A TUNING STRATEGY FOR FACE RECOGNITION IN ROBOTIC APPLICATION

Thierry Germa, Michel Devy, Romain Rioux and Frédéric Lerasle
LAAS-CNRS, Université de Toulouse; 7, Avenue du Colonel Roche, F-31077 Toulouse, France
Université de Toulouse, UPS

Keywords: Video-based face recognition, SVM, Tracking, Particle filtering.

Abstract: This paper deals with video-based face recognition and tracking from a camera mounted on a mobile robot companion. All persons must be logically identified before being authorized to interact with the robot while continuous tracking is compulsory in order to estimate the position of this person. A first contribution relates to experiments of still-image-based face recognition methods in order to check which image projection and classifier associations lead to the highest performance of the face database acquired from our robot. Our approach, based on Principal Component Analysis (PCA) and Support Vector Machines (SVM) improved by genetic algorithm optimization of the free-parameters, is found to outperform conventional appearance-based holistic classifiers (eigenface and Fisherface) which are used as benchmarks.
The integration of face recognition, dedicated to the previously identified person, as intermittent features in the particle filtering framework is well-suited to this context as it facilitates the fusion of different measurement sources by positioning the particles according to face classification probabilities in the importance function.
Evaluations on key-sequences acquired by the mobile robot in crowded and continuously changing indoor environments demonstrate the tracker robustness against such natural settings. The paper closes with a discussion of possible extensions.

1 INTRODUCTION

The development of autonomous robots acting as human companions is a motivating challenge and a considerable number of mature robotic systems have been implemented which claim to be companions, servants or assistants in private homes. The dedicated hardware and software of such robot companions are oriented mainly towards safety, mobility in human centered environments but also towards peer-to-peer interaction between the robot companion and its unengineered human user. The robot's interlocutor must be logically identified before being authorized to interact with the robot while his/her identity must be verified throughout the performance of any coordinated tasks. Automatic visual person recognition is therefore crucial to this process as well as person verification throughout his/her tracking in the video stream delivered by the onboarded camera. Our particle filtering-based tracker will bring spatio-temporal information in order to improve the FR robustness to populated and cluttered environment.

Visual person recognition from a mobile platform operating in a human-centered scene is a challenging task which imposes several requirements. First, onboard processing power must enable the concurrent execution of other non-visual functionalities as well as of decisional routines in the robot's architecture. Thus, care must be taken to design efficient vision algorithms. Contrary to conventional biometric systems, the embedded visual sensor is moving in uncooperative human centered settings where people stand at a few meters - approximately at social and intimate distances - when interacting with the robot. Because of this context dependence, we can't use well-known public face still images (MIT, CMU, Yale, ... face databases) for our evaluations.

Given this framework, our face recognition (FR) system must be capable of handling: (i) poor video quality and low image resolution which is computationally faster, (ii) heavier lighting changes, (iii) larger pose variations in the face images.

The remainder of the paper is organized as follows. Section 2 depicts the prior related work linked with our approach. Section 3 describes our still-face image recognition system in our robotic con-

Germa T., Rioux R., Devy M. and Lerasle F. (2009).
A TUNING STRATEGY FOR FACE RECOGNITION IN ROBOTIC APPLICATION.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 514-522
DOI: 10.5220/0001800105140522

text. For enhancing recognition performances, tuning of the classifier free-parameters is here focused. Section 4 depicts the evaluation of several still face image recognition systems while Section 5 shows the improvements brought in the face recognition process by such a stochastic framework as particle filtering. Lastly, section 6 summarizes our contributions and discuss future extensions.

2 PRIOR RELATED WORK

Still-image FR techniques can be classified into two broad categories : holistic and analytic strategies even if the two are sometimes combined to form a complete FR system (Lam and Yan, 98). We focus on the former as analytic or feature-based approaches are not really suited to our robotic context. In fact, possible small face (depending on the H/R distance) and low image quality of faces captured by the onboard camera increase the difficulty in extracting local facial features. On the contrary, holistic or appearance-based approaches (Belhumeur et al., 1996) consider the face as a whole and operate directly on pixel intensity array representation of faces without the detection of facial features.

For detecting faces, we apply the well known window scanning technique introduced in (Viola and Jones, 2003), which covers a range of $\pm 45^{\circ}$ out-of-plane rotation. The bounding boxes of faces segmented by the Viola's detector are then fed into the face recognition systems referred to below.

Since the 1990s, appearance-based methods have been dominant approaches in still-face image recognition systems. Classically, they involve three sequential processes (Figure 1): (1) image pre-processing, (2) image projection into subspaces to construct lower dimensional image representation, (3) final decision rule for classification purposes. (Adini et al., 1997) points out that there is no image representation that can be completely invariant to lighting conditions and image-preprocessing is usually necessary. Like (Jonsson et al., 2000; Heseltine et al., 2002), histogram equalization is here adopted.

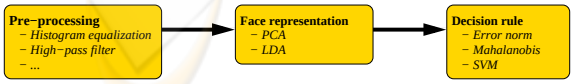


Figure 1: Face classification process.

Many popular linear techniques have been used in the literature for face representation. PCA (principal component analysis) uses image projection into PC (eigenface) to determine basis vectors that capture maximum image variance (Turk and Pentland, 1991) while LDA (linear discriminant analysis) determines a set of optimal discriminant basis vectors so that the ratio of the between-class and within-class scatters is maximized (Jonsson et al., 2000).

We design experiments in which faces are represented in both PC and LD subspaces, and focused thereafter on the free-parameter tuning of the overall FR system.

SVMs map the observations from input space into a higher dimensional feature space using a non-linear transformation, then find a hyperplane in this space which maximizes the margin of separation in order to minimize the risk of misclassification between faces. A RBF kernel is usually used for this transformation (Jonsson et al., 2000) where the width free-parameter γ controls the width of the Gaussian kernel. Another important free-parameter to tune is C, the upper bound of Lagrangian multipliers required for the minimization under constraints. Generally, the free-parameters are determined arbitrarily by trial and error norm. Genetic algorithms (GA) are well-known techniques for optimization problems, and have been proved for being effective for selecting SVM parameters (Seo, 2007).

3 OUR APPROACH

From these insights, we propose to focus our developments on the training process, involving both face representation and decision rule, but also the tuning of the parameters defined below. Figure 2 shows recognition process performed for histogram equalization-based preprocessing, two different representations (PC and LD basis) described in subsection 3.1, and three decision rules (error norm, Mahalanobis distance and SVM) depicted in subsection 3.2. In subsection 3.3, special emphasis concerns the tune of the free-parameters of each FR system. The face database is partitioned into four disjoint sets: (1) a training set #1 (8 users, 30 images per class), (2) a training set #2 (8 users, 30 images per class), (3) a training set #3 (8 users, 40 images per class). The training sets #1 and #2 are respectively used to learn the users' face representations and the class characteristics while the training set #3 allows us to estimate the free-parameters listed below.

Recall that the final goal is to classify facial regions $\mathcal{F}$, segmented from the input image, into either one class C_t out of the set $\{C_l\}_{l=1}^{M}$ of M subject faces using training algorithms.

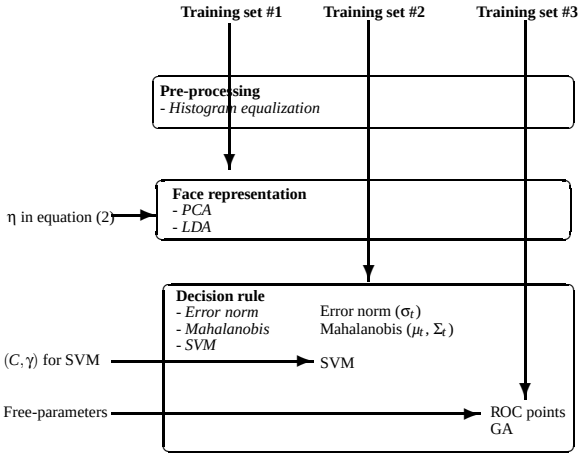


Figure 2: Face learning process.

3.1 Face Representation

Eigenface W_{pca} basis is deduced by solving

$$S_T.W_{pca} - W_{pca}.\Lambda = 0, \quad (1)$$

with S_T the scatter matrix, and Λ the ordered eigenvalue vector. We keep the first N_v eigenvectors as the eigenface basis such that

$$\frac{\sum_{i=0}^{N_v}\Lambda_i}{\sum\Lambda_i} \leq \eta, \quad (2)$$

accounting for a predefined ratio η of the total variance.

Another approach is to use Linear Discriminant Analysis (known as Fisherspace). ***Fisherface*** W_{lda} basis is deduced by solving

$$S_B.W_{lda} - S_W.W_{lda}.\Lambda = 0,$$

where S_B, and S_W are the between-class, and within-class scatter matrices while the eigenvectors selection follows also equation (2).

3.2 Decision Rule

Several methods are proposed to evaluate the decision rule which best fullfil our goals namely (i) Error norm, (ii) Mahalanobis distance, (iii) SVM.

The decision rule based on the ***error norm*** introduced in (Germa et al., 2007) is described as follow. Given an unknown test face $\mathcal{F} = \{\mathcal{F}(i)\}_{i=1}^{nm}$ and $\mathcal{F}_{r,t}$ the reconstructed face onto PC basis of the class C_t, this error norm is given by

$$\mathscr{D}(C_t, \mathcal{F}) = \sum_{i=1}^{nm}(\mathcal{F}(i) - \mathcal{F}_{r,t}(i) - \mu)^2,$$

and the associated likelihood follows

$$\mathscr{L}(C_t|\mathcal{F}) = \mathcal{N}(\mathscr{D}(C_t, \mathcal{F}); 0, \sigma_t),$$

where $\mathcal{F} - \mathcal{F}_{r,t}$ is the difference image of mean μ, σ_t terms the standard deviation of the error norms within the C_t's training set, and $\mathcal{N}(.; m, \sigma)$ is the Gaussian distribution with moments m and covariance σ. This error norm has been shown to outperform both the Euclidian distance and the DFFS.

The well-known ***Mahalanobis distance*** can be used in case of global space representation (Global PCA or LDA). It is defined as follow:

$$\mathscr{D}(C_t, \mathcal{F}) = \sqrt{(\mathcal{F}_t - \mu_t)^T\Sigma_t^{-1}(\mathcal{F}_t - \mu_t)},$$

where $\mathcal{F}_t$ is the vector resulting of the projection of $\mathcal{F}$ in W_t basis, and the class C_t is represented by μ_t and Σ_t, respectively its mean and covariance.

As described below, our motivation is to use a ***SVM*** as a base for our decision rule. The material about SVM framework will not be described hereafter for space reasons (see (Jonsson et al., 2000) for more details). The SVM method using RBF kernels needs two parameters to be fully defined. The first parameter is C, linked to the noise in the dataset. We choose its value as the greatest standard deviation computed on each class. The second, γ, is computed by heuristic methods from a set of tests on database called cross-validation.

The last issue concerns the appropriate decision rule. From a set of M learnt subjects/classes noted $\{C_l\}_{l=1}^{M}$ and a detected face $\mathcal{F}$, we can define for each class C_t the likelihood $\mathscr{L}^t = \mathscr{L}(C_t|\mathcal{F})$ for the detected face $\mathcal{F}$ and the posterior probability $P(C_t|\mathcal{F}, z)$ of labeling to C_t as

$$\begin{cases} \forall t\ P(C_t|\mathcal{F}, z) = 0 \text{ and } P(C_\emptyset|\mathcal{F}, z) = 1 \text{ when } \forall t\ \mathscr{L}^t < \tau \\ \forall t\ P(C_t|\mathcal{F}, z) = \frac{\mathscr{L}^t}{\sum_p \mathscr{L}^p} \text{ and } P(C_\emptyset|\mathcal{F}, z) = 0 \text{ otherwise}, \end{cases} \quad (3)$$

where $C_\emptyset$ refers to the void class, and τ is a predefined threshold for each class C_t.

3.3 Free Parameters Optimization

Several free parameters have to be tuned in order to optimze the FR process *i.e.* PC threshold η, SVM parameters C and γ and decision rule threshold τ.

The performances of the classifiers are analyzed by means of ***ROC*** when varying the free-parameter vector $\mathbf{q}$ subject to optimization for each classifier. The idea, pioneered by Provost *et al.* in (Provost and Fawcett, 2001), is outlined as follows. We search over a set of free-parameters by computing a ROC point *i.e.* the false rejection and false acceptance rates, namely FRR and FAR. For a given classifier, the set $\mathcal{Q}$ of all admissible parameter vectors $\mathbf{q}$ generates a set of ROC points, of which we seek the dominant, or optimal Pareto points along the ROC convex hull.

More formally, we seek for the subset $\mathcal{Q}^*_{1:n} \subset \mathcal{Q}_{1:n}$ of parameter vectors $\mathbf{q}_{1:n} = ((\gamma_1, C_1), \ldots, (\gamma_n, C_n))$ for which there is no other parameter vector that outperforms both objectives in $\mathcal{O} = \{FRR, FAR\}$:

$$\mathcal{Q}^* = \{\mathbf{q} \in \mathcal{Q} | \forall\, \mathbf{q}^{'} \in \mathcal{Q}, \\ \forall f_1 \in \mathcal{O}, f_1(\mathbf{q}) \geq f_1(\mathbf{q}^{'}) \wedge \exists f_2 \in \mathcal{O}, f_2(\mathbf{q}) > f_2(\mathbf{q}^{'})\} \quad (4)$$

Clearly, $\mathcal{Q}^*$ identifies the subset of parameter vectors that are potentially optimal for a given classifier.

Traditional methods using ***Genetic Algorithms*** are single-objective optimization problems (Seo, 2007). Non dominated sorting GA (NSGA-II) has been proved to be suited to multi-objective optimization problem (Xu and Li, 2006) as no solution can achieve a global optimum for several objectives, namely minimizing both the FRR and the FAR. If the value of first objective function cannot be improved without degrading the second objective function, the solution is referred to Pareto-optimal or non dominated ones (Gavrila and Munder, 2007). Algorithm 1 describes the steps of the process used for our free-parameters optimization. This algorithm is iterated on the overall FR process so as to find the best parameters for the complete system. This approach allows us to find the best compromise between FAR and FRR by finding the Pareto front

$$F_i = \{ f(\mathbf{q}) \in \mathcal{O} | \mathbf{q} \in \mathcal{Q}^* \}$$

with $\mathcal{Q}^*$ the Pareto optimal set.

4 EVALUATIONS AND RESULTS

We conducted FR experiments using the proposed framework on the face dataset composed of 5500 test examples including 8 possible human users and 3 impostors corresponding to unknown individuals for the robot. In this dataset, the subjects arbitrarily move

Figure 3: Samples for a given class.

their heads, possibly change their expressions while the ambient lighting, the background, and the relative distance might change. A few example images from this dataset are shown in figure 3 while the entire face gallery is available on demand.

Algorithm 1 NSGA-II algorithm.

1: Create a random parent population P_0 of size N. Set $t = 0$.
2: Apply crossover and mutation to P_t to create offspring population Q_t of size N.
3: **if** Equation (4) is not satisfied **then**
4: Set $R_t = P_t \cup Q_t$.
5: Identify the non-dominated fronts $F_1, F_2, \ldots, F_k$ in R_t.
6: **for** $i = 1, \ldots, k$ **do**
7: Calculate crowding distance of the solutions in F_i. Sort by crowding.
8: **if** $|P_{t+1}| + |F_i| > N$ **then**
9: Add the least crowded $N - |P_{t+1}|$ solutions from F_i to P_{t+1}.
10: **else**
11: Set $P_{t+1} = P_{t+1} \cup F_i$.
12: **end if**
13: **end for**
14: Use binary tournament selection based on the crowding distance to select parents from P_{t+1}.
15: Set $t = t + 1$. Go to step (2).
16: **end if**

4.1 Evaluated Recognition Systems

1. System FSS+EN: Face-Specific Subspace and Error Norm. As described in (Shan et al., 2003), for each class C_t, we compute $W_{pca,t}$ thanks to equation (1), and keep the $N_{v,t}$ eigenvectors (equation (2)). We use the predefined error norm as the decision rule.

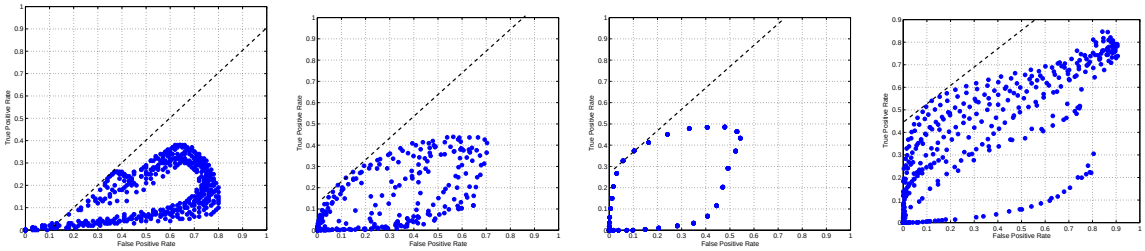
2. System GPCA+MD: Global PCA and Mahanalobis Distance. Here a single PC basis is estimated given equation (1) and the total scatter matrix S_T. The decision rule is based on the Mahanalobis distance.

3. System LDA+MD: Fisherface and Mahanalobis Distance. Fisherfaces are used thanks to equation (1) so as to get W_{lda} as the projection basis to compute Mahalanobis distance.

4. System GPCA+SVM: Global PCA and SVM. This system performs global PCA and SVM delivers probability estimates. The associated theory and implementation details are described in (Wu et al., 2004). This classifier model produces the free-parameters η, C, γ and τ.

4.2 Results

All the above classifiers lead to the same performances in terms of sensitivity ($\sim$ 75%) and selectivity ($\sim$ 90%). These results are very promising given the



(a) System 1, $\mathbf{q} = (\eta, \tau)'$: EER=0.51.

(b) System 2, $\mathbf{q} = (\eta, \tau)'$: EER=0.44.

(c) System 3, $\mathbf{q} = (\eta, \tau)'$: EER=0.37.

(d) System 4, $\mathbf{q} = (\eta, C, \gamma, \tau)'$: EER=0.29.

Figure 4: ROC points for each classifier and the associated isocost line for EER. Free-parameter vector $\mathbf{q}$ for optimization are listed under the corresponding classifier.

extent of pose, varying illumination, expression, and distance to individuals. On the contrary, significant differences between the face recognition systems are observed for FAR (misclassification) and FRR (misrejection). This is highlighted in the ROC analysis (figure 4) which shows the results of the different FR systems in terms of FAR and FRR when varying the corresponding free-parameters.

Figure 4 shows ROC points and the Pareto front when varying the free-parameters over their ranges. The subfigures, when plotting TPR and FPR on the Y- and X-axis, allows an informal visual comparison of the four classifiers. System 4 clearly dominates the other classifiers as its Pareto front lies in the nortwest corner of the ROC space (TPR is higher, FPR is lower). Considering the equal error rate (EER) leads to the same analysis. The best system, namely 4, provides a Pareto front with a lower EER, namely 0.29. Finally, note that its computational cost is 0.5 ms against 0.3 ms per image for systems 2-3. Unfortunately, an exhaustive search for the selection of all parameters, especially for model 4 which produces more free-parameters, is computationally intractable on a autonomous robot as the finality is to learn human faces on-the-fly when interacting with new persons. Consequently, we propose a genetic algorithm (GA) to discover optimal free-parameter vectors of system 4 more quickly due to its multi-objective optimization framework. By limiting the number of ROC points to be considered, GA renders the optimization procedure computationally feasible.

Figure 5 shows the evolution of the Pareto front when varying the population size in the range $[16, 20]$ and the preset generation count in the range $[1, 30]$.

Note that the generated solutions "move" so as to reduce both FPR and FRR objectives. This optimization strategy is no longer guaranteed to find the Pareto front optimum but there is an experimental evidence that the solution is close to optimal when increasing the preset generation count. Given a popula-

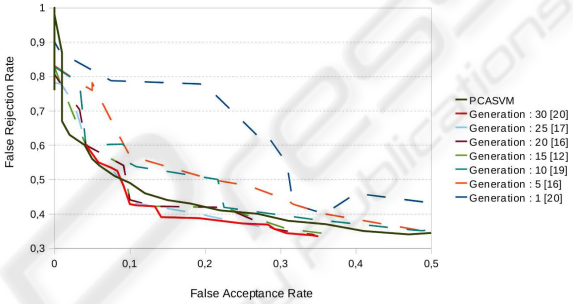


Figure 5: NSGA-II Pareto front evolution *vs.* PCA+SVM based system D.

tion initialized randomly (first generation in figure 5), we can see that after the first 10 generations, there is already one solution that outperforms the one without optimization while 30 generations increase the performance compared to ROC means slightly. Therefore, the minimum EER for 30 generations becomes 0.26 against 0.29 in subfigure 4(d).

5 FACE TRACKING AND ROBOTIC EXPERIMENTS

Spatiotemporal FR analysis is here considered even if FR and tracking process are split : the person-specific estimated dynamic characteristics helped the FR system and reciprocally. Solving these two tasks simultaneously by probabilistic reasoning (Zhou et al., 2004) has been proven to significantly enhance the recognition performances.

5.1 Tracking Framework based on Face Recognition

Particle filters (PF) aim to recursively approximate the posterior probability density function (pdf) $p(\mathbf{x}_k|z_{1:k})$ of the state vector $\mathbf{x}_k$ at time k conditioned on the set

of measurements $z_{1:k} = z_1, \ldots, z_k$. A linear point-mass combination

$$p(\mathbf{x}_k|z_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}), \ \sum_{i=1}^{N} w_k^{(i)} = 1, \quad (5)$$

is determined –with $\delta(.)$ the Dirac distribution– which expresses the selection of a value –or "particle"– $\mathbf{x}_k^{(i)}$ with probability –or "weight"– $w_k^{(i)}$, $i = 1, \ldots, N$. An approximation of the conditional expectation of any function of $\mathbf{x}_k$, such as the MMSE estimate $\mathrm{E}_{p(\mathbf{x}_k|z_{1:k})}[\mathbf{x}_k]$, then follows.

Recall that the SIR –or "Sampling Importance Resampling" – algorithm is fully described by the prior $p_0(\mathbf{x}_0)$, the dynamics pdf $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the observation pdf $p(z_k|\mathbf{x}_k)$. After initialization of independent identically distributed (i.i.d.) sequence drawn from $p_0(\mathbf{x})$, the particles evolve stochastically, being sampled from an importance function $q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, z_k)$. They are then suitably weighted so as to guarantee the consistency of the approximation (5). Then a weight $w_k^{(i)}$ is affected to each particle $\mathbf{x}_k^{(i)}$ involving its likelihood $p(z_k|\mathbf{x}_k^{(i)})$ w.r.t. the measurement z_k as well as the values of the dynamics pdf and importance function at $\mathbf{x}_k^{(i)}$. In order to limit the well-known degeneracy phenomenon (Arulampalam et al., 2002), a resampling stage is introduced so that the particles associated with high weights are duplicated while the others collapse and the resulting sequence $\{\mathbf{x}_k^{(i)}\}_{i=1}^N$ is i.i.d. according to (5).

With respect to our data fusion context, we opt for using ICONDENSATION (Isard and Blake, 1998), that consists in sampling some particles from the observation image (namely $\pi(.)$), some from the dynamics and some w.r.t. the prior $p_0(.)$ so that importance function reads as, with $\alpha, \beta \in [0;1]$

$$q(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}, z_k) =$$
$$\alpha\pi(\mathbf{x}_k^{(i)}|z_k) + \beta p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}) + (1-\alpha-\beta)p_0(\mathbf{x}_k). \quad (6)$$

where $\pi(.)$ relates to detector outputs which, despite their intermittent nature, are proved to be very discriminant when present (Pérez et al., 2004).

5.1.1 Tracking Implementation

The aim is to fit the template relative to the targeted person all along the video stream through the estimation of his/her image coordinates (u, v) and its scale factor s of his/her head. All these parameters are accounted for in the above state vector $\mathbf{x}_k$ related to the k-th frame. With regard to the dynamics $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, the image motions of humans are difficult to characterize over time. This weak knowledge is formalized by defining the state vector as $\mathbf{x}_k = [u_k, v_k, s_k]'$ and assuming that its entries evolve according to mutually independent random walk models, viz. $p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1}, \Sigma)$ where covariance $\Sigma = \mathrm{diag}(\sigma_x^2, \sigma_y^2, \sigma_s^2)$.

In both importance sampling and weight update steps, fusing multiple cues enables the tracker to better benefit from distinct information, and decrease its sensitivity to temporary failures in some of the measurement processes. The underlying unified likelihood in the weighting stage is more or less conventional. It is computed by means of several measurement functions introduced in (Germa et al., 2007), according to persistent visual cues, namely: (i) multiple color distributions to represent the person's appearance (both head and torso), (ii) edges to model the silhouette. Otherwise, our importance function is unique in the literature and so is detailed here below.

5.1.2 Importance Function based on Face Recognition

Recall that the function $\pi(.)$ in equation (6) offers a mathematically principled way of directing search according to multiple and possibly heterogeneous detectors and so to (re)-initialize the tracker. Given L independent detectors and κ their weights, the function $\pi(.)$ can be reformulated as

$$\pi(\mathbf{x}_k|z_k^1, \ldots, z_k^L) = \sum_{l=1}^{L} \kappa_l . \pi(\mathbf{x}_k|z_k^l), \text{ with } \sum \kappa_l = 1. \quad (7)$$

Two functions $\pi(\mathbf{x}_k|z_k^c)$ and $\pi(\mathbf{x}_k|z_k^s)$, respectively based on skin probability image (Lee et al., 2003) and face detector are here considered.

The importance function $\pi(\mathbf{x}_k|z_k^c)$ at location $\mathbf{x}_k = (u, v)$ is described by

$$\pi(\mathbf{x}|z^c) = \mathbf{h}(c_z(\mathbf{x})) \quad (8)$$

given that $c_z(\mathbf{x})$ is the color of the pixel situated in $\mathbf{x}$ in the input image z^c and $\mathbf{h}$ is the normalized histogram representing the color distribution of the skin learnt *a priori*. The function $\pi(\mathbf{x}_k|z_k^s)$ is based on a probabilistic image based on the well-known face detector pioneered in (Viola and Jones, 2003). Let N_B be the number of detected faces and $\mathbf{p}_i = (u_i, v_i), i = 1, \ldots, N_B$ the centroid coordinate of each such region. The function $\pi(.)$ at location $\mathbf{x} = (u, v)$ follows, as the Gaussian mixture proposal[1]

$$\pi(\mathbf{x}|z^s) \propto \sum_{j=1}^{N_B} P(C|\mathcal{F}_j, z^s).\mathcal{N}(\mathbf{x}; \mathbf{p}_j, \mathrm{diag}(\sigma_{u_j}^2, \sigma_{v_j}^2)), \quad (9)$$

[1]Index k and (i) are omitted for the sake of clarity and space.

Data fusion strategy	$t = 15$	$t = 81$	$t = 126$	$t = 284$
(a) $q(x_k\|x_{k-1},z_k) = \alpha\pi(.) + \beta p(.)$ with face detection				
(b) $q(x_k\|x_{k-1},z_k) = \alpha\pi(.) + \beta p(.)$ with face classification				

Table 1: Different data fusion strategies involved in importance sampling.

Figure 6: From left to right: original image, skin probability image (8), face recognition (9), unified importance function (6) (without dynamic), accepted particles (yellow dots) after rejection sampling.

with $P(C|\mathcal{F}_j,z)$ the face ID probabilities described in equation (3) for each detected face $\mathcal{F}_j$.

The particle sampling is done using the importance function $q(.)$ in equation (6) and a process of rejection sampling. This process constitutes an alternative when $q(.)$ is not analytically described. The principle is as follows with $Mg(.)$ an envelope distribution to make the sampling easier ($M > 1$):

Algorithm 2 Rejection sampling algorithm.

1: Draw $\mathbf{x}_k^{(i)}$ according to $Mg(\mathbf{x}_k)$
2: $r \leftarrow \frac{q(\mathbf{x}_k|\mathbf{x}_{k-1},z_k)}{Mg(\mathbf{x}_k^{(i)})}$
3: Draw u according to $\mathcal{U}_{[0,1]}$
4: **if** $u \leq r$ **then**
5: Accept $\mathbf{x}_k^{(i)}$
6: **else**
7: Reject it
8: **end if**

Figure 6 shows an illustration of the rejection sampling algorithm for a given image. Our importance function (6) combined with rejection sampling ensures that the particles will be (re)-placed in the relevant areas of the state space *i.e.* concentrated on the tracked person.

5.2 Live Experiments

The above tracker has been prototyped on a 1.8GHz Pentium Dual Core using Linux and the OpenCV library. Both quantitative and qualitative off-line evaluations on sequences are reported here below. This database of two different sequences (800 images) acquired from our mobile robot in a wide range of realistic conditions allows us to: (i) determine the optimal parameter values of the tracker, (ii) identify its strengths and weaknesses, and in particular characterize its robustness to environmental artifacts: clutter, occlusion or out-field of sight, lighting changes. Several filter runs per sequence are performed and analyzed.

The runs presented in Table 1 show the efficiency of the strategy of data fusion in both importance and measurement function. Let us comment these results. The template corresponding to the estimate of the position of the target is represented by the blue rectangles (color template) and the green curve (shape template) while the dots materialize the hypotheses and their weight after normalization (black is 0 and red is 1). The run (a) in Table 1 combines face and skin color detection with the random walk dynamic in the importance function in order to guide the particle sampling on specific additional areas of the current image (mainly on detected faces). We can see that this strategy is not sufficient to distinguish whether the template is on the right targeted person or not. The last run in Table 1(b) shows the complete system used in our experiments involving the face classification process in the importance function as described in (9). We can see, at time $t = 81$, that after a sporadic occlusion of the target by another person (with the black trousers), the face classification helps to direct the particle sampling only on the desired person and so enables the template to recover the target.

Quantitative performance evaluations summarized

below have been carried out on the sequence database. Table 2 consider the FR performance with or without tracking and presents the classification results. For each sequence, these results are compared to tracking results in terms of FAR (False Acceptance Rate) and FRR (False Rejection Rate). To be more consistent, the only images involving face detection have been taken into account. We note that the runs involving tracking are more robust to environmental changes, mainly due to spatio-temporal effects.

Table 2: Face classification performance for the database image subset involving detected frontal faces.

Face classif.	without tracking	with tracking
FAR	35.09%	26.47% ($\sigma = 1.97\%$)
FRR	60.22%	25.73% ($\sigma = 0.25\%$)

6 CONCLUSIONS AND PERSPECTIVES

This paper presented the development of a still-image FR system dedicated to Human/Robot interaction in a household framework. The main contribution is the improvement of the known FR algorithms thanks to a genetic algorithm for free-parameter optimization.

Off-line evaluations on sequences acquired from the robot show that the overall system enjoys the valuable capabilities: (1) efficiency of the recognition process against face pose changing, (2) robustness to illumination changes. Eigenface subspace and SVM makes it possible to avoid misclassification due to the environment while NSGA-II improves the FR process. Moreover, the fusion of FR outputs in the tracking loop enables the overall system to be more robust to natural and populated settings.

Several directions are studied regarding our still-image FR system. A first line of investigation concerns the fusion of heterogeneous information such as RFID or sound cues in order to keep the identification process more robust to the environment. Detection of an RFID tag worn by individuals will allow us to drive the camera thanks to a pan-tilt unit and so trigger tracker initialization, and will contribute as another measurement in the tracking loop. The sound cue will endow the tracker with the ability to switch its focus between known speakers.

REFERENCES

Adini, Y., Moses, Y., and Ullman, S. (1997). Face recognition: the problem of compensating for changes in illumination direction. 19(7):721–732.

Arulampalam, S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *Trans. on Signal Processing*, 2(50):174–188.

Belhumeur, P., Hespanha, J., and Kriegman, D. (1996). Eigenfaces vs. fisherfaces. In *European Conf. on Computer Vision (ECCV'96)*, pages 45–58.

Gavrila, D. and Munder, S. (2007). Multi-cue pedestrian detection and tracking from a moving vehicle. *Int. Journal of Computer Vision (IJCV'07)*, 73(1):41–59.

Germa, T., Brèthes, L., Lerasle, F., and Simon, T. (2007). Data fusion and eigenface based tracking dedicated to a tour-guide robot. In *Int. Conf. on Vision Systems (ICVS'07)*, Bielefeld, Germany.

Heseltine, T., Pears, N., and Austin, J. (2002). Evaluation of image pre-processing techniques for eigenface based recognition. In *SPIE: Image and Graphics*, pages 677–685.

Isard, M. and Blake, A. (1998). I-CONDENSATION: Unifying low-level and high-level tracking in a stochastic framework. In *European Conf. on Computer Vision (ECCV'98)*, pages 893–908.

Jonsson, K., Matas, J., Kittler, J., and Li, Y. (2000). Learning support vectors for face verification and recognition. In *Int. Conf. on Face and Gesture Recognition (FGR'00)*, pages 208–213, Grenoble, France.

Lam, K. and Yan, H. (98). An analytic-to-holistic approach fo face recognition based on a single frontal view. 7(20):673–686.

Lee, K., Ho, J., Yang, M., and Kriegman, D. (2003). Video-based face recognition using probabilistic appearance manifolds. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 1:I–313–I–320 vol.1.

Pérez, P., Vermaak, J., and Blake, A. (2004). Data fusion for visual tracking with particles. *Proc. IEEE*, 92(3):495–513.

Provost, F. and Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, 42(3):203–231.

Seo, K. (2007). A GA-based feature subset selection and parameter optimization of SVM for content-based image retrieval. In *Int. Conf. on Advanced Data Mining and Applications (ADMA'07)*, pages 594–604, Harbin, China.

Shan, S., Gao, W., and Zhao, D. (2003). Face recognition based on face-specific subspace. *Int. Journal of Imaging Systems and Technology*, 13(1):23–32.

Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'91)*, pages 586–591.

Viola, P. and Jones, M. (2003). Fast multi-view face detection. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR'03)*.

Wu, T., Lin, C., and Weng, R. (2004). Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005.

Xu, L. and Li, C. (2006). Multi-objective parameters selection for SVM classification using NSGA-II. In *Industrial Conference on Data Mining (ICDM'06)*, pages 365–376.

Zhou, S., Chellappa, R., and Moghaddam, B. (2004). Visual tracking and recognition using appearance-adaptive models in particle filters. *Trans. on Image Processing*, 13(11):1491–1506.