

センサデータとプライバシー

2013年2月28日

京都大学情報環境機構長
学術情報メディアセンター
美濃 導彦

本日の内容

- 内閣府連携施策群: 情報の巨大集積化と利活用基盤技術開発
 - 補完的課題: センサ情報の社会的利用のためのコンテンツ化(平成19~21年度)
- 科学技術戦略推進経費補助金: 安心・安全な社会のための犯罪・テロ対策技術等を実用化するプログラム
 - 環境適応型で実用的な人物照合システム(平成22~26年度(予定))

社会的背景

- WEBという膨大なデータベースにマルチメディアコンテンツが増加している
 - 画像、音声、音楽、映像などが増殖中
 - カメラによる実時間映像中継の普及
- 人間を観測するセンサネットワークが広がってきた
 - 現状把握に対する社会的要求
 - 防犯カメラの設置台数の増加
- プライバシ問題に対する関心の高まり
 - 監視社会への不安

コンテンツ流通の問題点

- 知的所有権
 - 人間が創造するコンテンツで発生
 - デジタルコンテンツのさまざまな権利問題
 - 権利を主張する人は明確
- プライバシ問題(人間を観測するセンサ)
 - センサ(カメラ)が勝手に取得する映像など
 - 意図せずに被撮影者にされる
 - プライバシの問題があるので勝手に公開できない
 - 相手は不特定多数

人間を観測したセンサ情報流通への対処

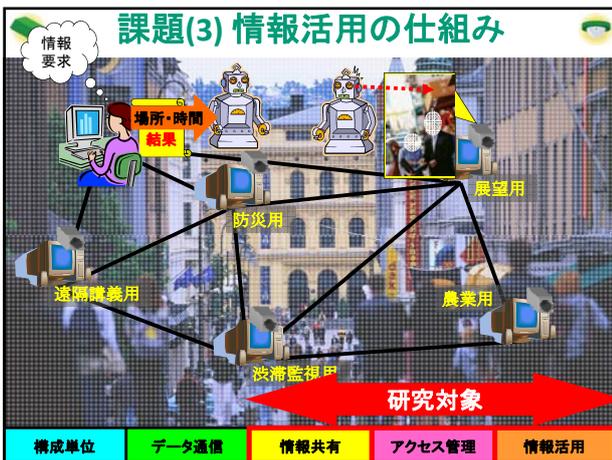
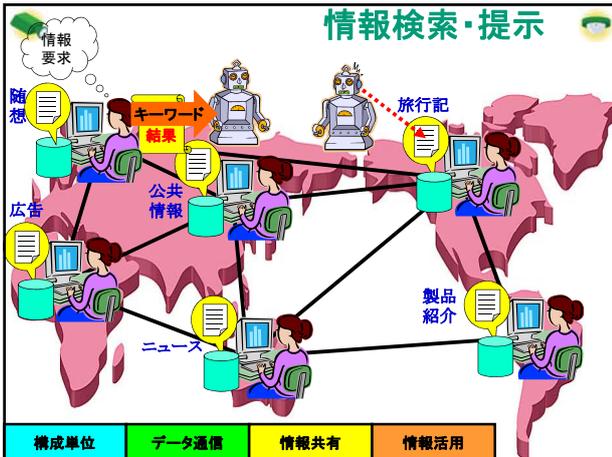
- センサネットワークは目的別に閉じており、センサデータの有効活用はなされていない
- 膨大な数の監視カメラが設置されているが、何が撮影されているかわからない不安



- センサ情報を社会利用できるコンテンツに発展させるセンシングWebの提案

Webの進展過程

第1回運営委員会 2007年10月19日



- ### 人間を観測するセンサの種類
- カメラ、マイクなど環境に設置され人間活動を記録するもの
 - 人間の外観、振舞、動き、話の内容、声など
 - 意図せずに取得される(プライバシー問題)
 - パターン情報であり時空間の膨大な情報を含む
 - スマホ、RFIDなど人間が使うためにもつもの
 - タグデータ、ログデータ、位置情報など
 - 装置を利用した/持っている時に取得される(個人情報)
 - シンボル情報であり、わかりやすい
 - バイオメトリックセンサなど身につけるもの
 - 人間の体からの信号を観測
 - 信号の時系列であるが、単純な処理が有効
- 第1回運営委員会 2007年10月19日
-

- ### 環境に設置されるセンサ情報の利活用
- センサ情報を共有するメカニズム(情報共有)
 - 時空間コンテキストを含むセンサ情報の記述法
 - 利用者からの要求の記述法
 - 必要なセンサの探索・発見方法
 - センサ情報特有の問題への対処(アクセス管理)
 - プライバシーの構造化とアクセス管理
 - ストリームデータからの情報抽出法
 - 分散したセンサ情報の検索・提示法(情報活用)
 - 大規模に分散しているセンサから情報を集める方法
 - 広大な領域に対する疎な観測環境での情報処理
-

環境に設置されたセンサ情報の特徴

- 実時間で定常的に取得される
 - ストリームデータ
- ほとんどが無駄な情報
 - 如何に捨てるかを考えることが重要
- 時空間的構造を持つ
 - 時間、場所、(3次元方向)などの属性
- パターンレベルの情報である
 - コンテキストの抽出が重要
 - プライバシの問題が存在

第1回運営委員会 2007年10月19日

視覚センサ情報取得法

① プライバシ情報を取得しない位置に設置する

設置法・制御法の工夫

人間がセンサを設置

センサ設置者

② 被写体との位置関係を利用してプライバシー情報を隠蔽する

第1回運営委員会 2007年10月19日

聴覚センサ情報取得法

人間がセンサを設置

センサ設置者

マイクはどこにおいても同じ音質変換、シンボル認識が必要

第1回運営委員会 2007年10月19日

Web情報とセンサ情報の違い

Web情報

手入力

自動獲得

プライバシー問題

センサ情報

プライバシー問題を避ける風景WEBカメラが中心

直接観測

実世界

実体験

第1回運営委員会 2007年10月19日

	文字情報 編集型、蓄積型	センサ情報 観測型、ストリーム型
閉じた系	情報処理システム	制御システム
開いた系	Web	センシングWeb

第1回運営委員会 2007年10月19日

研究内容

センサ情報コンテンツ化技術
・時空間コンテキストを含む
・センサ情報の記述法
・必要なセンサの探索・発見法

センサ情報コンテンツ化技術

コンテンツバス

センサ検索
場所

センサ応答

アクセス制御

データ取得

センサノード

プライバシー情報管理技術
・プライバシー保護処理
・プライバシーの構造化とアクセス管理

センサ情報の利用

クリックابلマップ

コンテンツ化システム

提示システム

観測型実世界コンテンツ提示技術
・分散センサからの情報収集法
・ユーザにわかりやすい提示法

センシングWeb

第1回運営委員会 2007年10月19日

プライバシー情報の扱い

第1回運営委員会 2007年10月19日

センシングWebにおけるプライバシー

- 被観測者のプライバシー
- センサ設置者のプライバシー
- センサ情報利用者のプライバシー



第1回運営委員会 2007年10月19日

被観測者のプライバシー

- 本人自身の情報
 - バイOMETリック情報
 - (映像)人相、顔色、表情、髪型、服装、体型
 - (音声)声
 - (その他)体温、体重、指紋、心拍数、発汗量
 - 行動
 - 行動内容、会話内容
 - 所有物
 - 所持金額、カード番号
 - 家の外観
 - 車
- etc.
- 位置・時刻に関する情報
 - 本人(または所有物)がその日その時刻にその場所に居たこと
 - そのとき同じ場所に誰が居たか

第1回運営委員会 2007年10月19日

センサ設置者のプライバシー

- 固定型 (あまり問題なし)
- 移動型
 - そのときセンサがどこにいてどういう状態にあったかが、センサ情報利用者にも伝わる
 - ⇒センサとセンサ設置者との位置関係が固定的なときに、設置者自身が観測の対象となる
- 保護すべきでない(または注意して保護すべき)プライバシー
 - センサを設置したのは誰か?
 - どういったポリシーで運用されているか?
 - センサの存在そのもの
 - 盗聴器、隠しカメラ
- センサ情報の信頼性
 - センサの故障への対応
 - 悪意のある偽データの流布による混乱の防止

第1回運営委員会 2007年10月19日

センサ情報利用者のプライバシー

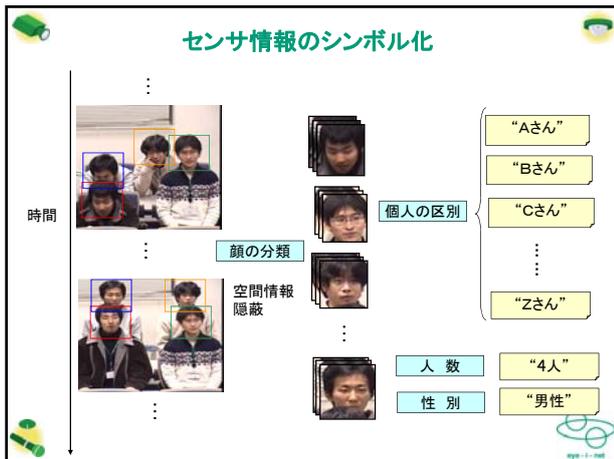
- 【保護すべきプライバシー】
 - 誰が、いつ、どこから
 - どこのセンサのどのようなデータにアクセスしたか
- 【保護に注意を要するプライバシー】
 - データをどこでどう二次利用したか
 - 電子透かし?

第1回運営委員会 2007年10月19日

カメラ情報のプライバシー構造化戦略

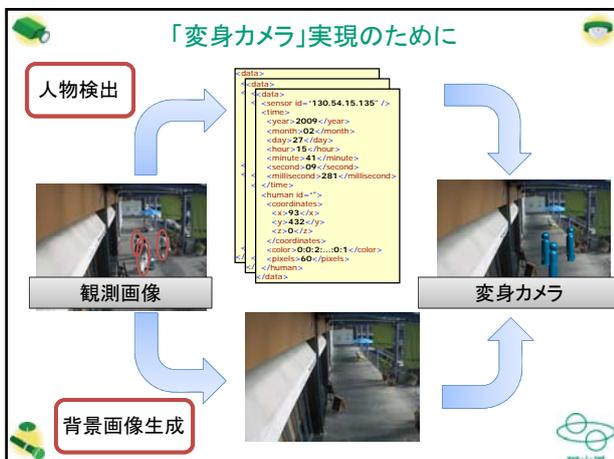
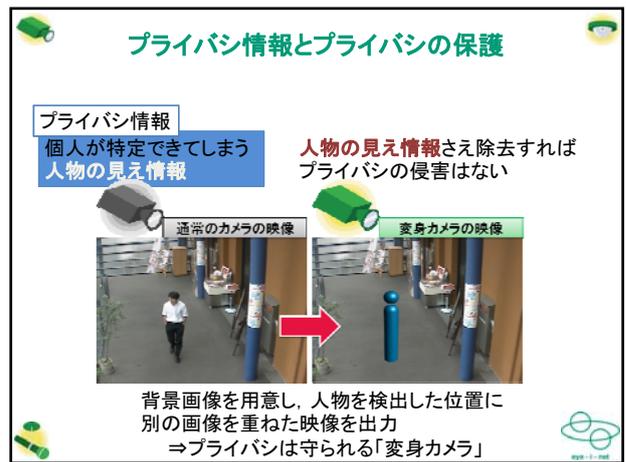
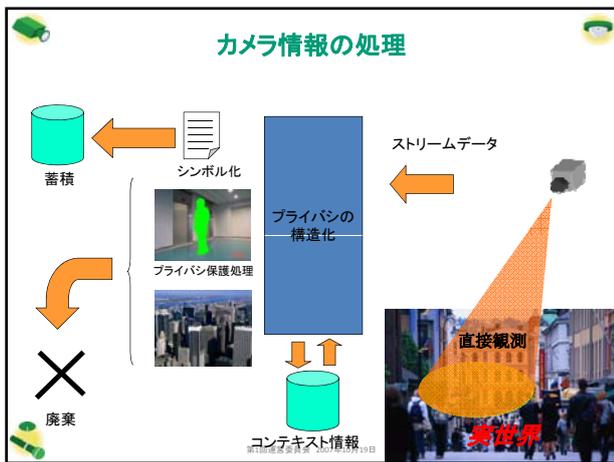
- 大前提
 - 人間が見るとプライバシーの問題が発生
 - 機械が処理し加工すると問題はない
- 人間に見せる場合
 - 認識処理によりセンサ情報をシンボル化
 - 生データの場合はプライバシー保護処理が必須
- 生データは蓄積しない
 - 蓄積するとプライバシー問題が発生
 - 蓄積していない、蓄積できない、という証明が必要

第1回運営委員会 2007年10月19日



カメラの“実世界センサ”化

	カメラ	実世界センサ	人感センサ
獲得データのタイプ	パターン	パターン	パターン
含まれる情報	rich	medium	poor
センサの出力情報	生データ (映像)	シンボル (実世界の状況)	シンボル (人の有無)



人物に関する情報以外を出来るだけ保存した背景画像生成

プライバシーは確実に保護したい

- 人は絶対に画像中に現れてはいけない

出来るだけ多くの情報を伝えたい

- 観測画像に出来るだけ近づける

観測画像をそのまま使うことはできない

観測画像の情報を出来るだけ生かしたい

背景画像生成手法

- 固有空間を使う手法
- 観測確率に基づく手法
- 背景変動の明示的なモデルに基づく手法
- 画像補完

固有空間法による背景画像生成

- 長時間観測した画像群から背景部分空間を学習

そのシーンで頻繁に現れるものが背景

↓ 1枚の画像を1つのベクトルとして主成分分析

固有空間

全体的な明るさが変動しても、それを反映した画像が生成できる

カメラの観測画像

生成した背景画像

固有空間法を用いた背景生成例と問題点

カメラの観測画像

生成した背景画像

背景生成失敗の種類とその要因

- ① 日照のエッジ部分がボケる
 - 頻繁・周期的に現れるが、連続的にゆっくり移動するようなもの
- ② 看板のゴーストが出る
 - 駐車場の車などもしばらくの間静止してこのケースに移動することがあるようなもの
- ③ 旗がボケる
 - 常にランダムに変動している木の揺れなどもこのケースにずっと存在しているもの

背景生成失敗の種類とその対処法

- ① 日照のエッジ部分がボケる
 - 超長時間(長期間)の観測結果に着目
- ② 看板のゴーストが出る
 - その物体が静止している程度の間(中期間)の観測結果に着目
- ③ 旗がボケる
 - 現在(短期間)の観測結果に着目

このように、異なる時間間隔のデータに着目し、それを解析することで背景のより高精度な生成を目指す

長期間にわたる超大量の観測画像

時間

◎構造 ×照明

×構造 ◎照明

照明の変化
時刻によって位置が変化
強さも変化

構造の変化
日によって変化
1日の間是不変とする

日

これらの成分を
独立に記述できるような画像モデルを提案
そのモデルを用いて独立に推定できるような処理を提案

実験(画像生成結果)

観測画像

長期間の平均化を用いる手法

時刻ごとの固有空間を用いる手法

今回の提案手法

同一時刻の画像群を用いた固有空間法

10:00 12:00 14:00 16:00

生成した背景画像

処理対象の画像と
同一時刻の画像を
グループ化して処理

日照の変化が再現できている
ことが確認できた

人物検出

第1回運営委員会 2007年10月19日

HOG特徴を用いた人物検出(1)

- HOG (Histograms of Oriented Gradients) 特徴

局所領域毎に線表勾配を算出し
線表勾配ヒストグラム H を獲得

領域毎の H を統合して
HOG特徴量 $V = \{v_1, v_2, \dots, v_n\}$ を獲得

HOG特徴は
対象の局所形状を反映

検出ウィンドウ

HOG特徴を用いた人物検出(2)

- HOG特徴を用いた人物検出の過程

探索範囲の設定
様々な大きさ・向きで全探索

検出ウィンドウ

HOG特徴量による検出
学習済み特徴との比較で判定

正事例 負事例

特徴空間

時系列方向の位置変化に注目した識別

• 検出結果の正誤を時系列情報から識別

- ① 瞬間的にのみ存在する検出結果は人物に対する誤検出
- ② 滑らかに位置変化している検出結果は正検出
- ③ 長時間位置変化がない検出結果は固定背景に対する誤検出
- ④ 前後のフレームから注目フレームの検出結果を予測し検出漏れを推定

43

正誤識別結果の各事例を用いたHOG特徴の追加学習

追加学習による新規検出器の生成

44

異なるカメラ間で同一人物の検出

異なるカメラ間での人物同一性判定

• 画像から得られる見え

- 色が同じ
- 顔検出により推定
 - 年齢
 - 性別
 - 髪型
 - ...
- 胴体
 - 身長
 - 色特徴 ⇒ 人物領域から抽出
 - ...

複数の特徴量の利用方法

信頼性の低い特徴量はいわずに適切な特徴量により判定

特徴量の選択利用

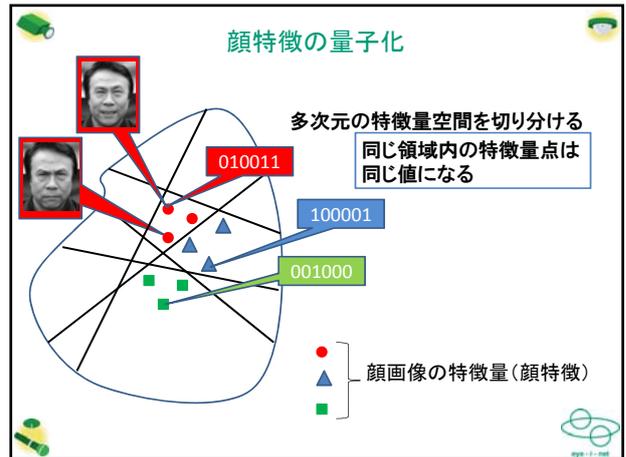
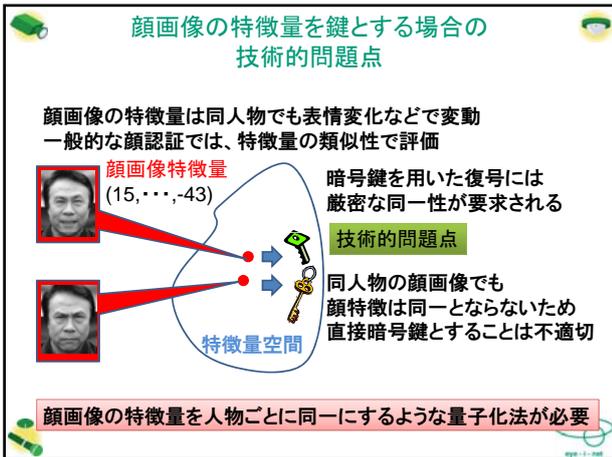
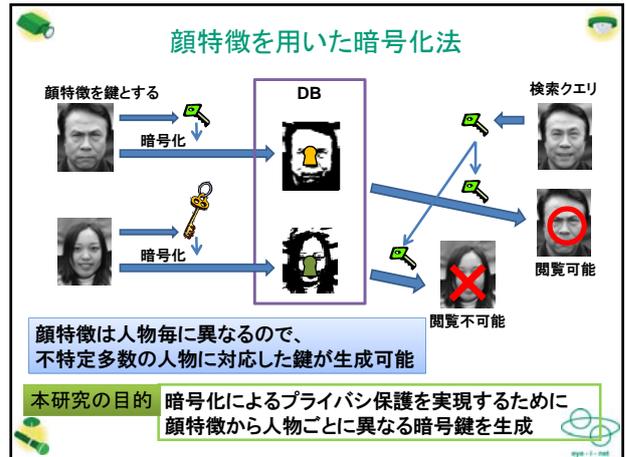
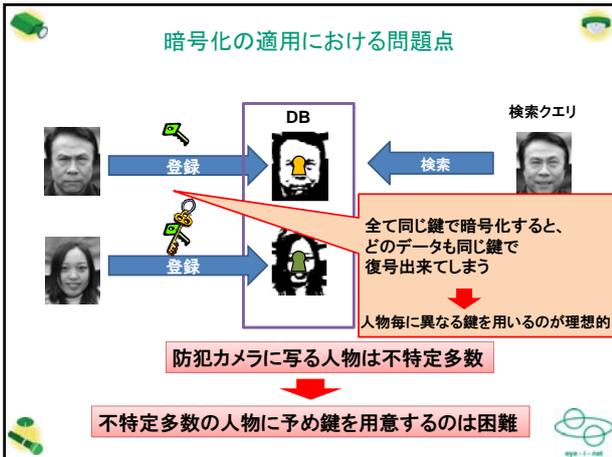
カメラ間の人物照合

カメラA

カメラB

対応付け

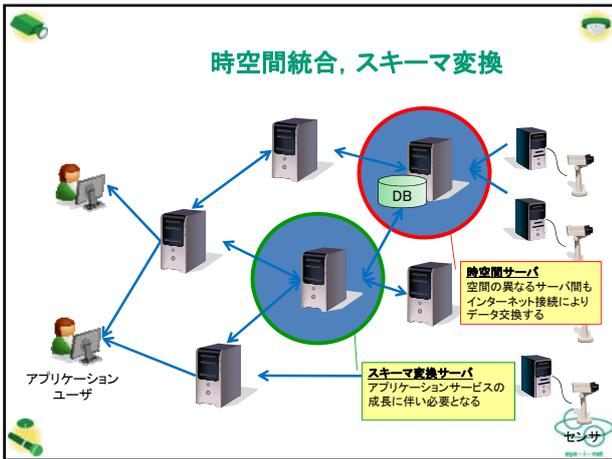
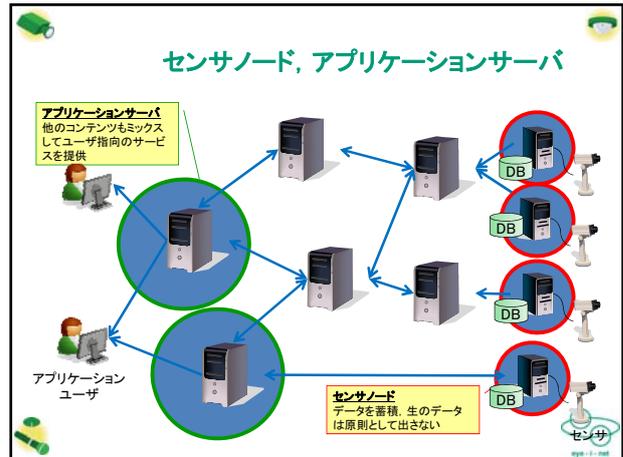
画像はVIPeRより転載(Grey, 2007)



- ### 情報共有のために必要なこと
- アーキテクチャ設計
 - 情報の取得から表示までの構成(器)
 - 入側・出側それぞれの独立性を保つこと
 - 入力がストリームデータ
 - データフォーマット(形式標準)
 - 情報の取得から表示までの流れ(中身)
 - 記述の自由度の加減がポイント
 - 情報要求(クエリ)の記述規定・記述支援
 - 中身の選り分け・切り出し

ザ・センシングウェブ・アーキテクチャ

- センサノード
 - プライバシー保護処理、ストリーム処理、シンボル化
 - 取得したデータを蓄積
- 時空間サーバ
 - オブジェクトとフレームの合成
- スキーマ変換サーバ
 - 需給者間のビュー変換
- アプリケーションサーバ
 - コンテンツの最終的な構成



データフォーマット

- コンテナ: アプリケーションに共通
 - 時空間情報 + メタデータ
 - オブジェクト: 空間情報を伴う
 - フレーム: 時間的スナップショットの単位
 - 空間優先、時間優先ともに記述可能
- コンテンツ: アプリケーションに依存
 - データの中身の構造化
 - 標準アプリケーション(トラッキングなど)を提供

オブジェクト優先とフレーム優先

- 1フレームに複数のオブジェクト
 - センシング空間のスナップショット

```

<frame timestamp="t0">
  <object>
    <coordinates>x1y1z1</coordinates>
    ...
  </object>
  <object>
    <coordinates>x2y2z2</coordinates>
    ...
  </object>
</frame>
    
```

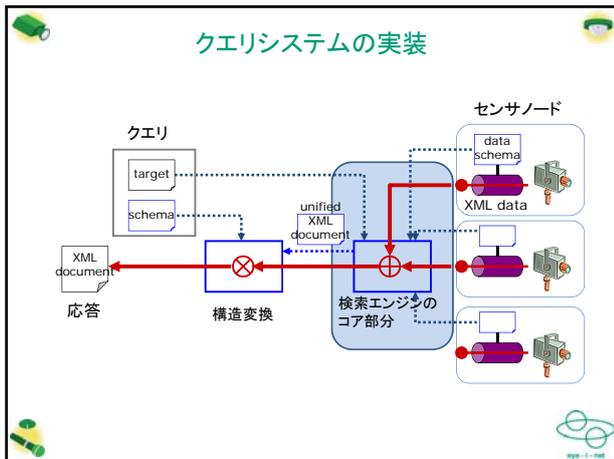
- 1オブジェクトに複数のフレーム
 - センシング対象の時間的な変遷

```

<object>
  <frame timestamp="t1">
    <coordinates>x1y1z1</coordinates>
    ...
  </frame>
  <frame timestamp="t2">
    <coordinates>x2y2z2</coordinates>
    ...
  </frame>
</object>
    
```

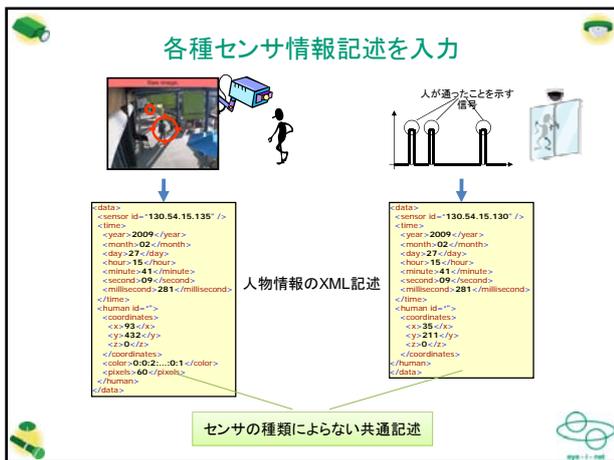
クエリ設計と記述支援

- 時空間指定 + 変換処理指定
 - 変換ルールセット(XSLTプログラム)
 - 記述順序の依存解析
- クエリ記述は抽象度低い
 - 要求仕様に対するプログラムのようにデバッグ支援にあたる記述支援が必要
 - 不適切な記述の検出
 - 要素選別の矛盾、データ取得と変換の不適用、構造変化の不一致など
 - 頻出タグ組合せの自動記述(ライブラリ)
 - 計数処理、カテゴリ選択、類似要素選択など



センサ情報の利活用

第1回運営委員会 2007年10月19日



実証実験

- ### 実証実験の目的
- 動くシステムを作って実証実験の実施
 - 公共の場での実験に関する問題点の解明
 - システム、手法の有効性の実証評価
 - プライバシーの保護の評価
 - 実証実験の場所: 新風館(京都、烏丸御池)
 - 実社会での一般の人々を対象とした実験
 - 屋内と屋外が混在(セミオープンスペース)
 - 中央にイベントスペースがあるので、イベント時の人の流れが変化
 - 期間: 2009年7月~12月



実証実験で使用する多種のセンサ



カメラ

RFIDタグリーダー

赤外線センサ

無線LANアクセスポイント

気象センサ

実証実験でのデモ内容

- 変身カメラ
- アプリケーション
 - にぎわいマップ
 - デジタルジオラマ
 - シースルービジョン

実験評価のアンケート調査

アンケートの目的

- プライバシー意識の調査
センサネットワークの社会的な配備について、利用者がどのようなプライバシー意識を持っているか調査する。
- プライバシー対処技術の有効性評価
SWebで行っているプライバシー対処のための処理が、プライバシー保護に有効に働いているか、利用者側からの評価を行う。
- SWebサービスの受容度調査
個人情報の開示と得られるサービスとの妥結点を調査する。また、音声情報に対するプライバシー意識を調査する。

実証実験のデモを通じて、プライバシーとサービスとの関連を体験した上で、アンケート調査を行う。

プライバシー意識の調査

プライバシー意識は多様

監視されることへの抵抗感	「分からない」ことへの不安感
<ul style="list-style-type: none"> 個人が同定される 行動を継続的に観察される カメラやマイク等で情報取得される 	<ul style="list-style-type: none"> センサの設置場所が明示されない 収集データの使用目的が明示されない 情報の利用者が限定されない

どのような点が重視されるかを調査する

プライバシー対処技術の有効性調査

SWebでは...

監視されることへの抵抗感	「分からない」ことへの不安感
個人情報の除去、統計情報化により対処	センサデータの公開、オープン化により対処

このような対処法を有効と感じるか調査する

本調査のねらい

- 防犯目的では、90%以上の方がプライバシー保護より防犯カメラの設置を支持。
- 防犯カメラでは、映像のオープン化は想定していない。

得られる情報をオープンにすることを前提に、開発した技術において、サービスの有用性とプライバシー開示とのトレードオフが成立しているか？

アンケート調査の概要

- 3回のイベント開催時(7/31, 9/19, 11/15)に、来場者にアンケートを実施。
- トークイベントまたはデモを見てもらった上で、アンケートに回答する形式。
- アンケート回収件数: **216件**

アンケートの結果(変身カメラ)

1. プライバシが保護されていると感じるか

あまり感じない	全く感じない	わからない	とても感じる	感じる
9 (4.2%)	1 (0.5%)	0 (0.0%)	82 (38.1%)	89 (41.1%)

2. この技術が社会的に広まることに心配か

とても心配	心配	わからない	全く心配ない	とても心配ない
4 (1.9%)	37 (17.2%)	39 (18.1%)	82 (38.1%)	89 (41.1%)

アンケートの結果(変身カメラ)

3. 防犯カメラと比べてプライバシー保護を重視しているか

あまり思わない	全く思わない	わからない	やや思う	思う
6 (3.0%)	0 (0.0%)	4 (2.0%)	82 (38.1%)	89 (41.1%)

新たな概念の構築

- 監視カメラ
 - 見ている人(公権力者、管理者)
 - 見られている人(人民)
 - 目的: 監視
- 実世界カメラ(ソーシャルセンシング)
 - 人民の人民による人民のためのカメラ
 - 見ている人(人民)
 - 見られている人(人民)
 - 目的: 社会統計の取得や現状把握

プライバシー管理方法

- 新しいプライバシーコントロール
 - 「環境オプトイン方式」
 - 個人が情報を収集される環境にあることを知りながら、あえてその環境下に自主的に入るのであれば、その情報補足に同意したと理解できる
 - その環境の情報収集レベルが一目で分かる工夫
 - 象徴的なマークの掲示、センサの色・形で識別
- センサ情報の蓄積とその利活用
 - 自己コントロールは困難
 - 個人情報と紐付けできない仕組みの構築

おわりに:ビッグデータの時代

- ビッグデータ時代の到来
 - 監視カメラの映像を共有化すればビッグデータとなる
 - さまざまな活動情報が計算機を利用するたびに取得され蓄積されビッグデータとなる
- ビッグデータでのプライバシー情報
 - 統計処理だけでは不十分
 - 同じ人間に紐づけられるから利用価値
 - 個人情報は不要(パーソナル情報)
 - どのようにプライバシーを確保するか?

個人健康記録

- 個人健康記録(Personal Health Record)
 - 個人の健康情報、病歴情報を生涯にわたって集め医療の質、効率を向上させる
 - 遺伝子情報との紐づけによる治療の確立
- 健康時のバイオメトリック情報の集積
 - 生活習慣病の早期発見
 - 心理状態の推定
- 生体情報による個人認証

個人学習記録

- 個人学習記録(Personal Learning Record)
 - 個人の学習記録を生涯にわたって集める
 - 個人の進路指導、落ちこぼれ等の異常の早期発見
- 教育の現状に基づくデータの収集は重要
 - 教育の状況の現状把握ができる
 - 政策に反映するための基礎データとなる
 - 高等教育については海外では集める準備はできている