Model based selection and classification of local features for recognition using Gabor filters ⋆

Plinio Moreno, Alexandre Bernardino, and José Santos-Victor
`{plinio, alex, jasv}@isr.ist.utl.pt`

Instituto Superior Técnico & Instituto de Sistemas e Robótica
1049-001 Lisboa - Portugal

Abstract. We propose models based on Gabor functions to address two related aspects in the object recognition problem: interest point selection and classification. We formulate the interest point selection problem by a cascade of bottom-up and top-down stages. We define a novel type of top-down saliency operator to incorporate low-level object related knowledge very soon in the recognition process, thus reducing the number of canditates. For the classification process, we represent each interest point by a vector of Gabor responses whose parameters are automatically selected. Both the selection and classification procedures are designed to be invariant to rotations and scaling. We apply the approach to the problem of facial landmark classification and present experimental result illustrating the performance of the proposed techniques.

1 Introduction

The object recognition problem has been tackled recently using the concept of low-level features with several successful results [1–4]. All of these works exploit the idea of selecting various points in the object and building up a local neighborhood representation for each one of the selected points. In this work we introduce models built with Gabor functions to unfold the following issues: (selection) which points are important to represent the object, and (classification) how to represent and match the information contained in each point's neighborhood.

The point selection problem, also called keypoint detection [1, 5], interest point detection[3], bottom-up saliency [6], and salient region detection [7], has been addressed in a bottom-up fashion. Bottom-up means that points selected are image-dependent, not task-dependent. Salient points are selected to be distinguishable from its neighbors and have good properties for matching, repeatability, and/or invariance to common image deformations. However, there is evidence of interaction between bottom-up and top-down processes in nearly every visual search model in the human visual system[8]. In guided visual search problems, where specific objects are searched in the scene, it is convenient to incorporate object related knowledge(top-down information) as soon as possible in the recognition process, to reduce the amount of possible candidates. The

objective of the saliency function is to remove points very different from the model and have very few rejections of "good points". We propose a novel rotation invariant saliency function based on the isotropic frequency characteristics of the feature to detect. We also propose a new method to compute each point's intrinsic scale, to provide scale invariance for both the saliency and detection processes. We will describe extensively this approach in Section 2.

In the classification stage we employ an image neighborhood representation built with Gabor filter responses proposed in [9]. The image regions are modeled by a feature vector formed from Gabor filter responses computed at the interest point, and the model parameters are automatically selected. In Section 3 we review this procedure, and we introduce rotation and scale invariance to the feature vector. In Section 4 we present tests in facial region recognition, followed by the conclusions and future work in Section 5.
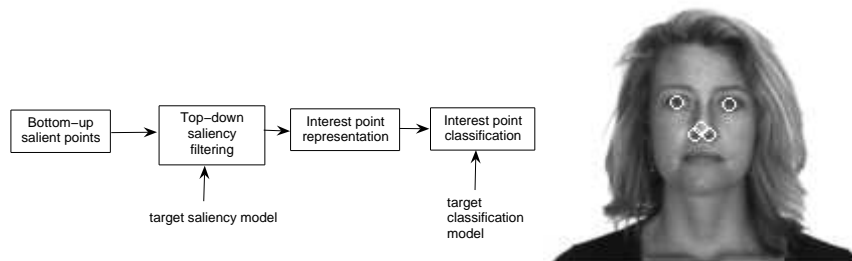

Fig. 1. In the left side, architecture of our recognition approach. In the right side, examples of facial landmarks

We observe in the left side of Figure 1 the proposal for our recognition approach. We remark the top-down filtering and the interest point classification steps as the bulk of our work.

2 Top-down interest point selection

When searching for a particular target[1] on a image, we must search, from a set of candidates points. Depending on the employed algorithms for searching and matching, this can be a computationally expensive procedure, and exhaustive search should be avoided. In this section we propose an intermediate step in the recognition process to reduce the space of candidates for matching, where a target-specific saliency operator is designed to encode low-level target information.

2.1 Appearance based saliency operator

We will exploit the good properties of Gabor Filters to represent texture, and introduce top-down information in interest point detection. The 2D zero mean isotropic Gabor

---

[1] We denote 'target' as a local feature belonging to the object of interest

function is:

$$g_{\theta,f,\sigma}(x,y) = \frac{e^{-\frac{x^2+y^2}{2\sigma^2}}}{2\pi\sigma^2}\left(e^{j2\pi f(x\cos(\theta)+y\sin(\theta))} - e^{-2\sigma^2 f^2\pi^2}\right) \quad (1)$$

By convolving the Gabor function with image patterns $I(x,y)$, we can evaluate their similarity. The Gabor response at point (x_0, y_0) is

$$G_{\theta,f,\sigma}(x_0,y_0) = \int\int I(x,y)g_{\theta,f,\sigma}(x_0-x, y_0-y)dxdy \quad (2)$$

In Fig. 2 we show the appearance of some Gabor kernels as a function of σ, θ, and f. These parameters can characterize the dominant texture of an interest point. One approach to characterize texture at an image point (x_j, y_j) would be to compute the response of several gabor filters, tuned to different orientations, frequencies and scales, and retain the parameters corresponding to the maximum response:

$$(\hat{\sigma}_j, \hat{f}_j, \hat{\theta}_j) = \arg\max_{\sigma,f,\theta}|G_{\theta,f,\sigma}(x_j, y_j)|. \quad (3)$$

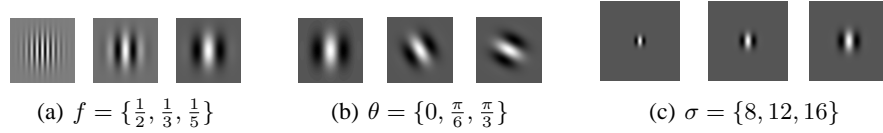However none of the parameters is invariant to common image transformation such as

(a) $f = \{\frac{1}{2}, \frac{1}{3}, \frac{1}{5}\}$ (b) $\theta = \{0, \frac{\pi}{6}, \frac{\pi}{3}\}$ (c) $\sigma = \{8, 12, 16\}$

Fig. 2. Examples of Gabor functions. Each sub-figure shows the real part of Gabor function for different values of f, θ, and σ

scalings and rotations. To obtain scale and rotation invariance we do as follows:

1. Sum the response of the Gabor filters for all orientations and scales:

$$GFS(x_j, y_j, f) = \int_0^\infty \int_{-\pi}^\pi G_{\theta,f,\sigma}(x_j, y_j)d\theta d\sigma \quad (4)$$

   For each point (x_j, y_j), this is a function of the frequency only, and is denoted by the *f-signature* of a point. This function will give us the "energy" of the image pattern for any frequency of interest.
2. The f-signature of an interest point is independent of the orientation but still depends on the scale. Therefore, we define a scale invariant parameter $\gamma = \frac{1}{\sigma f}$, where σ is the intrinsic [2] scale of the interest point. This parameter is the ratio between wavelength (inverse of frequency) and scale, and is proportional to the number of wave periods within the spatial support of the filter. Thus, it can be interpreted as a "scale invariant frequency" parameter.

---

[2] the concept of intrinsic scale will be explained later.

3. Finally, to obtain a scale invariant signature, we map the f-signature function to γ values, and compute the γ-signature function. This function will constitute the low-level scale and rotation invariant texture representation of an interest point.

In the next subsections we will describe how to compute and match γ-signatures.

2.2 Computing the f signature

To compute the f-signature of an image point one could use the direct implementation of Eq.(4). However this would require a significant amount of computation. To overcome this problem we define an equivalent kernel that filters the image just once for each frequency. The equivalent kernel is obtained by summing the gabor kernels for all scales and orientations, and is denoted "Gabor Frequency Saliency" kernel.

$$GFS_k(x_j, y_j, f) = \int_0^\infty \int_{-\pi}^{\pi} g_{\sigma,\theta,f}(x_j, y_j) d\theta d\sigma \qquad (5)$$

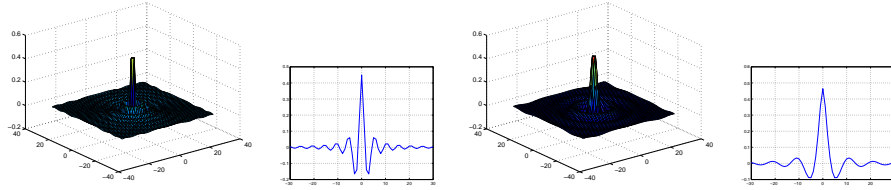The closed form expression for the frequency-space kernel is the following:

Fig. 3. Example of Gabor frequency-space kernel. Top figures, 3D plot and 1D slice of $GFS_k(x, y, 0.2)$. Bottom figures, 3D plot and 1D slice of $GFS_k(x, y, 0.1)$

$$GFS_k(r, f) = \frac{\sqrt{\pi/2}}{r} \left(-e^{-2\pi fr} + J_0(2\pi fr) \right) \qquad (6)$$

In Eq.(6), $r = \sqrt{x_j^2 + y_j^2}$, and $J_0(z)$ is the Bessel function of the first kind. In Figure 3 we can see an example of GFS_k, its shape is an exponentially decreasing 2D Bessel function, and it is rotationally and translation invariant. Therefore, the computation of the f-signature at point (x_j, y_j) can be performed by:

$$GFS(x_j, y_j, f) = I * GFS_k, \qquad (7)$$

The f-signature at a point is rotationally invariant. However, if we change the scale of the image, the f-signature will both translate in the frequency axis and its amplitude will change linearly with the scale factor. After amplitude normalization, f-signatures of the same point at different scales will be related by a single translation.

Amplitude normalization Let us consider two images: $I(x, y)$, and an homogeneously scaled version of $I(x, y)$. The new image is scaled by a factor a: $I_s(x, y) = I(ax, ay)$. The f-signature at point (x_j, y_j) is:

$$(I_s * GFS_k)(x_j, y_j, f) = (I * GFS_k)(ax_j, ay_j, f) \quad (8)$$

Now let $\tilde{x} = ax$, $\tilde{y} = ay$, and $\tilde{f} = f/a$. Then $dx = d\tilde{x}/a$, and $dy = d\tilde{y}/a$. By making substitutions in Eq. 6,

$$GFS^{I_s}(x_j, y_j, f) = (I * GFS)(ax_j, ay_j, f/a) = \frac{\tilde{f}}{f} GFS^I(ax_j, ay_j, f/a) \quad (9)$$

$$f GFS^{I_s}(x_j, y_j, \sigma) = \tilde{f} GFS^I(ax_j, ay_j, f/a) \quad (10)$$

From Eq. (10) we see that if we multiply the response of the kernel by the frequency, the f-signature amplitude becomes normalized with respect to scale changes. Thus, the normalized f-signature of an image point (x_j, y_j), is:

$$GFS_{norm}(x_j, y_j, f) = f(I * GFS_k) \quad (11)$$

In Figure 4 we can see an example of GFS_{norm} for the case of an eye's center point.

2.3 Computing the γ-signature

In order to compute the γ-signature of an image point, we perform two steps: (i) Compute the normalized f-signature, and (ii) map the frequency interval into a γ interval using the information of the image point intrinsic scale. The rationale is to obtain a signature that is invariant to image scale transformations.

Intrinsic scale from the f-signature A conventional method to compute the intrinsic scale of a point is to use the Lindeberg's method for blobs[10, 11]: the scale where the convolution with a normalized Laplacian of Gaussian functions is maximum.

Looking at the shape of the equivalent frequency kernels in Figure 3, we notice similarities between the Laplacian of Gaussian functions with different scales and GFS kernels of different frequencies (in fact frequency is related to the reciprocal of scale). Experimentally we noticed that the zero crossings of the f-signature function closest to the global maxima is very stable under image scale changes and is related reciprocally to the scale factor. Thus, we compute the intrinsic scale $\hat{\sigma}$ at a point (x_i, y_i) as:

$$\hat{\sigma} = 1/\hat{f}; \quad \hat{f} = \arg \min_{f, GFS_{norm}(x_i, y_i, f)=0} f - \tilde{f}; \quad \tilde{f} = \arg \max_f |GFS_{norm}| \quad (12)$$

Mapping f to γ values Let us define a set of frequency values $F = \{f_1, \ldots, f_i, \ldots, f_n\}$. To map the set of frequency values F into γ values, we compute the intrinsic scale($\hat{\sigma}$) from Eq. (12). Mapping the $f_i \in F$ values to $\gamma_i \in \Gamma$ values, the interval of γ-signature is $\Gamma = \{\gamma_1, \ldots, \gamma_i, \ldots, \gamma_n\} = \{1/f_1\hat{\sigma}, \ldots, 1/f_i\hat{\sigma}, \ldots, 1/f_n\hat{\sigma}\}$. Thus, the γ-signature(top-down saliency model) of an image point (x, y) is:

$$\Gamma S_{x,y}(\gamma_i) = FS_{x,y}(1/\gamma_i\hat{\sigma}), \gamma_i \in \Gamma \quad (13)$$

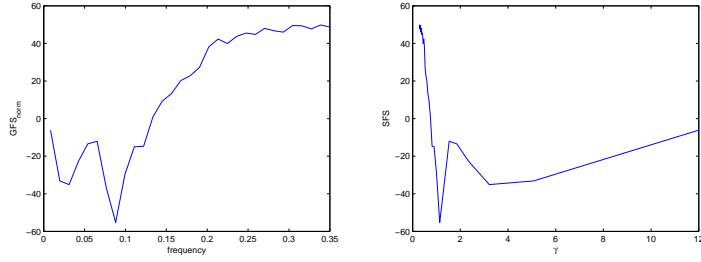As an example, in Figure 4 we show the γ-signature of an eye's center point.

Fig. 4. In the left side, GFS_{norm} and in the right side $\Gamma S_{x,y}$, for the eye's center point

2.4 Top-down Saliency Model and Matching

Let us assume an initial set of bottom-up salient points and assume we are interested in searching for a particular interest point. First we need a model of the point in terms of its γ-signature. This can be obtained from a single example of the point's neighbourhood or can be computed from the mean γ-signature values in a training set:

$$SM_{x,y}(\gamma_i) = \overline{\Gamma S}_{x,y}(\gamma_i), \gamma_i \in \Gamma \qquad (14)$$

After having an appropriate interest point model, in the form of a γ-signature, we can analize novel images and reject interest points not conforming to the model. The rejection of bad candidates is performed by matching the γ-signature of the test point $\Gamma S_{xt,yt}$ with the saliency model $SM_{x,y}$, doing the following steps: (i) Find the intersection of the γ intervals between the two signatures, (ii) subsample the longest signature, and (iii) compute the euclidean distance between signatures.

Let us define two intervals: $\Gamma_S = [\gamma_{i_S}, \gamma_{f_S}]$ of signature $\Gamma S_{xt,yt}$, and $\Gamma_{SM} = [\gamma_{i_{SM}}, \gamma_{f_{SM}}]$ of the object model $SM_{x,y}$, where i stands for initial value, and f stands for final value. The segment of the signature for computing the distance is the intersection of the two intervals, $\Gamma_S \cap \Gamma_{SM} = [\gamma_{i_\cap}, \gamma_{f_\cap}]$.

The number of signature elements within the interval $[\gamma_{i_\cap}, \gamma_{f_\cap}]$ could be different in $\Gamma S_{xt,yt}$ and $SM_{x,y}$. Therefore, we subsample the signature segment with more elements in $[\gamma_{i_\cap}, \gamma_{f_\cap}]$ to have equal sized signature segments. In the last step, we compute the Euclidean distance between the signatures.

3 Interest point classification

In this section we describe a method to automatically decide about the presence or absence of a certain target of interest in the image. A target is considered as a single point and a circular neighborhood of the size of its "intrinsic scale". An adequate target representation must have good matching properties, and invariance to image region transformations. In [9] a representation based on Gabor filter responses with good matching properties is presented. In this paper we add rotation and scale invariance, and further automate the recognition process.

3.1 Target Model and Parameter Selection

Each target is modeled as random vector containing the real and imaginary parts of Gabor responses with different parameters. We assume that the random feature vector follows a normal distribution with average $\bar{v}$ and covariance matrix Σ, $v_{(x,y)} \sim \mathcal{N}(\bar{v}_{(x,y)}, \Sigma_{(x,y)})$.

To select the set of dominant textures for a certain image region, we could select parameters (θ, f, σ) from Eq.(3). However this strategy do not perform well in the discrimination between different targets (the parameter distribution is concentrated in a narrow range which reduces the capability to discriminate the modeled object from others). To enforce variability in the parameter space and still be able to adapt the representation to the particular object under test, we will sample uniformly one of the parameters and perform a 2D search in the remaining dimensions. This strategy, used in [9], is denoted by "Extended Information Diagram", and is based in the concept of **Information Diagram**[12].

Extended Information Diagram The "Extended Information Diagram" function presented in [9] is defined as:

$$\mathrm{EID}_{x,y}(\theta, \sigma, \gamma) = |G_{\theta,\gamma,\sigma}(x, y)|$$

Then, in slices of EID are computed the local maxima. The coordinates of the local maxima are chosen as "good" Gabor function's parameters because they represent the target's dominant texture. In this work we prefer to consider the frequency f instead of γ, because there are much more and well located local maxima in (θ, σ, f). We redefine the Extended Information Diagram as:

$$\mathrm{EID}_{x,y}(\theta, \sigma, f) = |G_{\theta,f,\sigma}(x, y)| \tag{15}$$

There are three ways of slicing the EID function(θ slices, σ slices, and f slices), but we select the θ-ID because is the method with best classification performance in [9]. Then we denote θ-ID, as a slice of the EID function, keeping constant the orientation, $\theta = \theta_0$:

$$\theta\text{-}\mathrm{ID}_{x,y}^{\theta_0}(\sigma, f) = \mathrm{EID}_{x,y}(\theta_0, \sigma, f)$$

In Fig. 5 we show some examples of the θ-ID computed at an eye's center point.

Fig. 5. Three examples of θ-ID, and θ slices in the parameter space from left to right

Searching Multiple Information Diagrams The strategy to find good parameters for each target is based on uniformly discretizing θ, and search local maxima in the resulting set of θ-ID's. A set of θ-IDs for $\mathcal{T} = \{\theta_1, \cdots, \theta_i, \cdots, \theta_l\}$, at point (x, y) is given by:

$$\Theta\text{-ID}_{x,y}^{\mathcal{T}} = \{\theta\text{-ID}_{x,y}^{\theta_1}, \cdots, \theta\text{-ID}_{x,y}^{\theta_i}, \cdots, \theta\text{-ID}_{x,y}^{\theta_l}\} \qquad (16)$$

The several $\theta_i \in \mathcal{T}$ are uniformly spaced in the range $[0, \pi)$. Then we compute the parameters of the two higher local maxima:

$$(\hat{\sigma}_{i,1}^{\max}, \hat{f}_{i,1}^{\max}) = \arg \max_{\sigma, f} \theta\text{-ID}_{x,y}^{\theta_i}, i = 1, \cdots, l \qquad (17)$$

$$(\hat{\sigma}_{i,2}^{\max}, \hat{f}_{i,2}^{\max}) = \arg \max_{\sigma, f, \sigma \neq \hat{\sigma}_{i,1}, f \neq \hat{f}_{i,1}} \theta\text{-ID}_{x,y}^{\theta_i}, i = 1, \cdots, l \qquad (18)$$

Using the parameters computed in Eqs.(17) and (18), the feature vector is:

$$v_{(x,y)} = \left(v_{(x,y)}^1, \cdots, v_{(x,y)}^k, \cdots, v_{(x,y)}^{4m}\right)^T; \quad v_{(x,y)}^{4k-3} = \text{Re}(G_{\theta_i, \hat{f}_{i,1}^{\max}, \hat{\sigma}_{i,1}^{\max}}); \qquad (19)$$

$$v_{(x,y)}^{4k-2} = \text{Im}(G_{\theta_i, \hat{f}_{i,1}^{\max}, \hat{\sigma}_{i,1}^{\max}}); \quad v_{(x,y)}^{4k-1} = \text{Re}(G_{\theta_i, \hat{f}_{i,2}^{\max}, \hat{\sigma}_{i,2}^{\max}}); \quad v_{(x,y)}^{4k} = \text{Im}(G_{\theta_i, \hat{f}_{i,2}^{\max}, \hat{\sigma}_{i,2}^{\max}})$$

3.2 Interest point matching

In the training stage we compute the target model $(\bar{v}, \Sigma)$. Then, in the matching stage we compute the Mahalanobis distance between the target model and the interest point feature vector $v_{(x,y)}$. Because we assumed the feature vector follows a normal distribution, the Mahalanobis distance follows the chi-square distribution. By picking a confidence value from chi-square statistics table, we are also setting the Mahalanobis distance threshold to accept or reject the hypothesis of a target being located at interest point (x, y). The retrieved interest points are those below the Mahalanobis threshold.

3.3 Scale invariance

The feature vector in Eq. (19) is composed by Gabor filter responses. We want the feature vector scale invariant, so we must find the constraints for a Gabor response being scale invariant. Following the reasoning proposed in [13], consider two images: $I(x, y)$, and an homogeneously scaled version of $I(x, y)$. The new image is scaled by a factor a as $I_s(x, y) = I(ax, ay)$. The response of the scaled image at point (x_0, y_0), $G_{\theta, f, \sigma}^{I_s}(x_0, y_0)$ is

$$G_{\theta, f, \sigma}^{I_s}(x_0, y_0) = (I * g_{\theta, f/a, \sigma a})(ax_0, ay_0) = G_{\theta, f/a, \sigma a}^{I}(ax_0, ay_0) \qquad (20)$$

From Eq. (20) we can see that the Gabor response remains constant in the scaled image if we change the scale parameter σ of the Gabor filter to σa, and also changing the frequency f to f/a. If we see these value changings in terms of $\gamma = 1/\sigma f$, it means the scale invariant parameter γ mantains the same value.

We do not know the scale factor a in Eq.(20), so a common approach to solve this problem is to define an intrinsic scale of the interest point. We define the scale ratio

$$\rho_{i,j} = \frac{\hat{\sigma}_{i,j}^{\max}}{\sigma_{intrinsic}}, j = 1, 2; i = 1, \ldots, l \tag{21}$$

where $\hat{\sigma}_{i,j}^{\max}$ is computed in Eqs.(17) and (18), and $\sigma_{intrinsic}$ is computed in Eq.(12). The ratio $\rho_{i,j}$ is computed during the training stage, and is an indicator of the right scale parameter of Gabor filter in new images. Ratio values $\rho_{i,j}$ must keep the same value in scaled images, so

$$_s\hat{\sigma}_{i,j}^{\max} = {}_s\sigma_{intrinsic}\rho_{i,j}, j = 1, 2; i = 1, \ldots, l \tag{22}$$

$$_s\hat{\sigma}_{i,j}^{\max} = \hat{\sigma}_{i,j}^{\max} \frac{{}_s\sigma_{intrinsic}}{\sigma_{intrinsic}} \tag{23}$$

$$_s\hat{f}_{i,j}^{\max} = \hat{f}_{i,j}^{\max} / \left(\frac{{}_s\sigma_{intrinsic}}{\sigma_{intrinsic}} \right) \tag{24}$$

where the left subscript s in ${}_s\sigma_{intrinsic}$, ${}_s\hat{\sigma}_{i,j}$ and ${}_s\hat{f}_{i,j}^{\max}$ stands for scaled image. We see that the scale factor a is the ratio between intrinsic scales.

3.4 Rotation invariance

After the computation of the scale and frequency parameters in Eqs.(23) and (24), we must set the $\theta_i \in \mathcal{T}$ (i.e. select an appropiate $\Delta\theta$).

We tackle the rotation invariance by: (i) Selecting a small enough $\Delta\theta$, and (ii) comparing the target model with all the possible orientation shifts $\delta_i = i\Delta\theta, i = 0, \ldots, l-1$ of the orientation parameter of the feature vector $v_{(x,y)}$, and then picking the closest redirected vector to the target model. So we compute

$$\hat{\delta} = \arg \min_{\delta} (\text{shft}(v_{(x,y)}, \delta) - \bar{v}) \Sigma^{-1} (\text{shft}(v_{(x,y)}, \delta) - \bar{v})' \tag{25}$$

where $\text{shft}(v_{(x,y)}, \delta)$ is a function that performs a δ orientation shift of the vector $v_{(x,y)}$. The selected feature vector is $v_{(x,y)} = \text{shft}(v_{(x,y)}, \hat{\delta})$.

4 Results

We present tests in order to: (i) Verify the properties of the top-down saliency model, (ii) assess the invariance of the interest point model, and (iii) show the feasibility of the architecture presented in Section 1. The tests performed in this work use 82 subjects from the AR face database[14], where half of them are used for learning the saliency model and the target model, and the half remaining for the top-down guiding search and target classification. Both the saliency model and target model are learnt in a supervised manner, computing the models in groundtruth points. In Figure 1 we see an example of the points(regions) we are looking for in the images.

4.1 Saliency model tests

We present two tests, to assess the most important properties of our saliency model: (i) removal of points very different from the model with very few rejections, and (ii) scale invariance of the γ-signature. In both tests we perform an eye, nose, and nostril point pre-selection. The first step is a "generic" bottom-up procedure to select the initial set of salient points, computing the local maxima in space of the Difference of Gaussian operator applied at several scales in the image. In the training step we learn the mean γ-signature of the target, and the adequate distance threshold. Points with distance to the model less than the threshold are the candidates for further processing. To evaluate the performance of each experiment we count the number of hits (successful detections) in the test set. Given a saliency model, a distance function and an image point, a hit exists if there is a distance to the model below threshold inside a circle of radius r around the image point. We use the Euclidean distance, and $r = 4$ pixels.

Facial Point	Performance	% of bottom-up SP
Eye	100	20.36
Nose	100	19.97
Nostril	100	22.06

scale change(octaves)	Performance(%)
-0.5	93.49
-0.25	100
0	100
0.25	100
0.5	100

Table 1. Results of top-down guiding search of facial landmarks(left side) and scale invariance test(right side)

In the left side of Table 1 we can see that we do not miss points closer to the facial landmarks we are looking for, and we remove in average 79.21% of the "generic" bottom-up salient points.

To check the scale invariance of the γ-signature, we compute the success rate in rescaled images while mantaining the γ-signature model learned in the original scale images. In the right side of Table 1 we observe that the methodology proposed is tolerant to scale changes up to ± 0.5 octaves. Because of the very small size of the nostrils in the lowest resolution images, we miss some of them.

4.2 Target classification tests

We perform tests in order to verify: (i) the hypothesis confidence when accepting or rejecting targets, (ii) the rotation invariance of the feature vector, and (iii) the scale invariance of the feature vector. We compute values in whole images(i.e. Mahalanobis distance is computed for every pixel), and set $\Delta\theta = \pi/24 = 7.5°$.

Hypothesis confidence We set the confidence threshold to 99.9%, and compute the Mahalanobis distance in the test set images. We mark a hit if there is a selected point inside a circle of radius $r = 4$ pixels around the target. The points outside the circle are

marked as false positives. In table 2 we show the results when looking for eye, nose, and nostril points. It is important to remark the good recall rates achieved, showing experimentally the feature vector follows the Gaussian assumption.

Facial Point	Recall(%)	Precision(%)	130°Rot Recall
Left eye	100	64.36	97.56
Right eye	97.56	50.33	97.56
Nose	92.68	79	92.68
Left nostril	87.8	60.68	87.8
Right nostril	82.92	72.32	82.92

scale change(octaves)	Recall(%)
-0.5	83.19
-0.25	92.19
0	92.19
0.25	92.19
0.5	91.14

Table 2. Precision and recall rates of facial landmark matching(left side), and scale invariance test(right side)

Scale invariance To check the invariance to scale transformations, we compute the recall rate in rescaled images mantaining the object model learned in the original size images. In the right side of Table 2 we can see the average recall of all facial landmarks.

The decreasing performance when reducing image size is because of the tiny size of nostrils in the lowest resolution images. Because of the tiny size, in some cases we can not compute the intrinsic scale. For the remaining facial landmarks(eyes and nose) the recall remains constant.

Rotation invariance To verify the invariance to rotation transformations, we compute the recall rate in the image test set rotated by the angle 130°, keeping the object model learned in the standard pose images. In the left side of table 2, the rigthmost column shows that the recall remains almost constant for an angle that is not sampled in the model.

4.3 Whole architecture tests

The integrated test of the architecture performs: (i) local maxima in space of the Difference of Gaussians at several scales, (ii) point pre-selection using the saliency model, and (iii) in the selected points compute the interest point representation, followed by the hypothesis decision. We see in Table 3 that recall values remains almost the same when comparing vs. Table 2, but precision is substantially improved. The results show the feasibility of the proposed architecture, sketched in Figure 1.

5 Conclusions

In this paper we propose an architecture for interest point recognition. We show how to apply the Gabor filters in order to solve two of the most important issues in low-level recognition approaches: (i) saliency computation, and (ii) local feature classification. Using the Gabor function as source of information, we define a new saliency function that is able to introduce some prior(top-down) information during the recognition

Facial Point	Recall(%)	Precision(%)
Left eye	100	74.63
Right eye	97.56	57.99
Nose	90.24	100
Left nostril	87.8	67.94
Right nostril	82.92	94.47

Table 3. Performance of the whole architecture

process. This top-down information reduces the computational complexity in visual search tasks and improves recognition results. The appearance based saliency function presented removes points very different from the model and have very few rejections of "good" points. The function proposed is invariant to position, orientation and scale of the object we are searching. We describe a method to compute the intrinsic scale of an interest point using the saliency function. We propose to use the Gabor function also for representing and classifying targets. The presented representation is able to match successfully interest points, and is invariant to image rotations and scalings.

References

1. Lowe, D.: Object recognition from local scale-invariant features. In: Proc. IEEE Conf. on CVPR. (1999) 1150–1157
2. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE PAMI **19** (1997) 530–534
3. Mikolajcyzik, K., Schmid, C.: An affine invariant interest point detector. In: ECCV, Springer (2002) 128–142
4. Weber, M., Welling, M., Perona, P.: Towards automatic discovery of object categories. In: Proc. IEEE Conf. on CVPR. (2000)
5. Triggs, B.: Detecting keypoints with stable position, orientation, and scale under illumination changes. In: ECCV. Volume 3. (2004) 100–113
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE PAMI **20** (1998) 1254–1259
7. Kadir, T., Brady, M.: An affine invariant salient region detector. In: ECCV. Volume 1. (2004) 228–241
8. Chun, M., Wolfe, J.: Visual attention. In Goldstein, E., ed.: Blackwell handbook of perception. Blackwell Publishers (2000)
9. Moreno, P., Bernardino, A., Santos-Victor, J.: Gabor parameter selection for local feature detection. In: Proc. IbPRIA'05, Estoril, Portugal (2005)
10. Lindeberg, T.: Feature detection with automatic scale selection. International Journal of Computer Vision **30** (1998) 79–116
11. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. International Journal of Computer Vision **30** (1998) 117–154
12. Kamarainen, J.K., Kyrki, V., Kälviäinen, H.: Fundamental frequency gabor filters for object recognition. In: Proc. of the 16th ICPR. (2002)
13. Kyrki, V., Kamarainen, J.K., Kälviäinen, H.: Simple gabor feature space for invariant object recognition. Pattern Recognition Letters **25** (2004) 311–318
14. Martinez, A., Benavente, R.: The ar face database. Technical report, CVC (1998)