Appearance-based gaze tracking with spectral clustering and semi-supervised Gaussian process regression

Ke Liang, Youssef Chahir, Michele Molina, Charles Tijus, François Jouen

▶ **To cite this version:**

Ke Liang, Youssef Chahir, Michele Molina, Charles Tijus, François Jouen. Appearance-based gaze tracking with spectral clustering and semi-supervised Gaussian process regression. Conference on Eye Tracking South Africa (ETSA 2013), Aug 2013, Le Cap, South Africa. pp.29-31, 10.1145/2509315.2509318 . hal-01882823

HAL Id: hal-01882823
https://hal.science/hal-01882823

Submitted on 27 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Appearance-Based Gaze Tracking with Spectral Clustering and Semi-Supervised Gaussian Process Regression

Ke Liang
CHArt Laboratory
EPHE Paris
4-14 rue Ferrus 75014 Paris, France
ke.liang@etu.ephe.fr

Youssef Chahir
Computer Science Department,
GREYC-UMR CNRS 6072
University of Caen
Caen, France
youssef.chahir@unicaen.fr

Michèle Molina
PALM Laboratory EA 4649
University of Caen
Caen, France
michele.molina@unicaen.fr

Charles Tijus
CHArt Laboratory
EPHE Paris
4-14 rue Ferrus 75014 Paris, France
tijus@univ-paris8.fr

François Jouen
CHArt Laboratory
EPHE Paris
4-14 rue Ferrus 75014 Paris, France
Francois.Jouen@ephe.sorbonne.fr

ABSTRACT
Two of the challenges in appearance-based gaze tracking are: 1) prediction accuracy, 2) the efficiency of calibration process, which can be considered as the collection and analysis phase of labelled and unlabelled eye data. In this paper, we introduce an appearance-based gaze tracking model with a rapid calibration. First we propose to concatenate local eye appearance Center-Symmetric Local Binary Pattern(CS-LBP) descriptor for each subregion of eye image to form an eye appearance feature vector. The spectral clustering is then introduced to get the supervision information of eye manifolds on-line. Finally, taking advantage of eye manifold structure, a sparse semi-supervised Gaussian Process Regression(GPR) method is applied to estimate the subject's gaze coordinates. Experimental results demonstrate that our system with an efficient and accurate 5-points calibration not only can reduce the run-time cost but also can lead to a better accuracy result of 0.9°.

Keywords
Appearance-based gaze estimation, spectral clustering, Gaussian process regression

1. INTRODUCTION
As gaze tracking technology improves in the last 30 years, gaze tracker offers a powerful tool for diverse study fields, in particular eye movement analysis and human-computer interaction (HCI). Nowadays most commercial gaze trackers use feature-based method to estimate gaze coordinates, which relies on video-based pupil detection and the reflection of infra-red LEDs. In general, there are two principal methods: 1) Pupil-Corneal Reflection(P-CR) method [10], [24], [2]) 3D model based method [15], [19]. IR light and extraction of pupil and iris are important for these feature-based methods, and the calibration of cameras and geometry data of system is also required.

Appearance-based methods do not explicitly extract features like the feature-based method, but rather use the cropped eye images as input with the intention of mapping these directly to gaze coordinates [5]. The advantage is that they do not require calibration of cameras and geometry data like feature-based method. Moreover, they can be less expensive in materials than feature-based method since they don't have to work on the same quality images like feature-based method does. But they still need a relatively high number of calibration points to get accurate precision. Different works can be seen in multilayer networks [1], [16], [22], or Gaussian process [11], [21], or manifold learning [9], [17]. Williams et al. [21] introduces the sparse, semi-supervised Gaussian Process (S^3GP) to learn mappings from semi-supervised training sets. Fukuda et al. [4] propose a gaze-estimation method that uses both image processing and geometrical processing to reduce various kinds of noise in low-resolution eye-images and thereby achieve relatively high accuracy of gaze estimation.

Manifold learning is widely applied to solve many problems in computer vision, in pattern recognition etc [7], [13], [20], [23]. Manifold learning, often also referred to as non-linear dimensionality reduction, is also one of the approaches applied in appearance-based gaze tracking [17], and one of the reason to apply manifold learning techniques is to reduce computational costs. Manifold learning means the process of estimating a low-dimensional structure which underlies a collection of high-dimensional data, also preserves characteristic properties of the set of high-dimensional data. Here we are interested in the case where the manifold lies in a high dimensional space $\mathbb{R}^D$, but will be homeomorphic with a low dimensional space $\mathbb{R}^d$ ($d < D$). Laplacian Eigen maps [2], [3] most faithfully preserves proximity relations of a high-dimensional non-linear data set in the low dimensional space, by using spectral graph technique.

Another important technique in an appearance-based gaze tracking system is the predictive uncertainty. In other words, how to map a new image of eye or an eye manifold to 2D screen coordinates. Supervised learning requires an arduous calibration process to form the training set with the number of samples. In contrast, a large number of unlabelled samples can be easily collected. The semi-supervised learning has attracted an increasing amount of interest recently. It is a promising family of techniques that exploit the "manifold structure" of the data; such methods are generally based upon an assumption that similar unlabelled data should be given the same classification. In addition, sparse techniques such as the SVM and RVM have been proven efficient in the gaze tracking application. The related works can be seen in [8], [9], [12], [17], [21]. Lu et al.

[8] use 15D feature vector to represent eye image content, and propose an Adaptive Linear Regression via l_1-optimization to estimate gaze coordinates. Martinez et al.[9] employ multilevel Histograms of Oriented Gradients(HOG) features as appearance descriptor for eyes, and learn the mapping by Support Vector Regression(SVR). Noris et al.[12] present a wearable gaze estimation system based on Support Vector Machines(SVM), Gaussian Process Regression(GPR), and image appearance which is reduced by PCA (Principal Components Analysis). [21] uses a sparse regression model to infer eye-gaze mapping in real-time, which shows the performance of semi-supervised technique and the efficiency of sparsity in the experimental results.

Our contributions are:

- A subregion CS-LBP concatenated histogram is used as eye appearance feature which not only reduce the dimension of raw images, but also can be robust against the changes in illumination.
- Laplacian Eigen maps is introduced to project eye feature samples to a subspace in order to get the clusters of these samples.
- A sparse semi-supervised Gaussian process regression method infers the gaze coordinates by an active set which can be built on-line with limited numbers of samples and their manifold structures.

The rest of the paper is organized as follows. Section 2 describes the eye manifold learning to the proposed eye feature. Section 3 presents the proposed regression method to infer the gaze coordinates. Section 4 shows the experimental setup and results. Finally section 5 concludes the paper.

2. EYE APPEARANCE MANIFOLD LEARNING

2.1 Eye appearance descriptor

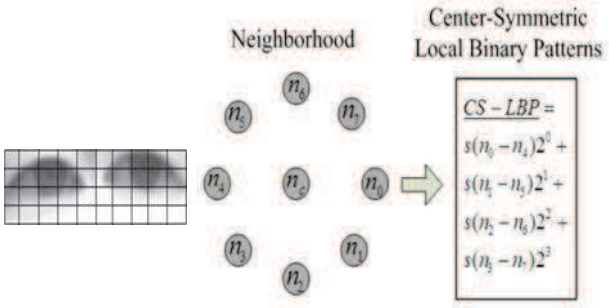
Let an eye image I be a two-dimensional M by N array of intensity values, and it may also be considered as a vector of dimension $M \times N$. The proposed gaze tracker captures left and right eyes together and combines them into one image. Our eye image of size 160 by 40 becomes a vector of dimension 6400. Appearance-based gaze tracking methods mostly rely on the eye images as input. Extracting eye appearance descriptor not only helps to reduce the dimension of eye images, but also preserves the feature and variation of eye movements.

There exist a number of eye appearance feature extraction methods for gaze tracking system, like multi-level HOG [9], eigeneyes by PCA [12], and subregions feature vector [8]. Lu et al. have proven the efficiency of using 15D subregions feature vector in [8]. To compute this feature vector, the eye image I_i is divided into N' subregions of size $w \times h$. Let S_j denote the sum of pixel intensities in j–th subregion, then feature vector X_i of the image I_i is represented by

$$X_i = \frac{[S_1, S_2, ..., S_j]}{\sum S_j} \qquad j \in N' \qquad (1)$$

Here we introduce our subregion methods with Center-Symmetric Local Binary Pattern (CS-LBP) to calculate low dimensional feature vector for raw eye image content. Local Binary Pattern (LBP) operator has been highly successful for various computer vision problems such as face recognition, texture classification etc. The histogram of the binary patterns computed over a region is used for feature vector. The operator describes each pixel by the relative graylevels of its neighbouring pixels. If the graylevel of the neighbouring pixel is higher or equal, the value is set to one, otherwise to zero.

We calculate the CS-LBP [6] histogram, which is a new texture feature based on the LBP operator, for each subregion in Fig.1(a) and concatenate them to form the eye appearance feature vector. Instead of describing a centre pixel by comparing its neighbouring pixels with it in LBP, CS-LBP compares the centre-symmetric pairs of pixels in Fig.1(b).



(a) (b)

Figure 1. a) 40 subregions of an eye image sample b) CS-LBP for a neighbourhood of eight pixels.

The CS-LBP value of a centre pixel in position (x,y) is calculated as follows:

$$CS - LBP_{R,N,T}(x,y) = \sum_{i=0}^{N/2-1} s(n_i - n_{i+(N/2)})$$

where $s(t) = \begin{cases} 1 & t > T \\ 0 & else \end{cases}$, n_i and $n_{i+(N/2)}$ are the gray values of centre-symmetric pairs of pixels of N equally spaced pixels on a circle with radius R , and the threshold T is a small value. From this equation, the value of CS-LBP may be any integer from 0 to $2^{N/2} - 1$, and the histogram dimension will be $2^{N/2}$. CS-LBP is fast to compute and its histogram has been proven to be robust against the changes in illumination as a texture descriptor [6].

2.2 Spectral clustering

Graph Laplacians are the main tools in spectral graph theory. Here we focus on two kinds of graph Laplacian:

- Unnormalized graph Laplacian.

  $$L_{un} = D - W$$

  where W is the symmetric weight matrix with positive entries for edge weights between vertices. If $w_{ij} = 0$, then vertices i and j are not connected.

D is the degree matrix: $d_{ii} = \sum_{j=1}^{n} w_{ij}$ and $d_{ij} = 0 \; \forall i \neq j$.

- Normalized graph Laplacian.

  $$L_{sym} = D^{-1/2} L_{un} D^{1/2} = I - D^{-1/2} W D^{1/2}$$
  $$L_{normalized} = D^{-1} L_{un} = I - D^{-1} W = I - L_{rw}$$

  where L_{sym} is a symmetric matrix, and L_{rw} is closely related to a random walk.

There are 3 properties:

1) λ is an eigenvalue of L_{rw} with eigenvector v if and only if λ and v solve the generalized eigenproblem $Lv = \lambda Dv$.

2) L_{rw} is positive semi-definite with the first engenvalue $\lambda_1 = 1$ and the constant one vector **1** the corresponding eigenvector.

3) All eigenvectors are real and it holds that: $1 = |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$.

Laplacian Eigen maps use spectral graph technique to compute the low-dimensional representation of a high-dimensional non-linear data set, and they most faithfully preserves proximity relations, mapping nearby input patterns to nearby outputs. The algorithm of Laplacian Eigen maps has a similar structure as LLE. First, one constructs the symmetric undirected graph $G = (V,E)$, whose vertices represent input patterns and whose edges indicate neighbourhood relations (in either direction). Second, one assigns positive weights W_{ij} to the edges of this graph; typically, the values of the weights are either chosen to be constant, say $W_{ij} = 1/k$, or a heat kernel, as $W_{ij} = exp(-\frac{||x_i - x_j||^2}{2l^2})$ where l is a scale parameter. In the third step of the algorithm, one obtains the embeddings $\psi_i \in \mathbb{R}^m$ by minimizing the cost function:

$$\mathcal{E}_L = \sum_{ij} \frac{W_{ij}||\psi_i - \psi_j||^2}{\sqrt{D_{ii}D_{jj}}}$$

This cost function encourages nearby input patterns to be mapped to nearby outputs, with "nearness" measured by the weight matrix W. To compute the embeddings, we find the eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and eigenvectors $v_1,\ldots,v_n$ of the generalized eigenproblem: $Lv = \lambda Dv$. The embeddings $\Psi : \psi_i \rightarrow (v_1(i),\ldots,v_m(i))$.

Spectral clustering refers to a class of techniques which rely on the eigen-structure of a similarity matrix to partition points into disjoint clusters with points in the same cluster having high similarity and points in different clusters having low similarity. We follow the works of Shi and Malik (2000). Their algorithm of spectral clustering computes the normalized graph Laplacian L_{rw}, and its first k generalized eigenvectors $v_1,\ldots,v_k$ as embeddings, and then utilise k-means to cluster the points.

From the section 2.1 we've introduced our subregion CS-LBP methods to extract the eye appearance feature descriptor. Here we'd like to at first obtain eye manifold distribution by using Laplacian Eigen maps, and then we apply the normalized spectral clustering. The Fig.2 shows eye samples of the subject's eye movements when the subject follows the visual pattern(green points) shown in the screen. The Fig.3 demonstrate the distribution of embeddings in the subspace. (a) gives the distribution of a dataset of 240 points which contains only the eye samples on the 8 points in the screen , while (b) contains only 120 eye samples from the 4 points in the corner(up right, up left, down left, down right). For a given number C of visual patterns, generally we can get l clusters $U = \{U_1,U_2,\ldots,U_l\}$ associated with weights $W = \{w_1,w2,\ldots,w_l\}$ by the size of cluster, where $C \leq l < n$.

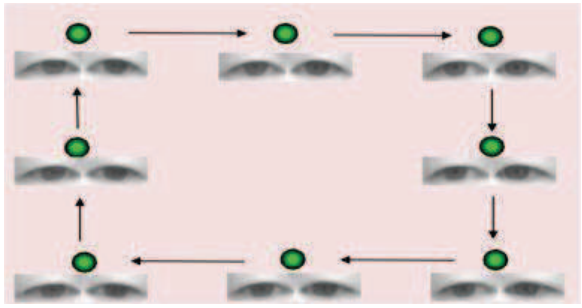
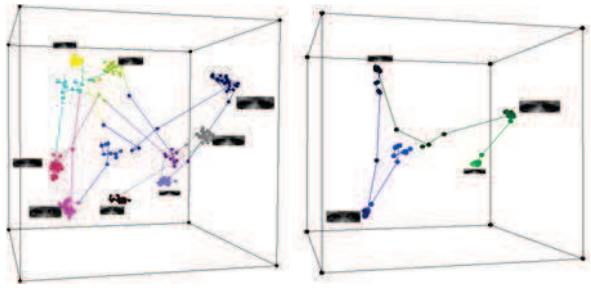


Figure 2. Eye calibration of 8 points



(a) (b)

Figure 3. (a) Eye manifolds in the phase of 8-points calibration
(b) Eye manifolds in the phase of 4-points calibration (up-right, up-left, down-left, down-right)

3. SPARSE SEMI-SUPERVISED GAUSSIAN PROCESS REGRESSION METHOD

In order to map feature vector $X \in \mathbb{R}^{M'}$ to gaze coordinate output $Y \in \mathbb{R} \times \mathbb{R}$, which has two values (x-coordinate, y-coordinate), we have a training set $\mathcal{D} = \{(\mathcal{X}_i, \mathcal{Y}_i)|i = 1, 2,\ldots,n\}$, where $\mathcal{X}_i$ denotes an input vector of dimension M' and $\mathcal{Y}_i$ denotes a scalar output, n is the number of observations. Given this training set $\mathcal{D}$, we wish to make predictions for the new inputs $\mathcal{X}'$ which is not in the training set. So we need to move from the finite training data $\mathcal{D}$ to a function f that makes predictions for all possible input values, where $\mathcal{Y}' \equiv f(\mathcal{X}') = P(\mathcal{Y}'|\mathcal{X}', \mathcal{D})$, and P is the *posterior* distribution for the training set $\mathcal{D}$). Gaussian Process (GP) is used to calculate this predictive distribution because it can be used as *prior* for Bayesian inference and it is known to be accurate both in terms of mean predictions and predictive uncertainty[14]. According to GP prior, joint distribution of $\mathcal{Y}$ and $\mathcal{Y}'$ is:

The joint distribution is a Gaussian process with zero mean and

$$\begin{bmatrix} \mathcal{Y} \\ \mathcal{Y}' \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\mathcal{X},\mathcal{X}) & K(\mathcal{X},\mathcal{X}') \\ K(\mathcal{X},\mathcal{X}') & \end{bmatrix}\right), \quad K(\mathcal{X},\mathcal{X}') = \sigma^2 exp(-\frac{(\mathcal{X} - \mathcal{X}')^2}{2l^2})$$

covariance function K, where

Given n data points from training set $\mathcal{D}$ and n' test data points, $K(\mathcal{X},\mathcal{X}')$ denotes the $n \times n'$ matrix of the covariances evaluated at all pairs of training and test points, and similarly for the other entries $K(\mathcal{X},\mathcal{X})$, $K(\mathcal{X}',\mathcal{X}')$ and $K(\mathcal{X}',\mathcal{X})$.

Taking a test data point x', the covariance function K is calculated among all possible combinations of these points:

$$K(\mathcal{X},\mathcal{X}) = \begin{bmatrix} K(x_1,x_1) & K(x_1,x_2) & \dots & K(x_1,x_n) \\ K(x_2,x_1) & K(x_2,x_2) & \dots & K(x_2,x_n) \\ \dots & \dots & \dots & \dots \\ K(x_n,x_1) & K(x_n,x_2) & \dots & K(x_n,x_n) \end{bmatrix}$$

$$K(\mathcal{X}',\mathcal{X}) = \begin{bmatrix} K(x',x_1) & K(x',x_2) & \dots & K(x',x_n) \end{bmatrix}$$

If $K(x_i,x_j)$ has a high positive value, our prior belief is that y_i and y_j are highly correlated.

In the training set $\mathcal{D}$, all the data $\mathcal{X}$ are labelled with $\mathcal{Y}$. In order to minimize the degree of uncertainty of distribution of function f, we need to have sufficient labelled data. But in reality, labelled data are often, however, very time consuming and expensive to obtain, as they require the efforts of human annotators, who must often be quite skilled. For example, the calibration phase of appearance-base gaze tracking needs to select the best eye data corresponding the coordinates of calibration point. Generally, a training set can have only some labelled data $\mathcal{X}$, and a number of unlabelled data $\mathcal{X}^*=\{(\mathcal{X}_i^*) | i = 1, 2,...,m\}$. So the problem of effectively combining unlabelled data with labelled data is therefore of central importance in machine learning.

Semi-supervised learning can be applied to have some supervision information on the distribution of the unlabelled data. In order to obtain a low run-time cost of the predictive uncertainty process, an active set $\mathcal{A} = \{\mathcal{X}_1,...,\mathcal{X}_k\}$ $(k < n)$ is proposed as a sparse solution, and is done by 4 steps:

- Sort the clusters U of each visual pattern in section 2.2 by their associated weights W and take the cluster with the most important size as $U_{labelled}$.
- Get the near centre exemplars set $X_{labelled} = \{\mathcal{X}_1,..., \mathcal{X}_i\}$ as labelled data (Fig. 4a) where $\|\mathcal{X}_i - \mu_i\| < T$, $\mathcal{X}_i \in U_{labelled}$, $i \in n$. μ_i is the centre of the cluster $U_{labelled}$ and T is a threshold.
- Form $X_{unlabelled} = \{\mathcal{X}_1,...\mathcal{X}_j\}$ (Fig. 4b) where $\mathcal{X}_j \notin U_{labelled}$, and $\|\mathcal{X}_j - \mathcal{X}_i\| > T_2$.
- For each unlabelled exemplars $\mathcal{X}_j$, make prediction $\mathcal{Y}_j = P(\mathcal{Y}_j|\mathcal{X}_j,X_{labelled})$, and $\mathcal{A} = X_{labelled} \cup X_{unlabelled}$ (Fig. 4c).

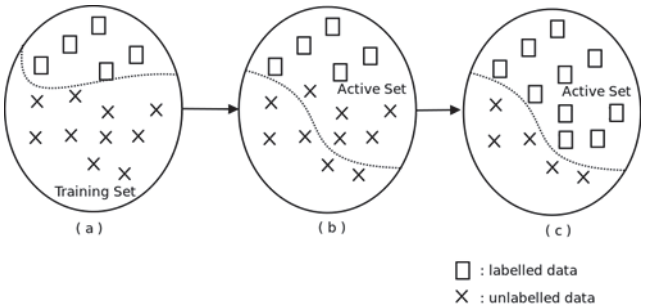


Figure 4. Generation of active set

4. EXPERIMENTATION

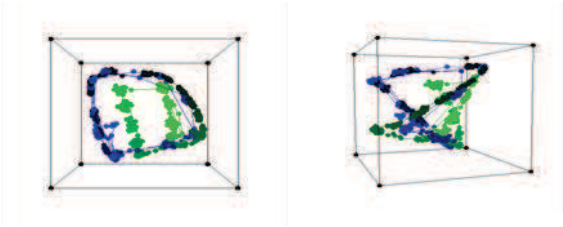
This section evaluates our proposed methods presented in the previous sections. Our experimentation is tested on MacBook Pro 8,1 with Intel Core i5-2415M CPU. Microsoft LifeCam HD-5000 is used for image acquisition of gaze tracking system in the experiments. The USB colour webcam captures 30 frames per second with a resolution of 640 × 480.

4.1 Eye detection and tracking

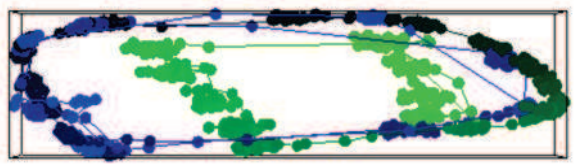
The distance between the subject and the camera is about 40 - 70 cm. To the entire RGB image captured from camera, we firstly use a face components detection model, which is based on Active Shape Model [18], to localize the eye regions and the corners. We introduce then Lukas-Kanade method to track the corner points in the following frames. Finally we combine the left and right eye regions to the eye appearance pattern, which is converted to grayscale and used as the input data for gaze estimation process. The process of eye detection and tracking is shown in Fig. 5. The eye appearance pattern is an image of 160×40. Fig. 6 shows eye samples of five subjects. As introduced in section 2.1, the pattern is divided into 40 subregions and we calculate CS-LBP histogram for each subregion. The size of the feature vector is 640.

Figure 5. Eye localization and tracking

Figure 6. Five eye samples in different light condition. The subjects have head-free movement and the distance between the subject and the camera is about 40 - 70 cm.



(a) (b)



(c)

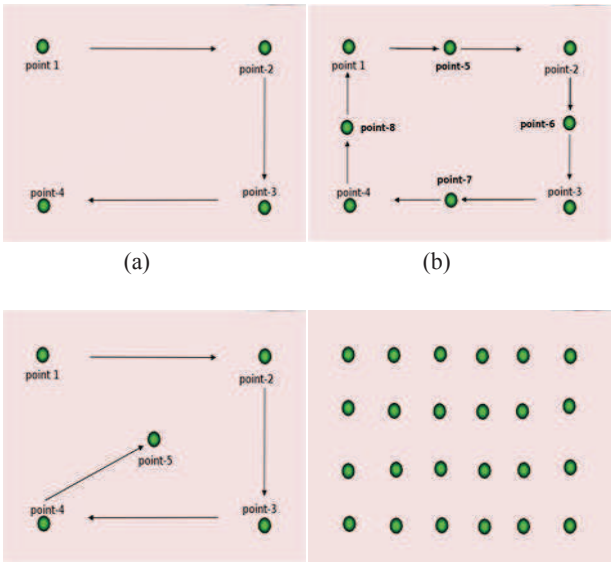
Figure 7. Projection of 990 eye gaze samples on 24 points in the screen (Fig. 8 d) by Laplacian Eigenmaps.

a & b) 3D eye manifolds $e_i = \{v_i^1, v_i^2, v_i^3\}$

c) 3D eye manifolds $e_i = \{\lambda_i^1 v_i^1, \lambda_i^2 v_i^2, \lambda_i^3 v_i^3\}$.

4.2 Calibration phase

The calibration phase will vary depending on application. Sometimes it takes more calibration points to get a more precise result, but sometimes it needs to be quick and efficient. Fig. 8 shows 3 different calibration plan: calibration of 4-points(up-right, up-left, down-right, down-left) such as Figure 8(a), 8-points calibration in Fig. 8(b), and the proposed 5-points calibration in Figure 8(c). Each point appears in the screen for one second, and the subject is expected to follow it. Each point is associated with a label which corresponds to the point's 2D screen coordinates. The 30Hz camera can capture 240 images during the 8-points calibration phase, while 150 images for 5-points calibration.



(a) (b)

(c) (d)

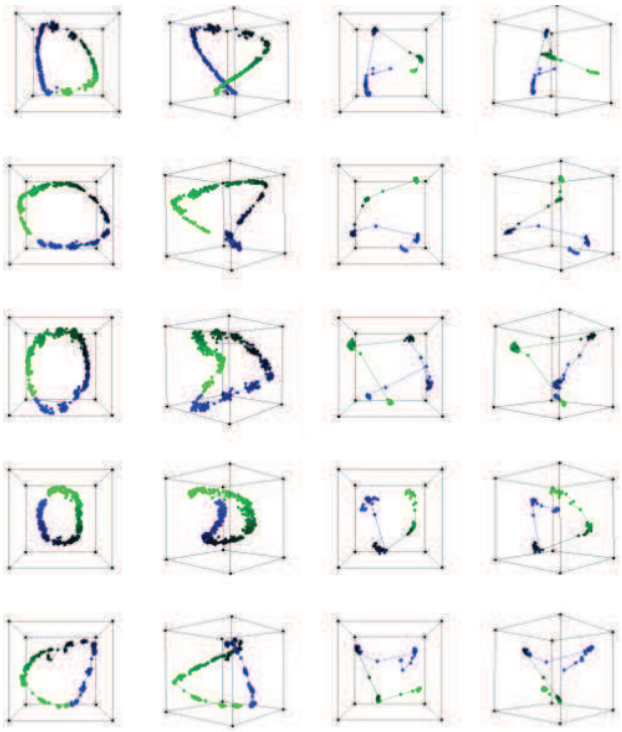
Figure 8. a) 4-points calibration b) 8-points calibration c) 5-points calibration d) 24-points visual pattern test scenario

4.3 Eye Manifolds

We propose a 24-points visual pattern scenario Fig. 8(d) as a test for our gaze estimation. The size is 754 × 519 pixels. With the screen in which pixel is 0.225mm/pixel, the real size is about 17cm × 11.6cm. Each point appears one by one in the screen for about 1 second. Figure 7 shows the distribution of 3D eye manifolds projected by Laplacian Eigen maps. Notice that there are some degrees of similarity between the manifold surface and position plane of the 24 points. Taking eye samples of five subjects such as Fig. 6 who follow the 16 points outside-round the screen, we can see that the distributions of their eye manifolds are relatively similar (Figure 9), despite the difference between their eyes' form. If taking eye samples

Here we analyse the structure of eye manifolds projected by Laplacian Eigen maps in 2 different conditions as the illumination changes and the head movement. We also compare our subregion CS-LBP descriptor with the original subregion method, which calculates the feature vector as the equation(1) shown in Sec. 2.1.

- Illumination changes
  Here the subject follows the 4 points in the screen (Figure 8a) 2 times within 20 seconds, while the indoor-illumination changes as Figure 11. We extract the eye appearance descriptor from the 500 eye images by our proposed CS-LBP descriptor, also by the original subregion descriptor as a comparative method. From the observations (Figure 12) of projection by Laplacian Eigen maps, CS-LBP descriptor gives the translation of eye movement structure for the changes of illumination, while sub-region descriptor is totally disturbed by the changes illumination.



(a) (b) (c) (d)

Figure 9. Each line of figures represents the eye manifold distribution for each subject mentioned above. $\alpha = 0.8$ $l = 100$
a & b) 750 eye samples
c & d) 120 eye samples

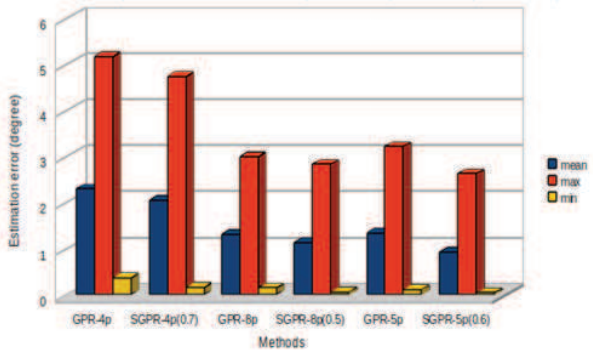
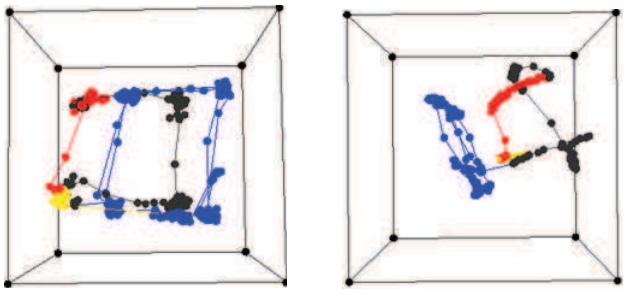


Figure 10. Comparison between our proposed sparse GPR method and conventional GPR method with different calibration plans.

Figure 11. Demonstration of the changes of illumination by 2 sample frames

- Head movements
  Different humans vary widely in the tendency to move the head for a given amplitude of gaze shift. We are interested in the difference of eye manifold structure between a limited natural head movement as Figure 13

and the movement keeping the head still. Here the subject is asked to follow a point which moves along the edge line of screen. The size of screen is 33cm × 22cm. The distance between the subject and the screen is about 60 cm. From the result as shown in Figure 14, we can see that both the descriptors can keep the structure of eye movement while moving the head slightly, but the scale of structure changes.



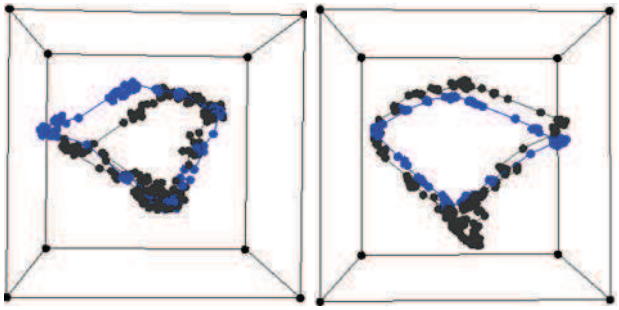
(a) (b)

Figure 12. Comparison of using CS-LBP and subregion methods as eye descriptor in the condition of changes of illumination (500 eye samples). The different colours show the changes of illumination.

a) $CS-LBP_{1,8,0.01}$ feature vector projected in 3D by Laplacian Eigen maps (l = 10000).

b) original subregion feature vector by Laplacian Eigen maps (l = 700).

Figure 13. Demonstration of the free-head movement while the subject follows the points in the screen. The distance between the subject and the screen is about 60 cm.



(a) (b)

Figure 14. Comparison of using CS-LBP and subregion methods as eye descriptor for the head movement (520 eye samples). Blue points represent the structure with fixed head and black points represent the structure with slight head movements.

a) $CS-LBP_{1,8,0.01}$ feature vector projected in 3D by Laplacian Eigen maps (l = 9000).

b) original subregion feature vector by Laplacian Eigen maps (l = 900).

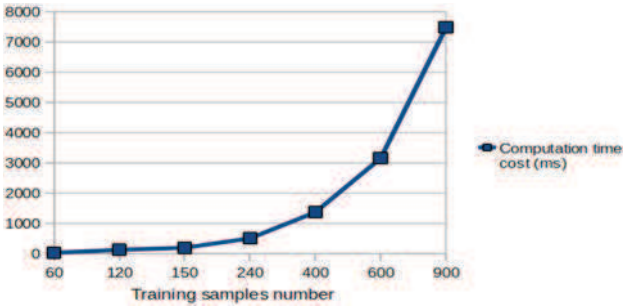


Figure 15. Consuming time of spectral clustering to different numbers of eye samples

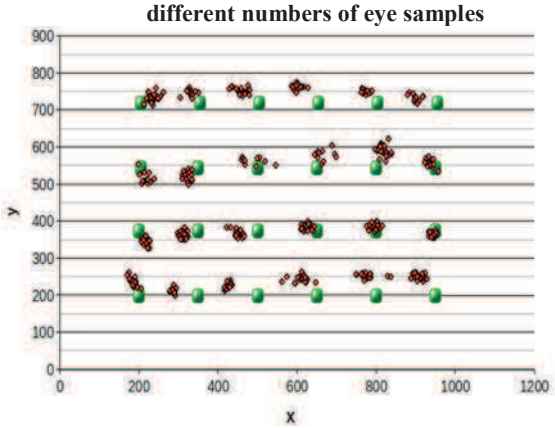


Figure 16. The gaze estimation result with a mean degree of 0.66, max degree of 1.45.

4.4 Gaze estimation results

The Fig. 10 shows the comparison between our sparse GPR method and conventional GPR method. The results are presented by $degree_{min}$, $degree_{mean}$ and $degree_{max}$. As for the sparsity, the number of unlabelled samples varies depending on the subject's behaviour during the calibration phase. Generally there are about 20∽50 samples for 5-points calibration. The result demonstrates also with more calibration point, the more efficient and accurate of using a sparse semi-supervised solution to predict the gaze. The 5-points calibration can be more efficient to learn eye manifold structure, and it provides a result of 0.9° as mean error degree, and 2.624° as max error degree. Moreover it takes about 20 -30 ms for the consuming time. The consuming time for spectral clustering to a 150 eye samples is about 206 ms (Figure 15). The sparse method takes about 0.55 ms to get the prediction for each eye data by a sparse degree of 0.3 - 0.6. The Figure 16 shows the visual gaze estimation results. Table 1 shows that our method offers little reduction in accuracy compared to the others by using less number of training samples.

Table 1. Comparison with other methods

Method	Error	Training samples
proposed	0.92°	5 labelled and 20 ~ 50 unlabelled
S^3GP+ edge+filter[21]	0.83°	16 labelled and 75 unlabelled
S^3GP[21]	1.32°	16 labelled and 75 unlabelled
Tan et al. [17]	0.5°	252 labelled
Baluja et al. [1]	1.5°	2000 labelled

5. CONCLUSIONS

We presented our appearance-based gaze tracker which uses subregion CS-LBP concatenated histogram as eye appearance feature. The feature not only can reduce the dimensionality of eye images, but also can be robust against the changes in illumination. Additionally, we introduced our sparse semi-supervised Gaussian Process Regression method with the supervision information of data by using spectral clustering. Spectral clustering to the eye data helps to learn about the "manifold structure" and give an efficient calibration phase. As a consequence, sparse semi-supervised GPR provides a more accurate prediction even when the number of calibration samples is limited. Here our gaze tracker can lead to a better result of 0.9° with 5-points calibration. The efficiency and reasonable accuracy can help to provide a real-time application.

6. ACKNOWLEDGMENTS

This work is supported by company UBIQUIET, and the French National Technology Research Agency (ANRT).

7. REFERENCES

[1] Baluja, S. and Pomerleau, D. 1994.. Non-intrusive gaze tracking using artificial neural networks. *Advances in Neural Information Processing Systems.*

[2] Belkin, M. and Niyogi, P. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. *NIPS*, 15, 6, 1373–1396.

[3] Belkin, M. and Niyogi, P. 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15, 6, 1373–1396.

[4] Fukuda, T., Morimoto, K. and Yamana, H. 2011. Model-based eye-tracking method for low-resolution eye-images. *2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction*.

[5] Hansen, D. W. and Ji, Q. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 32, 3, 478–500, Mar.

[6] Heikkilä, M.,Pietikäinen, M. and Schmid, C. 2009. Description of interest regions with local binary patterns. *Pattern Recogn.*, 42, 3, 425–436, Mar.

[7] Lee, K.-C. and Kriegman, D. 2005. Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, 852–859, Washington, DC, USA, IEEE Computer Society.

[8] Lu, F., Sugano, Y., Okabe, T. and Sato, Y. 2011. Inferring human gaze from appearance via adaptive linear regression. In D. N. Metaxas, L. Quan, A. Sanfeliu, and L. J. V. Gool, editors, *ICCV*, pages 153–160. IEEE, Jan.

[9] Martinez, F., Carbonne, A. and Pissaloux, E. 2012. Gaze estimation using local features and non-linear regression. *ICIP(International Conference on Image Processing)*.

[10] Morimoto, C. H., Koons, D., Amir, A. and Flickner, M. 2000. Pupil detection and tracking using multiple light sources. *Image and Vision Computing,* 331–335.

[11] Nguyen, B. L., Chahir, Y. and Jouen, F. 2009. Eye gaze tracking. *RIVF '*09.

[12] Noris, B., Benmachiche, K. and Billard, A. 2008. Calibration-free eye gaze direction detection with gaussian processes. *Proceedings of the International Conference on Computer Vision Theory and Application*.

[13] Rahimi, A., Recht, B. and Darrell, T. 2005. Learning appearance manifolds from video. *In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* - Volume 1 868–875, Washington, DC, USA, 2005. IEEE Computer Society.

[14] Rasmussen, C. E. and Williams, C. K. I. 2006. Gaussian processes for machine learning. MIT Press.

[15] Shih, S.-W. and Liu, J. 2004. A novel approach to 3d gaze tracking using stereo cameras. *IEEE Trans. Systems, Man, and Cybernetics*.

[16] Stiefelhagen, R.,Yang, J. and Waibel, A. 1997. Tracking eyes and monitoring eye gaze. *Proc. Workshop Perceptual User Interfaces*.

[17] Tan, K. H., Kriegman, D. J. and Ahuja, N. 2002.. Appearance-based eye gaze estimation. *Proc. Sixth IEEE Workshop Application of Computer Vision '*02.

[18] Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J. 1995. Active shape models– their training and application. Computer vision and image understanding, 61, 1, 38–59.

[19] Wang, J.-G., Sung, E. et al. 2005. Estimating the eye gaze from one eye. *Computer Vision and Image Understanding*.

[20] Weinberger, K. Q. and Saul, L. K. 2006. Unsupervised learning of image manifolds by semidefinite programming. *Int. J. Comput. Vision,* 70, 1, 77–90.

[21] Williams, O., Blake, A. and Cipolla, R. 2006. Sparse and semi-supervised visual mapping with the s3p. *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*.

[22] Xu, L.-Q., Machin, D. and Sheppard, P. 1998. A novel approach to real-time non-intrusive gaze finding. Proc. *British Machine Vision Conference*.

[23] Zhang, J., Li, S. Z. and Wang, J. 2004. Manifold learning and applications in recognition. In *Intelligent Multimedia Processing with Soft Computing,* 281–300. Springer-Verlag.

[24] Zhu, Z. and Ji, Q. 2007. Novel eye gaze tracking techniques under natural head movement. *IEEE TRANSACTIONS on biomedical engineering*.