

Mahsa Kiazadeh

**Assessing the Existence of Visual Clues
of Human Ovulation**



Universidade do Algarve

2019

Mahsa Kiazadeh

**Assessing the Existence of Visual Clues
of Human Ovulation**

Master Thesis in Informatics Engineering

Work done under the supervision of:

Prof. Hamid Reza Shahbazkia

Prof. Gabriela Gonçalves



Universidade do Algarve

2019

Statement of Originality

Assessing the Existence of Visual Clues of Human Ovulation

(Master Thesis in Informatics Engineering)

Statement of authorship: The work presented in this thesis is, to the best of my knowledge and belief, original, except as acknowledged in the text. The material has not been submitted, either in whole or in part, for a degree at this or any other university.

Candidate:

(Mahsa Kiazadeh)

Copyright© Mahsa Kiazadeh. A Universidade do Algarve tem o direito, perpetuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua copia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado credito ao autor e editor.

Abstract

Is the concealed human ovulation a myth? The author of this work tries to answer the above question by using a medium-size database of facial images specially created and tagged. Analyzing possible facial modifications during the mensal period is a formal tool to assess the veracity about the concealed ovulation. In normal view, the human ovulation remains concealed. In other words, there is no visible external sign of the mensal period in humans. These external signs are very much visible in many animals such as baboons, dogs or elephants. Some are visual (baboons) and others are biochemical (dogs). Insects use pheromones and other animals can use sounds to inform the partners of their fertility period.

The objective is not just to study the visual female ovulation signs but also to understand and explain automatic image processing methods which could be used to extract precise landmarks from the facial pictures. This could later be applied to the studies about the fluctuant asymmetry. The field of fluctuant asymmetry is a growing field in evolutionary biology but cannot be easily developed because of the necessary time to manually extract the landmarks.

In this work we have tried to see if any perceptible sign is present in human face during the ovulation and how we can detect formal changes, if any, in face appearance during the mensal period.

We have taken photography from 50 girls for 32 days. Each day we took many photos of each girl. At the end we chose a set of 30 photos per girl representing the whole mensal cycle. From these photos 600 were chosen to be manually tagged for verification issues. The photos were organized in a rating software to allow human raters to watch and choose the two best looking pictures for each girl. These results were then checked to highlight the relation between chosen photos and ovulation period in the cycle. Results were indicating that in fact there are some clues in the face of human which could eventually give a hint about their ovulation.

Later, different automatic landmark detection methods were applied to the pictures to highlight possible modifications in the face during the period. Although the precision of the tested methods, are far from being perfect, the comparison of these measurements to the state of art indexes of beauty shows a slight modification of the face towards a prettier face during the ovulation.

The automatic methods tested were Active Appearance Model (AAM), the neural deep learning and the regression trees. It was observed that for this kind of applications the best method was the regression trees.

Future work has to be conducted to firmly confirm these data, number of human raters should be augmented, and a proper learning data base should be developed to allow a learning process specific to this problematic. We also think that low level image processing will be necessary to achieve the final precision which could reveal more details of possible changes in human faces.

Keywords: Image processing, Facial landmark extraction, Learning based methods, Human ovulation, Fluctuant asymmetry.

Resumo

A ovulação no ser humano é, em geral, considerada “oculta”, ou seja, sem sinais exteriores. Mas a ovulação ou o período mensal é uma mudança hormonal extremamente importante que se repete em cada ciclo. Acreditar que esta mudança hormonal não tem nenhum sinal visível parece simplista. Estes sinais externos são muito visíveis em animais, como babuínos, cães ou elefantes. Alguns são visuais (babuínos) e outros são bioquímicos (cães). Insetos usam feromonas e outros animais podem usar sons para informar os parceiros do seu período de fertilidade. O ser humano tem vindo a esconder ou pelo menos camuflar sinais desses durante a evolução. As razões para esconder ou camuflar a ovulação no ser humano não são claras e não serão discutidos nesta dissertação.

Na primeira parte deste trabalho, a autora deste trabalho, depois de criar um base de dados de tamanho médio de imagens faciais e anotar as fotografias vai verificar se sinais de ovulação podem ser detetados por outras pessoas. Ou seja, se modificações que ‘as priori’ são invisíveis podem ser percebidas de maneira inconsciente pelo observador. Na segunda parte, a autora vai analisar as eventuais modificações faciais durante o período, de uma maneira formal, utilizando medidas faciais. Métodos automáticos de análise de imagem aplicados permitem obter os dados necessários.

Uma base de dados de imagens para efetuar este trabalho foi criado de raiz, uma vez que nenhuma base de dados existia na literatura. 50 raparigas aceitaram de participar na criação do base de dados. Durante 32 dias e diariamente, cada rapariga foi fotografada. Em cada sessão foi tirada várias fotos. As fotos foram depois apuradas para deixar só 30 fotos ao máximo, para cada rapariga. 600 fotos foram depois escolhidas para serem manualmente anotadas. Essas 600 fotos anotadas, definam a base de dados de verificação. Assim as medidas obtidas automaticamente podem ser verificadas comparando com a base de 600 fotos anotadas.

O objetivo deste trabalho não é apenas estudar os sinais visuais da ovulação feminina, mas também testar e explicar métodos de processamento automático de imagens que poderiam ser usados para extrair pontos de interesse, das imagens faciais. A automatização de extração dos pontos de interesse poderia mais tarde ser aplicado aos estudos sobre a assimetria flutuante. O campo da assimetria flutuante é um campo crescente na biologia evolucionária, mas não pode

ser desenvolvido facilmente. O tempo necessário para extrair referências e pontos de interesse é proibitivo. Por além disso, estudos de assimetria flutuante, muitas vezes, baseado numa só fotografia pode vier a ser inválido, se modificações faciais temporárias existirem. Modificações temporárias, tipo durante o período mensal, revela que estudos fenotípicos baseados numa só fotografia não pode constituir uma base viável para estabelecer ligas genótipo-fenótipo.

Para tentar ver se algum sinal perceptível está presente no rosto humano durante a ovulação, as fotos foram organizadas num software de apresentação para permitir o observador humano escolher duas fotos (as mais atraentes) de cada rapariga. Estes resultados foram então analisados para destacar a relação entre as fotos escolhidas e o período de ovulação no ciclo mensal. Os resultados sugeriam que, de facto, existem algumas indicações no rosto que poderiam eventualmente dar informações sobre o período de ovulação. Os observadores escolheram como mais atraente de cada rapariga, aquelas que tinham sido tiradas nos dias imediatos antes ou depois da ovulação. Ou seja, foi claramente estabelecido que a mesma rapariga parecia mais atraente durante os dias próximos da data da ovulação. O software também permite recolher dados sobre o observador para análise posterior de comportamento dos observadores perante as fotografias. Os dados dos observadores podem dar indicações sobre as razões da ovulação escondida que foi desenvolvida durante a evolução.

A seguir, diferentes métodos automáticos de deteção de pontos de interesse foram aplicados às imagens para detetar o tipo de modificações no rosto durante o período. A precisão dos métodos testados, apesar de não ser perfeita, permite observar algumas relações entre as modificações e os índices de atratividade.

Os métodos automáticos testados foram Active Appearance Model (AAM), Convolutional Neural Networks (CNN) e árvores de regressão (Dlib-Rt). AAM e CNN foram implementados em Python utilizando o módulo Keras library. Dlib-Rt foi implementado em C++ utilizando OpenCv. Os métodos utilizados, estão todos baseados em aprendizagem e sacrificam a precisão. Comparando os resultados dos métodos automáticos com os resultados manualmente obtidos, indicaram que os métodos baseados em aprendizagem podem não ter a precisão necessária para estudos em simetria flutuante ou para estudos de modificação faciais finas. Apesar de falta de precisão, observou-se que, para este tipo de aplicação, o melhor método (entre os testados) foi as árvores de regressão.

Os dados e medidas obtidas, constituíram uma base de dados com a data de período, medidas faciais, dados sociais e dados de atratividade que poderem ser utilizados para trabalhos posteriores.

O trabalho futuro tem de ser conduzido para confirmar firmemente estes dados, o número de avaliadores humanos deve ser aumentado, e uma base de dados de aprendizagem adequada deve ser desenvolvida para permitir a definição de um processo de aprendizagem específico para esta problemática. Também foi observado que o processamento de imagens de baixo nível será necessário para alcançar a precisão final que poderia revelar detalhes finos de mudanças em rostos humanos. Transcrever os dados e medidas para o índice de atratividade e aplicar métodos de data-mining pode revelar exatamente quais são as modificações implicadas durante o período mensal. A autora também prevê a utilização de uma câmara fotográfica tipo true-depth permite obter os dados de profundidade e volume que podem afinar os estudos. Os dados de pigmentação da pele e textura da mesma também devem ser considerados para obter e observar todos tipos de modificação facial durante o período mensal.

Os dados também devem separar raparigas com métodos químicos de contraceção, uma vez que estes métodos podem interferir com os níveis hormonais e introduzir erros de apreciação.

Por fim o mesmo estudo poderia ser efetuado nos homens, uma vez que homens não sofrem de mudanças hormonais, a aparição de qualquer modificação facial repetível pode indicar existência de fatos camuflados.

Termos chave: Processamento de imagem, Extração do marco facial, Métodos baseados na aprendizagem, Ovulação humana, Assimetria flutuante.

Table of Contents

Statement of Originality	II
Abstract	III
Resumo	V
1 Introduction	1
1.1 Motivation.....	1
1.2 Objective.....	2
1.3 Methodology.....	3
2 Background	3
2.1 Fluctuant Asymmetry overview.....	3
2.2 Concealed ovulation.....	6
2.3 Facial analysis and Landmark extraction.....	8
2.4 Expression analysis.....	10
3 Literature survey of localization techniques	11
3.1 Component based Deformable Mode.....	11
3.2 Cascade Deformable Shape model.....	12
3.3 Constrained Local Neural Field model (CLNF).....	13
3.4 Deep Convolutional Network Cascade.....	13
3.5 Coarse-to-fine Convolutional Network Cascade.....	14
3.6 Explicit Shape Regression.....	15
3.7 Constrained local models (CLM).....	15
3.8 Parameterized Kernel Principal Component Analysis (PKPCA).....	16
3.9 Multiple Kernel Learning.....	16
3.10 Semi - Supervised learning.....	17
3.11 Bayesian Approach.....	18
3.12 Conditional Regression Forests.....	18

3.13	Multi View.....	19
3.14	Combinatorial Search and Shape Regression.....	20
3.15	Tree-structured models.....	20
3.16	Supervised Descent Method.....	21
4	Methods and procedures	22
4.1	Introduction.....	22
4.2	Data collection.....	22
4.3	Human rating.....	22
4.4	Implementations.....	25
4.4.1	Active appearance models.....	25
4.4.2	Convolutional neural networks and deep learning.....	28
4.4.2.1	Architecture Overview.....	28
4.4.2.2	Learning process.....	31
4.4.2.3	Data preparation.....	34
4.4.2.4	Training.....	35
4.4.3	Regression trees using the Dlib library.....	37
4.5	Ground truth Hand marking of the data set.....	40
5	Results and discussion	42
5.1	Rating results.....	42
5.2	Automatic landmark extraction results.....	49
5.3	The AAM results.....	50
5.4	The CNN results.....	51
5.5	The Dlib results.....	53
5.6	Comparing the results between methods.....	56
5.7	Alternative approach to analysis of results.....	58
6	Conclusions	61
7	References	62

1 Introduction

1.1 Motivation

A living organism develops and creates a particular form following its genetic program. The type of the genetic program is of course different in each individual, not only between species but also within the species. This particular program is called the genotype. During the development of the fetus, the pre-programmed genetic path manages the form that the individual will have in its final form and this form is called the phenotype.

Following this definition and the reality, one can suppose that each phenotype represents a special genotype and therefore by analyzing only the forms of individual, genetic information and some clues about the genotype can be retrieved.

Going a bit deeper in this issue and following the evolutionary biology about the fetus and pre-puberty development, this genotype needs to defend the organism development against all environmental stress and to execute the program the best possible. For example, development of a pair of organs which normally have to be symmetrical and be distributed identically on either side of the body should be exactly the same. These organs such as eyes, ears, hands, legs, etc. appear in a symmetrical position. The capacity of the genotype to create this symmetry in the final form of the individuals is believed by biologists to be a sign of quality. In other word, a better symmetry means a better genotype. It is therefore normal that a lot of work and studies concentrate on understanding the links between this symmetry in the phenotype and the genotype. Possible links between facial asymmetry and predisposition to some genetic alterations or diseases is an important issue of such studies. A very large interval of research is possible. Such as sexual attraction passing by genetic illness, or mental disorder possible links with visual facial clues, etc.

Nevertheless, a bottleneck of this field of study is the time necessary for the researcher to pinpoint the landmarks on a face and to calculate manually the possible ratios and symmetry on each single picture. This time-consuming process of manual calculation and analysis have not yet allowed large scale studies in this domain and all possible studies done so far remain on a small number of cases (mostly less than 100). It is therefore obvious that an automatic tool for

extraction of facial landmarks and precise processing of visual clues is an absolute necessity to solve the main bottleneck in this field of study. If this tool is developed, then a full-scale analysis becomes possible and most of the theories in this field can either be confirmed or rejected in a robust manner.

Now what this facial asymmetry has to do with the topic of this work? Well, the problem of manual analysis is only one aspect that prohibits the verification of different theories on this field. Another point of discussion between some communities lays on the fact that most evolutionary biologists consider that the final form of an individual is achieved after the puberty and this form does not change in a short period of time. Therefore, most of the studies over the face asymmetry are done by working only over one picture of the individual. The proponents of this work, nevertheless, think that this is not the case. Studying one picture of an individual taken in a particular moment can introduce a lot of bias into the final study of the facial signs and their links to other abnormalities.

1.2 Objective

As everyone knows, on some days people look better and, on some days, they look worse. Taking one picture of this individual for analyzing its genotype where we all know that the study is based over very little differences seems a bit unexplainable. This single picture is not really the best way to study possible life term links between phenotype and genotype. To demonstrate and prove that individual faces can show minor differences even in short term, the proponents of this study have searched a special case of physiological changes in human body in short terms that may affect the facial form.

The best and straight forward case in human of a deep physiological change in a short period of time is the mensal cycle in women. This fix periodical physiological change modifies deeply the biochemical composition in the body by modification of different levels of hormones. The fact of the periodicity of these phenomena is also a positive point because it allows the total repeatability of the study whenever necessary. If a facial modification during the mensal period is proved and formally calculated, then it is clear that most studies over links between facial symmetry and genotype abnormalities are to be repeated as they are only based on single photography.

This in the beginning was the main motivation of this study. Nevertheless, when the study was initiated, and the idea clarified a much bigger domain is to be explored. This fact has now made this study a full possible multidisciplinary research subject which can have repercussions in basic beliefs over human behavior and well-established issues such as concealed ovulation. If it is proved that external signs of ovulation do exist in each mensal cycle, and then many other questions will be open such as are they for all? Who detects best? Is it in every cycle or only in some?

1.3 Methodology

To be able to manage this study in a statically correct manner we need of course to solve the manual analysis bottleneck and develop an automatic manner of measuring facial landmarks. The extraction of the measurements automatically will allow a swift path towards large scale study and eliminate all bias. The objective of the proposed work is to be divided into two main streams:

- To develop a simple program to allow the ratters to choose in a short time best pictures of each woman during the mensal cycle. This will allow us to access the existence of possible visual points which could appear during the hormonal modification in the body.
- To use image processing techniques, mainly active appearance models to understand exactly what these changes are in an automatic manner. The mathematical model of the faces can then be studied to explain why in this short period of time the same individual looks more attractive.

In the following parts of this report a background over fluctuant asymmetry and concealed ovulation as well as facial automatic analysis will be exposed.

2 Background

2.1 Fluctuant Asymmetry overview

Fluctuating asymmetry is a particular form of biological asymmetry, characterized by small random deviations from perfect symmetry. The fundamental basis for the study of fluctuating asymmetry is an a priori expectation that symmetry is the ideal state of bilaterally paired traits.

Fluctuating asymmetry measures deviations from the ideal state of symmetry and is therefore thought to reflect the level of genetic and environmental stress experienced by individuals or populations during development. It has attracted a great deal of attention because bilaterally symmetrical traits are extremely common in nature and because the measurement of fluctuating asymmetry appears to represent a relatively simple method of assessing biologically important stress at the individual and population levels.

Asymmetry of an individual is measured as the right minus the left value of the bilaterally paired trait. By studying the distribution of these asymmetries at the population level, we can distinguish between three types of biological asymmetry: fluctuating asymmetry, directional asymmetry, and symmetry. Fluctuating asymmetry is characterized by small random deviations from perfect bilateral symmetry. These small random deviations result in a normal or leptokurtic distribution of asymmetry around a mean of zero. Directional asymmetry is characterized by a symmetry distribution that is not centered around zero but is biased significantly, towards larger traits either on the left or the right side.

Anti-symmetry is characterized by being centered around a mean of zero; however, symmetric individuals are rarer than those seen in fluctuating asymmetry distributions, such that the distribution is platykurtic or, in the extreme, bimodal. Directional symmetry and antisymmetric are developmentally controlled and therefore likely to have adaptive significance. Fluctuating asymmetry, on the other hand, is not likely to be adaptive as symmetry is expected to be the ideal state [1][2] although subtracting the measurement of the right side of a trait from that of the left side forms the basis of the analysis, accurately quantifying fluctuating asymmetry is not simple.

The measurement of fluctuating asymmetry is complicated by the fact that its magnitude and distribution are the same as the magnitude and distribution of measurement error. Therefore, in order to establish that real differences in symmetry rather than just measurement error are being reported, it is imperative to establish that the measures of fluctuating asymmetry

explain a statistically significant proportion of the observed total variance between the sides. To achieve this, it is necessary to make repeated measures of the left and right sides of the trait. The repeated measures need to be made on the same subjects in ignorance of the initial measure, with the same equipment and under the same laboratory or field circumstances as those of the main data. Furthermore, to eliminate observer bias, ideally all measurements should be made in ignorance of the measurement recorded for the side's pair.

For analysis of a potential fluctuating asymmetry data set, certain criteria must be met: the measurements must represent actual deviations from symmetry and not measurement error, and the distribution of fluctuating asymmetry must conform to that expected for it, rather than for directional asymmetry or anti symmetry. Various procedures have been recommended for this stage of the analysis, the most widely used being a mixed model analysis of variance (ANOVA), which will determine whether fluctuating asymmetry is significantly different from measurement error, and whether the asymmetry distribution has a mean of zero [1]. The mixed model ANOVA does not reveal significant departures in the fluctuating asymmetry distribution towards platy kurtosis, characteristic of anti-symmetry, and therefore the asymmetry distribution should also be described statistically in terms of its kurtosis and assessed visually.

Fluctuating Asymmetry trait, data analysis should proceed by testing first for the relation between trait size and fluctuating asymmetry. If there is a relationship, this needs to be accounted for, and methods exist for controlling for size dependence [1]. Data can be analyzed between groups, treatments or populations by using methods for comparing trait variances. Relationships between individual fluctuating asymmetry and continuous variables such as fitness measures need to be conducted on absolute fluctuating asymmetry values, sometimes called 'unsigned fluctuating asymmetry'. Caution should be exercised when analyzing unsigned fluctuating asymmetry, as it has a half-normal distribution, which violates the assumptions of most parametric statistics. A more accurate assessment of developmental stability can be obtained by pooling the fluctuating asymmetry measures of several traits per individual. However, as different traits may have different selection pressures on their symmetry, the traits that are being pooled need to be chosen carefully. It is a common observation that there is at best a very weak relationship between fluctuating asymmetries measured from two different traits for the same individuals.

2.2 Concealed ovulation

Concealed ovulation or hidden estrus in a species is the lack of any perceptible change in an adult female (for instance, a change in appearance or scent) when she is "in heat" and near ovulation. Some examples of such changes are swelling and redness of the genitalia in baboons and bonobos *Pan paniscus*, and pheromone release in the feline family. In contrast, the females of humans and a few other species [3] have few external signs of fecundity, making it difficult for the male to consciously deduce, by means of external signs only, whether or not a female is near ovulation.

While women can be taught to recognize their own level of fertility (fertility awareness), whether men can detect fertility in women is highly debated. Several small studies have found that fertile women (compared to women in infertile portions of the menstrual cycle or using hormonal contraception) appear more attractive to men [4][5]. It has also been suggested that a woman's voice may become more attractive to men during this time [6]. Two small studies of monogamous human couples found that women-initiated sex significantly more frequently when fertile, but male-initiated sex occurred at a constant rate, without regard to the woman's phase of menstrual cycle [7]. It may be that a woman's awareness of men's courtship signals [8] increases during her highly fertile phase due to an enhanced olfactory awareness of chemicals specifically found in men's body odor [9][10]. Analyses of data provided by the post-1998 U.S. Demographic and Health Surveys found no variation in the occurrence of coitus in the menstrual phases (except during menstruation itself) [11]. This is contrary to other studies, which have found female sexual desire and extra-pair copulations ("EPC's") to increase during the mid-follicular to ovulatory phases (that is, the highly fertile phase) [12]. These findings of differences in woman-initiated versus man-initiated sex are likely caused by the woman's subconscious awareness of her ovulation cycle (because of hormone changes causing her to feel increased sexual desire), contrasting with the man's inability to detect ovulation because of its being "hidden".

In 2008, researchers announced the discovery in human semen of hormones usually found in ovulating women. They theorized that follicle stimulating hormone, luteinizing hormone, and estradiol may encourage ovulation in women exposed to semen. These hormones are not found in the semen of chimpanzees, suggesting this phenomenon may be a human male

counterstrategy to concealed ovulation in human females. Other researchers are skeptical that the low levels of hormones found in semen could have any effect on ovulation [13]. One group of authors has theorized that concealed ovulation and menstruation were key factors in the development of symbolic culture in early human society [14][15].



Figure 2.1: Image pairs of two women as examples of stimuli used: (a) is from Prague; and (b) is from Newcastle. One image in each pair was taken during the follicular phase (i) and one in the luteal phase (ii) of the cycle (in both cases, days 12 and 19, respectively)

Human differs from other mammals in their sexual behavior [1]. One of the most intriguing fact in this behavior is the concealed ovulation (The ovulation without any external visible signs). Many theories over the how and why of human move out from an advertised oestrus cycle to a concealed menstruation cycle has been developed [2] and for years it was conventionally accepted that the oestrus was either lost or completely concealed. Nevertheless, the idea of a completely hidden and sign-less period of fertility sounds contradictory to normal rules of reproductively and mate selection. Some recent studies using fine-tuned phenotype measurements have revealed that some signs appear at the fertile point of the cycle to make the female more attractive to male. Studies over body scent, voice frequency. Soft tissue symmetry, body ratio, creativity and fluidity in speech and facial attractiveness were assessed and studied as possible modifications that alerts a male over the fecundity of the female. Some other studies

over the number of tips gained by lap dancers in different period of their mensal cycle shows that the dancers earned much more during the oestrus or just before ovulation.

As one can see through these studies that hearing, smell and sight could all be used to receive fertility signs, but the human society defines attractiveness solely by means of sight. Therefore, if signs of oestrus do exist between human, they should mostly be present in visual cues. In this paper we focus on facial attractiveness during the cycle. The previous study on the subject have revealed possible signs but in a very limited proportion, Roberts & al used only 2 photos of women one in follicular phase and one in lethal phase to distinguish between possible visual cues. The results obtained by Roberts & al were just a bit over the random of 0.5 expected. The authors recognize also that their study could only suggest a fluctuant facial attractiveness during the mensal cycle.

2.3 Facial analysis and Landmark extraction

A facial biometric system is a computer application for automatically identifying or checking a person from a numeric image or from a video frame from a video source. One approach to do so, is comparing selected facial features from the image and a facial database. It is typically used in security systems and can be compared to other biometrics such as fingerprint or iris recognition systems. Some facial recognition algorithms identify facial features by extracting landmarks, or features, from an image of the subject's face. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features are then used to search for other images with matching features.

Other algorithms after normalizing a data base of images keep only the data necessary for the process of recognition and save them. A probe image is then compared with the face data. One of the earliest successful systems [18] is based on template matching techniques applied to a set of salient facial features, providing a sort of compressed face representation. Recognition algorithms can be divided into two main approaches, geometric, which look at distinguishing features, or photometric, which is a statistical approach that transforms an image into values and compares the values with templates. Popular recognition algorithms include Principal Component Analysis using eigenfaces, Linear Discriminate Analysis, Elastic Bunch Graph Matching using the Fisher-face algorithm, the Hidden Markov model, the Multilinear Subspace

Learning using tensor representation, and the neuronal motivated dynamic link matching.

We will be referring often in this document to the landmarks. A landmark is a recognizable natural or man-made feature that stands out from its near environment and is often visible from long distances. Facial landmarks are defined as key-points on the face with subsequent impact on the target task, like animation, face recognition, gaze detection, face tracking, expression recognition, gesture understanding etc. Facial landmarks are a prominent feature that can play a discriminative role or can serve as anchor points on a face graph. Facial landmarks such as the nose tip, eyes corners, chin, mouth corners, nostril corners, eyebrow arcs, ear lobes, are the common choices. For ease of analysis most landmark detection algorithm prefers an entire facial semantic region, such as the whole region of a mouth, the region of the nose, eyes, eyebrows, cheek or chin. The facial landmarks are classified in two groups, primary and secondary, or fiducial and ancillary. This distinction is based on reliability of image features detection techniques. For example, the corners of the mouth, of the eyes, the nose tips and eyebrows can be detected relatively easily by using low level image features, e.g. SIFT, HOG. The directly detected landmarks are referred as fiducial. The fiducial group of landmarks play a more prominent role in facial identity and face tracking. The search for secondary landmarks is guided by primary landmarks. The secondary landmarks are chin, cheek contours, eyebrow and lips midpoints, non-extremity points, nostrils. It takes more prominent role in facial expression [19].

Despite the conceptual simplicity of facial landmarks detection, in computer vision there are some challenges. The emerging applications like surveillance system, gesture recognition requires that landmark localization algorithms should run in real time parallel with the computational power of an embedded system, such as intelligent cameras. Such type of application requires a more robust algorithms against a confounding factor such as illumination effects, expression and out of plane pose. There are four main challenges in localizing facial landmarks [19]:

- Variability: Landmark appearances differ due to extrinsic factors such as partial occlusion, pose, illumination, camera resolution and expression, also due to intrinsic factors such as face variability between individuals. Facial landmarks can sometimes be only partially observed due to hand movements or self-occlusion due

to extensive head rotations or occlusions of hair. Also, facial landmark detections are difficult because of illumination artifacts and facial expressions. A facial landmark localization algorithm that delivers the target points in a time in an efficient manner and works well across all intrinsic variations of faces has not yet been feasible.

- **Accuracy and number of landmarks require:** Based on the intended application the number of landmarks and its accuracy varies. For example, in face recognition or in face detection tasks, primary landmarks like two mouth corner, four eyes corner and nose tips may be adequate. On the other hand, higher level tasks face animation or facial expression understanding require greater number of landmarks e.g. from 20-30 to 60 - 80 with higher accuracy. Fiducial landmarks are needed to be determined with more accuracy because they often guide the search of secondary landmarks.
- **Lack of globally accepted and error free dataset:** Most of the dataset provides annotations with different markups and accuracy of their fiducial point is questionable. The accuracy of landmark localization algorithm is largely depending on the data set used for training. Each algorithm uses different dataset to train and evaluate the performance, so it is difficult to compare algorithms.
- **Acquisition conditions:** Acquisition conditions, such as resolution, background clutter, and illumination can affect the landmark localization performance. The landmark localizers trained in one database have usually inferior performance when tested on another database. It is to mention that these challenges, are general case study challenges. In the particular case in scope of the present work we will not be dealing with variability or acquisition conditions because they were fixed in lab conditions and the external factors were kept as equal as possible between images. Nevertheless, we were not able to keep the exact lighting or contrast conditions.

2.4 Expression analysis

A facial expression is a visible manifestation of the affective state, cognitive activity,

intention, personality, and psychopathology of a person [20], it plays a communicative role in interpersonal relations. Facial expressions, and other gestures, convey non-verbal communication cues in face-to-face interactions. These cues may also complement speech by helping the listener to elicit the intended meaning of spoken words. As cited in [21], Mehrabian reported that facial expressions have a considerable effect on a listening interlocutor, the facial expression of a speaker accounts for about 55 percent of the effect, 38 percent of the latter is conveyed by voice intonation and 7 percent by the spoken words.

As a consequence of the information that they carry, facial expressions can play an important role wherever humans interact with machines. Automatic recognition of facial expressions may act as a component of natural human- machine interfaces [22] (some variants of which are called perceptual interfaces [23] or conversational [24] interfaces). Such interfaces would enable the automated provision of services that require a good appreciation of the emotional state of the service user, as would be the case in transactions that involve negotiation, for example. Some robots can also benefit from the ability to recognize expressions [25].

Automated analysis of facial expressions for behavioral science or medicine is another possible application domain [26][27]. From the viewpoint of automatic recognition, a facial expression can be considered to consist of deformations [28] of facial components and their spatial relations, or changes in the pigmentation of the face. Research into automatic recognition of facial expressions addresses the problems surrounding the representation and categorization of static or dynamic characteristics of these deformations or face pigmentation. Further details on the problem space for facial expression analysis are given in [29]. We will continue with a survey of some of the recent landmark localization techniques [19]

3 Literature survey of localization techniques

3.1 Component based Deformable Mode

Component based deformable model presented for generalized face alignment at 2007 [30], they used a novel bi-stage statistical framework to account for both local and global shape characteristics. It uses separate Gaussian models for shape components instead of using

statistical analysis on the entire shape which preserve more detailed local shape deformations. Each model of components used the Markov Network for search strategy. They used Gaussian Process Latent Variable Model to give control of full range shape variations, hence it makes out better description of the nonlinear interrelationships over shape components. This approach allows the system to preserve the full range of low frequency shape variations, and also the high frequency local deformations caused by exaggerated expression.



Figure 3.1: Output of component based deformable model – (Database: YALE FACE DATABASE B, Key point detected: 79 Image / video: Images only)

3.2 Cascade Deformable Shape model

Xiang, et al, [31] presents a two-stage cascaded deformable shape model to effectively and efficiently localize facial landmarks with large head pose variations. For face detection, they proposed a group sparse learning method to automatically select the most salient facial landmarks. 3D face shape model detects pose free facial landmark initialization. The deformation executes in two stages, the first step uses mean-shift local search with constrained local model to rapidly approach the global optimum. Second step uses component-wise active contours to discriminatively refine the subtle shape variation. It improves performance over CLM and multi-ASM in face landmark detection and tracking.



Figure 3.2: Output of cascade deformable shape model – (Database: MultiPIE, AR, LFPW, LFW and AFW, Talking Face Video, Key point detected: 65, Image / video: Images and video both)

3.3 Constrained Local Neural Field model (CLNF)

Baltrusaitis and et al, [32] presented the Constrained Local Neural Field model for facial landmark detection. This model uses probabilistic patch expert (landmark detector) that learn non-linear and spatial relationships between the input pixels and the probability of a landmark being aligned and Non-uniform Regularized Landmark Mean-Shift (NRLM) optimization technique, which takes the account the reliabilities of each patch expert leading to better accuracy.



Figure 3.3: Output of constrained local neural field model – (Database: LFPW and Helen, Key point detected: 65, Image / video: Images only)

3.4 Deep Convolutional Network Cascade

Yi Sun [33] presented three level deep convolutional networks cascade model. The output of multiple networks at each level is fused to give robust and accurate estimation. At first level, convolutional networks extracts high level features over whole face region. It has two advantages first, the texture context information over the entire face is utilized to locate key point. Second, the network is trained to predict all the key-points simultaneously, geometric

constraints among key-points are encoded. Therefore, method avoids local minimum caused by ambiguity and data corruption. In next two level the initially predicted key-points are finely-tune to achieve high accuracy. This method is not good for locating facial landmarks in large number.



Figure 3.4: Output of deep convolutional network cascade – (Database: LFW,LFPW and BioID Key point

3.5 Coarse-to-fine Convolutional Network Cascade

Zhou, et al [34] presented four level convolutional network cascades, which tackles the problem in coarse-to-fine manner. Each network level refines a subset of facial landmarks generated in previous network levels and predicts explicit geometric constraints like the position and rotation angles of specific facial component to rectify inputs of the current network level. The first level estimate bounding boxes for inner points and contour points separately. Second level predicts an initial estimation of the positions for the inner points which are refined in third level for each component. The last layer improves the predication of mouth and eyes by taking rotated image patch as new input. Two levels of separate networks are used for contour points. This method is good for locating facial landmarks in huge numbers (above 50).



Figure 3.5: Output of coarse-to-fine convolutional network cascade model – (Database: 300-W (300 faces in wild) Key point detected: 68 Image / video: Images Only)

3.6 Explicit Shape Regression

Xudong Cao [35] presented very efficient and highly accurate Explicit Shape Regression approach. This model directly learns a vectorial regression function to infer the whole facial shape and explicitly minimize localization error in training data. The inherent shape encoded into the regressor in learning framework applied from coarse to fine to during testing. This approach uses two level boosted regression, a correlation-based feature and shape-indexed features selection method hence regression is more effective. It shows highly accurate and efficient results. Regression process is extremely fast in test 15 MS for 87 landmarks shape.



Figure 3.6: Output of explicit shape regression model – (Database: BioID, LFPW and LFW87 Key point detected: 5,29 or 87 Image / video: Images Only)

3.7 Constrained local models (CLM)

This model [36] is developed on Active Appearance Model (AAM). CLM differs from AAM because it is not a generative model for the whole face, instead it produces landmark templates iteratively and use a shape constrained search technique. The position vectors of the landmark's templates are estimated using Bayesian formulation. The posterior distribution in Bayesian formula incorporates both the image information via template matching scores and the statistical shape information. Therefore, positions of new landmarks are predicted in the joint shape model and the light of the image, and then templates are updated by sampling from the training images.



Figure 3.7: Output of constrain local model – (Database: BioID and XM2VTS, Key point detected: 22 Image / video: Images Only)

3.8 Parameterized Kernel Principal Component Analysis (PKPCA)

Torre, et al [37] presented the parameterized kernel principle component analysis (PKPCA) model by extending KPCA to incorporate geometric transformation into formulation and applying gradient descent algorithm for fast alignment. This model differs from PCA because it can model non-linear structure in data in variant to rigid or non-rigid deformations. It does not require manually labeled training data.

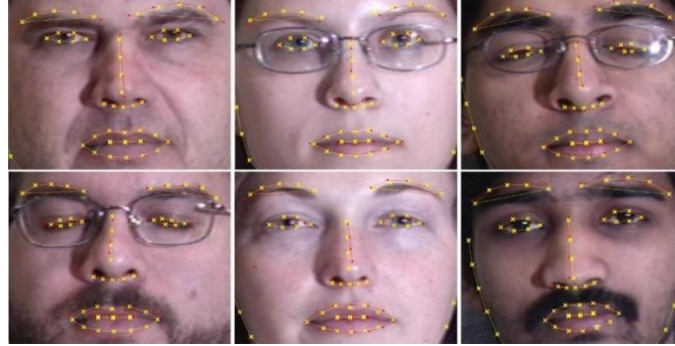


Figure 3.8: Output of PKPCA model – (Database: CMU Multi-PIE, Key point detected: 46, Image / video: Images Only)

3.9 Multiple Kernel Learning

Rapp, et al, in [38] presented multiple kernel, it uses two patches on the face. One covers the eye region and the other covering the mouth region. In testing, pixels of a respective regions should be part of a target landmark, using the multi-resolution windows (progressively smaller nested windows) texture data is extracted that captures information from global to local view range. Every resolution level, information is passed into a different kernel and the convex combination of these kernels, instead of concatenating pyramidal information. For each

resolution level there is dedicated kernel which forms a multi-kernel SVM. The multi-kernel SVM is trained using center-surround architecture and the surround windows forming the negative examples. After the discovery of landmark points, without having any spatial relationship among them initially. To reach the plausible shapes, point distribution models are invoked. The point distribution models are particularly focusing to the mouth and to the eye-eyebrow pairs. The shape alternatives are evaluated using Gaussian mixture models (GMMs), so that the point combinations that possess the highest sum of SVM scores and that fit best to the learned models are selected.



Figure 3.9: Output of multiple kernel learning – (Database: Cohn-Kanade and Pose, Illumination and Expression (PIE), Key point detected: 17, Image / video: Images only)

3.10 Semi - Supervised learning

Tong, et al [39] address to the often imperfect and tedious task of manual landmark labeling, and to overcome this suggest a scheme to partly automate land-marking. In their method, a small percentage (e.g., 3%) of faces needs to be hand labeled, while most the faces are automatically marked. This is done by propagating the few exemplars landmarking information to the whole set. On the minimization of the pair wise pixel differences resulting in two error terms, the learning is based. The penalty in one term makes the warping of each un-marked image toward all other un-marked images, irrespective of the content they become more alike. While penalty in the other term controls the warping of un-marked images toward marked images. The warping function itself can be a piecewise affine warp to model a nonrigid transformation or a global affine warp for the whole face.



Figure 3.10: Output of semi supervised learning – (Database: Notre Dame, Caltech 101 and FERE, Key point detected: 33, Image / video: Images only)

3.11 Bayesian Approach

Belhumeur in [40] presented a fully Bayesian approach to find landmark positions from local evidences. An interesting aspect about their work is that the local detector outputs are collected from a cohort of exemplars (sample faces with annotated landmarks), which provide non-parametrically the global model information. The local detector consists of a sliding window whose size is proportional to IOD and which collects SIFT features. In the next stage, the global detector models the configurational information of the ensemble of primary points. The joint probability of the location on ‘n’ landmarks and the vector of their local detectors outputs is maximized. This method surpasses in accuracy the performance of the manual landmarking in most of the 29 landmarks considered.



Figure 3.11: Output of Bayesian approach – (Database: BioID and LFPW, Key point detected: 29, Image / video: Images only)

3.12 Conditional Regression Forests

Dantone in [41] proposed pose-dependent landmark localization scheme that is achieved by conditional random forests. Conditional regression forests learn several conditional probabilities over the parameter space, and deals with facial variations in appearance and shape, while regression forests try to learn the probability over the parameter space from all face

images in the training set. The head pose is quantized into five segments of i) right profile ii) right iii) front iv) left and v) left profile faces and specific random forests are trained. Both texture and 2D displacement vectors that are defined from the centroid of each patch to the remaining ones described the local properties of a patch. Texture is described by normalized gray values in order to cope with illumination changes in addition to Gabor filter responses. Training of conditional random forests is very similar to random forests, except that the probability of assigning a patch to a class is conditioned on the given head pose. This method locates landmarks in a query image at real-time speed.

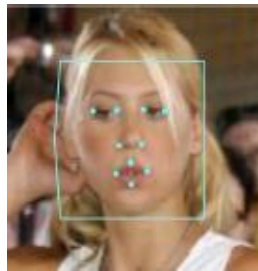


Figure 3.12: Output of conditional regression forests – (Database: LFW, Key point detected: 10, Image / video: Images Only)

3.13 Multi View

Zhanpeng Zhang, et al, in [42] presented a real time Multi-View facial landmark localization in RGB-D images. This system is able to estimate 3D head pose and 2D landmark localization. The model extracts random local binary patterns of different scales and estimate the facial parameters with hierarchical regression techniques. At first, 3D face positions and rotations are estimated via a random regression forest. Afterwards, 3D pose is refined by fusing the estimation from the RGB observation. The depth channel and RGB channel are used at different stages, the depth input is fed to the random forest for face detection and pose estimation at the beginning stage, the RGB input is fed to gradient boosted decision trees (GBDT) for head pose and hierarchical facial landmark location regression when the face is available. The pose estimation results from the depth and RGB inputs are weighted and combined to improve the precision.



Figure 3.13: Output of multi-view model – (Database: BIWI Kinect Head Pose, Key point detected: 13, Image / video: Images only)

3.14 Combinatorial Search and Shape Regression

Sukno in [43] have managed the combinatorial problem using RANSAC algorithm. First, they find out the reliable features using spin images as features and the missing are regressed using the multivariate Gaussian model encompassing all 3D landmark coordinates. The correct landmarks are sort out from the multitude of candidate points, all combinations of four points are used and RANSAC is used as the basis of the feature matching procedure. For the missing landmarks the median of the closest candidates is considered. The PCA instrumented shape fitting term for the detecting landmarks is has been used in the cost function consist of a part accounting for the reconstruction error, the other part accounts for the distance from the inferred landmarks to their closest candidates.



Figure 3.14: Output of combinatorial search and shape regression – (Database: FastSCAN, Key point detected: 8, Image / video: Images and video both)

3.15 Tree-structured models

Xiangxin Zhu, et al, in [44] address the three linked problems of face detection, face pose

estimation and facial landmarks localization. This algorithm is shape driven and local and global information are merged right from beginning. Since pose is part of estimation, the algorithm practically works as a multi-view algorithm. Multi-view implemented by considering several (30 to 60) local patches that are connected as a tree which collectively describe the landmark related region of the face.



Figure 3.15: Output of tree structured model – (Database: CMU MultiPIE and AFW, Key point detected: 61, Image / video: Images only)

3.16 Supervised Descent Method

Xuehan Xiong in [45] presented a Supervised Descent Method (SDM) for localizing facial landmarks. At first, it takes an image with manually labeled landmarks. Then it ran through training images to give initial configuration of landmarks. It uses SIFT function to extract initial landmarks. In training, SDM tries to minimize difference between manually labeled landmarks and initially located landmarks by SIFT (Δx). This method does not learn any shape or appearance model in advance from training data. The SDM learns a series of descent direction and re-scaling factors to produce a sequence of updates. SDM directly learns descent direction from training data by learning a linear regression between, (Δx) and difference of SIFT value of manually and extracted landmarks ($\Delta \Phi$). In testing, based on descent direction and re-scaling factors learn in training SDM estimates landmarks. SDM learns descent direction without computing Jacobian nor Hessian matrix, which are computationally expensive. SDM is fast and accurate.



Figure 3.16: Output of Supervised Descent Method – (Database: LFPW and LFW-A&C, RU-FACS, Key point detected: 49, Image / video: Images and video both)

4 Methods and procedures

4.1 Introduction

As the idea was to test and gather data over a mensal period and that there is no such database available, we had to start first by gathering the necessary data. The data needed to be collected over a whole period, and each day, the women who participated in the experience had to come to the lab to take a picture in particular setting.

Once the data was collected, we needed first to prove that there is a perceptible change in the visual aspects of the face. A simple analysis of the results could give a hint that a change occurs during the period and therefore, fluctuant asymmetry studies could not take place over a simple photo of one day. Then we needed to landmark all faces manually to create a ground truth data base for comparison with automatic methods. This process was very time consuming because all and every single picture had to be marked manually and then reviewed for a second time for correction. Finally, different automatic landmark extraction methods had to be implemented and tested over our own database.

4.2 Data collection

50 women were photographed for 32 days, about 10 photos each day. From this set of pictures, we have chosen one best photo for each day and for each woman. Then they are normalized as much as possible for light and contrast.

All women were chosen from Portuguese nationality. They were all aged between 20 and 35 and they were all university students. All data about their sexual orientation, partners and contraceptive methods was collected for further analysis. The data base is not yet publicly available as the study is ongoing and legal permissions are not yet obtained.

4.3 Human rating

Once the data was collected, we needed first to prove that there is a perceptible change in the visual aspects of the face. Therefore, human raters had to rate the pictures of the same

woman over her mensal period. To do this, we developed a simple application that provides a list of photos of each girl ordered in a form page. The rating app is developed using C# and Mysql. This is a window form application programmed by C# and one form per girl. Each form includes 15 check boxes, which is connected to the database in Mysql, a timer and conditional next bottom and skip bottom. The app runs as an EXE file and outputs results in an excel file.

The main goal is to allow the ratters to choose in a short time best pictures of each woman during the mensal cycle. This will allow us to assess the existence of possible visual points which could appear during the hormonal modification in the body. As the data base is not yet publicly available and legal permissions are not yet obtained, we developed a desktop windows application witch only works with local IP.

At the first page of the application we classified raters by 4 factors. Raters can introduce themselves by gender, age, education and marital status. Having 2 categories of males and females' raters allows us to make our database ready for comparing our results to other sexual attractiveness researches. Although most of our raters are university students or professors, regarding the fact that most of studies on attractiveness are based on age-controlled groups, we put 11 groups of ages between 15 and 75. In future work we will be able to find how much maturity affects the human opinion about attractiveness. Also, the last option is considered to determine if rater choices are influenced by marital status. The extraction of different classes of the ratters and the results they have obtained can also be a valuable data for future psychological work over who is really the receiver of the possible signs.

Each page includes 12 to 15 photos of one girl and each girl has 2 pages. All selected photos of each girl are placed in 2 pages randomly and each photo appears only one time. Therefore, the rater has a chance to see all photos of each girl taken over one month.

All photos are normalized as much as possible for light and contrast. But there is no preprocessing applied to change the skin color. All photos are face cropped to be in the same size and same head pose, to make the rater concentrate only on face attractions. Also, the app pages are very simple, neutral color and mostly filled by face photos, to avoid distractions.

In order to gather well-organized results, the rater can choose 2 photos in each page not

more and not less, and it is not possible for the rater to go to the next page without choosing but he can skip the app and save his rating. All photos on each page are connected to the original one in the data base containing the photo dates. Therefore, when the checkbox of a photo is marked by the rater in the platform, the photo and photo date get one 'plus' in database. At the end we have a list of data for each girl which contains photo name and photo dates. This list is ordered by number of times the photo was chosen.

Another feature of the app is a limited time for choosing. The rater has only 30 seconds on each page, to decide which photos are the most attractive ones. The idea behind the limited time is the fact that human brain makes up its mind up to ten seconds before actually realizing it [16].

Before starting the app, all raters are told that the goal is choosing the best photo of each girl in terms of attractiveness without considering any specific parameter, which means their first opinion is the best. That's why at first glance, the rater's brain would know what he is looking for. Therefor In the first few milliseconds of seeing each page, the rater realizes the best photo. But as long as the difference between photos in each page is very low, we gave 30 seconds to rostromedially prefrontal cortex of rater's brain go a bit deeper and evaluate the decision with all judgment parameters available in the brain [17].

The screenshot shows a web browser window with the title "<rating application>> supervised by Professor Hamidreza Shahbazi, Professor Gabriela Gonçalves". The address bar shows "Student No : a48655" and "Universidade do Algarve - FCT". The main content area has a heading "Please select your profile from the list below / Por favor, selecione o seu perfil a partir da lista abaixo". Below this, there are four dropdown menus for profile selection: "Gender / Género" (set to "Male"), "Age / Idade" (set to "30 - 35"), "Education / Educação" (set to "Undergraduate"), and "Marital Status / Estado Civil" (with a dropdown menu open showing options: "Single - Without fixed partner", "Single - With fixed partner", "Divorced", and "Widowed"). At the bottom, there is a button labeled "Next / Continuar".

Figure 4.1: rating application, first page, rater information

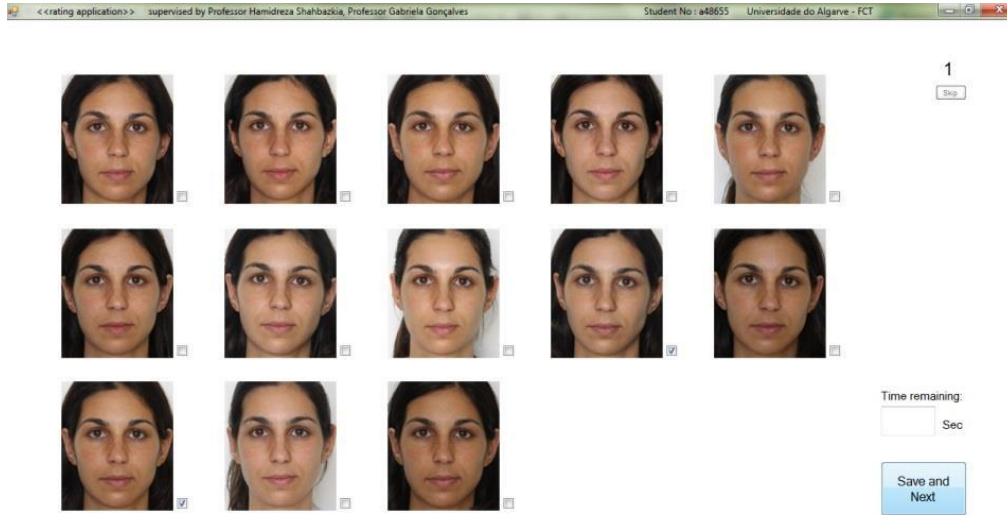


Figure 4.2: rating application, the example of girl's photos page

4.4 Implementations

After some verification of existing methods and the available results, we decided to implement and test 3 methods. The methods were chosen by the following criteria:

- Available and verifiable literature and previous results
- General towards specialized
- Estimated Precision

We, therefore, chose to test 3 methods:

- Active appearance Model -AAM
- Convolutional Neural Networks -CNN
- Regression Trees from Dlib library –Dlib

4.4.1 Active appearance models

Facial landmark point extraction is a key step in facial image representation and analysis. The Active Appearance Model (AAM) proposed by Cootes et al [46] is a powerful object description method that is commonly used for facial landmark points extraction AAM fitting is

a non-linear optimization problem. Different optimization approaches have been proposed to find the best model parameters that result in minimum error between the synthesized appearance models obtained from the AAM and the input image. In general, due to variation of camera view angle, resolution and focal distance, facial images have different scaling, rotation, and translations. In order to remove global shape variations, all shapes are normalized, and the modeling is only concerned with local shape deformation. Therefore, it is necessary to combine a global shape transformation with the normalized AAM. The global shape transformation is often a 2D similarity transformation. Finding optimal parameters of the global transformation improves the accuracy of fitting in representing novel facial images with different shape and pose variations. Traditionally, the stochastic gradient descent algorithm or iteratively incremental additive techniques are used to update the AAM parameters to fit onto novel images. The fitting problem can also be viewed as finding a model instance similar to the given facial image and therefore it can be considered as an image alignment problem.

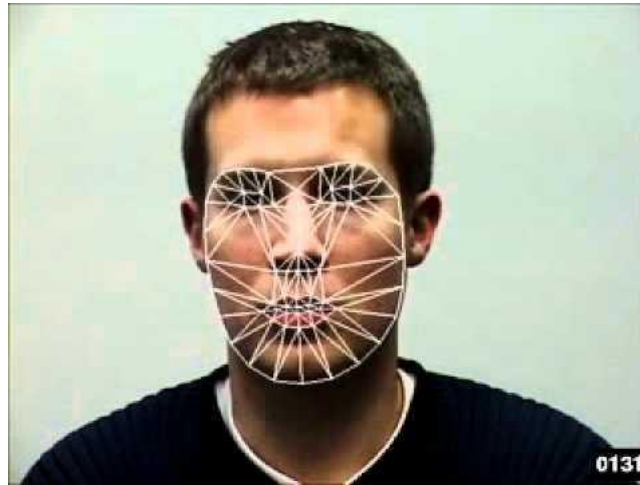


Figure 4.3: AAM fit on a given face

AAM consists of a shape component and an appearance component obtained from a set of annotated landmark points in training images. Let's assume we are given a training facial image set with annotated shapes defined as: $S = (X_1, Y_1, X_2, Y_2, \dots, X_v, Y_v)^T$. The training images are first normalized and aligned using iterative Procrustes analysis. This step removes variations due to a chosen global shape normalization transformation so that the resulting model can efficiently consider local and non-rigid shape deformation. We then can combine the resulting AAM with a global transformation. Afterwards, Principal Component Analysis

(PCA) is applied to the set of normalized training shapes and a shape model is defined as:

(1)

$$S = S_0 + \sum_{i=1}^n p_i s_i$$

where the base shape S_0 is the mean shape and the vectors S_i are n eigenvectors corresponding to the n largest eigenvalues. Then, all the training images are normalized by warping them into the base shape S_0 , using piecewise affine warp, and the appearance model is defined as:

(2)

$$A(x) = A_0(x) + \sum_{i=1}^m \lambda_i A_i(x) \quad \forall x \in s_0$$

where A_0 is the mean appearance and the vectors A_i are the m eigenvectors corresponding to the m largest eigenvalues. The goal of fitting is to find a model instance that can efficiently describe the object (e.g. face) in a given image. Thus, it can be considered as an image alignment problem. In other words, we want to find the model instance $M(W(x; p)) = A(x)$ as similar as the image $I(x)$. In general, facial images have different scaling, rotation, and translations. Therefore, it is necessary to combine a global shape transformation with the normalized AAM. If we consider the global shape transformation as $N(x; q)$, we want to minimize the error between the template and $I(N(W(x; p); q))$. Considering global shape transformation, the objective of the fitting process is to find p and q in order to minimize the error image as:

(3)

$$E(x) = \sum_{x \in s_0} [A_0(x) - I(N(W(x; p); q))]^2$$

Which is a non-linear least square problem. We can have different definitions for the global transformation $N(x; q)$. In [47], a set of 2D similarity transformations as a subset of piecewise affine warps is defined. This representation of $N(x; q)$ is similar to $W(x; p)$ and therefore similar analysis on the shape parameters p can be applied to q . If we assume that the

two sets of shape vectors S_i and S_i^* are orthogonal to each other, we can add the four 2D similarity vectors S_i^* to the beginning of AAM shape vectors S_i [47] and model any given shape as: $S = S_0 + \sum_{i=1}^{n+4} p_i s_i$ In practice, S_i and S_i^* are not quite orthogonal to each other.

This can either be ignored when the size of S_i is small or the complete set of S_i and S_i^* can be orthonormalized preferably.

In [48], Baker et al. relate AAM to the Lucas-Kanade algorithm. They proposed the Inverse Compositional Algorithm (ICA), in which they find shape variation on the template and compose the inverse of that with the current shape. Therefore, many computationally expensive tasks are recomputed. In [47], appearance variation is considered in the fitting by finding shape parameters in a linear subspace where the appearance variation is ignored and then “projected out” to the full space with respect to the appearance eigenvectors. The method is more generic compared with the ICA, but the fitting is not accurate when applied to subjects that are not similar to subjects in the training set. The “projecting out” approach is called PO in the rest of this paper. In [49] Simultaneously Inverse Compositional (SIC) method is introduced, which is more generic. In this method the fitting procedure minimizes the error between $[A_0(x) + \sum_{i=1}^m (\lambda_i + \Delta\lambda_i)A_i]$ and $I(N(W(x;p); q))$, where A_i are m appearance eigenvectors correspond to the m largest appearance eigenvalues, and $(\lambda_i + \Delta\lambda_i)$ are parameters of appearance that are found simultaneously with respect to the Δp . As the appearance parameters are optimized in each iteration, both steepest descent and the Hessian matrix (H) should be calculated in each iteration, and therefore the method is slower. In [49] the PO is compared with the SIC, and the SIC is reported more accurate in modeling unseen subjects.

4.4.2 Convolutional neural networks and deep learning

4.4.2.1 Architecture Overview

Recall: Regular Neural Nets: Neural Networks receive an input (a single vector) and transform it through a series of hidden layers. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer, and where neurons in a single layer function completely independently and do not share any connections. The last fully connected layer is called the “output layer” and in classification settings it represents the class scores.

Regular Neural Nets don’t scale well to full images. In CIFAR-10 dataset, images are only of size 32x32x3 (32 wide, 32 high, 3 color channels), so a single fully connected neuron in a

first hidden layer of a regular Neural Network would have $32 \times 32 \times 3 = 3072$ weights. This amount still seems manageable, but clearly this fully connected structure does not scale to larger images. For example, an image of more respectable size, e.g. $200 \times 200 \times 3$, would lead to neurons that have $200 \times 200 \times 3 = 120,000$ weights. Moreover, we would almost certainly want to have several such neurons, so the parameters would add up quickly! Clearly, this full connectivity is wasteful, and the huge number of parameters would quickly lead to overfitting.

3D volumes of neurons: Convolutional Neural Networks take advantage of the fact that the input consists of images and they constrain the architecture in a more sensible way. In particular, unlike a regular Neural Network, the layers of a ConvNet have neurons arranged in 3 dimensions: **width, height, depth**. (Note that the word *depth* here refers to the third dimension of an activation volume, not to the depth of a full Neural Network, which can refer to the total number of layers in a network.) For example, the input images in CIFAR-10 are an input volume of activations, and the volume has dimensions $32 \times 32 \times 3$ (width, height, depth respectively). As we will soon see, the neurons in a layer will only be connected to a small region of the layer before it, instead of all of the neurons in a fully connected manner. Moreover, the final output layer would for CIFAR-10 have dimensions $1 \times 1 \times 10$, because by the end of the ConvNet architecture we will reduce the full image into a single vector of class scores, arranged along the depth dimension. Here is a visualization:

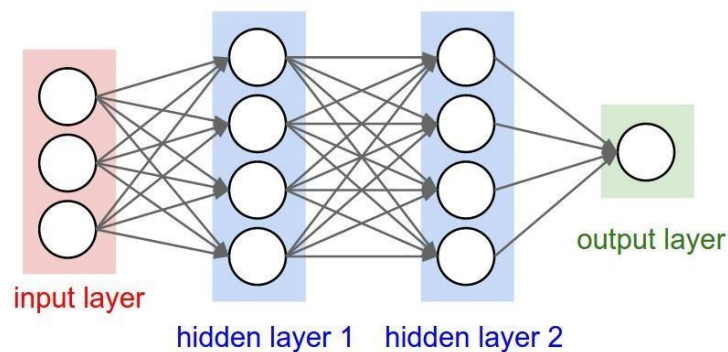


Figure 4.4: A regular 3-layer Neural Network

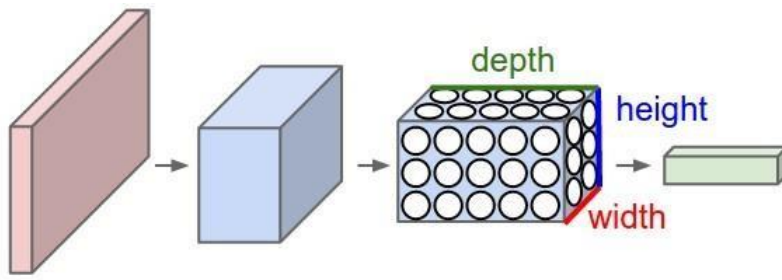


Figure 4.5: A ConvNet arranges its neurons in three dimensions (width, height, depth), as visualized in one of the layers. Every layer of a ConvNet transforms the 3D input volume to a 3D output volume of neuron activations. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be 3 (Red, Green, Blue channels)

Parameter Sharing. Parameter sharing scheme is used in Convolutional Layers to control the number of parameters. Using the real-world example above, we see that there are $55 \times 55 \times 96 = 290,400$ neurons in the first Conv Layer, and each has $11 \times 11 \times 3 = 363$ weights and 1 bias. Together, this adds up to $290400 \times 364 = 105,705,600$ parameters on the first layer of the ConvNet alone. Clearly, this number is very high.

It turns out that we can dramatically reduce the number of parameters by making one reasonable assumption: That if one feature is useful to compute at some spatial position (x,y) , then it should also be useful to compute at a different position (x_2,y_2) . In other words, denoting a single 2-dimensional slice of depth as a depth slice (e.g. a volume of size $[55 \times 55 \times 96]$ has 96 depth slices, each of size $[55 \times 55]$), we are going to constrain the neurons in each depth slice to use the same weights and bias. With this parameter sharing scheme, the first Conv Layer in our example would now have only 96 unique set of weights (one for each depth slice), for a total of $96 \times 11 \times 11 \times 3 = 34,848$ unique weights, or 34,944 parameters (+96 biases). Alternatively, all 55×55 neurons in each depth slice will now be using the same parameters. In practice during backpropagation, every neuron in the volume will compute the gradient for its weights, but these gradients will be added up across each depth slice and only update a single set of weights per slice.

Notice that if all neurons in a single depth slice are using the same weight vector, then the forward pass of the CONV layer can in each depth slice be computed as a **convolution** of the neuron's weights with the input volume (Hence the name: Convolutional Layer). This is why it is common to refer to the sets of weights as a **filter** (or a **kernel**), that is convolved with the input.



Figure 4.6: Example filters learned by Krizhevsky et al. Each of the 96 filters shown here is of size $[11 \times 11 \times 3]$, and each one is shared by the 55×55 neurons in one depth slice. Notice that the parameter sharing assumption is relatively reasonable: If detecting a horizontal edge is important at some location in the image, it should intuitively be useful at some other location as well due to the translationally invariant structure of images. There is therefore no need to relearn to detect a horizontal edge at every one of the 55×55 distinct locations in the Conv layer output volume.

4.4.2.2 Learning process

The learning process used is the backpropagation of error and the gradient descent. Gradient descent is based on the observation that if the multi-variable function $F(x)$ is differentiable and defined in a neighborhood of a point a , then $F(x)$ decreases fastest if one goes from a in the direction of the negative gradient of F at a , $-\nabla F(a)$

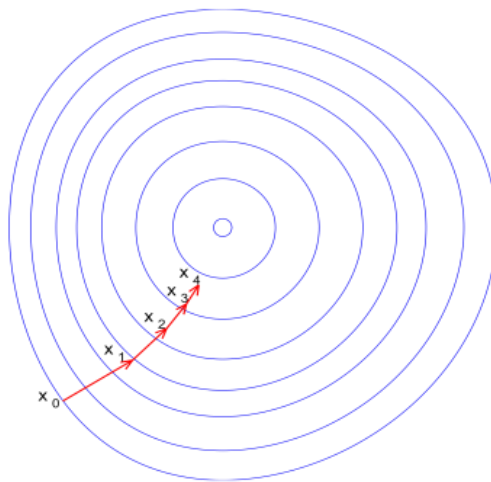


Figure 4.7: Illustration of gradient descent on a series of level sets

The algorithm is iterative, we start from a first guess and then iterate on each pass by moving into the opposite direction of the gradient by using

$$b = a - \gamma \nabla F(a) \quad (4)$$

Where b is step $x+1$ if a is the step x . This iteration can be seen in the following figure:

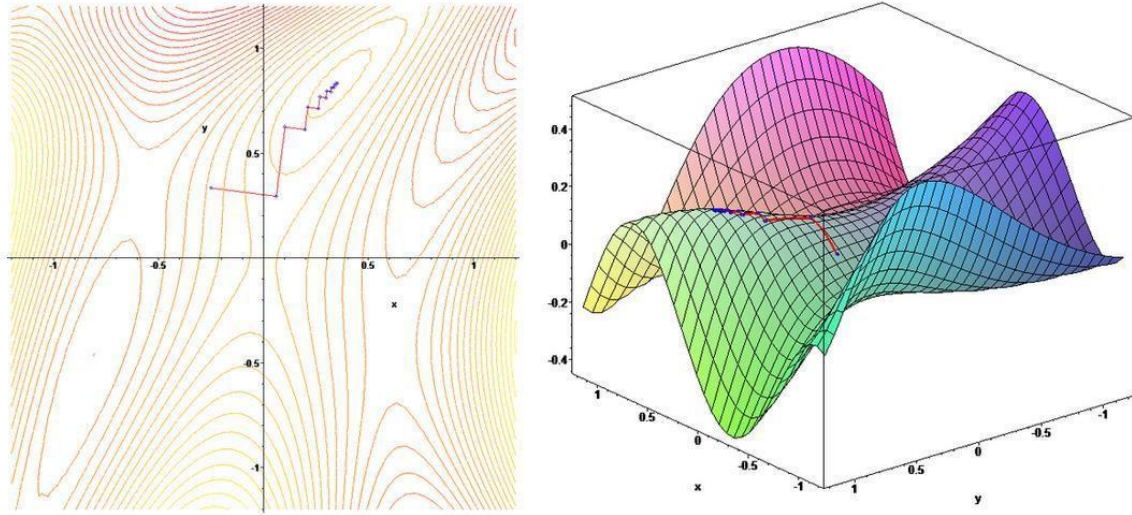


Figure 4.8: the iteration of CNN algorithm

The structure of the CNN used and implemented in this work:

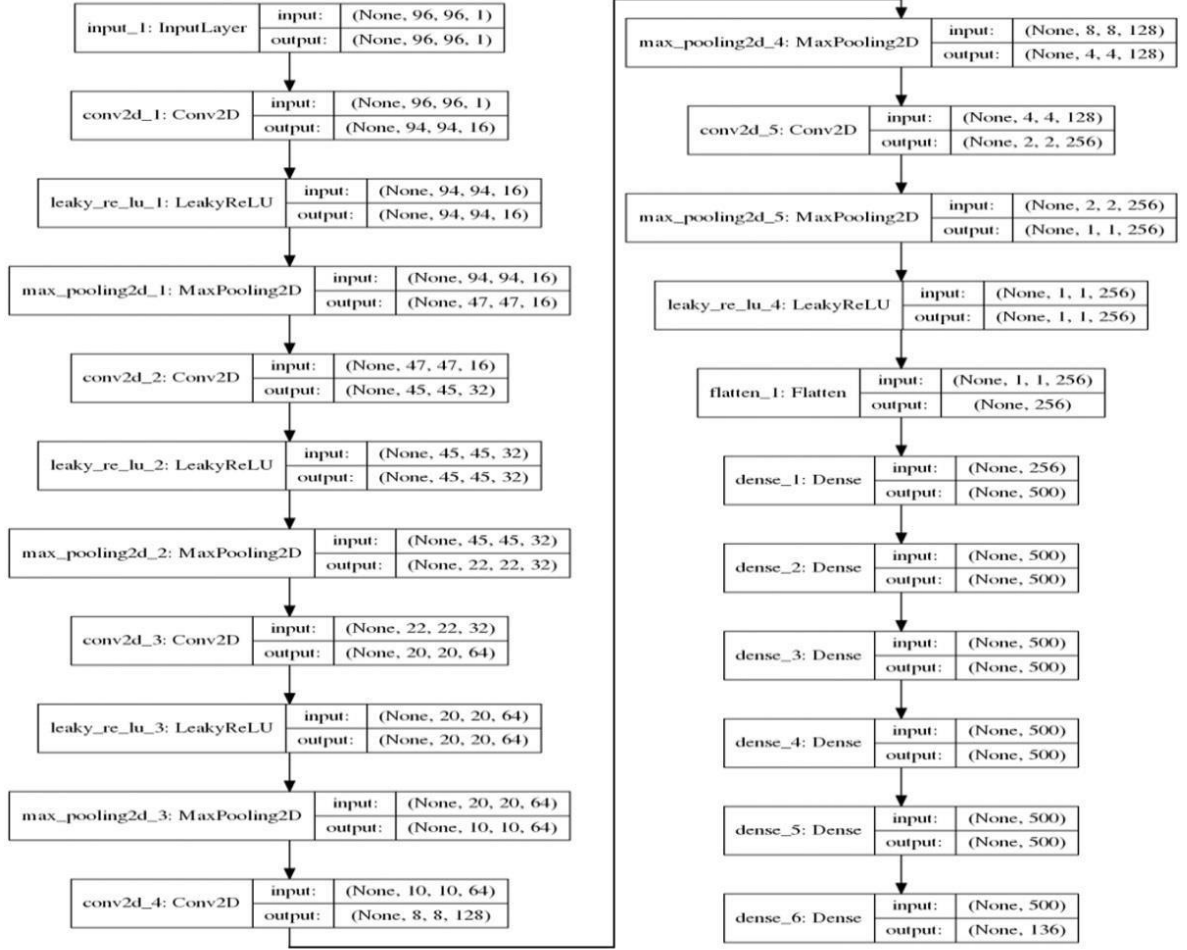


Figure 4.9: CNN structure

As we can see there are 136 outputs on this network where 68 are the x coordinates and 68 others are the y coordinates of the landmarks.

To activate the convolutional layers, we use the Leaky ReLU process and the loss is calculated as the total root mean squared error (RMSE), an elective measure of the deviations in distances between the 68 real and predicted facial point coordinates.

(5)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

4.4.2.3 Data preparation

A set of datasets is used for training the CNN which contains about 5000 images. We also applied following data augmentation techniques to expand the size of our training data:

4.4.2.3.1 Horizontal Reflection (Mirror)

The first technic used for expanding data size is mirroring the image. We reflected the image horizontally. Therefore, we have the similar key point labels in a different orientation.

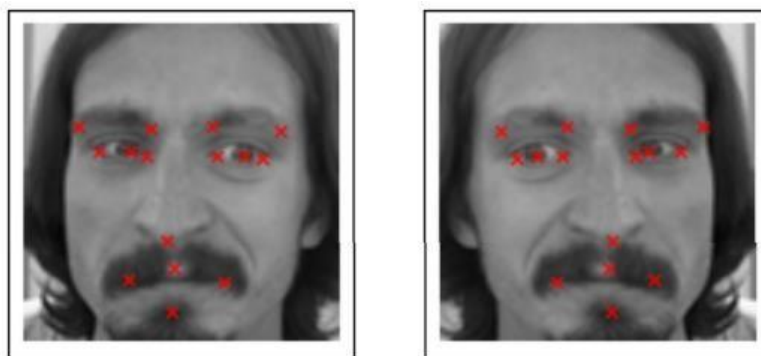


Figure 4.10: mirroring images

4.4.2.3.2 Rotation

The second method we used is randomly rotating images by the range of 10 to -10 degrees.



Figure 4.11: rotating images

4.4.2.3.3 Contrast Reduction (jittering)

The last data augmentation that we used is decreasing the contrast by applying the below formula:

$$CR(X) = (\delta * X) + (1 - \delta) * mean(X)$$

(6)

The idea is that pixel values are shifted slightly towards delta, which is the images mean pixel value.



Figure 4.12: Jittering image

All input images were reduced to 96X96 by maintaining the image's aspect ratio and cropped from face. Our training set images was transformed into around 12000 input images by applying data augmentation techniques

4.4.2.4 Training

Hyperparameters have been tuned properly before training the network. Our hyper parameters are as follows:

Learning rate, batch size, leaky ReLU and regularization parameters. In order to tune hyper parameters, we use the Grid search approach. We trained a model for each combination of hyper parameters values and score on the validation data. Following figures are some of plotted losses across epochs that we obtained.

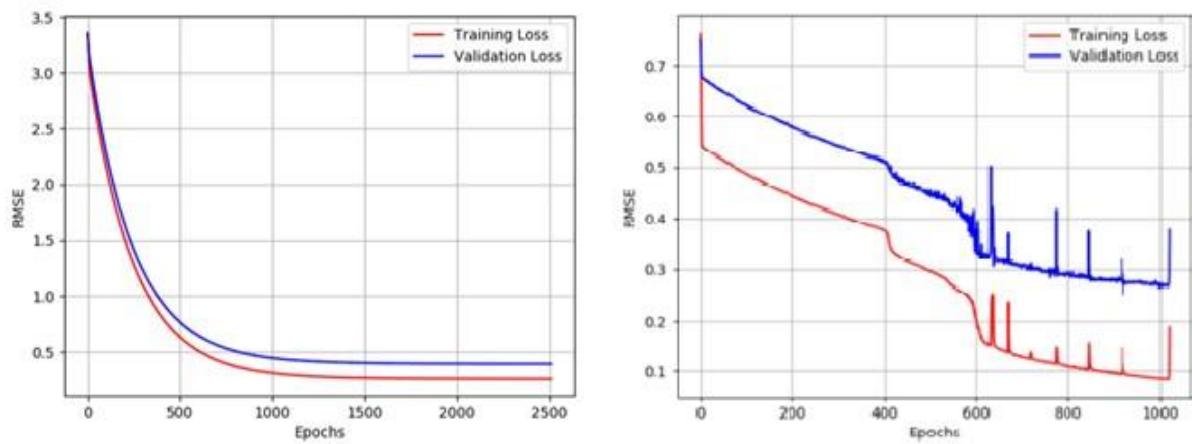


Figure 4.13: epoch graphs

At the end the model is defined by following hyper parameter values:

Learning rate: 0.01

Regularization parameter: $1e-5$

LeakyRelU parameter: 0.1

Batch size: 128

Here is the error function plot:

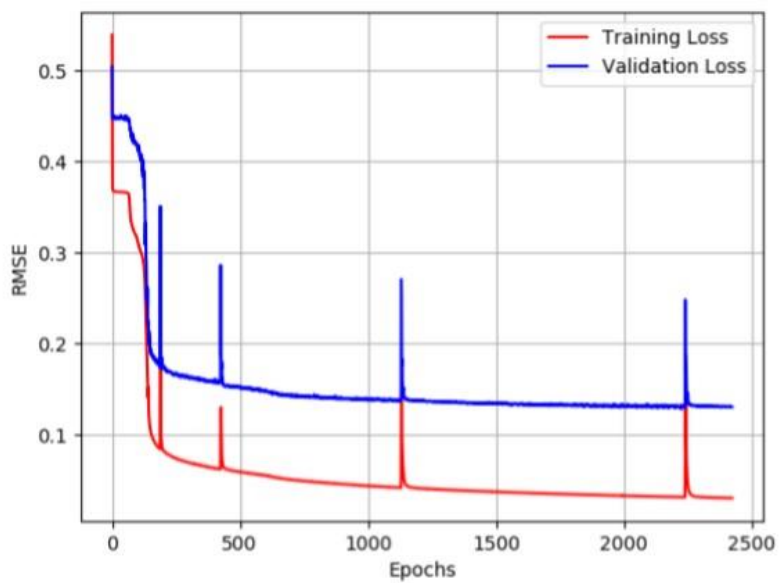


Figure 4.14: error function plot

These are the selected filters from first layer of CNN

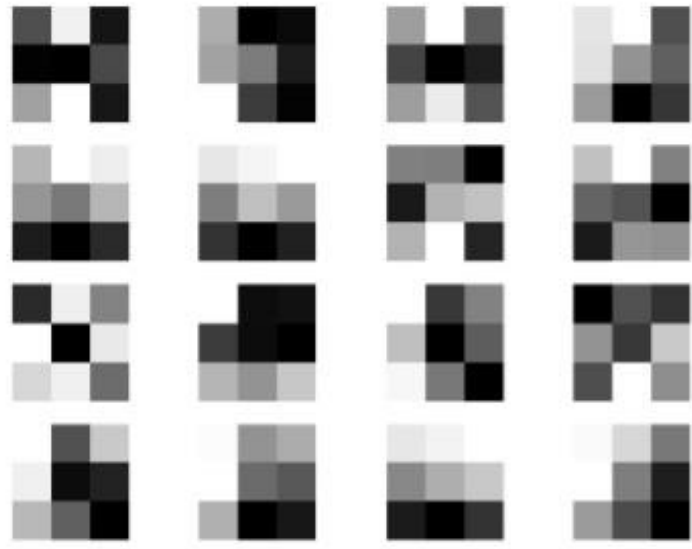


Figure 4.15: CNN first layer filters

4.4.3 Regression trees using the Dlib library

The detection is based on OpenCV and dlib (<http://dlib.net/>) library. The pre-trained landmark detection model is used to detect facial landmarks. Two landmark detection models are tested, and results will be provided:

- 68 facial landmarks points.
- 194 facial landmarks points.

The detection is based on OpenCV and dlib (<http://dlib.net/>) library. The pre-trained landmark detection model is used to detect facial landmarks. Two landmark detection models are tested, and results will be provided: a) 68 facial landmarks and b) 194 facial landmarks.

This dlib's face key-point detection implementation is based upon "One millisecond face alignment with an ensemble of regression trees. One millisecond face alignment [50] with an ensemble of regression trees: In this method an ensemble of regression trees can be used to estimate the face's landmark positions directly from a sparse subset of intensities, achieving

super-real-time performance with high quality predictions. An ensemble of regression trees is learnt based on gradient boosting optimizing the sum of square error loss and naturally handling the missing or partially labelled data [43]

The algorithm works as follows. Let $x_i \in \mathbb{R}$ be the x,y-coordinates of the i th facial landmark in an image I . Then the vector $S = (x_1, x_2, \dots, x_p)$ denotes the coordinates of the p facial landmarks in I .

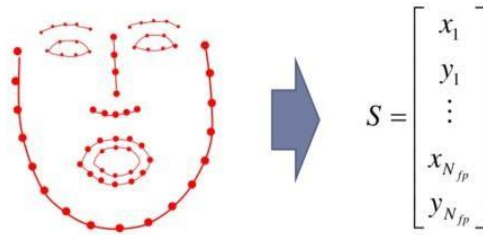


Figure 4.16: Dlib landmark vector

We refer vector S as shape and $\check{S}^{(t)}$ denotes current estimate of S . In each stage, each regressor $r^{(t)}$ in the cascade predicts an update vector from the image and previous estimate, which can be denoted as follows.

(6)

$$\check{S}_{(t+1)} = \check{S}_{(t)} + r_t(I, \check{S}_{(t)})$$

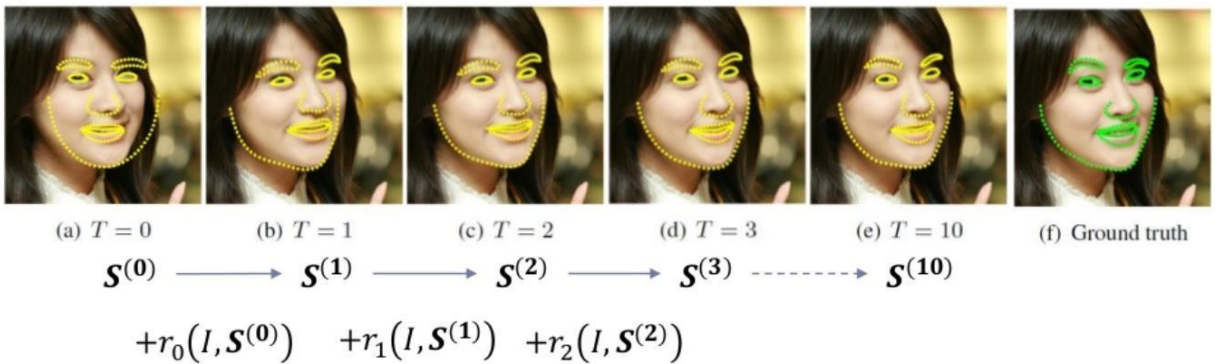


Figure 4.17: Dlib successive iterations

The critical point of the cascade is that the regressor r_t makes its predictions based on features, such as pixel intensity values, computed from I and indexed relative to the Current shape estimate $\check{S}^{(t)}$. This introduces some form of geometric invariance into the process and as

the cascade proceeds one can be more certain that a precise semantic location on the face is being indexed. Later we describe how this indexing is performed.

Note that the range of outputs expanded by the ensemble is ensured to lie in a linear subspace of training data if the initial estimate $S(0)$ belongs to this space. We therefore do not need to enforce additional constraints on the predictions which greatly simplifies our method. The initial shape can simply be chosen as the mean shape of the training data centered and scaled according to the bounding box output of a generic face detector. To train each r_i we use the gradient tree boosting algorithm with a sum of square error loss as described in [51]. The details of this process are given in the original paper.

The below algorithm demonstrates the learning r_i in the cascade. [52] [41]

Have training data $\{(I_{\pi i}, \hat{S}_i^{(t)}, \Delta S_i^{(t)})\}$ and learning rate (shrinking factor) $0 < \nu < 1$

1. Initialize

$$f_0(I, \hat{S}^{(t)}) = \arg \min_{\gamma \in \mathbb{R}^{2p}} \sum_{i=1}^N \|\Delta S_i^{(t)} - \gamma\|^2$$

2. For $k = 1, \dots, K$:

(a) Set for $i = 1, \dots, N$

$$r_{ik} = \Delta S_i^{(t)} - f_{k-1}(I_{\pi i}, \hat{S}_i^{(t)})$$

(b) Fit a regression tree to the targets r_{ik} giving a weak regression function $g_k(I, S^{(t)})$

(c) Update

$$f_k(I, S^{(t)}) = f_{k-1}(I, S^{(t)}) + \nu g_k(I, S^{(t)})$$

3. Output $r_k(I, S^{(t)}) = f_k(I, S^{(t)})$

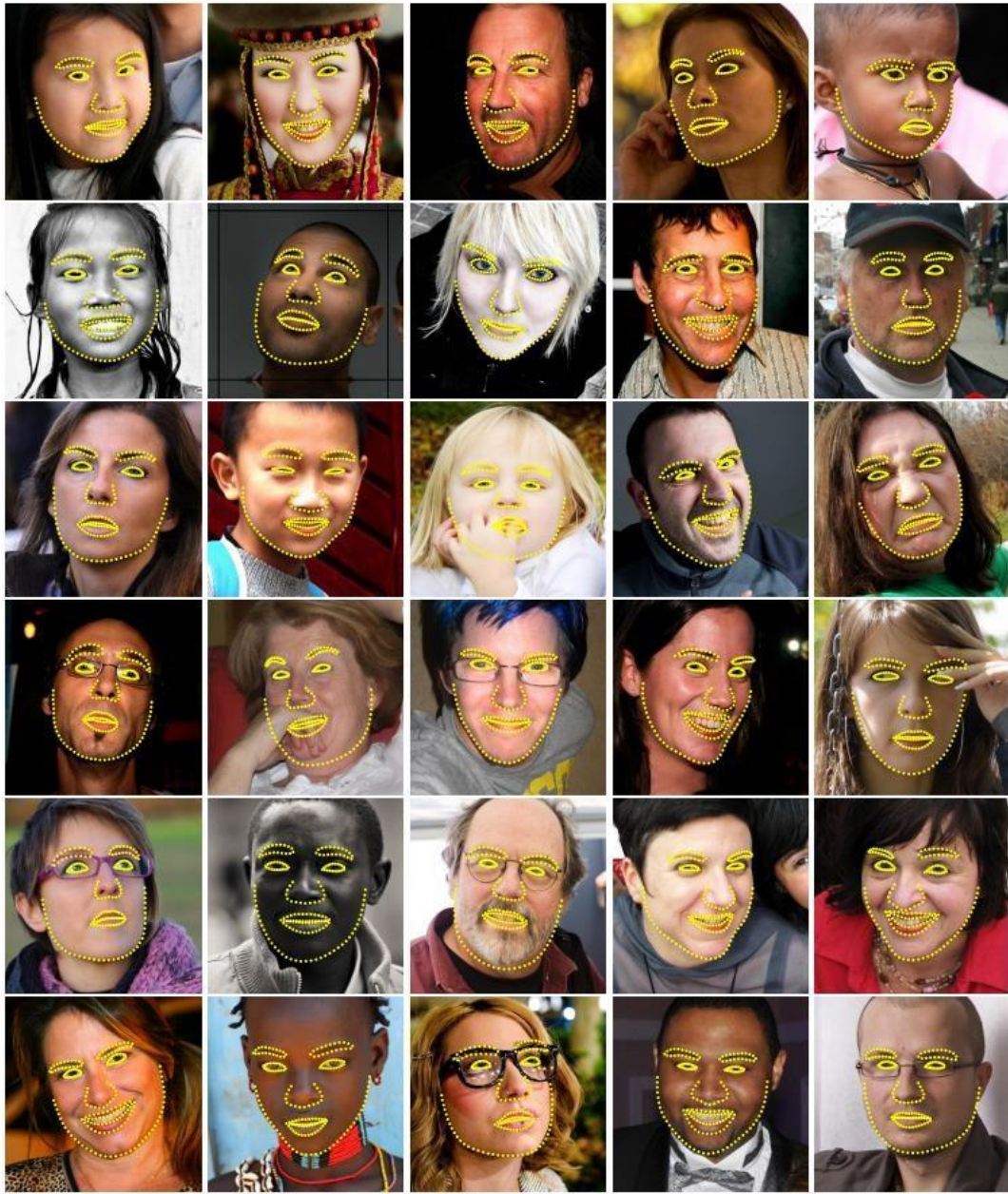


Figure 4.18: Some results in Helen dataset.

4.5 Ground truth Hand marking of the data set

All photos of the data set were manually tagged, and main facial landmarks were extracted. The manual process was repeated twice to eliminate human mistakes as much as possible. More than 600 photos were manually tagged and stored as a precision data base for facial landmarks. The process is extremely time consuming but allows a proper study over the precise positions of facial traits needed in this study.

The chosen landmarks are those that were judged most significant for this study. The obtained database is not yet publicly available due to legal reasons and lack of publication permission from the participants.



Figure 4.19: manual landmarking example image

Table 1 shows the extracted coordinate of manual tagging for only eye corners, iris tangent line and lips. We separated these coordinates for all methods in following chapters, to establish the comparison method.

Girl no.	41	Manual											
Date of Photo	rate number	Left eye				Right eye				Lip			
		Left corner		Right corner		Left corner		Right corner		Left corner		Right corner	
14-Nov	27	577	763	735	777	943	778	1103	762	698	1181	992	1177
16-Nov	45	574	751	733	762	947	765	1102	753	684	1163	996	1165
21-Nov	9	566	760	731	777	943	775	1109	756	685	1183	998	1176
28-Nov	27	568	771	724	786	932	785	1093	769	683	1182	977	1179
17-Nov	54	562	775	722	791	931	789	1095	764	684	1193	995	1180
22-Nov	36	570	741	732	754	941	760	1103	744	684	1160	988	1160
23-Nov	18	568	756	729	772	941	775	1103	759	683	1181	982	1183
29-Nov	9	565	755	730	769	945	768	1112	754	688	1181	994	1176
8-Nov	9	562	739	727	753	944	756	1108	741	678	1166	1006	1158
10-Nov	18	563	750	728	764	938	762	1108	745	682	1173	993	1170
5-Dec	27	565	768	727	788	942	792	1105	775	671	1200	987	1203
2-Dec	27	572	729	734	745	948	749	1114	734	687	1159	989	1160
6-Dec	27	567	724	728	742	948	749	1112	740	667	1160	980	1168
7-Dec	36	574	771	730	785	942	789	1108	778	678	1193	979	1196
12-Dec	9	576	765	736	779	949	784	1110	768	689	1187	987	1192
9-Dec	0	561	733	719	747	931	749	1093	738	680	1159	975	1154
14-Dec	0	570	726	730	753	937	758	1100	737	683	1153	978	1149
15-Dec	0	573	731	737	750	951	753	1120	735	692	1161	996	1157

Table 4-1: manual tagging coordinates

5 Results and discussion

5.1 Rating results

In this section the results are presented in different categories. The first and the most astonishing results are from the human rating process of photos of the same girl during the mensal period.

We divided the mensal period into 3 parts with day distances from the ovulation day. For example, if the ovulation date was 15 of November and the period time was 28 days, then the division was from day 1 to 5 and day 23 to 28 in one group, day 5 to 9 and 18 to 23 in second group and day 9 to 14 and 14 to 18 in the third group.

As the following charts show, a huge majority of rating people chose the attractive photo in the period 9 to 18. This means that majority of people find that photos in the range of 4 days before and after the ovulation time are the most attractive.

The following charts show the results for 9 samples and the last charts shows the results for the total dataset.

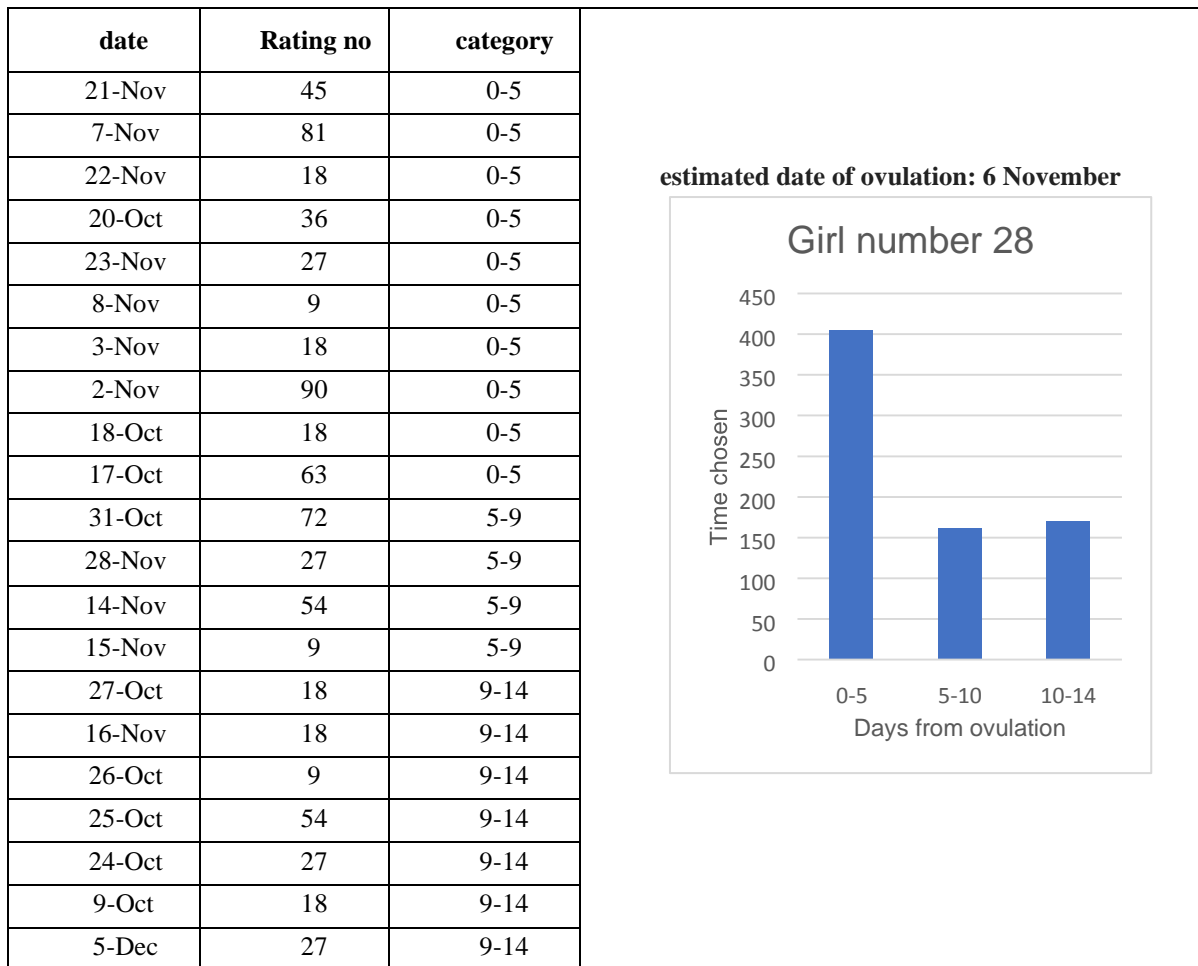


Table 5-1: Rating results extracted from the rating App for a set of images of one girl

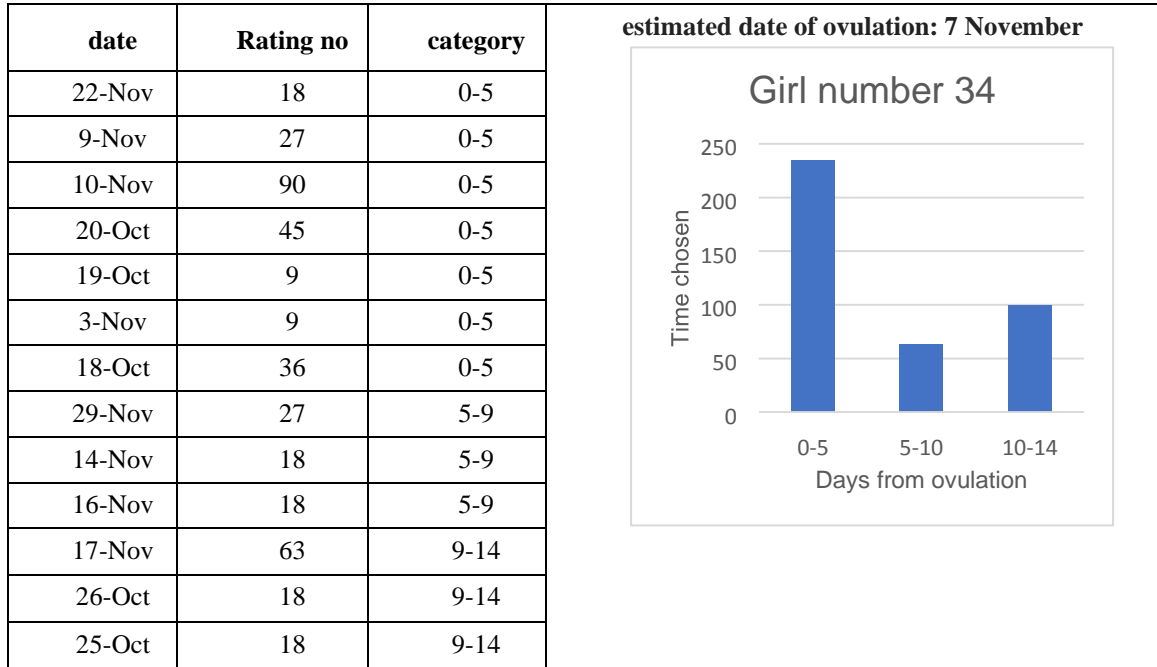


Table 5-2: Rating results extracted from the rating App for a set of images of one girl

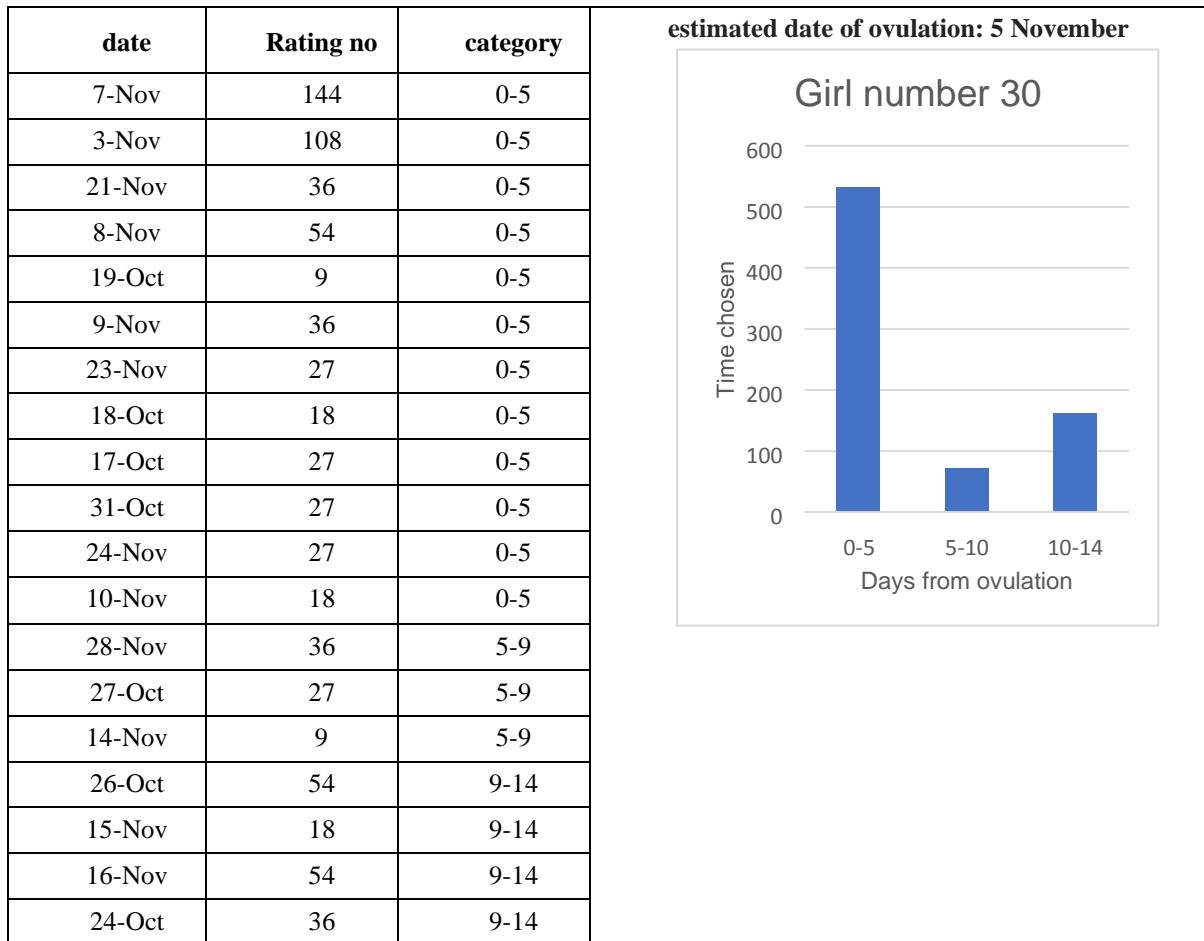


Table 5-3: Rating results extracted from the rating App for a set of images of one girl

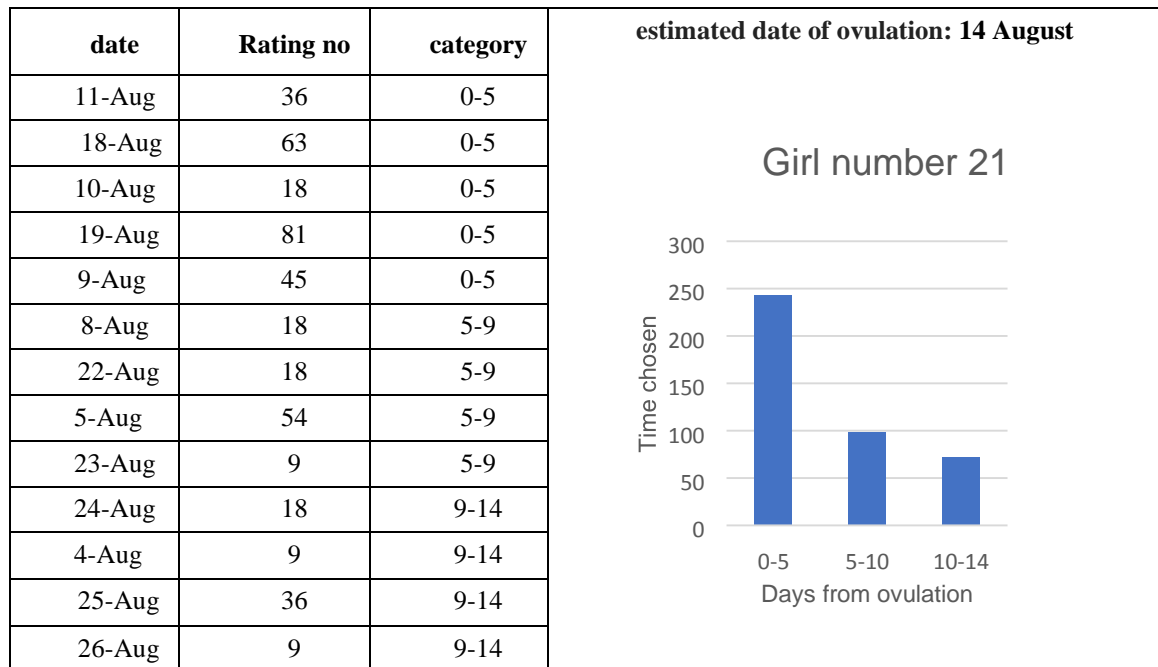


Table 5-4: Rating results extracted from the rating App for a set of images of one girl

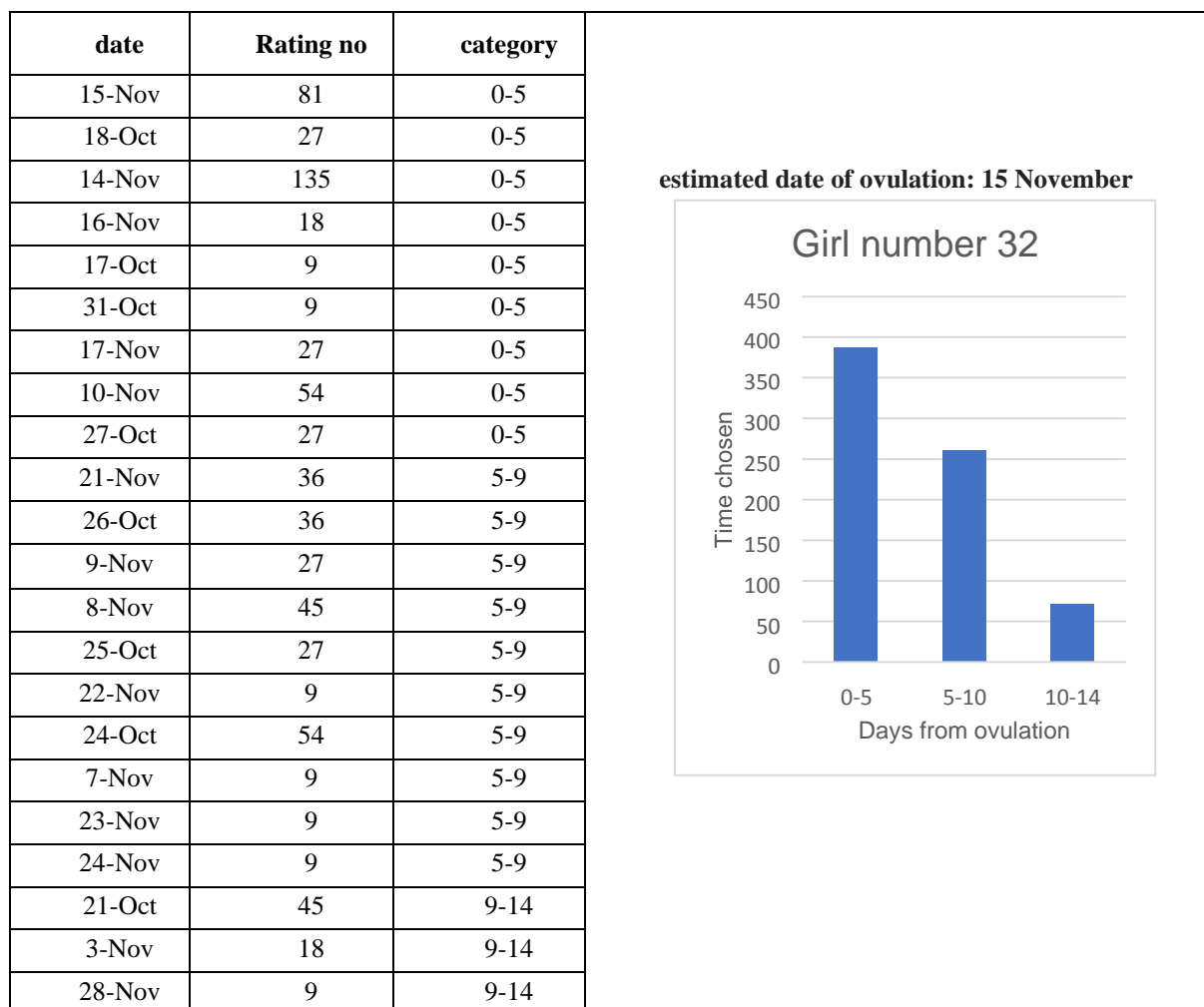


Table 5-5: Rating results extracted from the rating App for a set of images of one girl

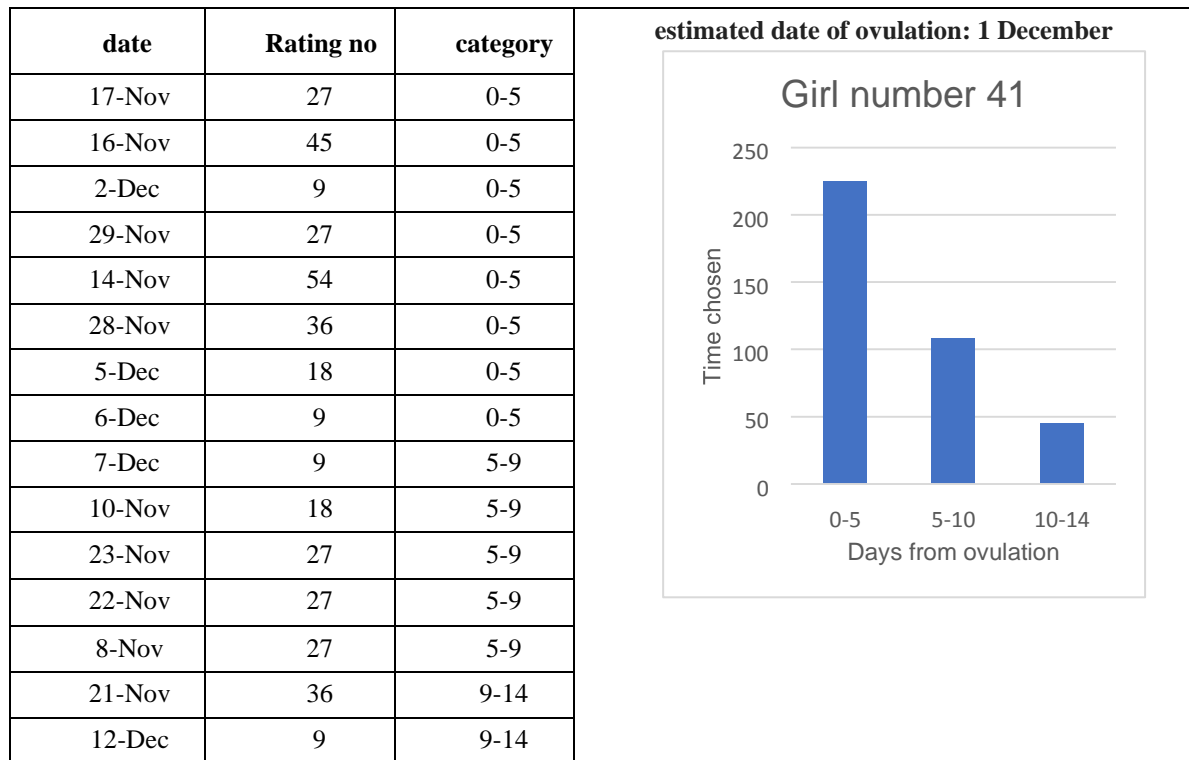


Table 5-6: Rating results extracted from the rating App for a set of images of one girl

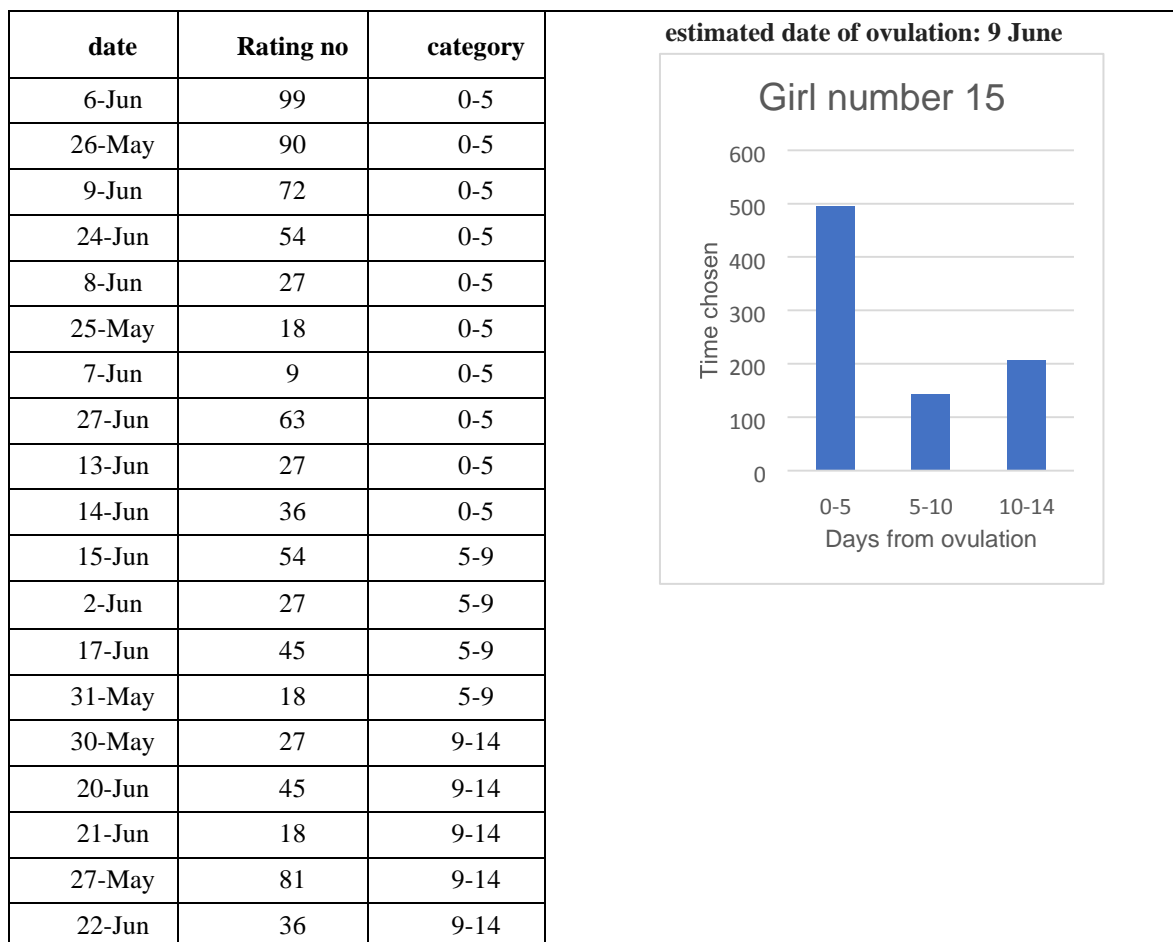


Table 5-7: Rating results extracted from the rating App for a set of images of one girl

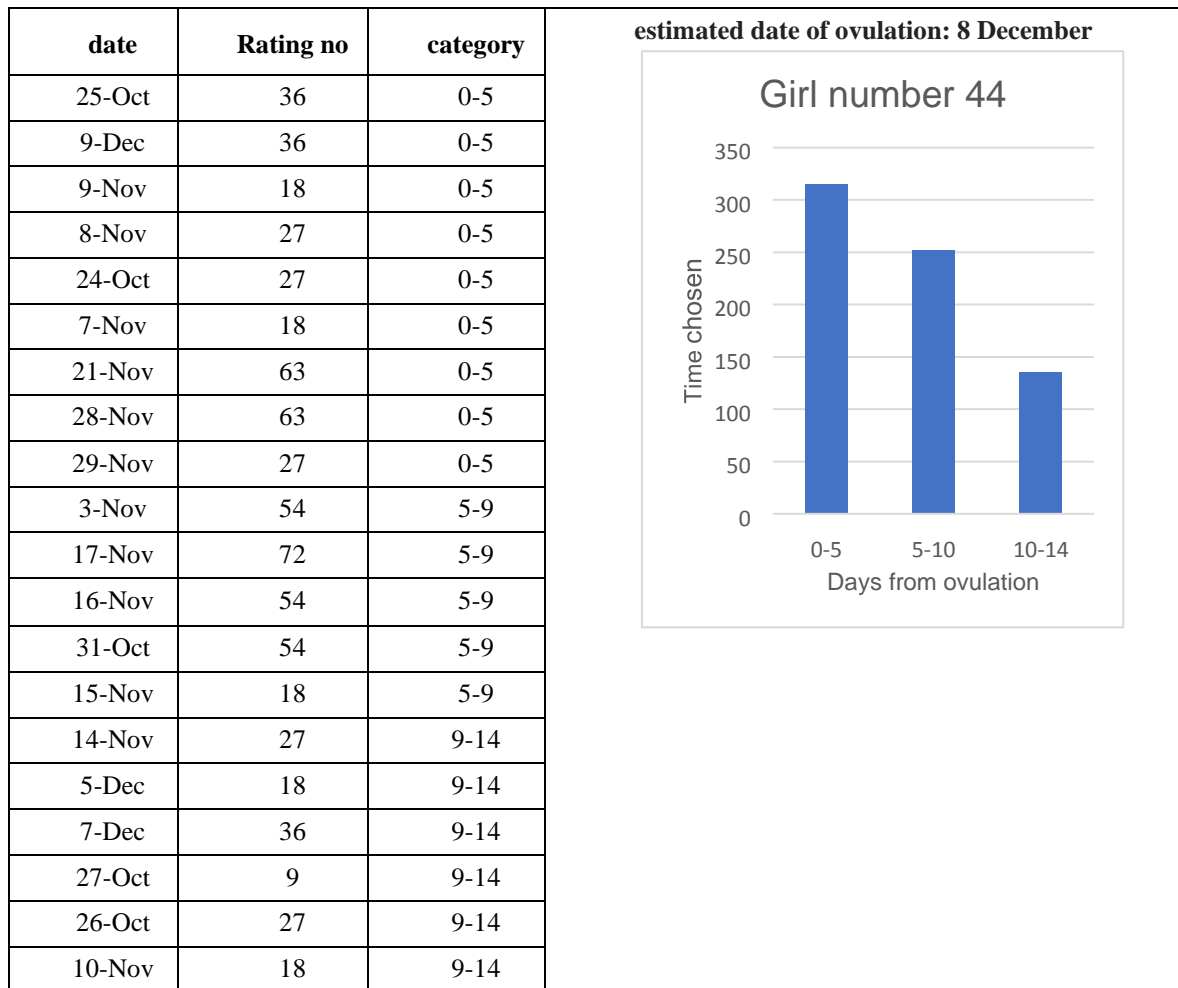


Table 5-8: Rating results extracted from the rating App for a set of images of one girl

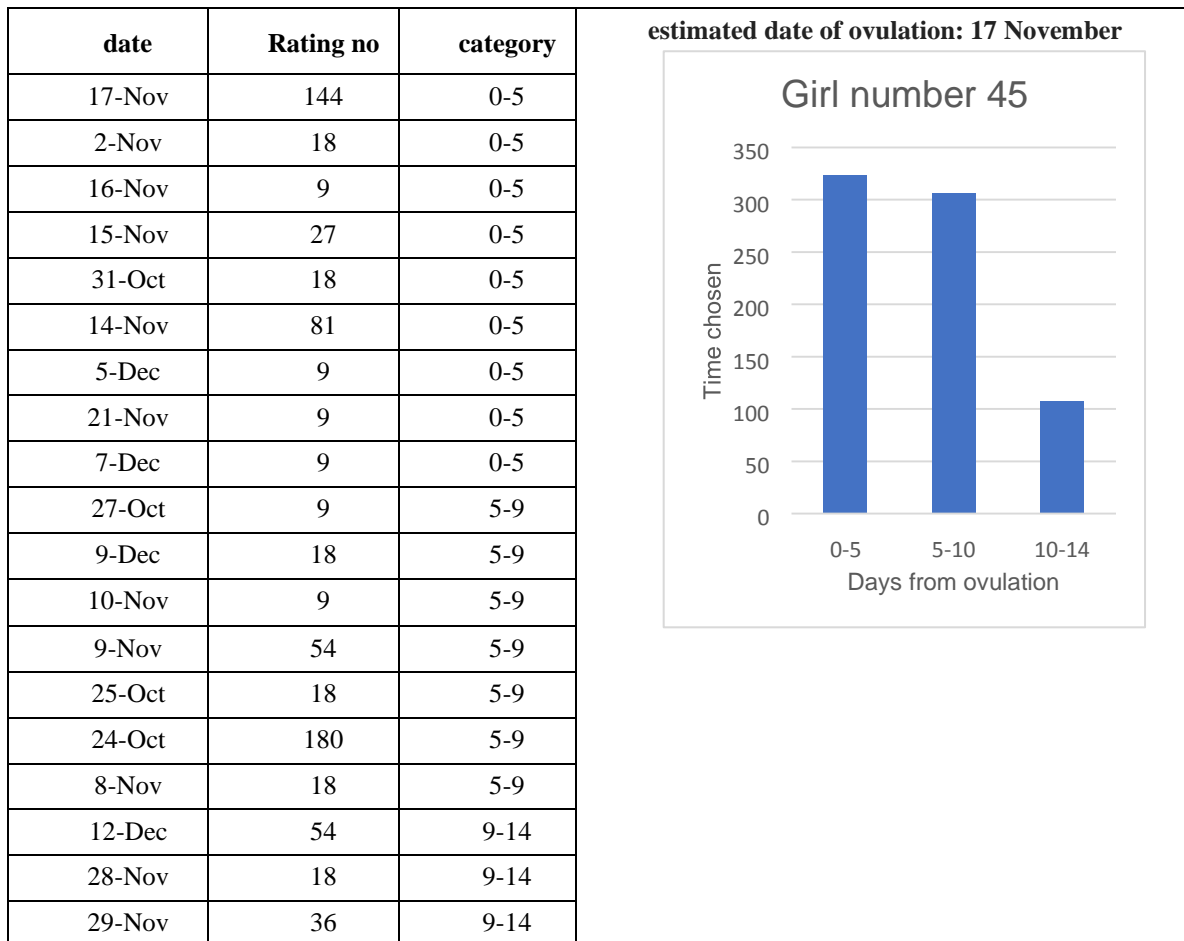


Table 5-9: Rating results extracted from the rating App for a set of images of one girl

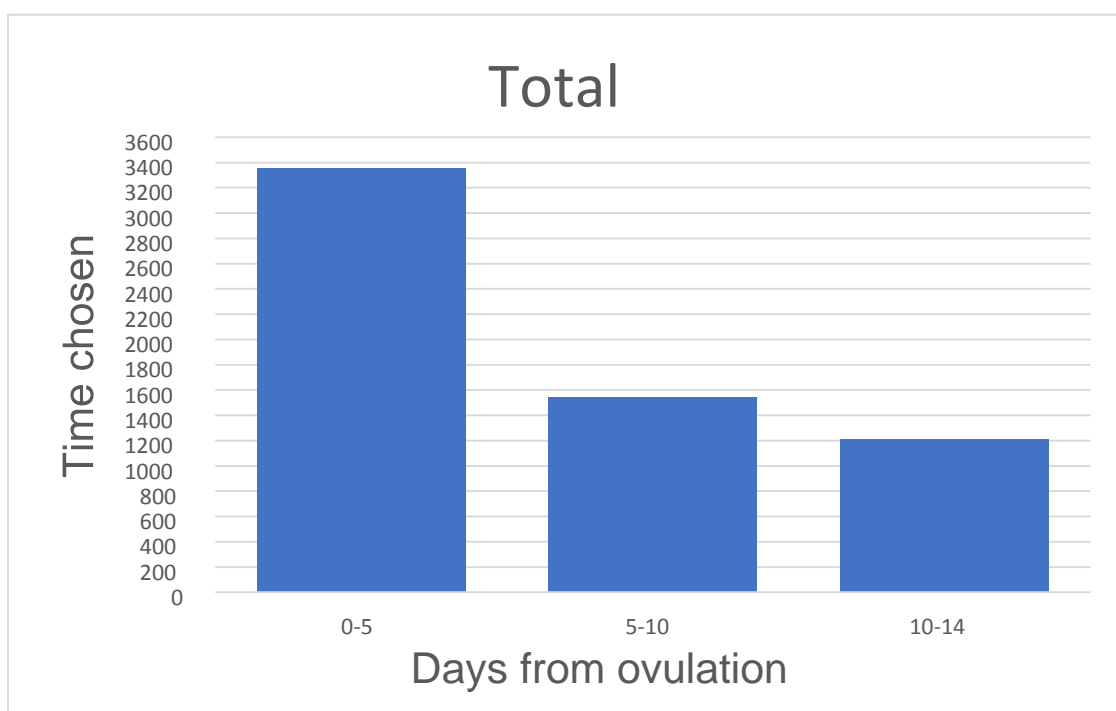


Table 5-10: Rating chart extracted from the rating App for all set of images of all girls

As it can be seen from the above distribution of the results, there is a clear distinction between the 3 different intervals that we used. This result has a very high impact as it is the first medium scale study showing that human ovulation is not completely concealed.

The impact of this first phase of the study is of importance to any evolutionary biologist or people working with fluctuant asymmetry. No previous study proved the existence of visual clues of human ovulation and a change in visual landmarks.

The studies of fluctuant asymmetry should also now be taken with more attention as we proved that human facial clues could have very quick changes and therefore introduce dynamic bias into the final conclusion of asymmetry studies.

5.2 Automatic landmark extraction results

We implemented and tested the 3 methods explained earlier in this chapter, a comparison method for the results is explained later in this section and the table is given. But before entering the details, a simple visual inspection of the results shows that the CNN has some strange and out-layered results that make it less trustable when it comes to fully automatic use. There is also a detail about the CNN method that is mostly forgotten, the network can work with very low-resolution pictures and get good results within a suboptimal interval. When it comes to precision around the landmarks, the low-resolution picture is not of much use. The CNN used can get a 94*94 input. Compared to the original picture quality, this 94*94 resolution has more than a five-to-one loss which makes it anyway the weakest of the methods when it comes to precision.

Figure 5.1 to 5.7 show the results of implemented methods on a set of 5 photos of 4 different girls. But the coordinate tables are only for one girl in all dates. The implementation code for the 3 methods can be found by contacting the supervisors of this work. The AAM and CNN were implemented in Python 3, using Keras library, and the Regression tree was implemented in C++ using the Dlib library.

5.3 The AAM results



Figure 5.1: example of the AAM output

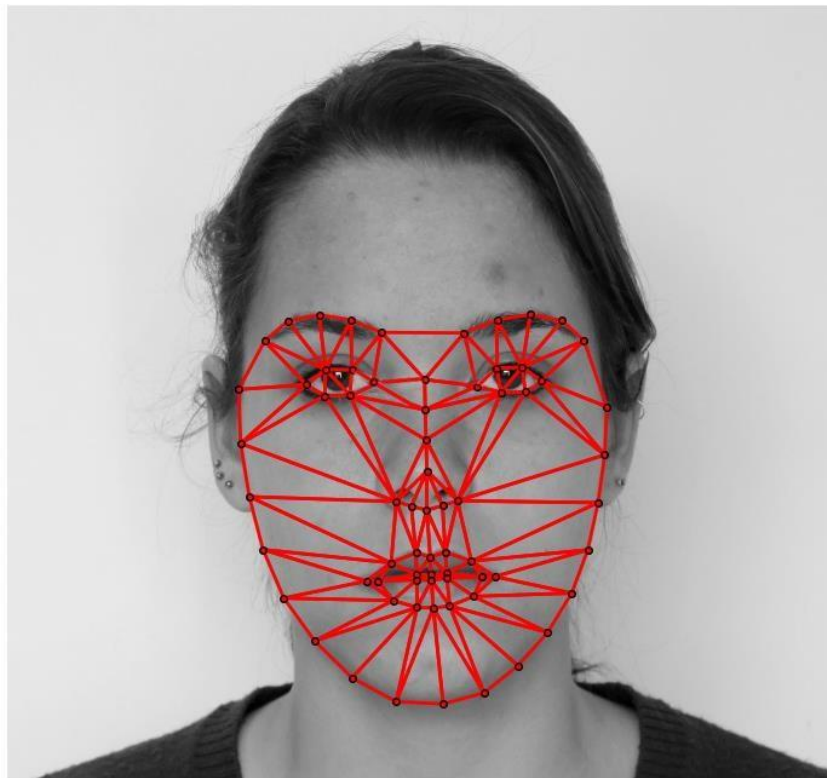


Figure 5.2: one example of the AAM output

Girl no.	41	AMM											
Date of Photo	rate number	Left eye				Right eye				Lip			
		Left corner		Right corner		Left corner		Right corner		Left corner		Right corner	
14-Nov	27	570	767	730	775	946	774	1109	764	690	1183	993	1181
16-Nov	45	568	756	725	761	953	760	1110	755	684	1169	997	1168
21-Nov	9	563	761	719	774	957	771	1114	755	684	1186	1000	1180
28-Nov	27	566	777	720	784	939	783	1100	772	677	1184	984	1181
17-Nov	54	556	778	712	789	944	782	1100	762	684	1197	995	1183
22-Nov	36	567	747	724	755	952	754	1108	746	684	1162	988	1161
23-Nov	18	565	758	726	772	948	772	1109	760	681	1183	986	1186
29-Nov	9	563	762	724	769	953	768	1116	760	685	1184	997	1180
8-Nov	9	559	748	721	753	950	749	1111	739	677	1164	1004	1156
10-Nov	18	556	757	720	761	948	757	1113	746	680	1172	999	1167
5-Dec	27	561	771	719	787	950	789	1112	777	677	1204	986	1206
2-Dec	27	568	731	727	745	959	745	1119	735	683	1162	996	1159
6-Dec	27	563	727	723	742	954	748	1114	740	670	1163	981	1169
7-Dec	36	571	777	730	787	947	789	1111	784	677	1196	983	1199
12-Dec	9	572	768	731	780	952	781	1116	772	689	1193	988	1194
9-Dec	0	559	740	712	747	941	747	1096	742	672	1162	979	1162
14-Dec	0	567	731	721	749	948	752	1105	738	676	1157	980	1152
15-Dec	0	569	737	728	747	963	747	1124	738	684	1167	1002	1163

Table 5-11: AAM extracted coordinates

5.4 The CNN results



Table 5-12: example of CNN output

Girl no.	41	CNN											
Date of Photo	rate number	Left eye				Right eye				Lip			
		Left corner		Right corner		Left corner		Right corner		Left corner		Right corner	
14-Nov	27	593	748	756	794	961	792	1121	783	696	1192	996	1188
16-Nov	45	591	743	751	780	968	778	1122	774	690	1178	1000	1175
21-Nov	9	586	737	745	793	972	789	1126	774	690	1195	1003	1187
28-Nov	27	589	738	746	803	954	801	1112	791	683	1193	987	1188
17-Nov	54	579	730	738	808	959	800	1112	781	690	1206	998	1190
22-Nov	36	590	742	750	774	967	772	1120	765	690	1171	991	1168
23-Nov	18	588	744	752	791	963	790	1121	779	687	1192	989	1193
29-Nov	9	586	742	750	788	968	786	1128	779	691	1193	1000	1187
8-Nov	9	582	739	747	772	965	767	1123	758	683	1173	1007	1163
10-Nov	18	579	738	746	780	963	775	1125	765	686	1181	1002	1174
5-Dec	27	584	737	745	806	965	807	1124	796	683	1213	989	1213
2-Dec	27	591	745	753	764	974	763	1131	754	689	1171	999	1166
6-Dec	27	586	741	749	761	969	766	1126	759	676	1172	984	1176
7-Dec	36	594	748	756	806	962	807	1123	803	683	1205	986	1206
12-Dec	9	595	749	757	799	967	799	1128	791	695	1202	991	1201
9-Dec	0	582	730	738	766	956	765	1108	761	678	1171	982	1169
14-Dec	0	590	739	747	768	963	770	1117	757	682	1166	983	1159
15-Dec	0	592	746	754	766	978	765	1136	757	690	1176	1005	1170

Table 5-13: CNN extracted coordinates

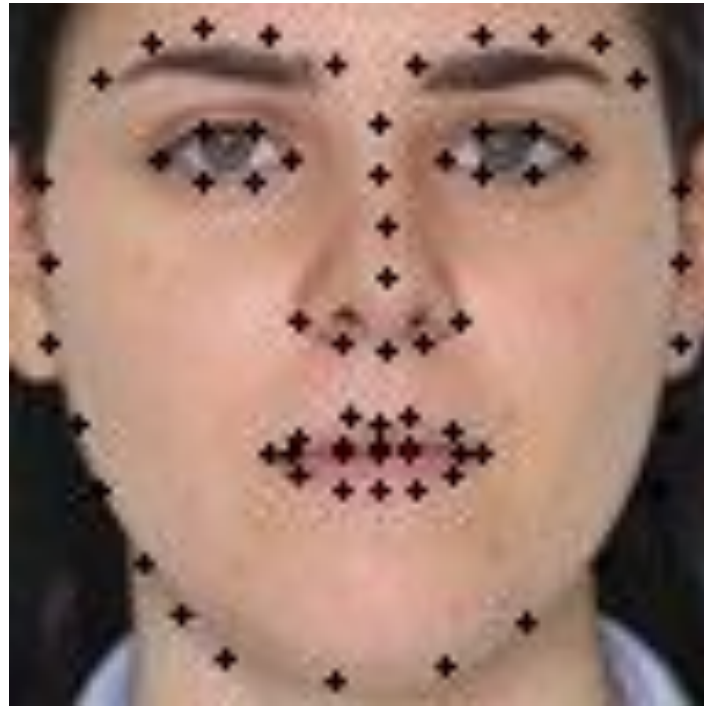


Figure 5.3: one example of the CNN output

5.5 The Dlib results



Figure 5.4: example of Dlib output with 68 points



Figure 5.5: one example of the dlib 68 points output

Girl no.	41	Dlib 68											
Date of Photo	rate number	Left eye				Right eye				Lip			
		Left corner		Right corner		Left corner		Right corner		Left corner		Right corner	
14-Nov	27	575	765	732	776	945	776	1105	764	694	1183	992	1180
16-Nov	45	571	756	730	761	949	762	1107	755	684	1165	997	1164
21-Nov	9	564	761	728	775	947	773	1110	755	684	1184	1000	1179
28-Nov	27	566	773	721	784	935	784	1097	769	680	1184	980	1180
17-Nov	54	560	778	719	789	930	784	1101	762	684	1196	995	1182
22-Nov	36	567	742	728	755	947	759	1106	746	684	1159	988	1161
23-Nov	18	566	757	727	772	944	773	1105	760	681	1182	985	1184
29-Nov	9	564	760	727	769	949	768	1115	759	686	1180	994	1179
8-Nov	9	561	742	723	753	947	753	1110	740	677	1164	1005	1156
10-Nov	18	560	752	726	762	941	759	1110	746	680	1172	995	1169
5-Dec	27	563	769	724	787	945	790	1108	777	674	1202	986	1205
2-Dec	27	571	729	732	745	951	747	1115	735	686	1160	991	1159
6-Dec	27	566	726	724	742	950	748	1112	740	669	1163	980	1169
7-Dec	36	573	774	730	785	944	789	1110	780	677	1195	981	1198
12-Dec	9	574	767	735	780	951	782	1112	771	689	1190	988	1192
9-Dec	0	561	735	715	747	937	747	1095	741	672	1159	976	1159
14-Dec	0	567	728	727	750	941	757	1104	738	680	1156	980	1151
15-Dec	0	572	733	735	748	957	750	1123	738	690	1165	1001	1161

Table 5-14: Dlib- 68 points extracted coordinates



Figure 5.6: example of Dlib output with 194 points



Figure 5.7: one example of the dlib 194 points output

Girl no.	41	Dlib 194											
Date of Photo	rate number	Left eye				Right eye				Lip			
		Left corner		Right corner		Left corner		Right corner		Left corner		Right corner	
14-Nov	27	578	764	733	778	946	777	1103	762	696	1180	993	1180
16-Nov	45	570	759	730	762	948	762	1108	755	683	1162	996	1165
21-Nov	9	568	759	726	775	947	770	1108	758	684	1185	998	1180
28-Nov	27	567	773	718	784	937	786	1095	772	680	1184	983	1180
17-Nov	54	561	778	719	789	928	784	1100	762	684	1196	997	1182
22-Nov	36	569	740	730	750	949	760	1102	742	684	1160	989	1159
23-Nov	18	569	750	724	769	945	770	1100	764	681	1181	986	1184
29-Nov	9	564	760	730	769	947	768	1113	759	682	1180	994	1179
8-Nov	9	560	742	724	754	946	752	1111	746	675	1163	1004	1157
10-Nov	18	563	750	727	762	941	758	1109	746	681	1172	996	1170
5-Dec	27	563	769	724	787	944	790	1106	777	674	1199	983	1197
2-Dec	27	575	729	730	745	951	747	1114	735	686	1160	991	1159
6-Dec	27	566	720	723	740	953	747	1110	742	670	1163	980	1169
7-Dec	36	570	777	730	785	946	789	1112	780	677	1195	979	1192
12-Dec	9	576	765	736	784	952	786	1116	776	686	1192	986	1196
9-Dec	0	559	736	715	743	935	742	1090	741	670	1159	976	1159
14-Dec	0	564	726	725	752	947	756	1104	738	680	1157	981	1152
15-Dec	0	570	730	736	747	956	752	1124	735	693	1160	1006	1160

Table 5-15: Dlib- 194 points extracted coordinates

5.6 Comparing the results between methods

Presenting the error from the automatic extraction methods is not an easy task when it comes to semantic notion of what is modified. To establish a normalized system in the face image and allow the comparison between the ground truth and the automatic results, we draw an imaginary vertical symmetry axis. This axis is obtained by the manual tagging process by adding considering a line between the two iris's midpoint and midpoint of upper lip

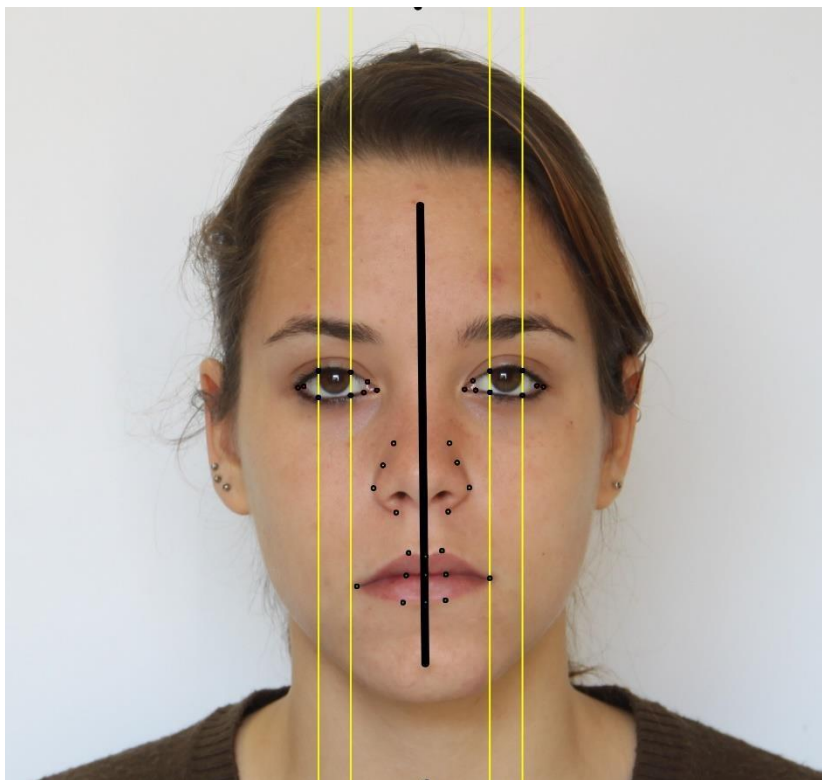


Figure 5.8: The symmetry axis. A line between the iris midpoint and midpoint of upper lip

Then we use an error calculation system by summing the distance differences in pixel unit, from the landmarks to the symmetry axis. The units are normalized with the resolution of the pictures. This system gives a general and item-based notion of error in each part of the face.

This means that for example if the Dlib error for the left eye is 10%, it means that compared to manual measurements it has a bias of 10% towards the symmetry axis of the face.

Methods	Left eye	Right eye	Lips	Nose	Max bias
DLIB	5.5%	5.6%	4%	9%	5%
AAM	8.4%	8.8%	4%	8%	8%
CNN	9.3%	9.1%	4.1%	8%	8%
Pix lvl	1%	<1%	NA	NA	NA

Figure 5.9: Error rate compare to manual marks

The results show that the Dlib regression trees provide a better fit over orthogonal pictures with fix gestures, but also demonstrates that for high precision landmark extraction none of the existing supervised learning methods can provide a reliable precision.

The methods based on learning in general struggle in the optimization process and can never converge to a 100% precision or a general optimal solution. But the studies about attractiveness or fluctuant asymmetry need strong and robust results with millimetric precision. For these studies none of the detection methods based on iteration optimization can be used.

It is clear that the suboptimal solutions, which are very efficient for many applications, are not robust enough to provide reliable solutions for biological related studies. The author believes that the solution is definitely, using low level image processing for the detection of precise position of few important landmarks rather than using learning methods with many points. In the team we have already a developed a low-level eye corner detection, the method should be expanded for lip points, nose limiting points and eyebrows.

As mentioned above the Regression trees present a possible choice for huge data sets as the introduced errors are around 5 percent in absolute measures. Nevertheless, the induced changes by physiological modifications are not very large and this small error can still present a huge bias in the final results.

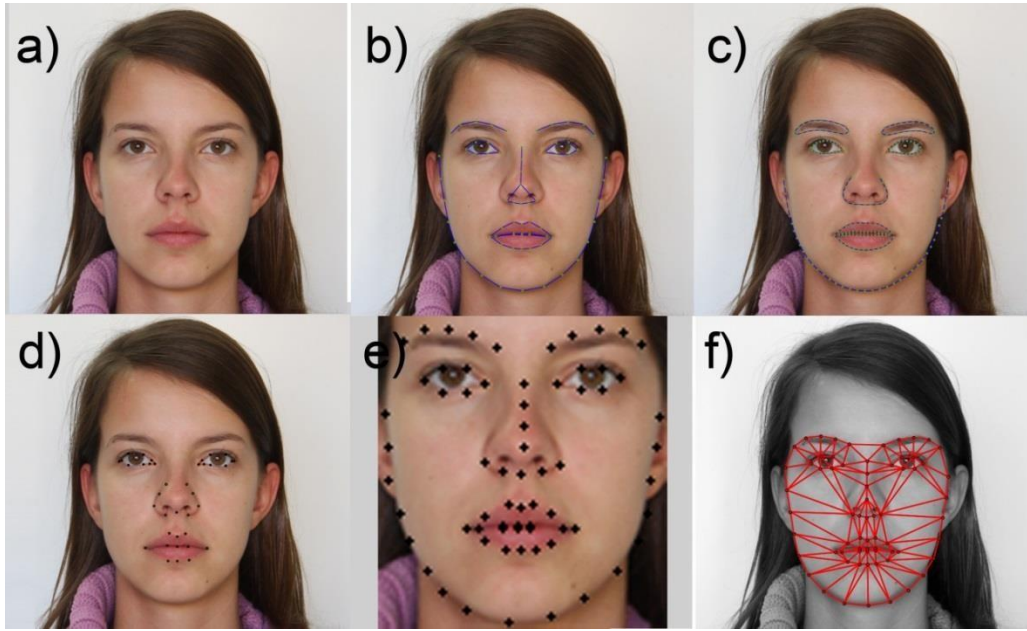


Figure 5.10: Girl no: 41, date: 23 November, a) original image, b) Dlib 68 point, C)Dlib 194 point, d)manual tagging, e) CNN, f) AAM

5.7 Alternative approach to analysis of results

Once we had all the results, we tried to see if there is any possible link that we could establish between the 2D landmarks that we had extracted and the results of the human rating. The idea was to check if we could establish a fix pattern about the landmarks and the beauty or attractiveness.

We used the self-organizing maps of Kohonen to map the landmarks of pictures of each girl during the mensal period. If any relation exists between the landmarks and the actual date of ovulation, we should get clusters of similar landmarks into one group and those clusters should fit the pictures chosen by the human raters. The facts unfortunately are not so simple. The attractiveness, even on a simple 2D picture is more sophisticated, as shown in [52] the index of attractiveness is much more complicated than only some landmarks and 2d measurements. Moreover, the projection on a 2D plan also reduces the available information.

As shown in the following self-organized maps, the direct relation between the mensal period and the distances of landmarks toward a symmetry axes are not clear. Nevertheless, the highest rated pictures mostly stay in the same clusters. Although other pictures are also in those clusters.

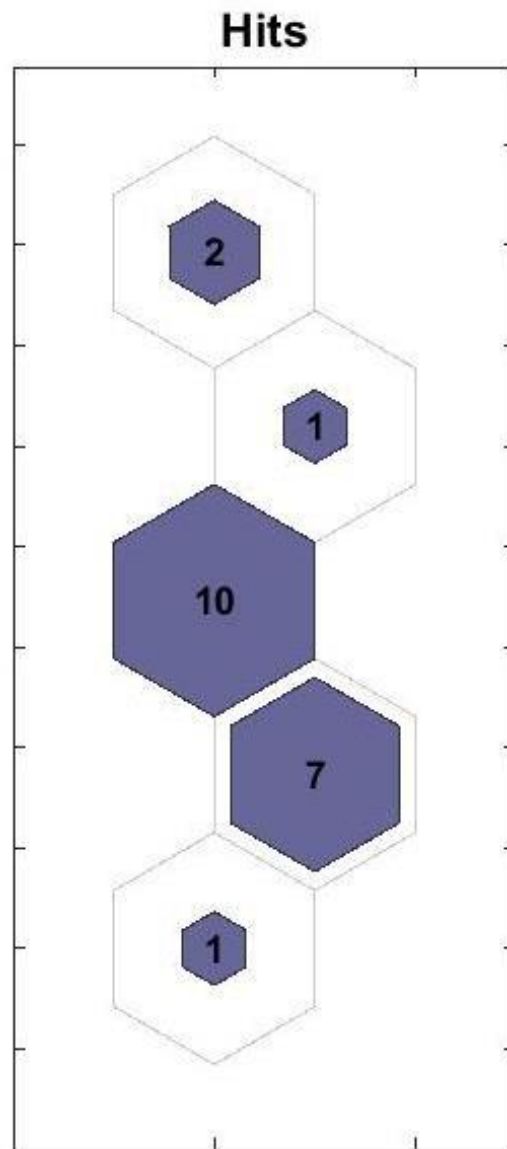


Figure 5.11: KOHONEN clustering plot

We can, analyze the above results in different ways. But the most probable information is that the distances of the landmarks to the symmetry axis may not be the only parameter or the sufficient parameter to cluster the pictures correctly. But they are part of the many other parameters involved in the visual clues which enables human to distinguish between pictures in terms of attractivity.

Here is the input data of kohonen algorithm:

NO 44	Left eye		Right eye		Lip		nose up		nose middle		nosepiece		nostril		nose base	Group Index
Date of Photo	Left corner	Right corner	Left corner	Right corner	Left corner	Right corner	left	right	left	right	left	right	left	right		
17-Nov	0.781	0.216	0.522	1.103	0.373	0.726	0.179	0.245	0.241	0.300	0.315	0.364	0.148	0.170	0.040	4
21-Nov	0.800	0.212	0.545	1.138	0.392	0.725	0.199	0.229	0.250	0.282	0.306	0.333	0.140	0.185	0.021	5
28-Nov	0.814	0.231	0.545	1.144	0.353	0.744	0.184	0.207	0.259	0.278	0.311	0.357	0.144	0.185	0.024	5
31-Oct	0.876	0.315	0.311	0.872	0.442	0.485	0.185	0.232	0.264	0.279	0.304	0.366	0.145	0.188	0.025	3
3-Nov	0.878	0.315	0.309	0.879	0.439	0.479	0.190	0.221	0.250	0.285	0.320	0.364	0.135	0.191	0.028	3
16-Nov	0.880	0.315	0.311	0.883	0.456	0.445	0.182	0.224	0.255	0.268	0.314	0.353	0.140	0.167	0.026	3
25-Oct	0.883	0.324	0.311	0.387	0.456	0.506	0.203	0.211	0.268	0.287	0.318	0.350	0.150	0.191	0.028	2
7-Dec	0.879	0.313	0.313	0.884	0.455	0.495	0.202	0.235	0.262	0.289	0.311	0.352	0.156	0.188	0.020	2
9-Dec	0.881	0.320	0.322	0.879	0.455	0.462	0.178	0.218	0.237	0.291	0.300	0.339	0.138	0.169	0.026	3
14-Nov	0.883	0.320	0.318	0.881	0.456	0.457	0.182	0.229	0.245	0.272	0.311	0.353	0.146	0.179	0.024	3
29-Nov	0.889	0.321	0.330	0.886	0.479	0.470	0.178	0.208	0.253	0.266	0.326	0.343	0.168	0.199	0.017	2
8-Nov	0.868	0.307	0.320	0.873	0.470	0.452	0.173	0.249	0.236	0.271	0.302	0.358	0.144	0.189	0.040	3
24-Oct	0.876	0.315	0.307	0.879	0.441	0.498	0.205	0.210	0.269	0.273	0.316	0.342	0.155	0.161	0.016	2
26-Oct	0.869	0.316	0.311	0.874	0.452	0.476	0.220	0.234	0.284	0.286	0.320	0.346	0.151	0.182	0.019	2
7-Nov	0.872	0.309	0.318	0.871	0.475	0.429	0.186	0.236	0.256	0.283	0.309	0.353	0.146	0.164	0.023	3
9-Nov	0.872	0.314	0.321	0.861	0.453	0.459	0.181	0.247	0.245	0.296	0.298	0.352	0.150	0.186	0.022	3
10-Nov	0.873	0.315	0.318	0.879	0.440	0.489	0.204	0.227	0.257	0.282	0.311	0.363	0.139	0.176	0.024	3
15-Nov	0.868	0.311	0.320	0.873	0.452	0.456	0.173	0.235	0.246	0.282	0.312	0.351	0.137	0.186	0.033	3
5-Dec	0.976	0.379	0.304	0.913	0.486	0.533	0.236	0.200	0.278	0.279	0.360	0.354	0.168	0.187	0.009	2
27-Oct	0.863	0.323	0.312	0.873	0.450	0.469	0.207	0.221	0.269	0.271	0.316	0.357	0.160	0.183	0.018	2
12-Dec	0.879	0.328	0.309	0.886	0.445	0.454	0.169	0.182	0.243	0.270	0.320	0.350	0.151	0.172	0.013	1

Table 5-16: input table of KOHONEN algorithm. (Distance ratio of coordinate to symmetry line)

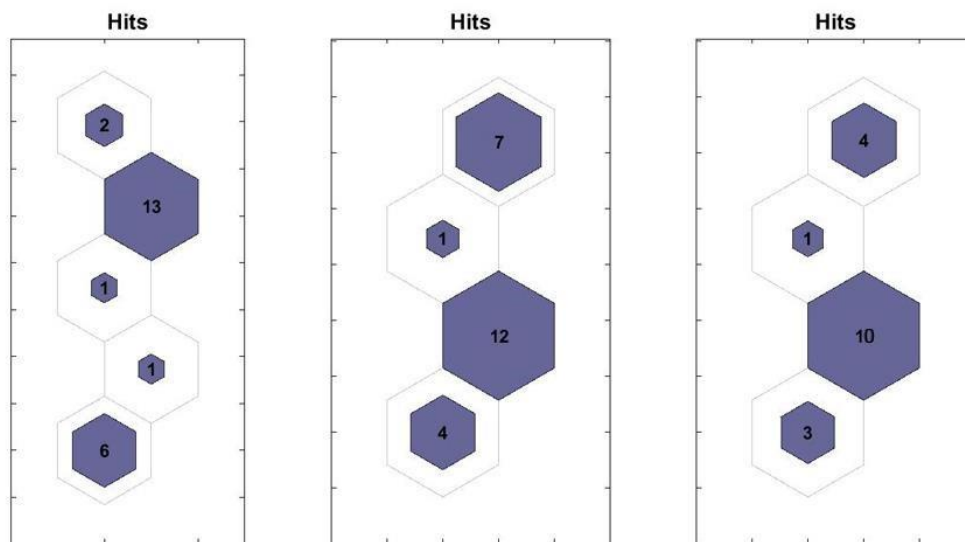


Figure 5.12: some more KOHONEN output example

6 Conclusions

The study carried out has potentially important impacts on different issues. This study has showed that visual clues, even very slight modifications can be perceived and interpreted by human as a sign of biological or metabolic modification. If we consider the changes in the nasal area during the time span of a mensal period, then we can see that the studying a face only over one photo obtained in one day cannot be a real tool for any serious asymmetry studies. If the asymmetry is around 5 % for example and that we get a bias of 3% from first day of ovulation to the 14th day, then obviously the induced error is far too high to allow any correct conclusion. But, for example, one particular case of nose area modification shows high sensitivity to physiologic modifications.

We have not presented all the conclusions and data in this thesis due to ongoing and pending ethical authorizations and data privacy issue. These data will be published as soon as the necessary legal issues are solved. The sociological reasons of why the human ovulation is following a semi-concealed process is to be investigated properly, but we had some strange results that could give us some clues.

One of these results (to be confirmed) was that the normal partner of the woman could not choose the correct period of the ovulation on the pictures, but others could do that better. Another possible but difficult to prove observation was that women who use contraceptive pills show less visual modification. This result is naturally understandable because the hormonal variation is reduced under effect of the contraceptive pills.

7 References

1. Palmer, A. R. Fluctuating asymmetry analyses: a primer. *Springer, Dordrecht*. **1994**. 335–364. http://doi.org/10.1007/978-94-0110830-0_26
2. Valen, L. Van. A Study of Fluctuating Asymmetry. *Evolution Vol. 16, No. 2*. **1962**. 125–142. <http://doi.org/10.2307/2406192>
3. Andelman, S. J. Evolution of Concealed Ovulation in Vervet Monkeys (*Cercopithecus aethiops*). *The American Naturalist* **1987**, 129, 785–799. <http://doi.org/10.1086/284675>
4. Roberts, S. C., Havlicek, J., Flegr, J., Hruskova, M., Little, A. C., Jones, B. C., Perrett, D. I. & Petrie, M. Female facial attractiveness increases during the fertile phase of the menstrual cycle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. **2004**. 271, <http://doi.org/10.1098/rsbl.2004.0174>
5. Miller, G., Tybur, J. M. & Jordan, B. D. Ovulatory cycle effects on tip earnings by lap dancers: economic evidence for human estrus. *Evolution and Human Behavior*. **2007**. 28, 375–381. <http://doi.org/10.1016/J.EVOLHUMBEHAV.2007.06.002>
6. Pipitone, R. N. & Gallup, G. G. Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*. **2008**. 29, 268–274. <http://doi.org/10.1016/J.EVOLHUMBEHAV.2008.02.001>
7. Bullivant, S. B., Sellergren, S. A., Stern, K., Spencer, N. A., Jacob, S., Mennella, J. A. & McClintock, M. K. Women's sexual experience during the menstrual cycle: Identification of the sexual phase by noninvasive measurement of luteinizing hormone. *Journal of Sex Research*. **2004**. 41, 82–93. <http://doi.org/10.1080/00224490409552216>
8. Renninger, L. A., Wade, T. J. & Grammer, K. Getting that female glance: Patterns and consequences of male nonverbal behavior in courtship contexts. *Evolution and Human Behavior*. **2004**. 25, 416–431. <http://doi.org/10.1016/J.EVOLHUMBEHAV.2004.08.006>
9. Thornhill, R. & Gangestad, S. W. The Scent of Symmetry: A Human Sex Pheromone that Signals Fitness? *Evolution and Human Behavior*. **1999**. 20, 175–201. [http://doi.org/10.1016/S1090-5138\(99\)00005-7](http://doi.org/10.1016/S1090-5138(99)00005-7)
10. Brooksbank, B. W. L., Brown, R. & Gustafsson, J.-A. The detection of 5 α -androst-16-en-3 α -ol in human male axillary sweat. *Experientia*. **1974**. 30, 864–865. <http://doi.org/10.1007/BF01938327>
11. Brewis, A. & Meyer, M. Demographic Evidence That Human Ovulation Is Undetectable (At Least in Pair Bonds). *Current Anthropology*. **2005**. 46, 465–471. <http://doi.org/10.1086/430016>
12. ovulatory_shifts_JSR_2004.pdf.
13. The secret life of semen. **2006**, New Scientist. <https://www.newscientist.com/article/mg19125633500-the-secret-life-of-semen/>
14. Watts, I. Chris Knight. 1991. Blood Relations: Menstruation and the Origins of Culture. *Papers from the Institute of Archaeology*. **1992**. 3, 84. <http://doi.org/10.5334/pia.26>
15. Knight, C., Power, C. & Watts, I. The Human Symbolic Revolution: A Darwinian Account. *Cambridge Archaeological Journal*. **1995**. 5, 75–114. <http://doi.org/10.1017/S0959774300001190>
16. Soon, C. S., Brass, M., Heinze, H.-J. & Haynes, J.-D. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*. **2008**. 11, 543–545. <http://doi.org/10.1038/nn.2112>
17. Cooper, J. C., Dunne, S., Furey, T. & O'Doherty, J. P. Dorsomedial prefrontal cortex mediates rapid evaluations predicting the outcome of romantic interactions. *The Journal of neuroscience: the official journal of the Society for Neuroscience*. **2012**. 32, 15647–56. <http://doi.org/10.1523/JNEUROSCI.2558-12.2012>
18. Advances in Computing and Communications, Part III. **2011**. <https://dl.acm.org/citation.cfm?id=2096004>
19. Rathod, D., Vinay, A. A. du texte, Shylaja, S. S. & Natarajan, S. Facial Landmark Localization - A Literature Survey. **2014**. <https://www.semanticscholar.org/paper/Facial-Landmark-Localization-A-Literature-Survey-Rathod-Vinay/6a7e464464f70afea78552c8386f4d2763ea1d9c>. doi:ND
20. Jianhua Tao, Tieniu Tan, Rosalind W. Picard. Affective Computing and Intelligent Interaction. *Springer, Berlin Heidelberg*. **2005**. 3784, <http://doi.org/10.1007/11573548>
21. Pantie, M. & Rothkrantz, L. J. M. Automatic analysis of facial expressions: the state of the art.

- IEEE Transactions on Pattern Analysis and Machine Intelligence*. **2000**. 22, 1424–1445.
<http://doi.org/10.1109/34.895976>
22. van Dam, A. Beyond WIMP. *IEEE Computer Graphics and Applications*. **2000**. 20, 50–51.
<http://doi.org/10.1109/38.814559>
 23. Pentland, A. Looking at people: sensing for ubiquitous and wearable computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **2000**. 22, 107–119.
<http://doi.org/10.1109/34.824823>
 24. Zue, V. W. & Glass, J. R. Conversational interfaces: advances and challenges. *Proceedings of the IEEE*. **2000**. 88, 1166–1180. <http://doi.org/10.1109/5.880078>
 25. Bruce, V. What the human face tells the human mind: some challenges for the robot-human interface. *Proceedings IEEE International Workshop on Robot and Human Communication*. **1992**. 44–51. <http://doi.org/10.1109/ROMAN.1992.253910>
 26. Donato, G., Bartlett, M. S., Hager, J. C., Ekman, P. & Sejnowski, T. J. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **1999**. 21, 974–989.
<http://doi.org/10.1109/34.799905>
 27. Essa, I. A. & Pentland, A. P. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **1997**. 19, 757–763.
<http://doi.org/10.1109/34.598232>
 28. Claude, F. B., Chibellushi, C., Facial Expression Recognition: A Brief Tutorial Overview, *Online Compendium of Computer Vision*. **2003**. doi:ND
 29. Kanade, T., Cohn, J. F. & Yingli Tian. Comprehensive database for facial expression analysis. **2002**. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. 46–53. <http://doi.org/10.1109/AFGR.2000.840611>
 30. Huang, Y., Liu, Q. & Metaxas, D. A Component Based Deformable Model for Generalized Face Alignment. **2007**. *IEEE 11th International Conference on Computer Vision (IEEE, 2007)*. 1–8.
<http://doi.org/10.1109/ICCV.2007.4409017>
 31. Yu, X., Huang, J., Zhang, S., Yan, W. & Metaxas, D. N. Pose-Free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model. **2013**. *IEEE International Conference on Computer Vision (IEEE, 2013)*. 1944–1951. <http://doi.org/10.1109/ICCV.2013.244>
 32. Baltrusaitis, T., Robinson, P. & Morency, L.-P. Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild. **2013**. *IEEE International Conference on Computer Vision Workshops (IEEE, 2013)*. 354–361. <http://doi.org/10.1109/ICCVW.2013.54>
 33. Sun, Y., Wang, X. & Tang, X. Deep Convolutional Network Cascade for Facial Point Detection. **2013**. *IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2013)*. 3476–3483.
<http://doi.org/10.1109/CVPR.2013.446>
 34. Zhou, E., Fan, H., Cao, Z., Jiang, Y. & Yin, Q. Extensive Facial Landmark Localization with Coarse-to-Fine Convolutional Network Cascade. **2013**. *IEEE International Conference on Computer Vision Workshops (IEEE, 2013)*. 386–391. <http://doi.org/10.1109/ICCVW.2013.58>
 35. Xudong Cao, Yichen Wei, Fang Wen & Jian Sun. Face alignment by Explicit Shape Regression. **2012**. *IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2012)*. 2887–2894.
<http://doi.org/10.1109/CVPR.2012.6248015>
 36. Cristinacce, D. & Cootes, T. Automatic feature localisation with constrained local models. *Pattern Recognition*. **2008**. 41, 3054–3067. <http://doi.org/10.1016/J.PATCOG.2008.01.024>
 37. De la Torre, F. & Minh Hoai Nguyen. Parameterized Kernel Principal Component Analysis: Theory and applications to supervised and unsupervised image alignment. **2008**. *IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2008)*. 1–8.
<http://doi.org/10.1109/CVPR.2008.4587523>
 38. Rapp, V., Senechal, T., Bailly, K. & Prevost, L. Multiple kernel learning SVM and statistical validation for facial landmark detection. **2011**. *Face and Gesture (IEEE, 2011)*. 265–271.
<http://doi.org/10.1109/FG.2011.5771409>
 39. Tong, Y., Liu, X., Wheeler, F. W. & Tu, P. H. Semi-supervised facial landmark annotation. *Computer Vision and Image Understanding*. **2012**. 116, 922–935.
<http://doi.org/10.1016/J.CVIU.2012.03.008>
 40. Belhumeur, P. N., Jacobs, D. W., Kriegman, D. J. & Kumar, N. Localizing Parts of Faces Using

- a Consensus of Exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **2013**. 35, 2930–2940. <http://doi.org/10.1109/TPAMI.2013.23>
41. Dantone, M., Gall, J., Fanelli, G. & Van Gool, L. Real-time facial feature detection using conditional regression forests. **2012**. *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012). 2578–2585. <http://doi.org/10.1109/CVPR.2012.6247976>
 42. Zhanpeng Zhang, Wei Zhang, Jianzhuang Liu & Xiaoou Tang. Multiview Facial Landmark Localization in RGB-D Images via Hierarchical Regression with Binary Patterns. *IEEE Transactions on Circuits and Systems for Video Technology*. **2014**. 24, 1475–1485. <http://doi.org/10.1109/TCSVT.2014.2308639>
 43. Sukno, F. M., Waddington, J. L. & Whelan, P. F. *Springer, Berlin, Heidelberg*. **2012**. 32–41. http://doi.org/10.1007/978-3-642-33863-2_4
 44. Xiangxin Zhu & Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. **2012**. *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012). 2879–2886. <http://doi.org/10.1109/CVPR.2012.6248014>
 45. Xiong, X. & De la Torre, F. Supervised Descent Method and Its Applications to Face Alignment. **2013**. *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2013). 532–539. <http://doi.org/10.1109/CVPR.2013.75>
 46. Cootes, T. F., Edwards, G. J. & Taylor, C. J. *Springer, Berlin, Heidelberg*. **1998**. 484–498. <http://doi.org/10.1007/BFb0054760>
 47. Matthews, I. & Baker, S. Active Appearance Models Revisited. *International Journal of Computer Vision*. **2004**. 60, 135–164. <http://doi.org/10.1023/B:VISI.0000029666.37597.d3>
 48. Baker, S. & Matthews, I. Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*. **2004**. 56, 221–255. <http://doi.org/10.1023/B:VISI.0000011205.11775.fd>
 49. Gross, R., Matthews, I. & Baker, S. Generic vs. person specific active appearance models. *Image and Vision Computing*. **2005**. 23, 1080–1093. <http://doi.org/10.1016/j.imavis.2005.07.009>
 50. Kazemi, V. & Sullivan, J. One millisecond face alignment with an ensemble of regression trees. **2014**. *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2014). 1867–1874. <http://doi.org/10.1109/CVPR.2014.241>
 51. Hastie, T., Tibshirani, R. & Friedman, J. H. Jerome H. The elements of statistical learning: data mining, inference, and prediction. **2009**. *springer series in statistics*, New York City, USA. doi:ND
 52. Eishenthal, Y., Dror, G. & Ruppel, E. Facial Attractiveness: Beauty and the Machine. *Neural Computation*. **2006**, 18, 119–142. <http://doi.org/10.1162/089976606774841602>