TUT Head Pose Estimation Challenge

***Abstract*—Determining where people look at in still images is one challenge in computer vision. As part of the head pose estimation competition which is hosted by Tampere University of Technology in https://inclass.kaggle.com/, we propose a solution to estimate the viewing angles of the human head. In this report, we explain the methods and tools that we use, trying to minimize the mean square error (MAE) to obtain the best estimation.**

Keywords—head post estimation; machine learning; PCA; libsvm; matlab

I. INTRODUCTION

In computer vision systems, head pose usually coincides with the gaze direction. In this competition, Head pose estimation is to estimate the viewing angles (yaw and pitch) of human head in a given facial images accordingly.

Machine learning tools are used to learn a model to map two-dimensional output labels from given imaginary features. After trying different methods we decide to use support vector machine method (SVM) for learning the model and predicting the two angles. An external library LIBSVM is used during the implementation. This is a popular open source machine learning library developed by National Taiwan University.

In order to obtain better performance in this competition, we mainly adopt the methods of Principal component analysis (PCA), cross-validation and parallel pool.

II. IMPLEMENTATIONS

By applying these methods in different ways, four different approaches are compared. The performance of an approach could be evaluated by mean square error, evaluation time and prediction time. These approaches are as follows:

A. Reduce the dimension of the original data by using PCA. Evaluate the performance of the model by cross-validation. Make prediction by using all the training data.

B. Evaluate the performance of the model by cross-validation. Make prediction by using all the training data.

C. Reduce the dimension of the original data by using PCA. In single cross-validation process, use 80% of the training data to predict. Repeat such cross-validation process for enough times. Use the mean/median/mode value as the final prediction value from all the cross-validation results.

D. In single cross-validation process, use 80% of the training data to predict. Repeat such cross-validation process for enough times. Use the mean/median/mode value as the final prediction value from all the cross-validation results.

III. RESULTS

After implementing the above four approaches, the results are shown in the following table:

Approach	A	B	C	D
Evaluation time (seconds)	30.19	275.44	N/A	N/A
Evaluation MAE	5.13	4.99	N/A	N/A
Prediction time (seconds)	10.00	87.86	413.85	2168.49
Score in Kaggle	4.94624	4.56989	Mean:4.81329 Median:4.62366 Mode:4.67742	Mean:4.68328 Median:4.38172 Mode:4.51613

Table 1 - Performance of different approaches

As can be seen in Table 1, the execution time could be decreased significantly by using PCA. However, it also has negative effect on the prediction which means that PCA decreases the result accuracy slightly. Meanwhile, repeating cross-validation process for enough times is better than simply using all the training data to predict. The performance of taking median value is better than taking mean and mode value.

IV. CONCLUSIONS

In this challenge, various kinds of machine learning methods could be used. From our implementations, the best performance is obtained by using the SVM with cross-validation and the median value.

REFERENCES

[1] TUT Head Pose Estimation Challenge, [online], Available: https://inclass.kaggle.com/c/tut-head-pose-estimation-challenge