Temporal Predictions with Bayesian Compositional Hierarchies

Bernd Neumann

Technical Note: FBI-HH-M-343/10

Universität Hamburg, Department Informatik,
Arbeitsbereich Kognitive Systeme

January 2010

Zusammenfassung

In dieser Mitteilung wird ein neuer Ansatz zur Modellierung und Auswertung probabilistischer Abhängigkeiten in kompositionellen Hierarchien für modellbasierte Szeneninterpretation vorgestellt. Mit einer "Bayes'schen Kompositionellen Hierarchie" (Bayesian Compositonal Hierarchy, BCH) werden Zusammenhänge innerhalb von objektzentrierten Aggregaten in uneingeschränkter Form repräsentiert, zwischen Aggregaten jedoch nur entlang der kompositionellen Beziehungen. Dadurch können probabilistische Inferenzen in Szenen sehr effizient berechnet werden. Die Prädiktionsleistungen einer BCH werden mit alternativen Modellen verglichen (reine Bayes-Netze, uneingeschränkte Verbundwahrscheinlichkeiten) und an Beispielen evaluiert. Als praktische Anwendung wird die Überwachung einer Flugzeugabfertigung vorgestellt. Mithilfe einer BCH können Vorhersagen über den zu erwartenden zeitlichen Ablauf einer Flugzeugabfertigung aus bereits vorliegenden zeitlichen Daten schritthaltend angepasst werden.

Abstract

In this note I describe a novel approach to modelling and exploiting probabilistic dependencies in compositional hierarchies for model-based scene interpretation. I present Bayesian Compositional Hierarchies (BCHs) which capture all probabilistic information about the objects of a compositional hierarchy in object-centered aggregate representations. BCHs extend typical Bayesian Network models by allowing arbitrary probabilistic dependencies within aggregates, yet providing efficient inference procedures. New closed-form solutions are presented for inferences in a multivariate Gaussian BCH. Results are presented comparing a BCH with existing methods (pure Bayesian Networks, unrestricted Joint Probability Distributions). Monitoring aircraft service operations is presented as a practical application. It is shown that predictions about the expected temporal development of service operations can be generated dynamically from available temporal data.

Temporal Predictions with Bayesian Compositional Hierarchies

Bernd Neumann

Abstract. In this note I describe a novel approach to modelling and exploiting probabilistic dependencies in compositional hierarchies for model-based scene interpretation. I present Bayesian Compositional Hierarchies (BCHs) which capture all probabilistic information about the objects of a compositional hierarchy in object-centered aggregate representations. BCHs extend typical Bayesian Network models by allowing arbitrary probabilistic dependencies within aggregates, yet providing efficient inference procedures. New closed-form solutions are presented for inferences in a multivariate Gaussian BCH. Results are presented comparing a BCH with existing methods (pure Bayesian Networks, unrestricted Joint Probability Distributions). Monitoring aircraft service operations is presented as a practical application. It is shown that predictions about the expected temporal development of service operations can be generated dynamically from available temporal data.

This work was partially supported by the EC, Grant 214975, Project Co-Friend

1. Introduction

Scene interpretation can be roughly defined as the task of assigning meaning to visual scenes beyond single-object recognition. From previous work [1, 2, 3, 4] it is evident that the compositional structure of higher-level entities is essential both for a well-structured organization of the model base and for an effective guidance of the interpretation process. Furthermore, judging from the success of probabilistic models in many subareas of Computer Vision, a sophisticated uncertainty management seems undispensable also for high-level scene interpretation.

In this article we present a new approach to probabilistic scene interpretation which combines the compositional structure of a model base with a reasonably general probabilistic framework, without presupposing a particular interpretation strategy. The main idea is to provide each node of the compositional hierarchy (called aggregate) with an expressive probabilistic description correlating its parts in a realistic way, yet preserving a probabilistic tree structure between the aggregates to enable efficient inferences. In its general form, the structure is called a Bayesian Compositional Hierarchy (BCH) because of its similarity to Bayesian Networks [5]. The main service of a BCH is to propagate the probabilistic effect of interpretation steps and provide dynamic priors for evidence assignment during an interpretation process. The focus of this contribution is on a BCH with multivariate Gaussian distributions which gives rise to particularly effective probability propagation procedures.

In the next section, we define a BCH and the probabilistic inference procedures supported by a BCH. In Section 3, we show how these models can be applied to spatial and temporal reasoning in scene interpretation. In Section 4, we derive closed-form solutions for a BCH where all aggregates have multivariate normal distributions. We then compare our aggregate models to existing Bayesian Network models and demonstrate their superior accuracy in typical prediction tasks.

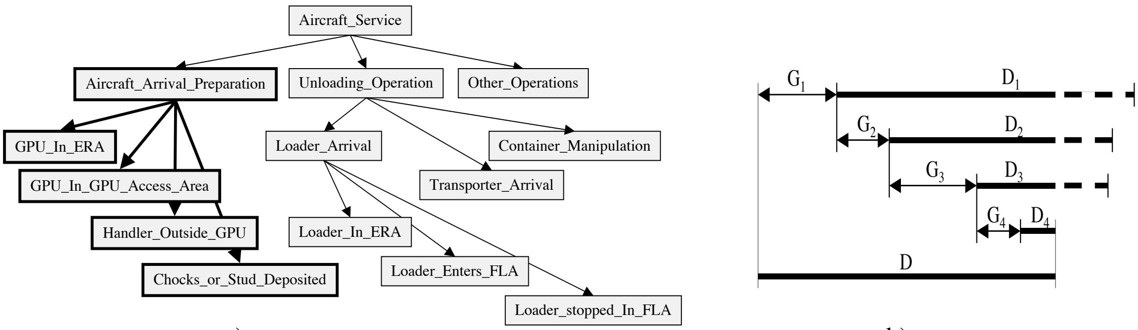
2. Bayesian Compositional Hierarchies

In this section we will briefly introduce Bayesian Compositional Hierarchies (BCHs) as probabilistic structures for scene interpretation. A compositional hierarchy is defined to be a tree of aggregates, each representing a meaningful entity composed of parts, except for the leaves of the hierarchy which are primitive aggregates. For each aggregate, probabilistic relations between parts can be modelled as accurately as desired. By postulating conditional-independence properties between aggregates, efficient inference procedures can be attained similar to tree-shaped Bayesian Networks.

An aggregate is represented by the following random vectors and a corresponding joint probability density $P(\underline{A}\ \underline{B}_1 .. \underline{B}_K\ \underline{C})$, understrokes denoting a vector:

$\underline{A}$	header of the aggregate, represents the aggregate as a whole
$\underline{B}_1 \dots \underline{B}_K$	headers of the parts of the aggregate
$\underline{C}$	properties relating the parts to the aggregate and to each other

The intuition is that the header of an aggregate is an external description which abstracts from details of the parts. As an example, consider the aggregate "Aircraft-Arrival-Preparation" in a compositional hierarchy for aircraft services (Fig. 1a). The aggregate describes four activities which together constitute the arrival preparation. The overall duration D is taken as the aggregate header, the parts are primitive with durations D_k as part headers and “gaps” G_k as properties relating the parts to each other (Fig. 1b). The GPU (Ground Power Unit) first enters the ERA (Entrance Restricted Area) and then the GAA (GPU Access Area). After a while the handler gets out and deposits chocks for the aircraft, concluding the arrival preparation.



a) b)

Fig. 1. (a) Compositional hierarchy of aircraft service operations. (b) Temporal structure of the aggregate “Aircraft-Arrival-Preparation”

We now introduce conditional independence requirements for a compositional hierarchy which reflect the intuition that the probabilistic influence of an aggregate is solely exerted by its header to the above and by the part headers to the below. Let $\underline{Z}$ represent an aggregate:

1. Given $\underline{Z}$, the properties of the successors of $\underline{Z}$ do not depend on the parent of $\underline{Z}$.
2. Given $\underline{Z}$, the properties of the successors of $\underline{Z}$ do not depend on the siblings of $\underline{Z}$.
3. Given the siblings of $\underline{Z}$, the successors of the siblings are statistically independent.

Given these abstraction requirements, evidence propagation will simply have to follow the structure of the compositional hierarchy.

Exploiting these abstraction requirements, a factorization formula for the JPD of a complete compositional hierarchy can be derived [5]. Let $\underline{A}^{(k)}$, k = 1 ... M, be all objects of the compositional hierarchy (represented by their headers) and $\underline{A}^{(1)}$ the root of the hierarchy. Let $P(\underline{A}^{(k)}\ \underline{B}^{(k)}\ \underline{C}^{(k)})$ be the JPD of the aggregate with header $\underline{A}^{(k)}$, $\underline{B}^{(k)} = \underline{B}_1^{(k)} \ldots \underline{B}_N^{(k)}$ its part headers, and $\underline{C}^{(k)}$ its relational part properties. Then

$$P(\underline{A}^{(1)} \ldots \underline{A}^{(M)}) = P(\underline{A}^{(1)}) \prod_{k=1 \ldots M} P(\underline{B}^{(k)} \underline{C}^{(k)} \mid A^{(k)}) \qquad \textbf{(1)}$$

Because of the remarkable similarity to the well-known Bayesian Network factorization formula, we call compositional hierarchies meeting the three abstraction requirements "Bayesian Compositional Hierarchies" (BCHs). The BCH factorization formula states that all probabilistic inferences can be carried out solely based on aggregate descriptions in terms of the JPDs of part headers and relational part properties, given aggregate headers. Furthermore, the root probability must be known, which will be 1 if the domain covered by the compositional hierarchy is known.

We now show how a BCH can be used for probabilistic inferences in a scene interpretation task. A typical interpretation step is to assign evidence $\underline{e}$ to one of the hierarchy leaves. Given several choices, the BCH provides priors in terms of the marginal probabilities. Let us assume that $\underline{e}$ is assigned to an aggregate $\underline{Z}^{(k)}$ such that the JPD of $\underline{Z}^{(k)}$ becomes $P'(A^{(k)}\ B^{(k)}\ C^{(k)})$, restricting the properties of $\underline{Z}^{(k)}$ according to the crispness of the evidence. To update the rest of the BCH, we propagate the change along the compositional structure of the hierarchy.

Propagating up: Let $\underline{Z}^{(k)}$ be the changed aggregate with JPD $P'(\underline{A}^{(k)}\ \underline{B}^{(k)}\ \underline{C}^{(k)})$. We marginalize to obtain $P'(\underline{A}^{(k)})$. $\underline{A}^{(k)}$ is simultaneouly the header part $\underline{B}_i^{(m)}$ of the parent aggregate $Z^{(m)}$ and hence $P'(\underline{B}_i^{(m)}) = P'(\underline{A}^{(k)})$. The parent aggregate is now updated as follows:

$$P'(\underline{A}^{(m)} \underline{B}_1^{(m)} \ldots \underline{B}_M^{(m)} \underline{C}^{(m)}) = P(\underline{A}^{(m)} \underline{B}_1^{(m)} \ldots \underline{B}_M^{(m)} \underline{C}^{(m)}) P'(\underline{B}_i^{(m)}) / P(\underline{B}_i^{(m)}) \qquad \textbf{(2)}$$

Propagating down: Let $\underline{Z}^{(k)}$ be the changed aggregate with JPD $P'(\underline{A}^{(k)}\ \underline{B}_1^{(k)} \ldots \underline{B}_N^{(k)}\ C^{(k)})$. We marginalize to obtain $P'(\underline{B}_i^{(k)})$, i = 1 .. N. Each $\underline{B}_i^{(k)}$ is simultaneously a header $\underline{A}^{(n_i)}$ of a child aggregate $\underline{Z}^{(n_i)}$ and hence $P'(\underline{A}^{(n_i)}) = P'(\underline{B}_i^{(k)})$. A child aggregate $\underline{Z}^{(n_i)}$ is now updated as follows:

$$P'(\underline{A}^{(n_i)} \underline{B}^{(n_i)} \underline{C}^{(n_i)}) = P(\underline{A}^{(n_i)} \underline{B}^{(n_i)} \underline{C}^{(n_i)}) P'(\underline{A}^{(n_i)}) / P(\underline{A}^{(n_i)}) \qquad \textbf{(3)}$$

Note that the assignment of crisp evidence can also be modelled by Eqs. 2 and 3.

Summarizing, a BCH provides a probabilistic description of a compositional hierarchy with three advantages: (i) All probabilistic information is captured in aggregate descriptions, preserving object-centered knowledge representation, (ii) arbitrary probabilistic dependencies can be modelled within an aggregate, and (iii) probabilities can be efficiently updated by propagation along the hierarchical compositional structure.

We now show how a BCH can be used to model compositional hierarchies of simple 2D shapes and - with one dimension omitted - of the temporal structure of activities.

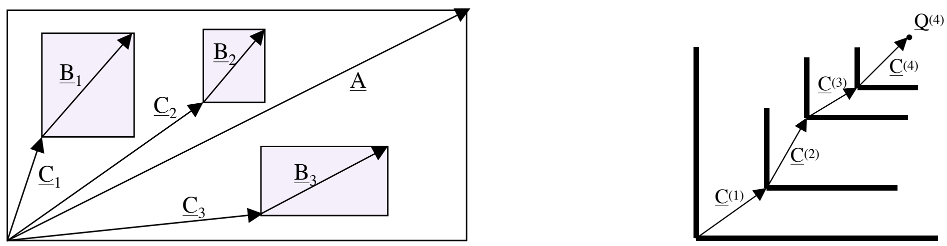
3. Spatial and Temporal Aggregate Models

In scene interpretation, spatial and temporal information play a prominent part, hence it is useful to consider probabilistic spatial and temporal models in more detail.

We first present box models as an instantiation of a 2D spatial BCH. Hierarchical box models are generally popular in Computer Vision and have been applied, for example, to facade interpretation [6]. The facade of a building can be modelled in terms of constituent structures like balconies, window arrays or entrance ensembles, which in turn can be further decomposed into parts such as doors, windows, canopies etc.

A box model for an aggregate is structured as shown in Figure 2a. Given a compositional hierarchy of aggregates, a BCH can be obtained by assigning JPDs $P(\underline{B}_k\ \underline{C}_k \mid \underline{A}_k)$ to each box model.

In Eqs. 2 and 3, we have considered probability update rules where evidence affects the properties of a *single* aggregate. Unfortunately, this is not the case when absolute location evidence is provided for a hierarchy of boxes with relative location variables. Absolute locations are typically given in scene or image coordinates corresponding to the reference frame of the root of the hierarchy. If provided for an object several levels down the hierarchy, the evidence constrains the chain of offsets between the nested reference frames as shown in Figure 2b.



a) b)

Fig. 2. (a) Box model for aggregates in a BCH. $\underline{A}$ represents the aggregate as a whole, $\underline{B}_1$ .. $\underline{B}_3$ represent parts and $\underline{C}_1$ .. $\underline{C}_3$ determine the position of the parts relative to the aggregate reference frame. (b) Offset chain corresponding to absolute location evidence in a compositional hierarchy

In order to maintain conditional independence between aggregates along the offset chain, we have to introduce additional random variables $\underline{Q}^{(k)} = \underline{C}^{(1)} + .. + \underline{C}^{(k)}$ in each aggregate along the offset chain, each representing the sum of the offsets from the root to this aggregate. In consequence, an observation of $\underline{Q}^{(k)}$ can be propagated with the propagation rules given in Equations 2 and 3.

It is easy to see that the box model described above can be applied to temporal aggregates by omitting one dimension. Dedicated variables for absolute time points are also required analogous to absolute location variables. Thus, probabilistic temporal properties of the aircraft service domain used as an example in Fig. 1 can be modelled by a BCH. The BCH can then be used for dynamically updated temporal predictions, for example about the duration of future activities or the conclusion of a service.

4. Gaussian Bayesian Compositional Hierarchies

A BCH for a small domain (such as aircraft service operations) may encompass a considerable number of probability distributions which must be maintained for each interpretation step. Even for moderate resolutions of spatial and temporal properties, the demands on storage and computation time may be excessive. We have therefore investigated the use of multivariate Gaussian distributions for implementing a BCH. Gaussian models may be a good approximation for sufficiently symmetric unimodal distributions in a range corresponding to -2σ .. $+2\sigma$ of a Gaussian distribution. It will be shown now that all propagation operations can be expressed by closed-form formulas which give rise to highly efficient propagation procedures.

Let $\underline{G} = [\underline{E}\ \underline{F}]$ be a vector of Gaussian random variables representing an aggregate. Let $\underline{F}$ be the subset whose distribution is changed by evidence or incoming propagation. $\underline{F}$ can be the aggregate header in the case of downward propagation or a part header in the case of upward propagation. We want to compute the effect of the changed distribution of $\underline{F}$ on $\underline{G}$. Before propagation, the distribution of $\underline{G}$ is $P(\underline{G}) = N(\underline{\mu}_G\ \Sigma_G)$ where $\underline{\mu}_G$ is the mean vector and Σ_G the covariance matrix. The partitions corresponding to $\underline{E}$ and $\underline{F}$, respectively, are denoted as shown:

$$\Sigma_G = \begin{bmatrix} \Sigma_E & \Sigma_{EF} \\ \Sigma_{EF}^T & \Sigma_F \end{bmatrix} \qquad \underline{\mu}_G = \begin{bmatrix} \underline{\mu}_E \\ \underline{\mu}_F \end{bmatrix}$$

For a probability update, we assume that the distribution of $\underline{F}$ is changed to $P'(\underline{F}) = N(\underline{\mu}'_F, \Sigma'_F)$. Then the new distribution is $P'(\underline{G}) = N(\underline{\mu}'_G, \Sigma'_G)$ with

$$\Sigma'_G = \begin{bmatrix} \Sigma'_E & \Sigma'_{EF} \\ \Sigma'^T_{EF} & \Sigma'_F \end{bmatrix} \qquad \underline{\mu}'_G = \begin{bmatrix} \underline{\mu}'_E \\ \underline{\mu}'_F \end{bmatrix} \quad \text{where}$$

$$\Sigma'_E = \Sigma_E - \Sigma_{EF}\,\Sigma_F^{-1}\,\Sigma_{EF}^T + \Sigma_{EF}\,\Sigma_F^{-1}\,\Sigma'_F\,\Sigma_F^{-1}\,\Sigma_{EF}^T \tag{4}$$

$$\Sigma'_{EF} = \Sigma_{EF}\,\Sigma_F^{-1}\,\Sigma'_F \tag{5}$$

$$\underline{\mu}'_E = \underline{\mu}_E + \Sigma_{EF}\,\Sigma_F^{-1}(\underline{\mu}'_F - \underline{\mu}_F) \tag{6}$$

To the best of our knowledge, the Gaussian updating rules in Eqs. 4 to 6 are new results. They can be derived by determining the resulting Gaussian distribution for $P'(\underline{G}) = P(\underline{E} \mid \underline{F})\,P'(\underline{F})$ using the formulas for multivariate Gaussian conditionals.

It is evident that both upward and downward propagation for an aggregate with random variables $\underline{A}$ $\underline{B}_1$ ... $\underline{B}_K$ $\underline{C}$ can be performed by fairly simple matrix computations. From the equations, we observe that the covariance matrix of $\underline{F}$ 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 aggregate header $\underline{A}$ and of the internal aggregate variables $\underline{B}_1$ ... $\underline{B}_K$ and $\underline{C}$ must each be non-singular. Note that this does not preclude deterministic mappings between $\underline{B}_1$ ... $\underline{B}_K$ and $\underline{A}$ which are natural for aggregate descriptions in a BCH.

(ii) Crisp evidence $\underline{F} = \underline{f}$ which is introduced by updating $P(\underline{F})$ with $P'(\underline{F}) = N(\underline{\mu}_F{}'=\underline{f},\ \Sigma_F{}'=0)$ may not be updated again (the inverse of $\Sigma_F{}'$does not exist, of course). This is naturally the case in a monotonic interpretation process where evidence may not be retracted.

Note that the box model and its temporal version presented in Section 3 can be modelled with multivariate Gaussians, if the aggregate header $\underline{A}$ is a linear combination of part headers $\underline{B}$ and relational part properties $\underline{C}$, for example, the sum of part dimensions and distances.

5. Evaluation

To demonstrate the properties of a BCH as compared to standard Bayesian Networks (BNs), we now compare three kinds of aggregate models for simple estimation tasks based on the temporal structure of "Aircraft Arrival Preparation" (Fig. 1), neglecting D_1 ... D_3 which are irrelevant here.

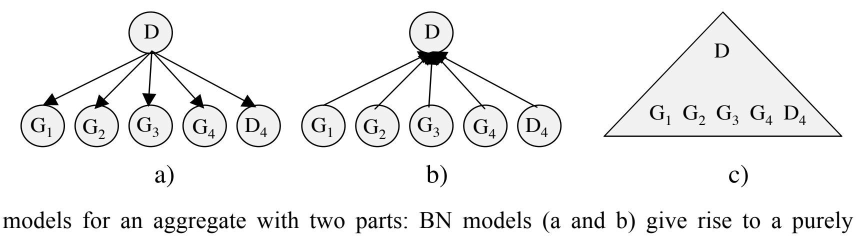


Fig. 3. Three models for an aggregate with two parts: BN models (a and b) give rise to a purely tree-shaped BN of a compositional hierarchy. The BCH aggregate model (c) allows an arbitrary aggregate JPD while preserving an overall tree-shaped compositional hierarchy.

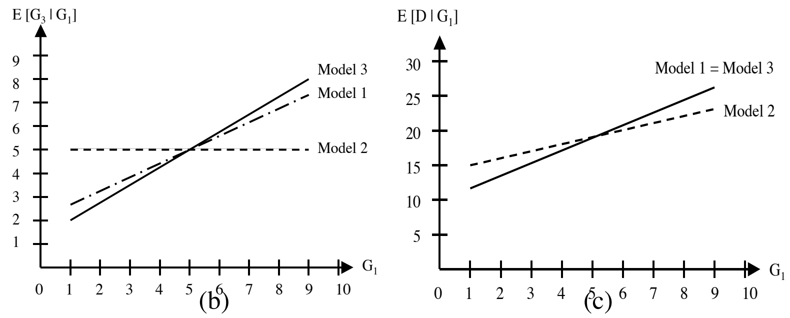
Model 1 (Fig. 3a) is a Bayesian Network (BN) model where an aggregate "causes" its parts, as proposed e.g. by [7, 8]. Model 2 (Fig. 3b) is a BN model where parts "cause" an aggregate. This is a plausible BN when parts are the primary meaningful entities. Model 3 (Fig. 3c) is a BCH aggregate model with an unrestricted JPD.
All random variables are modelled by a JPD P_Z of Gaussians with mean and covariance shown in Fig. 4a.. Only a single nonzero covariance value has been assumed for internal properties, correlating G_1 and G_3. To represent the aggregate as Model 1, the distributions $P(D)$ and $P(G_1|D)$ ... $P(D_4|D)$ were computed from P_Z, similarly $P(G_1)$ ... $P(D_4)$ and $P(D|G_1$ ... $D_4)$ were computed for Model 2. The three models were compared regarding the

prediction of G_3 and D after observing G_1 with values differing from the mean of G_1 up to twice the standard deviation.

μ	G_1	G_2	G_3	G_4	D_4	D
	5	2	5	5	2	19

Σ	G_1	G_2	G_3	G_4	D_4	D
G_1	4	0	3	0	0	7
G_2	0	1	0	0	0	1
G_3	3	0	4	0	0	7
G_4	0	0	0	4	0	4
D_4	0	0	0	0	1	1
D	7	1	7	4	1	20

(a)

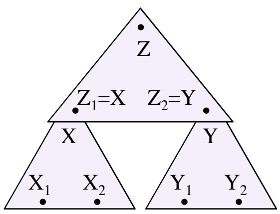


(b) (c)

Fig. 4. (a) Mean and covariance of multivariate Gaussian JPD for temporal structure "Aircraft Arrival Preparation" (s. Fig. 1). (b) Predictions of G_3 given G_1 based on the aggregate models presented in Fig. 3. (c) Predictions of D given G_1, based on the same three aggregate models.

As to be expected, the prediction error of BN models can be significant. For predicting G_3 the error is up to 28% for Model 1 and 150% for Model 2 (which always estimates the mean of G_1). For predicting D, Model 1 is error-free, but Model 2 deviates from the true estimate by up to 25%.

A BCH owes efficient propagation to the conditional independence requirements presented in Section 2. It is interesting to evaluate the estimation error of a BCH model if the independence requirements are not valid. Fig. 5a shows a simple BCH with three aggregates consisting of four primitive parts (Gaussian random variables X_1, X_2, Y_1, Y_2) and aggregate headers $X=X_1+X_2$, $Y=Y_1+Y_2$, $Z=Z_1+Z_2$. Fig. 5b shows mean and covariance of the primitive parts, with σ* as a variable parameter to model unequal individual dependencies between parts of different aggregates. In the BCH, these dependencies can only be represented by a single parameter, the covariance $\sigma_{Z1,Z2}$ in the parent aggregate.

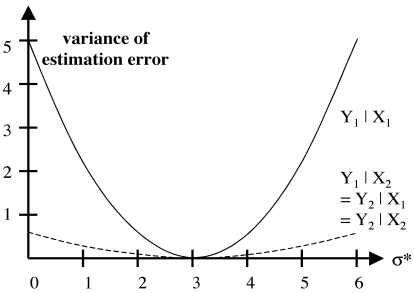


a)

μ	X_1	X_2	Y_1	Y_1
	10	10	10	10

Σ	X_1	X_2	Y_1	Y_1
X_1	25	5	σ*	3
X_2	5	25	3	3
Y_1	σ*	3	25	5
Y_1	3	3	5	25

b)



c)

Fig. 5. (a) Simple BCH structure with Gaussian variables X_1, X_2, Y_1, Y_2, $Z_1=X_1+X_2$, $Z_2=Y_1+Y_2$. (b) Mean and covariance of primitive parts. (c) Variance of estimation error using BCH approximation of original covariances for different choices of σ*.

An analysis reveals that for a given covariance matrix there is an optimal choice of $\sigma_{Z1\text{-}Z2}$ minimizing the error incurred by estimating Y_1 or Y_2 given X_1 or X_2 using the BCH instead of using directly the covariances of the primitive parts. Furthermore, the mean of these errors will always be zero. The variances of the errors depend very much on how well individual part covariances can be approximated by aggregate headers representing an abstraction in terms of the sum of the parts. For the example, the variances of estimates Y_1 or Y_2 given X_1 or X_2 are plotted against the parameter σ* in Fig. 5c. It can be seen that the errors remain small as long as the individual variations of inter-aggregate covariances are small.

Using the BCH structure, a probabilistic temporal model has been constructed for the aircraft service hierarchy shown in Fig. 1a. Each activity, primitive or aggregate, has been modelled by its duration D and offset C to the aggregate reference frame, comprising a total of 18 random variables. Estimations of various kinds have been performed, e.g. "Aircraft arrival preparation has begun at t=20, when is unloading likely to begin? When will the aircraft be ready for departure?" It should be clear, however, that the main intended use of a BCH is to support ambiguous interpretation steps where context-dependent priors - computed by a BCH - guide the decisions.

The computational effort using Gaussian distributions is very small and not worthy a discussion, it will easily scale to thousands of variables. Regarding space complexity, the BCH shows its advantages as space requirements grow linearly with the number of aggregates. This has to be compared with the space requirements of complete covariance matrices which grow with the square of the number of variables.

6. Discussion and Outlook

The BCH presented in this paper is a step towards probability-guided high-level scene interpretation where object-oriented knowledge at multiple abstraction levels must be brought to bear to provide a probabilistic context at each interpretation step. A BCH is in the spirit of tree-shaped BN models [7, 8, 9] but extends their expressivity significantly by allowing arbitrary distributions within aggregates, yet preserving efficient inference procedures. A BCH provides a general framework for modelling probabilistic information for a compositional hierarchy, but offers particularly elegant solutions in the case of multivariate Gaussian probabilities, as presented in this paper.

The evaluation has shown that the gain of accuracy compared to purely tree-shaped networks can be significant, while the loss of accuracy compared to completely unrestricted networks is tolerable as long as the abstractions in the compositional hierarchy are justified.

A BCH for probabilistic temporal dependencies will be employed for aircraft service operations in the joint project Co-Friend.

References

[1] Neumann, B.; Weiss, T.: Navigating through logic-based scene models for high-level scene interpretations, 3rd International Conference on Computer Vision Systems - ICVS 2003, Springer, 2003 , 212-222

[2] Arens, M., Ottlik, A. Nagel, H.-H.: Using Behavioral Knowledge for Situated Prediction of Movements. In: Proc. KI-2004, Springer, 2004, 141–155

[3] Bremond, F., Thonnat, M., Zuniga, M.: Video understanding framework for automatic behavior recognition. Behavior Research Methods 3(38), 2006, 416–426

[4] Zhu, S.C., Mumford, D.: A Stochastic Grammar of Images. In: Foundations and Trends in Computer Graphics and Vision, Vol.2, No.4, 2006, 259-362

[5] Neumann, B.: Bayesian Compositional Hierarchies - A Probabilistic Structure for Scene Interpretation. Report FBI-HH-B-282/08, Dep. of Informatics, Univ. of Hamburg,, 2008

[6] Hotz, L.; Neumann, B.: Scene Interpretation as a Configuration Task, Künstliche Intelligenz, 3/2005, BöttcherIT Verlag, Bremen, 59-65

[7] Binford, T.O., Levitt, T.S., Mann, W.B.: Bayesian inference in model based machine vision. In: Uncertainty in AI, Vol. 3, 1989, 73-94

[8] Rimey, R.D.: Control of Selective Perception Using Bayes Nets and Decision Theory. Dissertation, Univ. of Rochester, TR 468, 1993

[9] Todorovic, S., Nechyba, M.C.: Interpretation of complex scenes using dynamic tree-structure Bayesian networks. Computer Vision and Image Understanding, Vol. 106 No. 1, Elsevier, 2007