

# Indoor Environment Classification and Perceptual Matching

**Fiara Pirri**

Dipartimento di Informatica e Sistemistica  
via Salaria 113, 00198, Roma, Italy  
pirri@dis.uniroma1.it

## Abstract

In this paper we present an approach to indoor classification that presupposes a certain amount of prior information in terms of statistical information about possible interdependencies among objects and locations. A preselective perception process (that here is only hinted), using a database of textures, and built using the energy value computed with a tree-structured wavelet transform, selects regions in the image and, according to the database, builds an observation state including also a saliency map of the features of the image. The process of delivering this information is interleaved with a process that using a database of interdependencies between objects and locations, mimicking a memory, forms an hypothesis about the current location. We show that the process of hypothesis formation converges under specific constraints.

## Introduction

The observation process leading from attention to awareness of the surrounding environment deals with the ability of selecting task-relevant stimuli, while excluding irrelevant perception. This observation process, is preliminary to recognition, as it determines both the contextual conditions and the necessary stimuli load (see (Rees, Frith, & Lavie 1997; Lavie 1995)) for recognition. In this paper we report on a work done for the purpose of studying the observation process that leads from random visual perception to selective perception and attention. Without attention is not even possible to address the recognition problem, unless the system has already been driven toward a specific object. Attention is a well studied problem both in psychology and in computer vision. In the earliest approaches (see (Koch & Ullman 1985; Darrell & Pentland 1995)) attention was more concerned with the relevance of objects. For example (Koch & Ullman 1985) introduced the idea of a saliency map to encode the saliency of objects in the visual environment: the maximum value of the saliency map represents the most meaningful location in the acquired image. On the other hand the most recent approaches are more concerned with biologically plausible computational models for attention (Nikolaidis, Pitas, & Mallot 2000; Itti & Koch 1998; Itti, Koch, & Niebur 1998; Itti & Koch 2000) and sensors updating (Thrun 2002) to face the drawbacks of task

dependent processes in which the appearance of objects might suddenly change, because the observer's viewpoint had changed. Most of these approaches rely on specific methodologies in features extraction.

It is not yet clear how prior knowledge can affect attention, despite several experiments have reported that memory of prior perceptual experience plays an important role. However from a machine point of view any possible, and even simplest, computation requires prior knowledge of the operation to be performed, therefore the problem of prior knowledge is ill posed. There is however a question concerning the amount of constraints prior knowledge would impose on the selection of a specific feature better than another, if the prior knowledge is too constraining, especially in the early stage of perception, some important cue might be lost. For example if during the phase of selecting meaningful visual field, is it imposed that an object "cannot" have certain dimensions, or "cannot" be in a certain position, this constraint might induce the rejection of some information that could be trusted: for example a table turned upside down in a damaged environment. At the same time it is not possible to form an hypothesis about a current percept if the perceptual process is not provided with a basic ontology, i.e. a representation model handling the perceptual matching. The ontology is in turn influenced by the perceptual experience, or what in (Hurley 1998) is called perspectival self-consciousness ; however it is still controversial whether the perceiver needs, in order to have the experience, possess the concepts necessary to capture in thought the ways of filling out the space that would make the experience veridical (e.g. see (Noe, Pessoa, & Thompson 2000)).

In our approach, we pair a given prior knowledge with a class of statistical informations, that should play the role of the *perceptual experience*. In fact, we think that differently from the laws ruling actions and world relations and truth, the perceptual experience is statistical and essentially uncertain, therefore cannot be prescribe by rules of knowledge.

Still, determining a correct balance on both the amount of knowledge that can be a priori settled and its underlying ontology, is quite hard and needs a great amount of experimentation.

On the other hand a crucial component in perceptual matching is how the current task, as a focused stimulus, can affect both attention and early interpretation of elements in

the environment. For example (Lavie 1995) reports that distractors intervene in situation of low perceptual load, when just a few relevant stimuli are presented.

The problem here is to well understand the interplay between the stimulus and the expectation of what could be perceived. If the stimuli are “looking for something” then the stimuli are, indeed, the constraints we were discussing before. Since to look for something the target must be already known.

So it seems that knowledge of the task is helpful to keep attention and to disambiguate the perceptual information. A great amount of work has been done in this direction, from the early work of (Levesque & Scherl 1993; Bacchus, Halpern, & Levesque 1999; Golden & Weld 1996) to the more recent works of (Reiter 2001b; Shanahan 2002; 2004). From these and other works, it seems that there is some agreement about the way perception can affect reasoning and vice versa. A common result seems to be that the perceptual reasoning process is intrinsically hypothetical and abductive, as it is based upon beliefs and hypothesis formation imposing a continuum back and forth among focalization-attention-interpretation-action. The process interleaving these different activities serves to adjust parameters which is the well known hard problem in perception. Therefore a possible step toward a meaningful approach to perception, is to foresee a process that is back and forth and thus parameters might be adjusted according to the context, by this process itself.

To introduce our approach to this back and forth process using a statistical memory, playing the role of the perceiver perceptual experience, in this paper we present some ideas concerning a methodology to face the problem of indoor classification. A preselective perception process (that here is only hinted), performs image segmentation using as features arrays the energy value computed with a tree-structured wavelet transform (Chang & Kuo 1992). The resulting segmentation is then compared to a texture database. The output of this analysis is an observation state that is further processed by probabilistic updating, that uses the statistical memory, mimicking the perceptual experience, to output an hypothesis. We show that the process of hypothesis formation converges under specific constraints.

## Basic ontology

To represent and manage the reasoning process concerned with perception, so far we have been considering the language of the Situation Calculus (Reiter 2001a; Pirri & Reiter 1999) and some of its extensions (Pirri & Finzi 1999; Pirri & Romano 2003; 2002). In the attentional processing for environment localization awareness, to be described here, two reasoning components are involved: the state acquisition, or how the hypotheses computed according to the estimation process, that we shall illustrate in the sequel, are managed, and the back and forth communication to establish a selective perception. This process would then lead to a further one dealing with regions clustering and hence with recognition. The cycle of this specific component of the reasoning process is very simple: the information acquired and

elaborated by extracting salient image components is transformed into simple terms of the language, each term is a name for a category of objects, the knowledge about which it is needed to localize the current environment; note that a parameter (or name or denotation) serves the only purpose of indicating “there could be a desk”, and not the stronger notion “this is a desk”.

We limit here the discussion to the notion of state and observation state. The language we consider here is composed by a finite set of parameters  $\mathcal{A}_{par} = \{r_1, r_2, \dots, r_n, \dots\} \cup \{A_1, A_2, \dots, A_n\}$ , we call this set the *denotations* for all the elements that can be indicated in the environment. Note that these parameters are also handled to classify the textures database, for extracting the texture features. A second set is used for the classification purpose. The language also includes situations, and a countable set of variables to denote the parameters itself, and to denote the probabilities associated with each location. Here, in this context, we consider situations to be terms of the language constructed through observations. The initial situation is  $S_0$ , and in particular, a situation is a history of all the observation actions executed in a while. At the end of the sequence the database is progressed, and time is updated. In particular, if  $s_i$  is a situation, and  $\langle b_1, \dots, b_n, t \rangle$  are the data collected in the observation, then there is a (suitably defined in the underlining logic) situation  $s_j$ , such that  $s_j = observe(b_1, \dots, b_n, t) \circ s_i$ , where  $observe()$  is an action, taking as arguments the parameter of the observation, and  $\circ$  is a binary function for composing terms of the language, and satisfying the condition:

$$a \circ s \geq s' \equiv s \geq s' \text{ and } a = a' \wedge s = s' \equiv a \circ s = a' \circ s' \quad (1)$$

A situation can be interpreted as an index of the current state, and time is added under the conditions described in (Reiter 2001a; Pirri & Reiter 2000). We use subscripts to denote a specific situation, and in particular we assume an implicit ordering, so that  $s_i > s_j$  if  $i > j$ . For all what concerns a basic theory of actions we refer the interested reader to (Reiter 2001a).

In the current simplified language probabilities are terms of the language that can take as arguments situations, therefore we denote with  $P(r_i | X = A_j, s)$  the likelihood ratio, of the element  $r_i$  (i.e. an object in the domain), given that  $X = A_i$  in the situation  $s$ ; here  $s$  without subscript is a variable ranging over situations. By an estimation  $f$  we intend a function that behaves like a probability, i.e. it satisfy the probability axioms, however is generated by a subjective belief measure induced by observations.

We assume two concurrent processes in the Situation Calculus (Reiter 2001a; Pirri & Reiter 2000). The process concerning perceptual actions  $observe()$ , having only the effect of delivering an observation state, that we shall describe in the sequel, interleaves with the so called *effect producing actions*; in particular observation actions can be seen as special natural actions. The only further constraint needed here is that while executing observations, the concurrent effect producing actions should be restrained to some specific set, that would not affect the observations, for example the system should not change abruptly location while is still trying to form an hypothesis about where it currently is. To this

end we introduce a *flag* into the observation state, that notify when the hypothesis is formed, and a new exploration can be initiated. To comply with these constraints we introduce the two following concepts. An operator combining observation states, and saying that two pieces of information can be combined only if they have been delivered by an observation state, as follows:

$$\begin{aligned} A(a \circ s) \oplus B(\text{observe}(B, \dots, t) \circ s') = \\ a \circ \text{observe}(B, \dots, t) \circ s' \equiv \\ a = \text{observe}(A, \dots, t') \wedge t' > t \wedge \\ s = \text{observe}(B, \dots, t) \circ s' \end{aligned} \quad (2)$$

Here the ellipsis is about the parameters concerning the observation state. The above sentence means that the fluent  $A()$ , can be combined with the fluent  $B()$ , just in case they are part of the same observation sequence. The operator  $\oplus$  is non commutative.

Given the above definition, then we can specify how to restrain the set of actions interleaving with the observation sequence. Therefore let  $A_1 \dots, A_m$  be the list of admissible actions while observing, then the constraint on the interleaving between observations and actions can be expressed as follows:

$$\begin{aligned} \text{Poss}(\text{interleave}(a \circ s, B(\text{observe}(B, \dots, t) \circ s'))) \equiv \\ a \circ s = A(a \circ s) \oplus B(\text{observe}(B, \dots, t) \circ s') \vee \\ \bigvee_i a = A_i \end{aligned} \quad (3)$$

A very successful approach for estimating and explaining observations, generally used in speech recognition, is the Hidden Markov Model approach (see (Rabiner & Juang 1986)), that has also been extensively used in computer vision, both in image classification (see (Li, Najmi, & Gray 1999)) and also to capture the statistical structure of signals and images (see (Romberg, Choi, & Baraniuk 1999)). HMM have been further extended to multi-layered models to cope with processes in which variations happen at different scales. Because we were modeling hidden states (i.e. the current hypothesis), through observations, and because of the interest and the success of such an approach, we have been trying to adapt an extension of the HMM, together with the Viterbi algorithm (Viterbi 1967), to model the belief updating, through observation. Unfortunately it turned out that it was not appropriate and we could not suitably deal with the needed updating, according to the flush of information received by the observations. The main drawback being the stationarity hypothesis of the HMM, and the independence of each observation. HMM are stationary, which means that the transition model from a state to the other is predetermined a priori; so there is no way, once a certain amount of knowledge, e.g. about a given location or situation, is acquired, to change the transition from one state and the successive. While we need to determine the transition dynamically, depending on the current amount of information.

### The observation state

We consider an observation as the result of an early image processing, in which a set of images - acquired at a specified frame rate - are taken at a given direction. In terms of the

pan-tilt head actions, a direction is specified by the following simple actions *left, straight-left, straight, straight-right, right, straight-up, straight-down, up-left, up-right*, according to a precise pan-tilt angle-interval of rotation. Once a specific area has been analyzed an observation is returned, the direction observed is inhibited and the focus is shifted to the next direction. If the observation is too noisy and no relevant data is returned then the observation is repeated by a mild angular shift (note that this aspect that we do not treat here, concerns especially the navigation control in a small area navigation).

The early image processing based on texture classification (see Section ) returns a set of hypotheses concerning the presence of regions that could belong to specific objects, together with their position, with respect to the system coordinates, their bounding box, and the action that led to select these data from the acquired image (e.g. *straight-left*). This set of data, forms the *observation state*:

$$\begin{aligned} O_t = \\ \langle \langle f(R_1, \delta(C_1), B_i, p_1) = r_{i_1}, \dots, \\ f(R_k, \delta(C_k), B_k, p_k) = r_{i_k}, t, \alpha \rangle \rangle \end{aligned} \quad (4)$$

Here  $R_i$  is the processed region, with  $\delta_i$  its position in terms of the  $x, y, z$  coordinates of  $C_i$  (the position is acquired by a range sensor, namely a 'telemeter', coupled with the camera), the center of the region,  $B_i$  its bounding box,  $r_{i_j}$  is the  $j$ -th object, in the list of denotations, that is supposed to contain or to be contained in the identified region. And finally  $p_i$  is the saliency of the region. The notion of saliency, early introduced in (Baluja & Pommerleau 1995), in connection with the definition of a saliency map, is computed using a weighting function balancing context free and context dependent analysis (Pirrone 2003), that combines several image extracted features, associated with a single sub block (a cluster obtained from image tessellation by suitable region growing). The formula for saliency (that we do not report here) shows that among the significant features is symmetry. The saliency is a scalar reporting the relative importance of the specified element in the context of the observation, and in general it should indicate which regions of the input retina are important in the preselective perception process. The observation state is, thus, a saliency map of the visual field in the direction of the specific action performed by the pan-tilt head. The map is used by the hypothesis formation process to guide the selection of attended locations.

An example of an observation state is the following:

$$\begin{aligned} O_t = \\ \langle \langle f(R_1, \delta(65.3, 42, 54), \langle 60, 85, 110 \rangle, 0.3) = \text{desk}, \\ f(R_2, \delta(80, 140, 8), \langle 40, 85, 70 \rangle, 0.5) = \text{curtain}), \\ t, \text{straight} \rangle \rangle \end{aligned} \quad (5)$$

Note that the information collected in the information state, even if it is gathered by different images in a given direction, returns a region that could be just a small component of the object or, vice versa, that is contained in an object. These aspects that would eventually concern object recognition, are not relevant here. The only problem in the attentional context is to filter irrelevant perceptions for further image elaboration, by trying to generate an awareness about the loca-








Texture	Furniture			Wall			Pavement	...
	Closet	Bed	Bookcase	Curtain	Wallpaper	Carpet	Brick	Tile
	0.02	0.13	0	0.01	0.15	0.12	0	0.04
	0	0.1	0	0.2	0.1	0.1	0	0
	0	0.02	0.1	0.1	0	0.01	0.02	0.27
	0.14	0.02	0	0	0	0	0.01	0
	0	0	0	0	0.12	0.1	0.37	0
	0.01	0	0	0.01	0.01	0	0.12	0.23
	0	0.01	0.17	0.02	0	0.01	0.01	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Example from the Textures DB

tion, which is the precondition for attention, and hence for grouping and recognition. In the next sections we describe how the observation state is used to form an hypothesis that would guide the selection process.

### Assessing a dataset for prior information

As a data set for assessing the indoor classification theory we have chosen images of different home locations taken from several real estate web pages, and digest. The distribution of each ambient location is drawn with respect to 900 pictures of home interiors collected according to a real estate stratification of home typologies, made by ranking the ratios cost supply, e.g a supply of 100 at a cost of 20, and a supply of 10 at cost 200, with respect of several home features: living/dining room, 2 bedrooms, etc.

The need for the dataset is twofold:

1. Define a database of textures specialized on home interiors (in order not to resort to the Brodatz texture database, or others not specifically oriented). The dataset partitions the texture space into several subregions, each representing a cluster of visually similar patterns.
2. Define a prior knowledge on home features and conditional probability distributions of home artifacts. The likelihood ratios (see Table 2) play the role of prior memories of particular sensory objects, where the sensory perceptual episode is interpreted as the association object-location. Unfortunately being the process circular these associations are provided by a dataset instead of from direct learning.

The textures database, can be accessed by an index which is the texture itself, i.e. the energy value returned by applying a tree-structured wavelet transform (TSWT) to the texture (see (Chang & Kuo 1993)). For each texture, we keep a confidence vector with one element for each indoor object (see Table 1), denoted in the set of parameters  $\mathcal{A}_{par}$ . These

values interpret an object belonging probability distribution, i.e the frequency of objects with that specific texture, w.r.t. the acquired dataset.

To direct the hypothesis update, we are given a table drawing the probability distribution of home features (bathrooms, office, bedrooms, living/dining room, etc.) together with the probability of observing a specific object given the location, in the initial situation that we denote  $S_0$ . Therefore, looking at Table 2 each place has a distribution according to the strongly biased principle that the sample is representative of any home typology. Given the data, the probability model fits a normal distribution, both for each variable  $x_i$  ranging on the conditional distribution of the indoor objects, and for the variable  $X$  ranging over the locations. For both the databases we have not added noise, instead we have corrected the data with an empirical bias obtained by the consideration that the data from the real estate are the “best” available.

### Early selective perception based on visual features

The early selective perception has the role of shaping the visual information into possible fields of interest toward which directing the attention. It is therefore the first step in the perceptual awareness process. Detection of regions of interest is both context free (bottom up), and context dependent (top down). Context free analysis concerns the computation of a set of simple features, like color, motion, texture, and specific properties like symmetry. A context free image analysis involves only information related to the image raw data. In particular, in our work we have chosen to use color and intensity as image features and to characterize the acquired image in terms of them. Therefore, the image format more suitable for this kind of analysis is Hue, saturation, and brightness, because it incorporates di-

	X= BedR	X= Office	X= Kitchen	X= BathR	X= LivingR	X= Entr
$f(X, s)$						
$s_0$	0.29	0.12	0.25	0.29	0.03	0.02
$s_1$	0.10649	0.19931	0.20984	0.45799	0.02321	0.00314
$s_2$	0.07187	0.53079	0.2049	0.15456	0.03681	0.00106
$s_3$	0.03267	0.68427	0.23197	0.02727	0.02316	0.00064
$P(r_i X, s_0)$						
desk	0.1	0.5	0.1	0	0.2	0.1
window	0.21	0.27	0.3	0.1	0.1	0.02
stove	0	0.15	0.83	0	0.01	0.01
monitor	0.09	0.4	0.11	0	0.39	0.01
bed	0.61	0.2	0	0	0.18	0.01
bookcase	0.11	0.57	0.11	0.1	0.1	0.01
door	0.1	0.21	0.23	0.05	0.21	0.2
book	0.2	0.45	0.1	0.05	0.1	0.1
basin	0.09	0.01	0.43	0.46	0	0.01
sofa	0.1	0.2	0.01	0.01	0.5	0.18
	r1	p1	r2	p2	r3	p3
$O(s_0)$	desk	0.3	monitor	0.47	basin	0.58
$O(s_1)$	window	0.24	bookcase	0.32	sofa	0.22
$O(s_2)$	book	0.32	door	0.34	book	0.27

Table 2: A table describing a statistical memory including a probability distribution of likelihoods of home locations with respect to objects, the prior distribution of home locations, and the updates following a sequence of observations  $O_{s_0}, O_{s_1}, O_{s_2}$ .

rectly the features required (see (Pirrone 2003)). The context dependent analysis is based on the successful ideas of using textures to classify the image content, and in particular the wavelet transforms that measure the image properties over domains of varying sizes. The wavelet transform have been proved to be very useful for texture analysis (see e.g (Romberg, Choi, & Baraniuk 1999; Laine & Fan 1993; Chang & Kuo 1992; 1993)) as they cope with the need to describe the texture accurately by capturing its underlying structure, and the region boundaries. We use use them, in texture analysis to classify those areas selected in advance by the bottom-up approach, and retrieve the correct information from the database, in order to label each sub block in which the image is initially subdivided, with an object belonging probability distribution. The regions perceived in the specific image field are clustered according to a probabilistic region growing algorithm, with the K-mean. Finally, following well known psychological principles ( see (Siegel, Koerding, & Koenig 2000)), we have provided a weighted combination of the clustered regions, and the region saliency is the result of such a combination exploiting the following major factors: the entropy of the features, and the correlation among the resulting areas obtained from the context free and context dependent image analysis. The early pre-selective perceptual process is not further investigated here.

### Disclosing beliefs over a sequence of observations

Given a sequence of observations  $O_{s_1}, \dots, O_{s_n}$ , the first step toward the agent awareness is to recognize the loca-

tion it is stepping into. In our case, each observation should be seen as a sort of belief update, a further acquisition of awareness, despite all the possible errors drawn by the pre-selective perception delivering the observation itself. In other words, we want the transition, between a belief state and the successive, to be determined by the observation, according to the objects conditional distribution functions and the location distribution, given in the initial state.

For the reader convenience we summarize the notation in the following:

1.  $\mathbb{L} = \{A_i, \dots, A_m\}$ , with  $|\mathbb{L}| = M$ , are all the locations considered, with each  $A_i$  denoting a specific location, e.g. *bedroom, bathroom*.
2.  $\mathbb{E} = \{r_1, \dots, r_n\}$ , with  $|\mathbb{E}| = N$ , are all the possible elements, i.e. the objects in the vocabulary, denoting elements considered from the early texture analysis. E.g. *table, bookcase*, etc.
3.  $\mathcal{E}(s) = \{r_1, \dots, r_k\} \subset \mathbb{E}$ , denotes the set of regions that could be associated with objects, individuated in situation  $s$ .
4.  $O(s_1), \dots, O(s_n)$  is a sequence of observations, to each situation is associated an observation.

For example, as indicated in Table 2 there are six locations (namely bedroom, office, bathroom, kitchen, living-room, entrance), and the objects considered are ten. The  $A_i$  locations constitute the entire vocabulary of places that can be identified by the cognitive system, and analogously the set of  $r_i$  objects constitutes the whole set of parameters indicating an object in the home environment. For each  $r_i \in \mathbb{E}$  we

are given a conditional probability distribution, given that the state of nature is  $A_i$  and for the  $A_i$  we are given a prior probability distribution.

Given an observation  $O_s$ , in state  $s$ , and a set  $\mathcal{E}(s) = \{r_1, \dots, r_k\}$ , of elements reported in the observation state  $s$ , we want to determine both the best explanation for the current observation, that is the set of possible locations that explain where the system is, and an updating of the probabilities, in order to narrow the set of possible explanations. In other words we want that a sequence of observations, that are supposed to be accomplished in the same location, converges toward a minimal set of hypotheses, or even a single hypothesis, that would be available at the end of the observation sequence. The observation state will obviously be noisy, delivering information probably about non-existing regions, or definitely wrong information, however on the long run some reasonable information will be got. However not all the information can be accepted.

**1 Observation conditions** *We distinguish two cases:*

1. *The observation state is empty, nothing can be concluded, and thus  $s_{i+1} = s_i$ . Furthermore if  $\mathcal{E}(s_i) = \emptyset$  then we let  $\mathcal{E}(s_i) = \{nil\}$ , with  $f(nil, s_i) = \epsilon$ .*
2. *The observation state reports information about all the elements of the parameters set, in this case we consider the information inconsistent.*

Since the observation state includes a saliency map of the interesting region found, each element  $r_i$  is paired with its saliency, that we denote with  $p_i$ , observe that  $0 \leq p_i < 1$ , and  $\sum_i p_i = 1$ . The saliency, as noted above, is obtained from the probability distribution of the textures, that led to the decision of choosing  $r_i$ , and from other relevant estimations, such as the saliency of the region grown around the sub blocks chosen, and the symmetry of the region obtained.

The simplest method to get the hypothesis (i.e. the location) that best explains the observation, is the maximum a posteriori hypothesis, that is, for each  $r_i \in \mathcal{E}(s)$ , the hypothesis MAP:

$$P(X = A_j | r_i, s) = \frac{P(r_i | X = A_j, s)P(X = A_j, s)}{\sum_j P(r_i | X = A_j, s)P(X = A_j, s)} \quad r_i \in \mathcal{E} \quad (6)$$

and

$$H_{(A, r_i)} = \operatorname{argmax}_j P(X = A_j | r_i, s) \quad r_i \in \mathcal{E}$$

Here  $H_{A, r_i}$  is the hypothesis  $A$  that best explains the observation  $r_i$ . The drawback with this approach, is that it disregards the amount of information carried in from the observation itself, that is, the saliency and the congruence of the group of objects, and their relevance. Furthermore, by collecting the *MAP* hypothesis at each step  $s_i$ , we have no way of updating the probabilities, while we want to have an estimation of the current location, that could constitutes the basis for interpreting the next observation. In order to maximize the information received in the observation state we need to consider whether the element that is supposed to be “there” is relevant to a specific location, and sign it positive for this location, if this is the case. We state therefore the following notion of *relevance to an observation*. Given that

$\mathcal{E}(s) = \{r_1, \dots, r_k\}$  is the set of elements observed, in state  $s$ , the set of relevant observations  $\mathcal{R}^+(X, s)$ , for  $X \in \mathbb{L}$ , is defined as follows:

$$\mathcal{R}^+(X, s) = \{(X, y), X \in \mathbb{L}, y \in \mathcal{E}(s) \mid \neg \exists X'. X' \neq X \wedge P(y | X', S_0) > P(y | X, S_0)\} \quad (7)$$

An observed element  $y \in \mathcal{E}(s)$  is *relevant to  $A_j$*  in  $s$  if  $\langle X, y \rangle \in \mathcal{R}^+(X, s)$ .

Note that  $\forall r \in \mathcal{E}(s)$ , there is only one  $A_j$ , s.t.  $r$  is relevant to  $A_j$ , on the other hand each  $A_j$  can have one or more element  $r \in \mathcal{E}(s)$  which are relevant to it. The set of all the relevant pairs for an observation in  $s$  is the set

$$\mathcal{R}(s) = \bigcup \mathcal{R}^+(X, s)$$

So for example, suppose  $\mathcal{E}(s) = \{desk, monitor, basin\}$ , then according to the above definition  $\mathcal{R}^+(Office, s) = \{\langle Office, desk \rangle, \langle Office, monitor \rangle\}$ , and  $\mathcal{R}^+(Bathroom, s) = \{\langle Bathroom, basin \rangle\}$ . We shall use these positive observations to weight the initial probability distribution and build an hypothesis in the form of a probability estimation, that would then affect the interpretation of the next observations. In particular if  $\mathcal{R}^+(X, s_1)$  is the set of pairs given above, and all it is known about the current situation  $s_1$ , as a result of the first observation  $O_{S_0}$ , achieved in situation  $S_0$ , is that there might be a desk, a monitor, and a basin, then the best hypotheses are  $\{Office, Bathroom\}$ , note that the MAP hypothesis is *Bathroom*. On the other hand we do not want to withdraw the other hypotheses, because some hypothesis could come up in a further observation. To cope with these problems, we introduce a specific estimation of the state achieved in terms of values, obtained by the likelihood ratios and priors available in  $S_0$  (e.g. see Table 2). Let  $\mathcal{E}^-(X, s)$  be:

$$\frac{|\{\langle X, y \rangle \mid \langle X, y \rangle \notin \mathcal{R}(s)\}|}{N} \quad (8)$$

Where  $N$  is the cardinality of  $\mathbb{E}$  (i.e. the names for all the parameters denoting the objects that are looked for in the indoor environment). We define the following function  $\gamma(X, s)$

$$\gamma(X, s) = \frac{e^{\beta(X, s)}}{1 - \alpha(X, s)} \quad (9)$$

Here  $\alpha(X, s) = \prod_i p_i$ , with  $p_i$  the saliency of the element  $r_i \in \mathcal{E}(s)$ ,  $0 < p_i < 1$  and  $\beta$  is the entropy of the information contributed by the observation, i.e.  $\beta = -\mathcal{E}^-(X, s) \ln \mathcal{E}^-(X, s)$

The entropy about “what has not been observed” is meaningful, as in fact if there are more than  $N/2$  positive perceptions about the same location  $A_i$ , implying that  $\mathcal{E}^-(A_i, s) > 1/2$ , then the information can be disguising, because it would mean that several perceptions are either irrelevant or incorrect. Note also that while for a given  $A_i$ ,  $\mathcal{R}^+(A_i, s)$  can be empty,  $\mathcal{E}^-(X, s)$  can never be 0, otherwise the observation state is considered inconsistent (therefore we stipulate that when  $\mathcal{E}^-(X, s) = 0$ ,  $\ln(0) = 0$ ). Now, if for some  $A_i$

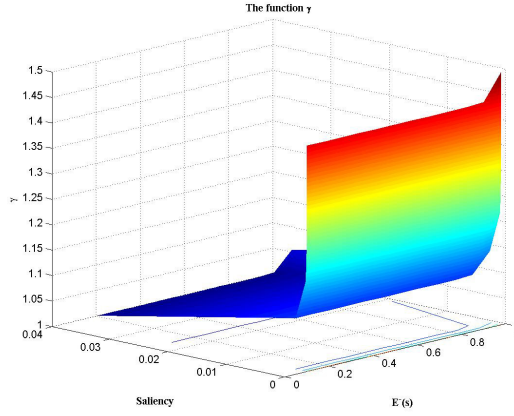


Figure 1: The function  $\gamma$

no positive element has been observed then  $e^\beta = 1$ . Moreover, because  $A_i$  has no relevant element, then  $p_{r_i} = 0$ , for each  $r_i$ , it follows that for such an  $A_i$ ,  $\gamma(A_i, s)$  is 1, which implies that the estimation will not alter the posterior ratios for such an  $A_i$ . On the other hand  $1 - \alpha(X, s_i)$  ensures that if the saliency is high then  $\gamma$  will increase. Therefore these considerations lead to the following lemma:

**Lemma 1** *Given the observation conditions (see Conditions 1), for all  $s_i$  and for all  $A_j$ ,  $\gamma(X, s_i) \geq 1$ . In particular,  $1 \leq e^\beta \leq \sqrt{2}$  and  $0 \leq \alpha < 1$ .*

A plot of the function  $\gamma$ , over the set  $\mathcal{E}^-(s_i)$  increasing from 0.1 to 1 and were the saliency of the observed elements is given a normal distribution, is shown in Figure 1.

To better illustrate the role of  $\gamma$ , consider the following example.

Let  $s = S_0$  and suppose that the information available in  $S_0$  is that illustrated in Table 2. Suppose that the observation state  $O_{S_0}$  includes the following information:

$$\{\langle desk, 0.3 \rangle, \langle basin, 0.58 \rangle, \langle monitor, 0.47 \rangle\}$$

Here the above numbers concern the saliency of each object, computed in the pre-selective attention step, as briefly explained in Section . Then  $s_1$  is the current state.  $R^+(Office, s_1) = \{\langle Office, desk \rangle, \langle Office, monitor \rangle\}$ ,  $R^+(Bathroom, s_1) = \{\langle Bathroom, basin \rangle\}$ .

For any other location  $X$ ,  $R^+(X, s_1)$  is empty.  $\mathcal{E}^-(Office, s_1) = 8/10$ ,  $\mathcal{E}^-(Bathroom, s_1) = 9/10$ , and for all other locations,  $\mathcal{E}^-(X, s_1) = 1$ . To compute  $\gamma(X, s)$  we need only to consider  $\gamma(Office, s_1)$  and  $\gamma(Bathroom, s_1)$ , because as we shall see, for all the others locations for which  $R^+(X, s_1)$  is empty,  $\gamma(X, s_1)$  is just 1. Now,  $\beta(Office, s_1) = -8/10 \ln 8/10 = 0.17851$ , and  $\beta(Bathroom, s_1) = -9/10 \ln 9/10 = 0.09482$ , for any other location  $X$  we have  $\beta(X, s_1) = -10/10 \ln 10/10 = 0$ . On the other hand,  $\alpha(Office, s_1) = 1 - (p_{desk} \times p_{monitor}) = 0.859$  and  $\alpha(Bathroom, s_1) = 1 - (p_{basin}) = 0.42$ , for any other location  $X$ ,  $\alpha(X, s_1) = 1 - 0 = 1$ . Therefore  $\gamma(Office, s_1) = e^{0.17851}/0.859 = 1.39167$  and

$\gamma(Bathroom, s_1) = e^{0.09482}/0.42 = 2.61778$ , for any other location  $X$ ,  $\gamma(X, s_1) = e^0/1 = 1$ .

Note also that  $\gamma(Bathroom, s_1) > \gamma(Office, s_1)$ , meaning that in this case  $\gamma$  captures the MAP hypothesis, and the reason why is so, it is because the saliency of *basin* is quite high.

The role of  $\gamma$  is to increase the posterior ratios for all locations involved by the perceptual matching between the observation and the conditional distributions stored in the statistical memory (i.e. those in  $\mathcal{R}(s)$ , for the current  $s$ , and for which  $\gamma > 1$ ), leaving unchanged the others. Note that we have defined  $\alpha(X, s) = \prod_j p_j$ , where  $p_j$  is the saliency; this choice makes the updating change very slow, because of the product. An alternative solution is to substitute  $\alpha$  with a Gaussian, taking  $\sigma$  to be the variance of the saliency of the observed  $r_i$  relevant to the specific  $A_j$ , and  $\mu$  their mean, but in the context of this example does not make much of sense. The above considerations lead to the following lemma:

**Lemma 2** *If  $\langle A_j, x \rangle \notin \mathcal{R}(s_i)$  then  $\gamma(A_j, s_i) = 1$ .*

*Proof.* If  $\langle A_j, x \rangle \notin \mathcal{R}(s_i)$  then  $\mathcal{E}^-(A_j, s_i) = N$  and  $\alpha(A_j, s_i) = 0$ , thus:

$$\gamma(A_j, s_i) = \frac{e^{-(1 \ln 1)}}{1 - 0} = \frac{1}{1} = 1$$

□

In what follows we make some implicit assumption on the distribution model of the class conditional probabilities, that we do not discuss here; this assumption amount to the fact that objects are evenly distributed, so that the system could use the statistical memory to “remember” each location for some specific elements, i.e a bedroom for the bed, a kitchen for the stove, and so on. In other words, given the above relevancy conditions we expect that if the state of nature is  $A_j$ , independently from  $A_j$  prior, there is some feature, object,  $r_i$  peculiar to  $A_j$ , so that after the observation of  $r_i$  the system could come to believe that  $A_j$  is an explanation for its observation.

We shall now consider three *update equations*, under the following:

**2 Updating conditions** *Any prior that should be considered as 0 is set to  $\epsilon$ .  $f(A_j, S_0) = P(A_j, S_0)$ ,  $P(r_i, S_0) = \sum_j P(r_i | A_j, S_0)P(A_j, s_0)$ .  $|\mathcal{E}(s_i)| \leq M/2$ , to ensure relevance of information (where  $M = |\mathbb{L}|$ , i.e. the number of locations).*

Then for all  $n > 0$ :

1.  $f(A_j | r_i, s_n) = \frac{P(r_i | A_j, S_0) f(A_j, s_{n-1}) \gamma(A_j, s_n)}{f(r_i, s_{n-1})}$
2.  $f(r_i, s_n) = \sum_j P(r_i | A_j, S_0) f(A_j, s_{n-1})$
3.  $f(A_j, s_n) = \frac{\sum_i f(A_j | r_i, s_n) f(r_i, s_n)}{\sum_j \sum_i f(A_j | r_i, s_n) f(r_i, s_n)} \quad r_i \in \mathcal{E}(s_n)$

(10)

If  $N$  is the number of parameters, and  $M$  is the number of locations, there are  $w = N \times M + N + M$  equations and  $w$  unknown, for which there is a real solution  $x$ . It is easy to see that for any  $n$ , each term in  $s_n$  can be computed by the initial distribution in  $S_0$  and equations (1-3) of 10.



**Lemma 3** For each  $n$ ,  $f(A_j, s_n)$  is a probability function, such that  $\sum_j f(A_j, s_n) = 1$  and  $0 \leq f(A_j, s_n) \leq 1$ .

*Proof.*  $f(A_j, s_n)$  is defined in equation (3) of (10). For this equation we have to consider that the  $\sum_i$  (appearing both in the denominator and the numerator) concerns only those  $r_i \in \mathcal{E}(s_1)$ , so in particular if  $\mathcal{E}(s_1) = \emptyset$  then  $f(A_j, s_1) = 0/0$ , but this is impossible, by the conditions on updating (see Conditions 2). However  $\sum_j \sum_i f(A_j | r_i, s_1) f(r_i, s_1)$  is a normalization factor, therefore the right hand side of equality (3) of (10) is always of the form  $K_j / K_1 + K_2 + \dots + K_j + \dots + K_q$ , from which the claim follows, given that the outcome space has cardinality  $q$ .  $\square$

Note that the likelihood ratios  $P(r_i | A_j, S_0)$  will remain the same in any state, as they are not affected by the observations. The last equation of (10), is quite relevant, because it updates the beliefs of the system about the current location it is in, and in fact it states that the estimation is established as if the  $r_i \in \mathcal{E}(s_n)$  would form a partition, bounding all it is known on the elements in  $\mathcal{E}(s_n)$ , i.e. making a closed world assumption on the observations.

It might be noted that  $\mathcal{E}(s_n)$  depends only on the current situation, therefore in the next step, the previous observations are some how forgotten. This choice can be justified by the fact that the information gain, obtained from the observation performed in the previous state is already in the probability update. Furthermore the history of observations is not lost because it is maintained in the history of actions, about which we could not say much in this presentation.

We claim that the probability update increases the values of those locations for which there have been relevant observations (see equation (7)).

**Lemma 4** Let  $r_k \in \mathcal{E}(s_n)$ . If  $r_k$  is relevant for  $A_j$ , in  $s_n$ , then  $f(A_j, s_n) > f(A_j, s_{n-1})$ .

*Proof.* First observe that:

- By the observation conditions given in 1, and the updating conditions (see 2 page 7)  $\mathcal{E}(s) \leq M/2$ , for all  $s$ .
- $\sum_j f(A_j, s) = 1$ , for all  $s$  and  $A_j$  (by Lemma 3).

We show the proof for a simpler case than the general, because there are so many terms involved in the proof: i.e we assume  $N = \{r_1, \dots, r_n\}$ ,  $M = \{A_1, A_2\}$ , and thus the set of observed elements  $\mathcal{E}(s_n) = \{r_k\}$ . The same proof can be generalized with some work, to the case  $|M| = m$  (this simplified version allows to eliminate the  $i$ -summation in equation (3) of (10)).

Let  $\mathcal{E}(s_n) = \{r_k\}$ . Suppose that  $\langle A_1, r_k \rangle \in R^+(A_1, s_n)$ , and that  $r_k$  is relevant for  $A_1$ , we want to show that  $f(A_1, s_n) > f(A_1, s_{n-1})$ . By equations 10, and the fact that  $\mathcal{E}(s_n) = \{r_k\}$  we have

$$\begin{aligned} 1. \quad & f(A_1 | r_k, s_n) = \frac{P(r_k | A_1, S_0) f(A_1, s_{n-1}) \gamma(A_1, s_1)}{f(r_k, s_{n-1})} \\ 2. \quad & f(r_k, s_n) = \sum_j P(r_k | A_j, S_0) f(A_j, s_{n-1}) \\ 3. \quad & f(A_1, s_n) = \frac{f(A_1 | r_k, s_n) f(r_k, s_n)}{\sum_j f(A_j | r_k, s_n) f(r_k, s_n)} \quad r_k \in \mathcal{E}(s_n) \end{aligned} \quad (11)$$

We now write equation 3, by substituting each term for its definition:

$$\begin{aligned} f(A_1, s_n) &= \frac{P(r_k | A_1, S_0) f(A_1, s_{n-1}) \gamma(A_1, s_1) f(r_k, s_n)}{f(r_k, s_{n-1})} \\ &= \sum_j \left[ \frac{P(r_k | A_j, S_0) f(A_j, s_{n-1}) \gamma(A_j, s_1) f(r_k, s_n)}{f(r_k, s_{n-1})} \right]_{r_k \in \mathcal{E}(s_n)} \end{aligned} \quad (12)$$

Which simplifies to:

$$\begin{aligned} f(A_1, s_n) &= \frac{P(r_k | A_1, S_0) f(A_1, s_{n-1}) \gamma(A_1, s_1)}{P(r_k | A_1, S_0) f(A_1, s_{n-1}) \gamma + P(r_k | A_2, S_0) f(A_2, s_{n-1})} \\ & \quad r_k \in \mathcal{E}(s_n) \end{aligned} \quad (13)$$

Note that, we have dropped the arguments  $(A_j, s_n)$  from  $\gamma$  in the above denominator, for lack of space, and also because the term with  $A_2$  does not multiply for  $\gamma$ : in fact, since  $\langle A_2, r_k \rangle \notin R(s_n)$ , by Lemma 2,  $\gamma(A_2, s_n) = 1$ . To simplify notation we rename the terms mentioned in (13), as follows:

- $\gamma(A_1, s_n) \rightarrow c$ ;
- $P(r_k | A_1, S_0) \rightarrow a$ ;
- $f(A_1, s_{n-1}) \rightarrow b$ .

Furthermore, since  $M = \{A_1, A_2\}$  and by Lemma 3,  $f(A_2, s_n) = 1 - f(A_1, s_n)$  and  $P(r_k | A_2, S_0) = 1 - P(r_k | A_1, S_0)$ , according to the above renaming, we shall rename the first term as  $1 - b$  and the second as  $1 - a$ . Therefore, we can re-write (13) as follows:

$$f(A_1, s_n) = \frac{abc}{abc + (1-a)(1-b)} \quad (14)$$

Rearranging the denominator we get  $(abc + 1 - b - a + ab)$ , and since  $b$  is  $f(A_1, s_{n-1})$ , we divide both the terms of the equality in (14) by  $b$  and we get:

$$\frac{f(A_1, s_n)}{f(A_1, s_{n-1})} = \frac{ac}{abc + 1 - a - b + ab} \quad (15)$$

We have, thus, to show that:

$$ac > abc + 1 - a - b + ab \quad (16)$$

Under the relevance condition that  $a > (1 - a)$ , and  $c > 1$ , consider the following transformations of the inequality (16), taking into account that both  $a < 1$  and  $b < 1$ :

$$\begin{aligned} ac - (1 - a) &> abc - b + ab \\ a(c - \frac{(1-a)}{a}) &> ab(c - \frac{(1-a)}{a}) \\ c > 1 \text{ and } a > (1 - a) &\text{ implies } c > \frac{1-a}{a} \end{aligned} \quad (17)$$

Since  $a > ab$  (because both  $a < 1$  and  $b < 1$ ), the claim is verified, and thus we have proved that  $f(A_1, s_n) > f(A_1, s_{n-1})$ .  $\square$

On the other hand if an observation does not affect a location  $A_i$ , then by equation (3),  $f(A_i, s_{n+1}) < f(A_i, s_n)$ :



**Lemma 5** Let  $A_j \in \mathcal{E}^-(A_j, s_n)$ , then  $f(A_j, s_n) < f(A_j, s_{n+1})$ .

*Proof.* Let  $\mathcal{E}(s_n) = \{r_k\}$ , with  $r_k$  not relevant to  $A_j$ , because  $A_j \in \mathcal{E}^-(A_j, s_n)$  (see (8) page 6). We shall make the same simplification on the number of locations and observations used in Lemma 4. Since  $r_k$  is not relevant for  $A_1$ , we have to show that  $f(A_1, s_{n+1}) < f(A_1, s_n)$ . Consider equation (12) and its simplification (13), now  $\gamma(A_1, s_n) = 1$ , and thus we can drop it, however since  $r_k$  is not relevant for  $A_1$  it must be that it is relevant for  $A_2$ , and thus using the simplified notation, dividing both the terms of the equality by  $b$ , and rearranging the denominator, we get:

$$\frac{f(A_1, s_n)}{f(A_1, s_{n-1})} = \frac{a}{ab + c - ac - bc + abc} \quad (18)$$

We have, thus, to show that:

$$a < ab + c - ac - bc + abc \quad (19)$$

Under the condition that  $a < 1, b < 1$  and  $a < 1 - a$  and  $1 < c$ . Consider the following inequalities:

$$\begin{aligned} 1 < c; \quad \frac{1-b}{c} < 1-b; \quad (1-b)a < (c-bc)(1-a) \\ a - ab < c - ac - bc + abc \end{aligned} \quad (20)$$

The above lemmas show that relevant observations tend to enforce hypotheses. We want to show under which conditions one hypothesis can emerge over all the other ones. Our claim is as follows. Suppose that for a given sequence of observation states  $O_{s_1}, O_{s_2}, \dots, O_{s_k}$ , there are observations  $\{r_{j1}, \dots, r_{jn}\} \subset \bigcup_{i=1}^k \mathcal{E}(s_i)$  relevant to  $A_j$  (that is,  $\langle A_j, r_{jp} \rangle \in \mathcal{R}^+(A_j, s_i), 1 \leq p \leq n$ , and  $1 \leq i \leq k$ ); then we expect that if the saliency of each  $r_{jp}, 1 \leq p \leq n$ , is meaningful, and  $A_j$  is at each step a best explanation for the current observations, then  $f(A_j, s_k)$  will be maximal with respect to all the other  $A_w \neq A_j$ , and it will maintain this position. Which, in the specific example, means that the system becomes aware of the location it is currently standing in. Now, we have been able to prove only the sufficient conditions, not the necessary ones, which means that the conditions we put forward might be unnecessarily stronger than needed. To this end we first need to introduce a suitable notion of growth for each function in the observation process, as follows:

**3 Growth** The specific increment of  $A_j$  in situation  $s_i$ , is defined to be  $\delta(A_j, s_i) = \gamma(A_j, s_i) \sum_k P(r_k | A_j, s_0) f(r_k, s_i)$ , with  $\langle A_j, r_k \rangle \in \mathcal{R}^+(A_j, s_i)$ .  $\delta(A_j, s_i)$  reflects the amount of observed elements relevant to  $A_j$  at each step, and their saliency. Furthermore we denote with  $\Delta(A_j, s_i)$  the total increment of  $A_j$  in  $s_i$ , i.e.  $f(A_j, s_i) - f(A_j, s_{i-1})$ .

The conditions for convergence are as follows:

**4 Convergence conditions** Given a sequence  $O_{s_1}, O_{s_2}, \dots, O_{s_k}$  of observations, we say that  $A_j$  is maximal in  $\bigcup_i \mathcal{R}(s_i)$ , if for all  $1 \leq i \leq k$ :

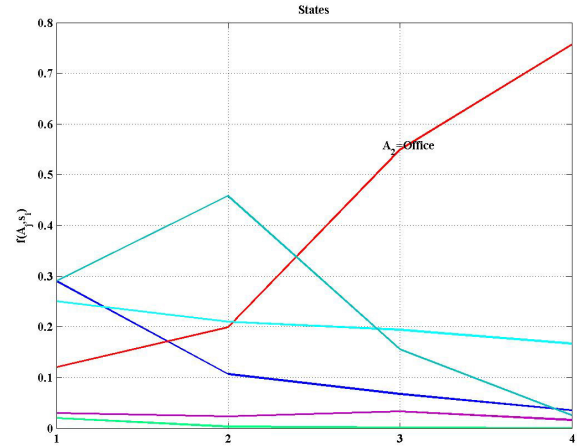


Figure 2: Surpass point

1. For all  $A_w \neq A_j$ , s.t.  $P(A_j, S_0) \leq P(A_w, S_0)$ , then there exists  $s_i < s_k$  s.t.  $P(A_w, S_0) - \Delta(A_j, s_i) \leq P(A_j, S_0)$ .
2.  $|R^+(A_j, s_i)| > |R^+(A_w, s_i)|$ , for all  $A_w \neq A_j$ .
3.  $\gamma(A_j, s_i) > \gamma(A_w, s_i)$ , for all  $A_w \neq A_j$ .

It follows that if  $A_j$  is maximal then for all  $s_i$  in the sequence,  $\delta(A_j, s_i) > \delta(A_w, s_i)$  for all  $A_w \neq A_j$ .

Observe that the first condition is very strong, and it says that if a state of nature is very unlikely then its growth slope must compensate it. In fact, to cope with a sequence of noisy observations, strong conditions seem necessary. For, suppose that there is a location  $A_w$ , with  $P(A_w, S_0) = 0.5$ , then any noisy observation concerning  $A_w$ , would make the system believe in  $A_w$  just because it is initially so privileged. Therefore this condition is also an indication on the initial structure of the statistical memory (i.e. the class-conditional probabilities): it should not be biased towards a specific location. The other two items say that at each observation step  $O_{s_i}$  what is noted makes  $A_j$  more likely, and what is observed has a meaningful saliency, because  $\gamma(A_j, s_i)$  is greater than all the other  $\gamma$ 's.

The above notion is needed to prove that even if  $P(A_j, s_0) < P(A_w, s_0)$ , there is a surpass point, that is a point  $s_m$  at which  $f(A_j, s_m) > f(A_w, s_m)$ . This is illustrated in Figure 2.

**Theorem 1** Let  $O_{S_0}, \dots, O_{S_t}$  be a sequence of observations, and let  $A_j$  be maximal on the whole sequence. Then the sequence converges on  $A_j$  as a most likely hypothesis, having maximal value for  $f(A_j, s_t)$ , that is,  $f(A_j, s_t) > f(A_q, s_t)$ , for all  $A_q$ , with  $q \neq j$ .

*Proof.* We give only a sketch of the proof and all its steps.

First observe that if  $A_j$  is maximal on the whole sequence, then in particular  $f(A_j, s_i) \in \bigcup_{i=1}^t \mathcal{R}(s_i)$ , hence it is monotonically increasing, by Lemma 4. Furthermore at each observation step  $s_i$ , some  $A_w \notin \mathcal{R}(s_i)$ , because  $\mathcal{E}(s_i) < M/2$  (see 2 page 7), and thus for Lemma

5, for some  $A_w$ ,  $f(A_w, s_i) < f(A_w, s_{i-1})$ . By the convergence conditions and at each step  $A_j$  will increase more than any other increasing function, because of items 2 and 3 of the convergence condition. Finally, even if  $\lim_{i \rightarrow \infty} f(A_w, s_i) = 1$  for all  $f(A_w, s_i)$  increasing on the sequence, since  $\sum_j f(A_j, s_i) = 1$ , the increasing must have very different  $\delta$ , so the set  $\mathbb{L}$  could be partitioned in three subsets: the monotonically increasing the monotonically decreasing, and those oscillating. For this reason we have first to show that, if  $A_j$  is maximal in the sequence, then as soon  $A_j$  surpasses any other increasing or oscillating  $A_w$ , then  $A_w$  will never more reach  $A_j$ , independently if it is increasing, decreasing or oscillating; in so we get read of the behaviour of any other function, whose prior is less or equal that of  $A_j$ . of  $A_j$ . Let  $A_j$  be maximal on the whole sequence of observation states, we need to show that:

1. If  $f(A_j, s_i) \geq f(A_w, s_i)$  then  $f(A_j, s_k) > f(A_w, s_k)$ , for all  $s_k > s_i$ .
2. If  $P(A_j, S_0) < P(A_w, S_0)$ , then there exists an  $s_m$  such that  $f(A_j, s_m) > f(A_w, s_m)$ ,  $S_0 < s_m \leq s_t$ .

1. Let  $f(A_j, s_i)$  be maximal in the sequence, and suppose that  $f(A_j, s_i) \geq f(A_w, s_i)$ , we show that  $f(A_j, s_{i+1}) > f(A_w, s_{i+1})$  and the dominance will persist, for the whole sequence. We sketch the proof by induction on  $i$  with basic step  $s_1$ . By substituting in equation (3) in (10) each term for its definition and considering that both in equation (1) and (3) of (10), the normalization factors are the same for all  $A_w$ , we can substitute them with a constant  $S$ , that we omit, then we have to show:

$$\frac{\gamma(A_j, s_1)P(A_j, S_0) \times \sum_k P(r_k | A_j, S_0) f(r_k, s_1)}{\gamma(A_w, s_1)P(A_w, S_0) \times \sum_k P(r_k | A_w, S_0) f(r_k, s_1)} > \quad (21)$$

Which is verified because by the hypothesis  $P(A_j, S_0) \geq P(A_w, S_0)$ , because  $\gamma(A_j, s_1) > \gamma(A_w, s_1) > 1$ , by the convergence conditions, and because  $|R^+(A_j, s_1)| > |R^+(A_w, s_1)|$  together with the hypothesis of  $A_j$  maximal, implies that  $\sum_k P(r_k | A_j, S_0) > \sum_k P(r_k | A_w, S_0)$ . For the induction let us write  $f(A_j, s_{i+1})$  as:

$$\frac{1}{S} f(A_j, s_i) (\delta(A_j, s_{i+1}) + \gamma(A_j, s_{i+1}) \sum_k P(r_k | A_j, S_0) f(r_k, s_i)) \quad (22)$$

Where  $\sum_k P(r_k | A_j, S_0) f(r_k, s_1)$  denotes the set of elements on which there is the increment of  $A_w$  and of all any other increasing  $A_p$ . Note that  $\delta(A_j, s_{i+1}) > \gamma(A_j, s_{i+1}) \sum_k P(r_k | A_j, S_0) f(r_k, s_i)$ . We can write  $f(A_w, s_{i+1})$ , in an analogous way:

$$\frac{1}{S} f(A_w, s_i) (\delta(A_w, s_{i+1}) + \gamma(A_w, s_{i+1}) \sum_k P(r_k | A_w, S_0) f(r_k, s_i)) \quad (23)$$

By induction hypothesis,  $f(A_j, s_i) \geq f(A_w, s_i)$ ; by maximality  $\gamma(A_j, s_{i+1}) > \gamma(A_w, s_{i+1})$ , and again by maximality,  $\delta(A_j, s_{i+1}) > \delta(A_w, s_{i+1}) > \gamma(A_w, s_{i+1}) \sum_k P(r_k | A_w, S_0) f(r_k, s_i)$ . Observe that the terms in both  $\sum_k P(r_k | A_j, S_0) f(r_k, s_i)$  and

$\sum_k P(r_k | A_w, S_0) f(r_k, s_i)$  are of no account and they do not weight on the whole summation, so they could even be eliminated. Therefore we have proved that  $f(A_j, s_{i+1}) > f(A_w, s_{i+1})$ , if  $f(A_j, s_i) \geq f(A_w, s_i)$ . This, in particular means that, once a maximal  $f(A_j, s_i)$  becomes greater than some  $f(A_w, s_i)$ , it will remain so for the whole sequence.

2. We want to show that, if  $P(A_j, S_0) < P(A_w, s_i)$  and  $A_j$  is maximal, then the sequence converges to  $f(A_j, s_i)$ . By Lemma 4, it follows that  $\Delta(A_j, s_i) > 0$  for all  $s_i$ , and by the convergence conditions, it follows that there is an  $s_i$  such that  $f(A_j, s_1) + \Delta(A_j, s_i) \geq f(A_w, s_1)$ , therefore by the first item shown above,  $f(A_j, s_{i+1}) > f(A_w, s_{i+1})$ , and it will remain dominant.  $\square$

The above theorem states that, given a maximality criterion, it is possible to establish after a given sequence  $k$  that the updating did converge.

Table (2), illustrates the outcomes of a sequence of three observations, leading to the hypothesis that the current location is an office. In Figure 2, is shown that after the first observation  $f(Office, s_1)$  becomes increasing. So we end up with a value of  $f(Office, s_3) = 0.68427$ , despite it never has been maximal according to all the convergence conditions. In fact, given that the first condition is satisfied, in the first observation *Office* satisfies the second but not the third condition. In the second observation, *Office* does not satisfy neither the second nor the third, and in the third observation state, *Office* satisfies the second but not the third condition. Observe that there is no  $A_j$ , in this example which is maximal along the whole run. This can be interpreted as follows: we do not know, as far as what has been proved here, if after  $n$  more observations the system could end up with a conclusion far different from the one established after the three shown observations. This suggests that a sequence starting point could be considered the one from which one of the posteriors begin to have a maximal behaviour.

At the end of the process the result is reported to the pre-selective process in order to proceed to a further verification. Two possible actions are in order: either the situation is progressed, and therefore the updated table is recovered as in the initial situation  $S_0$ , or a further analysis is performed, and in such a case the memory maintained is necessary to proceed toward a verification of the hypotheses.

A video about the indoor classification problem presented here, and implemented on a pioneer 3DX, can be found at the page: [www.dis.uniroma1.it/~Alcor](http://www.dis.uniroma1.it/~Alcor), by going on the "events" page. The work has been exhibited in Milan at the IST European exposition.

## References

- Bacchus, F.; Halpern, J.; and Levesque, H. 1999. Reasoning about noisy sensors in the situation calculus. *Artificial Intelligence* 111:171–208.
- Baluja, S., and Pommerleau, D. A. 1995. Using a saliency map for active spatial selective attention: Implementation & initial results. In Tesauro, G.; Touretzky, D.; and Leen,

- T., eds., *Advances in Neural Information Processing Systems*, volume 7, 451–458. The MIT Press.
- Chang, T., and Kuo, C. 1992. Tree-structured wavelet transform for textured image segmentation. *SPIE* 1770:394–405.
- Chang, T., and Kuo, C. 1993. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on Image Processing* 2(4):429–441.
- Darrell, T., and Pentland, A. 1995. Attention-driven expression and gesture analysis in an interactive environment. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, 135–140.
- Golden, K., and Weld, D. 1996. Representing sensing actions: The middle ground revisited. In *KR'96: Principles of Knowledge Representation and Reasoning*. 174–185.
- Hurley, S. 1998. *Consciousness in Action*. Harvard University Press, Cambridge, MA.
- Itti, L., and Koch, C. 1998. Learning to detect salient objects in natural scenes using visual attention. In *DARPA98*, 1201–1206.
- Itti, L., and Koch, C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention.
- Itti, L.; Koch, C.; and Niebur, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(11):1254–1259.
- Koch, C., and Ullman, S. 1985. Shifts in selective visual-attention: towards the underlying neural circuitry. *Human Neurobiology* 4(4):219–227.
- Laine, A., and Fan, J. 1993. Texture classification by wavelet packet signatures. *PAMI* 15(11):1186–1191.
- Lavie, N. 1995. Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance* 21:451–468.
- Levesque, H. J., and Scherl, R. B. 1993. The frame problem and knowledge-producing actions. In *Proceedings of the National Conference on Artificial Intelligence (AAAI'93)*, 689–695.
- Li, J.; Najmi, A.; and Gray, R. 1999. Image classification by a two dimensional Hidden Markov Model. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*.
- Nikolaidis, N.; Pitas, I.; and Mallot, H. 2000. *Computational Vision: Information Processing in Perception and Visual Behavior*. MIT Press.
- Noe, A.; Pessoa, L.; and Thompson, E. 2000. Beyond the grand illusion: what change blindness really teaches us about vision. *Visual Cognition* 7(1-3):93–106.
- Pirri, F., and Finzi, A. 1999. An approach to perception in theory of actions: Part I. *ETAI* 3:19–61.
- Pirri, F., and Reiter, R. 1999. Some contributions to the metatheory of the situation calculus. *Journal of the ACM* 46(3):325–361.
- Pirri, F., and Reiter, R. 2000. Planning with natural actions in the situation calculus. 213–231.
- Pirri, F., and Romano, M. 2002. A Situation-Bayes view of object recognition based on symgeons. In *The Third International Cognitive Robotics Workshop*.
- Pirri, F., and Romano, M. 2003. 2d qualitative recognition of symgeon aspects. In *Soft Computing Techniques for 3D Vision - KES2003*, 1187–1194.
- Pirrone, M. 2003. Active vision and visual attention for indoor environment classification. Technical report, University of Roma 'La Sapienza'.
- Rabiner, L. R., and Juang, B. H. 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine* 4–15.
- Rees, G.; Frith, C.; and Lavie, N. 1997. Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science* 278(5343):1616–1619.
- Reiter, R. 2001a. *KNOWLEDGE IN ACTION: Logical Foundations for Building Dynamic Systems*. MIT press.
- Reiter, R. 2001b. On knowledge-based programming with sensing in the situation calculus. *ACM Trans. Comput. Logic* 2(4):433–457.
- Romberg, J. K.; Choi, H.; and Baraniuk, R. G. 1999. Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. In *SPIE Technical Conference on Mathematical Modeling, Bayesian Estimation, and Inverse Problems*, volume 3816, 31–44.
- Shanahan, M. 2002. A logical account of perception incorporating feedback and expectation. In *Proceedings of KR2002*. 3–13.
- Shanahan, M. 2004. A logic based formulation of active visual perception. In *Proceedings of KR2004*.
- Siegel, M.; Koerding, K. P.; and Koenig, P. 2000. Integrating top-down and bottom-up sensory processing by somato-dendritic interactions. *Journal of Computational Neuroscience* 8:161–173.
- Thrun, S. 2002. Particle filters in robotics. In *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI)*.
- Viterbi, A. 1967. Error bounds for convolutional codes an an asymptotically optimal decoding algorithm. *IEEE Transaction on Information Theory* IT-13:260–269.