

視覚情報から抽出した文脈を用いた情報アクセス・システムの提案

An Information Access System based on Context extracted from Visual Data

佐藤 真^{*1} 田中 克明^{*2} 赤石 美奈^{*2} 堀 浩一^{*2}
 Makoto Sato Katsuaki Tanaka Mina Akaishi Koichi Hori

^{*1}東京大学大学院工学系研究科航空宇宙工学専攻
 Department of Aeronautics and Astronautics, School of Engineering, University of Tokyo

^{*2}東京大学先端科学技術研究センター
 Research Center for Advanced Science and Technology, University of Tokyo

In this paper, we propose a information access system based on context extracted from visual data. The system computes the co-occurrence relation of objects in one's vision based on Narrative Structure Model and accesses the knowledge base with the relation as query.

1. はじめに

本研究は、議事録などの蓄えられた設計情報を処理することにより、新しい設計手法や問題点を見出す機会を増やすことを目的とし、そのためのシステムを提案するものである。

人間は、日常的に、自らの知識や経験を利用しながら、ある事象を解釈し、問題として表現し、それに対する方法論を用いる。自分の知識や経験が足りないときは、文献を調べる、他人に聞くなどの様々な対処法を考える。しかし、思いつく対処法には限度がある。この原因のひとつに入手できるデータは莫大で人間がすべてを把握することは困難であるということがある。人間は莫大なデータを記録、処理、伝達するとき、意識的に経験や知識に基づいてデータを取捨選択し、情報を整理するが、それ以前の知覚の段階では、無意識のうちに多くのデータを削っている。本研究では、対処法の選択肢の幅を広げるために、そもそも意識すらされていなかった視覚情報に着目する。

一般的に、冗長なデータや混沌としたデータから情報を得ることはもちろん困難であるが、秩序を持ったデータからでも、そのままの状態では意味のある情報を見出しにくい。データに対する多様な意味づけが、データに価値を付与する。よって、状況に応じた情報を産み出すためには、その文脈に沿ってデータを再構成することが必要である。

これらを踏まえて、本研究では、以下のような情報アクセス・システムを提案する(図1)。まず、設計の場において視界にあって失われそうな情報、たとえば、意識していたにも関わらず忘れてしまったことや意識すること自体が少ないことをあえて指摘するために、設計過程の状況をカメラで機械的に記録し、視覚的なデータを収集する。つぎに、収集した視覚的なデータに意味を付与するために、視覚情報の文脈を抽出する。さらに、システムによって状況に応じた情報を産み出すために、その文脈に基づいて蓄積された設計議事録などの知識のデータベースにアクセスし再構成を行う。

本稿では、提案するシステムに関する基礎設計について述べる。

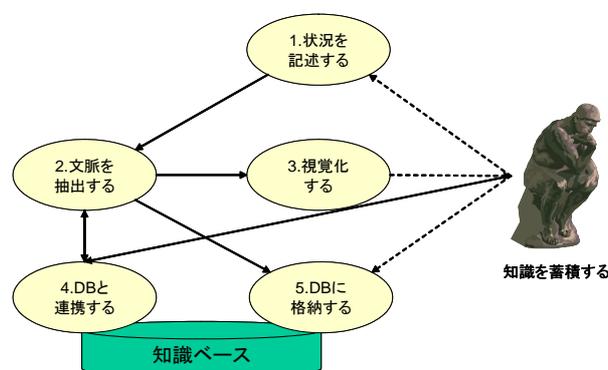


図1: システムの構成

2. 視覚情報から紡ぎ出す文脈

本研究では、「視覚情報の文脈」を「視覚情報の流れの中にある意味内容のつながり」と定義し、視覚情報の文脈を把握するために、視覚内の場面と場面の論理的関係、オブジェクト（人や物）とオブジェクトの意味関連を抽出することを試みる。

一般的に、機械的に論理的関係や意味関連を解釈することは困難である。そこで本研究では、映像内のオブジェクトとオブジェクトの意味関連をオブジェクトの出現依存関係に、場面と場面の論理的関係を場面ごとのオブジェクトの出現依存関係の相違にそれぞれ見出せることを期待する。

2.1 視覚データの文書化

本研究では記録すべき視覚情報として、設計者の視点からの映像、設計の場の風景の映像を検討している。

映像データの記述と解析のために物語構造モデル [赤石 06] を導入する。物語構造モデルはことばの意味や物語の内容を解釈せず、表層的な特徴量のみで機械的に文書の分解・再構成を行うためのモデルである。このモデルでは、線形に記述された文書集合を、語と語の連鎖関係を表すグラフ構造に変換し、グラフの構造に対する操作を通じて、文書群の分解・再構成を実現している。物語構造モデルにおける構成要素と文書の構成要素の対応を表1に示す。

カメラで撮った映像に対し、この物語構造モデルを適用する。映像内に出現するオブジェクトをキャラクター（単語）と

表 1: 物語構造モデルと構成要素

world model (世界構造)	set of stories
story (物語)	sequence of scenes (document)
scenes (場面)	chunk of event
event (出来事)	set of terms
character (登場人物)	term

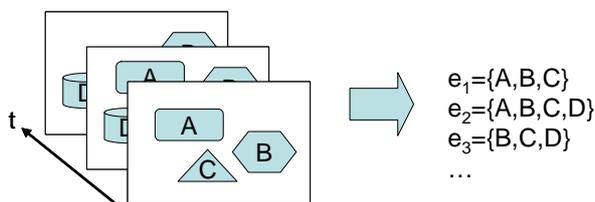


図 2: 映像データの文書化

して、ある時間内に出現したのオブジェクトの集合をイベント (文) として、さらにそのイベントの集合をシーン (章・節・段落) として扱う。視覚データを以下の手順で文書化する (図 2)。

1. 事前知識に基づきオブジェクトと名前の対応を定義する。
2. カメラを用いて映像を記録する。
3. 対応の定義を参照しながら、カメラの映像にあるオブジェクトを認識する。
4. 認識したオブジェクトの名前とタイムスタンプをテキストとして記述する。

なお、映像のオブジェクトの認識には既存の映像認識エンジンを用いる。

2.2 出現依存度と吸引力

文書化された視覚データのオブジェクトの連鎖関係を、「出現依存度」「吸引力」という概念を用いて考察する。

出現依存度 td を、語 t を含む文中に語 t' の出現する条件付確率で定義する。

$$td(t, t') = \frac{\text{sentences}(t \cap t')}{\text{sentences}(t)} \quad (1)$$

sentences は文書中に語 t を含む文の数を表す。たとえば、語 t から語 t' への出現依存度 $td(t, t')$ が大きいということは、語 t は語 t' に依存していることを示す。

吸引力 $attr$ を、ある語 t への出現依存度の総和で定義する。

$$attr(t) = \sum_{t \neq t''} td(t'', t) \quad (2)$$

吸引力が大きければ大きいほど他の語から依存されていることを示す。

オブジェクト間の関係を、出現依存度に基づきエッジの重み・向きを定義し、吸引力に基づきノードの大きさを定義した重み付き有向グラフ構造として表現できる。グラフ構造に変換されたオブジェクト間の関係は、視覚データから機械的に抽出された文脈として見做す事ができると考える (図 3)。

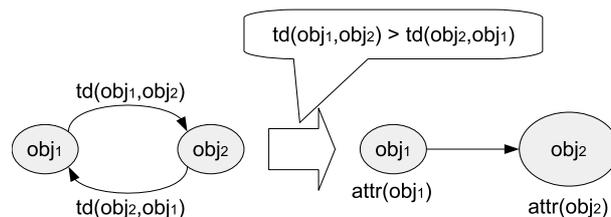


図 3: 有向グラフ構造表現

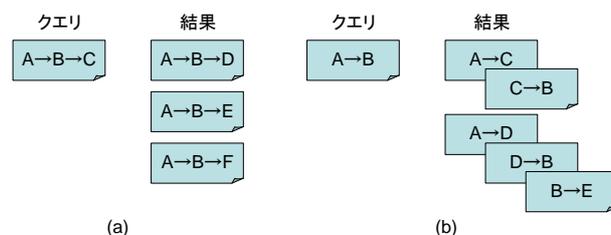


図 4: セグメントの選択 (a) 類似を利用した選択 (b) 連鎖を利用した選択

3. 文脈に基づく情報アクセス

前章の手順で抽出した視覚情報の文脈に基づき知識のデータベースにアクセスする。ここでは、視覚内のオブジェクトの出現依存関係 (および検索語) をクエリとして、ここでは、視覚内のオブジェクトの出現依存関係 (および検索語) をクエリとして、知識データベースの情報の再構成を行う。

3.1 対象とする知識のデータベース

知識の再構成を行うために知識のデータベースとして議事録文書などを主張に応じて適当な粒度でセグメント化する。そして各セグメントについて、出現依存度、吸引力を算出する。このような処理を施しておくことにより、視覚情報の文脈と既存の知識を同次元で扱うことができる。

3.2 セグメントの選択

クエリに対して知識のデータベースから、文脈に沿っているか、価値のある情報が含まれているか、という基準に従い、有益であると推定されるセグメントを選択する。現状では、類似した出現依存関係や吸引力分布を含むセグメント、出現依存の連鎖関係を含むセグメントを選択し、利用者に適当な形式で提示することを検討している (図 4)。

4. おわりに

視覚内のオブジェクトの出現依存関係に基づき、視覚情報の文脈把握するための手法を提案し、また、抽出した出現依存関係に基づく情報を既存の知識のデータベースから再構成するシステムを提案し、基礎設計を行った。今後、以上のシステムの実装と評価を行っていく予定である。

参考文献

- [赤石 06] 赤石 美奈: "文書群に対する物語構造の動的分解・再構成フレームワーク", 人工知能学会論文誌, Vol. 21, No. 5, pp.428-438 (2006)。