IMAGINATION

Image-based Navigation in Multimedia Archives

FP6 - 034626

Deliverable

WP 6: Image Recognition Algorithms
D11
Report on Image Recognition Algorithms

Dr. Christian Küblbeck

April 30th, 2008

EXECUTIVE SUMMARY

This deliverable deals with the algorithms developed in the field of image recognition. It covers especially the algorithms for face detection and for face recognition. Since these algorithms are working with so-called models that means templates that are used for the detection, they can be extended to other objects as long as those object are of a typical structure.
Face detection and identification are playing a significant part within the imagination project in order to automate the process of annotation and the creation of knowledge. These tools are combined with person segmentation tools and text mining tools in order to enhance metadata information for images either they have or don't have textual captions.

Document Information

IST Project Number	FP6 – 034626	**Acronym**	IMAGINATION
Full title	Image-based Navigation in Multimedia Archives		
Project URL	http://www.imagination-project.org/		
Document URL			
EU Project officer	Claude Poliart		

Deliverable	**Number**	11	**Title**	Report on Image Recognition Algorithms
Work package	**Number**	6	**Title**	Automatic Metadata Generation

Date of delivery	**Contractual**	M 24	**Actual**	30-April-08
Status	version. 1.0		final ☑	
Nature	Prototype ☐ Report ☑ Dissemination ☐			
Dissemination Level	Public ☑ Consortium ☐			

Authors (Partner)	Fraunhofer IIS			
Responsible Author	Christian Küblbeck		**Email**	Christian.kueblbeck@iis.fraunhofer.de
	Partner	IIS	**Phone**	+49-9131-776-5160

Abstract (for dissemination)	This deliverable deals with the algorithms developed in the field of image recognition. It covers especially the algorithms for face detection and for face recognition. Since these algorithms are working with so-called models that means templates that are used for the detection, they can be extended to other objects as long as those object are of a typical structure. Face detection and identification plays a significant part within imagination in order to automate the process of annotation and the creation of knowledge. These tools are combined with person segmentation tools and text mining tools in order to get the maximum information out of images with textual captions.
Keywords	Face detection, face recognition

Version Log			
Issue Date	**Rev No.**	**Author**	**Change**
<dd-mmm-yy>	<nnn starting 001>	<author name>	<Description of the changes that were made to the preceding revision>

01-Apr-08	**001**	KUE	Initial version
20-Apr-08	**002**	KEU	Incorporated internal review comments
30-Apr-08	**003**	Nagypal	final formatting

Project Consortium Information

Partner	Acronym	Contact
FZI Forschungszentrum Informatik an der Universität Kalrsruhe	FZI	**Clemens van Dinther** FZI Forschungszetrum Informatik Haid-und-Neu-Str. 10-14 76131 Karlsruhe Germany email: dinther@fzi.de phone: +49 721 9654 862 fax: +49 721 9654 863
Fraunhofer Institute for Integrated Circuits	IIS	**Dr. Christian Küblbeck** Fraunhofer-Institut für Integrierte Schaltungen IIS Am Wolfsmantel 33 91058 Erlangen Germany email: christian.kueblbeck@iis.fraunhofer.de phone: +49 9131 776 549 fax: +49 9131 776 588
R un I Demo	RIDemo	**Edvins Snore** R un I Demo Brivibas iela 183 Riga, LV1012 email: edvins@ri.lv phone: +3717378155 fax: +3717378130
Jozef Stefan Institute	JSI	**Dunja Mladenic** Jozef Stefan Institute Jamova 39 1000 Ljubljana Slovenia email: dunja.mladenic@ijs.si phone: +386 1 4773900 fax: +386 1 4773131
Photos12	Photo12	**Aline Lauvergeon** PHOTOS 12 SA 33, avenue du Maine 75015 Paris France email: aline@photo12.com phone: +33 1 56 80 14 42 fax: +33 1 56 80 14 41
National Technical University of Athens	NTUA	**Yannis Avrithis** Image, Video and Multimedia Systems Laboratory School of Electrical and Computer Engineering National Technical University of Athens 9 Iroon Polytechniou Str. 157 73 Athens Greece email: iavr@image.ntua.gr phone: +30 210 772 4350 fax: +30 210 772 2492

Partner	Acronym	Contact
disy Informationssysteme GmbH	disy 	**Dr. Wassilios Kazakos** disy Informationssysteme GmbH Erbprinzenstr. 4-12, Eingang B 76133 Karlsruhe email: kazakos@disy.net phone: +49 721 16006260 fax: +49 721 1600605
Biblioteca di Storia Moderna e Contemporanea **Italian State Library of modern and contemporary history**	BSMC 	**Dr. Rossella Caffo** Biblioteca di Storia Moderna e Contemporanea Via Michelangelo Caetani, 32 00186 Roma Italy email : rossella.caffo@bsmc.it phone: +39 06 69291723 fax: +39 06 68807662

LIST OF KEY WORDS/ABBREVIATIONS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The following paragraphs describe in a specific way what kinds of algorithms have been used in the field of object recognition. Object recognition means in a first step detecting a certain object type and probably in a second step assigning a specific subtype to the object. The type of objects mainly focused on within Imagination is the human face. In this context the meaning of the subtype classification is for example a gender classification or the identification of the person.

2 FACE AND OBJECT DETECTION

The approach chosen to detect objects and especially faces is based on an approach outlined below, but described in more detail in **[1]**.

The principle of the used classification mechanisms can be applied for the detection of any type of object as long as the object consists of a typical structure. The information about the structure of the object type is stored in a so called *model*. The model is built from a large number of typical representatives that are adjusted manually by marking typical points (like the eyes in a face). These representatives then are cut out in a pre-given size (for example 24 pixels x 24 pixels for face detection).

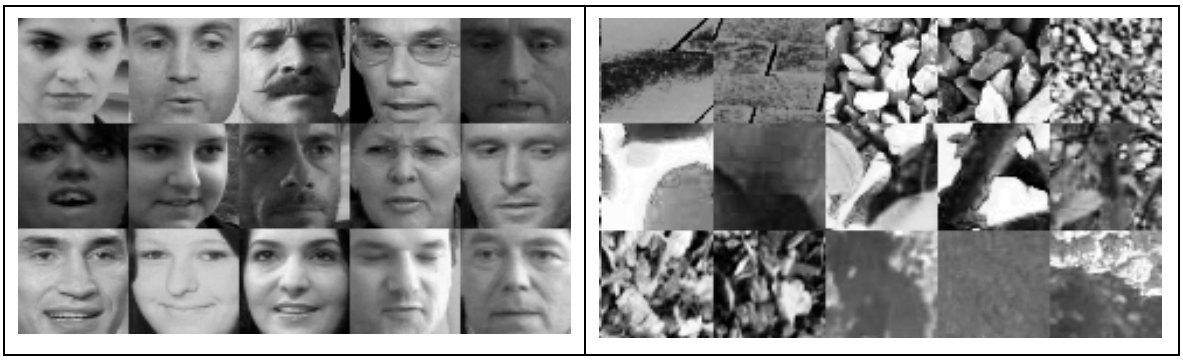

Figure 1 Typical samples of faces and non-faces. The faces all are aligned in a way that the eye positions are the same within each image. Thousands of typical representations of each object are needed in order to train the model.

2.1 Training

We have implemented a so called feature pool that extracts a total of approximately 30 000 features all over the region. These features contain so called Census features **[3]**, gradient features and local structure features. The features are selected in such a way that they are robust against variations in lightning as shown in Figure 2.

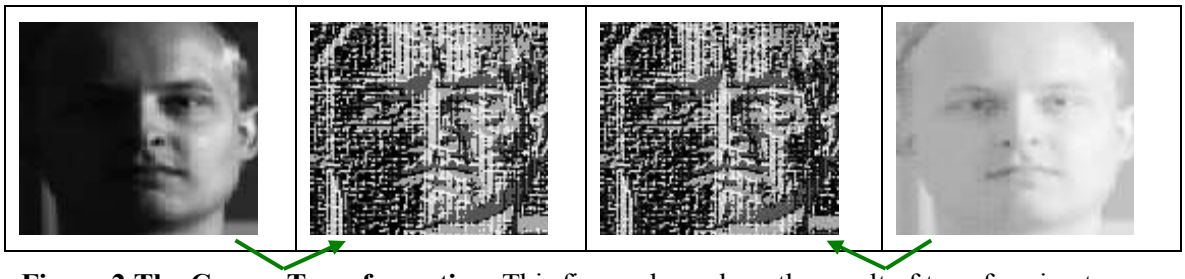

Figure 2 The Census Transformation. This figure shows how the result of transforming two significantly differently illuminated images (leftmost and rightmost face) result in very similar transformation results (left middle and right middle)

Starting from these 30 000 features then in a training process those features are selected that are most feasible are selected with respect to their ability to distinguish between the given samples. The selection and training process is conducted using the Adaboost Algorithm that is described more precisely in **[1]**. A schematic view of the whole training process can be seen in Figure 3.

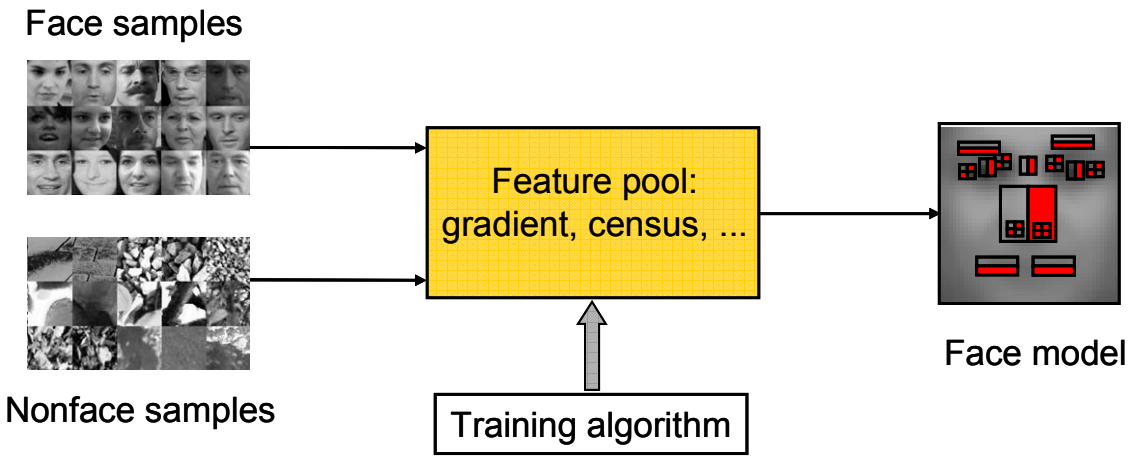

Figure 3 Training of the face Model. From a feature pool those features are selected that enable a maximum separation between the different sample classes. That from the face model is built.

2.2 Classification

During the classification process each image is scanned using the model that is shifted all over the image. Hence each image region has to be compared to the model. In addition several downscaled images of different size have to be analyzed. This is necessary because the model consists of a fixed size but it is the intention to find faces of different size.

In order to improve the speed of the classification process the model is built as a multi-stage cascade as described in Figure 5.

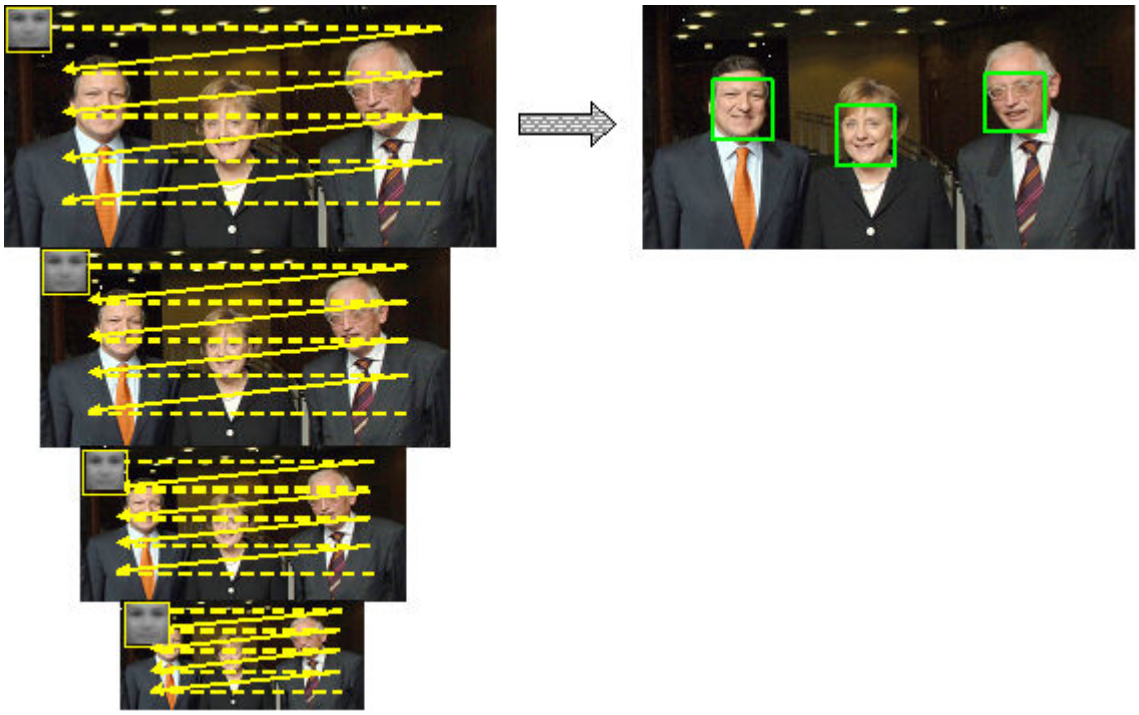

Figure 4 Scanning Images for Faces. Each location of the image has to be searched as well as subsequent down-scaled images.

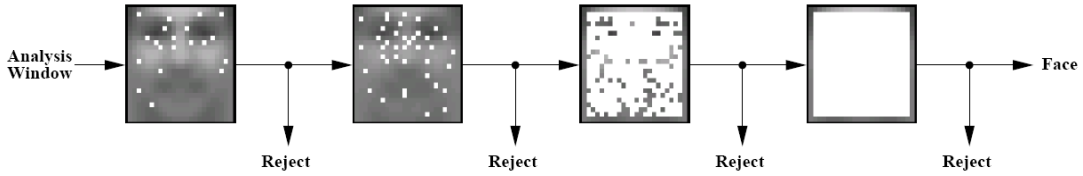

Figure 5 Classification Cascade. Each window is at first analysed with a very small number of features. In this way a huge amount of windows can already be rejected while a small amount of face candidate regions are further analysed with more features. This process is performed several times and results in a great speed-up of the algorithm.

2.3 Gender Classification

Gender Classification works in principle in the same way as face detection does. We build a model from a huge training set showing male and female faces. From that training data a model is derived that consists of that selected features that are capable of distinguishing the two classes. With this model a detected face then can be classified as male or female. At the moment the gender classification works best with European faces, since the training images were selected from mainly European faces. On the BioID dataset **[4]** this model achieves a recognition rate of 90%.

3 FACE IDENTIFICATION

Identifying a face means to assign a certain ID to the face which is stored in some kind of database and that corresponds to an already know identity. As soon as the face detection has been able to detect one or more faces within an image, the next step of the face identification is started to propose which face belongs to which person.

The challenge within the imagination project is that the faces are not always orientated exactly frontally — the pose of each face may vary within a quite huge range (about ±45° out-of-plane rotation). Hence the idea of estimating in a first step the pose of the head and estimating the full frontal face from the estimated pose in beginning of the identification step appears to be promising although the implementation is sophisticated and quite complex.

Another approach is to use a window based approach in which features are calculated. The feature values can then be accumulated in histograms and these histograms can then be used for training and classification. This approach was used within the imagination project and is explained in more detail in following paragraphs.

The training process for identification is similar to that for detection. Again a general model is built using two opposite classes, consisting of a large pool of face images. Image pairs are built on the one side from the same person and on the other side from different persons respectively (see Figure 6). On each image pair a "feature pool of face regions" is calculated using a histogram based approach that counts the outcomes of feature calculation within certain regions of the face. Using this method results in greater robustness against varying poses. The modified features then are used to train an identification model in the same way as explained above for face detection. For the training of this generic model a huge amount of approximately 1000 face pairs for each class is needed.

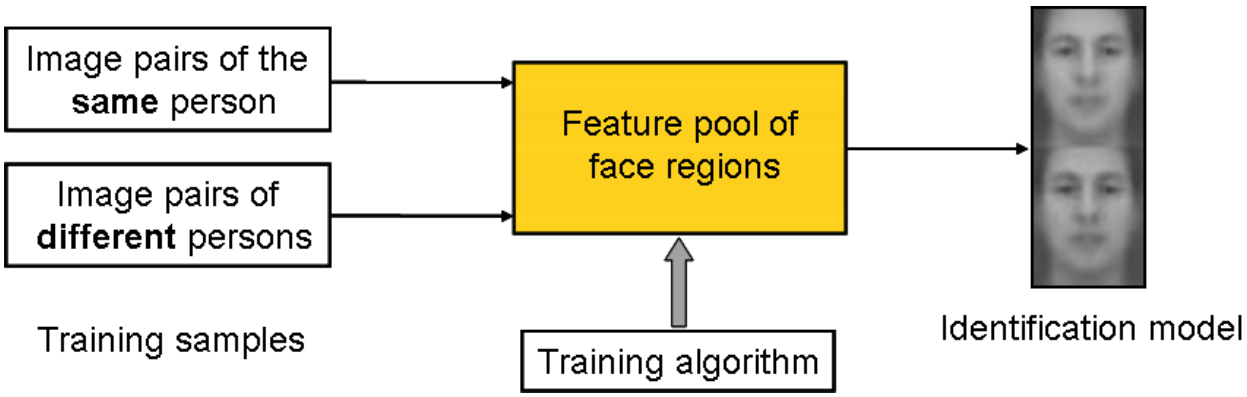

Figure 6 Training of the Identification Model. A huge set of image pairs (pairs of the same person and pairs of different persons) is used for training. Compared to the training for face detection an adapted feature pool of face regions is used. The rest of the training is working in the same way as the training for face detection.

The identification works as following: a probe (the face where we want to get the ID from) is stitched together with each of the faces stored in the gallery of know faces. Each thus created face pair is evaluated using the identification model. The out-coming

score is a similarity measure between the probe and the gallery face. Thus an ID or a ranking of the best 5 IDs can be created (see also Figure 7). Hence in order to identify a certain face, at least one sample of the face has to be in the gallery, but of course several faces at differing poses will increase the robustness.

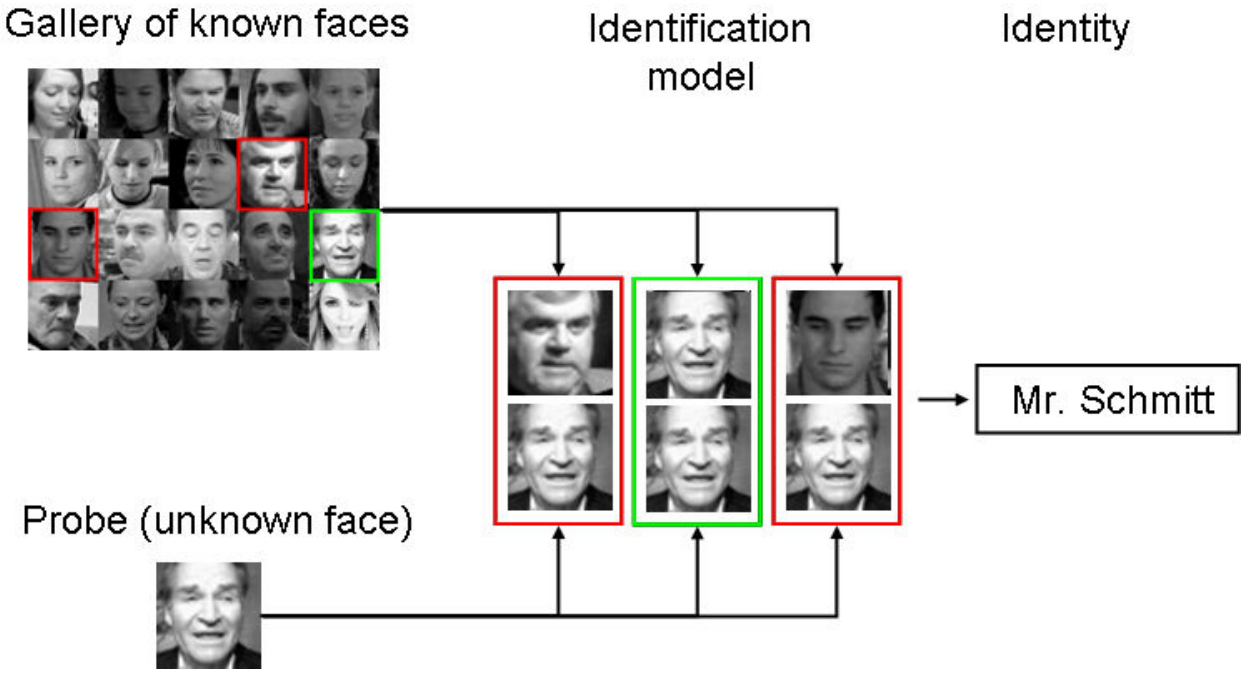

Figure 7 Identification Process. At first pairs of images are created using the probe and the gallery of known images. With the identification model each pair is investigated and the returned score indicates which identity the probe belongs to.

4 CONCLUSION

Algorithms for fast and robust face detection and face identification have been implemented. First results show that especially the face detection algorithms perform very well on a broad variety of images and hence are especially suitable in the context of the Imagination project. The implemented identification algorithms show up to also perform quite well, but their effectiveness has to be evaluated further on.

REFERENCES

[1] Yoav Freund and Robert E. Shapire. A short introduction to boosting. In Journal of Japanese Society for Artificial Intelligence, number 14, pages 771–780, September 1999.

[2] Christian Kueblbeck, Andreas Ernst: Face detection and tracking in video sequences using the modified census transformation, Image and Vision Computing, Volume 24, Issue 6, June 2006, Pages 564-572

[3] Ramin Zabih and John Woodfill. A non-parametric approach to visual correspondence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996.

[4] The BioID dataset. It can be downloaded from http://www.bioid.com/downloads/facedb/index.php