Colour Invariant Head Pose Classification in Low Resolution Video

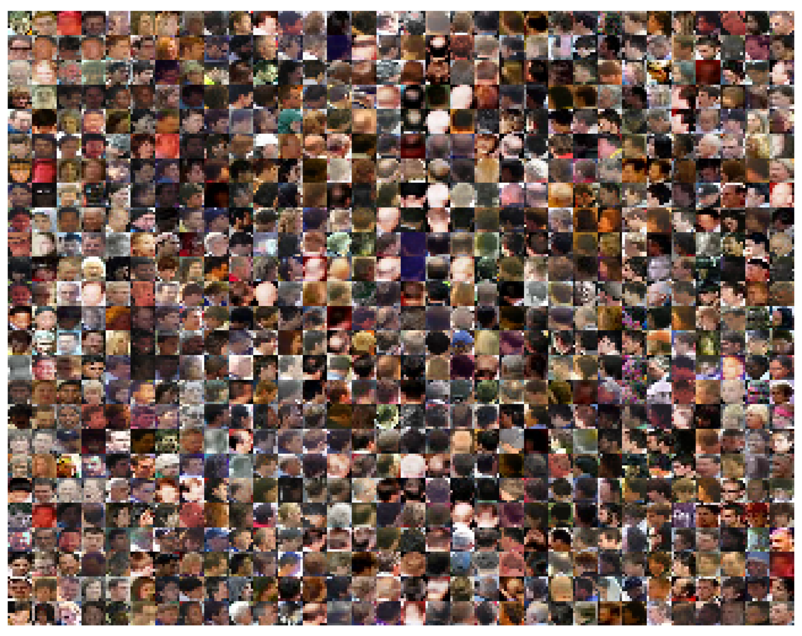
Ben Benfold and Ian Reid Oxford University

Problem

We aim to estimate the head pose of people in security video. There are a number challenges that a head pose estimator must cope with to be effective in real-world situations:

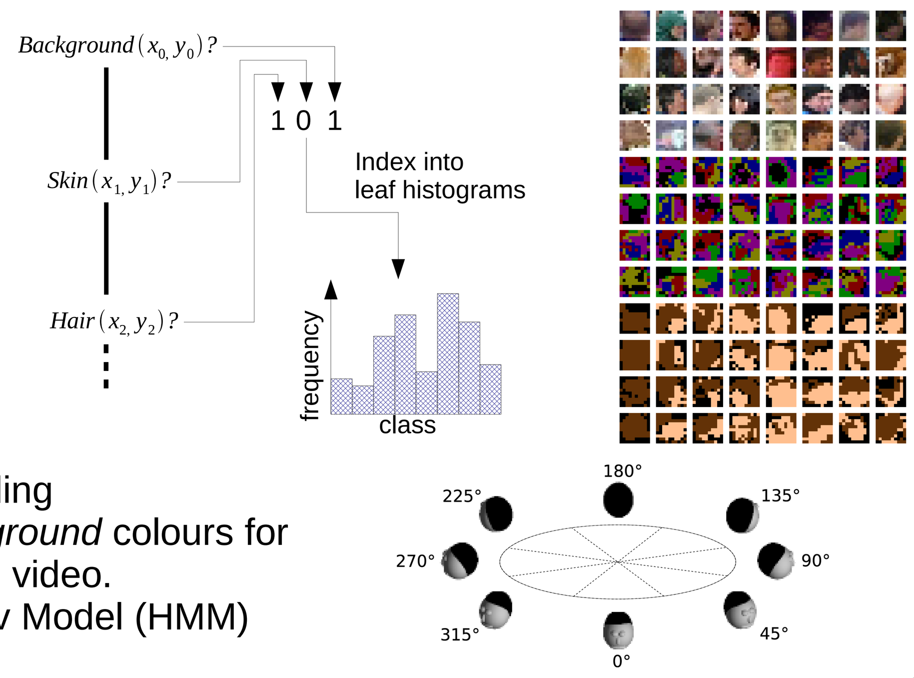
- Small head images due to wide fields of view
- A wide variety of hair and skin colours
- Different hair styles, clothing and backgrounds
- Variations in lighting conditions

Most existing classifiers are susceptible to these variations - we aim to find a solution that can cope with them.



Solution

- Use randomised ferns to categorise images into one of eight direction classes depending on the head pan angle.
- Segment head images before classification to provide colour invariance.
- Train ferns using data that has *hair*, *skin* and *background* regions hand labelled.
- Test labelling hypotheses for test images using the ferns to find the most likely labelling
- Use online learning of *hair*, *skin* and *background* colours for individual people to improve estimations in video.
- Model head motion using a Hidden Markov Model (HMM)



Classification

Notation

h	A hypothesis which maps each segment to one of *hair*, *skin* or *background*.
d_h	The branch outcomes resulting in an image reaching a particular leaf node.
S, s	The set of all segments and a single segment respectively.
C, c	The set of all direction classes and a single class respectively.
Y	The learnt colour histograms for the three labels.
$Y_{h(s)}$	The colour histogram for segment *s* under hypothesis *h*.
$B(a,b)$	The Bhattacharyya coefficient calculated between histograms and *b*.

1 Identify Colour Segments

The image is first segmented to remove instance-specific colour information whilst leaving most of the important structural information.

Head images are cropped and scaled to 10 pixels square before being segmented into six different regions based on colour. Six groups were formed using k-means clustering in a normalised YUV colour space to ensure that the hair and skin were separated in cases where lighting conditions caused their colour distributions to be multimodal.

2 Find Best Labelling Hypothesis

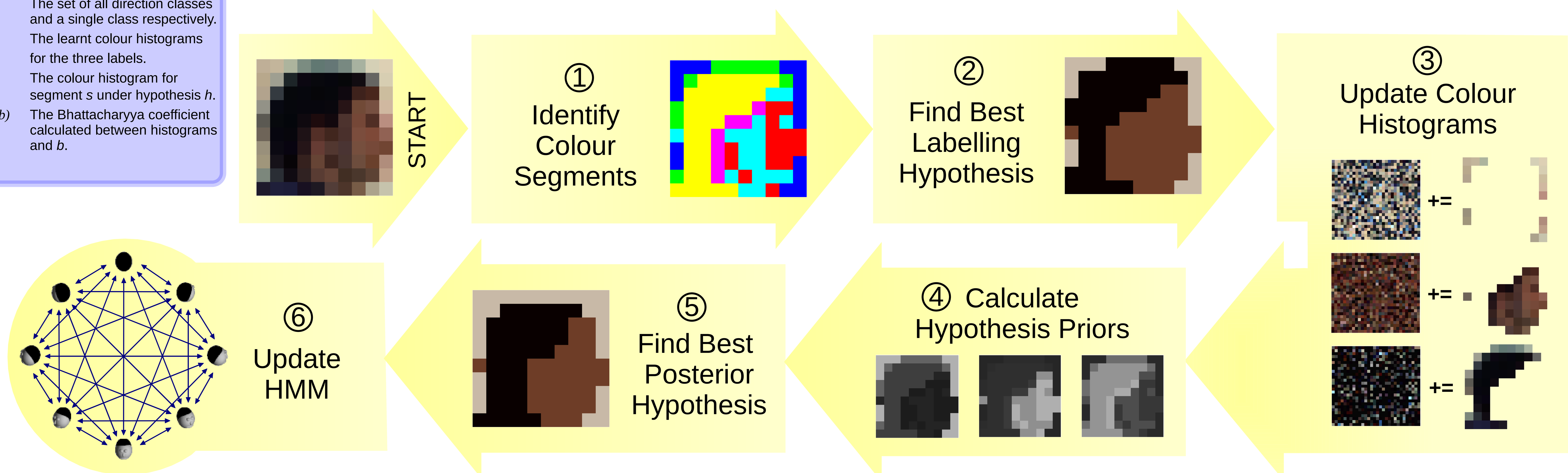
Head images reach different fern leaf nodes under different labelling hypotheses. The proportion of training data that reached the leaf provides an estimate of the probability that the hypothesis is correct and the amount in each histogram bin indicates the relative probabilities of the image belonging to each class. The joint probability for hypotheses and classes is calculated using:

$$P(h, c|d_h) = \frac{P(d_h|h, c)P(c)P(h)}{\sum_{c_i \in C} P(d_h|h, c_i)P(c_i)P(h) + P(d_h|\overline{h}, c_i)P(c_i)P(\overline{h})}$$

The outputs from 20 different ferns were marginalised over *c* and then

3 Update Colour Histograms

The classification accuracy can be improved by learning the colours represented by the labelled image segments. For every head image, the pixels labelled as *hair*, *skin* and *background* by the best hypothesis are added to corresponding colour histograms.



6 Update HMM

The pose of a head at any time provides information relating to the pose shortly after due to the physical constraints of human head motion. A Hidden Markov Model allows transitions between poses to be modelled

The class probabilities that are marginalised over all hypotheses provide observations in each frame. These observations are used to update the model to provide a filtered pose for every frame.

5 Find Best Posterior Hypothesis

The next step is to calculate an improved posterior hypothesis using the priors based on the learnt colour distributions. This time, the joint distribution is marginalised over all hypotheses to find the probabilities that the image belongs to each of the eight classes.

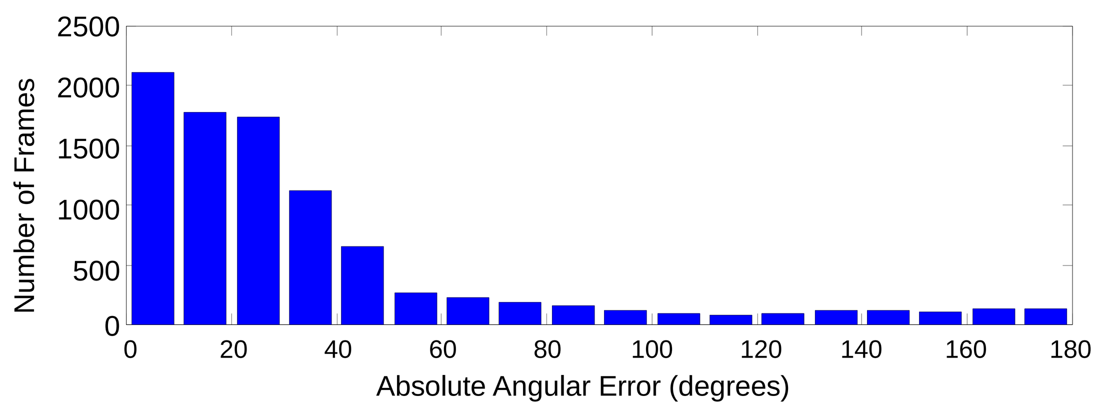
$$P(h, c|d_h, Y) = \frac{P(d_h|h, c)P(c)P(h|Y)}{\sum_{c_i \in C} P(d_h|h, c_i)P(c_i)P(h|Y) + P(d_h|\overline{h}, c_i)P(c_i)P(\overline{h}|Y)}$$

4 Calculate Hypothesis Priors

The colour histograms allow the hypotheses for the current frame to be weighted by comparing the colour distribution of each segment to each of the colour histograms using the Bhattacharyya coefficient. The overall probability that a hypothesis is correct is proportional to the product of the coefficients for each segment:

$$P(h|Y) \propto \prod_{s \in S} B(s, Y_{h(s)})$$

Evaluation



A classifier was trained using approximately 1000 images from still photos and then tested on a number of video sequences:

- Head pose was estimated in a total of 9260 head images
- Video sequences included sixteen actors with different hair and skin colours, and different lighting conditions
- Ground truth angles and head regions were hand labelled
- Twelve humans also estimated the head pose in every 100th frame to provide a performance comparison

The angular error in each frame was recorded both before and after the colour priors were introduced and both with and without HMM filtering.

Test Details	Mean Absolute Angular Error (degrees)
Initial Estimation	43.5
Estimation using priors	38.4
Estimation using priors and HMM	37.9
Human Performance	26.6

Conclusion

There are a number of key points which the research has demonstrated:

- Head pose can be accurately measured without making assumptions about hair and skin colours.
- Pose can be measured in images as small as ten pixels square.
- Hair and skin colour histograms can be obtained through online learning without any prior knowledge of their distributions.
- Humans still outperform computer vision based classifiers in low resolution images.
- Human performance provides a useful means by which datasets can be compared
- Real-time performance is easily achievable with each pose estimated in less than ten milliseconds on a modern desktop computer.

Contact Info

Ben Benfold bbenfold@robots.ox.ac.uk
Ian Reid ian@robots.ox.ac.uk

Department of Engineering Science
University of Oxford
OX1 3PJ
Oxford
UK

http://www.robots.ox.ac.uk/ActiveVision/

HERMES

Human Expressive Representations of Motion and their Evaluation in Sequences

This research was funded as part of the EU project HERMES (IST-027110)