Domain Adaptation for Upper Body Pose Tracking in Signed TV Broadcasts

James Charles
j.charles@leeds.ac.uk
Tomas Pfister
tp@robots.ox.ac.uk
Derek Magee
d.r.magee@leeds.ac.uk
David Hogg
d.c.hogg@leeds.ac.uk
Andrew Zisserman
az@robots.ox.ac.uk

The objective of this work is to estimate upper body pose for signers in TV broadcasts. Given suitable training data, the pose is estimated using a random forest body joint detector (similar to that used by Shotton et al. [4], though without requiring depth measurements). However, obtaining such training data can be costly.

The novelty of this paper is a method of transfer learning that is able to harness existing training data and use it for new domains. Our contributions are: (i) a method for adapting existing training data to generate new training data by synthesis for signers with different appearances, and (ii) a method for personalising training data. In addition we add features to the tracker of Pfister et al. [3] to improve tracking performance. As a case study we show how the appearance of the arms for different clothing, specifically short and long sleeved clothes, can be modelled to obtain person-specific trackers.

Motivation. The main motivation for tracking the pose is to automatically learn to recognise sign language [1], where upper body layout [2] is of great importance. Tracking is non-trivial due to changing background (the signer is overlaid on the broadcast video, as shown in Figure 1), and also because the signer changes between broadcast and so there are variations in shape and clothing. Furthermore, large quantities of video data are required to learn even a modest number of signs, implying the tracker has to be reliable over long video sequences and for multiple signers, and require little or no human supervision.

Summary of method. The tracker in our previous work [3] was only designed for application on a domain of signers wearing long sleeved clothing as shown in Figure 1(a). Our target domain is signers wearing short sleeved clothes, shown in Figure 1(b). We generate training data for the target domain from a source domain of long sleeved clothes in two stages. Stage 1 of our method (output shown in Figure 2) uses material from the source domain to generate semi-synthetic training data of signers wearing sleeves of a specific length. With this, one can retrain multiple general upper body pose trackers for a particular sleeve length, as shown in Figure 1(c). Stage 2 (shown in Figures 3 and 4) re-synthesises the training data to become signer specific. Then the refined semi-synthetic data is used to train a personalised tracker tuned to a particular person’s arm shape and sleeve length, as shown in Figure 1(d). Each step contributes a significant boost to pose estimation performance, as shown in Figure 1(e).