Eötvös Loránd Tudományegyetem
Természettudományi Kar

Retina alapú mintavételezés arckomponens detektálási feladaton

Szakdolgozat

Szabó Zoltán
Alkalmazott matematikus hallgató

Témavezető: Szatmáry Botond
Informatika Ph.D. hallgató
ELTE TTK Információs Rendszerek Tanszék

2003.

Köszönetnyilvánítás

Elsősorban szeretnék köszönetet mondani témavezetőmnek, Szatmáry Botondnak, aki rendkívül sok segítséget nyújtott szakdolgozatom elkészítésében. Folyamatosan tanácsokkal látott el, gondosan és türelmesen válaszolt a felmerülő kérdéseimre, fáradhatatlanul támogatott munkámban.

Hálás vagyok Lőrincz Andrásnak, az egyetemen működő NIPG csoport vezetőjének. Ő ismertetett meg a csapatmunka szellemével, az együttes alkotás örömével. Megtanított a tudományok összefonódásának fontosságára, elképesztő munkabírása példát mutatott. Csapatának tagjaiban valódi barátokra leltem.

Köszönettel tartozom az ELTE tanárainak, akik rendíthetetlenül kalauzoltak a matematika rögös ösvényein.

Hálámat szeretném kifejezni szüleimnek, akik biztosították tanulmányaimhoz a családi békét, és harmóniát.

Köszönet testvéremnek és parányi, huncut kislányának, kik vidámságukkal állandó fényt csempésztek életembe.

Tézisek [1]

A retina egy közismert jellemvonása a változó mintavételezési sűrűség: a középen levő nagyfelbontású rész (fovea), a periféria irányába haladva log-polár struktúrával leírható környezettel párosul. Az az általános vélekedés, hogy a log-polár szerkezet elsődleges oka a mozgásfelismerésben van: az idegrendszer előnyben részesíti ezt a nemlineáris transzformációt, hogy a fontos mozgások, mint például a közeledés (nagyítás), elfordulás (forgás) során fellépő változásokat egyszerű transzlációval (eltolással) közelíthesse.

Dolgozatomban azt a kérdést vizsgáltam meg, hogy a retina ismertetett tulajdonsága alapján kialakított mintavételés okoz-e számottevő hatékonyságbeli romlást. Tesztfeladatul a viselkedés szempontjából fontos, arckomponens detektálási feladatot választottam.

(1) A létrehozott biologilag motivált mintavételezési technikát a főkomponens analízis (PCA) eszköztárával összekötve, arckomponens felismerő rendszert építettem és a nyilvános FERET arc-adatbázison teszteltem.

(2) Önálló értékelési elveken alapuló összehasonlító eljárással vizsgáltam a retina alapú módszer, a log-polár illetve az egyenletes mintavételezési módok hatékonyságát. (A változó fovea méret két extrém esetének tekinthető a log-polár és az egyenletes mintavételezési mód.)

(3) Eredményül azt kaptam, hogy a kérdéses biológia analógiára épülő mintavételezés alkalmazása nem okoz jelentős romlást, sőt az általam vizsgált feladat esetén előnyösnek bizonyult. Az elforgatás és transzláció invariáns log-polár módszert felülmúlta, az egyenletes technika eredményességét megtartotta, helyenként javította.

(4) A javasolt módszer kis képterületeken hatékony keresést valósít meg, egyfajta hasonlósági mértéket nyújt, így lehetővé téve sztochasztikus szűrős keretbe való integrálását, amely speciálisan arcrészek esetén valós idejű követésre alkalmazható.

[1] A szakdolgozat magvát képező cikk az *Image and Vision Computing* szaklapba lett beküldve.

Contents

List of Figures

Does retina based architecture cause significant drawback in face component recognition?

Zoltán Szabó [a], Botond Szatmáry [a], András Lőrincz [a,*]

[a] *Eötvös Loránd University, Department of Information Systems, Pázmány Péter sétány 1/C, Budapest, Hungary H-1117*

Abstract

We examine a combined sampling technique suggested by well-known properties of the human retina, such as the greatest visual acuity in the center (fovea) and exponentially decreasing resolution toward the periphery. We modelled the high resolution part with uniform sampling, and log-polar sampling was applied at the periphery. The aim of the paper is to investigate whether this retinotopic sampling gives rise to a considerable deterioration in efficiency compared with the commonly used uniform and log-polar techniques in case of face component recognition task on the FERET database. We have found that it certainly does not have considerable drawbacks, moreover it seems to be favourable in the present task.

Key words: retinotopic sampling, uniform fovea with log-polar periphery, face component detection, PCA reconstruction

1 Introduction

The make-up of the retina proposes a sampling structure that has a high resolution central region and a less detailed periphery. A number of approaches approximate this property and describe the sampling method of the retina with log-polar technique (Wilson (1983); Boluda and Domingo (2001); Bernardino et al. (2002); Koh et al. (2002); Smeraldi and Bigun (2002)). The log-polar construction can be mainly rooted in motion detection according to the common view. It has been widely accepted that the nervous system prefers this

* Corresponding author

Email addresses: szzoli@cs.elte.hu (Zoltán Szabó), botond@inf.elte.hu (Botond Szatmáry), lorincz@inf.elte.hu (András Lőrincz).

URL: http://people.inf.elte.hu/lorincz (András Lőrincz).

nonlinear transformation, because it allows important operations, such as rotation and scaling generated changes to be approximated by translations. We compare the efficiency of a combined sampling method, 'uniform fovea with log-polar periphery' (UFLP) with the commonly used log-polar and uniform sampling techniques in a face component recognition task on the FERET face database[1] (Phillips et al. (1998)).

The organization of the paper is as follows. In Section 2 the basic features of the applied mathematical transformation are reviewed, the examined sampling techniques are presented, the preprocessing and the training methods are also described. Testing results are shown in Section 3. Section 4 summarizes the paper.

2 Methods

2.1 Log-polar representation

Log-polar transformation (L) is a mapping from the Cartesian plane onto the cortical plane:

$$L(x,y) = \begin{bmatrix} \log\left(\sqrt{x^2+y^2}\right) \\ \arctan\left(y/x\right) \end{bmatrix} = \begin{bmatrix} \log r(x,y) \\ \varphi(x,y) \end{bmatrix}.$$

The major advantage afforded by the space-invariant log-polar geometry is the coarser representation near the periphery of the grid, in other words it has exponentially decreasing resolution towards its edge. Moreover it has other attractive features: By its effective data reduction ability the volume of calculations can be diminished; It also allows multi-resolution analysis, which, together with divergency handling and the wide field of view supplied by the periphery, play a substantial role in motion detection (Juday and Weiman (1990); Ferrari et al. (1995)).

2.2 Sampling

Sampling was performed in a 85 × 85 pixel size image window to completely cover the examined face components. We compared three different sampling

[1] In our work, three face components are investigated, namely right eye, left eye and nose.

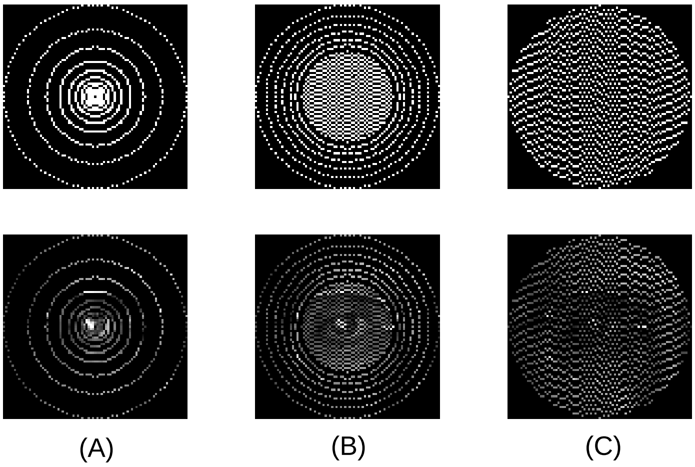

Figure 1. The compared sampling strategies. In the first row the distribution of the sampling points for different sampling techniques are displayed: (A) log-polar sampling, (B) 'uniform fovea with log-polar periphery' sampling, (C) uniform sampling. In the second row the sampled pixels of a right eye are shown.

techniques with fixed sampling point numbers (the number of sampling points was 1368): (i) In case of log-polar sampling 12 circles with the same center and exponentially increasing radii were laid on the inspected area and 114 sampling points were placed on each circle uniformly (Fig. 1(A)). The radius of the outside circle was 42 (in pixel units); (ii) The UFLP method approximates the retinotopic distribution: 684 uniformly distributed sampling points were settled in the middle of the inspected area in a circle with $r = 21$ pixel units, corresponding to the fovea, and the rest 684 points were placed on the anchoring ring (Fig. 1(B)) to get log-polar representation (6 circles, 114 point on each circle); (iii) In the third case all the 1368 sampling points were uniformly distributed in a circle with $2r = 42$ pixel units (Fig. 1(C)).

2.3 Database

We used the frontal images of the FacE Recognition Technology (FERET) database (Phillips et al. (1998)) for learning and testing. Every image in this collection is 8-bit gray-scale, has a 384×256 pixel size, containing a centered human head. In order to reduce the effect of variation in illumination on these images, histogram equalization was employed as a preprocessing step. The goal of this method is to achieve uniform distribution on the gray-level values of the original image.

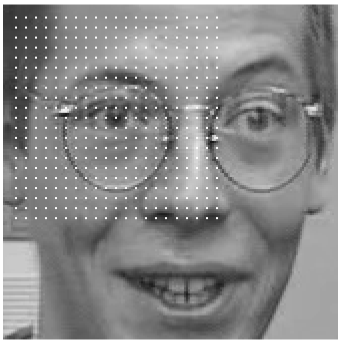

Figure 2. A test grid was placed around the exact position of the studied face component (FERET database provides this data). For testing the different sampling techniques, in every node of the grid a probability value was calculated expressing the chance (probability) of being the center of the actual face component.

2.4 Training

Principle component analysis (PCA) was used in our face component recognition task, i.e., principal components were derived for the different face components on the histogram equalized images employing the presented sampling techniques using intensity values of the images. PCA is a projection technique (employing orthonormal basis), it is computationally efficient, it has effective data reduction capability and it provides naturally defined similarity measure (L_2 norm) between the original and the reconstructed image. For details about PCA the interested reader is referred to Movellan (1997).

For this training process 2061 frontal images were chosen from the FERET database. In the database, the true location of the face components are also available making it unnecessary to find them manually. These face components were cut from the faces, sampled according to the already described sampling methods, then principal components were developed on them.

2.5 Testing principles

Around the center of the inspected face component [2] a quadratic test grid was laid down with $step = 4$ pixel in both directions (see Fig. 2). The nodes of the grid were used to find the approximate position of the requested component [3]. In every node a certain fitting examination was executed utilizing the reconstruction property of PCA: an image window was used around the grid points, this image underwent histogram equalization, it was sampled according to the

[2] Similar results were obtained for left and right eye. From now on, eye stands for both left and right eyes.

[3] The grid with $step = 4$ pixel unit was satisfactory and computationally efficient.

different sampling methods, and it was filtered and reconstructed by means of the previously trained PCA basis. A natural similarity measure between the original and the reconstructed image offered by PCA is the reconstruction error:

$$r_e := \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_2 ,$$

where $\mathbf{X}$ stands for the original, and $\hat{\mathbf{X}}$ for the reconstructed image.

An error surface was created using the above described reconstruction error. Around every node of the grid three different image window was cut: (1) one with 85×85 pixel size (as introduced in Subsection Sampling), (2) a 68×68 pixel size one and (3) a 102×102 pixel size one. The smaller and greater image windows were resized to 85×85, so we obtained three images with the same size representing three different scales (80%, 100%, 120%). For these three images the reconstruction error values were calculated and the minima of these error values was assigned to the corresponding node. This technique made it possible to rectify the scaling variation being present in the FERET database.

The values of the error surface were linearly rescaled to interval [0,1] and then '1 - error surface' was computed. All the further analysis were carried out on the obtained surface, whose points can be interpreted as probability values. High probability suggests the presence of the desired face component at around the corresponding node, whereas low probability proposes the opposite. Henceforth the received surface is referred to as probability surface. The examined image window is accepted to be a face component candidate, if the probability value of its correspondent node is strict local maxima on the probability surface and if it is above a predefined threshold.

In our simulations, reconstruction with the first three principal components were studied (Koh et al. (2002) applied 2-4 basis vectors for face reconstruction) and the threshold value was 0.9.

3 Results

500 randomly chosen frontal images of the FERET database were used to compare the performance of the UFLP method with standard log-polar and uniform sampling techniques. The probability surface and the potential face components were calculated for each image.

After ordering the potential face component candidates on the ground of their probability values, the distance between the most probable position (first candidate) and the true location was evaluated. According to Fig. 3, the retina

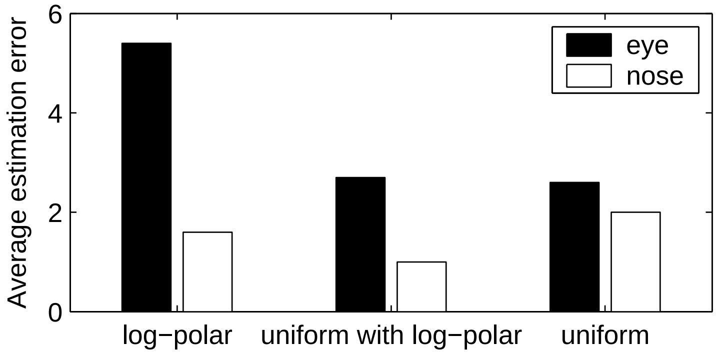

Figure 3. The average distance between the position of the best candidate (the candidate with the lowest reconstruction error) and the true location of the searched component are plotted as a function of the chosen representation. The black and white bars correspond to eye and nose, respectively. The retina-like UFLP sampling method has favorable properties (is more precise) on the face component task.

based UFLP sampling structure has considerable advantages over log-polar sampling and is better by a small margin than the uniform method.

The average number of candidates were also evaluated yielding similar results (see Fig. 4). The UFLP sampling method doesn't seem to suffer from serious disadvantages. It considerably surpasses the log-polar technique, and proposes on average 1.6 candidates for eye, and 1.2 for nose, likewise the uniform method. The larger average number of candidates for eye than for nose is due to the glasses: people with glasses were not excluded from the training and testing phases.

The average probability surfaces are plotted in Fig. 5, reflecting all the discussed features. The retina-like UFLP sampling structure has smoother probability surfaces than the log-polar sampling. Peaks are steepest at the exact location of the face components, suggesting more accurate estimation for the position of the searched component.

4 Discussion

Our experiments were based on widely known properties of the human retina: it has high resolution at the center, called the fovea, and the sampling density (i.e. the actual resolution) falls-off quickly towards the edges. The approximately log-polar structure of the periphery approximates essential transformations, such as rotation and scaling, by simple translations. We have studied if a uniform fovea with log-polar periphery exhibits considerable changes on the reconstruction abilities.

As a test bed, face component recognition problem was chosen using the

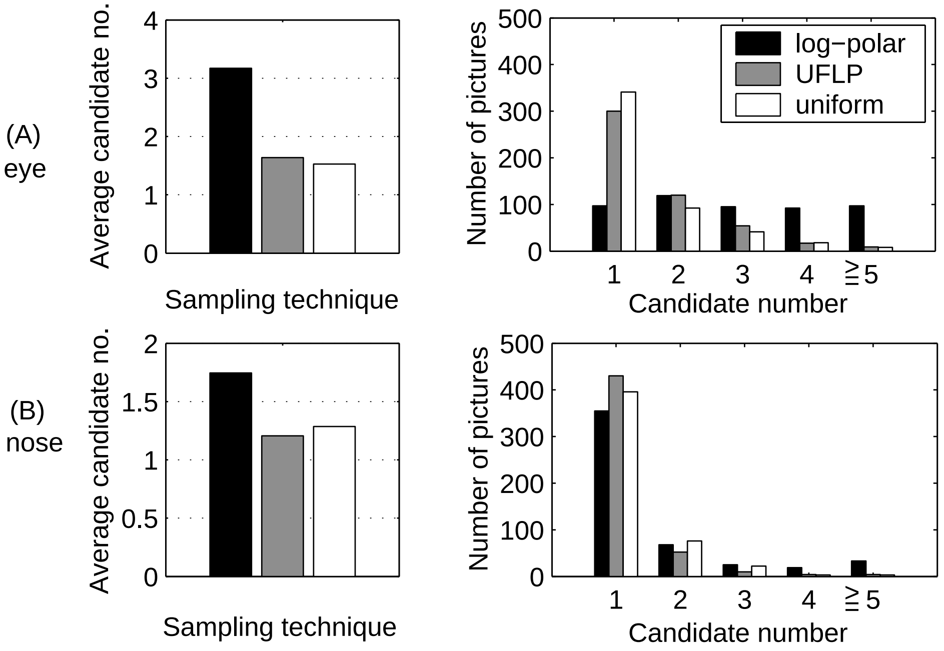

Figure 4. Left hand side: the number of average eye candidates compared for different – log-polar, UFLP and uniform – sampling techniques. Right hand side: Same with more details. The histograms show the number of pictures, where $1, 2, 3, 4, \geq 5$ candidates were predicted. Gray scales indicate different sampling methods. (A) and (B) refer to eye and nose.

FERET gallery. According to the estimation error concerning the exact location of the face part and the number of the potential candidates, the efficiency of the retina based sampling was compared to log-polar and uniform techniques. It was found that the biologically motivated UFLP approach seems favourable on face component recognition. As it was demonstrated, the UFLP sampling technique combined with PCA provided an efficient similarity measure for face components.

Acknowledgements

Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

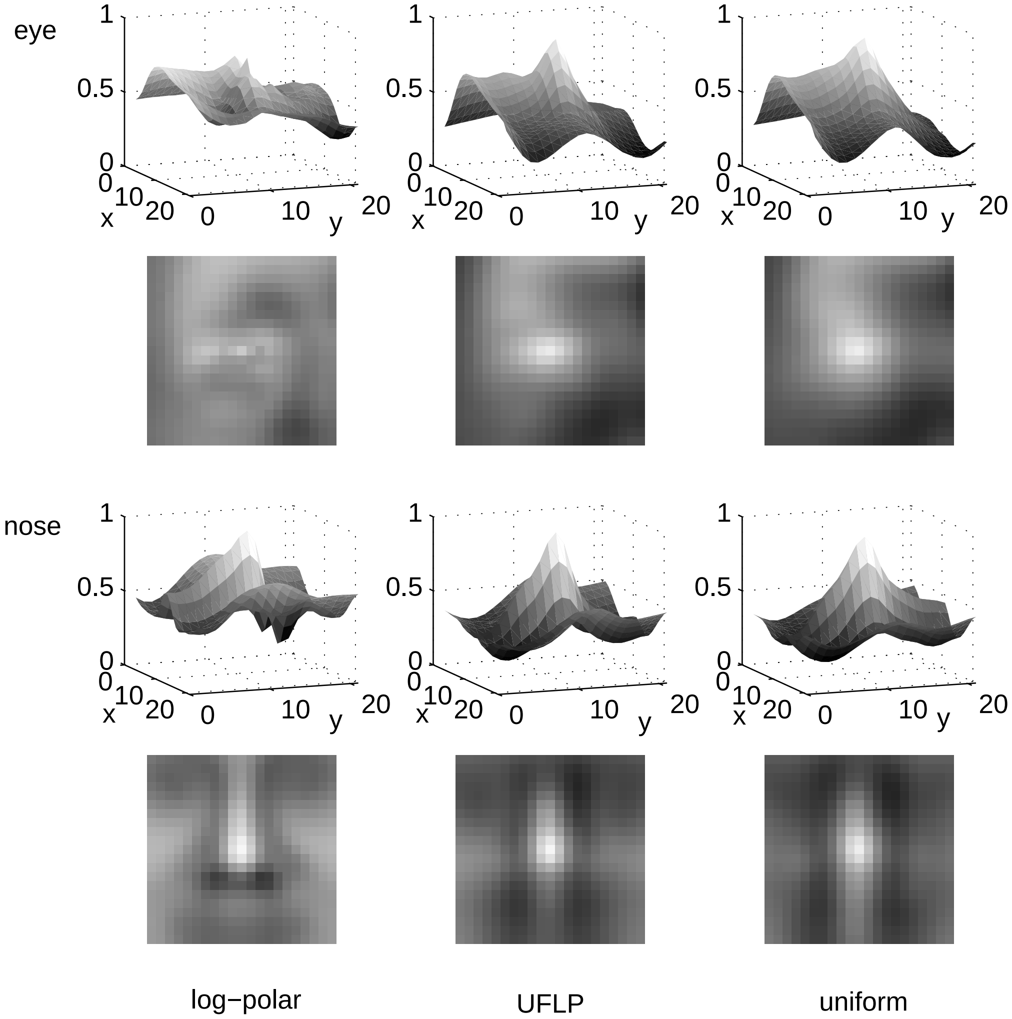

Figure 5. The average probability surfaces for the compared sampling techniques are drawn from different points of view for eye and nose. The point with the largest probability value is the lightest, whereas the darkest pixel denotes the smallest chance. Results for eye and nose are shown for all three sampling methods.

A Appendix

A.1 Object (particularly face) detection

The current evolution of computer technologies has created an accelerating world with machines employing artificial intelligence to facilitate human life. An essential task in these systems is object, or particularly face (part) detection, which is one of the visual tasks which humans can do absolutely effortlessly, however in computer vision terms it is not really easy. Anyone can put the question: how these methods work, or first of all what kind of techniques exist?

A.1.1 A short historical review

Labeled Graph Matching (LGM) from von der Malsburg and L. Shams (1988), is a well-known algorithm addressing object detection as a general recognition task. LGM represents each pattern as a labeled graph, where labels encode features, and links between nodes express topological relationships. In this formulation of visual pattern, an input can be recognized by finding the best, approximate neighbourhood and feature preserving correspondence with a stored graph. In most of the implementations, magnitudes (see von der Malsburg et al. (1998)) of Gabor wavelets (for details, study Grossmann and Mortlett (1985)) with different orientations and frequencies constitute the labels of the nodes in the superimposed graph. The explanation of Gabor wavelets beyond its biologically motivation, is the ability that it encodes local gray level distribution of an image with greater emphasis.

Neural implementation of LGM is exemplified in the Dynamic Link Architecture (DLA) by von der Malsburg (1981) with great success in object recognition.

An extension of the standard LGM, called LGM+ with direct biological analogy (based on findings of the visual cortex) has been introduced by Shams et al. (2001) and successfully tested on the task of finding 3-D stochastically generated, digital embryos in complex scenes. It is compared to a statistical approach called Mutual Information Maximization (MIM), which has enjoyed much attention in recent years and has been adopted by several groups (for example, see Viola (1995); McGarry et al. (1997); Pluim et al. (2000)). From the wide spectrum of other statistical approaches to the problem of pattern recognition Shannon information as described by Becker (1995), description length applied by Bienenstock and von der Malsburg (1987), and support vector machines (SVM) introduced by Vapnik (1995) cover three more examples.

The problem of facial feature detection is solved with employing multiscale and multiorientation Gabor kernel on a log-polar grid combined with SVM to model the Saccadic Search strategy by Smeraldi and Bigun (2002). Also Gabor decomposition is a basis of an eye detection and Saccade modelling system demonstrated by Smeraldi and Bigün (1998). While in the aforementioned studies all the Gabor kernels obtained unified role, a method has been developed by Kalocsai et al. (2000) to weight the contribution of the these kernels according to their predicting power and been evaluated in a special, facial object recognition task.

Neural networks, which are more than simply Multilayer perceptrons (MLP), have become a popular technique for pattern recognition problems. Modular architectures, autoassociative and compression networks, network evolved with genetic algorithms are all examples of their widespread use in this topic. The first neural approaches were based on MLPs implemented by Burel and Carel (1994); Juell and Marsh (1996); Propp and Samal (1992). To see more examples, Lin et al. (1997) created a fully automatic recognition system based on probabilistic decision-based neural network (PDNN), and a new learning architecture called SNoW (sparse network of winnows) by Roth (1999) was successfully applied to face detection task.

A naturally emerging idea is to conceive to a kind of object as it is lying in the overall image space. To represent this subspace, standard multivariate approaches can be applied including independent component analysis (ICA) to extract features (for discussion, examine Hyvarinen (1999)), or factor analysis (FA) as proposed by Yang et al. (2000), or principal component analysis (PCA), where each individual 'face' is approximated by the linear combination of the so called eigenvectors, using appropriate weights, as a it has been done in the present work.

As it can be seen recognition of faces (or face components) is a rapidly progressing research area and its application field covers a quite broad spectrum.

A.1.2 Possible application areas of object detection

A.1.2.1 Biometrics Among the recently emerging technologies used in artificial intelligence, biometrics (which intensively makes use of the various object detection solutions) is one of the most exciting and appealing one. Some significant advantages afforded by an ideal biometric system are the followings:

- All members of a population possess several unique keys, such as irises or fingerprints.
- These unique features provide biometric passwords, which can't be stolen, forgotten or lost.

Plenty of biometric security systems have been actively used in our life (for details, see Phillips et al. (2000)). Fingerprint or iris pattern recognition based biometric identification can be imagined in high-level security buildings (these biometric passwords exchange their traditional alphanumeric counterparts). Protection of private information carried in mobile electronic devices are also ideal test-bed for similar techniques (e.g. face recognition screen-savers in commercial handhold devices, i.e., laptops). The dramatically decreasing size of the fingerprint-capture device (not larger than a postage map) and its cost (less than $100) make this technique more widely spread in the private sector (next to the earlier, exclusive government applications), where convenience and security are both significant standpoints.

A bit futuristic, but practical solution for personal fingerprint authentication would be a universal key facilitating the access to everything from front doors to car doors, bank machines and computers.

Police is also employing biometric based techniques, such as recognition of subjects from mug-shots, passport photos and scanned fingerprint, or from latent fingerprint left at a crime scene.

A.1.2.2 FACS The objective description of facial behaviour from a video, or specially facial expression recognition, is a widely known and challenging problem in computer society. For example, measurements of facial behaviour at the level of detail of FACS[4] provide information for detection of deceit, including information about whether an expression is posed or genuine and leakage of emotional signals that an individual is only attempting to suppress (for complete discussion, see Ekmann (2001)).

I mention two more interesting facts:

- Recognition of 'Duchene' and non Duchene smiles can also be handled in this framework. Genuine, happy smiles can be differentated from posed, or elsewhere called social smiles by the contradiction of muscles (encoded by the action unit 6) circling eyes.
- An experiment was carried out to test the efficiency of FACS in lie detection. Surprisingly the detection rate based on this technique was significantly higher than the detection rate of both naive human subjects and police officers watching the same video.

Among the numerous applications of object (human face) recognition, the followings are further examples:

[4] FACS is the abbreviation of Face Action Coding System, which decomposes facial motion into component actions, i.e., describes face motions as it was generated by approximately independent muscles.

- Automatic visual surveillance systems.
- Monitoring of the alertness and anxiety of a pilot.
- Indexing of image and video databases.
- Nowadays, repelling of terrorist attacks has also unfortunately got to the center of attention, consisting of the task of picking out faces from crowd, or in other words searching for faces and extracting them from crowded pictures.
- In intelligent human-computer interfaces (HCI), detection and tracking of faces is the first step in building up interaction between the human and the computer.
- The wide spreading of Internet commerce and tele-banking require privileged and remote access to resources. In this environment, face authentication is a naturally emerging idea comprising the problem of face detection.
- Face detection technology can be useful and necessary in video conferencing, where there is a need to control the camera in such a way that the current speaker always has the focus.

A.2 Description of the employed mathematical procedures

A.2.1 Principal component analysis (PCA)

A key problem in statistical pattern recognition is that of feature selection or extraction. This task refers to a process, where the original data space is transformed into the so called feature space, which has exactly the same dimension as the original input space. The purpose is to find a coordinate transform, so that with reduced number of features the intrinsic information in the data can still be retained, supplying an informative and compact description of the inputs. The problem can be formulated and considered in several ways:

(1) The basic idea behind the PCA method is the following. Possessing a random variable X with mean μ[5] and variance matrix Σ, given a constant k, the purpose is to find a k dimension linear subspace denoted by S and an orthonormal basis (called principal components) in it with the property that the projection of X into S (indicated with $\hat{X}$) minimizes the reconstruction error in least-squares sense:

$$\min_{\mathrm{S:dim(S)=k}} E\left\|X-\hat{X}\right\|^2, \text{ where } \hat{X}=U\cdot U^T\cdot X.$$

The columns of matrix U are the orthonormal basis vectors and E denotes the statistical expectation operator. An interesting fact, that regardless of the distribution of X the optimal solution can be achieved using only the

[5] For simplicity of notations, without loss of generality, we can assume that $\mu=0$.

mentioned first and the second order statistics of X, declares Movellan (1997), where the reader can also find more details about PCA.

(2) Let's suppose an n dimension random variable X with zero mean[6] and variance matrix Σ. Let's denote u a unit vector ($\|u\|_2 = u^T \cdot u = 1$) with the same dimension as X. Taken the projection of X onto u, we get a random variable a:

$$a = u^T \cdot X = X^T \cdot u,$$

with mean and variance related to the original data vector. Its mean value is zero, indeed:

$$E(a) = E(u^T \cdot X) = u^T \cdot E(X) = 0.$$

Consequently, the variance of a is equal to its second momentum, which has the implication, that

$$\begin{aligned}\psi(u) := \sigma^2 = E(a^2) &= E((u^T \cdot X) \cdot (X^T \cdot u)) = u^T \cdot E(X \cdot X^T) \cdot u \\ &= u^T \cdot E(X \cdot X^T) \cdot u = u^T \cdot \Sigma \cdot u.\end{aligned}$$

Now our purpose is to find the extremal or stationary values of ψ. As it can be seen, this is a quadratic optimization problem with constraints, in other words, we are to look for the extreme values of a quadratic form on the unit sphere in $\mathbb{R}^n$. Asking for the help of the method of the Lagrange multipliers, the following equation governs the optimal solution:

$$\Sigma \cdot u = \lambda \cdot u,$$

in which the eigenvalue problem can recognized, commonly encountered in linear algebra. The nontrivial solutions of this problem (i.e., $u \neq 0$) exist only for special values of λ (these numbers are called the eigenvalues (λ_i) of Σ), with the associated eigenvectors (u_i), which are unique, if the eigenvalues are different (as it's further assumed)[7]. Introducing the notation

$$U = [u_1, \ldots, u_n],$$

an interesting fact, that U is orthogonal ($U^T \cdot U = I$).

To reconstruct the original data vector (X) from the a_i principal components, we combine them to a vector

$$a = [a_1, \ldots a_n],$$

and get

$$x = U \cdot a = U \cdot U^T \cdot X = \sum_{i=1}^{n} a_i \cdot u_i,$$

[6] If X has non zero mean, then we simply subtract the mean from it and continue the analysis with the resulting variable.

[7] The extreme values of the variances (ψ) are equal to λ_i.

which can be viewed as the synthesis formula. In this sense, u_i are the basis vectors of the data space with the correspondent basis coefficients (a_i), or in other words the X data point has been transformed into an a point lying in the feature space.

The number of features can be reduced, by discarding components inducing small variances (σ_i). Indeed, recovering the data using only the first m principal components[8]

$$\sum_{i=m+1}^{p} \sigma_i^2 = \sum_{i=m+1}^{p} \lambda_i$$

total variance occurs for the approximating vector, so the closer these eigenvalues are to 0 (the variance of X is only concentrated into few directions), the more effective data reduction can be achieved.

As an illustration, I show, how PCA can accomplish image compression in practise. The test image was divided into small square pathes with side of 15 pixels, constituting the input samples of the random variable X. The experimental variance matrix was determined, and all the patches (consequently the full image) were recovered using different number of principal components (see Fig. A.1) achieving dimension reduction.

(3) We can look at the sketched compression problem from a neural network point of view too (see Fig. A.2). An input pattern (X) is transformed into a set of activations (a) in the hidden layer. The weight matrix U^T make up the connections between the input and the hidden layer ($a = U^T \cdot X$), and U creates the activation of the output layer from the hidden values ($\hat{X} = U \cdot a$). The path to the hidden can be seen as the analysis, while from the hidden to the output as the reconstruction step. The goal in this linear feed-forward network is to reconstruct the input data, as well as it is possible through the optimal U matrix (in mean square sense).

As PCA based reconstruction of face components serves as a basis of this work, for the sake of illustration of the eigenvectors, I have developed PCA components for images depicting human eyes using all the pixels in the sampling square (see Fig. A.3).

A.2.2 Log-polar transformation

In the examined, biologically inspired retina model, the periphery of the visual field around the fovea in the human visual system was represented involving log-polar transformation (see Fig. 1). As it was defined earlier, it's a mapping (L) from the points of the Cartesian plane to the cortical plane:

[8] Without loss of generality it can be assumed, that λ_i-s are arranged in decreasing order.

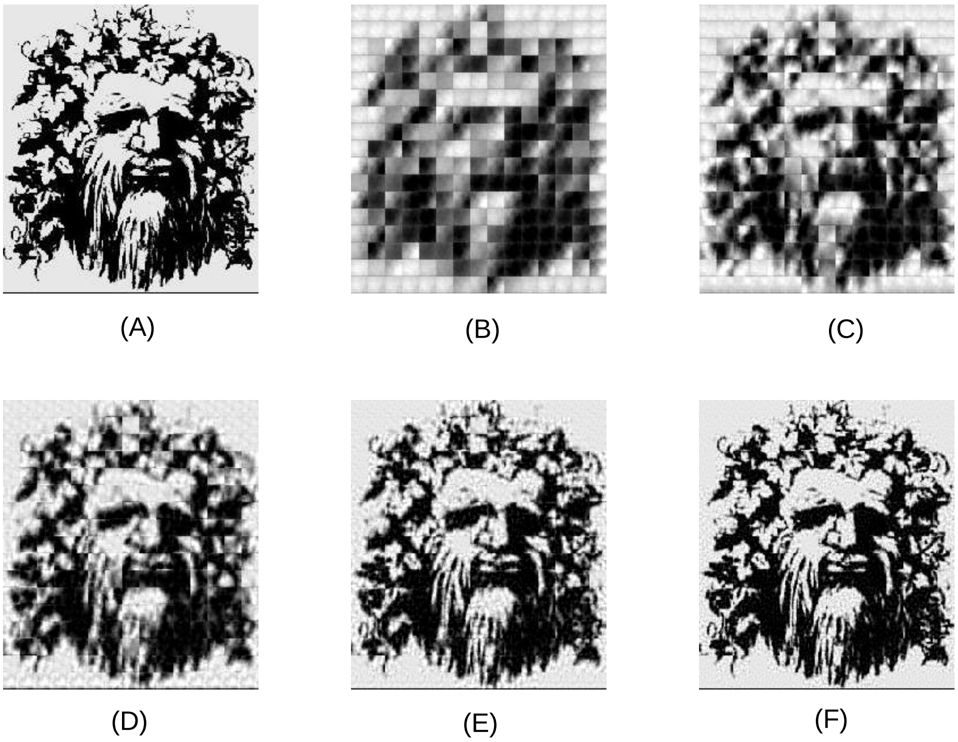

Figure A.1. I have executed PCA based image compression on the test picture (A), representing an exciting illusion (Can you find the hidden figures by staring at it? Never mind, it's not easy.). Square patches of (A) constituted the basis of the training procedure, as it can be inspected in (B). Using 1,2,5,10,20% of the developed principal components for recovery, respectively, I could achieve compression with percentage rates of 1,2,5,10,20% (B,C,D,E,F). As it can be seen, with the ratio of 20% (×5 compression) almost perfect reconstruction is possible.

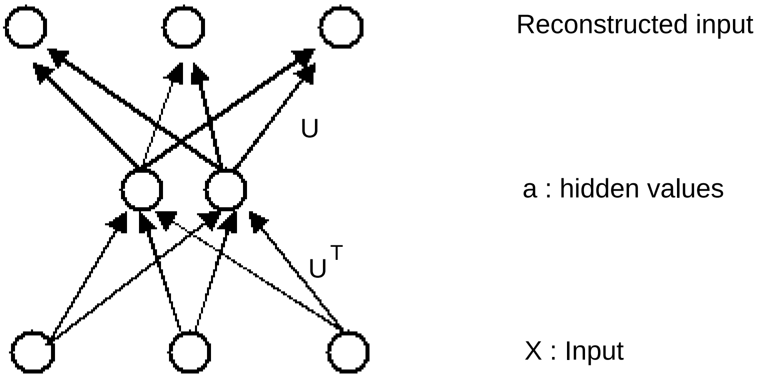

Figure A.2. Neural network interpretation of the compression problem

$$L(x,y) = \begin{bmatrix} \log\left(\sqrt{x^2+y^2}\right) \\ \arctan\left(y/x\right) \end{bmatrix} = \begin{bmatrix} \log r(x,y) \\ \varphi(x,y) \end{bmatrix}.$$

In fact, it's a coordinate transform, in other words, it makes possible to de-

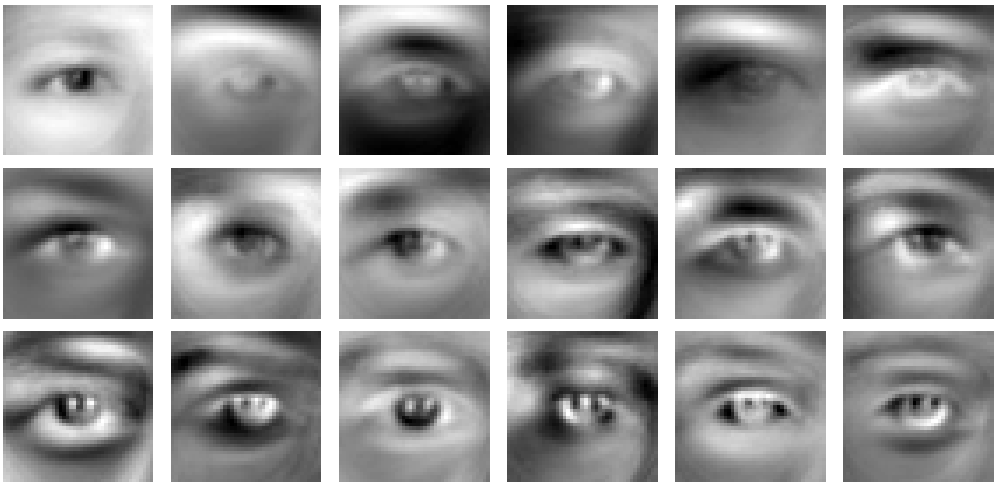

Figure A.3. This figure serves as an illustration of the eigenvectors (specially eigeneyes). Here, I have developed principal components from images containing eyes, and ordered them on the grounds of the corresponding eigenvalues (decreasingly) in row wise manner (the first 18 can be investigated). These are the perpendicular directions, which preserve most of the information content of the input data, i.e., give the best (linear) approximation in L_2 sense.

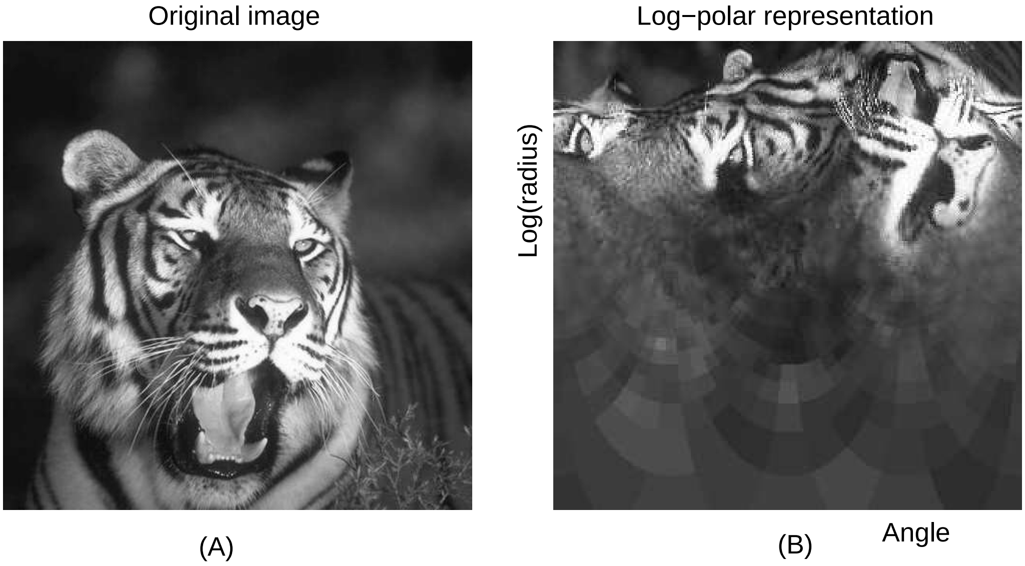

Figure A.4. A sample image (A) is plotted with its log-polar representation (B), for the sake of better understanding.

scribe or handle an f function (an image) by the use of a new coordinate system (example can be seen in Fig. A.4):

$$f(x, y) = f(e^r \cdot \cos(\varphi), e^r \cdot \sin(\varphi)).$$

Log-polar mapping has appealing properties:

- It provides a one-to-one correspondence with space varying geometry.

- Two fundamental transformations, namely scaling and rotation is converted to translation.

A.3 Biological analogy

A.3.1 Anatomy of the eye (retina)

Since the examined sampling technique was born on grounds of biological analogy, it's important to understand some essential concepts in connection with the human eye and become acquainted with the fundamental features of the retina.

The retina is a light-sensitive, complex layer at the back of the eye covering about 65 percent of the interior surface. Photosensitive cells (photoreceptors) called rods and cones in the retina convert incident light energy into signals that are carried to the brain by the optic nerve (see Fig. A.5). Making a study of rods and cones, the following can be declared:

- The rods are most sensitive to light and dark changes, shape and movement and only contain one type of light-sensitive pigment. Rods are not good for color vision (in a dim room, we use mainly our rods, but we are 'color blind'). Rods are more numerous than cones in the periphery of the retina. There are about 120 million rods in the human retina.
- The cones are not as sensitive to light as the rods. However, cones are most sensitive to one of three different colors (green, red or blue) and they only work in bright light. That's why you cannot see color very well in dark places. So, the cones are used for color vision and are better suited for detecting fine details. There are about 6 million cones in the human retina. Some people cannot tell some colors from others ('color blinds'). They don't not have a particular type of cone in the retina or one type of cone may be weak[9].

Cones and rods intermingle nonuniformly over the retina. At the edge of the retina there are only rodes (which are responsible for light-dark and motion detection). Coming closer to the center, we can find less rodes, while the number of the cones are growing, which results an increasing color sensitivity and augmenting resolution. In the middle of the retina there is a small dimple called the fovea (only cones) with quite high resolution and most color perception (Fig. A.6).

Summarizing, the retina is very different in terms of scene digitization from

[9] In the general population, about 8% of all males and 0.5% of all females are color blind.

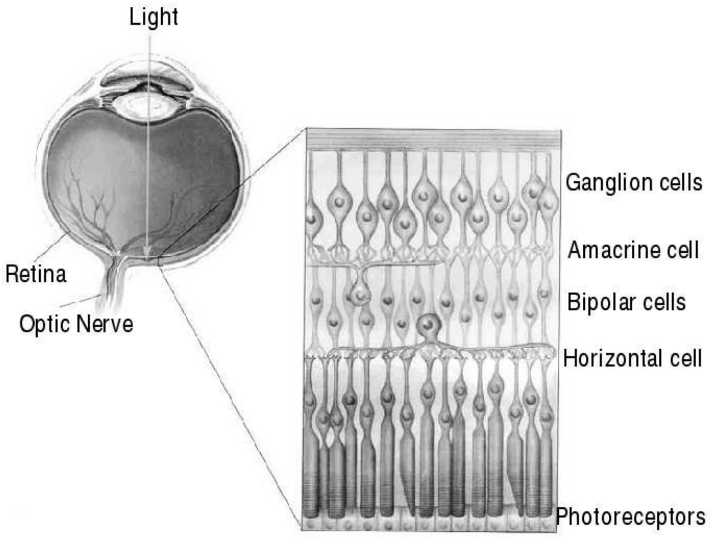

Figure A.5. The layered structure of the retina can be observed in the figure. Surprisingly, in most of the retina the cells responsible for capturing light and initiating neural signals – rod and cone photoreceptors – lie underneath a dense network of blood vessels and neurons. In other words, light must pass through several layers before it can be captured by the receptors.

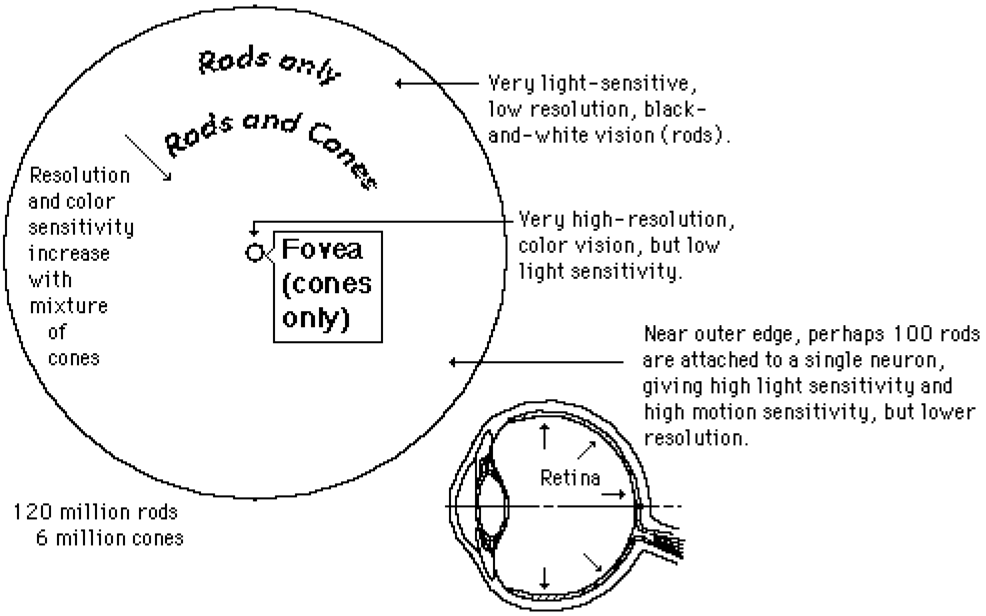
Rods only
Rods and Cones
Resolution and color sensitivity increase with mixture of cones
Fovea (cones only)
Very light-sensitive, low resolution, black-and-white vision (rods).
Very high-resolution, color vision, but low light sensitivity.
Near outer edge, perhaps 100 rods are attached to a single neuron, giving high light sensitivity and high motion sensitivity, but lower resolution.
Retina
120 million rods
6 million cones

Figure A.6. The varying color sensitivity, resolution and photoreceptor distribution is plotted in the figure with the location of the fovea.

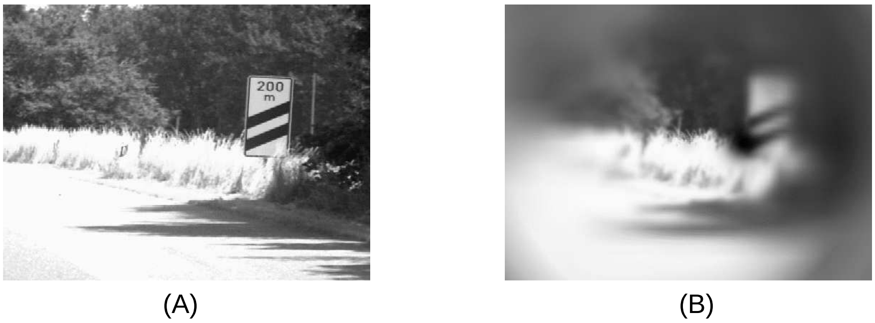

Figure A.7. Seeing the world through the retina. The original image from CCD camera (uniform sampling) can be inspected in (A), with the same image seen with a retina (right eye, fixation at center) in (B), where it's illustrated, that retina images with a very high resolution at the center (fovea) and the sampling density (i.e., the actual resolution) falls-off quickly towards the periphery.

a CCD camera which uniformly samples a visual scene. The retina with its specialized central region (fovea) images the immediate neighborhood of the point of fixation with a much higher resolution than peripheral regions of the visual field (see Fig. A.7).

A.3.2 Neurobiological analogy

The fovea is the area of the retina with the greatest acuity. It has the greatest density of ganglion cells and therefore has a much larger representation. Approximately half of the neural mass in the lateral geniculate nucleus and in the primary visual cortex represent the fovea and the region just around it, while the peripheral portion of the retina is less well represented, claims Kandel et al. (1991).

Summarizing, the examined retina based sampling technique was created considering the outlined biological constraints:

- Varying sampling density (depending on the spatial position of the sampling points, i.e., decreasing resolution towards the periphery).
- Representation of the retina (fovea and surrounding region) in the primary visual cortex.

A.4 Future work

As the studied, 'artificial retina' provides similarity measure for face components, instead of the direct, deterministic, candidate-like approach, it would be possible to integrate this 'sensor' into a stochastic filtering framework (see Isard (1998)), which performs significantly better than the most robust de-

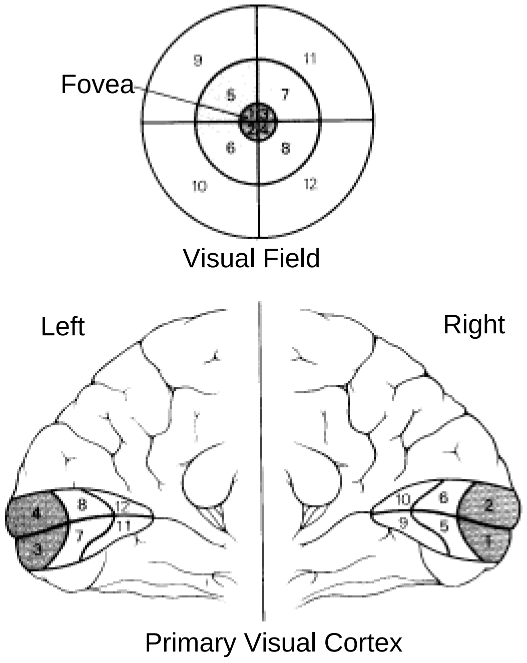

Figure A.8. Areas in the primary visual cortex are devoted to specific parts of the visual field, as indicated by the corresponding numbers ($1, 2, 3, 4 \leftrightarrow$ fovea, $5 - 12 \leftrightarrow$ peripheral part). The striking aspect of this map is that about half of the neural mass is devoted to representation of the fovea and the region just around it, which has the greatest visual acuity.

terministic algorithms in numerous object vision tasks (for example, study Bartlett et al. (2001)). This constitutes the essence of my future research.

References

Bartlett, M. S., Braathen, B., Littlewort-Ford, G., Hershey, J., Fasel, I., Marks, T., Smith, E., Sejnowski, T. J., Movellan, J. R., October 2001. Automatic analysis of spontaneous facial behavior: A final project report. Tech. rep., University of California.

Becker, S., 1995. Learning to recognize moving objects as a modell fitting problem. Advances in Neural Information Processing 7, 1–9.

Bernardino, A., Santos-Victor, J., Sandini, G., 2002. Foveated active tracking with redundant 2d motion parameters. Robotics and Autonomous Systems 39, 205–221.

Bienenstock, E., von der Malsburg, C., 1987. A neural network for invariant pattern recognition. Europhysics Letters 4, 121–126.

Boluda, J. A., Domingo, J., 2001. On the advantages of combining differential algorithms and log-polar vision for detection of self-motion from a mobile robot. Robotics and Autonomous Systems 37, 283–296.

Burel, G., Carel, D., 1994. Detection and localization of faces on digital images. Pattern Recognition Letters 15, 963–967.

Ekmann, P., 2001. Telling lies: Clues to deceit in the marketplace, politics, and marriage. New York:W.W. Norton, 3rd edition.

Ferrari, F., Nielsen, P., Sandini, G., 1995. Space variant imaging. Sensor Review 15 (2), 17–20.

Grossmann, A., Mortlett, J., 1985. Decomposition of functions into wavelets of constant shape, and related transforms, mathematics and physics, lecture on recent results. Singapure:World Scientic publishing.

Hyvarinen, A., 1999. Survey on independent component analysis. Neural Computing Surveys 2, 94–128.

Isard, M., 1998. Visual motion analysis by probabilistic propagation of conditional density. Ph.D. thesis, Oxford University.

Juday, R., Weiman, C., 1990. Tracking algorithms using log-polar mapped image coordinates. In: SPIE Int. Conf. on Intelligent Robots and Computer Vision VIII: Algorithms and Techniques. Vol. 1192. Philadelphia (PA), pp. 843–853.

Juell, P., Marsh, R., 1996. A hierachical neural network for human face detection. Pattern Recognition 29, 781–787.

Kalocsai, P., von der Malsburg, C., Horn, J., 2000. Face recognition by statistical analysis of feature detectors. Image and Vision Computing 18, 273–278.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., 1991. Principles of Neural Science, 3rd Edition. McGraw-Hill/Appleton and Lange.

Koh, L. H., Ranganath, S., Venkatesh, Y. V., 2002. An integrated automatic face detection and recognition system. Pattern Recognition , 1259–1273.

Lin, S. H., Kung, S. Y., Lin, L. J., 1997. Face recognition/detection by probabilistic decision-based neural network. IEEE Trans. Neural Networks 8, 114–132.

McGarry, D. P., Plantec, T. R., Kassel, M. B., N.F., Downs, J. H., 1997. Registration of functional magnetic resonance imagery using mutual information. Paper presented at teh SPIE Medical Imaging.

Movellan, J. R., 1997. Tutorial on principal component analysis.

Phillips, P. J., Martin, A., Wilson, C., Przybocki, M., February 2000. An introduction to evaluating biometric systems. Computer , 56–63.

Phillips, P. J., Wechsler, H., Huang, J., Rauss, P., 1998. The feret database and evaluation procedure for face recognition algorithms. Image and Vision Computing 16 (5), 295–306.

Pluim, J. P. W., Maintz, J. B. A., Viergever, M. A., 2000. Image registration by maximiation of combined mutual information and gradient information. IEEE Transactions on Medical Imaging 19, 809–814.

Propp, M., Samal, A., 1992. Artificial neural network architecture for human face detection. Intell. Eng. Systems Artificial Neural Networks 2, 535–540.

Roth, D., 1999. The snow learning architecture. Tech. rep., UIUC Computer Science Department.

Shams, L. B., Brady, M. J., Schaal, S., 2001. Graph matching vs mutual information maximization for object detection. Neural Networks 14, 345–354.

Smeraldi, F., Bigün, J., October 1998. Facial feature detection by saccadic exploration of the gabor decomposition. Proceedings of the 1998 International Conference on Image Processing 3, 163–167.

Smeraldi, F., Bigun, J., 2002. Retinal vision applied to facial feature detection and face authentcation. Pattern Recognition Letters 23, 463–475.

Vapnik, V., 1995. The nature of statistical learning theory. Springer-Verlag, New York.

Viola, P., 1995. Alignment by maximization of mutual information. Ph.D. thesis, MIT, Cambridge, unpublished PhD Thesis.

von der Malsburg, C., 1981. The correlation theory of brain function. Internal report 81-2, Dept. Neurobiology, Max-Planck-Institute for Biophysical Chemistry, P.O. Box 2841, Göttingen, Germany.

von der Malsburg, C., L. Shams, U. E., 1988. Pattern recognition by labeled graph matching. Neural Networks 1, 141–148.

von der Malsburg, C., Shams, L., Eysel, U., November 1998. Recognition of images from complex cell responses, papers presented at the Society for Neuroscience Meeting, Los Angeles, CA.

Wilson, S. W., 1983. On the retino-cortical mapping. Internal Journal of Man Machines Studies 18, 361–389.

Yang, M. H., Ahuja, N., Kriegman, D., 2000. Proceedings fourth ieee international conference on automatic face and gestue recognition.