Bayesian Compositional Hierarchies - A Probabilistic Structure for Scene Interpretation

Bernd Neumann

CSL, University of Hamburg
neumann@informatik.uni-hamburg.de

Abstract. In high-level vision, it is often useful to organize conceptual models in compositional hierarchies. For example, models of building facades (which are used here as examples) can be described in terms of constituent parts such as balconies or window arrays which in turn may be further decomposed. While compositional hierarchies are widely used in scene interpretation, it is not clear how to model and exploit probabilistic dependencies which may exist within and between aggregates. A probabilistic framework has to meet the challenge that probabilities must be continually updated as evidence becomes available and incremental interpretation steps are performed. Hence computational efficiency is mandatory. In this report I present Bayesian Compositional Hierarchies as a means to capture probabilistic dependencies in an aggregate hierarchy. The formalism integrates well with object-centered representations and extends Bayesian Networks by allowing arbitrary probabilistic dependencies within aggregates. To obtain efficient inference procedures, the aggregate structure must possess abstraction properties which ensure that internal aggregate properties are only affected in accordance with the hierarchical structure. Using examples from the building domain, it is shown that probabilistic aggregate information can thus be integrated into a logic-based scene interpretation system and provide a preference measure for interpretation steps.

Keywords: Scene interpretation, probabilistic inferences, compositional hierarchies

1. Introduction

Interpretations are generally ambiguous and not clearly defined with respect to a task. When constructing an explanation for evidence one often has the choice between alternatives. For example, given a straight knowledge base about building facades, the image section marked in Fig. 1 can be interpreted both as an entrance or a balcony. In the course of a stepwise interpretation, there can be many more decision points where multiple choices are available. As humans, we seem to exploit our experiences for such decisions and prefer the most likely choice given all we know about the domain

Dagstuhl Seminar Proceedings 08091
Logic and Probability for Scene Interpretation
http://drops.dagstuhl.de/opus/volltexte/2008/1605

and the current scenario. Hence it appears natural to provide a probabilistic model for the uncertainty of logically ambiguous choices.

Fig. 1. Facade component in box may be both a balcony or an entrance. (It is an entrance of a house in Montreal)

The basic idea is to consider concepts as random variables with probability distributions which govern the likelihood of possible instantiations represented by the concept. A general approach to construct Bayesian Networks for first-order logic expressions is presented in [1]. Our approach, first sketched out in [2], exploits the fact that aggregates are the concepts of interest for an interpretation task and dependencies between objects can effectively be encapsulated in aggregates. Within aggregates, we do not require conditional independence of parts given aggregate properties as in the pioneering work of [3] but allow arbitrary distributions. We will, however, impose certain abstraction requirements in order to ensure that efficient propagation mechanisms can be used.

To simplify the presentation, let us assume that each part may occur at most once in an aggregate. Aggregates with repeated parts must be described by giving every possible part its own representation within the aggregate. Alternatively, aggregates with different part configurations can be treated as different concepts. We will be able to incorporate taxonomical branchings in our model, so alternative aggregates do not pose problems.

This way, the space of all interpretations has an AND-OR node structure, with aggregate nodes indicating an AND relation between parts, and concept specialisation nodes (representing taxonomical branchings) an OR relation between specialisations.

The task is now to assign probability distributions to aggregates and their parts such that the probability of any object (regarding its existence, location and other properties) can be computed at any time during the interpretation process conditioned on the evidence which has been incorporated so far. In other words: We want to be able to provide dynamic priors exploiting high-level context and partial evidence.

Let A be an aggregate concept and B1 ... BN its part concepts. An aggregate will be represented by the following random variables (the understroke denotes a vector):

Ax	boolean random variable representing the existence probability of A
$\underline{A}$	vector-valued random variable representing simple properties of A
B1x ... BNx	boolean random variables representing the existence probabilities of the parts

$\underline{B}1 \ldots \underline{B}N$	vector-valued random variables representing the properties of the parts

Properties are assumed to map into a fixed domain of values, not into structured objects (called simple functions in [1], page 520). Similar models have been proposed by [4] and [5] for situation modelling.

The probabilistic dependencies between the random variables are described by the following joint probability distribution (JPD):

$$P(\underline{B}1 \ldots \underline{B}N\ B1x \ldots BNx \mid Ax = T)$$

This distribution of part properties, called parts distribution, reflects all constraints imposed by the aggregate definition. Note that the existence properties B1x ... BNx allow to model aggregate configurations with varying numbers of parts. Note also, that there is no meaningful distribution for parts if Ax = F.

Part properties are mapped into external properties by the function $\underline{f}a$:

$$\underline{a} = \underline{f}a(\underline{b}1 \ldots \underline{b}N\ b1x \ldots bNx)$$

The function $\underline{f}a$ is the abstraction function of aggregate A. In general, $\underline{f}a$ maps detailed part properties into less detailed aggregate properties. For example, $\underline{f}a$ could compute the bounding-box coordinates of an aggregate from the bounding box coordinates of its parts.

From the parts distribution one can compute the JPD of aggregate properties $\underline{A}$:

$$P(\underline{A} \mid \underline{B}1 \ldots \underline{B}N\ B1x \ldots BNx\ Ax = T)$$

The properties $\underline{A}$ are called external aggregate properties. They represent the aggregate as a whole when it is part of a higher-level aggregate. Correspondingly, $\underline{B}1 \ldots \underline{B}N$ are called internal aggregate properties. Each $\underline{B}k$ may simultaneously describe the external properties of a lower-level aggregate. Parts which do not decompose further are called primitive parts.

An aggregate can be graphically represented as shown in Fig. 2. Each part of an aggregate can be the root of further decompositions, hence aggregates give rise to a compositional hierarchy as illustrated in Fig. 3.

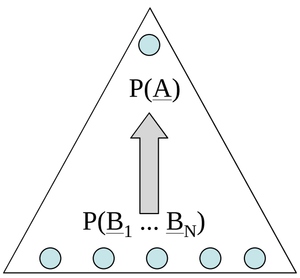

Fig. 2. Probabilistic aggregate structure

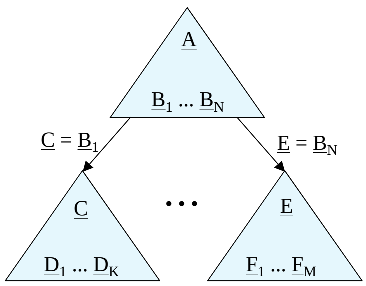

Fig. 3. Aggregates form a compositional hierarchy

2. Bayesian Compositional Hierarchies

For probabilistic inferences in a compositional hierarchy, we have to be precise about probabilistic dependencies beyond those expressed by the aggregate specifications. How does evidence for one aggregate influence the probabilities in another aggregate? Note that the aggregate hierarchy is not a Bayesian Network: Given the external aggregate properties, parts are in general not statistically independent, hence subtrees below the parts will in general also be dependent.

Intuitively, external aggregate properties should represent all information relevant for probabilistic dependencies concerning the aggregate as a whole, abstracting from irrelevant parts properties. We now state conditional independence requirements which reflect this intuition, and show that the requirements give rise to an interesting factorisation theorem, enabling efficient probabilistic inference procedures.

Let us simplify the notation by denoting the existence variable Ax of an object together with its property values $\underline{\mathrm{A}}$ by the augmented property vector $\underline{A}$ (in italic):

$$\underline{A} = [\mathrm{Ax}\ \underline{\mathrm{A}}]$$

In the following, we always refer to objects in terms of their augmented property values. Let parts($\underline{A}$) = $\underline{B}1$ ... $\underline{B}N$ be the parts of an aggregate $\underline{A}$ (empty, if $\underline{A}$ is primitive) and succ($\underline{A}$) be all objects in the hierarchy below $\underline{A}$ (including its parts).

We postulate that the following abstraction requirements are fulfilled:

Requirement 1:

$$\mathrm{P}(\mathrm{succ}(\mathrm{parts}(\underline{A}))\mid \mathrm{parts}(\underline{A})\ \underline{A}) = \mathrm{P}(\mathrm{succ}(\mathrm{parts}(\underline{A}))\mid \mathrm{parts}(\underline{A})) \qquad (\mathbf{1})$$

Given properties of parts of an aggregate, the properties of successors of the parts do not depend on the external properties of the aggregate.

Requirement 2:
Let $\underline{B}1$ ... $\underline{B}N$ be the parts of an aggregate $\underline{A}$.

$$\mathrm{P}(\mathrm{succ}(\underline{B}k)\mid \underline{B}1 \ldots \underline{B}N) = \mathrm{P}(\mathrm{succ}(\underline{B}k)\mid \underline{B}k) \qquad (\mathbf{2})$$

Given aggregate properties, then properties of its parts do not depend on siblings of the aggregate

Requirement 3:
Let $\underline{B}1$ ... $\underline{B}N$ be the parts of an aggregate $\underline{A}$.

$$\mathrm{P}(\mathrm{succ}(\underline{B}1 \ldots \underline{B}N)\mid \underline{B}1 \ldots \underline{B}N) = \Pi\, \mathrm{P}(\mathrm{succ}(\underline{B}k)\mid \underline{B}1 \ldots \underline{B}N) \qquad (\mathbf{3})$$

Given their mother aggregates, parts of different aggregates are statistically independent.

From requirements 2 and 3 it follows that

$$\mathrm{P}(\mathrm{succ}(\underline{B}1 \ldots \underline{B}N)\mid \underline{B}1 \ldots \underline{B}N) = \Pi\, \mathrm{P}(\mathrm{succ}(\underline{B}k)\mid \underline{B}k)$$

The three abstraction requirements express that the JPD of any object in the compositional hierarchy is only affected via its immediately connected hierarchy neighbours. Hence evidence propagation will simply have to follow the hierarchical structure.

Exploiting these abstraction requirements, we can derive a factorisation formula for the JPD of a complete compositional hierarchy. Let $\underline{Z}k$, k = 0 ... M be all objects of the compositional hierarchy and $\underline{Z}0$ its root, then

$$P(\underline{Z}0 \ldots \underline{Z}M) = P(\underline{Z}0)\, P(\mathrm{succ}(\underline{Z}0) \mid \underline{Z}0) \tag{4}$$

$$= P(\underline{Z}0)\, P(\mathrm{parts}(\underline{Z}0)\ \mathrm{succ}(\mathrm{parts}(\underline{Z}0)) \mid \underline{Z}0)$$
$$= P(\underline{Z}0)\, P(\mathrm{parts}(\underline{Z}0)) \mid \underline{Z}0)\, P(\mathrm{succ}(\mathrm{parts}(\underline{Z}0)) \mid \mathrm{parts}(\underline{Z}0)\ \underline{Z}0)$$

$$= P(\underline{Z}0)\, P(\mathrm{parts}(\underline{Z}0)) \mid \underline{Z}0)\, P(\mathrm{succ}(\mathrm{parts}(\underline{Z}0)) \mid \mathrm{parts}(\underline{Z}0)) \tag{5}$$

The last step exploits Requirement 1. Let part($\underline{Z}0i$) be the ith part of $\underline{Z}0$, then Eq. 2 can be rewritten using Requirements 2 and 3:

$$P(\underline{Z}0 \ldots \underline{Z}M) = P(\underline{Z}0)\, P(\mathrm{parts}(\underline{Z}0)) \mid \underline{Z}0)\, \Pi\, P(\mathrm{succ}(\mathrm{part}(\underline{Z}0i)) \mid \mathrm{part}(\underline{Z}0i)) \tag{6}$$

Now the derivation steps from Eqs. 4 to 6 can be applied recursively, and we obtain

$$P(\underline{Z}0 \ldots \underline{Z}M) = P(\underline{Z}0)\, \Pi\, P(\mathrm{parts}(\underline{Z}k) \mid \underline{Z}k), \qquad k = 1 \ldots M \tag{7}$$

Because of the remarkable similarity to the well-known Bayesian Network factorisation formula, we call compositional hierarchies meeting the three abstraction conditions "Bayesian Compositional Hierarchies" (BCHs). The BCH factorisation formula Eq. 7 states that all probabilistic inferences can be carried out solely based on aggregate descriptions in terms of the JPD of internal aggregate properties given external aggregate properties.

It is interesting to rephrase the BCH factorisation formula in terms of aggregate descriptions based on $P(\underline{Z}k \mid \mathrm{parts}(\underline{Z}k))$. As shown earlier, these conditional probabilities are in fact deterministic mappings from internal to external aggregate properties. We get the following alternative formula:

$$P(\underline{Z}0 \ldots \underline{Z}M) = \Pi\, P(\underline{Z}k \mid \mathrm{parts}(\underline{Z}k)\, C(\mathrm{parts}(\underline{Z}k)) \qquad k = 1 \ldots M \tag{8}$$

with $C(\mathrm{parts}(\underline{Z}k)) = P(\mathrm{parts}(\underline{Z}k)) \,/\, \Pi\, P(\underline{Z}ki)$, where the $\underline{Z}ki$ are all parts of $\underline{Z}k$. $C(\mathrm{parts}(\underline{Z}k))$ reflects the correlation between the parts of an aggregate and equals 1 for uncorrelated parts.

The alternative factorisation formula shows an intuitive way for determining the probabilities of a BCH. Starting with the JPDs of primitive aggregate parts, the JPDs of external aggregate properties are determined using the abstraction function which maps internal into external property values. This process is continued incrementally for higher abstraction levels.

A challenging goal, of course, would be to learn aggregate definitions which meet the abstraction conditions. At this time we are not aware of a clustering approach for finding a BCH which approximates a given JPD of primitive parts.

3. Taxonomical Aggregate Relations

For model-based scene interpretation, it is also necessary to structure aggregate concepts in taxonomical hierarchies based on specialisation relations, and a probabilistic model must include such relations. Fortunately, the probabilistic model introduced for compositional hierarchies can also be used for taxonomical hierarchies. A concept A and its specialisations B1 ... BN are described as follows:

Ax	boolean random variable representing the existence probability of A
$\underline{A}$	vector-valued random variable representing simple properties of A
B1x ... BNx	boolean random variables representing the existence probabilities of the specialisations
$\underline{B}1$ ... $\underline{B}N$	vector-valued random variables representing the properties of the parts

The probabilistic dependencies can be described by the JPD P($\underline{B}1$ ... $\underline{B}N$ B1x ... BNx | $\underline{A}$ Ax). Here P(B1x ... BNx | $\underline{A}$ Ax) models the probabilities for each of the possible specialisations given properties of the mother concept A. Note that, in general, specialisations need not be disjunctive, i.e. the concepts $\underline{B}1$ ... $\underline{B}N$ may overlap. P($\underline{B}k$ Bk=T | $\underline{A}$ Ax) models the dependencies between properties of a specialisation and the mother concept. Since all properties of A are inherited, one can think of $\underline{B}k$ as a property vector which refines and extends $\underline{A}$.

For disjunctive specialisations, the JPDs of the specialisations are independent given the mother concept:

P($\underline{B}1$ ... $\underline{B}N$ B1x ... BNx | $\underline{A}$ Ax) = P($\underline{B}1$ B1x | $\underline{A}$ Ax) ... P($\underline{B}N$ BNx | $\underline{A}$ Ax)

For the BCH factorisation formula to hold and for the validity of a probabilistic inference scheme based on the three abstraction requirements, we must show that the abstraction requirements also hold for taxonomical relations.

Consider a tree-shaped specialisation hierarchy where each concept has a single parent (except the root) and specialisations of a concept are disjunctive. Let us call the immediate specialisations of a concept "children" (replacing "parts" used for compositional hierarchies). Then the first abstraction requirement is:

Requirement 1:

P(succ(children($\underline{A}$)) | children($\underline{A}$) $\underline{A}$) = P(succ(children($\underline{A}$)) | children($\underline{A}$))

Given properties of concept children, then properties of their successors do not depend on the properties of the mother concept. This requirement is always fulfilled as children include all properties of their mother concept by definition of a specialisation, and the mother concept does not add new information.

Requirement 2:
Let $\underline{B1}$ ... $\underline{BN}$ be the children of a concept $\underline{A}$.

P(succ($\underline{Bk}$) | $\underline{B1}$ ... $\underline{BN}$) = P(succ($\underline{Bk}$) | $\underline{Bk}$)

Given properties of a mother concept, then properties of its specialisations do not depend on siblings of the mother concept. This is always the case as long as children of a concept are disjunctive. If not, then information about a sibling of the mother concept could influence expectations about the children.

Requirement 3:
Let $\underline{B}1 \ldots \underline{B}N$ be the children of a concept $\underline{A}$.

$$P(\text{succ}(\underline{B}1 \ldots \underline{B}N) \mid \underline{B}1 \ldots \underline{B}N) = P\ P(\text{succ}(\underline{B}k) \mid \underline{B}1 \ldots \underline{B}N)$$

Given their mother concepts, specialisations of different concepts are statistically independent. Again, for this requirement to hold, children must be disjunctive.

Summarising this section, we have shown that for disjunctive specialisation, taxonomical relations can also be modelled within the framework of Bayesian Compositional Hierarchies.

4. Probability Propagation

We have shown that in a BCH, aggregates influence each only via the connections in the tree structure. We now describe probability propagation in detail. As an illustrating example consider Fig. 4.

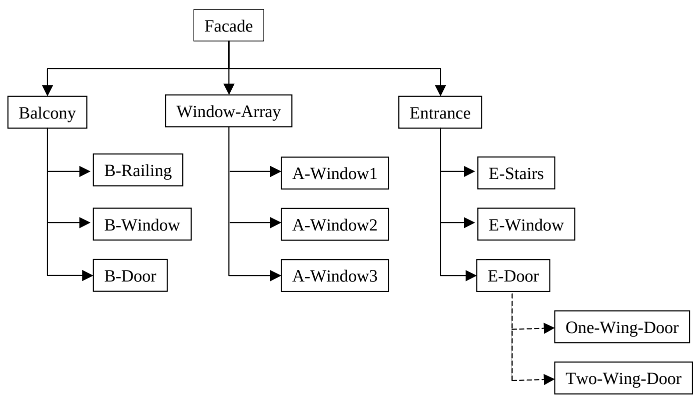

Fig. 4. Simple compositional hierarchy (solid arrows) for a facade including a taxonomical refinement (dotted arrows)

Let us assume that all aggregates $\underline{Z}k$ are described by $P(\text{parts}(\underline{Z}k) \mid \underline{Z}k)$ which specifies the JPD of aggregate parts given the external aggregate properties. In order to determine the prior probabilities for all objects, we have to provide the prior root probability of the hierarchy $P(\underline{Z}0)$. In the example, this could be the probability distribution for existence, location and size of a facade. We now determine

$$P(\text{parts}(\underline{Z}0)\ \underline{Z}0) = P(\text{parts}(\underline{Z}0) \mid \underline{Z}0)\ P(\underline{Z}0)$$

and then $P(\underline{Z}0k)$ for all parts of $\underline{Z}0$ by marginalising. Proceeding top-down in the same manner, we obtain priors for all objects of the hierarchy.

Let us now assume that evidence for a leaf object has been found. For our example, this could be evidence for the Two-Wing-Door in terms of specific values for position and size. We want to determine the influence of this evidence on the remaining random variables of the hierarchy. As propagation will follow the hierarchical structure, it suffices to show how the changed JPD of a part affects the JPD of the aggregate which contains it (bottom-up propagation) and how the changed JPD of an aggregate affects its parts (top-down propagation). To specify the propagation rules, we will denote the external properties of an aggregate by $\underline{A}$ and the properties of its kth part by $\underline{B}k$.

For bottom-up propagation, let us assume that the JPD of $\underline{B}k$ changes from $P(\underline{B}k)$ to $P'(\underline{B}k)$. Then the changed JPD of $\underline{A}$ is determined by

$$P'(\underline{A}\ \underline{B}1 \ldots \underline{B}N) = P(\underline{A}\ \underline{B}1 \ldots \underline{B}N)\ P'(\underline{B}k) / P(\underline{B}k) \qquad (\mathbf{9})$$

followed by marginalisations. Similarly, for top-down propagation we assume that $P(\underline{A})$ has changed to $P'(\underline{A})$. Then the changed JPD of the parts $\underline{B}1 \ldots \underline{B}N$ is determined by

$$P'(\underline{A}\ \underline{B}1 \ldots \underline{B}N) = P(\underline{A}\ \underline{B}1 \ldots \underline{B}N)\ P'(\underline{A}) / P(\underline{A}) \qquad (\mathbf{10})$$

followed by marginalisations. It is convenient to model the introduction of crisp evidence also as a change of a JPD. Thus if the evidence $\underline{B} = \underline{b}$ becomes available for an object with JPD $P(\underline{B})$, then the changed JPD of $\underline{B}$ is $P'(\underline{B}) = 1$ for $\underline{B} = \underline{b}$ und 0 otherwise.

In our example, after receiving evidence for the Two-Wing-Door, the probabilities of the superconcept E-Door and of the aggregates Entrance, Facade, Window-Array and Balcony have to be recomputed, requiring five propagation steps.

This process is repeated whenever new evidence forces the change of a marginal probability.

5. Scene Interpretation with Probabilistic Guidance

The rationale of the BCH is to provide context-sensitive and dynamic priors for all objects for which evidence may become available. In order to clarify the role of evidence, we have to refine the hierarchy shown in Fig. 4. Every physical object concept will be connected to a corresponding view concept which describes possible appearances of the physical object (Fig. 5). A view concept is modelled probabilistically as another concept, and its relation to the physical object concept is expressed analog to the relation of a part to an aggregate containing the part.

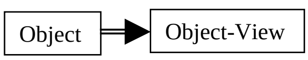

Fig. 5. Refined object representation with attached view concept

An alternative model would be to include appearance properties in the physical-object concept. We prefer a separate view concept to emphasise the distinction between physical and image objects.

An extended object concept is now modelled as follows:

Ax	boolean random variable representing the existence probability of a physical object A
A	vector-valued random variable representing simple properties of A
Bx	boolean random variable representing the existence probability of an object view
B	vector-valued random variables representing the properties of the view

Note that this representation may be easily extended to describe multiple views by several cameras or evidence by multimodal sensors.

Concrete evidence is considered as an instantiation of the random vector $\underline{B}$. As in other aggregates, $\underline{B}$ is related to $\underline{A}$ by a JPD $P(\underline{B} \mid \underline{A})$ where $\underline{A}$ and $\underline{B}$ are taken to include the existential variable Ax and Bx, respectively. This JPD allows to model the dependency of views from properties of the physical object.

Obviously, camera parameters also play a part in determining the relation between an object and its appearance. Hence $\underline{B}$ must be assumed to encompass such information. While this information is not a natural part of the physical object representation $\underline{A}$, the abstraction properties of a BCH require that this information must be channelled to $\underline{B}$ via $\underline{A}$ and its compositional parents. This is the price one has to pay for the tree-shaped propagation structure.

Given this extension of the BCH formalism to include view concepts, the dynamic state of a BCH during interpretation can be described as follows. Let $\{\underline{Z}0 \ldots \underline{Z}M\}$ be all concepts of the BCH and $\{\underline{Z}0 \ldots \underline{Z}M\} = \{\underline{X}1 \ldots \underline{X}N\ \underline{Y}1 \ldots \underline{Y}K\}$ where $\underline{Y}1 \ldots \underline{Y}K$ denote concepts with assigned evidences $\underline{y}1 \ldots \underline{y}k$. Then the JPD of the BCH is

$$P(\underline{X}1 \ldots \underline{X}N \mid \underline{Y}1=\underline{y}1 \ldots \underline{Y}K=\underline{y}k)$$

and the dynamic priors of object classes $\underline{X}i$ are given by the marginalisations

$$P(\underline{X}i \mid \underline{Y}1=\underline{y}1 \ldots \underline{Y}K=\underline{y}k)$$

To provide these dynamic marginalisations for all potential objects of a domain using the propagation procedure may seem a monstrous task, but the abstraction hierarchy allows to perform valid probabilistic inferences without considering every branch of the interpretation space in full detail, as will be shown in the following.

Consider a BCH structured as shown in Fig. 4, and a situation where some rectangular evidence $\underline{e}1$ is available which may be either a one-wing entrance door or a balcony door. For an optimal decision, we have to compare the posterior probabilities of One-Wing-Door-View and B-Door-View given the evidence.

Let us assume that this is the first evidence in the interpretation process, then the probabilities are immediately available from the initialised values of the BCH, and we can compare P(One-Wing-Door-View = $\underline{e}1$) and P(B-Door-View = $\underline{e}1$) and choose the most likely.

Assume now, that P(B-Door-View = $\underline{e}1$) is larger and we decide that $\underline{e}1$ is a Balcony-Door-View. In order to compute the effect of this decision, we have to propagate this decision only in the subtree of the BCH which may concern the next interpretation steps. For example, to compute the effect of the balcony-door decision on the other parts of the balcony, we can restrict propagation to the balcony subtree. Similarly, if we want to determine the effect on the probabilities of the entrance, we have to propagate within the facade aggregate, but not in a larger BCH of which the facade may be a part. This suggests that an efficient interpretation process will employ a strategy which one may call "lazy propagation": Effects of interpretation decisions are only propagated as far as needed, and preference decisions can be made by comparison within common subtrees, without knowledge of absolute probability values.

Another effort-saving idea is to stop propagation when changes are negligible. For example, if the position of the balcony door does not significantly affect the expected position of objects in another storey of the building, propagation may be restricted accordingly.

6. Propagation of Location Information

So far, it has been assumed that evidence is related to an object property represented by a random variable in our probabilistic aggregate representation. For example, evidence in terms of a rectangular image segment would instantiate random variables for width, height and colour of a door model. However, image analysis also provides evidence in terms of absolute image locations which cannot be directly related to the property of a single object within a BCH where location properties are specified relative to the object´s parent aggregate. Fig. 6 illustrates the situation for bounding-box aggregate representations in the building domain. Bold lines represent bounding box vertices used as local reference frames. Object locations are represented by offsets $\underline{d}i$ relative to the reference frame of the parent aggregate.

Note that mapping between image coordinates and scene coordinates (measured in the reference frame of the root node "scene" of the BCH) is assumed to be known. Hence evidence in terms of absolute image locations corresponds to locations in the "scene" reference frame. For objects modelled several levels down in the aggregate hierarchy, absolute image locations therefore correspond to the sum of the offsets between the local reference frame and the scene reference frame. We want to determine now which probability updates must be performed if evidence for a location property within the aggregate hierarchy is provided in absolute image coordinates.

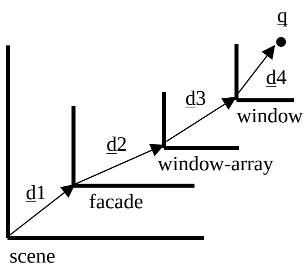

Fig. 6. Relative location specification of a location $\underline{q}$ in a 2D aggregate hierarchy.

To simplify the presentation, let us assume a 2D domain (such as the facades in the building domain) where image coordinates directly represent coordinates in the "scene" reference frame. Let $\underline{q}$ be a location of an object defined in a reference frame k levels deep in the compositional hierarchy. Let $\underline{q}1$ be the location of this point in absolute image coordinates and $\underline{D}1$ ... $\underline{D}k$ be the random variables representing the offsets between the nested reference frames. Then obtaining the absolute position of $\underline{q}$ amounts to obtaining the value for

$$\underline{q}1 = \underline{d}1 + \underline{d}2 + ... + \underline{d}k$$

Let $\underline{Q}1$ be the random variable defined by

$$\underline{Q}1 = \underline{D}1 + \underline{D}2 + ... + \underline{D}k$$

then after observing $\underline{q}1$, the updated joint distribution $P'(\underline{Z})$ (where Z is any set of nodes of the BCH) is defined by

$$P'(\underline{Z}) = P(\underline{Z} \mid \underline{Q}1=\underline{q}1)$$

To compute the update, we have to obtain the joint distribution of $\underline{D}1$, $\underline{D}2$ ... $\underline{D}k$ and then derive the distribution of the sum. Let us denote the aggregate descriptions containing $\underline{D}1$ ... $\underline{D}k$ by $P(\underline{B}1 \mid \underline{A}1)$, $P(\underline{B}2 \mid \underline{A}2)$ ... $P(\underline{B}k \mid \underline{A}k)$. Each internal property vector $\underline{B}i$ contains the offset $\underline{D}i$ to the next nested aggregate and its external properties $\underline{A}i+1$, among others, so by marginalisation one gets $P(\underline{D}1\ \underline{A}2 \mid \underline{A}1)$, $P(\underline{D}2\ \underline{A}3 \mid \underline{A}2)$ ... $P(\underline{D}k\ \underline{A}k+1 \mid \underline{A}k)$. From this, using the conditional independence assumption expressed in the abstraction Requirement 1, one gets the joint distribution

$$\begin{aligned} &P(\underline{A}1\ \underline{D}1\ \underline{A}2\ \underline{D}2\ ...\ \underline{A}k\ \underline{D}k\ \underline{A}k+1) \\ &\quad = P(\underline{A}1)\ \Pi\ P(\underline{D}1\ \underline{A}2 \mid \underline{A}1)\ ...\ P(\underline{D}k\ \underline{A}k+1 \mid \underline{A}k) \end{aligned} \qquad (\mathbf{11})$$

and by marginalisation $P(\underline{D}1, \underline{D}2 ... \underline{D}k)$. In the situation depicted in Fig. 6, $P(\underline{A}1)$ is a known factor since we assume that the properties of the scene reference frame are known. As shown below, however, $P(\underline{A}1)$ has to be taken into account in more general situations.

We can summarize now the steps which must be carried out to compute the probabilistic effect of observing $\underline{q}1$:

Absolute Location Update Procedure

A Determine $P(\underline{A}1)$, $P(\underline{D}1\ \underline{A}2 \mid \underline{A}1)$, $P(\underline{D}2\ \underline{A}3 \mid \underline{A}2)$ ... $P(\underline{D}k\ \underline{A}k+1 \mid \underline{A}k)$ for all aggregates containing $\underline{D}1$ ... $\underline{D}k$.

B Determine $P(\underline{D}1, \underline{D}2 ... \underline{D}k)$ using Eq. 11 and marginalising.

C Determine $P(\underline{Q}1 = \underline{D}1+\underline{D}2+ ... + \underline{D}k)$ and $P(\underline{D}1 ... \underline{D}k \mid \underline{Q}1)$ from $P(\underline{D}1 ... \underline{D}k)$. From this, one gets the updated distributions $P'(\underline{D}1)$ ... $P'(\underline{D}k)$ and the updated aggregate descriptions $P'(\underline{B}1 \mid \underline{A}1)$, ... $P'(\underline{B}k \mid \underline{A}k)$.

D From the updated aggregate descriptions, propagate the changes into all other aggregates using the regular propagation formulas Eqs. 9 and 10.

As more absolute image locations become known, the update procedure becomes more complex. Fig. 7 illustrates the general situation. $\underline{q}^{(0)}$... $\underline{q}^{(3)}$ represent locations in absolute coordinates whose probabilistic effect has already been incorporated into the BCH. $\underline{q}^{(4)}$ is a new location evidence in absolute coordinates.

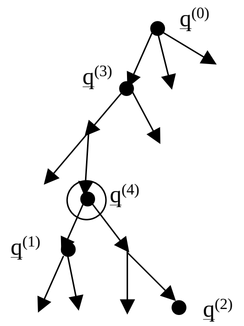

Fig. 7. Updating probabilities in a BCH for evidence in absolute coordinates

Observing $\underline{q}^{(4)}$ directly affects all ancestor aggregates along the path up to the nearest absolute position value, in the example $\underline{q}^{(3)}$, and all descendant aggregates along the paths down to the nearest absolute position values, in the example the aggregates on the paths from $\underline{q}^{(4)}$ to $\underline{q}^{(1)}$ and $\underline{q}^{(2)}$. Descendant branches without previous absolute position values remain unaffected. For each of these paths, the Absolute Location Update Procedure must be applied.

It is apparent that evidence in terms of absolute position information is against the grain of a hierarchical model. The larger the hierarchical distance between successive evidence, the higher is the computational cost for updating the aggregates on the paths between absolute values. Hence interpretation procedures should collect evidence in spatially coherent regions rather than shifting attention too often.

On the other hand, the computations of the updating procedure can be reasonably efficient as shown in the following section where the distributions are assumed to be multivariate Gaussians.

7. Propagation with Multivariate Gaussian Distributions

Gaussian densities are often acceptable approximations of unimodal probability distributions as long as quantities are centered around a most likely mean, and deviations beyond a certain distance from the mean are negligible. This is true for many random quantities which play a part in the composition of a facade, e.g. window sizes, distances between windows, heights of storeys etc. Simultaneously, it is obvious that none of these quantities are unlimited as required for true Gaussians. Hence Gaussian models can be taken seriously only within certain limits, say the range of -2σ ... $+2\sigma$ of each variable. The biggest advantage of Gaussian models, of course, is their compact representation in terms of two parameters for a univariate variable, and N parameters for the means and N^2 parameters for the covariances of N multivariate Gaussian variables.

Regarding the probability propagation in a BCH, Gaussians also offer considerable simplifications. First, marginalisations of a multivariate density can be obtained directly from the covariance matrix and the mean vector, and second, the parameters of conditional densities can be computed from a multivariate density by closed-form formulas.

It must be noted, however, that taxonomical branchings can only be modelled by Gaussians if the external properties of alternative specialisations have the same distribution or can be approximated by a single Gaussian (instead of a Gaussian mixture distribution). This is, for example, the case for different types of facades if they have a similar bounding box as external representation.

For multivariate Gaussian densities in a BCH, the propagation formulas can be specified as follows. To simplify the presentation, we describe how an arbitrary subset of multivariate Gaussians is updated. Let $\underline{G} = [\underline{C}\ \underline{D}]$ be a vector of Gaussian random variables where $\underline{D}$ is the subset whose distribution is changed by propagation. Before propagation, the distribution of $\underline{G}$ is

$$P(\underline{G}) = N(\mu_G, \Sigma_G)$$

where μ_G is the mean vector and Σ_G the covariance matrix. The partitions corresponding to $\underline{C}$ and $\underline{D}$, respectively, are denoted as shown:

$$\Sigma_G = \begin{vmatrix} \Sigma_C & \Sigma_{CD} \\ \Sigma_{CD}^T & \Sigma_D \end{vmatrix} \qquad \mu_G = \begin{vmatrix} \mu_C \\ \mu_D \end{vmatrix}$$

For a probability update, we assume that the distribution of $\underline{D}$ is changed to

$$P(\underline{D}') = N(\mu_D', \Sigma_D').$$

Then the new distribution $P'(\underline{G})$ is

$$P'(\underline{G}) = N(\mu_G', \Sigma_G')$$

with $$\Sigma_G' = \begin{vmatrix} \Sigma_C' & \Sigma_{CD}' \\ \Sigma_{CD}'^T & \Sigma_D' \end{vmatrix} \qquad \mu_G' = \begin{vmatrix} \mu_C' \\ \mu_D' \end{vmatrix}$$

where

$$\Sigma_C' = \Sigma_C - \Sigma_{CD}\, \Sigma_D^{-1}\, \Sigma_{CD}^T + \Sigma_{CD}\, \Sigma_D^{-1}\, \Sigma_D'\, \Sigma_D^{-1}\, \Sigma_{CD}^T \quad \textbf{(12)}$$

$$\Sigma_{CD}' = \Sigma_{CD}\, \Sigma_D^{-1}\, \Sigma_D' \quad \textbf{(13)}$$

$$\mu_C' = \mu_C + \Sigma_{CD}\, \Sigma_D^{-1}\, (\mu_D' - \mu_D) \quad \textbf{(14)}$$

Eqs. 12 to 14 can be derived by determining the resulting Gaussian distribution for

$$P'(\underline{G}) = P(\underline{C} \mid \underline{D})\, P(\underline{D}')$$

using the formulas for multivariate Gaussian conditionals:

$$P(C \mid D) = N(\mu_{C|D}, \Sigma_{C\{D})$$

with

$$\mu_{C|D} = \mu_C + \Sigma_{CD}\, \Sigma_D^{-1}\, (\underline{d} - \mu_D) \quad \textbf{(15)}$$

and

$$\Sigma_{C|D} = \Sigma_C - \Sigma_{CD} \Sigma_D^{-1} \Sigma_{CD}^T \qquad (\mathbf{16})$$

Eqs. 12 to 14 show that both upward and downward propagation for an aggregate with random variables $\underline{A}$ $\underline{B}1$ ... $\underline{B}N$ can be performed by fairly simple computations. For upward propagation, $\underline{D}$ represents a subset of $\underline{B}1$ ... $\underline{B}N$, for downward propagation $\underline{D}$ represents a subset of $\underline{A}$.

From the equations, we observe that the covariance matrix of $\underline{D}$ must be non-singular for its inverse to exist. This is the case if two conditions are fulfilled:

(i) The prior covariance matrices of the external aggregate variables $\underline{A}$ and the internal aggregate variable $\underline{B}1$ ... $\underline{B}N$ must each be non-singular. Note that this does not preclude deterministic mappings between $\underline{B}1$ ... $\underline{B}N$ and $\underline{A}$ which are natural for aggregate descriptions in a BCH.

(ii) Crisp evidence $\underline{D} = \underline{e}$ which is introduced by updating $\underline{D}$ with

$$P(\underline{D}) = N(\mu_D{}' = \underline{e}, \Sigma_D{}' = 0)$$

may not be updated again (the inverse of $\Sigma_D{}'$ does not exist, of course). This is naturally the case in a monotonic interpretation process where evidence may not be retracted.

Absolute Position Values

The computation of the probabilistic effect of absolute position values which was treated in Section 6, can also take a simplified form in the case of multivariate Gaussians. We assume that a new absolute value of an object is observed and hence offset chains in the aggregate hierarchy are constrained by knowledge of their sum. Let $\underline{D}1$ ... $\underline{D}k$ be such an offset chain between nested reference frames with known absolute position values at the beginning and at the end., and $\underline{d}_{1k}$ be the known value for the sum of the offsets. The essential goal then is to compute

$$P(\underline{D}1\ \underline{D}2\ ...\ \underline{D}k \mid \underline{D}1+\underline{D}2+ ... +\underline{D}k = \underline{d}_{1k})$$

Steps A and B of the update procedure call for the computation of the joint distribution of the corresponding aggregate descriptions. For Gaussians, it is convenient to describe an aggregate by the joint probability distribution of its internal and external properties P($\underline{A}$i $\underline{B}$i), specified my mean and covariance. To compute P($\underline{D}1$ $\underline{D}2$ ... $\underline{D}k$), we only need the components $\underline{D}$i and $\underline{A}$i+1 which are contained in $\underline{B}$i, and can reduce the aggregate covariance matrices accordingly. The covariance matrix $\Sigma_{Dk..D1}$ can be recursively determined as follows.

Let $\Sigma_{Dk..Di\ Ai}$ be the covariance matrix for $\underline{D}$k, $\underline{D}$k-1, ... $\underline{D}$i, $\underline{A}$i. We want to extend it to include $\underline{D}$i-1 and $\underline{A}$i-1. Fig. 8 illustrates this recursive situation. From the aggregate descriptions we know the covariance matrix of $\underline{A}$i, $\underline{A}$i-1 and $\underline{D}$i-1 (the box in the lower right corner). To compute the submatrices in the shaded areas, we exploit the conditional independence requirement for a BCH (Eq. 1). Note that for a multivariate Gaussian P($\underline{ABC}$) with covariance

$$\begin{vmatrix} \Sigma_A & \Sigma_{AB} & \Sigma_{AC} \\ \Sigma_{AB} & \Sigma_B & \Sigma_{BC} \\ \Sigma_{AC} & \Sigma_{BC} & \Sigma_C \end{vmatrix}$$

the conditional independence condition $P(\underline{A} \mid \underline{B}) = P(\underline{A} \mid \underline{BC})$ holds iff

$$\Sigma_{AC} = \Sigma_{AB}\, \Sigma_{B}^{-1}\, \Sigma_{BC}.$$

From this, the submatrices in the shaded area can be determined as

$$\Sigma_{(Di\text{-}1\ Ai\text{-}1)(Dk..Di)} = \Sigma_{Ai\ (Dk..Di)}\, \Sigma_{Ai}^{-1}\, \Sigma_{(Di\text{-}1\ Ai\text{-}1)\ Ai}$$

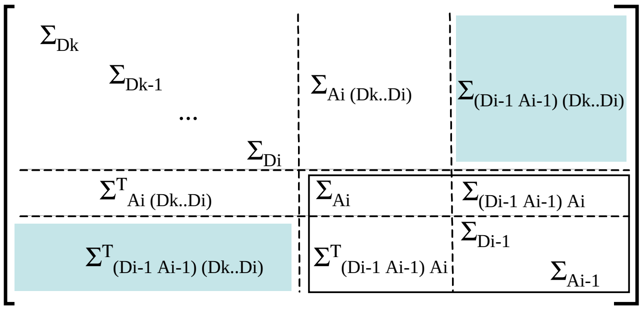

Fig. 8. Recursive computation of the covariance matrix of $P(\underline{D}1\ \underline{D}2\ ...\ \underline{D}k)$ - see text.

The recursive step is concluded by deleting the row and column for $\underline{A}i$, resulting in $\Sigma_{Dk..Di\text{-}1\ A\text{-}1}$. After completion of all recursive steps, the covariance of $P(\underline{D}1\ \underline{D}2\ ...\ \underline{D}k)$ is determined, and Step C and D of the Absolute Position Update Procedure can be carried out. The updated mean $\mu'_{D1..Dk}$ and covariance $\Sigma'_{D1..Dk}$ of

$$P'(\underline{D}1\ ...\ \underline{D}k) = P(\underline{D}1\ ...\ \underline{D}k \mid \underline{D}1+...+\underline{D}k)$$

can be determined from the covariance matrix $\Sigma_{D1..Dk}$ based on the sums Σ_{S1} ... Σ_{Sk} of pairwise covariances as follows. Let

$$\Sigma_{Si} = \mathrm{sum}(\Sigma_{D1\ Di}\ \Sigma_{D2\ Di}\ ...\ \Sigma_{Di}\ ...\ \Sigma_{Dk\ Di})$$

then

$$\Sigma'_{D1..Dk} = \Sigma_{D1..Dk} - [\Sigma_{S1}\ ...\ \Sigma_{Sk}]\ [\mathrm{sum}(\Sigma_{S1}\ ...\ \Sigma_{Sk})]^{-1}\ [\Sigma_{S1}\ ...\ \Sigma_{Sk}]^{T} \qquad \mathbf{(17)}$$

$$\mu'_{D1..Dk} = \mu_{D1..Dk} + [\Sigma_{S1}\ ...\ \Sigma_{Sk}]\ [\mathrm{sum}(\Sigma_{S1}\ ...\ \Sigma_{Sk})]^{-1}\ (\underline{d}_{1k} - \mu_{1k}) \qquad \mathbf{(18)}$$

The update formulas Eqs. 17 and 18 have been derived for vector-valued offsets $\underline{d}i$ which can be 3D vectors for general aggregate models, 2D vectors for special domains such us the facade domain, or scalars if only one dimension has been observed, e.g. the horizontal position of a facade boundary.

8. Conclusions and Outlook

We have presented a probabilistic framework for computing dynamic priors based on probabilistic dependencies between objects embedded in a compositional hierarchy. By requiring certain abstraction properties, probability changes induced by evidence can be propagated along the tree-shaped structure of the compositional hierarchy, and, in the case of disjunctive specialisations, also along taxonomical branchings. A factorisation theorem similar to the Bayesian Network factorisation formula has been derived which generalises a conventional Bayesian Network representation of a compositional hierarchy by allowing arbitrary probabilistic dependencies between the parts of an aggregate.

Implementations of the probabilistic framework are currently underway. In one approach, objects are modelled by location and bounding box parameters, and Gaussian distributions will be assumed. In this case, the abstraction function $\underline{f}a$ maps the bounding-box parameters of the parts into the resulting bounding-box parameters of the aggregate as a whole. If this mapping is linear (which is the case for many realistic aggregates), Gaussian distributions for primitive objects map into Gaussians at higher compositional levels, and probability updates can be performed as shown in Section 7.

Another approach is to allow arbitrary distributions. However, preliminary work shows that the probability tables for realistic aggregates tend to be very large, and modelling has to be done with special consideration of this aspect.

References

1. Russell, S., Norvig, P.: Artificial Intelligence - A Modern Approach. Prentice-Hall (2003)
2. Neumann, B., Moeller, R.: On Scene Interpretation with Description Logics. In: Cognitive Vision Systems, H.-H. Nagel and H. Christensen (Eds.), Springer, LNCS 3948, pp. 247-275 (2006)
3. Rimey, R.D.: Control of Selective Perception Using Bayes Nets and Decision Theory. Dissertation, Univ. of Rochester, TR 468 (1993)
4. Laskey, K., Mahoney, S., Wright, E.: Hypothesis Management in Situation-Specific Network Construction. In Proc. 17th Conf. on Uncertainty in Artificial Intelligence, Morgan Kaufmann, pp. 301-309 (2001)
5. Gyftodimos, E., Flach, P.: Hierarchical Bayesian Networks: A Probabilistic Resoning Model for Structured Domains. In E. de Jong and T. Oates (eds.), Proc. Workshop on Development of Representations (ICML-2002), pp. 23-30 (2002)