# A Practical Case Study: Face Recognition on Low Quality Images Using Gabor Wavelet and Support Vector Machines

**Enrique D. Martí · Miguel A. Patricio · J. M. Molina**

**Abstract**  Face recognition is a problem that arises on many real world applications, such as those related with Ambient Intelligence (AmI). The specific nature and goals of AmI applications, however, requires minimizing the invasiveness of data collection methods, often resulting in a drastic reduction of data quality and a plague of unforeseen effects which can put standard face recognition systems out of action. In order to deal with this, a face recognition system for AmI applications must not only be carefully designed but also subject to an exhaustive configuration plan to ensure it offers the required accuracy, robustness and real-time performance. This document covers the design and tuning of a holistic face recognition system targeting an Ambient Intelligence scenario. It has to work under partially uncontrolled capturing conditions: frontal images with pose variation up to 40 degrees, changing illumination, variable image size and degraded quality. The proposed system is based on Support Vector Machine (SVM) classifiers and applies Gabor Filters intensively. A complete sensitivity analysis shows how the recognition accuracy can be boosted through careful configuration and proper parameter setting, although the most adequate setting depends on the requirements for the final system.

**Keywords**  Face recognition · Gabor filters · Support vector machine · Implementation · Low quality image

E. D. Martí (✉) · M. A. Patricio · J. M. Molina
Applied Artificial Intelligence Group, University Carlos III de Madrid, Colmenarejo, Spain
e-mail: emarti@inf.uc3m.es

M. A. Patricio
e-mail: mpatrici@inf.uc3m.es

J. M. Molina
e-mail: molina@ia.uc3m.es

# 1 Introduction

Some voices claim against face recognition as a biometric for critical systems due to its low accuracy. But in the last decade, research on Ambient Intelligence (AmI) has brought non-critical person identification into focus [21].

Whilst face recognition accuracy is still below other biometrics as iris or fingerprint recognition, it is a very suitable approach to avoid intrusiveness, either as a stand-alone method or to be fused in a multimodal system.

But non-intrusiveness is not for free. It usually means that the available face images will suffer from poor quality (low definition, lack of color fidelity and others) and variable resolution, a problem analyzed in [12].

In those cases where the enemy to beat is the lack of resolution, reconstruction techniques have demonstrated to be a suitable approach. For instance, Face Hallucination [1] generates super-resolution images from a small input face, using a system especially trained for preserving (or recreating) facial features. Several hallucination algorithms [15, 27] have been developed since that pioneer work was published. The empirical results show that the generated images improve face recognition accuracy.

Hallucination recreates the lost high frequency components by exploiting the knowledge about human facial structure acquired during training. This could be a problem, because the added information is invented based on what we have learned from a reduced dataset: compared with low-resolution input images, the appearance of hallucinated faces is more similar to those in the training dataset, thus increasing the probability of false positives. Another negative side of this technique is that, in some cases, the reconstructed images show artifacts and deformations which lead to misclassifications during recognition.

An alternative technique related with reconstruction consists on synthesizing 3D views of the training dataset faces, from one [14] or several 2D images. The recognition error rate when classifying profile poses is about half the obtained without using this technique [20].

Back to our smart-environment scenario, the major obstacle for seamless face recognition is not resolution but rather the capturing conditions. They are uncontrolled –variability in face lightning, poses and size–, may gradually change over time –dust in the camera lens can lead to noise and unfocused images– and are heterogeneous due to the characteristic mix of devices in cooperative environments. Furthermore, the hardware for acquiring and processing the data is very likely to have restricted computational power – [21] analyzes and evaluates the face recognition performance of different wearable devices, including mobile phones–.

Looking at the big picture, this problem can be defined as non-critical face recognition using images around $100 \times 100$ pixels with degraded quality and highly variable face attributes, using low power hardware.

Computer-based face recognition has reached quite satisfactory results in the last years thanks to the efforts of an active community, even surpassing human capacities with non-familiar faces [19]. Reviewing the existing literature, it is not difficult to find a number of algorithms suitable for building the proposed system. Nonetheless, the implementation of a high-end face recognition system satisfying all the requirements exposed above is a costly project.

Now, let us raise a question: taking into account the performance gain margin and the requirements for this concrete scenario, does building the best possible

system worth the associated costs? Our hypothesis is that we can use one among the largely studied simpler approaches, and still obtain good results just by fine-tuning the algorithm.

On this paper we tackle the practical task of configuring a 2D face recognition system (based on a SVM classifier and a bank of Gabor filters) under partially controlled conditions when the image quality is degraded by a wide variety of image artifacts due to light configuration, movement and compression.

Since the publication of Dynamic Link Architecture [13], Gabor filters have demonstrated to be a valuable tool for facial identification. A large number of researches point that it increases the robustness of pattern recognition algorithms further than many other information transform methods [6, 9, 27].

The decision of using an SVM as classifier is based on the performance it has demonstrated with this kind of problems in several researches [7, 22, 28]. Furthermore, it reduces the vast experimental workload of this project thank to being an almost auto-configurable algorithm.

Once the architecture is selected, the configuration parameters will be exhaustively analyzed in order to know how they affect the result and extract the whole potential of the system.

An important decision is how to use the available face information, either in a component-based or a holistic (the image as a whole) approach. Component based classifiers are superior over holistic in the general case [7], but they require (a) a greater development effort and (b) to take into account contingencies as occluded or lost facial features that could spoil their performance. In our case, it is impossible to identify in advance all the factors causing variations in the captured face images, and so a holistic classifier is a better option thanks to their independence of punctual traits and features.

The two following sections of this document will introduce briefly the main algorithms used in our research: Gabor wavelet and SVM classifiers.

A third section, depicted to face datasets used in the experiments, puts special attention in assuring that they correctly mimic real conditions, as their fidelity is key for experimental validity. The design of the face recognition system will be detailed straight afterwards, followed by the experiments performed on it as well as the result obtained. Finally, the last section encloses the conclusions extracted from the experimental part.

## 1.1 Gabor Wavelet Filters

The idea behind any pre-process is to transform the input into something more adequate for solving the problem. In signal analysis, it is known that information available in the frequency domain is quite important; this information is intrinsic to the structure of the raw signal, but it is difficult to work with it directly. It is often necessary to apply different mathematical tools to obtain a synthesized representation of it.

Fourier transform is probably one of the best examples of a tool for transforming data into the frequency domain. It takes a signal in the time domain (for instance, a sound record) and calculates its decomposition in an infinite sum of simple sinusoidal waves of different period: this gives us perfect information about the frequency components of the whole signal, but completely delocalized in time.

Instead of using sinusoidal components for decomposing the input, we can use other signals providing also spatial information (i.e. are located in the space). A Gabor wavelet (Fig. 1) satisfies these criteria.

When a Gabor wavelet is convolved with an input signal (i.e. a face image), we obtain a map with the "amount" of a certain frequency (that of the wavelet) at each point of the signal, as denoted in Fig. 2. A very important feature of Gabor wavelets is that they equal the uncertainty about obtained spatial and spectral information [5].

Vision in mammals has been proven to be related with Gabor transform [8, 11], due to the distribution and response of retinal cells (bars and cones) to the light. Artificial vision and general pattern recognition algorithms as [10, 18, 26] have also been benefited by the use of Gabor transform, as its worthy output contains much more information than raw images in the same amount of data.

A Gabor filter is characterized by several parameters as orientation or frequency. When it is applied to an image, the result describes the input in terms of that particular configuration. For this reason, Gabor wavelets have to be grouped in banks covering all the interesting combinations of parameters in order to obtain a complete description of the input image.

As it was previously said, the pioneer work on Gabor wavelets applied to face recognition was the Dynamic Link Architecture, which used a bank of 40 filters of $128 \times 128$ pixels, covering 5 different frequencies (consecutive frequencies separated a single octave) and 8 orientations (one each 45° degrees) that rapidly became a *de facto* standard [17, 30]. We will analyze whether this is actually the best available option, or the useful information about extra-personal variation is distributed on any other way.

## 1.2 Support Vector Machine Classifier

SVM is a binary maximum-margin linear classifier [4]. It tries to calculate the hyperplane that separates instances of two different classes, in such way that the distance from the decision surface to the nearest instance of both classes is the
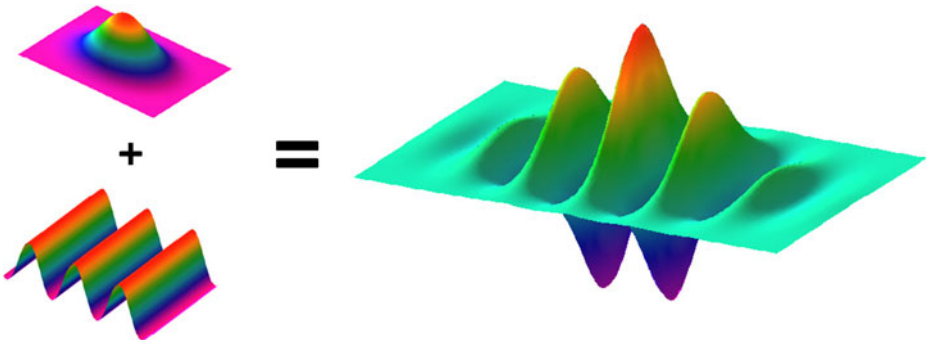


**Fig. 1** 2D Gabor wavelet. It is composed by a 2D Gauss function (for the spatial location) multiplied by a harmonic function (for the frequency analysis; introduces the frequency and orientation parameters). A Gabor wavelet uses a complex function for the harmonic component, though this figure shows only the real part: a cosine
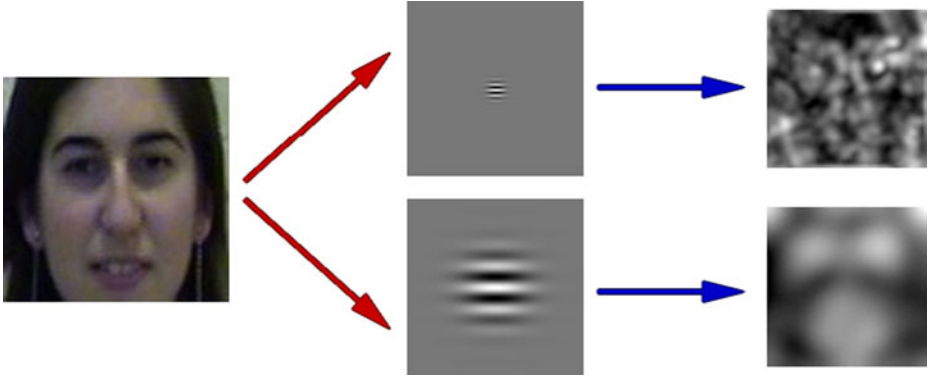
**Fig. 2** Filtering an image (*left*) with Gabor wavelets of different frequencies (*central column*) produce different outputs (*right*). A high frequency filter highlight borders and rapid tone changes, while the low frequency one remarks subtle gradients in wide areas

maximum possible. This way, the structural risk of the classifier is minimized [24], increasing the probability of correctly categorizing unseen instances.

In the problem to solve, we initially have a set of *m* elements in an n-dimensional space $x_i \ \varepsilon \ \mathbb{R}^n$, $i=1,\ldots,m$. Each sample belongs to one of the two possible classes (values and respectively), resulting in a dataset $\{x_i, y_i\}$, $y_i \ \varepsilon \{-1,1\}$.

Let $x \cdot w + b = 0$ be the equation of the hyperplane which is the solution of our problem (*x* are the points in the hyperplane, which satisfy the equation above; *w* is the normal to the plane and *b* is the origin). Such hyperplane, for effectively being the solution of the problem, must hold the following conditions:

The instances $x_i$ belonging to the class $y_i$ must satisfy

$$y_i \cdot (x_i \cdot w + b) - 1 \geq 0 \forall i \tag{1}$$

Distance between hyperplane and support vectors has to be maximum:

$$\text{maximize } d = 2/\|w\| \tag{2}$$

With these conditions, the problem has to be solved by moving the separating hyperplane, or what is the same, changing the values of *w* and *b*. This will be expressed as an optimization problem.

If both classes were not separable, the problem would not have a solution. This difficult can be overcome by introducing one "slack variable" $\xi_i$ per sample in the dataset, allowing some of them to lie in the incorrect side of the classification hyperplane at the cost of a certain penalization (controlled by parameter *C*) in the function to optimize.

A formal specification is obtained by combining Eqs. 1 and 2 properly modified to take into account $\xi_i$:

$$\text{Minimize } \|w\|^2 + \sum_i \xi_i \cdot C \text{ subject to :}$$

$$y_i \cdot (x_i \cdot w + b) - 1 \geq \xi_i \forall i \tag{3}$$

This is a quadratic optimization problem that can be reformulated –using the Karush-Kuhn-Tucker conditions and the Wolfe dual formulation– to maximize:

$$L\left(\alpha_i\right) = -\frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j\left(x_i \cdot x_j\right) + \sum_{i=1}^{m}\alpha_i$$

$$\text{Subject to: } 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{m}\alpha_i y_i = 0 \tag{4}$$

Once maximized, the normal vector $w$ can be calculated as:

$$w = \sum_{i=1}^{m}\alpha_i y_i x_i \tag{5}$$

SVMs can be extended to deal with non-linear problems, but the linear formulation is enough for us, as it will be shown later.

## 1.3 Datasets

Three different databases have been used for development and testing phases: the first of them was specifically created for this research, with the purpose of being used in the majority of the experiments; the second one is an external face database which image quality somehow resembles the first one, in order to be used as a set of unknown individuals (for experiments related with false positives); the third and last one is the Extended Yale Face Database B, on its cropped version, used to compare our results with previous papers (a kind of test for the validity of the implementation).

Right below these lines, the aforementioned databases will be briefly described.

### 1.3.1 Self-made Database

A set of face images of the staff of the GIAA department at the campus of Colmenarejo –they will be the subjects to be ultimately recognized by the system–. Video recording was used to register a great amount of images of each individual, under conditions very similar to those of the real scenario.

The shot was taken with day light plus artificial light lamps, creating lighting conditions very similar to those present in the real scenario; we allowed great variations in facial expressions and moderate in pose.

After processing the video, we obtained a total of 4293 images for the 17 members of the lab. The number of images for each person is not equally distributed: it varies from 26 to 754, although 14 of the members are in the range 100-300 images. Their dimensions are variable; face images go from less than 80 to approximately 240 pixels high. These results were obtained applying an open source face detection tool [2].

The frames were captured with a digital video (DV) camera, recording 25 frames per second with a total frame size of $720 \times 576$ pixels (although the target faces are more than 3 meters distant, making them much smaller); some of the images have different types of artefacts such as interlaced field problems, where the even lines are horizontally displaced with respect to the odd, as seen on Fig. 3.

The frames were split into three separate sets of data (sample images in Fig. 4) according to their quality, in order to better describe the behaviour of the system:

**Fig. 3** Still images from an interlaced field video suffer from loss of high frequency details and duplication of some features –as in the third image, with a vertical displacement

*Easy* Contains images with tiny pose variation and no difficult expressions, for a total of 1756 faces.

*General* All expressions are allowed, as well as pose variations up to 40°. This angle is close to the limit were some features can be occluded, and others –as the nose– have a completely different appearance). The amount of images ascends to 1777.

*Extreme* Encloses all the images outside the previous categories. Part of them are images that can be considered very similar to those in the general set, but other images are face captures totally outside of the scope of the project: profile views, severely occluded faces and wrong-framed faces amongst others. It contains a total of 760 images.

Additionally, a training dataset was extracted from the Easy subset. It contains 7 to 20 images for each individual for a total of 185 faces.

Each one of the sets have their particular purpose: the first one is simply a general validation set, whilst the second one is expected to be much more similar to images in a real scenario. The extreme set will be used just to push the classifiers to the limit of their capabilities, and for very particular tests in an image-per-image.
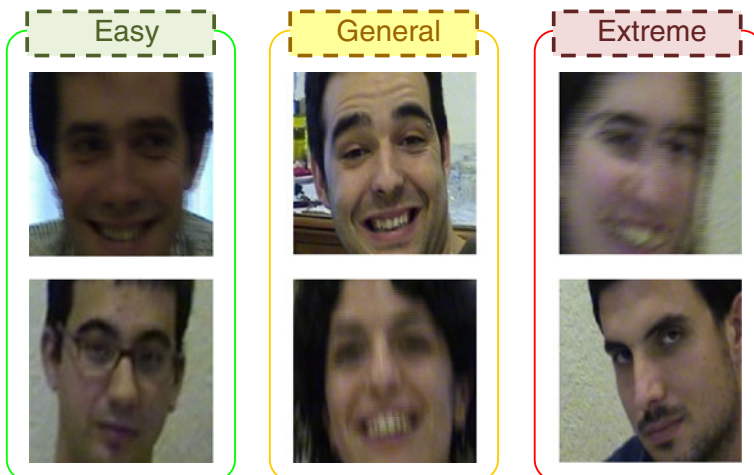


**Fig. 4** Typical face images composing each of the three subsets of our self-made database

### 1.4 Dr. Libor Spacek Face Database

Compiled by Dr. Spacek for his research activity, and made public on its personal webpage at the Essex University portal[1] to "encourage competitive research".

We are interested in the subset "face95", consisting in 20 images for each of the 72 individuals (a total of 1440 faces). Images are $180 \times 200$ pixels, 24-bit colour. Our face detector rejected some of the images, so we finally obtained 946 samples unevenly distributed amongst the 72 persons.

### 1.5 Extended Yale Face Database B "Cropped"

Hosted by the Comp. Sci. and Eng. Department at U.C. San Diego, this project extends the original Yale Face Database B. It contains the same 585 images per individual (9 camera positions, 65 different lighting configurations), but adding 28 new persons to the 10 in the original set. The final database contains 16128 new images.

We have used the "cropped version",[2] which has only 65 images of each individual (the 64 illuminations for the pure frontal pose plus an additional image with the ambient as unique source of light) and each image has been reduced to contain only the face of the person on it.

### 1.6 System Architecture

The proposed face recognition system follows the simple architecture shown in Fig. 5. The purpose of this work consists in adjusting each part to obtain the best results. Input images receive several pre-processing operations –resizing, face detection, histogram equalization–, and then are filtered with the bank of Gabor wavelets just before being used to feed the SVM classifier.

The features resulting from applying a bank of Gabor filters to a single image are usually reported as having a huge dimensionality and, thus, representing an unmanageable load for latter steps. Some efforts have been made to decrease the amount of information to be processed, either by downsampling (scaling) the resulting features, by performing a dimensionality reduction technique such as PCA [16] or by selecting the most interesting features as in [23].

Our election has been downsampling, as the size is not a limiting factor in our system: we will show that the information loss does only have negative effects when using images below $8 \times 8$ pixels.

The second step of dimensionality consists in combining in a per-pixel basis the features obtained with filters of similar frequency and different orientation (Fig. 6). This way we are not just reducing the dimensionality by a factor of 8, but also the resulting image will offer improved robustness against minor changes as small deformations and/or rotations.

---

[1]http://dces.essex.ac.uk/mv/allfaces

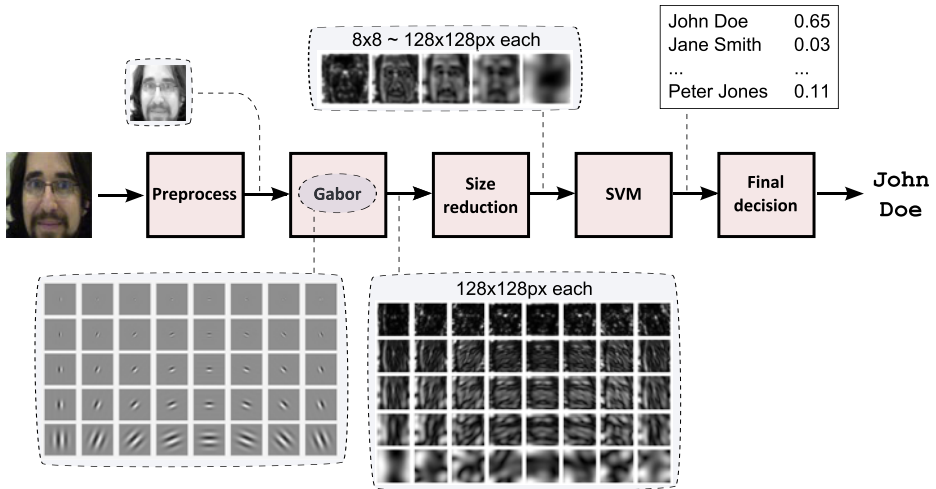[2]http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html

**Fig. 5** Schema showing all the phases since an image is taken until it is presented to the SVM classifier

The use of any other technique is not necessary: even with ten thousand dimensions per sample, the training of a SVM classifier with 300 images of 17 individuals is completed within a minute in a medium-end modern computer. Classification can be performed at a pace of several tens of samples per second. All the experiments and results of this paper were obtained using a custom modified version of LIBSVM [3] library for Support Vector Machines.
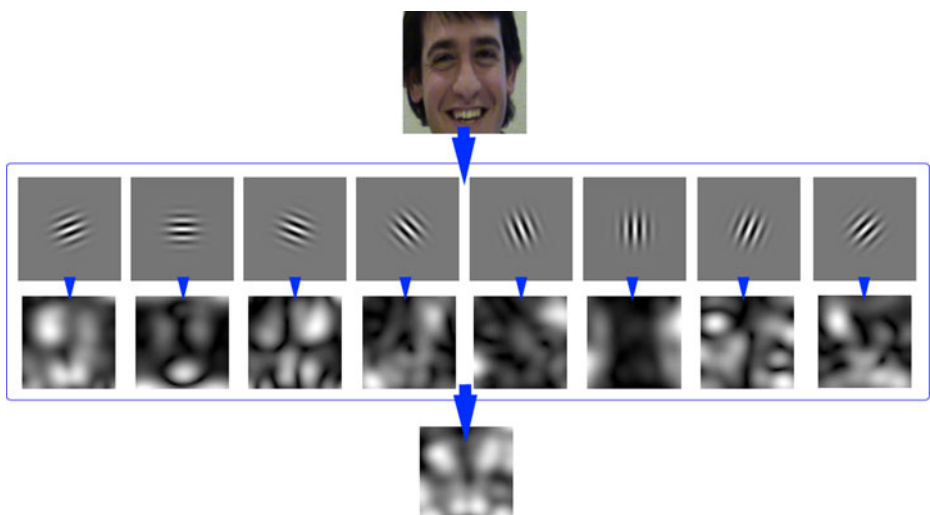


**Fig. 6** Gabor features combined using L2 norm

1.7 Experimental Results

Experimental part has two clear goals. In first place, it has to perform a sensitivity analysis to help determining how changes on each configuration variable affect the behavior of the system. The experiments are also focused in evaluating the global effective performance of the proposed system under those conditions assumed for an AmI scenario.

The sensitivity analysis treats each part of the face recognition system individually, studying no more than one variable at a time. Figure 7 shows the complete plan, where the three first steps are depicted to explore the different configurable parameters.

In the fourth step, the whole system performance is reinterpreted by introducing unknown individuals that should be rejected in a real scenario. Finally, the experiments of previous papers will be reproduced to establish a comparison with other results.

As a clarification regarding the notation: when describing input image formats, "raw" is referred to face images that have not been filtered with a bank of Gabor wavelets, and the pair of numbers following it (e.g. "16 × 16") makes reference to the size in pixels of the image. On the other hand, "gabor" is usually followed by three numbers ("32 × 32 × 5") which meaning is that input images have been processed with a bank of Gabor wavelets with a certain number of frequencies (following with the provided example, 5 frequencies) and then have been downsampled to the specified size (32 × 32).

### 1.7.1 1 – SVM Classifier Configuration

We have chosen to adjust the SVM in first place for two reasons: its relevance in the process, and because it has a very small number of parameters –making the process quite simple–.

The very first step is to decide the best kernel for this problem. Experiments reveal that SVM classifiers perform almost identically with independence of the underlying kernel. This is true also for the dot product kernel, which results in a linear classifier, in spite that a non-linear kernel would give a classifier with superior capacity. Appendix B in [4] explains that when the number of features is much larger than the number of samples, it is usually not necessary to map the data to an even higher dimensional space; here we have several thousand features, which is a number
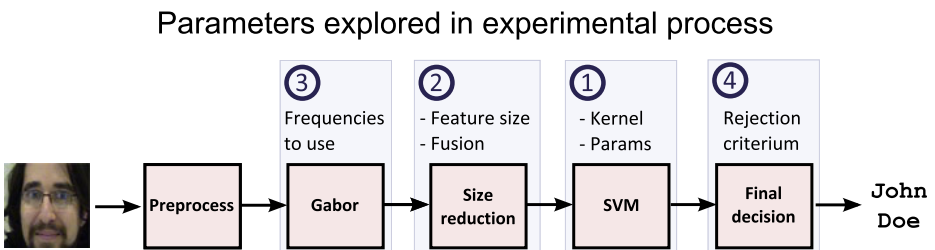
## Parameters explored in experimental process



**Fig. 7** Experimentation plan. Numbers indicate the order in which experiments were executed

clearly higher than our 185 training samples, so linear SVMs will be, in general, enough for our average case.

Once decided that a dot product kernel will be used, we can start with the second part of this experiment: calculate the optimal value for parameter $C$. This variable represents the penalization to be introduced by the slack variables in the non-separable case of SVM classification –see Eq. 3.

Figure 8 shows that it has to be configured with a high value to avoid misclassifications during the training phase. The reason is the same that lead us to use a linear classifier instead of a kernel: the high dimensionality of the input data makes possible to solve the problem with a hyperplane even considering noise in the samples. When the penalization to misclassifications is relaxed, the training phase always results in defective classifiers.

### 1.7.2 2 – Size Reduction Phase

This process was composed by two parts. The first was deciding the downsample ratio for the obtained Gabor features, and the second consists in combining those downsampled Gabor features corresponding to the same frequency on a per-pixel basis –as shown in Fig. 6.

Let us start with the second step. Our results show that this fusion does not only decrease the dimensionality of the resulting data, but also improves classification quality. Regarding the way on which features must be combined, we tested three different criteria: maximum value, average value and L2 norm. The last one offered the best results, followed closely by the average value.

Back to the feature size reduction, Fig. 9 reveals that a smaller size does not necessarily mean a negative lose of information. Instead, Gabor features with a size of $32 \times 32$ help to better discern the identity of a person given its face. This could be somehow related with the low quality of the images, where the fine detail either follows a pattern or consists on noise, but does not provide useful information.

**Fig. 8** Evolution of cross-validation accuracy during training, depending on the value for the penalization parameter C. It can be seen that the optimal setting is different for each image size and type of input, but all of them have in common that from a certain value in advance, the recognition rate is maximum and stable
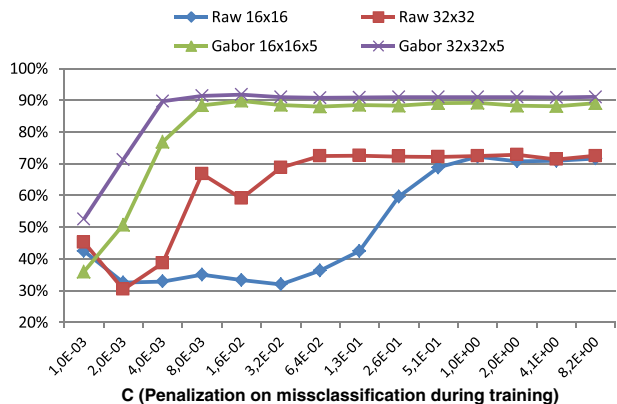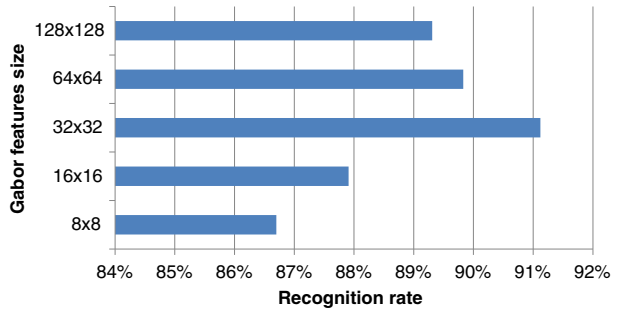
**Fig. 9** Recognition rate over General subset of the self-made database for various sizes of the downsampled Gabor features (using a Gabor wavelet bank covering 5 frequencies). Apparently, a linear downsampling factor of 4, which yields Gabor features with 32 × 32 values, offers the best results, with a 91% of correctly classified instances

*1.7.3 3 – Gabor Wavelet Bank Configuration*

A bank of Gabor wavelet filters can be configured according to a large number of parameters which comprise those related with the morphology of Gabor wavelets and also those describing which filters will compose the bank.

Regarding the first type, our intention has been to fix the maximum amount of them (in fact, all of them except frequency): size 128 × 128 pixels, no ellipticity for the Gaussian envelope, fixed ratio between the sinusoidal component wavelength and the Gaussian envelope size, and 8 orientations.

On their paper, Lades et al. argue that in spite of being 7 the maximum number of frequencies that can be used without incurring in aliasing (on 128 × 128 pixel images, with frequency increments of one octave each time, and according to Nyquist theorem), the Gaussian envelope makes nonsense to use a number over 5.

We have tested how the system performs when 3, 5 and 7 frequencies are used. The results (Fig. 10) show that using 7 frequencies actually provides a noticeable increase in the recognition rate over using 5 frequencies, in every of the 3 tested datasets. Nonetheless, the greatest increase happens when testing the Extreme dataset, which is out of our scope; the difference is practically inexistent in the General subset, the one of greatest interest.



**Fig. 10** Performance of the proposed face recognition system against the three subsets of the self-made database, when the input images are filtered using a bank of Gabor filters covering several numbers of frequencies. Results are also compared with using the raw images directly with the classifier (no Gabor filtering)
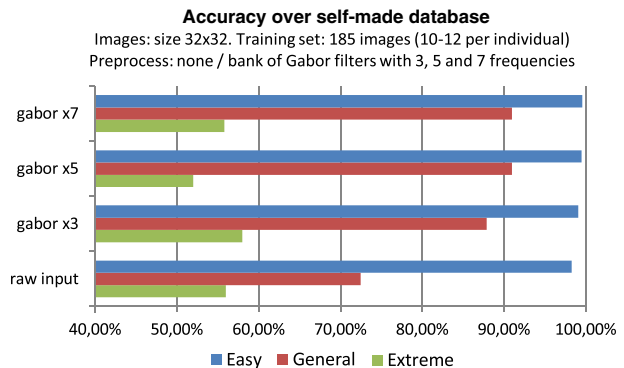
Some researches [25] point to the hypothesis that low frequency features are useful to differentiate between faces and objects that are not faces, while high frequency components contain the information about extra-personal variations (this is, the key to identify persons). In order to evaluate the validity of the hypothesis, two Gabor wavelet banks were built with the three lowest and the three highest frequencies of the 5 frequencies bank respectively. Our experiments show that the version with all the 5 frequencies clearly outperforms the two others, obtaining both banks with 3 frequencies comparable accuracies. This does not invalidate the hypothesis, but rather points us that under the imposed conditions it is better to have all the frequency information available

### 1.7.4 4 – Rejection Criteria, or Classification Confidence

All the previous results were obtained by forcing the classifier to emit a guess about the identity of the person provided as input, no matter if it actually knows it or it is a stranger. Nevertheless, in a real case the classifier has to be configured in such way that it correctly detects when an input image represent an unknown person.

This scenario can be modeled by introducing a new dataset with faces of persons that were not presented during the training phase. We used the database offered by Dr. Libor Spacek because the quality and aspect of the images on it is somehow similar to the self-made dataset.

The output of a SVM is a set of numbers whose meaning is the probability for the input sample to belong to each one of the classes known by the classifier. This way, the class with the largest probability is always the best candidate for classifying the sample, but depending on how large this probability is we can have an idea about how confident the classifier is about its own verdict.

After training a classifier, the procedure for adjusting a rejection threshold is as follows: the dataset of unknown samples is classified using the trained system, saving the probability of the winner class on each case; then, a probability is taken as discriminating threshold, such that only a certain amount of the unknown samples would be accepted as known individuals (false positive). Using this threshold, a validation dataset of known persons is classified.

The different configurations for the Gabor wavelet bank that were explored on previous section are revisited here to confirm the obtained results. Figure 11 analyzes the accuracy of the system depending on the final size of the Gabor features for a bank covering 5 frequencies (only $16 \times 16 \times 5$ and $64 \times 64 \times 5$ are shown for the sake of clarity. $32 \times 32 \times 5$ version obtain a result more or less in the middle of both, and the two extremes, $8 \times 8$ and $128 \times 128$, obtain the poorest results).

It can be seen that, as the figure footnote says, the version with a size of $16 \times 16$ features obtain almost 10% more accuracy than the $64 \times 64$ one, but at the price of having more errors. Nonetheless, in both cases the amount of errors is practically zero –the erroneous predictions are usually discarded as unknown persons due to its low confidence–, and because of this the more downsampled version appears to be the most robust for an access control system.

This result changes what we seen on the first experiments, where the Gabor features with a size of $16 \times 16$ were the second worst option: here, having a large number of features give "cautious" classifiers, which hardly ever fail but easily reject a known person.
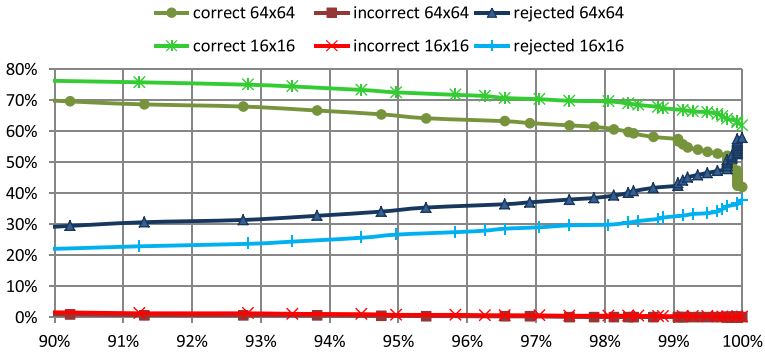
**Fig. 11** Percentage of samples from the General dataset that were correctly classified, incorrectly classified and rejected as not known (curves in the plot, percentage shown in *vertical scale*) depending on the value of the rejection threshold. Instead of indication the threshold explicitly, the horizontal scale shows the fraction of samples from the unknown dataset that has been rejected (*horizontal scale*). The zone of interest for a real scenario is that one over 99% of false positives rejected, where the classifier using $16 \times 16$ Gabor features obtain a 67% recognition rate and barely a 0,35% of errors, against the 57.5%/0.06% for the $64 \times 64$ version

When the number of frequencies is varied, we obtain Fig. 12. Here the difference between versions with 5 and 7 frequencies is null, supporting the theory that maybe the highest of the 7 frequencies are producing aliased Gabor feature vectors (although it is slightly more robust against failure). The accuracy using 3 frequencies is significantly lower, and using the raw images instead of Gabor features is revealed as a completely useless option.

### 1.7.5 Comparison with Previous Results

Using the Extended Yale Face Database B it was possible for us to reproduce the experiments at [29]. Two randomly chosen training datasets were used: the first uses 10 out of the 65 sample images per individual, and the second one increases the number to 30. The whole database is taken as validation set.

**Fig. 12** Classification accuracy when using Gabor wavelet banks covering different numbers of frequencies, varying with the rejection threshold
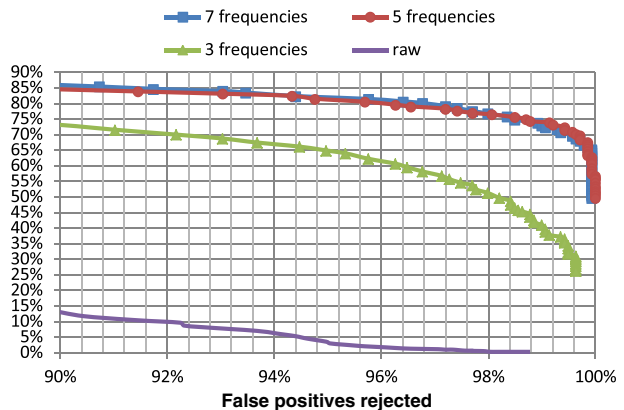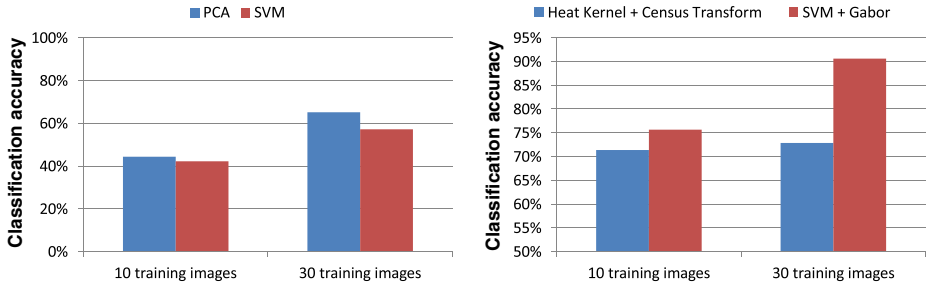
**Fig. 13** Comparison of our results (*red*) against those stated in (Whittier, et al., 2006) (*blue*). At the left, the classification accuracy is shown when only the classifier is used, without pre-processing input data. The right plot compares results when the pre-processing techniques are applied (Heat Kernel+Modified Census Transform on their case, against Gabor wavelet filtering in ours)

When we reproduce these experiments using our system, the results summarized in Fig. 13 are obtained. It is important to note that the basic version of the algorithms, just using the proposed classifiers but without applying any pre-process to input data, shows that PCA is superior to SVM classifiers. Apart from the possible discussions regarding this concrete result, it shows one more time that the real power of our system depends more on the Gabor wavelet filtering than on the SVM classifier.

In fact, when it is applied, the obtained accuracy is superior to that attributed to the combination Heat Kernel+Modified Census Transform. This is especially true when a large training dataset is used: with 30 training images, our best configured system obtains 91% classification accuracy, against the 72.6% from the paper.

## 2 Conclusions

This paper covers the detailed tuning of a face recognition system based on SVM classifiers and Gabor filters, to be used in a non-critical application where the input images have low and highly variable quality –making it possible to use a wide range of capturing devices, even webcams.

In spite of its simplicity, the combination Gabor+SVM can outperform some modern pre-process techniques focused on lighting invariance. It is robust against many forms of noise and image artifacts as the interlaced field problems shown in the employed dataset.

It has been shown that the proposed architecture can be used for non critical face recognition applications as those found in AmI projects. More concretely, the system proved to be valid in a video based application, processing face images around 100 × 100 pixels under slightly constrained conditions (frontal pose with up to 30–40° rotation, which is fairly common in many scenarios). Given an expected false acceptance rate of 1%, it can achieve around 70% recognition accuracy with faces varying in expression and having small pose changes under variable lighting. The error rate with known persons is below 0,5%.

Regarding the tuning of the system, downsampling Gabor features is an effective method for performing dimensionality reduction. In fact, thanks to the smooth

appearance of Gabor features, it often leads to more robust results in all the scenarios. The optimal size for the extracted Gabor features varies between 16 × 16 and 32 × 32.

We have found that a bank of Gabor filter covering 7 frequencies behaves at least as well as a version with 5 frequencies, but can outperform it largely when the SVM classifier approaches its limits. This fact contradicts the conclusion remarked in the work that pioneered the application of Gabor filters in face recognition [13]. Since almost all posterior works have used the same configuration, our result suggests that further analysis of this question can be of interest for researchers.

# References

1. Baker, S., Kanade, T.: Hallucinating Faces. Robotics Institute, Carnegie Mellon University. 1999. CMU-RI-TR-99-32 (1999)
2. Chesnokov, Y.: Face Detection C+Library with Skin and Motion Analysis. The Code Project: your development resource. [En línea] 2007. [Citado el: 13 de oct de 2008.] http://www.codeproject.com/KB/audio-video/face_detection.aspx (2007)
3. Chih-Chung, C., Chih-Jen, L.: LIBSVM: a library for support vector machines. [Online] 2001. http://www.csie.ntu.edu.tw/~cjlin/libsvm (2001)
4. Chih-Wei, H., Chih-Chung, C., Chih-Jen, L.: A Practical Guide to Support Vector Classification. [Online] July 2007. [Cited: 8 April 2009.] http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (2007)
5. Daugman, J.G.: Uncertainty relation for resolution in space, spatial-frequency and orientation optimized by two dimensional visual cortical filters. J. Opt. Soc. Am. A. Opt. Image Sce. Vis. **2**(7), (1985). doi:10.1364/JOSAA.2.001160
6. Hamamoto, Y., et al.: A Gabor filter-based for recognizing handwritten numerals. Pattern Recogn. **31**(4), 395–400 (1998)
7. Heisele, B., Ho, P., Poggio, T.: Face recognition with Support Vector Machines: global vs. component-based approach. International Conference on Computer Vision. pp. 688–694 (2001)
8. Field, D.J.: Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. **4**, 12 (1987)
9. Jain, A.K., Farrokhnia, F.: Unsupervised texture segmentation using Gabor filters. Pattern Recogn. **24**(12), 1167–1186 (1991)
10. Jain, A.K., Ratha, N.K.: Object detection using Gabor filters. Pattern Recogn. **30**, 295–309 (1997)
11. Jones, J., Palmer, L.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. **58**(6), 1233–1258 (1987)
12. Kemal Ekenel, H., Fischer, M., Stiefelhagen, R.: Face recognition in smart rooms. Lecture Notes in Computer Science. s.l. : Springer Berlin/Heidelberg. **4892/2008**, 120–131 (2008). doi:10.1007/978-3-540-78155-4_11
13. Lades, M., et al.: Distortion invariant object recognition in the Dynamic Link Architecture. IEEE Trans. Comput. **42**(3), 300–311 (1993)
14. Levine, M.D., Yu, Y.: State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person. Pattern Recogn. Lett. **30**(10), 908–913 (2009). doi:10.1016/j.patrec.2009.03.011
15. Li, B., Chang, H., Shan, S., Chen, X., Gao, W.: Hallucinating facial images and features. In: 2008 19th International Conference on Pattern Recognition, pp. 1–4. IEEE (2008)
16. Liang, Y., et al.: Gabor features-based classification using SVM for face recognition. Advances in Neural Networks-ISNN 2005. vol. 2, pp. 118–123. 2nd China International Symposium on Neural Networks (2005)
17. Liu, C.J., Wechsler, H.: Gabor feature based classification using the enhanced Fisher Linear Discriminant model for face recognition. IEEE Trans. Image Process. **11**(4), 467–476 (2002)

18. Mehrotra, R., Namuduri, K.R., Ranganathan, N.: Gabor filter-based edge detection. Pattern Recogn. **25**, 12 (1992). doi:10.1016/0031-3203(92)90121-X
19. O'toole, A.J., et al.: Face Recognition Algorithms surpass Humans Matching Faces over Changes in Illumination. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1642–1646 (2005)
20. Park, U., Jain, J.A.: 3D model-based face recognition in video. 2nd International Conference on Biometrics. Seoul : s.n. (2007)
21. Pentland, A., Choudhury, T.: Face recognition for smart environments. Computer. **33**(2), 50–55 (2000). doi:10.1109/2.820039
22. Phillips, P.J.: Support vector machines applied to face recognition. Adv. Neural Inf. Process. Syst. **11**, 803–809 (1999)
23. Shen, L., et al.: Gabor feature selection for face recognition using improved adaboost learning. Advances in Biometric Person Authentication. s.l.: Springer Berlin / Heidelberg **3781/2005**, 39–49 (2005)
24. Vapnik, V.: The nature of statistical learning theory (1995)
25. Vuilleumier, P., et al.: Distinct spatial frequency sensitivities for processing faces and emotional expressions. Nat. Neurosci. **6**, 624–631 (2003). doi:10.1038/nn1057
26. Wang, J., Asundi, A.K.: A computer vision system for wineglass defect inspection via Gabor-filter-based texture features. Inf. Sci. **127**(3–4), 157–171 (2000). doi:10.1016/S0020-0255(00)00036-0
27. Wang, X., Xiaoou, T.: Face Hallucination and Recognition. (ed.) Springer Berlin/Heidelberg. Lecture Notes in Computer Science. ISBN 978-3-540-40302-9 (2003)
28. Weldon, T.P., Higgins, W.E., Dunn, D.F.: Efficion Gabor filter design for texture segmentation. Pattern Recogn. **29**(12), 2005–2015 (1996)
29. Whittier, C., Xiaojun, Q.: Supervised heat kernel LPP method for face recognition. Computer Science Department-Utah State University. [Online]. http://www.cs.usu.edu/~xqi/Teaching/REU06/Website/Crystal/CrystalFinalPaper.pdf (2006)
30. Wiskott, L., et al.: Face recognition by elastic bunch graph matching. IEEE Trans. PAMI **19**(7), 775–779 (1997)