LABELLING IMAGE REGIONS USING SPATIAL PROTOTYPES

Carsten Saathoff, Marcin Grzegorzek, and Steffen Staab

University of Koblenz
{saathoff,marcin,staab}@uni-koblenz.de

ABSTRACT

In this paper we present an approach for introducing spatial context into image region labelling. We combine low-level classification with spatial reasoning based on explicitly represented spatial arrangements of labels. We formalise the problem using *Linear Programming*, and provide an evaluation on a set of 923 images.

1. INTRODUCTION

Exploiting solely low-level features for automatic image region labelling often leads to unsatisfactory results, and recent studies [1] show the importance of contextual and spatial information. In this paper, we propose an approach based on [2] that integrates a wavelet-based low-level classification [3] and spatial reasoning based on *Linear Programming*.

During the training phase we train the classifiers and acquire our background knowledge. In the classification phase, each image is first segmented by an automatic segmentation algorithm. The low-level classification produces for each region s_i and each supported label l_j a probability $\theta_i(l_j)$. Then relative (e.g. *above-of, left-of*) and absolute (e.g. *above-all*) spatial relations are extracted, and are processed by the spatial reasoning together with the probabilities. The output is a final labelling that is optimal with respect to both our spatial background knowledge and the probabilities.

We provide results of a number of experiments showing that our approach provides comparable performance with low numbers of training examples. Due to length constraints, we will not detail the low-level processing at all.

2. SPATIAL REASONING BASED ON CONSTRAINTS

The goal of the spatial reasoning step is to exploit background knowledge about the typical spatial arrangements of objects in images in order to improve the labelling accuracy compared to pure local, low-level feature-based approaches. We will first discuss the acquisition of constraint templates from a set of spatial prototypes, and then describe the formalisation of the problem as a Linear Program.

2.1. Constraint acquisition

Spatial constraint templates constitute the background knowledge in our approach. We acquire these templates from so-called spatial prototypes, which are manually labelled images. We mine the prototypes using support and confidence as selection criteria, and come up with a set of templates representing typical spatial arrangements.

For each label l we have to determine in what spatial relation to other labels it might be found. Therefore, for each spatial relation type t, we consider the relation set $R_{t\downarrow l}$, which contains the relations of type t from images depicting l. We then define $R_{t\downarrow l}^{l,l'}$ to be the set of relations between segments s, s' depicting l and l', respectively, and finally $R_{t\downarrow l}^{*,l'}$ to denote all relations between an arbitrary region and a region depicting l'. The confidence of a label arrangement is then defined as $\gamma_t(l,l') = \frac{|R_{t\downarrow l}^{l,l'}|}{|R_{t\downarrow l}^{*,l'}|}$, and the support as $\sigma_t(l,l') = \frac{|R_t^{l,l'}|}{|R_{t\downarrow l}|}$.

Finally, we define the template $\mathcal{T}_t$ for the spatial relation type t as $\mathcal{T}_t(l,l') = 1$ iff $\sigma_t(l,l') > th_\sigma$ and $\gamma_t(l,l') > th_\gamma$, and $\mathcal{T}_t(l,l') = 0$ otherwise. For absolute spatial relations we define support, confidence, and the template accordingly.

2.2. Spatial reasoning with linear programming

We will show in the following how to formalize image labelling with spatial constraints as a linear program. We consider Binary Integer Programs, which have the form

$$\begin{aligned} \text{maximize} \quad & Z && = \mathbf{c^T x} \\ \text{subject to} \quad & \mathbf{Ax} && = \mathbf{b} \\ & \mathbf{x} && \in \{0,1\} \end{aligned} \tag{1}$$

Goal of the solving process is to find a set of assignments to the integer variables in $\mathbf{x}$ with a maximum evaluation score Z that satisfy all the constraints.

In order to represent the image labelling problem as a linear program, we create a set of linear constraints from each spatial relation in the image, and determine the objective coefficients based on the hypotheses sets and the constraint templates. Let $O_i \subseteq R$ be the set of outgoing relations for region $s_i \in S$, i.e. $O_i = \{r \in R | \exists s \in S, s \neq s_i : r = (s_i, s)\}$, and $E_i \subseteq R$ the set of incoming spatial relations, i.e. $E_i = \{r \in R | \exists s \in S, s \neq s_i : r = (s, s_i)\}$. Then, for each possible pair of label assignments to the regions, we create a variable c_{itj}^{ko}, representing the possible assignment of l_k to s_i and l_o to s_j with respect to the relation r with type $t \in T$. Each c_{itj}^{ko} is an integer variable and $c_{itj}^{ko} = 1$ represents the assignments $s_i = l_k$ and $s_j = l_o$, while $c_{itj}^{ko} = 0$ means that these assignments are not made. Since every such variable represents exactly one assignment of labels to the involved regions, and only one label might be assigned to a region in the final solution, we have to add this restriction as linear constraints. The constraints are formalised as $\forall r \in R : r = (s_i, s_j) \in R \rightarrow \sum_{l_k \in L} \sum_{l_o \in L} c_{itj}^{ko} = 1$. These constraints assure that there is only one pair of labels assigned to a pair of regions per spatial relation, but it still there could be two variables c_{itj}^{ko} and $c_{it'j'}^{k'o'}$ both being set to 1, which would result in both k and k' assigned to s_i.

Since our solution requires that there is only one label assigned to a region, we have to add constraints that "link" the variables accordingly. This can be accomplished by linking pairs of relations, and start by defining the constraints for the outgoing relations. We arbitrarily take one base relation $r_O \in O_i$ and then create constraints for all $r \in O_i \setminus r_O$. Let $r_O = (s_i, s_j)$ with type t_O, and $r = (s_i, s_{j'})$ with type t be the two relations to be linked. Then, the constraints are $\forall l_k \in L : \sum_{l_o \in L} c_{it_O j}^{ko} - \sum_{l'_o \in L} c_{itj'}^{ko'} = 0$. The first sum can either take the value 0 if l_k is not assigned to s_i by the relation r, or one if it is assigned, and basically the same applies for the second sum. Since both are subtracted and the whole expression has to evaluate to 0, either both equal 1 or both equal 0 and subsequently, if one of the relations assigns l_k to s_i, the others have to do the same. The constraints for the incoming relations are defined accordingly, where r_E is the base relation.

Finally we have to link the outgoing to the incoming relations. Since the same label assignment is already enforced within those two types of relations, we only have to link r_O and r_E, using the following set of constraints: $\forall l_k \in L : \sum_{l_o \in L} c_{it_O j}^{ko} - \sum_{l'_o \in L} c_{j' t_E i}^{o'k} = 0$ Absolute relations are formalized and linked accordingly.

Eventually, let t_r and t_a refer to the type of the relative relation r and the absolute relation a, respectively, then the objective function is defined as

$$\sum_{r=(s_i,s_j)} \sum_{l_k \in L} \sum_{l_o \in L} \min(\theta_i(l_k), \theta_j(l_o)) * \mathcal{T}_{t_r}(l_k, l_o) * c_{it_r j}^{ko} + \sum_{a=s_i} \sum_{l_k \in L} \theta_i(l_k) * \mathcal{T}_{t_a}(l_k) * c_{it_a}^{k}. \quad (2)$$

This function rewards label assignments that satisfy the background knowledge and that involve labels with a high confidence score provided during the classification step.

3. EXPERIMENTS AND RESULTS

We evaluated the approach on a set of 923 images depicting outdoor scenes. We used the labels *building, foliage, mountain, person, road, sailing-boat, sand, sea, sky, snow*. In our evaluation we used the spatial relations *above-of, below-of, left-of* and *right-of*, the absolute spatial relations *above-all* and *below-all*, and we used the thresholds $\sigma = 0.001$ and $\gamma = 0.2$ for both relative and absolute spatial relations. We compared the performance of the low-level classification with the spatial reasoning on different training set sizes and measured *precision (p)*, *recall (r)* and the *classification rate (c)*. Further we computed the *F-Measure (f)*. In Table 1 the average for each of these measures is given.

	Low-Level				BIP			
set size	p	r	f	c	p	r	f	c
50	.63	.65	.57	.60	.77	.75	.73	.75
100	**.70**	.67	**.65**	**.69**	.78	**.77**	.75	.80
150	.67	.63	.61	.66	.74	.71	.70	.75
200	.69	.65	.63	.67	.80	.75	.76	.80
250	.69	.64	.60	.66	.78	.73	.72	.77
300	.68	.63	.61	.66	**.82**	**.77**	**.78**	**.82**
350	.63	**.68**	.61	.66	.80	.75	.76	.80
400	.68	.63	.61	.66	.80	.75	.75	.79

Table 1. Overall results for the two approaches.

The best overall classification rate is achieved with the binary integer programming approach on the data set with 300 training images. However, with only 100 training examples we achieve nearly the same performance, indicating that 100 training examples are a good size for training a well performing classifier using our approach.

4. CONCLUSIONS

In this paper we have introduced a novel spatial reasoning approach based on an explicit model of spatial context. Our results show a good classification rate compared to results in the literature, while requiring only a low number of training data.

5. REFERENCES

[1] M. Grzegorzek and E. Izquierdo, "Statistical 3d object classification and localization with context modeling," in *15th European Signal Processing Conference*, 2007.

[2] Carsten Saathoff and Steffen Staab, "Exploiting spatial context in image region labelling using fuzzy constraint reasoning," in *Proc. of WIAMIS*, 2008.

[3] M. Grzegorzek and H. Niemann, "Statistical object recognition including color modeling," in *2nd International Conference on Image Analysis and Recognition*, 2005.