

# A MAP approach to landmarking

Gert Beumer

University of Twente

Fac. EEMCS, Signals and Systems Group  
Hogekamp Building, 7522 NB Enschede  
The Netherlands

`g.m.beumer@ewi.utwente.nl`

Raymond Veldhuis

University of Twente

Fac. EEMCS, Signals and Systems Group  
Hogekamp Building, 7522 NB Enschede  
The Netherlands

`r.n.j.veldhuis@ewi.utwente.nl`

## Abstract

Landmarking can be seen as calculating the maximum a posteriori probability of a certain set of landmarks given a certain image (texture) which contains a face. The MLLL plus BILBO algorithm by Beumer et al [1] is one method. An improvement to this existing landmarking algorithm, MLLL in combination with BILBO, is presented. MLLL uses a likelihood ratio based similarity score to mark candidate landmark locations. BILBO uses a statistical method on the shape to correct outliers. A theoretical analysis of this intuitive approach shows new insight.

In order to verify this theory we performed a simple experiment. The results show that using the new method performs significantly better than when only using the similarity score method (MLLL) in combination with outlier removal (BILBO).

## 1 Introduction

Face recognition is often done on still images., because photographs of the face of an individual are in many situations the only data that are available for recognition. In order to do a proper recognition, the face first needs to be registered. Registration is the alignment of the face to a fixed position, scale and orientation. This applies to all images used for training, enrollment and recognition. Registration in face recognition is usually based on landmarks. Landmarks are stable points in the face which can be found with sufficient accuracy. Having a reliable and stable method for automatic landmarking and registration is essential for the automation of a face recognition system. It has been shown that the accuracy of the landmarking has a strong relation with the final recognition result [2]. Using a log likelihood ratio based similarity score method like MLLL has shown to be a useful tool to find landmarks on a face [1].

In this paper we will propose a theory for improving already developed methods based on maximizing the likelihood ratio. In order to verify this theory we performed a simple experiment. Results show that using the new method performs significantly better than when only using the likelihood ratio.

A theoretical analysis of the landmarking brings the new insight, that landmarking can be seen as calculating the maximum a posteriori probability of a certain set of landmarks (shape) given a certain image (texture) which contains a face. According to Bayes' rule this is the same as the likelihood ratio multiplied with the probability of that certain shape. In this paper the method by Beumer et al [1] is expanded so that takes the probability of the shape into account.

---

<sup>1</sup>Werkgemeenschap voor Informatie- en Communicatietheorie

## 2 Previous and related work

A likelihood ratio based facial feature detector was proposed by Bazén et al [3], which maximized the likelihood that a certain location or pixel is the landmark. The importance of accurate landmarking has been shown by Beumer et al in [4]. Later the Most Likely Landmark Locator (MLLL) and a subspace outlier correction method called BILBO were introduced by Beumer et al [1].

MLLL calculates a log likelihood ratio based similarity score  $S$ . MLLL is a classifier which is trained to discriminate between a landmark and non-landmark using a PCA and LDA based algorithm. Thresholding  $S_{u,v}$  determines if at location  $(u, v)$  there is a landmark. MLLL however takes the intuitive approach of looking for the highest value for  $S$  in a certain region. This location then is considered the landmark location.  $S$  is calculated for all locations in a region where the landmark is expected to be.

$$S_{u,v} = -(y_{u,v} - \mu_L)^T \Sigma_L^{-1} (y_{u,v} - \mu_L) + (y_{u,v} - \mu_{\bar{L}})^T \Sigma_{\bar{L}}^{-1} (y_{u,v} - \mu_{\bar{L}}). \quad (1)$$

where  $\Sigma$  denotes a covariance matrix and  $y_{u,v} = T(x_{u,v} - x_{0,u,v})$  where  $x$  is the texture surrounding  $(u, v)$  and  $T$  a PCA/LDA feature reduction transformation matrix. For more details see [1]

BILBO uses a lower dimensional space to project the shape *there and back again* in order to remove outliers. The improvement gained by BILBO suggested that using the shape during landmarking has significant advantages over uncorrelated detection of correlated landmarks.

Other work, amongst others, includes Everingham [5] who uses a regression method, Bayesian methods and a discriminative method for landmarking. They not explicitly use the shape or the distribution of the shape. Work by Cristinacce [6] focuses on both multiple templates of the landmark and the shape to constrain the search area.

## 3 Theory

Assume a collection of pixel values,  $\vec{x}$ , around a set of landmarks (shape),  $\vec{s}$ , both with a known probability density. Calculating the maximum a posteriori (MAP) probability gives the most likely shape  $\vec{s}$  given a certain texture  $\vec{x}$ . This is found by maximizing the  $p(\vec{s}|\vec{x})$ . Bayes rule states that

$$\operatorname{argmax}_{\vec{s}} p(\vec{s}|\vec{x}) = \operatorname{argmax}_{\vec{s}} \frac{p(\vec{x}|\vec{s})}{p(\vec{x})} p(\vec{s}). \quad (2)$$

In the first term of (2) we recognise the likelihood ratio

$$L = \frac{p(\vec{x}|\vec{s})}{p(\vec{x})} \quad (3)$$

that was used in [1]. In [1] the implicit assumption was made that  $p(\vec{s})$  is uniform. Maximizing (3) is how MLLL [1] finds landmarks by maximizing the likelihood ratio for each landmark location  $s_i$  with surrounding texture  $x_i$ . Equation (2) also shows that the  $\vec{s}$  for which (2) is maximal depends on the distribution of texture given a certain shape,  $p(\vec{s}|\vec{x})$ , the overall distribution of the texture,  $p(\vec{x})$  and the distribution of the shapes  $p(\vec{s})$ .

Assume an image contains  $l$  landmarks, i.e. a face with eyes, nose, mouth etc. Also assume that the neighbourhoods of the landmarks (the texture) are not overlapping,

and independent. Equation (2) then turns into

$$\operatorname{argmax}_{\vec{s}} \prod_{i=1}^l \frac{p(\vec{x}_i | \vec{s}_i)}{p(\vec{x}_i)} p(\vec{s}_1 \dots \vec{s}_l) \quad (4)$$

where  $\vec{x}_i$  is a vector containing all the pixel values and  $\vec{s}_i$  is a 2 element vector containing the coordinate.

For simplicity reasons, we assume that in the initial experiment an image contains only one landmark, or equivalently that the landmarks are fully independent i.e.  $p(\vec{s}_1 \dots \vec{s}_l) = p(\vec{s}_1) \dots p(\vec{s}_l)$ . The final choice for the location of the  $i$ -th landmark is where

$$\operatorname{argmax}_{\vec{s}} \prod_{i=1}^l \frac{p(\vec{x}_i | \vec{s}_i)}{p(\vec{x}_i)} p(\vec{s}_i) \quad (5)$$

is maximum. This is the element wise product of the likelihood landscape and the probability landscape of the landmark distribution or the sum of the log likelihood based similarity score and the log of the distribution.

Implementing this makes that (1) must be modified to

$$S_{i,u,v} = -(y_{u,v} - \mu_L)^T \Sigma_L^{-1} (y_{u,v} - \mu_L) + (y_{u,v} - \mu_{\bar{L}})^T \Sigma_{\bar{L}}^{-1} (y_{u,v} - \mu_{\bar{L}}) + \log(p(\vec{s}_i)). \quad (6)$$

## 4 Experiment

This experiment is the first step towards an improved version of the combination of both MLLL and BILBO. As said we assume that future experiments will be less constrained by the assumption that the landmarks are independent. For the experiment we used a MLLL algorithm which was trained on the BioID [7] database. The train data are used for training MLLL, as well as the shape data for BILBO and to compute the estimation of the local distribution of  $P(s)$ . The BioID database consists of 1521 images of 23 persons. The database is provided with ground truth data for 20 landmarks of which 17 were used. These can be seen as green crosses in figure 1(a).

The algorithm was then tested on the 5647 images of the FRGC-version 1 [8] database. This database is provided with ground truth data for 4 landmarks, being both eyes, the nose and the centre of the mouth. These can be seen as the red dots in figure 1(a).

MLLL will calculate all 17 landmarks but the results can only be compared to the four for which there is ground truth data. When comparing to the centre of the mouth, the four found landmark location from MLLL, both corners and the centres of both upper lip and lower lip. BILBO uses all 17 landmarks found by both MLLL and the MAP approach.

The results of the experiment will be given as the RMS value of the distance of a found landmark to the ground truth data.

### 4.1 Estimating $p(s)$

In figure 1(b) it can clearly be seen that the spatial distribution of the landmarks is Gaussian nor uniform. In order to estimate the distributions the following algorithm was used.

1. Calculate average shape.

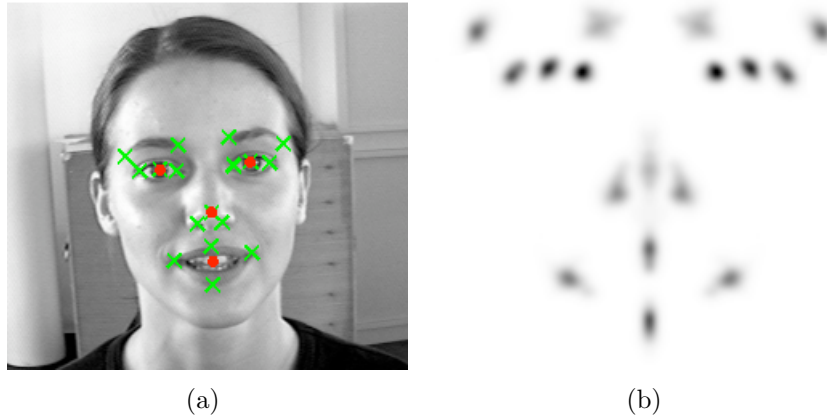


Figure 1: a: The landmarks used from the BioID database (green crosses) and from the FRGC (red dots). b: The shape distributions map showing the 17 landmarks. Black for high probability and white for low. All probabilities estimated on the BioID database ground truth data.

2. Scale the average so that the eye are 100 pixels apart.
3. Align all shapes to the average.
4. Go to 1 until stable.
5. Create a 2D histogram of all the landmarks.
6. Flip the histogram over the symmetry axis of the face and add them.
7. Take the log of the resulting histogram.

The resulting distribution is shown in Figure 1(b). The errors are presented as RMS value of the distance between the found landmark and the ground truth data.

## 5 Results

The results can be seen in Table 1 and Figure 2. The new method clearly outperforms MLLL on all landmarks. It is interesting to show that the impact of BILBO is no longer visible. This is caused by the fact that the new methods does not generate outliers which are corrected by BILBO. Because of the setup of the experiment there it is not yet possible to compare the results to those of others.

## 6 Conclusion and discussion

Using the MAP estimator of the landmarks gives more accurate estimates of the landmark locations. This however is only a first step to show that using the MAP probability actually improves the accuracy of the MLLL and MLLL with BILBO algorithm.

Landmark	MLLL [px]	MLLL+BILBO [px]	MAP [px]	MAP+BILBO [px]
Right eye	6.3	5.6	4.0	4.0
Left eye	6.2	5.1	3.9	3.9
Nose	17.1	8.0	5.5	5.5
Mouth	7.7	5.6	3.0	2.9
Total	10.4	6.2	4.2	4.2

Table 1: RMS error results on the FRGC database. The error is relative to 100 pixels between de centres of the eyes.

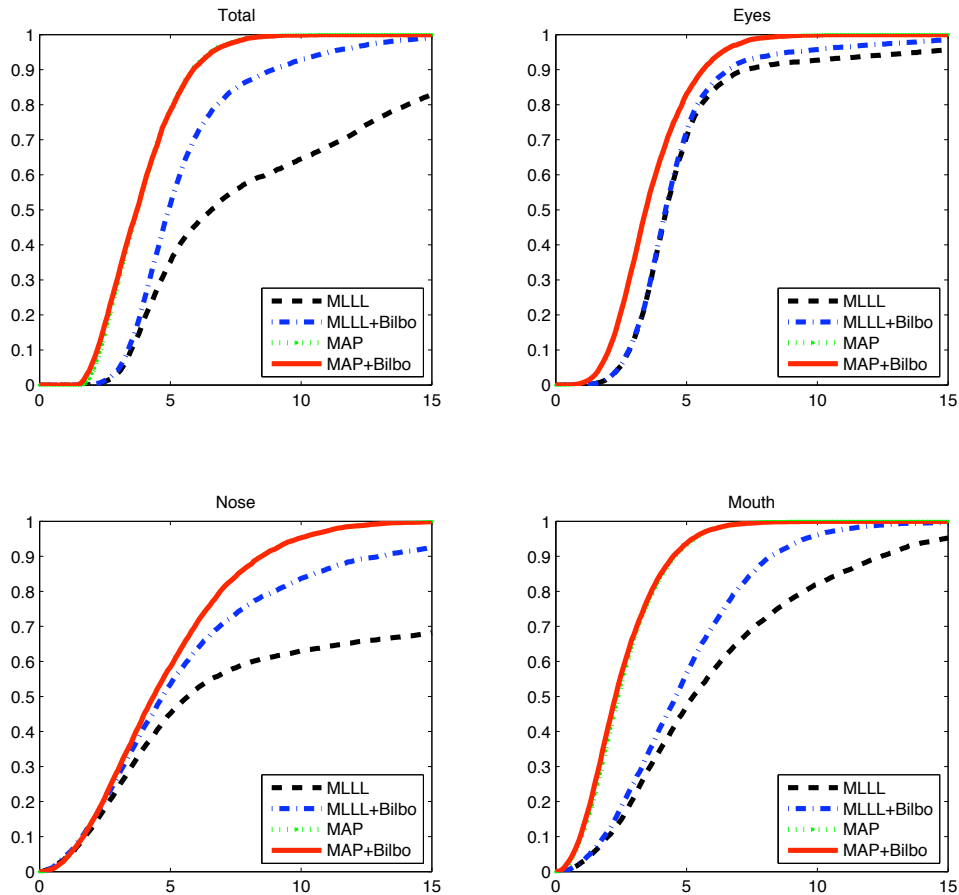


Figure 2: Cumulative RMS error results on the FRGC [8] database. The error is relative to 100 pixels between de centres of the eyes.

The experiment only assumed one landmark in the shape and not yet the full shape. Currently the assumption is that the landmark locations are completely independent. This is however not true. The next step will be to introduce the dependencies between the landmarks in order to improve the estimates of  $p(\vec{s})$ . Still the results are promising. Further research is needed to both improve the results even more and make it a workable algorithm for landmarking. Especially on how to estimate  $p(\vec{s})$  for shapes that consist of then one landmark and how to choose the texture,  $x$ , from the image.

## Acknowledgment

This work was done within the IOP-GenCom Project IGC03003: BASIS sponsored by Senter Novem.

## References

- [1] G. M. Beumer, Q. Tao, A. M. Bazen, and R. N. J. Veldhuis, "A landmark paper in face recognition," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, Southampton, UK*, Los Alamitos, April 2006, IEEE Computer Society Press.
- [2] G.M. Beumer, A.M.Bazen, and R.N.J. Veldhuis, "On the accuracy of eers in face recognition and the importance of reliable registration.," in *SPS 2005*. IEEE Benelux/DSP Valley, April 2005.
- [3] A. M. Bazen, R. N. J. Veldhuis, and G. H. Croonen, "Likelihood ratio-based detection of facial features," in *Proc. ProRISC 2003, 14th Annual Workshop on Circuits, Systems and Signal Processing*, Veldhoven, The Netherlands, nov 2003, pp. 323–329.
- [4] G. M. Beumer, A. M. Bazen, and R. N. J. Veldhuis, "On the accuracy of eers in face recognition and the importance of reliable registration," in *5th IEEE Benelux Signal Processing Symposium (SPS-2005), Antwerp, Belgium*, secretariaat in Delft, April 2005, pp. 85–88, IEEE Benelux Signal Processing Chapter.
- [5] Mark Everingham and Andrew Zisserman, "Regression and classification approaches to eye localization in face images," in *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, Washington, DC, USA, 2006, pp. 441–448, IEEE Computer Society.
- [6] D. Cristinacce and T. F. Cootes, "Facial feature detection and tracking with automatic template selection," in *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, Washington, DC, USA, 2006, pp. 429–434, IEEE Computer Society.
- [7] HumanScan, "Bioid face db," .
- [8] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005.