Domain Adaptation for Upper Body Pose Tracking in Signed TV Broadcasts

James Charles¹, Tomas Pfister², Derek Magee¹, David Hogg¹, Andrew Zisserman²

¹School of Computing
University of Leeds

²Visual Geometry Group
University of Oxford
Motivation: **Automatic sign language recognition**

**Idea:** automatically learn signs by watching TV

using weak supervision from subtitles
Motivation: Automatic sign language recognition

Idea: automatically learn signs by watching TV
using weak supervision from subtitles
Introduction

Given a large amount of annotated joint training data

Accurate real-time joint tracking with Random Forest

Method is similar to Kinect, but without depth
Problem

There is only training data for one *type of signer* (domain)
Problem

There is only training data for one type of signer (domain)

Can’t generalise to some new signers (new domains)

Different build
Different hair length
Different sleeve length
Problem

There is only training data for one type of signer (domain)

Can’t generalise to some new signers (new domains)

Different build  Different hair length  Different sleeve length

Example: signers with short sleeves
One solution

Annotate new training data
One solution

Annotate new training data

**Problem:** annotation is very expensive!
One solution

**Annotate new training data**

**Problem:** annotation is very expensive!

**Several hours** of manual annotation per video
Key idea in this paper

Synthesise training data for new domain by reusing existing training data

Several hours of manual annotation per video
Key idea in this paper

Synthesise training data for new domain by reusing existing training data

Several hours of manual annotation per video

Annotation = 1 NUMBER
Example: Tracking signers wearing short sleeves

Original forests
Example: Tracking signers wearing short sleeves

Personalised forests

Training COMPLETELY AUTOMATIC
Related work on synthesising training data for discriminative prediction

- **Kinect (CVPR’11):** synthesise depth images

- **Everingham & Zisserman ICCV’05:** synthesise using head model

- **Lepetit & Fua PAMI’06:** synthesise affine distortions to classify interest points

- **Agarwal & Triggs PAMI’06:** synthesise human pose for pose estimation

- **Pishchulin et al. CVPR’12:** synthesise photorealistic images with different human poses
METHOD

Running example: signers wearing sleeves of various length

Sleeve length (decreasing)
Existing method for signers with long sleeves

✅ Accurate real-time tracking with Random Forest

Original  Skin/Torso colour posterior  Joint tracking
Existing method for signers with long sleeves

✓ Accurate real-time tracking with Random Forest

Why colour posterior?

✓ Abstracts away colour of clothes
Existing method for signers with long sleeves

✓ Accurate real-time tracking with Random Forest

Why colour posterior?

Why not just segmentation?

✓ Abstracts away colour of clothes

✗ Cannot locate hands
Existing method for signers with long sleeves

✓ Accurate real-time tracking with Random Forest

Why colour posterior?
✓ Abstracts away colour of clothes

Why not just segmentation?
× Cannot locate hands

Problem with short sleeves:
× Colour posterior looks very different
Existing method for signers with long sleeves

Solution:
Synthesise training data for short sleeved signers using already tracked long sleeved videos

+ joint positions

Abstracts away colour of clothes

Cannot locate hands

Colour posterior looks very different
Method overview

Input:

- Video from target domain
- Source domain colour posterior
- Target label (sleeve length)
- Colour posterior & joint positions in source domain

0.5
Method overview

Input:
- Video from target domain
- Source domain colour posterior
- Target label (sleeve length)
- Colour posterior & joint positions in source domain

Output: synthesised training data (with target sleeve length)
Method overview

**Input:**
- Video from target domain
- Source domain colour posterior
- Target label (sleeve length)
- Colour posterior & joint positions in source domain

**Output:** synthesised training data (with target sleeve length)

1. Train **joint tracker** for target domain
Method overview

**Input:**
- Video from target domain
- Source domain colour posterior
- Target label (sleeve length)
- Colour posterior & joint positions in source domain

**Output:** synthesised training data (with target sleeve length)

1. Train joint tracker for target domain
2. Apply tracker to target domain
Method detail

1. Colour posterior synthesis
   Synthesise bare arms on signers wearing long sleeves

   Initial arm templates
   Source signer (Known joint locations)
   Initial synthesised CP image (Known joint locations)
1. Colour posterior synthesis

Synthesise bare arms on signers wearing long sleeves

- Initial arm templates
  - Upper
  - Lower

Source signer (Known joint locations)

Initial synthesised CP image (Known joint locations)

2. Template personalisation

Personalise arm templates by refining initial colour posterior

- Initial synthesised CP image

- Initial arm templates
  - Upper
  - Lower

- Personalised arm templates
  - Upper
  - Lower
1. Colour posterior synthesis

Synthesise bare arms on signers wearing long sleeves

Initial arm templates

Source signer (Known joint locations)

Initial synthesised CP image (Known joint locations)
1. Colour posterior synthesis

Synthesise bare arms on signers wearing long sleeves

Source signer (Known joint locations)

Initial synthesised CP image (Known joint locations)
1. Colour posterior synthesis

Synthesise bare arms on signers wearing long sleeves

2. Template personalisation

Personalise arm templates by refining initial colour posterior
2. Template personalisation

**Personalise** arm templates by refining initial colour posterior

- Initial synthesised CP image
- Initial arm templates (Upper, Lower)
- Personalised arm templates (Upper, Lower)
Method detail

2. Template personalisation

Personalise arm templates by refining initial colour posterior

Initial synthesised CP image

Initial arm templates

Personalised arm templates
Method – Template personalisation

1: Train Random Forest on synthetic colour posterior & apply to target signer
Method – Template personalisation

1: Train Random Forest on synthetic colour posterior & apply to target signer

2: Sample and verify joint locations
Method – Template personalisation

1: Train **Random Forest** on synthetic colour posterior & **apply to target** signer

2: **Sample and verify** joint locations

- RGB input
- Colour posterior image
- Joint confidence map
- Most confident joints
Method – Template personalisation

1: Train Random Forest on synthetic colour posterior & apply to target signer

2: Sample and verify joint locations

RGB input  Colour posterior image  Joint confidence map  Most confident joints

Sampling

Initial arm templates
Upper  Lower

Torso/head
Method – Template personalisation

1: Train **Random Forest** on synthetic colour posterior & apply to target signer

2: Sample and verify joint locations

- RGB input
- Colour posterior image
- Joint confidence map
- Most confident joints

- Best sample
- Best sample (most similar to colour posterior image)

- Initial arm templates
- Upper
- Lower
- Torso/head

Tuesday, 10 September 13
Method – Template personalisation

1: Train Random Forest on synthetic colour posterior & apply to target signer

2: Sample and verify joint locations

- RGB input
- Colour posterior image
- Joint confidence map
- Most confident joints

Best sample
Best sample (most similar to colour posterior image)

3: Personalise arm templates
Use current detected body joints to extract windows of arms for training
Method – Template personalisation

1: Train Random Forest on synthetic colour posterior & apply to target signer

2: Sample and verify joint locations

- RGB input
- Colour posterior image
- Joint confidence map
- Most confident joints

Best sample
Best sample (most similar to colour posterior image)

3: Personalise arm templates
Use current detected body joints to extract windows of arms for training
Method detail

1. Colour posterior synthesis
   Synthesise bare arms on signers wearing long sleeves

   ![Initial arm templates](image)
   +
   ![Source signer](image)
   (Known joint locations)
   →
   ![Initial synthesised CP image](image)
   (Known joint locations)

2. Template personalisation
   Personlise arm templates by refining initial colour posterior

   ![Initial synthesised CP image](image)
   +
   ![Initial arm templates](image)
   →
   ![Personalised arm templates](image)
Method detail

1. Colour posterior synthesis
   Synthesise bare arms on signers wearing long sleeves

   ![Personalised arm templates](image1)
   ![Source signer](image2)
   ![Personalised CP image](image3)

2. Template personalisation
   Personalise arm templates by refining initial colour posterior

   ![Initial synthesised CP image](image4)
   ![Initial arm templates](image5)
   ![Personalised arm templates](image6)
Method detail

Iterate until convergence

1. Colour posterior synthesis
   Synthesise bare arms on signers wearing long sleeves

   Personalised arm templates
   Upper
   Lower

   +

   Source signer
   (Known joint locations)

   →

   Personalised CP image

2. Template personalisation
   Personalise arm templates by refining initial colour posterior

   Initial synthesised CP image

   Initial arm templates
   Upper
   Lower

   +

   →

   Personalised arm templates
   Upper
   Lower
Method – example output

Target

Sleeve length = 0.2

Colour posterior image

Sleeve length = 0.5

Synthesis of target's arms
RESULTS

Running example: signers wearing sleeves of various length

Sleeve length (decreasing)
Dataset

Source domain (training data)
Quantitative results

Evaluation measure

5 pixels
Quantitative results

Evaluation measure

Effect of template refinement

Wrist

Accuracy (%)

Distance from GT (px)

Original
Initial
Personal

5 pixels
Quantitative results

Evaluation measure

Accuracy (averaged over all videos)
Conclusion

Track upper body pose in new domains by transferring from existing training material

**Future work:**

Automatic method for measuring side-information (e.g. sleeve length)