An Extension of the Radial Line Model to Predict Spatial Relations

Logan Servant, Camille Kurtz and Laurent Wendling
LIPADE, Université Paris Cité, France
{firstname.lastname}@u-paris.fr

Keywords: Spatial Relations, Reference Point, Radial Line Model, Image Understanding, Dataset Denoising.

Abstract: Analysing the spatial organization of objects in images is fundamental to increasing both the understanding of a scene and the explicability of perceived similarity between images. In this article, we propose to describe the spatial positioning of objects by an extension of the original Radial Line Model to any pair of objects present in an image, by defining a reference point from the convex hulls and not the enclosing rectangles, as done in the initial version of this descriptor. The recognition of spatial configurations is then considered as a classification task where the achieved descriptors can be embedded in a neural learning mechanism to predict from object pairs their directional spatial relationships. An experimental study, carried out on different image datasets, highlights the interest of this approach and also shows that such a representation makes it possible to automatically correct or denoise datasets whose construction has been rendered ambiguous by the human evaluation of 2D/3D views. Source code: https://github.com/Logan-wilson/extendedRLM.

1 INTRODUCTION

The description of a scene or its components is fundamental for its understanding. This often requires a recognition of the different objects or regions that constitute it but also of their spatial arrangement. Numerous studies have been carried out for the modeling of spatial relations between objects, in various fields of application of pattern recognition and computer vision. The first notable formalization of spatial relations was proposed by (Freeman, 1975) in the form of 13 spatial relations divided into 3 categories (directional, topological and geometric). (Egenhofer and Franzosa, 1991) proposed an other formalism than Freeman's elementary spatial relations to describe the topological relations between objects, called the 9-intersection model, depending on their position in space. This work, which is the basis of the topological relations RCC8 (for Region Connection Calculus), was then extended to distinguish more topological configurations (Cohn et al., 1997) in order to consider spatial configurations where the objects can be concave and where the intersection of their convex hull is not empty (e.g. RCC23). Early work showed that standard mathematical ”all or nothing” relations are clearly not sufficient to describe spatial relations, and (Freeman, 1975) suggested to use fuzzy relations to refine their evaluation. Many works still assimilate 2D objects with very elementary entities such as a point (centroid) or a bounding rectangle. This procedure is convenient and useful in most cases, but one cannot expect satisfactory modeling (Rosenfeld and Klette, 1985).

Quantitative spatial relationships have been widely studied to model the gradation of the relationship between object pairs according to two dual concepts. (1) The evaluation of the spatial relationship for two objects is based on fuzzy modelling directly from the image space (i.e. a fuzzy landscape), using morphological operations (Bloch, 1999); (2) A typical relative position descriptor is the force histogram (Matsakis and Wendling, 1999), a generalization of the angle histogram. It has the advantage of being isotropic and less sensitive to noise, while allowing the distance between objects to be taken into account explicitly depending on the type of application. More recent work has introduced the ϕ-descriptor (Matsakis and Naeem, 2016) based on Allen's time intervals, providing a generic framework to evaluate usual spatial relations. Finally, other approaches have also been proposed to model more specific spatial relationships such as ”surrounded by” (Vanegas et al., 2011) or ”enlaced by” (Clément et al., 2017). Although these families of relationships provide an accurate assessment of spatial relationships between objects in a very large number of cases, they often require a high processing time, which is not practical when considering large datasets, and a potential learning step.

Servant, L., Kurtz, C. and Wendling, L.
An Extension of the Radial Line Model to Predict Spatial Relations.
DOI: 10.5220/0011644300003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 4: VISAPP*, pages 187-195
ISBN: 978-989-758-634-7; ISSN: 2184-4321

Early work had already shown the interest of fuzzy neural network approaches (Wang and Keller, 1999) for learning and predicting spatial relationships, also incorporating linguistic quantifiers (Matsakis et al., 2001). The advent of deep learning architectures and the creation of large annotated datasets such as *Visual Genome* or *SpatialSense* from crowd-sourcing campaigns have allowed us to reconsider spatial relationship recognition as a classification task. Most of the recent approaches (Peyre et al., 2019) rely on CNN architectures where a model is trained from the coordinates of the objects' bounding boxes, and potentially their semantics, to predict spatial relationships. For example, the *SpatialSense* dataset (Yang et al., 2019), is built using an adversarial crowd-sourcing method, in which humans are asked to find hard-to-predict spatial relationships using bounding boxes. One of the drawbacks of these approaches is that they often only consider the bounding boxes of the objects present in the image and not the segmentation masks, whereas most state-of-the-art computer vision approaches rely on these features. Recently, a powerful new representation called Force Banner, combined with CNN training, has been used to better predict spatial relationships in complex situations (Deléarde et al., 2022). One of the current challenges is the cleaning of erroneous annotations in datasets. This problem is even more crucial for annotations on spatial relationships which are often ambiguous because they depend on human interpretation. In particular, directional spatial relations such as "to the left of" are very sensitive to the viewpoints (2D/3D) and the representation of the scene (e.g. front or back person). This can lead to inconsistent and erroneous predictions in many situations.

We focus in this article on an original and fast approach where the spatial positioning of objects is no longer considered in a relative way (i.e. for one object in relation to another), but from a reference point. After a quick presentation of the original Radial Line Model (RLM) (Santosh et al., 2012) (Sec. 2), we show how to wisely extend it to any pair of objects by defining a reference point from the convex hulls and not the enclosing rectangles (Sec. 3). We propose to integrate other mathematical functions into its modelling in order to provide a more complete spatial interpretation via a viewpoint. The achieved descriptors are then embedded in a neural learning mechanism via a multilayer perceptron to efficiently predict from object pairs their directional spatial relationships (Sec. 4). We also show that such a representation makes it possible to automatically correct datasets whose construction has been rendered ambiguous by the human evaluation of 2D/3D views.

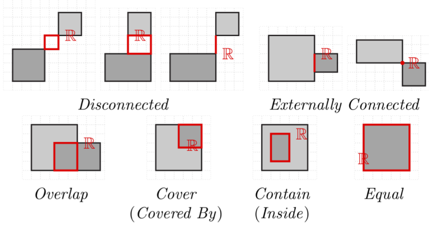


Figure 1: Illustration of the minimum boundary rectangle (MBR) between object pairs (Santosh et al., 2012).

2 RADIAL LINE MODEL: BACKGROUND

The spatial positioning of objects is no longer considered relatively (i.e. one object $\mathbb{A}$ in relation to another object $\mathbb{B}$), but from a reference point $\mathbb{R}_p$ located in the image support. This reference point is defined as the centroid of a region determined by the topological characteristics between the two objects $\mathbb{A}$, $\mathbb{B}$, obtained from the 9-intersection formalism (Egenhofer and Franzosa, 1991).

As seen in (Santosh et al., 2012), minimum boundary rectangle (MBR) can be used to determine the position of this point (see Fig. 1). From this reference point $\mathbb{R}_p$, a pencil of lines along a set of discrete directions is calculated to provide angular coverage of both objects (see Fig. 2).

Let us consider an object $\mathbb{X}$, $\Theta = 2\pi/m$ a constant discretization angle and $line(\mathbb{R}_p, \theta_i)$ being the half-line defined from $\mathbb{R}_p$ and of direction θ_i. The histogram of the angular model $\mathcal{H}$ is defined as:

$$\mathcal{H}(\mathbb{X}, \mathbb{R}_p) = [M(\mathbb{R}_p, j.\Theta)]_{j=0,m-1} \tag{1}$$

where

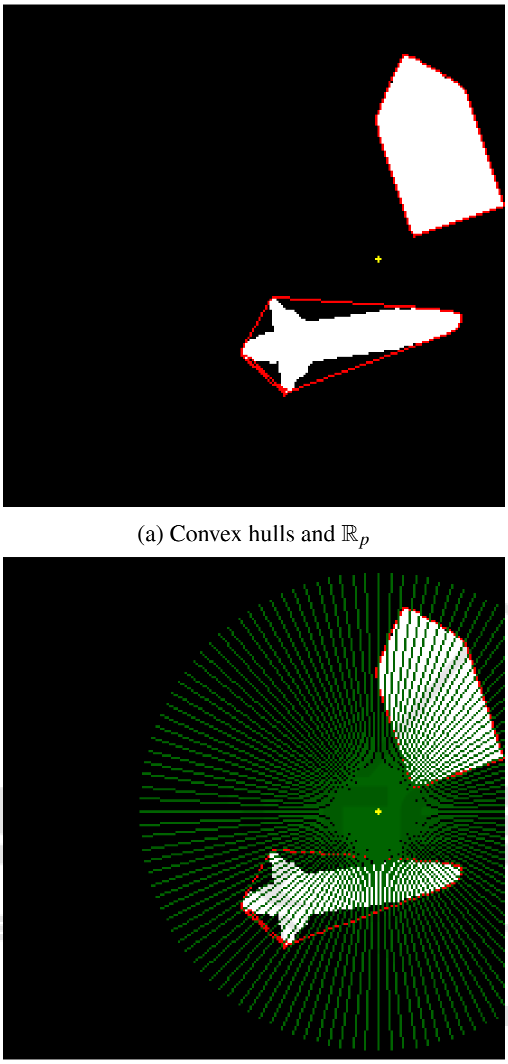
$$M(\mathbb{R}_p, \theta_i) = |line(\mathbb{R}_p, \theta_i) \cap \mathbb{X}| \tag{2}$$

Considering both objects $\mathbb{A}$ and $\mathbb{B}$, a global spatial relation signature $\mathfrak{R}(\mathbb{X}, \mathbb{R}_p)$, denoted as Radial Line Model (RLM), is obtained:

$$\mathfrak{R}(\mathbb{X}, \mathbb{R}_p) = \{\mathcal{H}(\mathbb{A}, \mathbb{R}_p), \mathcal{H}(\mathbb{B}, \mathbb{R}_p)\} \tag{3}$$

Finally, the model can be refined by considering the number of points belonging to $\mathbb{X}$, included in a sector delimited by two consecutive half-lines $line(\mathbb{R}_p, \theta_i)$ and $line(\mathbb{R}_p, \theta_{i+1})$, and then normalized by the size of the object $\mathbb{X}$ to better take into account its shape (Santosh et al., 2014).

It is important to note that this spatial relation model was first defined to solve document analysis problems where similar symbols are usually described by considering four main directions. We show here that this model can be extended and adapted to consider pairwise broad objects and that it can also be integrated into a classification task.



(a) Convex hulls and $\mathbb{R}_p$

(b) Radial lines

Figure 2: Radial Line Model (RLM) computation illustration. From an image containing a pair of segmented objects, the reference point $\mathbb{R}_p$ (in yellow) is computed using convex hulls (in red) of both objects (a). According to the chosen value of $\Theta = 3°$, the radial lines are then created (b).

3 FROM RADIAL LINE MODEL TO SPATIAL RELATION PREDICTION

3.1 Reference Point

The reference point $\mathbb{R}_p$ of the RLM should provide meaningful information about the two objects $\mathbb{A}$ and $\mathbb{B}$ contained in the image (see Fig. 2). The optimal point for any given pair of objects is therefore located between them. As seen in Sec. 2, minimum boundary rectangle (MBR) can be used to determine the position of the point. However, objects may have a topology that leads to the calculation of an eccentric point, which makes the histogram difficult to analyse.

In this article, we focus on the analogical convex hull calculation described in (Freeman, 1975) and the barycenter of each object in order to calculate a new reference point $\mathbb{R}_p$. It is important to note that an object can be defined by disconnected parts. Let us consider an object $\mathbb{X}$ defined from a set of finite points of size $s = |\mathbb{X}|$. By definition, the convex hull of $\mathbb{X}$ forms a 2D convex polygon, denoted $\mathbb{X}_\mathbb{P}$ and by the upper bound theorem, the number of faces of the convex hull of s points is linear in $O(s)$. In this work, we consider the well-known Graham scan algorithm for computing the convex hull in $O(s \ln s)$ time. The Kirkpatrick-Seidel algorithm can also be considered to reduce this complexity. If an object consists of a single connected component without holes, the convex hull can be constructed in linear time from the set of contour points.

Let us consider $C_\mathbb{A}$ the centroid (or barycenter) of object $\mathbb{A}$ (resp. $C_\mathbb{B}$ for object $\mathbb{B}$). If the points coincide, we assume that $\mathbb{R}_p$ is equal to a centroid; otherwise, we focus on the possible intersection points between the line $\mathcal{D}(C_\mathbb{A}, C_\mathbb{B})$ and the two convex hulls $\mathbb{A}_\mathbb{P}$ and $\mathbb{B}_\mathbb{P}$. If $\mathbb{A}_\mathbb{P} \cap \mathbb{B}_\mathbb{P} = \emptyset$ the intersection points closest to $\mathcal{D}(C_\mathbb{A}, C_\mathbb{B})$ and each convex hull, belonging to the segment $[C_\mathbb{A}, C_\mathbb{B}]$ are kept to define a new segment $[I_\mathbb{A}, I_\mathbb{B}]$ whose size corresponds to the minimum distance between $\mathbb{A}_\mathbb{P}$ and $\mathbb{B}_\mathbb{P}$ from $\mathcal{D}(C_\mathbb{A}, C_\mathbb{B})$ that is (reciprocally for $I_\mathbb{B}$):

$$I_\mathbb{A} = \arg\min_{x_i} \left\{ ||x_i – C_\mathbb{B}|| \,|\, x_i \in \{\mathcal{D}(C_\mathbb{A}, C_\mathbb{B}) \cap \mathbb{A}_\mathbb{P}\} \right\} \quad (4)$$

Then, the reference point $\mathbb{R}_p$ is defined as the middle of the segment $[I_\mathbb{A}, I_\mathbb{B}]$. After applying (4) if both points $I_\mathbb{A}$ and $I_\mathbb{B}$ belong to $[C_\mathbb{A}, C_\mathbb{B}]$ (i.e., a weak overlapping of convex hulls) the coordinates of the reference point are calculated as before; otherwise (high overlapping) if $I_\mathbb{A} \in [C_\mathbb{A}, C_\mathbb{B}]$ (resp. $I_\mathbb{B}$) the reference point is set in the middle of $[I_\mathbb{A}, C_\mathbb{B}]$ (resp. $[C_\mathbb{A}, I_\mathbb{B}]$).

The proposed method for determining the reference point has the advantage of being fast, although a little slower than the original MBR method, and allows to limit the number of occurrences where the reference point overlaps with objects when the considered objects are close thanks to convex hulls.

3.2 Extended Radial Line Model

Without loss of generality, the original Radial Line Model described in Sec. 2 can be easily rewritten from a set of m half-lines defined from polar coordinates as follows:

$$D^{\theta_i}(x,y) = (x_p + r\cos\theta_i, x_p + r\sin\theta_i) \qquad (5)$$

with a constant angular step $\Theta = 2\pi/m$ calculated from $m \in \mathbb{N}$, an even integer value representing all the studied directions $\theta_{i=0,m-1} \in [0, 2\pi[$ and $r = (x - x_p)/\cos\theta_i = (y - y_p)/\sin\theta_i$. Then an accumulator is applied as in Eq. (2) but we normalized the reached value by $\sin(\theta_i)$ (resp. $\cos(\theta_i)$) following the sector in a isotropic way.

Additionally, we propose to integrate in our global spatial relation signature $\mathfrak{R}$ (Eq. 3) a new spatial relation distribution by considering attractive forces (Matsakis and Wendling, 1999) between two objects but limited to the set of lines of the radial model. This amounts to evaluating a new type of spatial relations from a reference point of view defined between the pair of objects.

As in (Matsakis and Wendling, 1999) the base of the model relies on the definition of a force of attraction between points, extended to the handling of segments. Given a point of object $\mathbb{A}$ and a point of object $\mathbb{B}$ located at a distance d from each other, their force of attraction is computed as:

$$\varphi_r(d) = \frac{1}{d^r} \qquad (6)$$

where r characterizes the kind of force processed. Let $I^{\mathbb{A}}$ and $J^{\mathbb{B}}$ be two segments on a line of angle θ_i belonging to a pencil of lines describing the RLM defined previously. $D^{\theta_i}_{I^{\mathbb{A}}J^{\mathbb{B}}}$ is the distance between them and $|.|$ the segment length. The force of attraction f_r of segment $I^{\mathbb{A}}$ with regards to segment $J^{\mathbb{B}}$ is given by:

$$f_r(I^{\mathbb{A}}, J^{\mathbb{B}}) = \int_{D^{\theta_i}_{I^{\mathbb{A}}J^{\mathbb{B}}}+|J^{\mathbb{B}}|}^{|I^{\mathbb{A}}|+D^{\theta_i}_{I^{\mathbb{A}}J^{\mathbb{B}}}+|J^{\mathbb{B}}|} \int_0^{|J^{\mathbb{B}}|} \varphi_r(u-v)\,\mathrm{d}v\,\mathrm{d}u. \qquad (7)$$

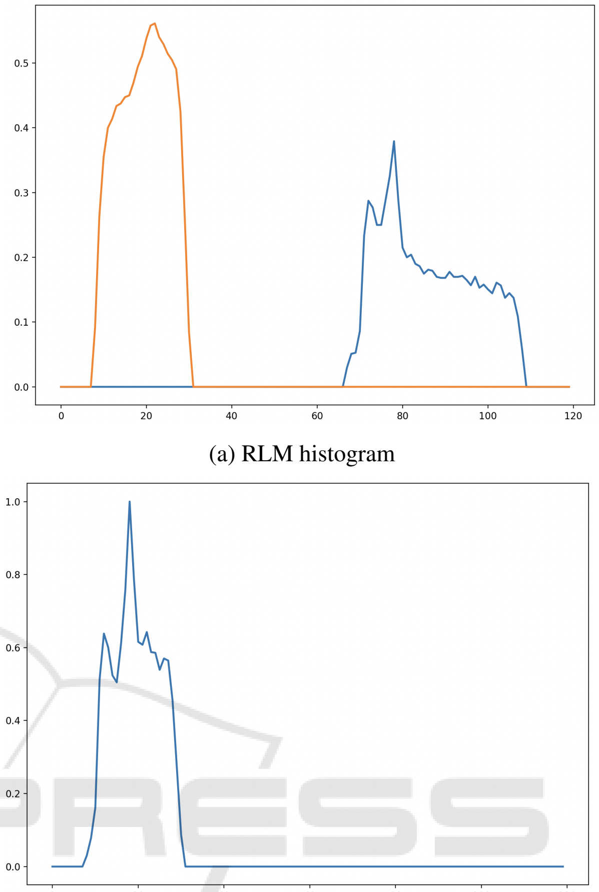
Given two binary objects $\mathbb{A}$ and $\mathbb{B}$, each θ_i-oriented line of the RLM forms two sets of segments belonging to each object: $\mathcal{C}_{\mathbb{A}} = \cup\{I^{\mathbb{A}}_j\}_{j=1,n^{\mathbb{A}}_{\theta_i}}$ and $\mathcal{C}_{\mathbb{B}} = \cup\{I^{\mathbb{B}}_j\}_{j=1,n^{\mathbb{B}}_{\theta_i}}$. The mutual attraction between these segments is defined as:

$$F_r(\theta_i, \mathcal{C}_{\mathbb{A}}, \mathcal{C}_{\mathbb{B}}) = \sum_{I^{\mathbb{A}}\in\mathcal{C}_A}\ \sum_{J^{\mathbb{B}}\in\mathcal{C}^{\mathbb{B}}_J} f_r(I^{\mathbb{A}}, J^{\mathbb{B}}). \qquad (8)$$

Considering all the m oriented lines of direction θ_i allows to define a new signature similar to Eq. (1):

$$\mathcal{H}((\mathbb{A},\mathbb{B}),\mathbb{R}_p) = [F_r(\theta_i, \mathcal{C}_{\mathbb{A}}, \mathcal{C}_{\mathbb{B}})]_{i=0,m-1} \qquad (9)$$

Then we focus on two levels of forces widely used in the literature and efficient to assess spatial relation between a couple of objects:



(a) RLM histogram

(b) Force histogram

Figure 3: Computed descriptors for the spatial configuration presented in Fig. 2. The three histograms use $\Theta = 3°$. (a) Radial Line Model from each object, using the middle point between the two objects of the image. Object A corresponds to the referent, object B, to the argument. (b) Force histogram using f_0 around a reference point after Max normalization.

- $f_{r=0}$ relies on constant forces which are independent of the distance between objects. In some extent this approach is based on the handling of an isotropic histogram of angles;
- $f_{r=2}$ relies on gravitational forces where more importance is given to closer points.

Fig. 3 provides some examples of RLM and force histograms for the spatial configuration from Fig. 2.

3.3 Recognition of Spatial Relations

Based on the achieved global spatial relation signatures (extended RLM + forces) between pairs of objects, we propose a generic framework to translate our descriptor into spatial relations expressed in natural language, and / or potentially into more compact spatial features which can be used in a larger recognition tasks, for example when a scene is composed of multiple pairs of objects.

Two main options are possible to translate relative position descriptors into spatial relations in natural language: relying on machine learning to automatically generate the transformations, as in (Wang and Keller, 1999) for the histogram of angles, or using predefined evaluation rules from theoretical analysis, as in (Matsakis et al., 2001) for the force histogram. We propose to use machine learning for our descriptor, by learning the transformation from a dataset annotated with object pairs and their spatial relations.

4 EXPERIMENTAL STUDY

We develop three ways to showcase the interest of our method: the first way aims to show the ability of our descriptors to capture enough spatial information to predict spatial relations between object pairs by training a model on a given dataset. The second one aims to learn a spatial model from synthetic images and predict spatial relations in satellite images, while the third way deals with the denoising of a given ambiguous dataset.

4.1 Datasets

Different datasets of (synthetic or natural) images were considered in this study. Each image depicts a scene containing a specific spatial configuration between a pair of crisp objects, including corresponding annotations. SimpleShapes dataset contains 2280 synthetic images, divided in two distinct sub-datasets named SimpleShape1 (S1) and SimpleShape2 (S2). S1 comprises masks of complex objects such as boats and cars (see Fig. 2), while S2 is composed of convex and concave geometric objects such as triangles, and ellipses. Images have been synthesised in a random way, with no background, and created by generating random orientation, scale and place on the image. The GIS dataset is composed of 211 images representing spatial configurations of geographical objects (e.g., houses, river) sensed from aerial images.

Each image of these three datasets contains annotations in which their spatial relations has been assessed by three different experts using the four main directions (North, West, South, East) (Deléarde et al., 2022). Images were also ranked from N1 to N4 according to the difficulty to determine the spatial relations between the two objects (N1 corresponds to easiest, N4 to ambiguous and/or undecidable). For this experimental study, N4 images were rejected from the datasets, lowering the total images used of SimpleShapes to 1993, and to 190 images for GIS.

Table 1: Comparison of different methods to compute non-overlapping reference points ($\mathbb{R}_p$) in the 2280 images of the SimpleShapes dataset. Overlapped $\mathbb{R}_p$ does not prevent computation of RLM and forces but decreases the quality of the obtained descriptors.

$\mathbb{R}_p$ computation	% of $\mathbb{R}_p$ overlapping w/ objects
Straight MBR	71.1%
Oriented MBR	73.7%
Mean of centroids	74.4%
Convex hulls	95.8%

For some experiments, we also considered a subset of the SpatialSense dataset (Yang et al., 2019), composed of 11570 natural images representing everyday life scenes (S3), see e.g. Fig. 4. For each image, SpatialSense provides different spatial annotations (bounding boxes of objects and spatial configurations between them), with spatial relations between object pairs. We restrict the dataset to images presenting *to the left of*, *above*, *to the right of*, or *below* spatial relations, thus reducing the size to 2290 images. However, some spatial relations are given in a 3-dimensional space. The orientation of subjects and objects are taken into account when given a spatial relation, which means the spatial relation may vary a lot depending on the point of view (2D or 3D). According to our experiences, we may need segmented objects. We have then pre-processed this dataset to obtain regions corresponding to the objects of interest via a segmentation performed in the bounding boxes provided in the annotations (Deléarde et al., 2021).

4.2 Directional Relation Classification

We aim in this preliminary experiment to showcase the ability of the proposed method to predict directional spatial relations from the images characterized by our descriptors, and the importance of using convex hulls as the basis to obtain the reference point $\mathbb{R}_p$.

4.2.1 Experimental Protocol

As mentioned in Sec. 2 different methods can be employed to determine the reference point $\mathbb{R}_p$, such as straight MBR, oriented MBR and mean distance to barycenters of both objects. In cases where $\mathbb{R}_p$ overlaps with an existing object, results from the RLM may vary a lot and can create errors in classification. We aim to minimize the average number of images where the reference point $\mathbb{R}_p$ overlaps with an object while still retaining a correct position to obtain the necessary information to predict the spatial relation of the two objects. A comparative study of the

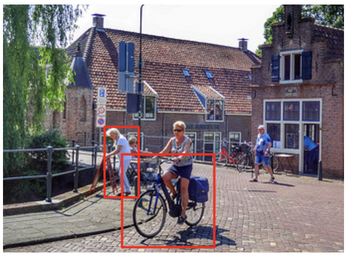


Figure 4: A SpatialSense image containing an ambiguous annotation. The image connects the bicycle in the foreground, to the woman on the left, considers in the annotations "Bicycle to the left of woman", but the bicycle is on the right of the woman.

behavior of the different methods possible to compute $\mathbb{R}_p$ is provided in Tab. 1, using the SimpleShapes dataset. This preliminary result confirms our intuition and highlights the interest of using convex hulls as the basis to obtain the reference point. We will then only consider this approach to compute our descriptors.

We now illustrate the ability of the proposed method to predict directional spatial relations. For this experiment, we use the whole S1 and S2 datasets individually. In order to showcase the ability of our method to predict correct spatial relations from these object configurations, we compute for each pair the extended RLM of each object and the forces associated with the relation (RLM + f_2). For RLM and force histograms, we considered 120 values ($\Theta = 3°$) leading to a final feature vector of dimension 360.

Given these image representations, we trained different models to predict the correct spatial relation: Support-Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Random Forests (RF). The three models were implemented with default parameters. As a comparison method, we implemented another strategy relying only on the bounding boxes of the object pairs (B-Box approach), which is similar to the method "2D-only" of SpatialSense (Yang et al., 2019). Technically, an image containing a spatial configuration can be characterized by considering the coordinates of all bounding boxes of objects. For this approach, the same classifiers have been considered.

For each tested approach, we used a 5-fold cross-validation (to learn on 80% of the data), and computed accuracy and standard deviation of cross-validation on test sets (20% of the data).

4.2.2 Results

The obtained results are presented in Tab. 2. Overall, the models trained from our descriptors or the comparative B-Box approach offer very good performance (with an accuracy greater than 0.9 and a low standard deviation). However, using the extended RLM with forces allowed to average +0.02 in accuracy with almost no changes in the standard deviation of the cross-validation method. The only model performing worse compared to the bounding box coordinates is the SVM on the S1 dataset.

Table 2: Classification of directional relations. Comparison of accuracy and standard deviation of cross-validation on test sets using different supervised models fed with radial line models and forces (RLM + f_2) and models with bounding box coordinates (B-Box).

	SVM	k-NN	RF
S1 (B-Box)	**0.95**±0.02	0.90±0.02	0.91±0.02
S2 (B-Box)	0.96±0.01	0.93±0.02	0.94±0.01
S1 (RLM)	0.93±0.03	0.91±0.03	0.92±0.03
S2 (RLM)	0.95±0.02	0.94±0.03	0.94±0.02
S1 (RLM + f_2)	0.93±0.02	**0.93**±0.03	**0.95**±0.01
S2 (RLM + f_2)	**0.97**±0.01	**0.96**±0.01	**0.97**±0.01

This first study shows the discriminability capacity of directional spatial configurations between objects offered by our descriptors as well as its comparability to a (naive but) state-of-the-art approach.

4.3 Synthetic to Natural Configurations

In a second experiment, we present a practical use case of our spatial relationship prediction method. Despite the lack of annotated data (with this type of spatial information) in the literature, we want here to show that it is possible to learn powerful spatial models from synthetic data and to transfer these models to deal with natural images containing realistic spatial configurations.

4.3.1 Experimental Protocol

This experiment uses the same (synthetic) datasets as the former experiment with the addition of the GIS dataset composed of natural images. This experiment aims to show the ability of the extended Radial Line Model to predict spatial relations of natural images from a learning on another dataset.

As discussed in Sec. 3.2, for each object pair, we compute the spatial relation signature $\mathfrak{R}(\mathbb{X}, \mathbb{R}_p)$, composed of the both radial line models of each object, and the force histogram between the referent and the argument. We use a value of Θ=3°, which means each histogram comprises 120 values, leading to a final feature vector of dimension 360. The tested approach uses a Multi-Layer Perceptron (MLP) of 4 hidden layers of 448 units each with a ReLU activation function, and Adam Optimizer (with initial learning rate of 10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$).

For comparative purposes, different other models have been implemented and evaluated:

Table 3: Transfer: From synthetic configurations to natural ones. Comparison of different models trained on SS synthetic datasets and then used to predict spatial relations from the GIS dataset.

Model	B-Box	RLM$_{init}$	f_0	f_2	RLM	RLM + f_2	RLM + f_0 + f_2
Accuracy	75%	83%	72%	79%	87%	90%	93%

- as baseline, the B-Box approach and the initial RLM model (Santosh et al., 2012);
- a model using a 240-dimensional vector built from the spatial relation signature $\mathfrak{R}(\mathbb{X},\mathbb{R}_p)$ (without force information);
- models using only a 120-dimensional vector, the force histogram (f_0, f_2).

This choice aims to demonstrate the relevance of combining different information to obtain a model more robust than each descriptor considered individually. The parameters of the MLP remain the same.

We train each model on the two synthetic datasets (S1+S2) and employ a cross-validation strategy, with five subsets (of 20% of the samples each) as test sets.

4.3.2 Results

Each model trained is then applied to predict spatial relations on the 190 GIS segmented images. Most of the image contain easily found relationships (118 images with N1 level) and contain mostly horizontal relationships, with 86 *to the left of* and 49 *to the right of* (see (Deléarde et al., 2022)). Tab. 3 presents the percentage of correctly found relations for each tested model. Using only forces, first with f_0 and then with f_2, the accuracy of the models reach between 72% and 79%, while using only RLM of each object reaches an accuracy of 87%. f_2 providing better overall results, we used this type of forces to combine with the RLM. The accuracy of the prediction shows slightly better although considerable improvements. By combining the Radial Line Model of both objects, with both types of forces used before, we obtain a more robust and better model with better accuracy.

4.4 Dataset Denoising

In this last experiment, we show how our method can be practically implemented to correct datasets containing (human) spatial annotations that may be erroneous or ambiguous.

4.4.1 Experimental Protocol

The last experiment aims to provide a correction of the SpatialSense (S3) dataset. As mentioned earlier, this dataset contains annotations about the spatial relations between objects present in the images. However, as it has been annotated by humans, via crowd-sourcing campaigns, some annotations can be ambiguous because they depend on human interpretation. In particular, directional spatial relations are very sensitive to the viewpoints (2D/3D) and the representation of the scene (e.g. Fig. 4). This can lead to inconsistent learning and erroneous predictions.

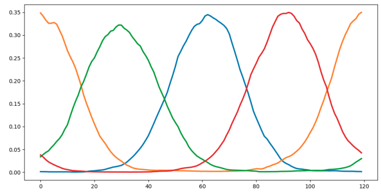


Figure 5: Prototypes of each spatial relation class in the SimpleShapes dataset (*to the right of* in orange, *above* in green, *to the left of* in blue, *under* in red).

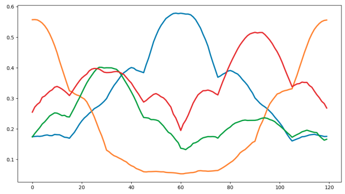


Figure 6: The prototypes of the *to the left of* and *to the right of* each mix two curves, the bigger one corresponds to the 2D viewpoint, while the smaller one regroups all of the wrongly annotated spatial relations (3D viewpoint). The prototypes of the *above* and *under* although not as affected, still display forms of noise created by 3D viewpoints.

To further illustrate this problem of ambiguity, we propose in a preliminary experiment to build and visualize, for each class of spatial relationship of a dataset, the prototypes obtained by averaging, for all the images belonging to a class, the proposed spatial relation descriptors. It is interesting to compare the prototypes obtained from the *clean* SimpleShapes dataset (Fig. 5) to the ones obtained on the *noisy* SpatialSense dataset (Fig. 6) to further motivate the need of denoising it.

The goal of this experiment is then to provide automatic corrections of these spatial relations. As previously, we focus here solely on the four main directional spatial relations. Out of all available images in the test set of S3, 100 labels have been manually corrected to match spatial relations in a 2-dimensional space. Most of the corrections are done on the left-right axis, as most of the ambiguity in the dataset are based on the orientation of objects in the space. Few objects in the images had an orientation which lead to a different interpretation depending on the number of dimension chosen. This manual correction leads to a cleaned test set of S3 that we will consider later as a ground truth to evaluate the impact of our dataset denoising strategy.

Table 4: Cleaning process of the training set of SpatialSense (S3). For each class of the initial (noisy) dataset, the table illustrates the distribution of the new spatial labels assigned to the configurations.

		Cleaned labels			
Initial label	# images	Under	Right	Above	Left
Under	597	197	194	104	102
Right	498	52	236	58	152
Above	429	68	60	191	110
Left	666	70	241	84	271

Table 5: Classification results of each model on the 100 hand-made corrections of S3. The models were trained either on the noisy version of S3, or on its clean version.

Training	Model	Accuracy
Noisy S3	B-Box	0.36
	RLM + forces (bboxes)	**0.46**
	RLM + forces (segment.)	0.29
Cleaned S3	B-Box	0.83
	RLM + forces (bboxes)	0.93
	RLM + forces (segment.)	**0.95**

To denoise a dataset like SpatialSense, we propose to learn a clean model for predicting spatial relationships from a synthetic dataset like SimpleShapes. We propose to use as models the prototypes stated above (Fig. 5) computed by aggregating the spatial signatures for each class on S1+S2. From this model, the next step is to apply it to the S3 training set to correct one by one the annotations of the images that may contain errors. This can be done by a calculation of affinity scores between the representation of the image tested and the set of prototypes of the model. Depending on the obtained result, the spatial label of the image is corrected or not. This process leads to a cleaned S3 dataset, which can then be considered to train a model of spatial relationship predictions, which can be tested on the ground truth set.

To showcase the added value of this process, we can compare the performance on the ground truth of a model trained on the cleaned version of the dataset versus a model trained on the initial noisy version of the dataset. Here different representations were compared, with the objective to highlight the interest of the force information and the radial line models: (1) A strategy relying only on the coordinates of the bounding boxes of the object pairs (B-Box approach). A scene is then described by 8 integers: 4 for each object (x_{min},x_{max},y_{min}, and y_{max}); (2) A strategy that considers the bounding box of each object (initially provided in the annotations of S3) as crisp regions on which to apply combination of the forces and the RLM (same parameters as in Sec. 4.2). Note that simplifying the object as a bounding box leads in some cases to boxes overlapping each other and creating wrong relations. In some other cases, the bounding boxes created for objects did not fit correctly the objects whose topology had a center of gravity far away from the center of the corresponding bounding box; (3) A strategy to correct the problems listed above which was implemented using relatively precise object segmentations of the S3 images thanks to (Deléarde et al., 2021). It allows for more precise topology of objects to be used. The forces and the RLM are then computed from these regions. As in Sec. 4.3, we use a MLP to learn a model from these representations to predict the relations.

4.4.2 Results

As preliminary results, Tab. 4 presents the outputs of the cleaning process applied to the training set of S3 thanks to the prototypes learnt on S1+S2. As expected, most of the relationships that were corrected were horizontal (left, right) and were ambiguous due to the point of view considered.

To evaluate the impact of the denoising step and the interest of learning on a clean dataset, Tab. 5 provides the classification results (accuracy) of each assessed model on the 100 hand-made corrections of S3 which is considered as ground-truth. As observation, the combination of RLM and force information carries more useful spatial information to discriminate the different spatial configurations than the B-Box model. Furthermore, on a natural image dataset such as SpatialSense, we also note the interest of considering regions that realistically approximate the objects of interest whose spatial configurations are studied.

5 CONCLUSION

To model spatial relations between object pairs, the proposed approach combines the Radial Line Model and the forces histogram computed from a reference point, computed from the convex hulls of objects. Used as an image representation, this model outperforms state-of-the-art models in classifying spatial configurations of objects, albeit in a negligible slower time. The extended RLM can transfer spatial relations from one dataset to another one, and also denoise datasets subjected to the human error of viewpoint on images. This novel approach could further be improved by employing a convolutional neural network architecture to learn the representation in an end-to-end fashion. Another improvement could be made by using a morphological operator such as the geodesic distance to extract the reference point, to better take into account concave objects.

REFERENCES

Bloch, I. (1999). Fuzzy Relative Position between Objects in Image Processing: A Morphological Approach. *IEEE TPAMI*, 21(7):657–664.

Clément, M., Poulenard, A., Kurtz, C., and Wendling, L. (2017). Directional Enlacement Histograms for the Description of Complex Spatial Configurations between Objects. *IEEE TPAMI*, 39(12):2366–2380.

Cohn, A., Bennett, B., Gooday, J., and Gotts, N. (1997). Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 3(1):275–316.

Deléarde, R., Kurtz, C., Dejean, P., and Wendling, L. (2021). Segment my object: A pipeline to extract segmented objects in images based on labels or bounding boxes. In *VISAPP*, pages 618–625.

Deléarde, R., Kurtz, C., and Wendling, L. (2022). Description and recognition of complex spatial configurations of object pairs with force banner 2d features. *PR*, 123:108410.

Egenhofer, M. and Franzosa, R. (1991). Point-set topological spatial relations. *Int. J. GIS*, 2(5):161–174.

Freeman, J. (1975). The Modelling of Spatial Relations. *CGIP*, 4(2):156–171.

Matsakis, P., Keller, J. M., Wendling, L., Marjamaa, J., and Sjahputera, O. (2001). Linguistic description of relative positions in images. *IEEE TSMC*, 31(4):573–88.

Matsakis, P. and Naeem, M. (2016). Fuzzy Models of Topological Relationships Based on the PHI-Descriptor. In *FUZZ-IEEE*, pages 1096–1104.

Matsakis, P. and Wendling, L. (1999). A new way to represent the relative position between areal objects. *IEEE TPAMI*, 21(7):634–643.

Peyre, J., Laptev, I., Schmid, C., and Sivic, J. (2019). Weakly-Supervised Learning of Visual Relations. In *ICCV*, pages 5189–5198.

Rosenfeld, A. and Klette, R. (1985). Degree of adjacency or surroundedness. *PR*, 18(2):169–177.

Santosh, K. C., Lamiroy, B., and Wendling, L. (2012). Symbol recognition using spatial relations. *Pattern Recognit. Lett.*, 33(3):331–341.

Santosh, K. C., Wendling, L., and Lamiroy, B. (2014). Bor: Bag-of-relations for symbol retrieval. *Int. J. Pattern Recognit. Artif. Intell.*, 28(6).

Vanegas, M. C., Bloch, I., and Inglada, J. (2011). A fuzzy definition of the spatial relation “surround” - Application to complex shapes. In *EUSFLAT*, pages 844–851.

Wang, X. and Keller, J. M. (1999). Human-based spatial relationship generalization through neural/fuzzy approaches. *Fuzzy Sets Syst.*, 101(1):5–20.

Yang, K., Russakovsky, O., and Deng, J. (2019). Spatialsense: An adversarially crowdsourced benchmark for spatial relation recognition. In *ICCV*, pages 2051–2060.