Kinect for Windows による手話認識

みずほ情報総研会社
サイエンスソリューション部
下元　正義

Kinect for Windowsを用いた日本手話の認識研究は千葉大学とみずほ情報総研が共同で実施しています。

Kinect®は米国マイクロソフトの商品です。

本日の内容

- Kinect for Windowsの紹介
- 骨格追跡
- 手話と手話認識
- Kinect for Windows ver.2 preview

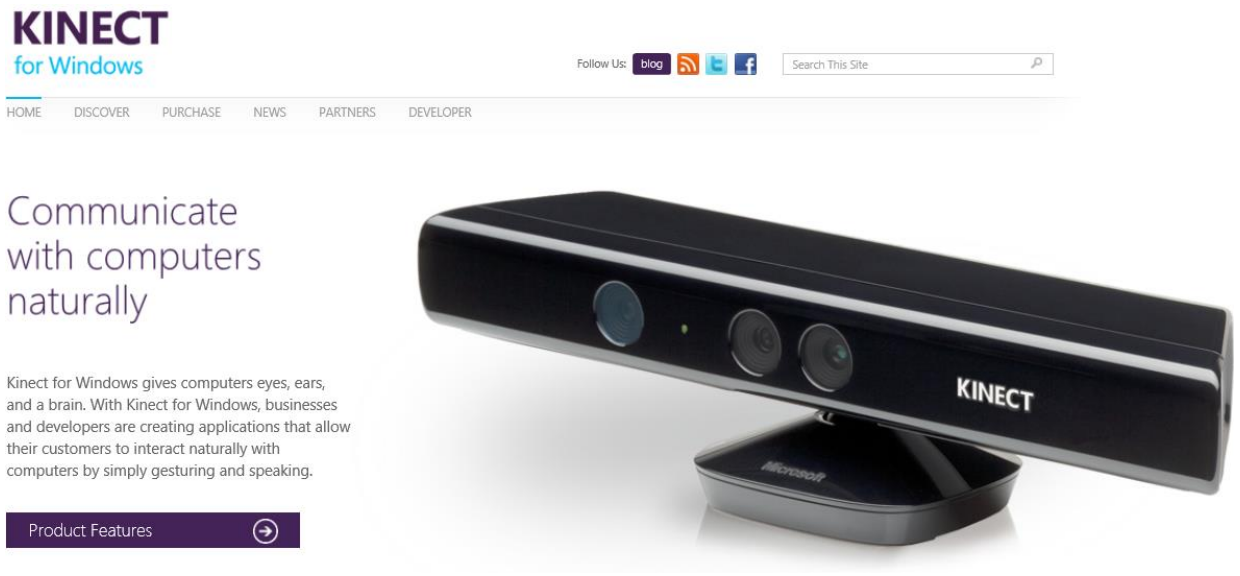
Kinect

- マイクロソフトが開発
- 2010年Xboxのコントローラとして発売
- 2012年商用利用可能なKINECT for Windowsが販売開始
- 世界累計販売台数1000万台突破（2012年4月時点）
- 2014年Next Generation Kinect（ver.2）がリリース

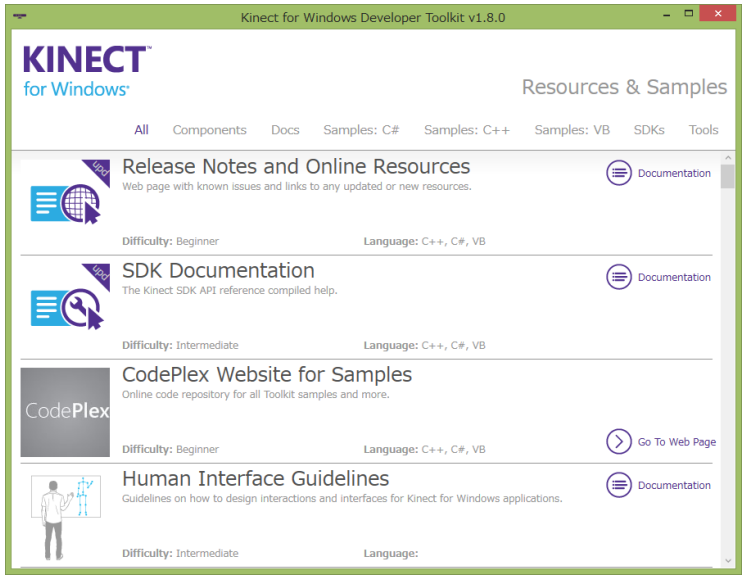
http://www.microsoft.com/en-us/kinectforwindows/

Kinect for Windows

- 商用利用可能なKinect
- SDKをウェブサイトから入手
- サンプルプログラム・サンプルコードが利用可能



http://www.microsoft.com/en-us/kinectforwindows/



Developer Toolkit

Kinect for XBOX

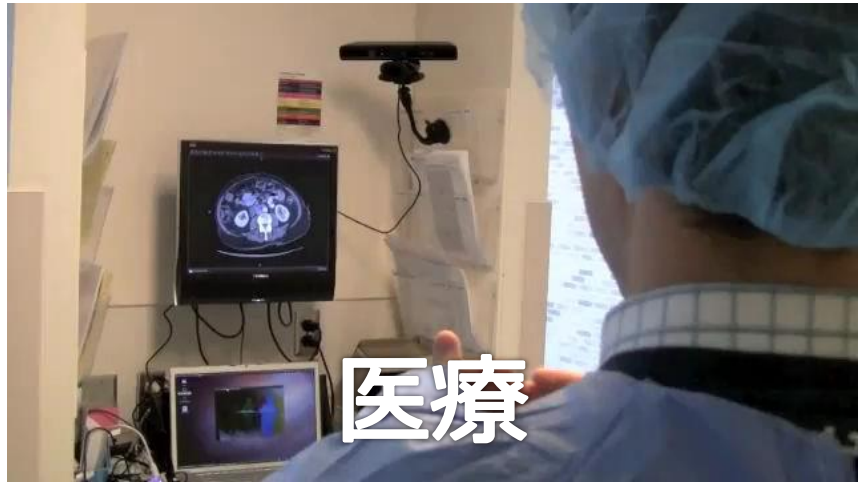
- 商用利用は不可

Kinect for Windowsはどこで買えるの？

- 家電量販店、ネット通販などで購入
  →Windows Developer Days(2012年4月)の帰りに購入しました。
  →2万円ちょっとくらいです。

NUIを利用したアプリケーション

- KINECT SDK, Open CV
- C#,C++,Java Script
- 骨格追跡を利用したプログラム
- 音声認識を利用したプログラム



外科手術現場（カナダの事例）

恋するマリオネット（ユナイテッドアローズ）

Kinectの魅力

- コンピュータが人の動きを理解することが可能
- 骨格追跡により人の動きを利用したアプリ開発が容易

Kinectはどうやって骨格追跡をしているのか？

- 骨格を簡単にとれることで面白いアプリが開発できる

一方で、

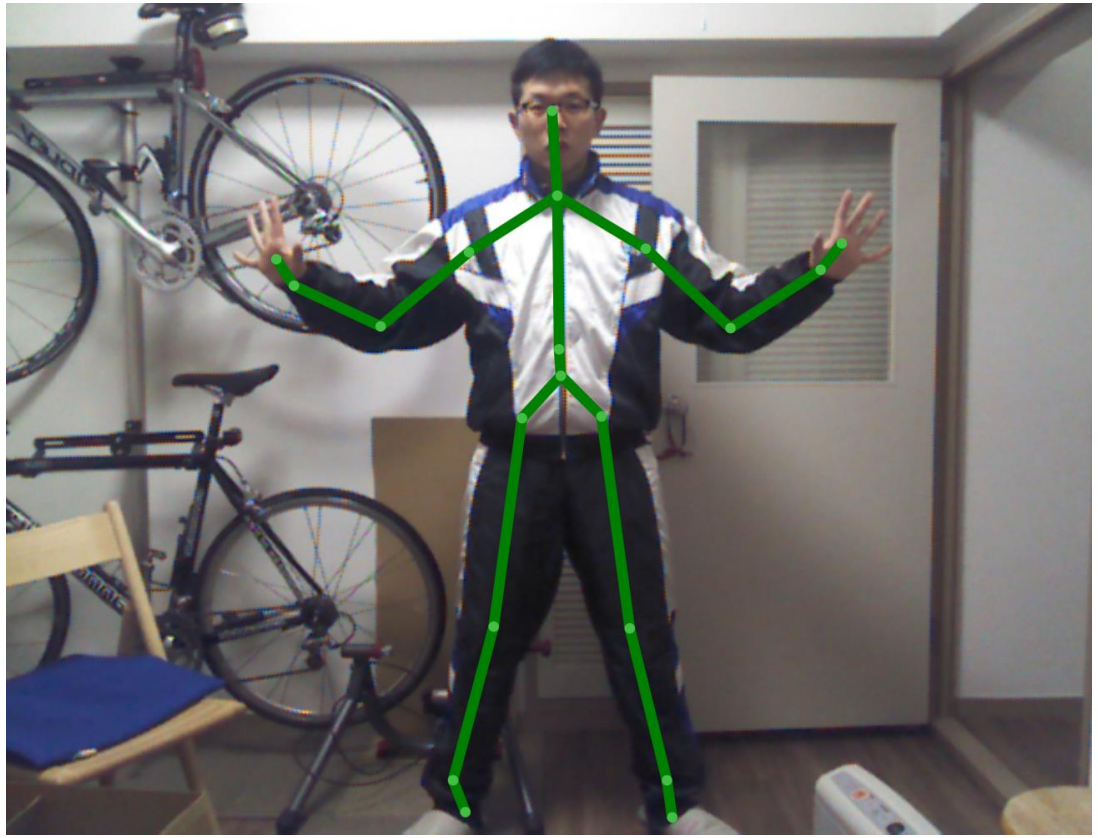
- 追跡に失敗することもある
- 何をやっているか分からないものに頼りたくない

骨格追跡機能にフィーチャー

- 関節位置を推定
- 手話認識で利用

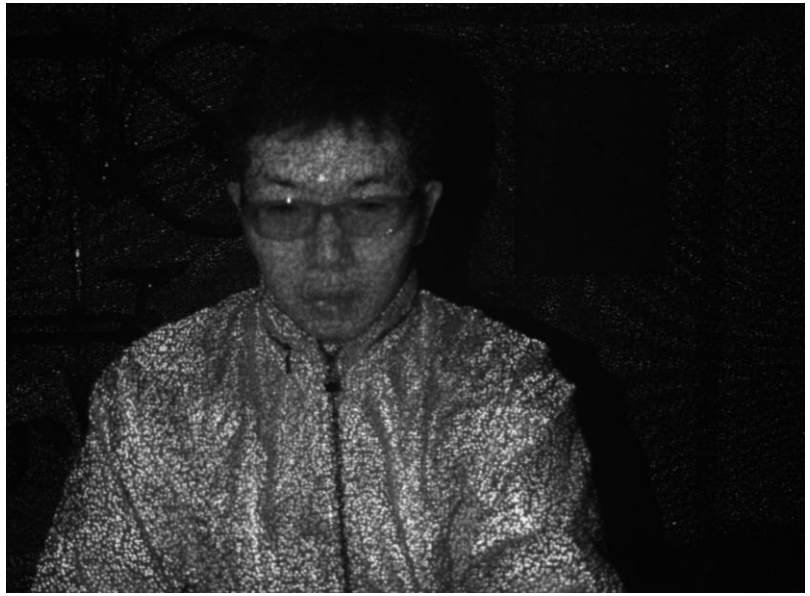
骨格追跡デモ

- スケルトン取得＋RGB



距離の取得　Light Coding

- Prime Senseの特許技術
- 光の粒を放射
- 反射パターンから距離を推定

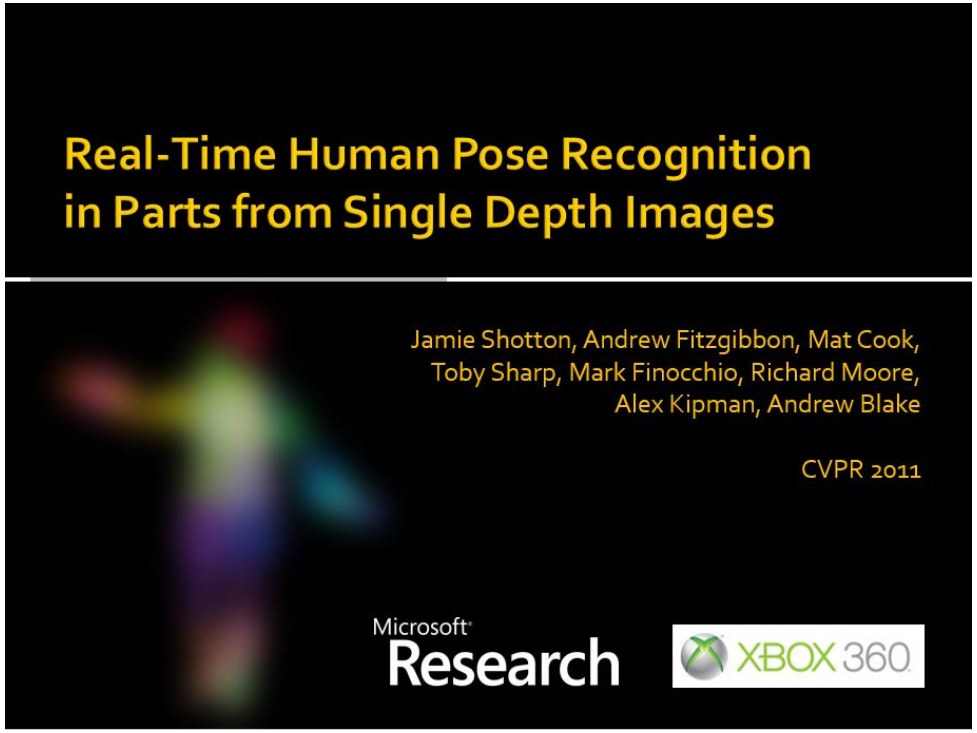


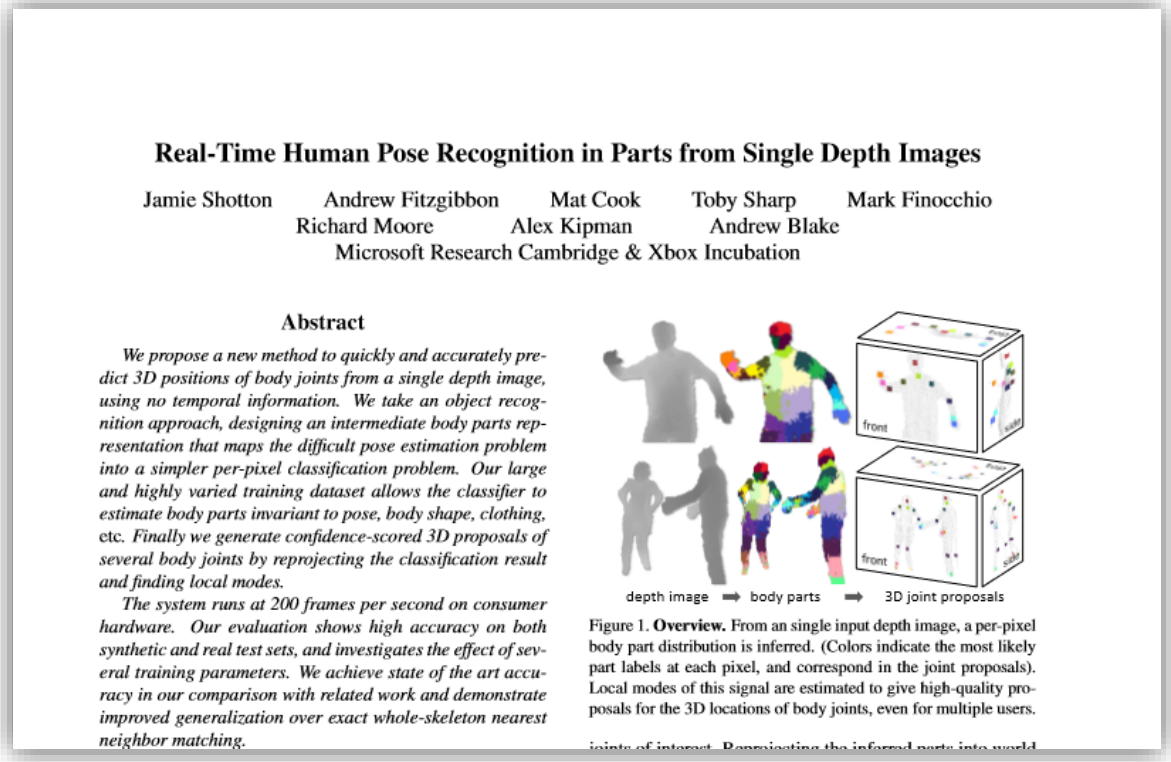
赤外線カメラで撮影した赤外線のパターン画像の例

http://www.youtube.com/watch?v=dTKlNGSH9Po

骨格追跡技術

- Microsoft Researchが研究・開発





Real-Time Human Pose Recognition in Parts from Single Depth Images

Jamie Shotton Andrew Fitzgibbon Mat Cook Toby Sharp Mark Finocchio
Richard Moore Alex Kipman Andrew Blake
Microsoft Research Cambridge & Xbox Incubation

Abstract

We propose a new method to quickly and accurately predict 3D positions of body joints from a single depth image, using no temporal information. We take an object recognition approach, designing an intermediate body parts representation that maps the difficult pose estimation problem into a simpler per-pixel classification problem. Our large and highly varied training dataset allows the classifier to estimate body parts invariant to pose, body shape, clothing, etc. *Finally we generate confidence-scored 3D proposals of several body joints by reprojecting the classification result and finding local modes.*

The system runs at 200 frames per second on consumer hardware. Our evaluation shows high accuracy on both synthetic and real test sets, and investigates the effect of several training parameters. We achieve state of the art accuracy in our comparison with related work and demonstrate improved generalization over exact whole-skeleton nearest neighbor matching.

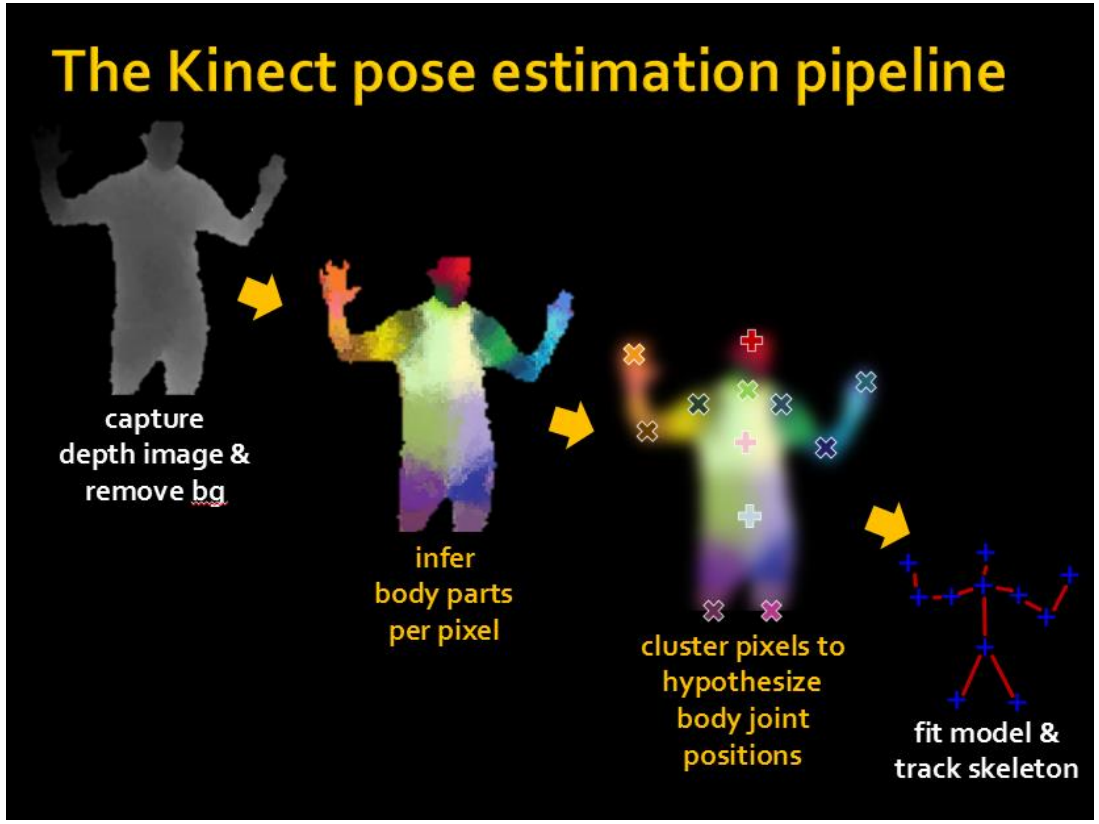
depth image ➡ body parts ➡ 3D joint proposals

Figure 1. **Overview.** From an single input depth image, a per-pixel body part distribution is inferred. (Colors indicate the most likely part labels at each pixel, and correspond in the joint proposals). Local modes of this signal are estimated to give high-quality proposals for the 3D locations of body joints, even for multiple users.

http://research.microsoft.com/apps/pubs/?id=145347

骨格位置推定の概要

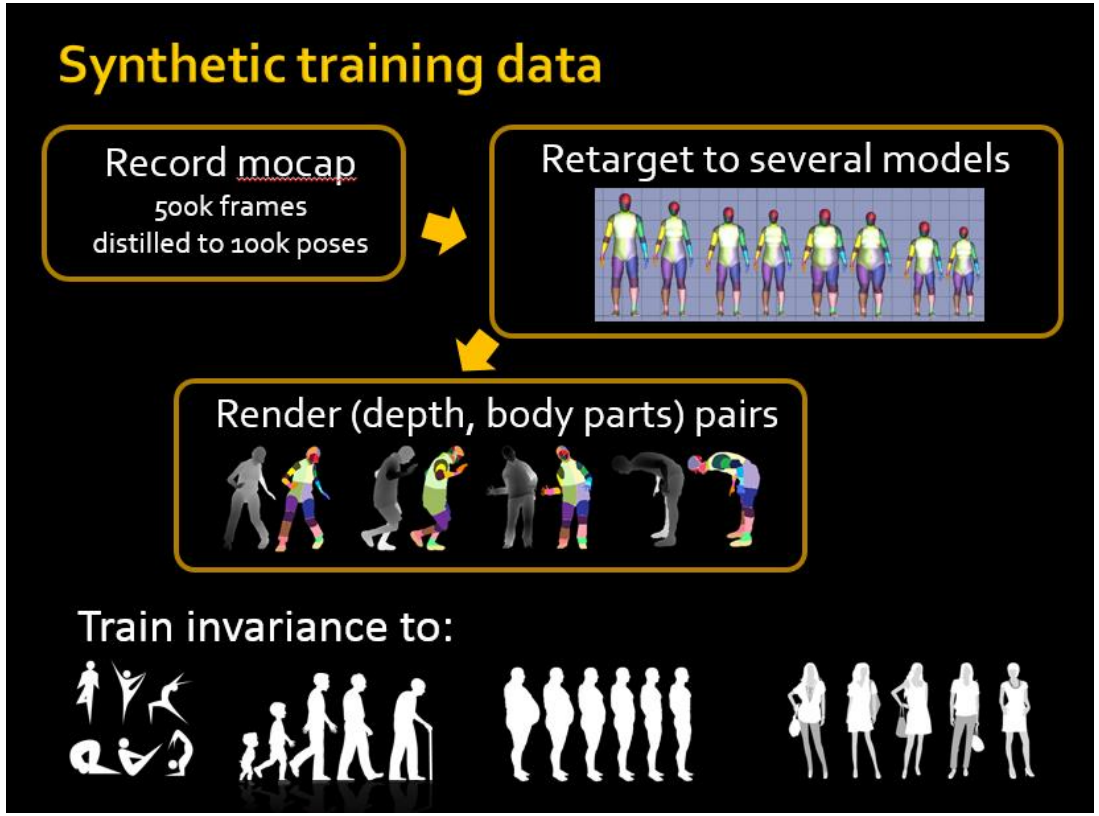
- 1フレームの深度イメージを使用
  →時系列データを使用していない
- 人型を抽出
- ピクセル単位でパーツに分類
- 関節位置を推定



http://research.microsoft.com/apps/pubs/?id=145347

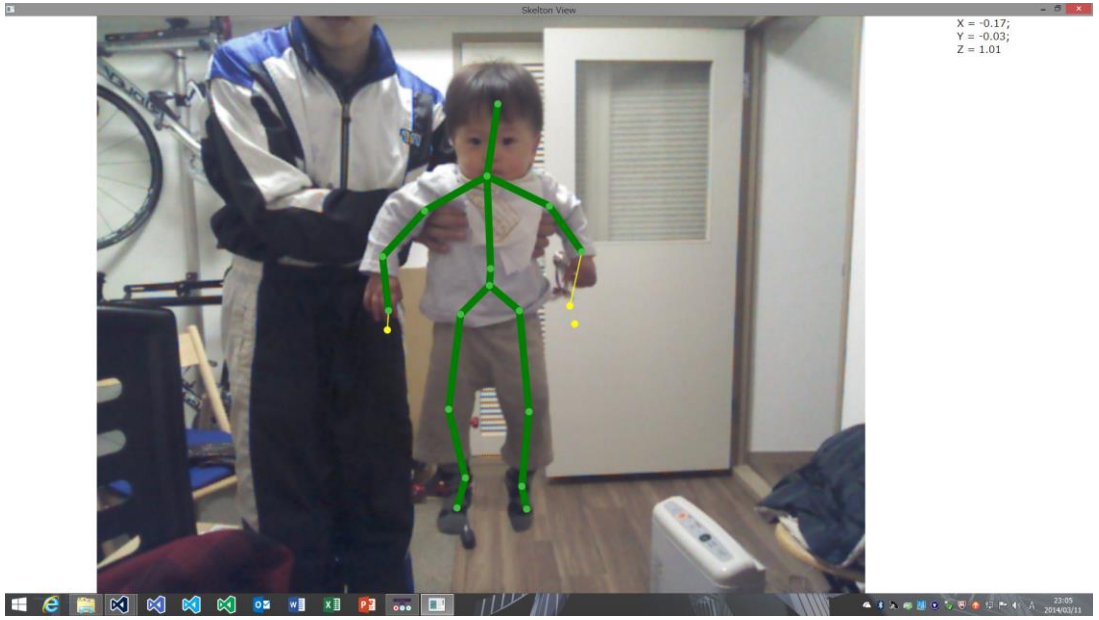
機械学習データの作成

- モーションキャプチャ50万→10万を使用
- 3次元モデル化
- 深度とパーツの対応付け
- モデルを使用した機械学習
- 身長、ポーズ、服装、体型を学習



http://research.microsoft.com/apps/pubs/?id=145347

Kinectの骨格追跡



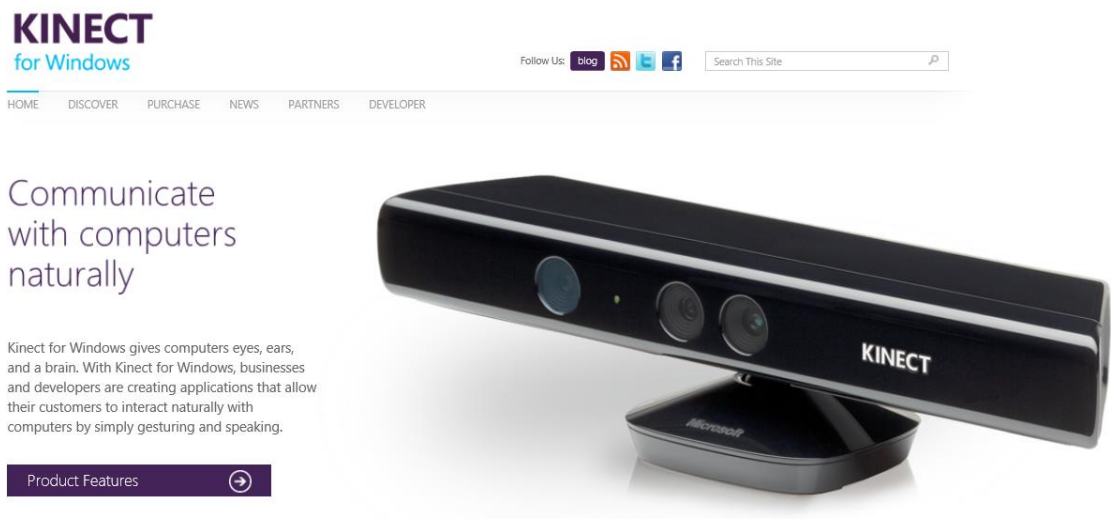
- １フレームのみを使用
- ピクセルごとにパーツに分類
- モデルから骨格位置を推定

１フレーム→ロバスト
１ピクセル単位→高速処理

インタラクティブなアプリケーションとしてNUIが利用可能

Kinectを用いた開発

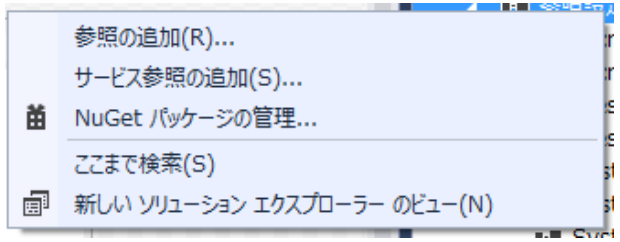
- Kinect for Windowsを購入
- SDK(1.8)をダウンロード
- Visual Studio（ExpressもOK）
  - ✓For Windows Desktop

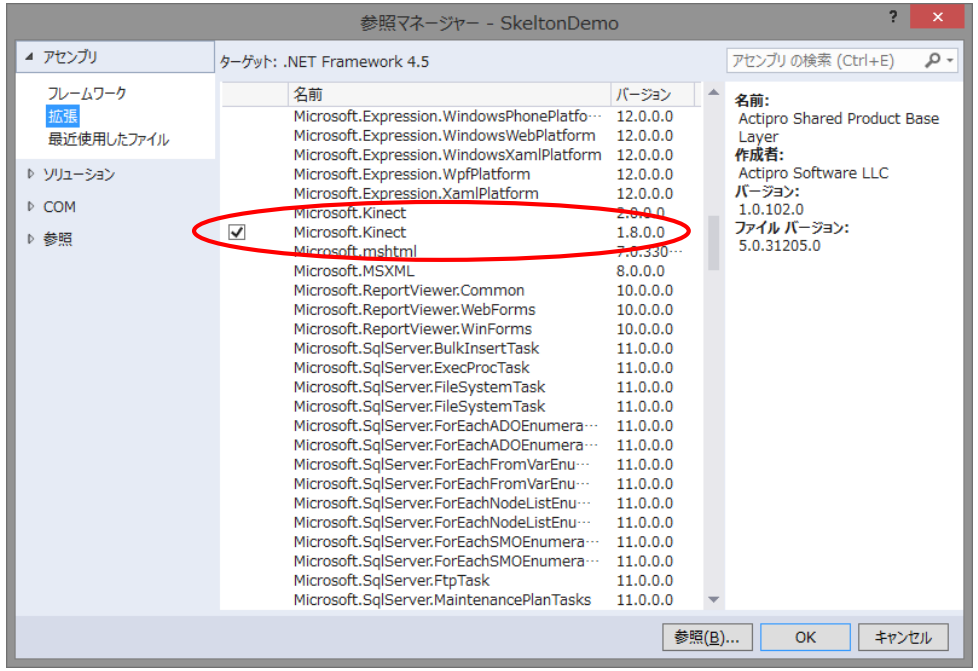


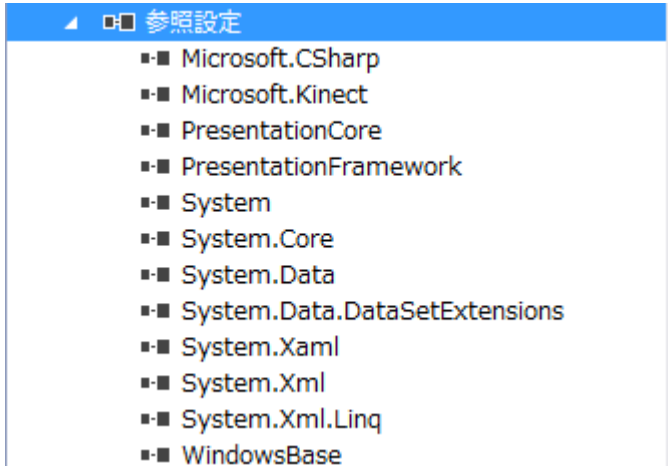
Visual Studio

ストアアプリもWeb Socketを使うと利用できるようです。

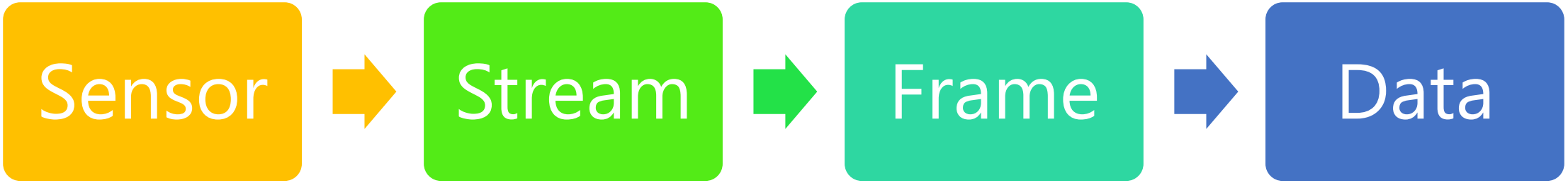
Kinectを参照に追加する





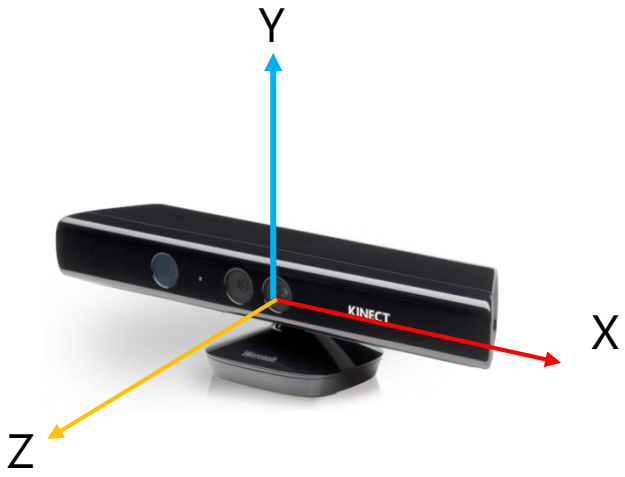


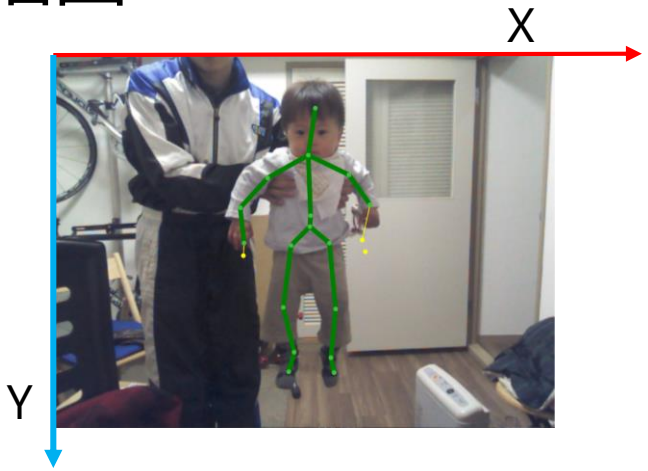
Kinectからのデータ取得



デモコードの概要

- C#を使用
- Kinectを１台取得
- RGBストリーム、Skeltonストリームをイベントハンドラで取得
- RGBを描画
- SkeltonをRGBの座標に変換して描画





Kinectによる手話認識

- 人とコンピュータの未来
- 世の中の先進的な開発者はすでに様々なアプリを公開
- 今からできることは？→手話認識できないかな？

聴覚障碍者にとっての窓口の不便さ

	社会的な情勢	金融機関の状況
現状	・改正障害者基本法では、手話が「言語」と規定 ・公共施設でのバリアフリー化、障害を持つ方が情報を円滑に利用できるようにすることが求められる	・金融機関ではインターネットでサービスの予約や申し込みが出来る機会が増加してきた
課題	・行政に比べ民間の窓口では、手話に対応できるスタッフが少ないなど、対応が追いついていないのが現状	・本人確認のために未だ窓口での手続きが必要なケースも多く存在している

聴覚障害者が金融機関の対面窓口を利用する際に、より円滑な意思疎通が促進され、バリアフリーな社会の構築に貢献できるよう、手話を日本語へと自動変換する「コミュニケーション支援システム」を紹介します

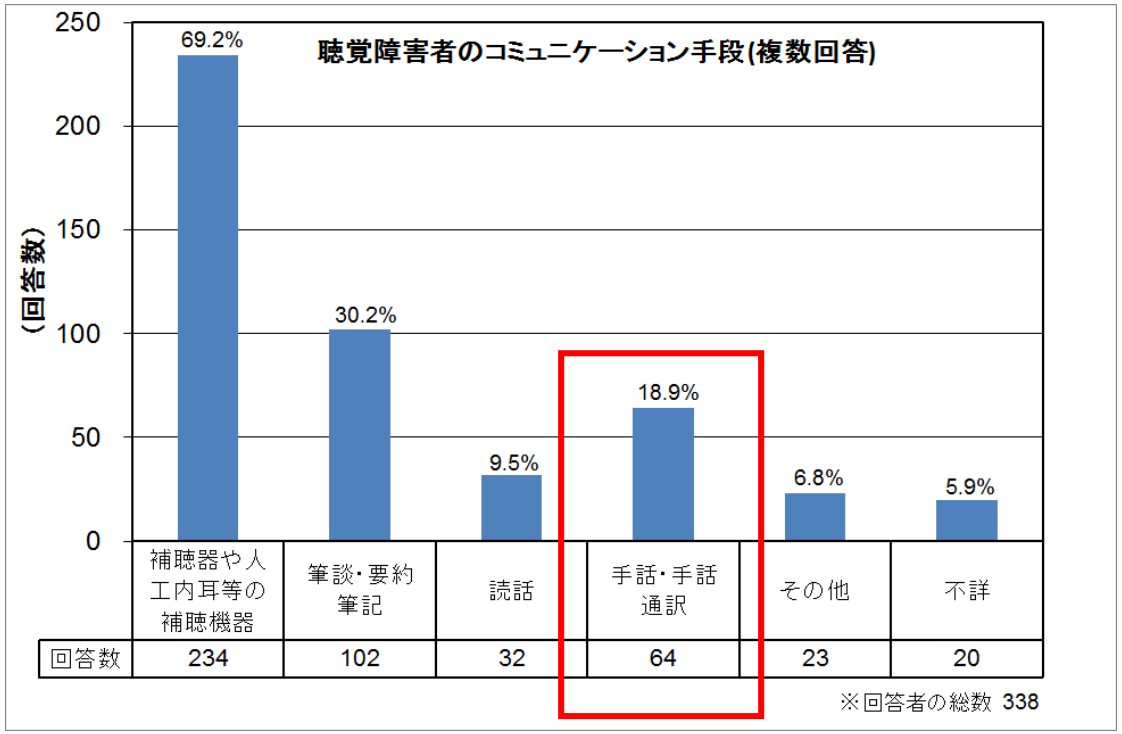
ご参考：ライフデザインレポート　第一生命経済研究所　水野映子氏「対面・非対面窓口の障害者等にとっての不便さ」（2010年1月）
―一般消費者・聴覚障害者・視覚障害者を対象とするアンケート調査結果より―

みなさんに質問です

- 手話を母語として使用している方は日本に何人くらいいると思いますか？

おおよそ６万人です

- 32.4万人中の18.9%

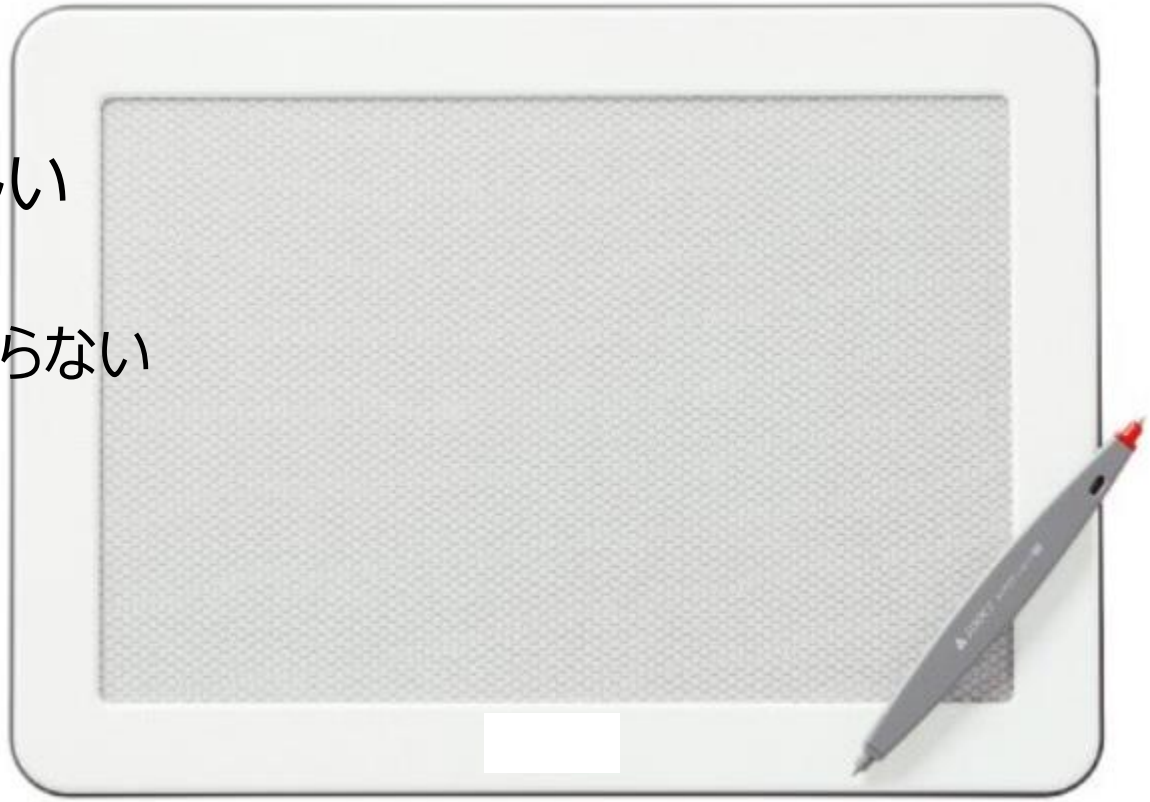


厚生労働省「平成１８年身体障害児・者実態調査結果」（２００８年）から作成

手話は日本語とは異なる「言語」

- 語順が異なる
- 助詞にあたるものがない

- 筆談ボードでのコミュニケーションは難しい
  →日本語を書く必要がある→ストレス
  →簡単なことも時間が非常にかかる→伝わらない



日本における手話に関する取組

- 筆談ボード
- リレー電話サービス
- 手話通訳
- タブレットを利用した手話通訳サービス（ShuR）

シュアール

http://shur.jp/

障碍者権利保護条約

2013年12月4日締結の国会承認

日本と国際社会の平和と安定に向けた取組

障害者の権利に関する条約
(Convention on the Rights of Persons with Disabilities)

平成26年2月21日

ツイート 14 いいね! 416
メール

障害者権利条約は，障害者の人権及び基本的自由の享有を確保し，障害者の固有の尊厳の尊重を促進することを目的として，障害者の権利の実現のための措置等について定める条約です。

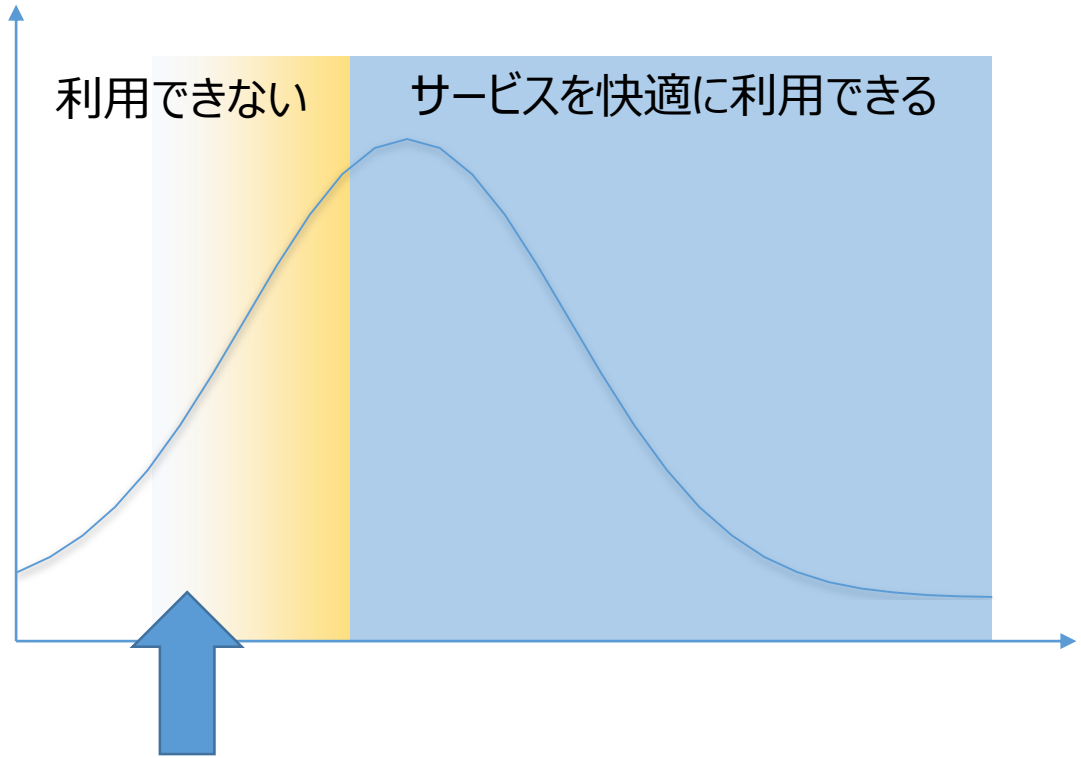
この条約の主な内容としては，(1) 一般原則(障害者の尊厳，自律及び自立の尊重，無差別，社会への完全かつ効果的な参加及び包容等)，(2) 一般的義務(合理的配慮の実施を怠ることを含め，障害に基づくいかなる差別もなしに，すべての障害者のあらゆる人権及び基本的自由を完全に実現することを確保し，及び促進すること等)，(3) 障害者の権利実現のための措置(身体の自由，拷問の禁止，表現の自由等の自由権的権利及び教育，労働等の社会権的権利について締約国がとるべき措置等を規定。社会権的権利の実現については漸進的に達成することを許容)，(4) 条約の実施のための仕組み(条約の実施及び監視のための国内の枠組みの設置。障害者の権利に関する委員会における各締約国からの報告の検討)，となっています。

障害者権利条約は，2006年12月13日に国連総会において採択され，2008年5月3日に発効しました。我が国は2007年9月28日に，高村正彦外務大臣(当時)がこの条約に署名し，2014年1月20日に，批准書を寄託しました。

http://www.mofa.go.jp/mofaj/gaiko/jinken/index_shogaisha.html

世の中のサービス

- 事業者は過度の負担にならない限り障碍を持つ人の権利を守る



権利を保障する合理的配慮が求められる

手話に関する行政の取り組み

- 鳥取県の手話言語条例、石狩市の手話基本条例
- 手話講習会（江戸川区など）
- 実社会における手話に対する理解、対応は今後の課題

http://www.pref.tottori.lg.jp/220879.htm

http://www.city.ishikari.hokkaido.jp/citizen/life/syougais05016.html

http://www.city.edogawa.tokyo.jp/kenko/shogaisha/shuwamanabi.html

手話の例

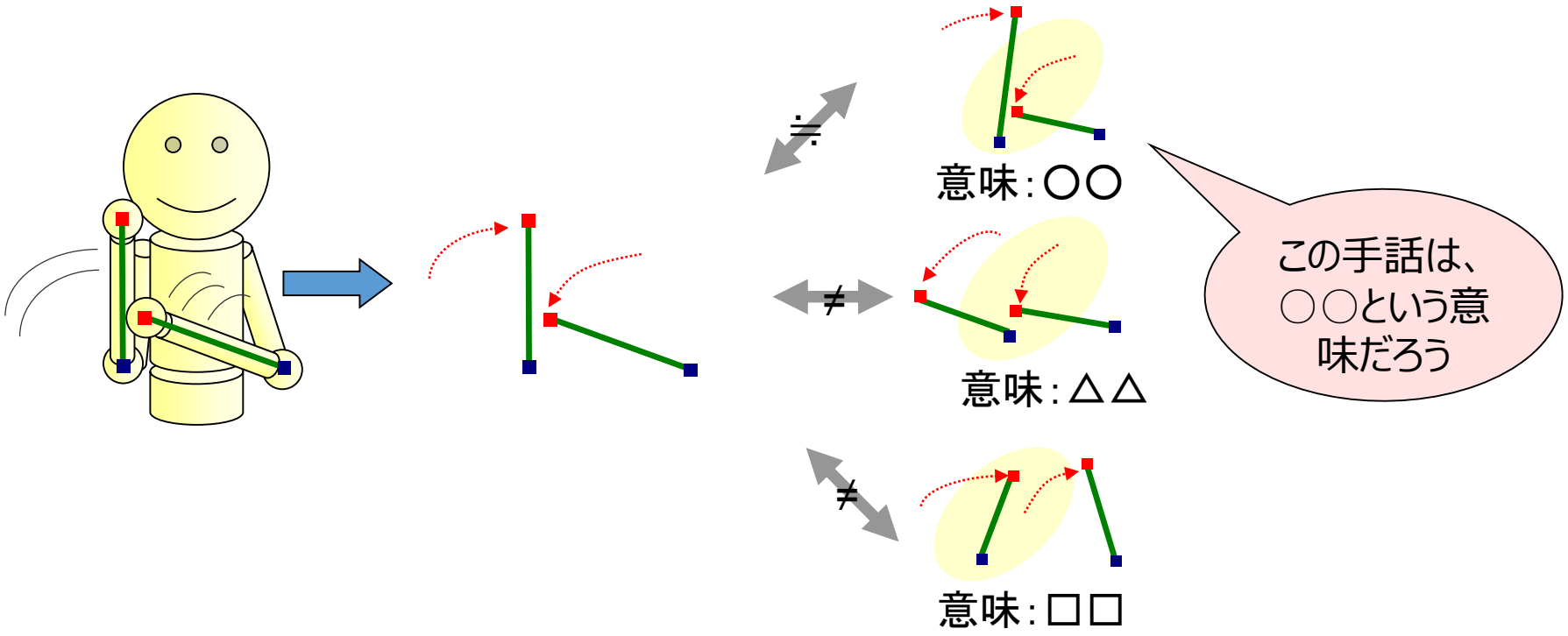
- いくつか手話を紹介します。

手話

- 手話の基本要素は位置・動き・形
- 表情やうなずきも重要な要素
- 口の形を使用することもある

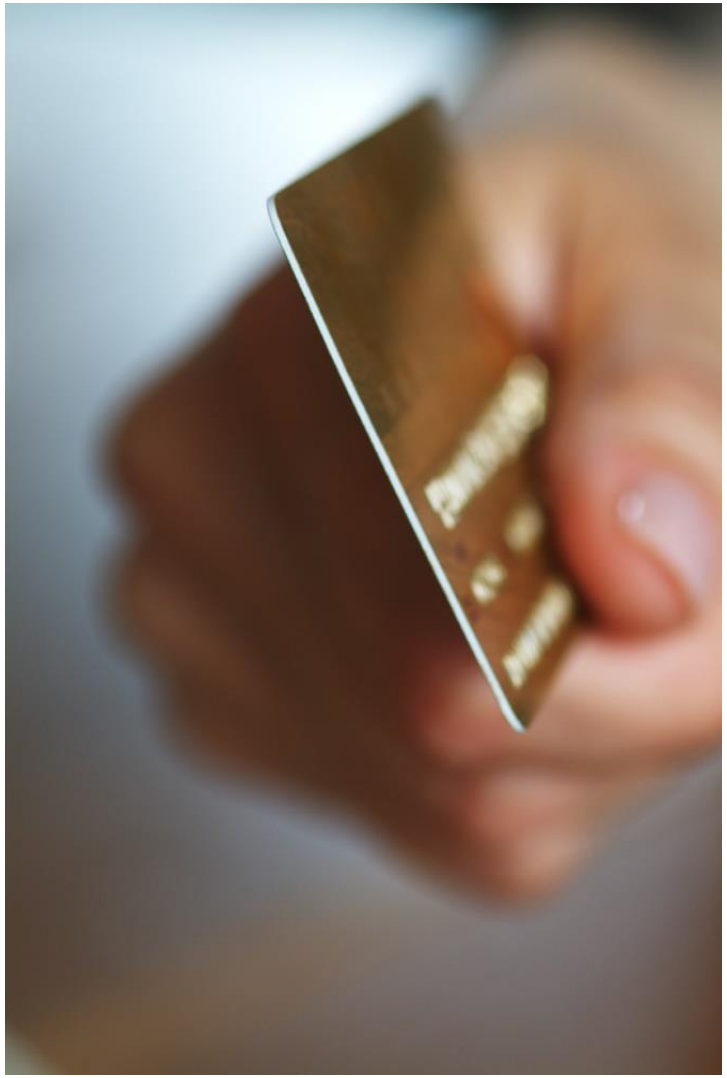
腕の「動き」のみを使用

- 動きだけでも会話の８割程度はカバー可能
- 数字は難しい



デモ

- 銀行窓口
- キャッシュカードを紛失
- 電話で手続き可能だが、、、
- 窓口に来るしかない



デモの説明

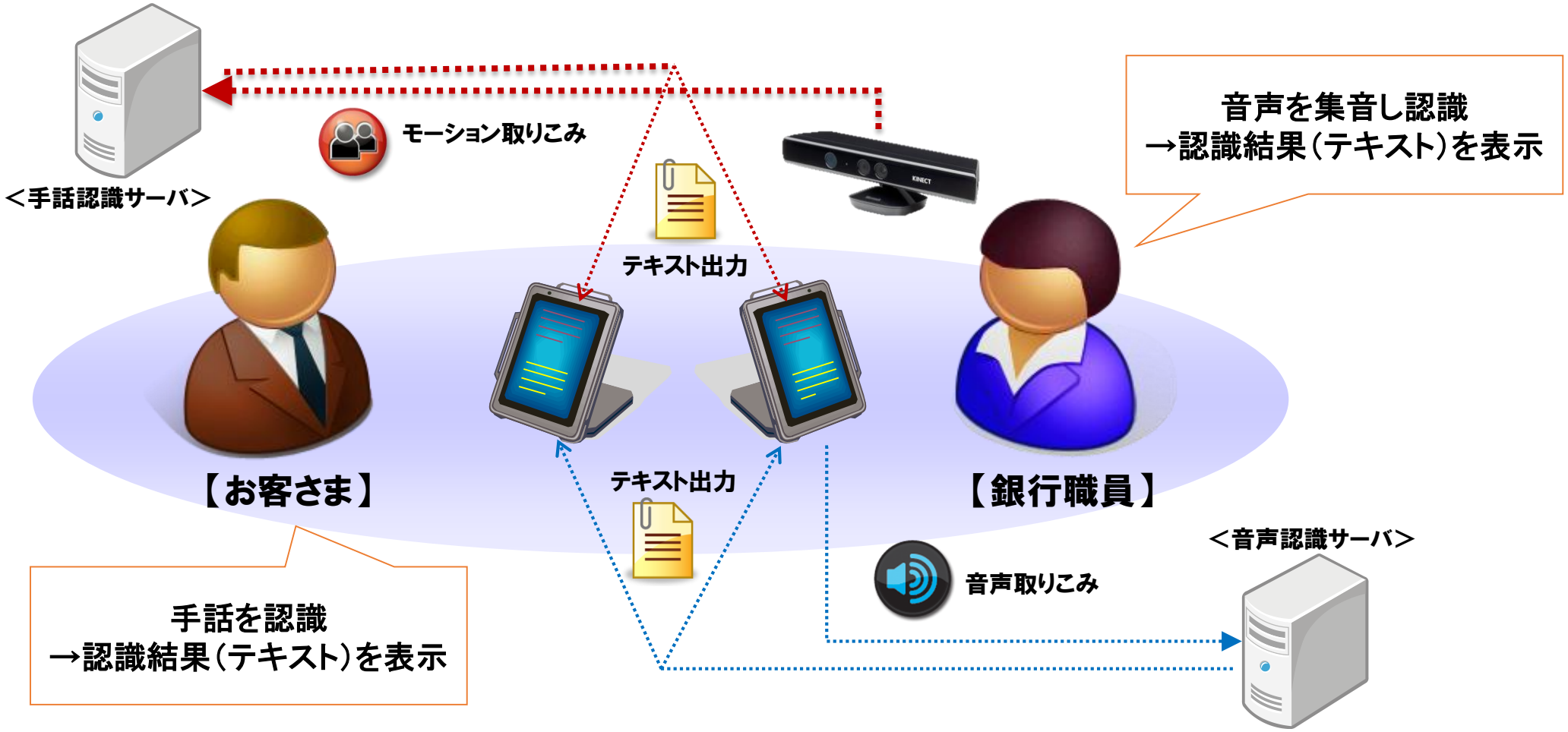
- キャッシュカードをなくしたので再発行をお願いしたい

→カード　なくなる　また　作る　お願い

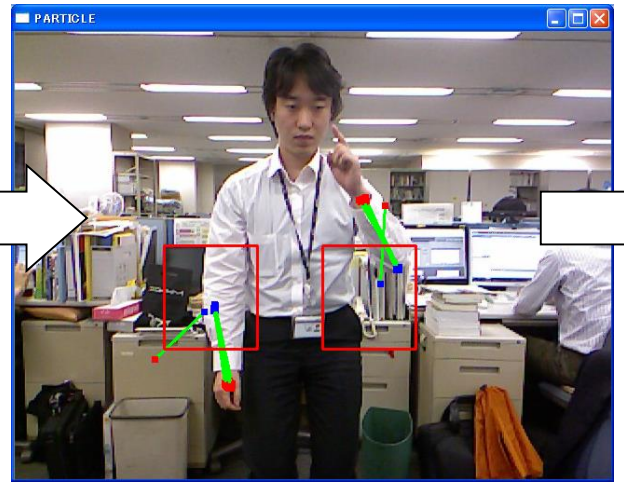
- 手話は語順や助詞の使い方などが日本語と大きく異なる

想定する利用状況

- 音声認識と組み合わせることで聴者とろう者の対話を可能とする

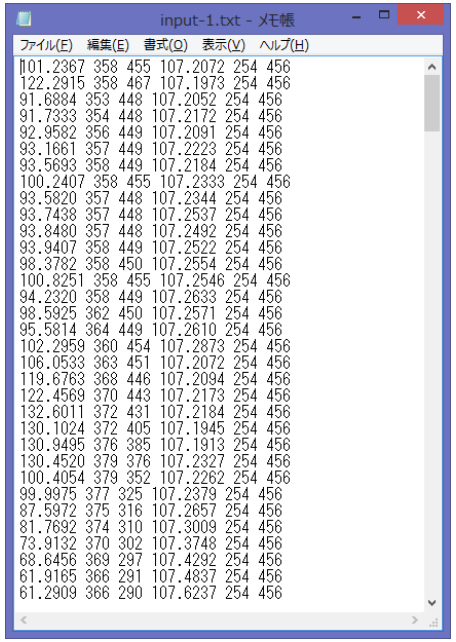


手話認識の流れ



肘位置からみた手の相対座標を取得

テキストファイル



差分を計算
(特徴量の計算)

HTKへ

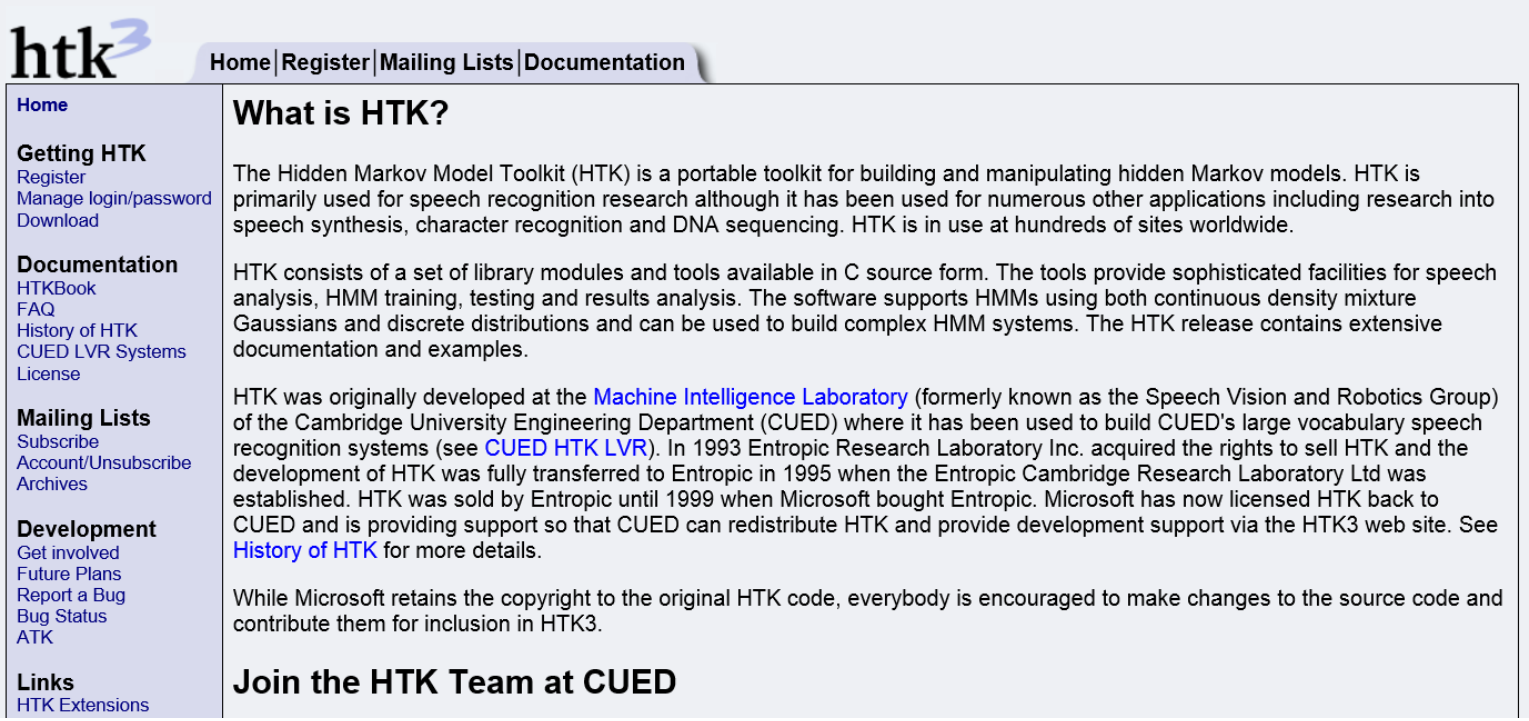
学習データの作成

- 学習させる文章を作成
- 手話翻訳者に通訳を依頼
- 翻訳した手話を撮影
- 別の人で手話を撮影（１０人×１０回）
- 撮影した手話を単語ごとに区切る（ラベリング）
- 機械学習を一昼夜

手話認識に利用しているツール　HTK

- HTK(Hidden Markov Model Toolkit)

→研究目的のみ利用可能



htk3

Home | Register | Mailing Lists | Documentation

Home

Getting HTK
Register
Manage login/password
Download

Documentation
HTKBook
FAQ
History of HTK
CUED LVR Systems
License

Mailing Lists
Subscribe
Account/Unsubscribe
Archives

Development
Get involved
Future Plans
Report a Bug
Bug Status
ATK

Links
HTK Extensions

What is HTK?

The Hidden Markov Model Toolkit (HTK) is a portable toolkit for building and manipulating hidden Markov models. HTK is primarily used for speech recognition research although it has been used for numerous other applications including research into speech synthesis, character recognition and DNA sequencing. HTK is in use at hundreds of sites worldwide.

HTK consists of a set of library modules and tools available in C source form. The tools provide sophisticated facilities for speech analysis, HMM training, testing and results analysis. The software supports HMMs using both continuous density mixture Gaussians and discrete distributions and can be used to build complex HMM systems. The HTK release contains extensive documentation and examples.

HTK was originally developed at the Machine Intelligence Laboratory (formerly known as the Speech Vision and Robotics Group) of the Cambridge University Engineering Department (CUED) where it has been used to build CUED's large vocabulary speech recognition systems (see CUED HTK LVR). In 1993 Entropic Research Laboratory Inc. acquired the rights to sell HTK and the development of HTK was fully transferred to Entropic in 1995 when the Entropic Cambridge Research Laboratory Ltd was established. HTK was sold by Entropic until 1999 when Microsoft bought Entropic. Microsoft has now licensed HTK back to CUED and is providing support so that CUED can redistribute HTK and provide development support via the HTK3 web site. See History of HTK for more details.

While Microsoft retains the copyright to the original HTK code, everybody is encouraged to make changes to the source code and contribute them for inclusion in HTK3.

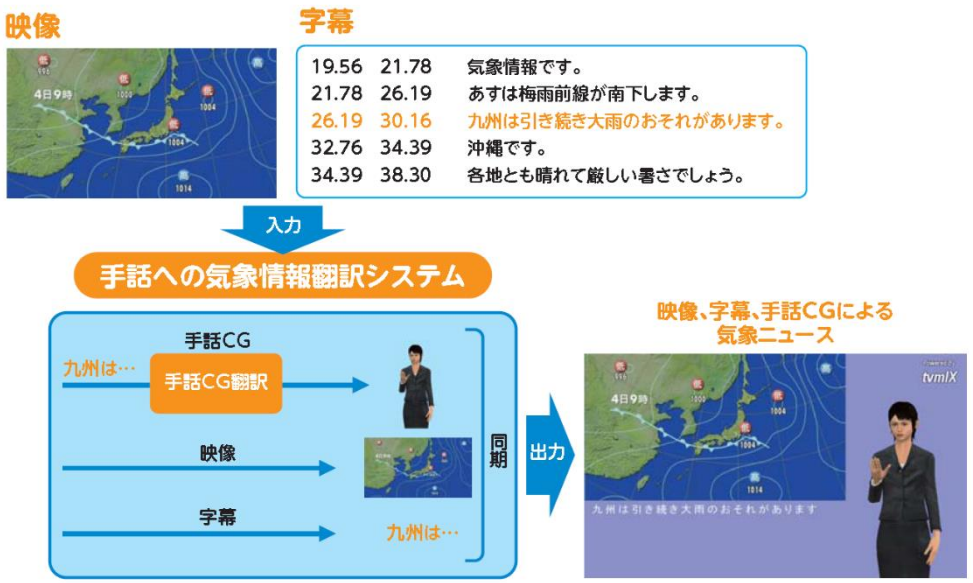
Join the HTK Team at CUED

http://htk.eng.cam.ac.uk/

日本語→手話

- NHK技研と工学院大学
- 天気予報を手話で伝える研究
- 現状は天気予報に特化

NHK技研の事例



映像と字幕から手話CGを生成する気象情報翻訳システム

http://www.nhk.or.jp/strl/open2013/index2.html

Microsoft Research

- 中国手話の認識と翻訳

Microsoft Translator | Choose language

Microsoft Research

Search Microsoft Research

Our research Connections Careers About us

What we do Opportunities Research tools

Kinect Sign Language Translator expands communication possibilities

Publication date: October 30, 2013

Kinect Sign Language Translator user interface

An estimated 360 million people worldwide suffer from hearing loss. Because the majority of hearing individuals do not understand sign language, communication between the hearing and the deaf can be challenging. While other communication methods exist, researchers hope to make translation even easier with a cost-effective, efficient prototype that translates sign language into spoken language—and spoken language into sign language—in real time.

Asia answers the call

Students of the special education school at Beijing Union University try out the Kinect Sign Language Translator prototype.

Dedicated researchers in China have created the Kinect Sign Language Translator, a prototype system that understands the gestures of sign language and converts them to spoken and written language—and vice versa. The system captures a conversation from both sides: it displays the signer and renders a written and spoken translation of the sign language in real-time, and it also takes the non-signer's spoken words and turns them into accurate, understandable sign language. An avatar on the screen represents the non-signer and makes the appropriate sign languages gestures.

This project was a result of collaboration, facilitated by Microsoft Research Connections, between the Chinese Academy of Sciences, Beijing Union University, and Microsoft Research Asia, each of which made crucial contributions.

A Microsoft Research Connections-funded project supporting advanced technology research

Read the complete case study (PDF file, 682 KB)

Video

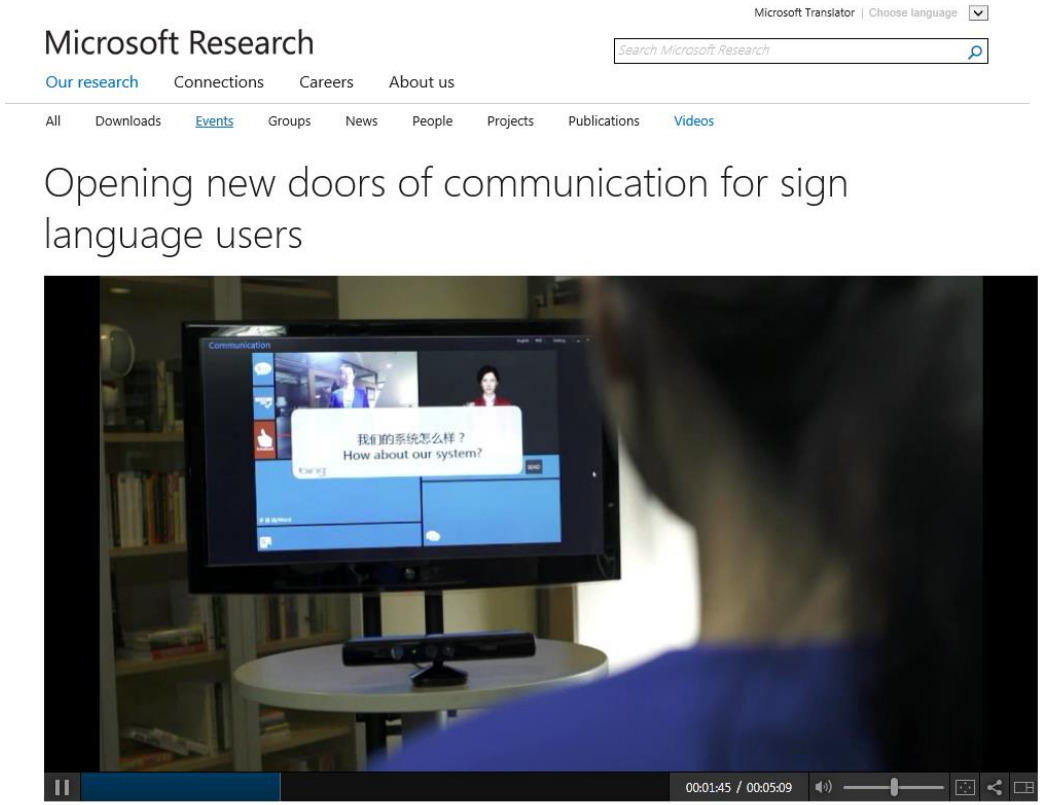
Opening new doors of communication for sign language users

Technology

- Microsoft Kinect for Windows
- Kinect for Windows Dev Center

More about the translator

- Kinect Sign Language Translator - part 1 (Microsoft Research Connections blog)
- Kinect Sign Language Translator - part 2 (Microsoft Research Connections blog)
- Digital Assistance for Sign Language Users



Microsoft Translator | Choose language

Microsoft Research

Search Microsoft Research

Our research Connections Careers About us

All Downloads Events Groups News People Projects Publications Videos

Opening new doors of communication for sign language users

http://research.microsoft.com/en-us/collaboration/stories/kinect-sign-language-translator.aspx

Kinect ver.2 Preview

- プレビュープログラムに参加
- $399でKinect ver.2 ~~alpha~~版とSDKを試用可能

beta

This is preliminary software and/or hardware and APIs are preliminary and subject to change.

どんなことができる？

- Ver.1で可能なことはVer.2でも可能
- 関節に首、親指が追加
- 関節情報に向きが追加
- グーチョキパーがとれる
- 顔の状態をいろいろと取得できる

関節向きについてはSDK
ver.1.5以降でも取得可能

手話認識とKinect ver.2

- 表現の範囲が大きく広がった！
- Ver.2を用いた認識を前提に学習データを作成
- 最初は手の形、関節の向きを利用
- 順次、うなづき、表情を追加していく予定

学習データの作成が大きな課題！