# Reconstructing High Quality Face-Surfaces using Model Based Stereo

Brian Amberg†, Andrew Blake‡, Andrew Fitzgibbon‡, Sami Romdhani†, and Thomas Vetter†
University of Basel, Switzerland† Microsoft Research, Cambridge‡

## Contribution

We present a method to fit a detailed 3D morphable model to multiple images. Our formulation allows the fitting of the model without determining the lighting conditions and albedo of the face, making the system robust against difficult lighting situations and unmodelled albedo variations such as skin colour, moles, freckles and cast shadows.
The cost function employs

- The model shape prior
- A small number of landmarks for initialization
- A monocular silhouette distance cost
- A stereo colour cost

The optimisation consists of multiple runs of a non-linear minimizer. During each run the visibility of all sample points is assumed to stay constant. After some iterations the minimizer is stopped and visibility is reevaluated.

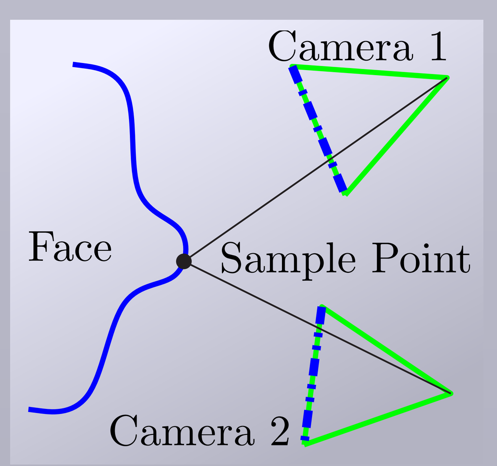
## Model

The linear morphable face model was created by registering 200 face scans and performing a PCA on the data matrix to fit a Gaussian probability to the data and reduce the dimensionality of the model.

## Silhouette Cost

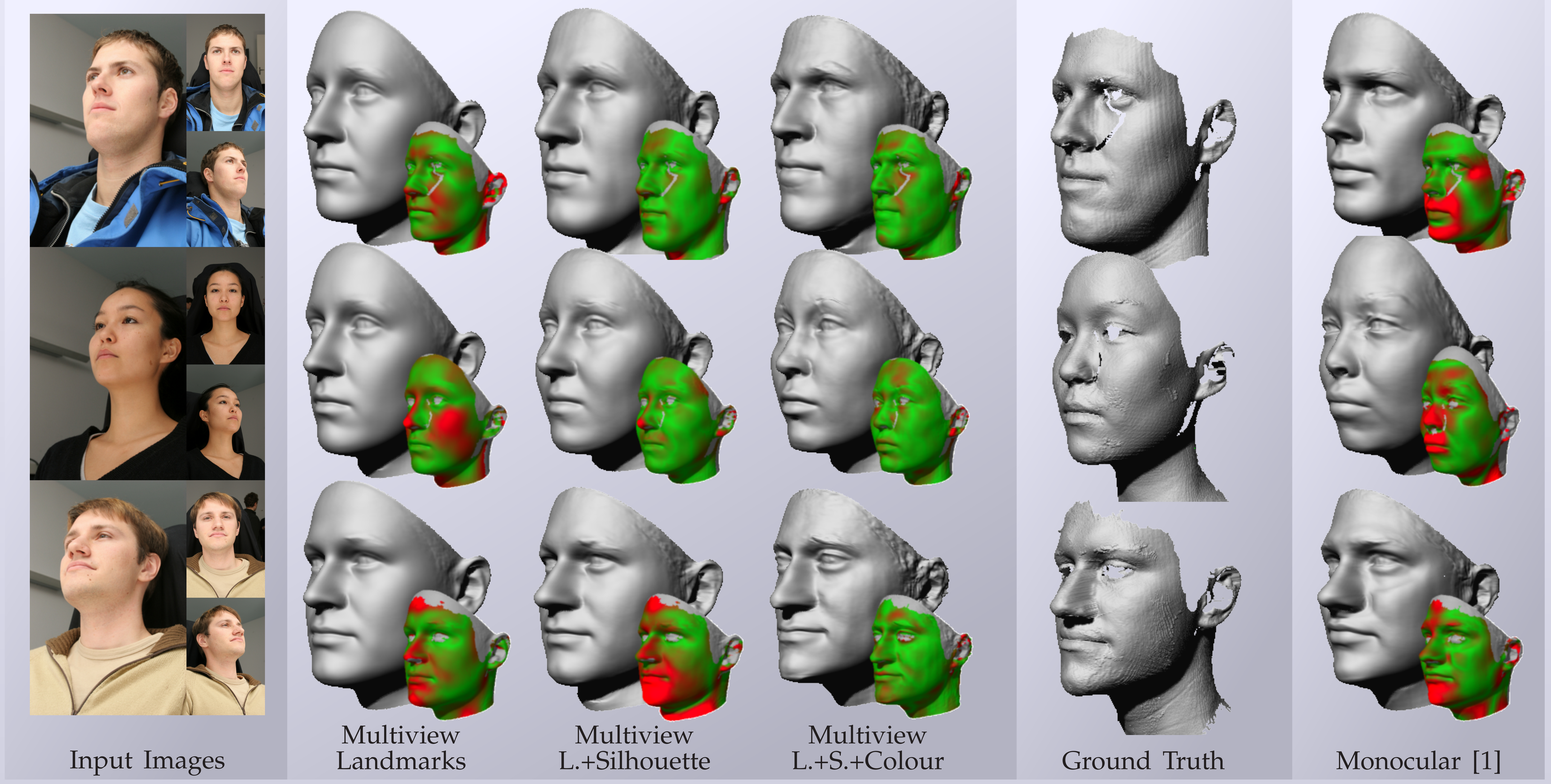

The silhouette cost measures the distance of the silhouette to image edges. An edge cost surface is created from the image, by combining the distance transforms of edge detections with different thresholds. The cost ist integrated over the projection of 3D sample points at the silhouette of the hypotheses.

## Colour Reprojection Cost



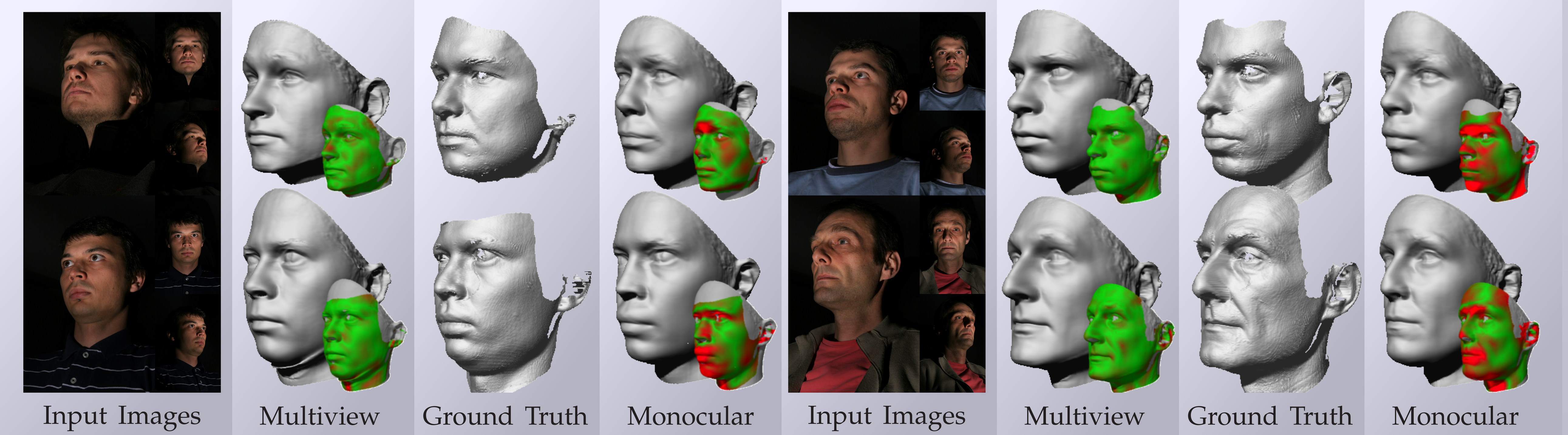

The colour reprojection cost measures the image colour difference between the projected positions of sample points in two images. The sample points are spaced out regularly in the projected images.

## Ambient Lighting




Each cue increases the reconstruction accuracy, leading to significantly better result than possible with the state of the art monocular system [1]. Reconstructions of the face surface are compared to ground truth data acquired with a structured light system. The point wise distance from the reconstruction to the ground truth is shown in the inset head renderings. Here green is a perfect match, and red denotes a distance of 3mm or more. The best of the three monocular results is shown.

## Directed Lighting




The new stereo algorithm is robust under directed lighting and yields significantly more accurate surface reconstructions than the monocular algorithm. Again the distance to the ground truth is shown for green=0mm and red=3mm in the insets. Future work will include a skin and lighting model, hopefully improving speed and accuracy of the method. All cues were used.

## Funding

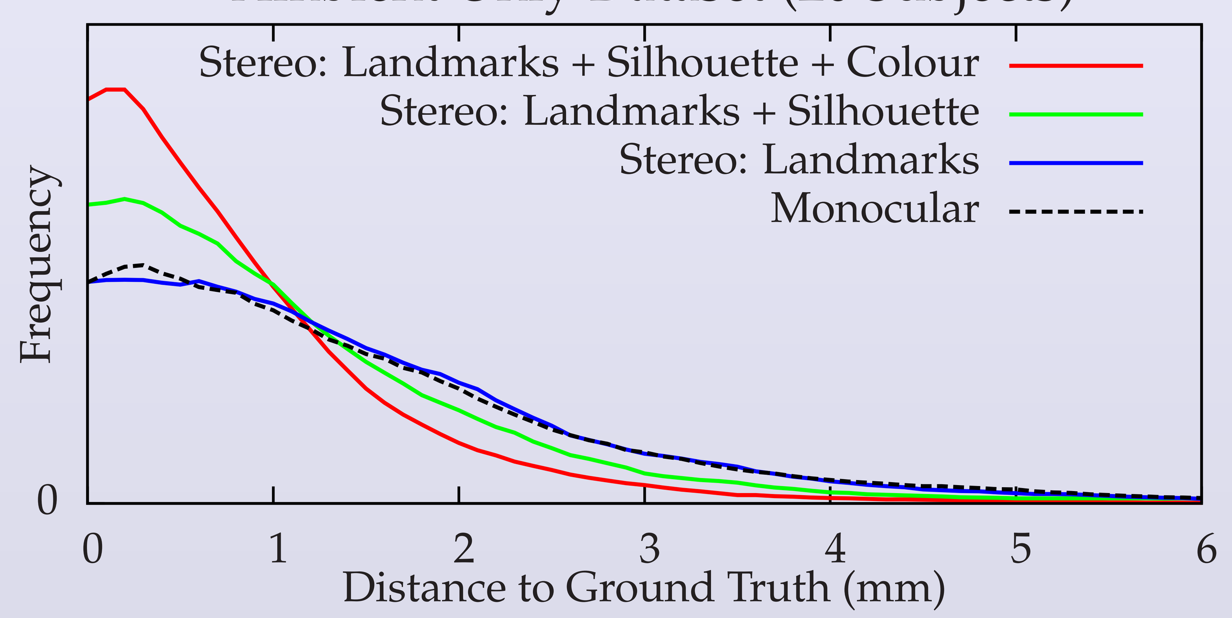
This work was supported in part by Microsoft Research through the European PhD Scholarship Programme.

## References

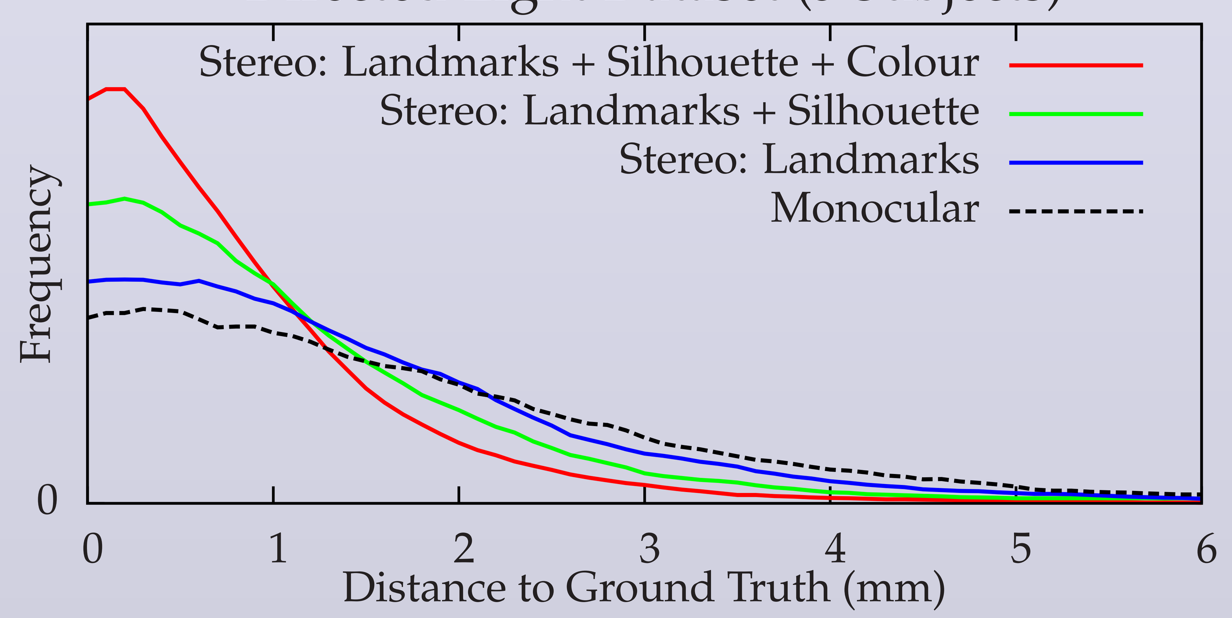
[1] S. Romdhani and T. Vetter. Estimating 3D Shape and Texture Using Pixel Intensity, Edges, Specular Highlights, Texture Constraints and a Prior. In *CVPR 2005*

## Evaluation: Gold Standard

Ambient Only Dataset (20 Subjects)




Directed Light Dataset (5 Subjects)




The use of multi-view information results in a much higher accuracy than achievable by the monocular method. A higher frequency of lower residuals is better.

## Evaluation: Face Recognition

To test the method on a difficult dataset, a face recognition experiment on the PIE dataset was performed. The results show, that the extracted surfaces are consistent over variations in viewpoint and that the reconstruction quality increases with an increasing number of images.

| View-points | Landmark | | + Silhouette | | + Colour | |
|---|---|---|---|---|---|---|
| | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| 2 | 10% | 18% | 50% | 68% | 63% | 82% |
| 3 | 7% | 18% | 62% | 74% | 74% | 85% |
| 4 | 19% | 37% | 76% | 82% | 87% | 94% |

The columns labelled "1st" show the frequency of correct results, "2nd" is the frequency with which the correct result was within the first two subjects returned. The angle between the shape coefficients was used as the distance measure.
Texture information should be used to achieve state of the art recognition results.