FORCE HISTOGRAMS AND NEURAL NETWORKS FOR HUMAN-BASED SPATIAL RELATIONSHIP GENERALIZATION

R. Bondugula [1], P. Matsakis [2], J. Keller [1]

[1] University of Missouri-Columbia
Dept. of Computer Science
Columbia, MO 65211-2060
USA

[2] University of Guelph
Dept. of Computing and Information Science
Guelph, ON N1G 2W1
Canada

Abstract

The aim of this paper is the design of a system that can learn any individual user's perception of four subjective and key spatial relationships between image objects: "to the right of," "above," "to the left of" and "below." The proposed approach is based on the utilization of artificial Neural Networks (NNs) and the modeling of relative positions between 2D objects through histograms of forces. The NNs are fed by features extracted from the histograms and are trained to numerically assess the four relationships according to the target perception.

Key Words

High-level computer vision, human spatial perception, spatial relationships, force histograms, neural networks.

1. Introduction

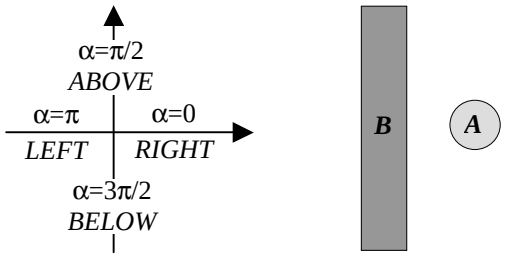
The modeling of spatial relationships—and especially of the four relationships "to the right of," "above," "to the left of" and "below"—is an important challenge in computer vision [1] [2] [3]. Applications can be found in robotics, scene description and spatial databases, to name a few. Due to the subjective nature of the four relationships above, several attempts were made to learn them from examples [4] [5] [6]. In this paper, we revisit the work in [6]. A *directional spatial relation* is a fuzzy binary relation between 2D image objects that depends on two parameters: a *direction* (or angle) α and a *perception*. With any pair *(A,B)* of objects, the relation associates a value between 0 and 1. This value is interpreted as the degree of truth, according to the given perception, of the proposition "*A* is in direction α of *B*." Object *A* is the *argument* and object B the *referent*. The angles 0, $\pi/2$, π and $3\pi/2$ correspond to the directions *RIGHT*, *ABOVE*, *LEFT* and *BELOW*, which are the four *primitive directions* (Fig. 1a). For instance, the propositions "*A* is in direction π of *B*" and "*A* is to the *LEFT* of *B*" are equivalent. Their degree of truth is denoted by $d^{AB}(LEFT)$. If $d^{AB}(LEFT)$ is 0 then *A* is "not at all" to the left of *B* and if $d^{AB}(LEFT)$ is 1 then *A* is "perfectly" to the left of *B*. Fig. 1 shows the values of $d^{AB}(RIGHT)$, $d^{AB}(ABOVE)$, $d^{AB}(LEFT)$ and $d^{AB}(BELOW)$ for a given pair of image objects and according to two different perceptions. Our goal is to build a system that can learn any individual user's perception of the four primitive directional spatial relations. Like in [6], the proposed approach is based on the modeling of relative positions between image objects through histograms; the histograms provide inputs to Neural Networks (NNs) which are trained to assess the four degrees of truth according to the target perception. Contrary to [6], however, the histograms are force histograms [7], not angle histograms [8]; the NN inputs are histogram features, not histogram values (Fig. 2). In the end, important properties of directional relations are preserved, the number of image object pairs for training is drastically cut down, and the system learns faster and generalizes better. Section 2 goes over the notion of the histogram of forces and discusses feature extraction. Section 3 describes the NNs and the structure of the input and output vectors. Section 4 introduces the users whose perceptions we attempt to capture. It also presents the training and testing sets of image object pairs. The results of all experiments are displayed and discussed in Section 5. Conclusion is given in Section 6.

2. Force Histogram Features

The systems described in this paper use different tuples of features extracted from different types of force histograms. The notion of the histogram of forces [7] is briefly presented in Section 2.1. The features likely to serve as NN input values are described in Section 2.2. A normalization procedure of the histograms is then explained in Section 2.3. The procedure will be used in Section 3 for dynamic feature selection and ordering.

2.1. The Histogram of Forces

The relative position of a 2D image object *A* with regard to another object *B* is represented by a function F^{AB} from $\mathbb{R}$ (the set of real numbers) into $\mathbb{R}_+$ (the set of non-negative real numbers). *A* and *B* are seen as two flat metal plates of uniform density and constant and negligible



(a)

(b)

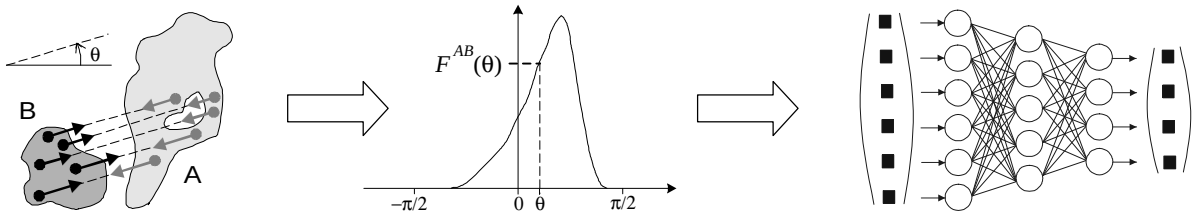
Jane's perception	
$d^{AB}(RIGHT)$	1.0
$d^{AB}(ABOVE)$	0.0
$d^{AB}(LEFT)$	0.0
$d^{AB}(BELOW)$	0.0

(c)

John's perception	
$d^{AB}(RIGHT)$	1.0
$d^{AB}(ABOVE)$	0.3
$d^{AB}(LEFT)$	0.0
$d^{AB}(BELOW)$	0.3

(d)

Fig. 1. Directional spatial relations. (a) The four primitive directions. (b) A pair *(A,B)* of image objects. (c) Values of $d^{AB}(RIGHT)$, $d^{AB}(ABOVE)$, $d^{AB}(LEFT)$ and $d^{AB}(BELOW)$ according to Jane, student at the University of Missouri. (d) Values according to John. Both students agree that *A* is perfectly to the right of *B*. However, John also considers that *A* is a little above *B* and a little below.



(a) (b) (c) (d)

Fig. 2. Synoptic diagram of the systems tested in this paper. (a) A pair *(A,B)* of image objects is presented to the system. (b) The force histogram that represents the relative position of the two objects is computed. (c) Features are extracted from the histogram to constitute the neural network input vector. (d) The network outputs a vector whose components are the degrees of truth of the propositions "*A* is to the *RIGHT* of *B*," "*A* is *ABOVE B*," "*A* is to the *LEFT* of *B*" and "*A* is *BELOW B*."

thickness—a kind of objects commonly considered in physics. For any direction θ, the value $F^{AB}(\theta)$ is the scalar resultant of elementary forces. These forces are exerted by the particles of *A* on those of *B*, and each tends to move *B* in direction θ (black arrows in Fig. 2a). F^{AB} is the *histogram of forces associated with (A,B) via F*, or the *F–histogram associated with (A,B)*. Actually, *F* denotes a numerical function and defines the force fields. Let *r* be a real number. If the elementary forces are in inverse ratio to d^r, where *d* represents the distance between the particles considered, then *F* is denoted by F_r. The F_0-histogram (histogram of constant forces) and F_2-histogram (histogram of gravitational forces) have very different and interesting characteristics. The former considers the closest parts and the farthest parts of the objects equally, whereas the F_2-histogram focuses on the closest parts. Details can be found in [7].

2.2. Feature Extraction

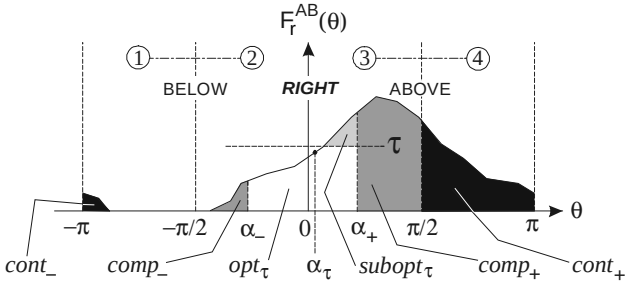


Fig. 3. Force typology associated with the proposition "*A* is to the *RIGHT* of *B*."

In practice, forces are computed in a finite number of evenly distributed directions. Although the computed values can serve as direct inputs to the neural network a given system relies on, other histogram features can be proposed. Consider, for instance, the proposition "*A* is to the *RIGHT* of *B*." First, the set of directions is divided into four quadrants as shown in Fig. 3. The forces $F_r^{AB}(\theta)$ of the first quadrant ($\theta \in [-\pi,-\pi/2]$) are *contradictory* forces. They weaken the proposition. The amount of these forces (i.e., area of the black region on the left of Fig. 3) is denoted by $cont_-$. The forces of the second quadrant ($\theta \in [-\pi/2,0]$) are elements that support the proposition. Some are used to compensate the contradictory forces mentioned above. The amount of these *compensatory* forces (i.e., area of the dark gray region on the left of Fig. 3) is denoted by $comp_-$ and determined by some angle α_-. The values $cont_+$, $comp_+$ and α_+ are defined in the same way (see Fig. 3, quadrants 3 and 4). The remaining forces ($\theta \in [\alpha_-, \alpha_+]$) are the *effective* forces. A threshold τ divides them into *optimal* and *sub-optimal* components. The amount of the optimal components (area of the white region in Fig. 3) is denoted by opt_τ. The amount of the sub-optimal components (area of the light gray region) is denoted by $subopt_\tau$. The optimal components support the idea that *A* is "perfectly" to the right of *B*: whatever their direction, they are regarded as horizontal and pointing to the right. The "average" direction α_τ of the effective forces is computed in conformity with this agreement. It allows the degree of truth d_τ of the proposition "*A* is to the *RIGHT* of *B*" to be assessed. d_τ is a decreasing function of $|\alpha_\tau|$ and an increasing function of the percentage of the

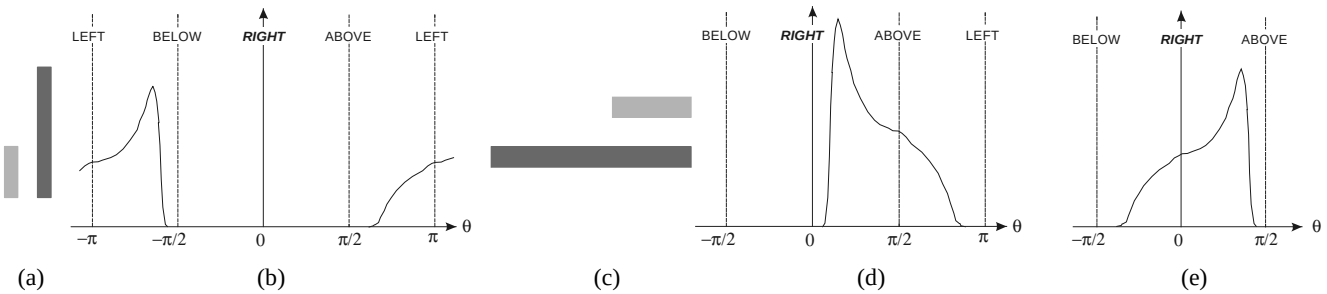


Fig. 4. (a) The reference image. (b) The associated histogram. The primary direction is *LEFT* and the secondary direction is *BELOW*, i.e., $(\delta_1,\delta_2,\delta_3,\delta_4)$ is *(LEFT,BELOW,RIGHT,ABOVE)*. (c) Reference image rotated through π/2, mirrored about the horizontal axis and rescaled. (d) The associated histogram. The primary direction is *ABOVE* and the secondary direction is *RIGHT*, i.e., $(\delta_1,\delta_2,\delta_3,\delta_4)$ is *(ABOVE,RIGHT,BELOW,LEFT)*. (e) The normalized histogram associated with both (a) and (c). The area under the curve is 1. The primary direction is *RIGHT* and the secondary direction is *ABOVE*. Note that swapping the referent and argument objects in (a) or (c) would give the same normalized histogram.

effective forces. For instance, if all forces are contradictory then d_τ is 0 ("*A* is to the *RIGHT* of *B*" is judged to be completely false) and if all forces are effective and optimal then d_τ is 1 (proposition completely true). There are of course many ways of choosing the threshold τ. If τ is $+\infty$, all the effective forces are optimal and $subopt_{+\infty}$ is 0 (most optimistic point of view). Conversely, if τ is 0, all the effective forces are sub-optimal and opt_0 is 0 (most pessimistic point of view). Between these two extremes, one can choose a threshold τ_S that depends on $\{(\theta, F_r^{AB}(\theta))\}_{\theta\in[\alpha_-,\alpha_+]}$ and such that neither $subopt_{\tau_S}$ nor opt_{τ_S} are 0 [9]. Finally, eighteen values are associated with the direction *RIGHT*. Ten of them are forces: $cont_-$, $cont_+$, $comp_-$, $comp_+$, opt_0, opt_{τ_S}, $opt_{+\infty}$, $subopt_0$, $subopt_{\tau_S}$ and $subopt_{+\infty}$. Five are angles: α_-, α_0, α_{τ_S}, $\alpha_{+\infty}$ and α_+. Three are degrees of truth: d_0, d_{τ_S} and $d_{+\infty}$. Details on the computation of all these values can be found in [9]. Obviously, the other primitive directions, *ABOVE*, *LEFT* and *BELOW*, can be handled by applying exactly the same procedure to the shifted histograms $F_r^{AB}(\theta+\pi/2)$, $F_r^{AB}(\theta+\pi)$ and $F_r^{AB}(\theta+3\pi/2)$. In all, seventy-two numerical values (18×4) are extracted from F_r^{AB}. It can easily be shown, however, that some are redundant, some are always zero, etc. After feature reduction, only seventeen forces and angles are retained:

$$cont_-(RIGHT),\ comp_-(RIGHT),\ comp_+(RIGHT),$$
$$subopt_{\tau_S}(RIGHT),\ cont_-(ABOVE),\ comp_-(ABOVE),$$
$$comp_+(ABOVE),\ subopt_{\tau_S}(ABOVE),\ cont_-(LEFT),$$
$$\alpha_-(RIGHT),\ \alpha_0(RIGHT),\ \alpha_{\tau_S}(RIGHT),\ \alpha_+(RIGHT),$$
$$\alpha_-(ABOVE),\ \alpha_0(ABOVE),\ \alpha_{\tau_S}(ABOVE),\ \alpha_+(ABOVE);$$

and only six degrees of truth:

$$d_0(RIGHT),\ d_{\tau_S}(RIGHT),\ d_{+\infty}(RIGHT),$$
$$d_0(ABOVE),\ d_{\tau_S}(ABOVE),\ d_{+\infty}(ABOVE).$$

It is convenient, at this point, to introduce the following notations and definitions. The *primary direction* of F_r^{AB} is the primitive direction δ_1 for which the amount of effective forces, $opt_{+\infty}(\delta_1)$, is maximum. The *secondary direction*, δ_2, is the neighboring direction of δ_1—*ABOVE* or *BELOW* if δ_1 is *RIGHT* or *LEFT*, *RIGHT* or *LEFT* if δ_1 is *ABOVE* or *BELOW*—with the larger amount of effective forces. The *ternary direction*, δ_3, is the neighboring direction of δ_2 different than δ_1. The remaining primitive direction is the *quaternary direction*, δ_4.

2.3. Histogram Normalization

It is reasonable to state that the identity $d^{AB}(RIGHT) = d^{t(A)t(B)}(RIGHT)$ should hold whenever *t* is a translation, a scaling or a reflection about a horizontal (left/right) axis; the identity $d^{AB}(RIGHT) = d^{t(A)t(B)}(ABOVE)$ should hold for any rotation through π/2; we should also have $d^{AB}(RIGHT) = d^{BA}(LEFT)$ (*A* is to the right of *B* as *B* is to the left of *A*). In [6], part of the problem was ignored, and part of it was addressed by multiplying the number of image object pairs for training. The approach proposed in the present paper avoids such a heavy manipulation and allows all the identities above to be perfectly retrieved. It is based on the following normalization procedure. First, the histogram F_r^{AB} is "rearranged" such that the primary direction δ_1 is *RIGHT* and the secondary direction δ_2 is *ABOVE*. The rearrangement consists in mirroring F_r^{AB} about the Y-axis and/or shifting it along the X-axis by a multiple of π/2. Then, the obtained histogram is divided by its mean (computed over an interval of length 2π) and by 2π such that the contradictory, compensatory and effective forces sum to 1. Let $\langle F_r^{AB}\rangle$ be the normalized histogram and let *t* be any translation, or scaling, or horizontal or vertical reflection, or rotation through a multiple of π/2. We have: $\langle F_r^{t(A)t(B)}\rangle = \langle F_r^{BA}\rangle = \langle F_r^{AB}\rangle$. This is illustrated by Fig. 4. The formal justification relies on geometric properties discussed in [7].

3. Neural Networks

Nine systems, denoted by 6-F_0, 6-F_2, 6-F_0/F_2, 17-F_0, 17-F_2, 17-F_0/F_2, 180-F_0, 180-F_2 and 180-F_0/F_2, were considered in our experiments. Each incorporates a

standard one hidden layer perceptron neural network (NN). In preliminary tests, we considered different training algorithms (resilient backpropagation learning, conjugate gradient learning with Fletcher-Reeves update, one step secant learning, etc.). We eventually adopted scaled conjugate gradient learning, which offered the best compromise between convergence time and Mean Square Error (MSE) performance. For each run, the weights were initialized randomly and the training stopped when the MSE was less than 0.0001 or the number of epochs reached 3500 (whichever occurred first). The NN input vectors are described in Section 3.1 and the output vectors in Section 3.2. All input and output values are between 0 and 1, and the number of hidden neurons for each NN is roughly equal to half the number of input neurons.

3.1. Input Vectors

For 180-F_0, 180-F_2 and 180-F_0/F_2, forces are computed in 180 evenly distributed directions and the computed values serve as NN inputs. The 180-F_0 NN is fed with the values $<F_0^{AB}>(2\pi i/180)$, where i belongs to {0;1;…;179}. Similarly, the 180-F_2 NN is fed with the 180 values that represent $<F_2^{AB}>$, and the 180-F_0/F_2 NN is fed with the 180 values that represent $<F_0^{AB}>$ and the 180 values that represent $<F_2^{AB}>_0$ (for a total of 360 input neurons). The histogram $<F_2^{AB}>_0$ differs from $<F_2^{AB}>$ in that the rearrangement procedure applied to F_2^{AB} is guided by F_0^{AB}. For instance, if F_0^{AB} had to be mirrored about the Y-axis and shifted along the X-axis by $-\pi/2$ in order to get $<F_0^{AB}>$, then F_2^{AB} is also mirrored about the Y-axis and shifted along the X-axis by $-\pi/2$. The idea is to avoid conflicting orderings of the two sets of 180 input values ($<F_2^{AB}>$ and $<F_2^{AB}>_0$ are not always equal). As shown in Section 2.2, seventeen forces and angles can be extracted from any force histogram. The 17-F_0 NN is fed with the 17 values extracted from $<F_0^{AB}>$, the 17-F_2 NN is fed with the 17 values extracted from $<F_2^{AB}>$, and the 17-F_0/F_2 NN is fed with the 17 values extracted from $<F_0^{AB}>$ and the 17 values extracted from $<F_2^{AB}>_0$ (34 input neurons). Note that the range for the angles $\alpha_{-}(RIGHT)$, $\alpha_0(RIGHT)$, etc., is linearly mapped from $[-\pi/2;\pi/2]$ to [0;1]. Besides forces and angles, six degrees of truth can also be extracted from any force histogram (Section 2.2). The 6-F_0 NN is fed with the values extracted from $<F_0^{AB}>$, the 6-F_2 NN is fed with the values extracted from $<F_2^{AB}>$, and the 6-F_0/F_2 NN is fed with the values extracted from $<F_0^{AB}>$ and the values extracted from $<F_2^{AB}>_0$.

3.2. Output Vectors

All NNs have 4 output neurons. Output j, where j belongs to {1,2,3,4}, is dedicated to the assessment of the proposition "A is in direction δ_j of B." The letters A and B denote the objects under consideration and δ_1, δ_2, δ_3 and δ_4 denote the primary, secondary, ternary and quaternary directions of either F_2^{AB} (for 180-F_2, 17-F_2 and 6-F_2) or F_0^{AB} (for all other systems). Note that δ_1, δ_2, δ_3 and δ_4 depend on A and B (since they are extracted from a non-normalized histogram). Therefore, for a given NN, the primitive direction a given output is associated with varies with the objects considered. This dynamic ordering of the output values matches the ordering of the input values induced by the histogram normalization.

4. Data

The training and testing sets of image object pairs are described in Section 4.1. The target outputs depend, of course, on the perception to be captured. Section 4.2 introduces the two virtual users whose perceptions we are interested in.

4.1. Input Images

Three sets of image object pairs are considered in our experiments: the *Animation*, the *Survey* and the *Power Plant* data sets. The *Animation* data set is composed of 2131 configurations (i.e., pairs of objects). It was introduced in [9] with the aim of evaluating a system for linguistic scene description. The animation, which lasts about three and a half minutes, is structured around thirty-five key configurations (Fig. 5). A large part of it is covered by six short movies that supplement the electronic version of [9]. The *Survey* data set (Fig. 6) was used in a survey conducted at the University of Missouri-Columbia in an attempt to find out how humans assess spatial relationships. It is composed of 69 configurations. Some of them look roughly similar to configurations from the *Animation* data set, as can be noticed from Fig. 5 and Fig. 6. Finally, the *Power Plant* data set (Fig. 7) is composed of 210 configurations extracted from a hand-segmented LADAR range image of a power plant near China Lake, CA. The image was supplied by the Naval Air Warfare Center. For all experiments, the even-numbered *Animation* frames constituted the training data set (TR) whereas the odd-numbered *Animation* frames and the *Survey* and *Power Plant* configurations constituted the testing data sets (T_A, T_S, T_P).

4.2. Target Outputs

All nine systems are involved in two sets of experiments. In the first set, we attempt to teach the systems—i.e, the corresponding NNs—*Sam*'s perception of the four primitive directional spatial relations. The target outputs are the degrees of truth $d^{AB}(RIGHT)$, $d^{AB}(ABOVE)$, $d^{AB}(LEFT)$ and $d^{AB}(BELOW)$ as estimated by *Sam*. *Sam* is a virtual user whose perception engine is the centroid (*Sam*troid) method, a well-known method based on the computation

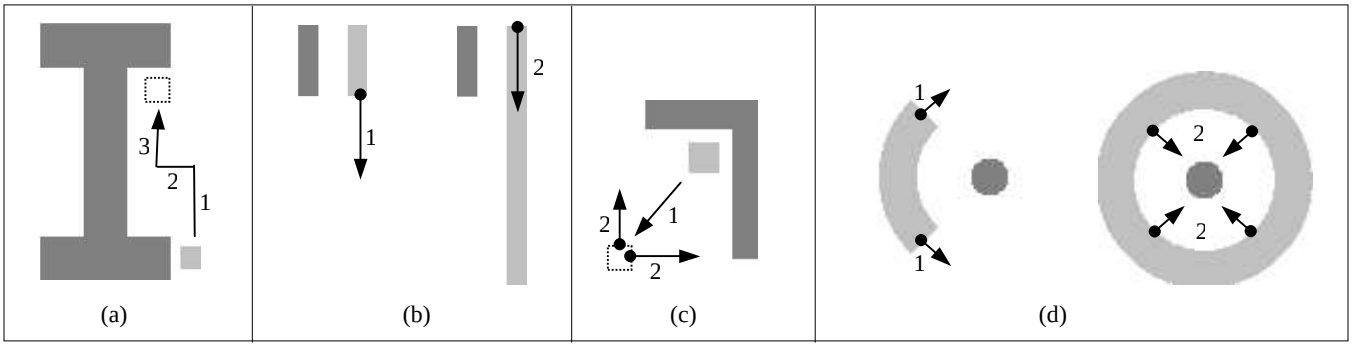


Fig. 5. *Animation* data set. Some of the thirty-five key configurations the animation is structured around. The arrows indicate the motion (→) or deformation (●→) applied to the argument object between two consecutive key configurations.

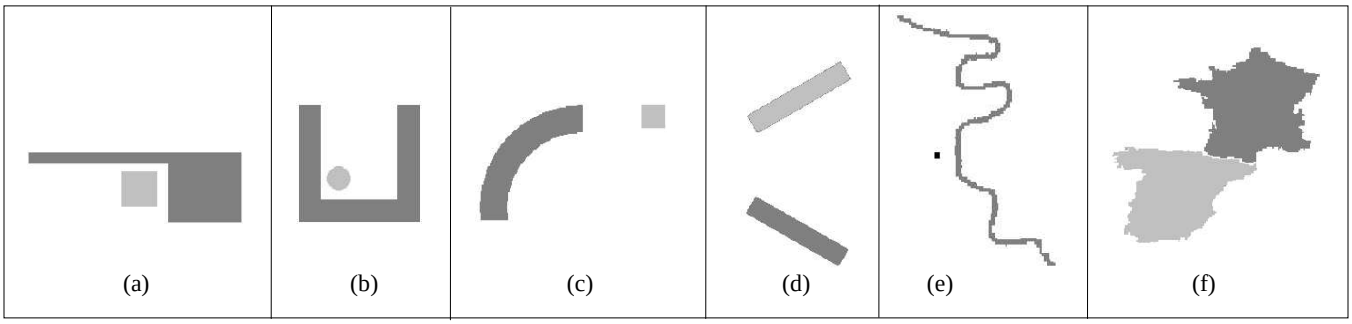


Fig. 6. *Survey* data set. Some of its sixty-nine configurations.

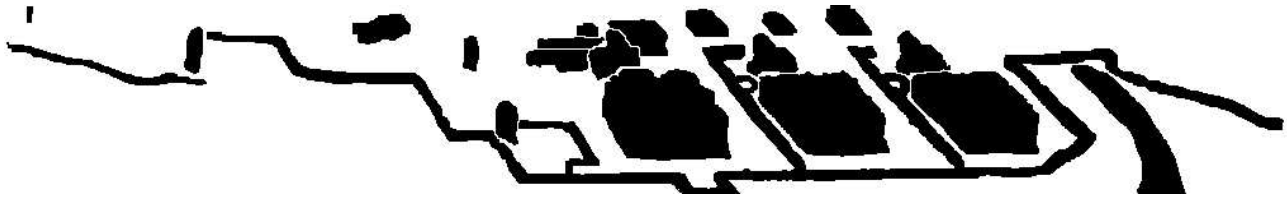


Fig. 7. *Power Plant* data set. The twenty-one objects in the scene define two hundred ten configurations (pairs of objects).

of the centroid of each object (see, e.g., [10]). In the second set of experiments, we attempt to teach the NNs *Angel*'s perception. The target outputs are generated by the aggregation method [10], which is based on the computation of angle (*Angel*) histograms. *Sam* and *Angel* have very different perceptions of the directional spatial relations. Contrary to flesh and blood users, they don't mind examining thousands of object pairs and are always consistent.

5. Results

Ninety instances of each system were created for *Sam*. In other words, each one of the nine systems (6-F_0, 6-F_2, 6-F_0/F_2, 17-F_0, 17-F_2, 17-F_0/F_2, 180-F_0, 180-F_2 and 180-F_0/F_2) was trained ninety different times with *Sam*'s outputs on the training data. Each group of ninety instances was then divided into ten clusters of nine instances each. For each pair *(A,B)* of objects and each primitive direction δ, a given cluster therefore outputs nine degrees of truth $d^{AB}(\delta)$. The median value defines the output of a new system instance: the cluster representative. The idea, of course, is to minimize the effect of random NN weight initialization. In the end, ten representatives of each system try to capture *Sam*'s perception of the four primitive directional spatial relations. Statistics on these ninety representatives (10×9) are given in Table 1. Ninety other representatives are defined in the same way with the aim of capturing *Angel*'s perception. The corresponding statistics are given in Table 2. The 180- systems give the lowest Mean Absolute Errors (MAEs) on the training data set TR and on the very similar testing data set T_A, but the highest errors on T_P and T_S. The -F_0/F_2 systems perform better than the -F_0 and -F_2 systems that use the same set of histogram features. Overall, the 17-F_0/F_2 system generalizes better than any other system and performs the best. Its representatives give errors way below 5%, whatever the data set and the virtual user. The resubstitution errors are less than 1%.

TABLE 1 — Capturing *Sam*'s perception: for each system, the table shows the min, mean and max Mean Absolute Errors (in %) of the ten representatives, over each data set.

		6-				17-				180-			
		TR	T_A	T_P	T_S	TR	T_A	T_P	T_S	TR	T_A	T_P	T_S
F_0	min	3.2	3.2	2.5	3.7	1.0	1.0	1.9	2.2	0.6	0.6	4.4	3.4
	mean	**3.4**	**3.4**	**3.0**	**3.9**	**1.1**	**1.1**	**2.1**	**2.4**	**0.7**	**0.7**	**5.0**	**3.8**
	max	3.8	3.8	4.7	4.2	1.2	1.2	2.6	2.7	0.7	0.7	5.5	4.2
F_2	min	5.4	5.3	5.4	5.7	2.7	2.7	4.6	5.2	0.6	0.6	4.9	3.7
	mean	**5.7**	**5.7**	**5.9**	**5.8**	**2.8**	**2.9**	**5.3**	**5.5**	**0.7**	**0.7**	**5.5**	**4.1**
	max	6.0	6.0	7.0	5.9	3.0	3.0	6.6	6.0	0.7	0.7	6.0	4.3
F_0 / F_2	min	2.2	2.2	2.1	3.4	0.5	0.6	1.8	2.4	0.6	0.6	4.9	3.7
	mean	**2.3**	**2.3**	**2.5**	**3.7**	**0.6**	**0.7**	**2.0**	**2.6**	**0.7**	**0.7**	**5.3**	**4.0**
	max	2.5	2.5	2.7	4.1	0.7	0.7	2.1	2.9	0.7	0.7	5.6	4.2

TABLE 2 — Capturing *Angel*'s perception: for each system, the table shows the min, mean and max Mean Absolute Errors (in %) of the ten representatives, over each data set.

		6-				17-				180-			
		TR	T_A	T_P	T_S	TR	T_A	T_P	T_S	TR	T_A	T_P	T_S
F_0	min	2.8	2.7	2.7	3.3	1.3	1.3	2.8	2.9	0.4	0.4	5.2	4.6
	mean	**3.1**	**3.1**	**3.2**	**3.6**	**1.4**	**1.4**	**3.2**	**3.0**	**0.4**	**0.5**	**5.7**	**4.9**
	max	3.4	3.4	4.2	3.7	1.5	1.6	3.8	3.3	0.5	0.5	6.3	5.6
F_2	min	6.4	6.4	6.1	5.7	3.3	3.3	5.9	6.8	0.4	0.4	6.3	5.2
	mean	**6.8**	**6.8**	**7.0**	**5.9**	**3.5**	**3.6**	**6.3**	**7.4**	**0. 5**	**0.5**	**6.7**	**5.5**
	max	7.2	7.2	8.6	6.1	4.0	4.1	6.6	8.1	0.5	0.5	7.4	6.0
F_0 / F_2	min	1.5	1.5	2.8	2.5	0.6	0.6	2.7	2.2	0.4	0.4	5.2	4.4
	mean	**1.6**	**1.6**	**3.1**	**2.8**	**0.6**	**0.7**	**2.8**	**2.4**	**0.4**	**0.4**	**5.6**	**4.9**
	max	1.8	1.8	3.3	3.0	0.7	0.8	3.1	2.6	0.5	0.5	5.9	5.2

6. Conclusion

We have designed systems that can learn individual perceptions of the four primitive directional spatial relations "to the *RIGHT* of," "*ABOVE*," "to the *LEFT* of " and "*BELOW* ". The proposed approach is based on the utilization of artificial Neural Networks (NNs) and the modeling of relative positions between image objects through histograms of forces. The NNs are fed by features extracted from the histograms and are trained to assess the directional relationships according to the target perception. Two types of histograms and three sets of histogram features have been considered. The histograms are histograms of constant forces (F_0) and histograms of gravitational forces (F_2). They have very different and interesting characteristics, as shown in previous work. The feature sets are respectively composed of degrees of truth (6 values), forces and angles (17), and histogram bins (180). A normalization procedure allows the systems to efficiently handle object pairs related through basic geometric transformations. Thanks to this procedure, important properties of directional relations are preserved, the number of needed training configurations is cut down, and the systems learn faster and generalize better. The system based on the computation of forces and angles extracted from both F_0 and F_2 histograms performed the best and was able to capture equally well two very different perceptions of the four primitive directional relationships. The results are encouraging and should now be validated by further experiments involving other perceptions and testing data sets. In addition, many issues still have to be addressed. For instance, training a system such as the ones described in this paper requires hundreds of configurations to be evaluated—a very long, boring task for the everyday user. Moreover, in practice, new configurations might need to be regularly added to the training data set, and some configurations might need to be reevaluated because of human inconsistencies. At this time, however, altering or adding to the knowledge already acquired cannot be done without completely retraining the system.

References

[1] J. Freeman, The modeling of spatial relations, *Computer Graphics and Image Processing*, *4*, 1975, 156-171.

[2] D.J. Peuquet, Z. Ci-Xiang, An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane, *Pattern Recognition*, *20*(1), 1987, 65-74.

[3] I. Bloch, Fuzzy relative position between objects in image processing: new definition and properties based on a morphological approach, *Int. J. of Uncertainty Fuzziness & Knowledge-Based Systems*, *7*(2), 1999, 99-133.

[4] P. Winston, Learning structural descriptions from examples, in *The psychology of computer vision* (P. Winston, ed.; McGraw-Hill, New York, 1975).

[5] S. Satoh, T. Satou, M. Sakauchi, One method of structural description rule extraction based on graphical and spatial relations, *Proc. 11th IAPR Int. Conf. on Pattern Recognition Methodology and Systems*, 1992, 281-284.

[6] X. Wang, J.M. Keller, Human-based spatial relationship generalization through neural fuzzy approaches, *Fuzzy Sets and Systems*, *101*(1), 1999, 5-20.

[7] P. Matsakis, L. Wendling, A new way to represent the relative position between areal objects, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *21*(7), 1999, 634-643.

[8] K. Miyajima, A. Ralescu, Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations, *Fuzzy Sets and Systems*, *65*(2/3), 1994, 225-236.

[9] P. Matsakis, J. Keller, L. Wendling, J. Marjamaa, O. Sjahputera, Linguistic description of relative positions of objects in images, *IEEE Trans. on Systems, Man, and Cybernetics*, *31*(4), 2001, 573-588.

[10] R. Krishnapuram, J.M. Keller, Y. Ma, Quantitative analysis of properties and spatial relations of fuzzy image regions, *IEEE Trans. on Fuzzy Systems*, *1*(3), 1993, 222-233.