight factor is learned for adjusting the impact of the spatial versus the visual cues on the final outcome.

3.2. Spatial Constraints Acquisition and Evaluation

In order to acquire the appropriate spatial constraints that will facilitate towards the discrimination between concepts that exhibit similar visual characteristics, a statistical learning approach is followed. For that purpose, a set of manually annotated image content, denoted by D_{tr}^1 and for which the fuzzy directional relations have been computed, is assembled. Then, for every ordered concept pair (c_k, c_l) the mean vector $\bar{\mathbf{r}}^{kl}$ and the corresponding covariance matrix $cov(\mathbf{r}^{kl})$, with respect to relations r_γ , are calculated as follows:

$$\begin{aligned} \mathbf{r}_{n,m} &= [r_1(s_n, s_m), r_2(s_n, s_m) \dots r_\Gamma(s_n, s_m)]^T \\ \bar{\mathbf{r}}^{kl} &= [\bar{r}_1^{kl}, \bar{r}_2^{kl} \dots \bar{r}_\Gamma^{kl}]^T = E[\mathbf{r}_{n,m}] \\ cov(\mathbf{r}^{kl}) &= E[(\mathbf{r}_{n,m} - \bar{\mathbf{r}}^{kl})(\mathbf{r}_{n,m} - \bar{\mathbf{r}}^{kl})^T], \end{aligned} \quad (2)$$

where for the calculations the spatial relations $r_\gamma(s_n, s_m)$ which have been computed for all region pairs (s_n, s_m) , $n \neq m$, that are assigned to the concepts (c_k, c_l) , respectively, are taken into account. The set of values $\bar{\mathbf{r}}^{kl}$ and $cov(\mathbf{r}^{kl})$ obtained for concept pair (c_k, c_l) define a spatial constraint, denoted by u^{kl} , which represents the 'allowed' spatial topology of concepts c_k and c_l .

For evaluating the agreement of a given pair of region to concept mappings (g_{nk}, g_{ml}) with spatial constraint u^{kl} , the following mahalanobis distance-based expression is used:

$$Y_{u^{kl}}(g_{nk}, g_{ml}) = \frac{1}{1 + \sqrt{\mathbf{p}_{n,m}^T cov^{-1}(\mathbf{r}^{kl}) \mathbf{p}_{n,m}}}, \quad (3)$$

where $\mathbf{p}_{n,m} = (\mathbf{r}_{n,m} - \bar{\mathbf{r}}^{kl})$. $Y_{u^{kl}}(g_{nk}, g_{ml}) \in [0, 1]$ denotes the degree to which the pair of mappings (g_{nk}, g_{ml}) is consistent with the acquired spatial contextual information. Greater values of $Y_{u^{kl}}(g_{nk}, g_{ml})$ indicate more plausible spatial arrangements.

3.3. Combination of Spatial, Visual and Co-occurrence Information

BNs constitute an efficient methodology for learning complex probabilistic relationships among a set of random variables [9]. Under the proposed approach, BNs are employed for automatically adjusting the impact of the available spatial, visual and concepts' co-occurrence information on the detection of each pair of concepts (c_k, c_l) . Combining this information, a BN estimates the value of $V(g_{nk}, g_{ml})$ (Eq. (1)), which measures how plausible a given pair of region to concept mappings (g_{nk}, g_{ml}) is. To this end, a series of K^2 BNs are constructed, where an individual BN is introduced for every possible ordered pair of concepts (c_k, c_l) to learn the respective correlations. The general structure of each BN is described in the sequel. It must be highlighted that in the presented work discrete space BNs are employed, since they are less prone to under-training occurrences compared to the continuous space ones [9].

The first step in the development of any BN is the definition of the random variables that are of interest for the given application. For the task at hand, the following random variables are defined:

- variables CA_{nk} and CA_{ml} , which correspond to the mappings g_{nk} and g_{ml} , respectively. Variable CA_{nk} denotes the fact of assigning concept c_k to region s_n ; similarly for CA_{ml} .
- variable SC_{nm}^{kl} , which indicates the consistency of the aforementioned mappings with respect to the acquired spatial knowledge (Section 3.2). This variable denotes the value of the spatial constraint verification factor $Y_{u^{kl}}(g_{nk}, g_{ml})$.
- variables VA_{nk} and VA_{ml} , which represent the visual analysis results for concepts c_k and c_l (Section 2), respectively. Variable VA_{nk} denotes the feasibility of the mapping g_{nk} based

on visual cues, i.e. the value of the estimated posterior probability h_{nk} ; similarly for VA_{ml} .

Subsequently, the space of every introduced random variable, i.e. the set of possible values that it can receive, needs to be defined. In particular, for variables CA_{nk} and CA_{ml} the set of values that they can receive is chosen equal to $\{ca_{nk1}, ca_{nk2}\} = \{ca_{ml1}, ca_{ml2}\} = \{True, False\}$, where *True* denotes the assignment of concepts c_k, c_l to regions s_n, s_m , respectively, and *False* the opposite. On the other hand, a discretization step is applied to the values $Y_{u^{kl}}(g_{nk}, g_{ml}), h_{nk}$ and h_{ml} for defining the spaces of variables SC_{nm}^{kl}, VA_{nk} and VA_{ml} , respectively. The aim of the selected discretization procedure is to compute a close to uniform discrete distribution for each of the aforementioned variables, which was experimentally shown to better facilitate the BN inference, compared to discretization with constant step or other common distributions like gaussian and poisson.

The discretization is defined as follows: initially, a set of annotated image content, denoted by D_{tr}^2 , is formed (similarly to the D_{tr}^1 set described in Section 3.2). Then, for every possible ordered region pair (s_n, s_m) in D_{tr}^2 , the posterior probabilities h_{nk}, h_{ml} and the verification factor $Y_{u^{kl}}(g_{nk}, g_{ml})$ are estimated. Subsequently, the aforementioned values are grouped, forming sets $L_1 = \{h_{nk}\} \equiv \{\lambda_{1i}\}, L_2 = \{h_{ml}\} \equiv \{\lambda_{2i}\}$ and $L_3 = \{Y_{u^{kl}}(g_{nk}, g_{ml})\} \equiv \{\lambda_{3i}\}$ (for $1 \leq i \leq I$), where I denotes the total number of ordered region pairs in D_{tr}^2 . Then, the elements of the aforementioned sets are sorted in ascending order, and the resulting sets are denoted by \mathcal{L}_j ($j = 1, 2, 3$). If Q denotes the number of possible discrete values of every corresponding random variable, these are defined according to the following equations:

$$B_j = \begin{cases} b_{j1} & \text{if } \lambda_{ji} \in [0, \mathcal{L}_j(\phi)) \\ b_{jq} & \text{if } \lambda_{ji} \in [\mathcal{L}_j(\phi \cdot (q-1)), \mathcal{L}_j(\phi \cdot q)), \quad q \in [2, Q-1] \\ b_{jQ} & \text{if } \lambda_{ji} \in [\mathcal{L}_j(\phi \cdot (Q-1)), 1] \end{cases} \quad (4)$$

where $\phi = \lfloor \frac{I}{Q} \rfloor$, $\mathcal{L}_j(o)$ denotes the o^{th} element of the ascending sorted set \mathcal{L}_j , and $b_{j1}, b_{j2}, \dots, b_{jQ}$ denote the values of variable B_j ($B_j \in \{VA_{nk}, VA_{ml}, SC_{nm}^{kl}\}$). From the above equations, it can be seen that although the number of possible values for all random variables B_j is equal to Q , the corresponding value ranges with which they are associated are generally different.

The next step in the development of a BN structure is to define a Directed Acyclic Graph (DAG), which represents the causality relations among the introduced random variables. For the problem of concern, the causal DAG \mathbb{G}_{kl} , which is illustrated in Fig 1, is constructed. The direction of the arcs in the proposed BN structure defines explicitly the causal relationships / conditional independence assumptions among the defined variables. In particular, it is considered that: a) variables VA_{nk} and VA_{ml} are conditionally dependent only on variables CA_{nk} and CA_{ml} , respectively (i.e. the semantic concept that is present in an image region fully determines the observed visual features), and b) variable CA_{nk} has a causal influence on CA_{ml} both directly (co-occurrence information) as well as transitively through variable SC_{nm}^{kl} (spatial constraint verification factor).

From the developed causal DAG \mathbb{G}_{kl} and the conditional independence assumptions that it represents, the joint probability distribution of the random variables that are included in \mathbb{G}_{kl} , which is denoted by P_{joint} and satisfies the Markov condition [9] with \mathbb{G}_{kl} , is defined as follows:

$$\begin{aligned} P_{joint}(ca_{nk}, ca_{ml}, va_{nk}, va_{ml}, sc_{nm}^{kl}) &= P_1 \cdot P_2 \\ P_1 &= P(ca_{nk}) \cdot P(ca_{ml}|ca_{nk}, sc_{nm}^{kl}) \cdot P(sc_{nm}^{kl}|ca_{nk}) \\ P_2 &= P(va_{nk}|ca_{nk}) \cdot P(va_{ml}|ca_{ml}), \end{aligned} \quad (5)$$

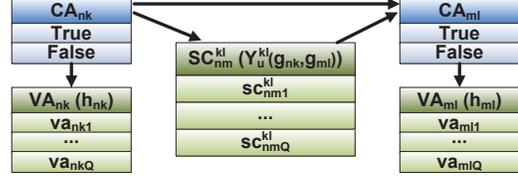


Fig. 1. Developed BN structure \mathbb{G}_{kl} .

where $ca_{nk}, ca_{ml}, va_{nk}, va_{ml}, sc_{nm}^{kl}$ are the values of the variables $CA_{nk}, CA_{ml}, VA_{nk}, VA_{ml}, SC_{nm}^{kl}$, respectively. The pair $(\mathbb{G}_{kl}, P_{joint})$ constitutes the developed BN. From the above equations, it can be seen that the proposed BN probabilistically learns the impact that the spatial, visual and co-occurrence information should have on the calculation of the degree of plausibility for the pair of mappings (g_{nk}, g_{ml}) . More specifically, it is capable of learning the importance of the visual cues on the assignment of concepts c_k and c_l to regions s_n and s_m , respectively, and in particular it adds variable significance to every corresponding analysis value (i.e. values h_{nk} and h_{ml}), by calculating the conditional probabilities $P(va_{nk}|ca_{nk})$ and $P(va_{ml}|ca_{ml})$ in term P_2 , respectively. Similarly, the developed BN also encodes the complex correlations between the mappings g_{nk} and g_{ml} , by adaptively adjusting the degree to which spatial-related cues and concept co-occurrence information are taken into account. The latter is realized by calculating the probabilities $P(ca_{ml}|ca_{nk}, sc_{nm}^{kl})$ and $P(sc_{nm}^{kl}|ca_{nk})$ in term P_1 .

Regarding the training process of the developed BN, the set of all conditional probabilities among the defined conditionally-dependent random variables of \mathbb{G}_{kl} (Eq. (5)), are estimated from the set of annotated image content D_{tr}^2 , which was also used for input variable discretization. At the evaluation stage, the BN receives as input the visual analysis results (i.e. posterior probabilities h_{nk} and h_{ml}) and the corresponding spatial constraint verification factor $Y_{u^{kl}}(g_{nk}, g_{ml})$. These constitute the so called evidence data that a BN requires for performing inference. Then, the BN estimates the following posterior probability (degree of belief), making use of all the pre-computed conditional probabilities and the defined local dependencies among the random variables of \mathbb{G}_{kl} : $P(ca_{nk} = True, ca_{ml} = True|va_{nk}, va_{ml}, sc_{nm}^{kl})$. This probability constitutes a quantitative indication of how plausible the pair of region to concept mappings (g_{nk}, g_{ml}) is, based on spatial, visual and co-occurrence information; the value of $V(g_{nk}, g_{ml})$ in Eq. (1) is set equal to this probability.

4. EXPERIMENTAL RESULTS

In this section, experimental results from the application of the proposed approach to two publicly available datasets, denoted by D_1 and D_2 , are presented. In particular, the SCEF¹ dataset, which is denoted by D_1 (922 images) and was introduced in [5], is used for experimentation. For this, the following 10 concepts are defined: *Building, Foliage, Mountain, Person, Road, Sailing-boat, Sand, Sea, Sky* and *Snow*. The aforementioned image set was divided into three sub-sets, namely D_{tr}^1 (230 images), D_{tr}^2 (230 images) and D_{te} (462 images). The first one, D_{tr}^1 , was used for training the SVM-based classification algorithm and acquiring the spatial constraints. D_{tr}^2 was utilized for training the proposed BNs, while D_{te} was used for evaluation. Regarding the dataset D_2 (591 images), the MSRC² v2 dataset was utilized. For the latter dataset, the following 21 concepts

¹<http://mklab.iti.gr/project/scef>

²<http://research.microsoft.com/en-us/projects/objectclassrecognition/>

are supported: *Building, Grass, Tree, Cow, Sheep, Sky, Aeroplane, Water, Face, Car, Bicycle, Flower, Sign, Bird, Book, Chair, Road, Cat, Dog, Body* and *Boat*. Moreover, the corresponding sets D_{tr}^1 , D_{tr}^2 and D_{te} , including 148, 147 and 296 images, respectively, were also formed.

In Fig. 2, quantitative performance measures from the application of the proposed approach to the utilized datasets are presented in terms of the difference in concept detection accuracy. The latter is calculated by subtracting the detection accuracy accomplished based solely on visual features from the corresponding one obtained after the application of the proposed spatial context exploitation approach. The initial classification results computed based on visual information are depicted in parentheses. It has been considered that for each region s_n , $argmax_k(h_{nk})$ indicates its concept assignment based solely on visual features. Accuracy is defined as the percentage of the image regions that are assigned to the correct semantic concept. It must be noted that the value of variable Q in Eq. (4), which defines the number of possible values for variables VA_{nk} , VA_{ml} and SC_{nm}^{kl} , was set equal to 19 and 24 for the D_1 and D_2 datasets, respectively; it has been observed that values of Q greater than 10, i.e. when the selected discretization was not coarse, led to marginal changes in the overall detection accuracy for both datasets.

From the presented results, it can be seen that the proposed approach achieves an overall performance improvement of 7.94% and 5.21% in the D_1 and D_2 datasets, respectively, compared to the initial classification results. Additionally, the detection rates for most of the supported concepts are significantly increased in both datasets. In particular, it is shown that concepts exhibiting more well-defined spatial configuration are substantially favored, such as concepts Building, Person in D_1 and Tree, Road in D_2 . Concept c_k is considered to have well-defined spatial context if the sum $\sum_l tr(cov(\mathbf{r}^{kl}))$ receives relatively low values (where $tr(\cdot)$ denotes the trace of a matrix), i.e. the spatial relations of concept c_k with all other concepts c_l of the respective dataset do not present significant variations in their values. On the other hand, the detection rate of concepts that present less well-defined spatial context is also increased (for example concepts Snow, Foliage and Sheep, Chair in D_1 and D_2 , respectively). For the latter set of concepts, this performance improvement is mainly due to the incorporation of the concepts' co-occurrence information in the developed BNs. Moreover, it can be seen that significant performance improvement can be obtained for concepts that present low initial classification rate (e.g. concepts Road, Sailing-boat and Aeroplane, Car in D_1 and D_2 , respectively). Significant contribution towards this performance improvement is induced by the probabilistic approach that is followed for adjusting the impact that the visual cues should have on the detection of every supported concept. On the contrary, small decrease in the detection performance may be observed for a few concepts that either: a) present significantly increased initial classification rate (e.g. concept Sky in both datasets), or b) have less well-defined spatial context and the visual / co-occurrence information can not facilitate towards their discrimination (e.g. concepts Water and Face in D_2). These results demonstrate the efficiency of the proposed approach in improving the region classification results that have been computed based solely on visual information, by probabilistically combining spatial context with visual and co-occurrence information.

The performance of the proposed approach was also compared with the spatial context exploitation techniques presented in [5]. In particular, it was shown that the proposed method outperforms the methods of [5] by approximately 5% and 3% in the D_1 and D_2 datasets, respectively, in terms of overall concept detection accuracy.

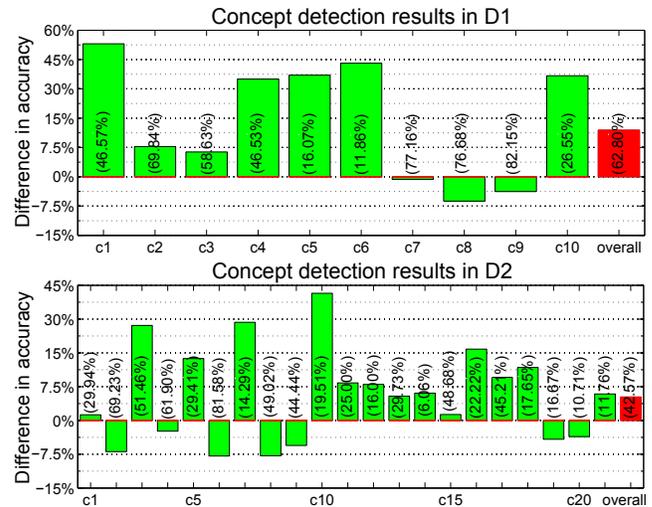


Fig. 2. Concept detection results.

This difference in performance is due to the more elaborate approach followed by the proposed method for probabilistically combining the available spatial, visual and co-occurrence information, contrary to the simpler methodologies that the methods of [5] adopt.

5. CONCLUSIONS

In this paper, a probabilistic approach to semantic image analysis, which combines spatial context with visual and co-occurrence information, was presented and evaluated on two publicly available datasets. Future work includes the investigation of additional information sources (e.g. scene-level information) and their integration in the developed framework.

6. REFERENCES

- [1] A. Smeulders et al., "Content-Based Image Retrieval at the End of the Early Years," *IEEE TPAMI*, pp. 1349–1380, 2000.
- [2] J. Yuan, J. Li, and B. Zhang, "Exploiting spatial context constraints for automatic image region annotation," *Proc. ACM Int. Conf. on Multimedia*, pp. 595–604, 2007.
- [3] C. Galleguillos et al., "Object categorization using co-occurrence, location and appearance," in *IEEE CVPR*, 2008.
- [4] Z. Wang, D.D. Feng, Z. Chi, and T. Xia, "Annotating image regions using spatial context," in *Eighth IEEE Int. Symposium on Multimedia, ISM'06*, 2006, pp. 55–61.
- [5] G. Th. Papadopoulos et al., "Comparative evaluation of spatial context techniques for semantic image analysis," in *Proc. WIAMIS '09*, pp. 161–164, 2009.
- [6] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis, "Still Image Segmentation Tools for Object-Based Multimedia Applications," *Int. J. of Patt. Recog. and Artif. Intell.*, vol. 18, 2004.
- [7] G. Csorika et al., "Visual categorization with bags of keypoints," in *Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, vol. 1, p. 22.
- [8] G. Th. Papadopoulos et al., "Combining Global and Local Information for Knowledge-Assisted Image Analysis and Classification," *EURASIP J. on Adv. in Signal Proc.*, 2007.
- [9] R.E. Neapolitan, *Learning bayesian networks*, Prentice Hall Upper Saddle River, NJ, 2003.