Article

Symmetry as an Intrinsically Dynamic Feature

Vito Di Gesu [1,2,†], **Marco E. Tabacchi** [1,3,]*** **and Bertrand Zavidovique** [4]

[1]   DMA, Università degli Studi di Palermo, via Archirafi 34, 90123 Palermo, Italy
[2]   CITC, Università degli Studi di Palermo, via Archirafi 34, 90123 Palermo, Italy
[3]   Istituto Nazionale di Ricerche Demopolis, via Col. Romey 7, 91100 Trapani, Italy
[4]   IEF, Université Paris IX–Orsay, Paris, France; E-Mail: bertrand.zavidovique@u-psud.fr (B.Z.)

[†] Deceased on 15 March 2009.

[*] Author to whom correspondence should be addressed; E-Mail: metabacchi@demopolis.it.

Abstract: Symmetry is one of the most prominent spatial relations perceived by humans, and has a relevant role in attentive mechanisms regarding both visual and auditory systems. The aim of this paper is to establish symmetry, among the likes of motion, depth or range, as a dynamic feature in artificial vision. This is achieved in the first instance by assessing symmetry estimation by means of algorithms, putting emphasis on erosion and multi-resolution approaches, and confronting two ensuing problems: the isolation of objects from the context, and the pertinence (or lack thereof) of some salient points, such as the centre of mass. Next a geometric model is illustrated and detailed, and the problem of measuring symmetry in a world where symmetry is not perfect nor the only attention trigger is tackled. Two algorithmic lines, based on the so-called symmetry kernel and its evolution with pattern warping, and by correlation of blocks with varying sizes and positions, are proposed and investigated. An extended illustration of the power of symmetry as a feature, based on face expression recognition, concludes the paper.

1. Introduction

Image processing by computer is the first step towards artificial vision by systems after data acquisition. It consists mainly of extracting primitive variables–so called features–to be further gathered and sorted into projections of candidate objects or phenomena, and then situations. These projections will in turn be inverted to obtain real world correspondents in order to trigger some action by the system. Thus, such features can be viewed as principal axes in a multidimensional perception/decision space that most often remains implicit. Basic features are related to shape (points of interest, edges and regions) or displacement (depth and motion). More evolved features rely on earlier variables, like region or edge, which may easily stem more from color or texture than from the mere intensity. Such features are often conjectured as biological instances (e.g. points of interest *vs.* fixation points) likely localized in a subset of the brain, but usually obey independent models–geometric, analytic or informatics–and are associated to sets of specific algorithms to minimize a distance to the latter models.

Our research for the last seven years confirms that symmetry can be considered one of these features, in the same vein as motion or range are: first, it is a property that characterizes the invariance of a given system, giving range to some description of it. Second, it is one of the most prominent spatial relations perceived by human beings. Psychologists of perception, following the Gestalt school, assign a relevant role to symmetry for attentive mechanisms in both visual and auditory systems (See [1] for historical background, and [2] for a review). Indeed human beings are far more sensitive to symmetry than to translation (for instance see Figure 1). A significant bibliography on parallel human and machine symmetry detections can be found in [3,4].

Figure 1. Left and right halves are symmetric (left) and shifted (right).
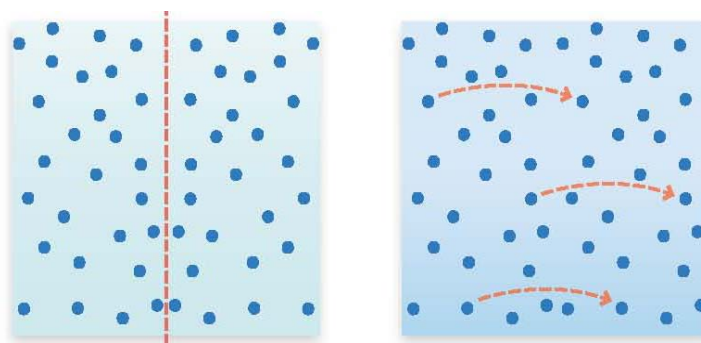
Now, invariance refers to a given transformation:

let be S a system, p a variable and T a transform, S is symmetric in p iff. $S(p) = S(T(p))$.

And then the question arises to know which kind of transformation. For instance, the first guess in dealing with 2 or 3-D patterns would be to pick isometries, and an object is symmetric if it remains unchanged when permuting one of its parts. Figure 2 shows three examples of axial symmetry combined with local operations that turn out to mislead humans: Respectively concavity/convexity, black first /white first and color/texture. In the following will be shown that they may happen not to disturb a system, depending on how their symmetry detector is programmed.

Figure 2. Symmetry requires bumps to become holes (left), dark parts to be seen first (middle), pop-up of texture for vertical elements and color for diagonal elements (right).

Conversely, Figure 3 shows three examples of transforms combined with the axial symmetry that would not perturb any human and would puzzle systems more: respectively adding noise or texture, bending, projecting.

Figure 3. Noise or bending do not disturb global symmetry, neither do projections for local symmetry.

(a)   (b)   (c)
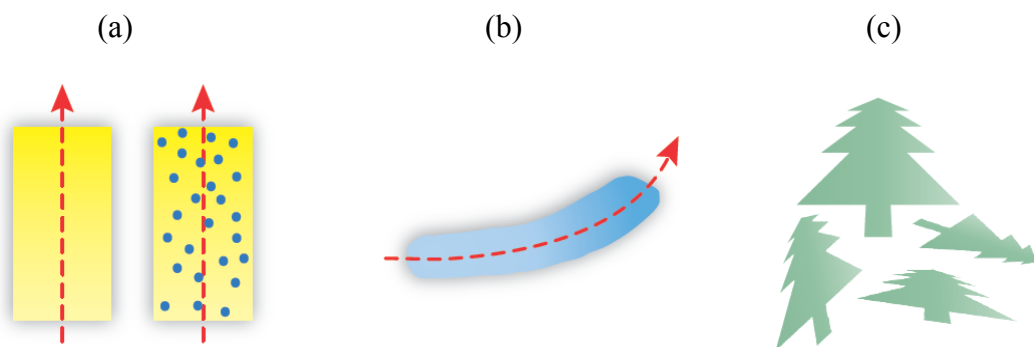
The following concept is the cornerstone of the present paper: symmetry can be conceived as a feature in artificial vision—our studies have confirmed this, as summarized in the following sections, something which is an advance in understanding the key role of the symmetry transform. To this aim we tackled the potential contradiction between local and global symmetry in a given vision task. In terms of computational paradigm, it translates into a contradiction between symmetry from edges and symmetry from grey levels or texture. Let us stress that in psycho-physiology of vision it is common knowledge that parts near edges contribute more to symmetry detection by humans than parts near a conjectured axis, themselves contributing more than parts in between [5].

The paper is organized as follows: Sections 1 and 2 concern the estimation of symmetry by computer, targeting machine attention. The algorithms there outlined were designed to be adapted on demand, and are more sensitive to contours and inner regions. A continuous link between them in the process of symmetry evaluation is made explicit. Two modes of warping patterns that both highlight the role of symmetry as a dynamic feature have been presented, namely erosion and multi-resolution. This study will be related in Section 1. Some applications to clustering from symmetry indexes strengthen the idea that symmetry is definitely a feature, similar to motion or depth. But the first programs raise two new problems that are explained and discussed in Section 2: when analyzing a scene, objects, symmetric or not, are not in isolation at first glance. Also, even if patterns of interest

could be easily extracted, there is no real reason to focus on some specific geometric points like the centre of mass to derive the symmetry. This agains leads to put forward the transform triggering symmetry by performing it explicitly, before detection by mere correlation. Such procedure is well grounded in a geometric model that is detailed and illustrated through detection results in sub-section 2.1. Subsection 2.2 is then devoted to the question of measuring symmetry, since in real life perfect symmetry is rare, and it does not usually serve as the sole attention triggering. For an extended illustration, Section 3 details an application to face expressions clustering—A fundamental and permanent human activity-based on a collection of partial symmetry measures. Results of two concurrent processes are evaluated with regards to the ideas proposed in the paper.
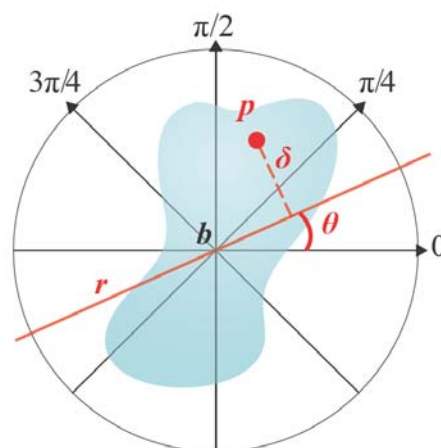
2. Symmetry Constancy

In order to introduce some dynamics in the symmetry detection process by integrating edge symmetry (symmetry computed using the profile of an object) to region symmetry (symmetry computed used the whole mass information), we choose to work with a local detector, the Symmetry Transform, based on the notion of inertia. This operator, proposed by DiGesù and Valenti in 1996 [6], works according to Figure 4: given a center–e.g., the center of mass–of an object O and the direction θ of an axis through this point, the operator outputs the momentum of inertia respective to the axis $T(O,r)$:

$$T(O,r) = \int_O \delta^2\left(\underline{p}, r(\underline{b},\theta)\right) \sum m\left(\underline{p}\right) \sum d\underline{p}$$

(1)

where O is the pixmap object, r is the axis passing through b at angle θ, δ is a function of the pixel p and its geometric distance to the axis (b,θ), m is a measure of intensity or texture in p.

Figure 4. Symmetry and momentum.

The circular symmetry can be evaluated for instance in computing the variance σ_T (or 1- σ_T) of T over θ as shown by the three examples in Figure 5.

One interesting feature of this operator is that its discrete version, actually used in image processing, amounts to a convolution, the so called Axial Moment Filter. Indeed the distance part becomes a weight w and the mass is the grey level g:
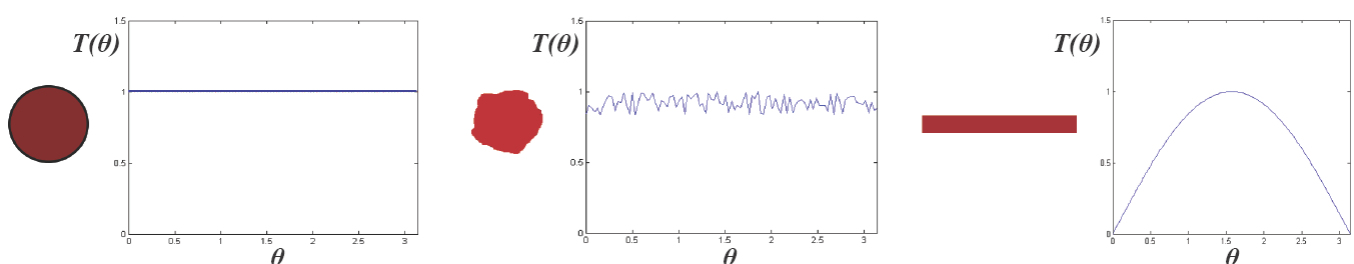
$$T_{i,j}(k) = \sum_{r=-h}^{+h}\sum_{s=-h}^{+h} w_{r,s}(k) \times g_{i-r,j-s} \qquad w_{r,s} = \left(r \times \sin\frac{k\pi}{n} + s \times \cos\frac{k\pi}{n} \right)^2 \tag{2}$$

with the circular version:

$$T_{i,j} = 1 - \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(T_{i,j}^k\right)^2 - \left(\frac{1}{n}\sum_{k=1}^{n}T_{i,j}^k\right)^2} \tag{3}$$

We will prove in Subsection 1.2 that this AMF filter shows interesting properties in terms of balancing between the respective influences of edges and mass.

Figure 5. Three patterns and their respective variations of *T* over *θ*.

2.1. Dynamics Through Erosion

Symmetry Transform from Equation 1 can be combined with iterated erosion to give the so called Iterated Object Transform (*IOT*) that we proposed in 2004 [7]:

$$IOT_{\theta,1}(X) = S_\theta(X)$$
$$IOT_{\theta,m}(X) = S \circ (E)_\theta^{m-1}(X) \quad for \ m > 1 \tag{4}$$

where o^n denotes the *n*-iterative application of an operator, *S* is the Symmetry Transform, *E* is the classical Erosion operator and *θ* is the slope of the axis.

Thus, *IOT* computes the Symmetry Transform, *S*, on steadily intensity reduced versions of the input image. Indeed the morphological erosion is classically given by:

$$E(X) = \{x \in X \mid C_1(x) \subseteq X\} \tag{5}$$

while the Soille operator was further used for its fuzzier nature:

$$E(X) = \left\{ x \in X \mid \min_{y \in C_1(x)}\left\{1 - \delta(g_y, f_y)\right\} \right\} \tag{6}$$

More global indexes can be computed to characterize the symmetry tendencies of patterns (e.g., the directions of maximal or minimal symmetry), as illustrated by Figure 6.

But we rather profit from the symmetry constancy along erosion to relate the overall symmetry of a given object. To that aim two variables are estimated, qualitatively complementary one to the other:

$$\eta_n(X) = \frac{\min_{\theta \in [0,\pi[}\left\{IOT_{\theta,n}(X)\right\}}{\max_{\theta \in [0,\pi[}\left\{IOT_{\theta,n}(X)\right\}}, \ \ circle_n(X) = 1 - var\left\{IOT_{\theta,n}(X) \mid \theta \in [0,\pi[\right\} \tag{8}$$

the first derivatives indicate the variation between successive erosions:

$$\partial(IOT_{\theta,n}(X)) = IOT_{\theta,n}(X) - IOT_{\theta,n-1}(X)$$
$$\partial(\eta_n(X)) = \eta_n(X) - \eta_{n-1}(X)$$
$$\partial(circle_n(X)) = circle_n(X) - circle_{n-1}(X) \tag{9}$$

Figure 6. Variations of IOT with the number of iterations and the angle for an image and its binary version.
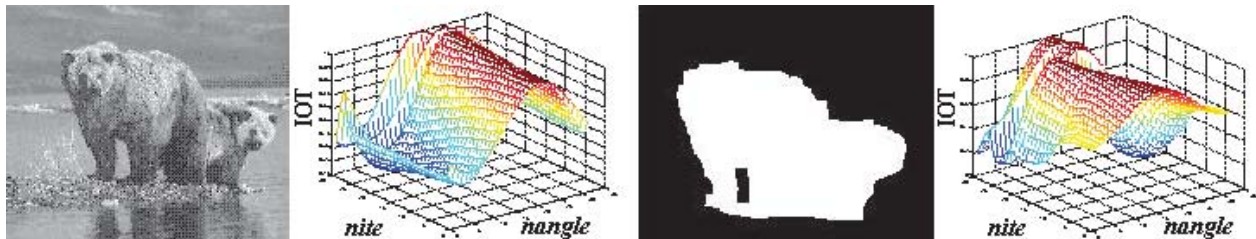
It should be noted that being S a convolution, ∂IOT outputs a High-Pass filtered version of $[S(X)]$. The edges are now local maxima of a High-Pass filtered version of *X*. As maximum finding does not commute with linear operations, *S(edges)* will not coincide with *max(∂IOT)*. Experiments have been completed in a systematic manner on synthetic and real images, addressing the shape parameter dynamics, as shown in Figures 7 and 8, respectively (see [8,9] for further details).

Figure 7. Elongation (red) and circle (green) variations with the number of erosions for the upper left patterns (circle, perturbed circle, rectangle, perturbed rectangle, star and perturbed star). Third row; the first derivatives in the rectangle cases and then the variations and their derivative for the random shape.
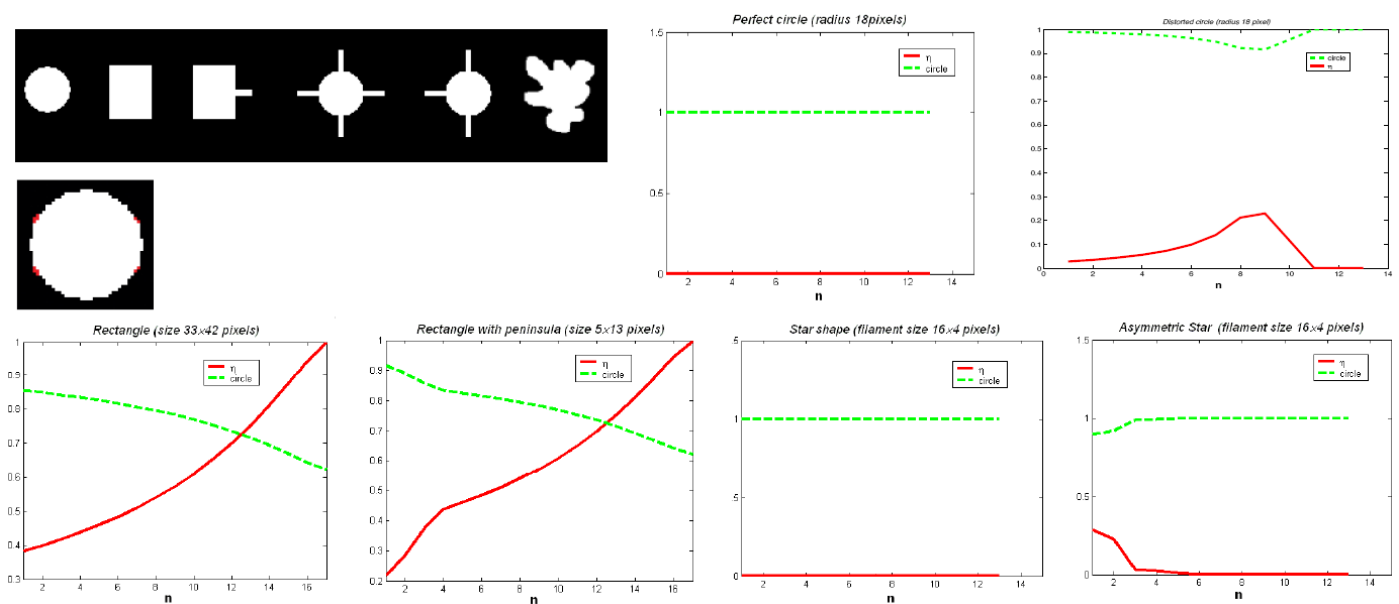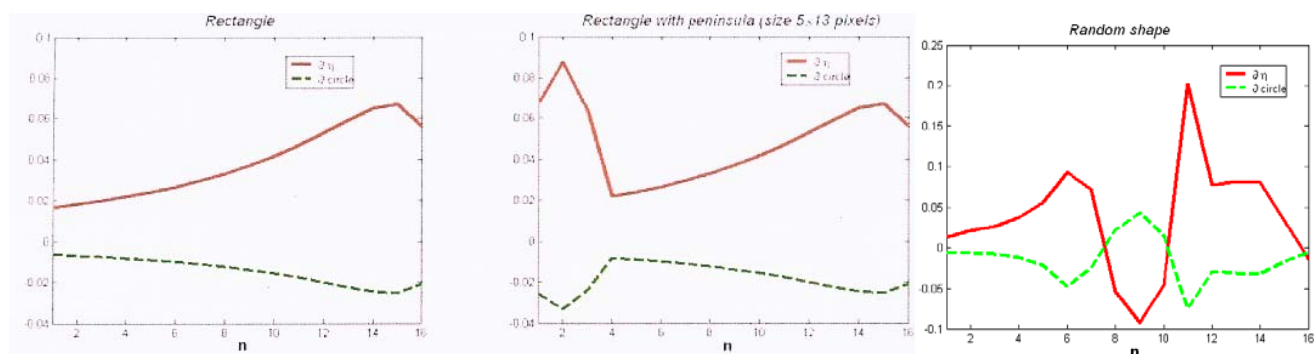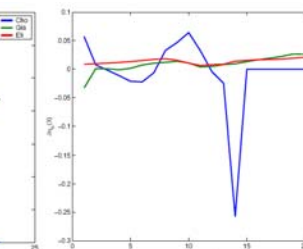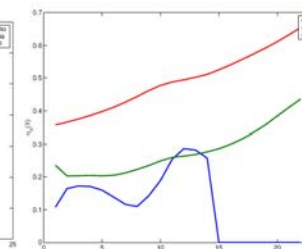
Figure 7. *Cont.*

Figure 8. Elongation (left) and circle (right) variations, along with the number of erosions for the corresponding images. Fourth row; circle of the cell images (blue, green, red in the order), their first derivatives and then again for elongations.
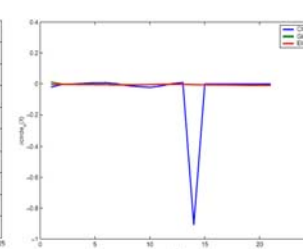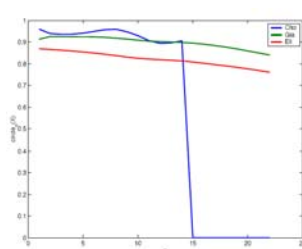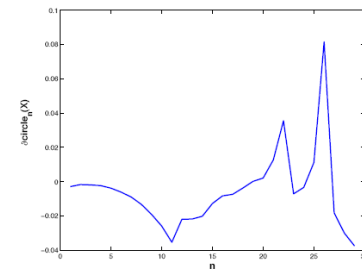
Another representation can be proposed in order to promote different pattern properties bound to the symmetry constancy, along with pattern warping. For instance, the beauty of a body is said to stem from a perfect symmetry. If the elongation, as defined above, is displayed in a polar diagram, steady maximal symmetry over erosions can translate the corresponding quality of a face, or, more likely, its pose. Figure 9a shows this representation of beauty where both min (green) and max (red) elongation are perf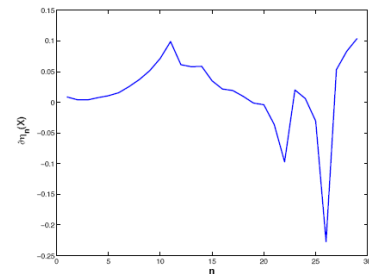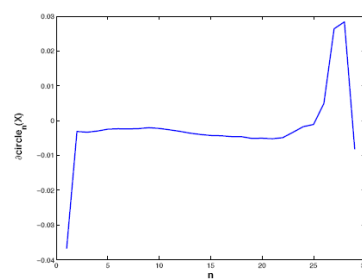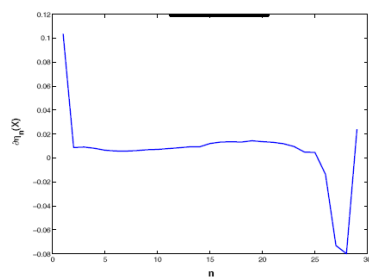ectly stable over erosion. Figure 9b, although representing a very pretty face too (according to alleged European standards), does not behave alike. Actually due to the head tilt in association with the decreasing importance of inner (symmetric) features along grey level erosion, min and max elongations are subject to bifurcation at the second and third steps.

Figure 9. Elongation in polar coordinates superposed to the corresponding pictures after erosion. The max and min symmetry axes are: (a) perfectly stable in case of perfect symmetry; (b) switching in the case of a tilted (asymmetric) head.

(a)

(b)
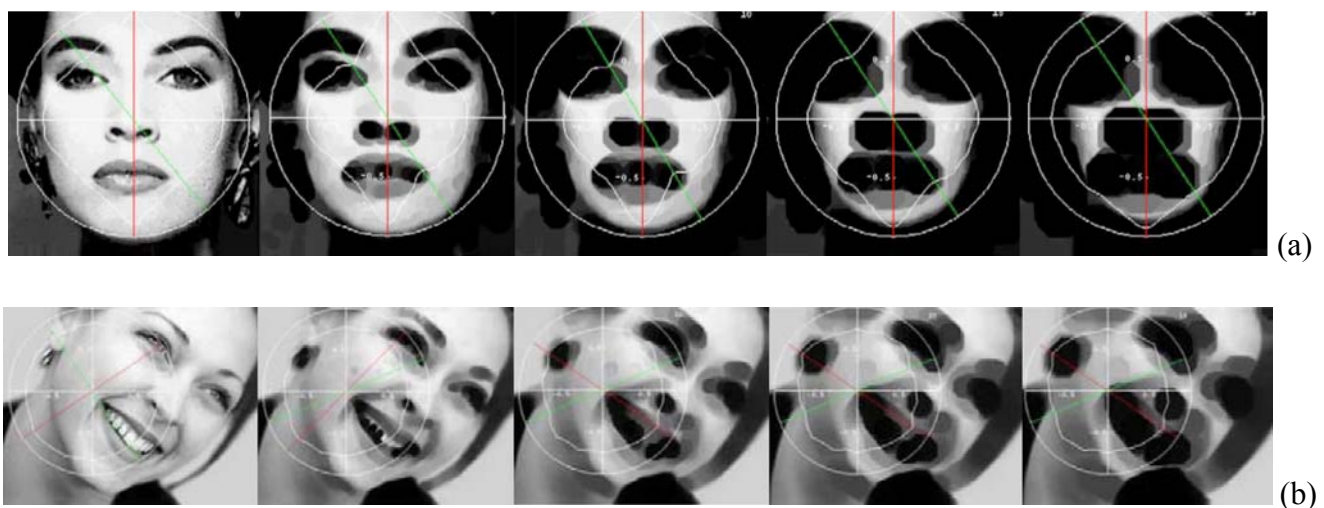
Many similar examples, while confirming that solid symmetry easily translates into a constancy of symmetry indexes along erosion, already suggest important properties such as:

1. When going from a regular pattern to a modified version of it, any perturbation gives a signature. Moreover, experiments about ratios of pattern size over defect size, or about the number of defects, allow precision assessment.
2. Curve variations follow, at least in a qualitative manner, tendencies of symmetry variations with the angle: constancy, monotony, smoothness *etc.*

After having reviewed enough of such basic cases, human beings can develop an expertise that leads to guess such properties from a hidden pattern, with results that are very close to automatic recognition in the simplest cases (e.g., quasi circular, pin like *etc.*) with respect to the angular sampling and the number of iterations. To confirm the latter ability and check its feasibility on machines, a clustering process was performed using elongation or circularity. Application is the recognition of *300* biological cells of the three types *A*, *B*, *C* shown in Figure 8–chondrocyte, giarda protozoan, myeloid leukemia. The automatic process, a tree classifier, is exactly identical to a previous one performed on

more classical parameters like convex hull, moments and eccentricity (both are detailed in [8]). Tables 1a and 1b display the confusion matrices, showing an average improvement from *0.70* to *0.83*.

Eventually a program based on adapted tiling of the picture was developed to extract faces from group photos. Although too much a priori knowledge is injected in the process, results for detection of objects from their symmetry are interesting (as shown in Figure 17, for instance) More details and references on this part are in [4] for the study and in [8] for the application.

**Table 1.** Confusion matrices for the cell recognition process.

	A	*B*	*C*			*A*	*B*	*C*
A	**0,65**	0,15	0,20		*A*	**0,72**	0,21	0,07
B	0,10	**0,70**	0,20		*B*	0,20	**0,80**	0,00
C	0,10	0,15	**0,75**		*C*	0,01	0,02	**0,97**

a   **b**

2.2. Dynamics Through Multi-Resolution

Among all the possible continuous links between edges and regions, erosion is rather edge-oriented, as it gnaws at outlines. A second classical process for warping patterns is more bound to regions as it consists in sub-sampling the whole object and consider patterns at various resolutions. This approach exploits the pyramidal data structure, created by stacking multiple versions of the same picture, progressively reduced by increasing the pixel size either physically (e.g., defocusing) or by means of computing (the so called "numeric filtering"). Thus this structure is defined by two mappings: one for topology F, that determines what neighborhood at layer n will generate the pixel (father) at layer $n+1$ above (see Figure 10a for two examples), the other for intensity V that computes the father's value from the sons' value, using filters such as max, min, and, or, average, median, laplacian. Figure 10b compares respective results of erosion (using a disk kernel) and sub-sampling, Figure 10c displays a pyramid of similar patterns. Here again the first idea is to check parts where symmetry maintains itself over layers.

Figure 10. Defining the pyramidal warping: (a) examples of father/sons topology; (b)comparison of three evolutions of a same image under erosion (top) and multiresolution (bottom); (c) symmetry constancy over pyramid layers.
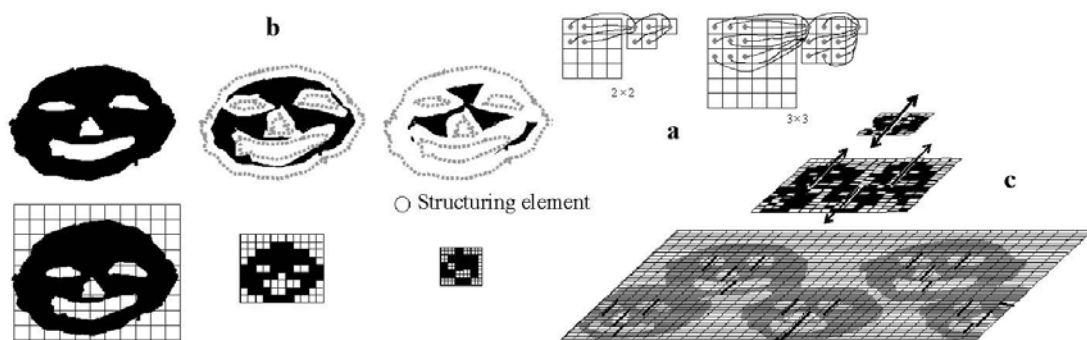
Two classes of problems should be addressed when working with this symmetry detection paradigm: first, some frequency and phase shifts can be introduced, due to sub-sampling; second, which local detector to choose and how to combine it with the pyramid. Figure 11 summarizes how an adequate choice of frequency-*i.e.*, layer in the pyramid - preserves more or less information. As an example it could be reasonable to question how much information is left in the three last layers of the pyramid of Figure 11.

Figure 11. A well tuned in Frequency multiresolution.

Figure 12 illustrates how a shift by a half block can jeopardize the whole symmetry depending only on whether blocks overlap (b) or not (a).

Figure 12. (a) and (b) phase problems *vs.* symmetry evolution; (c) global symmetry of local asymmetries
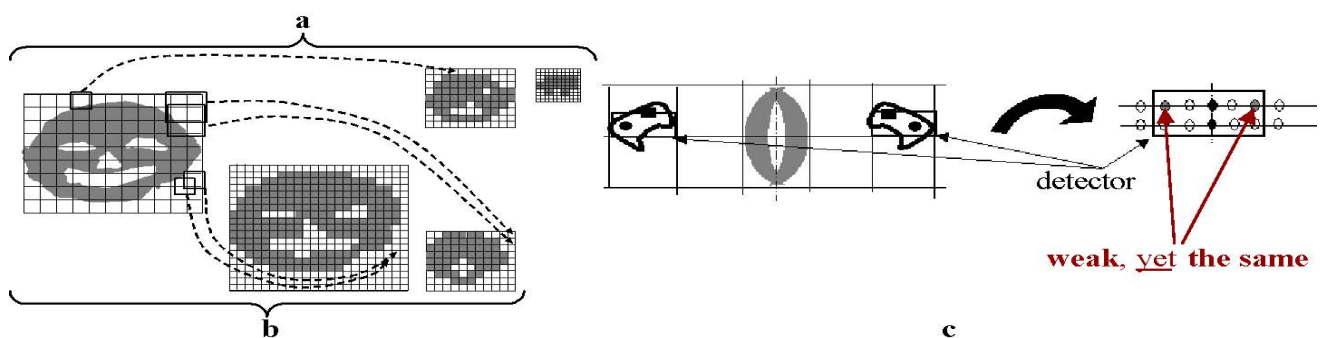
In order to actually design the pyramidal process, after frequency and phases have been tuned, only the combination of symmetry with the pyramid remains to be done. There are three sensible choices:
- direct: building the pyramid and running the symmetry operator *S* layer by layer;
- indirect: running *S* at the bottom (image) and then building the pyramid from these values;
- hierarchical: recursively running *S* to build the pyramid.

Let us underline that the choices above are independent from the question of local *vs.* global symmetry mentioned in introduction of the paper; a detector may give a weak response locally and yet one high enough for global symmetry, as illustrated in Figure 12c. This makes the selection of a local detector much easier. Within a frame of template matching it becomes obvious that any convolution kernel showing symmetry, either axial or radial depending on the application, would do fine. Examples of such kernels are:

$$\begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 2 & 3 \\ 2 & 0 & 2 \\ -3 & -2 & -3 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Nevertheless, the axial moment filter has interesting properties in terms of weight repartition. Trivial combinatory computing exhibits a recursive expression of the mask expansion as outlined here below:

...	...	...	...	...	...	...	...	...	...	...	...
...	2	1	2	5	10	...	...	...	...	...	...
...	1	0	1	4	9	16	25	...	...	x_n	x_n+2n-1
...	2	1	2	5	10	17	26	...	...	x_n+1	x_n+2n
...	5	4	5	8	13	20	29	...	...	x_n+4	x_n+2n+3
...	10	9	10	13	18	25	34	...	...	x_n+9	x_n+2n+8
...	...	16	17	20	25	32	41	...	...	x_n+16	...
...	...	...	...	...	...	...	...	...	...	x_n+25	...
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	$2n^2$	...
...	...	...	...	...	...	...	...	...	...	...	$(2n+1)^2$

Considering that the expectation of the response is the result of the convolution with the uniform image, it is enough to compare the sum of coefficients on external fringes to that on the whole mask. Whence the formulas:

$$\Sigma_n = 8\sum_{k=1}^{n}(2k-1)^2 + 4n = \frac{4n}{3}(8n^2+1) \approx 10n^3 + n$$

$$M_n = \sum_{k=1}^{n}\Sigma_k = \frac{8n^2}{3}(n+1)^2 + \frac{2}{3}n(n+1) \approx 3n^4 + 5n^3$$

(10)

So the ratio Σ_n/M_n is around one half for usual mask sizes between *5* and *15*. If we consider the *11×11* version, the external two-pixels-large belt weighs *90%*. That allows balancing the region orientation of the multi-resolution process.

Extensive experimentations have been completed again both on synthetic images of various complexities, and then on real images. Synthetic images help checking all expected detectors' behavior w.r.t. frequency, phase and symmetry direction or precision, also making the processes more automatic. Figure 13 gives an idea of such tries.

Real pictures, mostly faces and crowded images, were used as trials in a frame of interest paradigm: any pixel or image part is considered interesting either if it is a local maximum of the symmetry response or if it leads to a response of the symmetry detector over a threshold (e.g., $\mu+2\sigma$, considering the distribution of responses over the whole image set). For example Figure 14a shows the evolution of points of interest with the sampling (*i.e.*, while climbing the pyramid) and Figure 14b shows again the impact of the computing scheme, with threshold and then local maxima.

Figure 13. Comparing pyramidal symmetry detection schemes on test images: (a) direct; (b) hierarchy; (c) hierarchy with overlap.

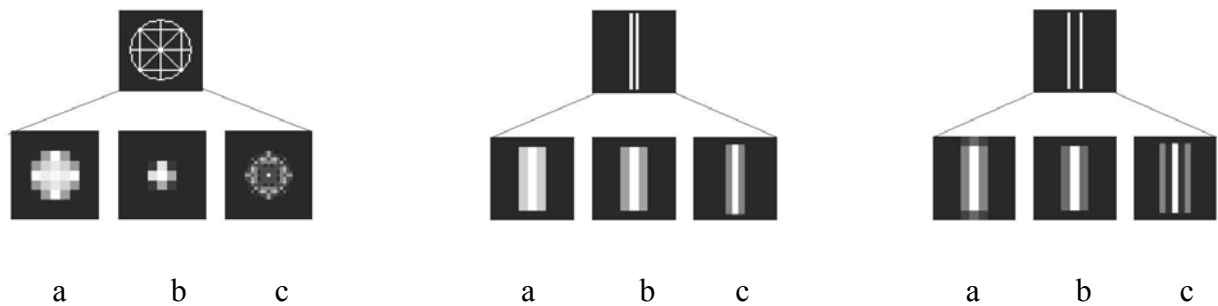a   b   c   a   b   c   a   b   c
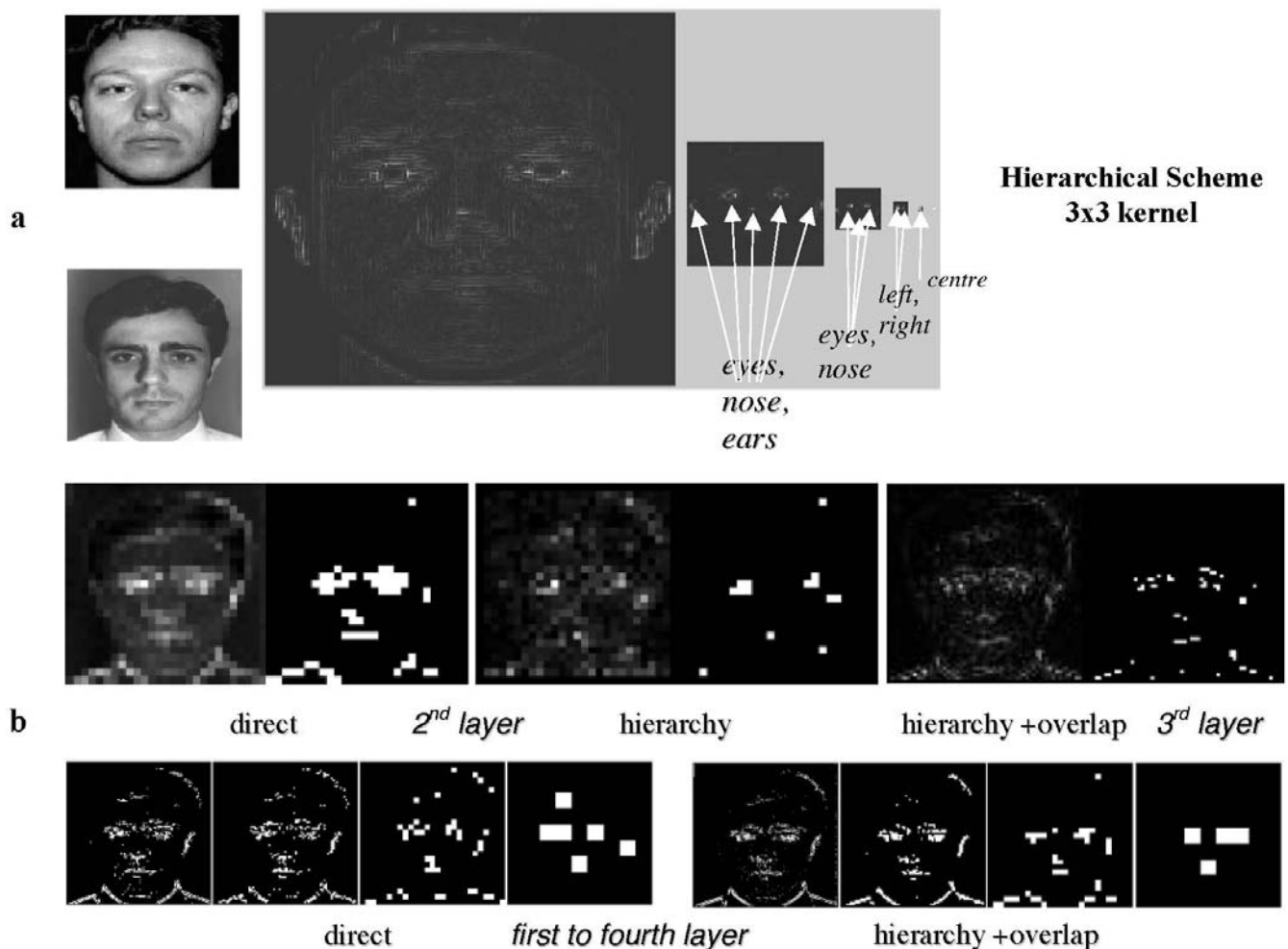
Figure 14. Points of interest from local symmetry of human faces: (a) the best evolution; (b) comparing schemes over layers.

Eventually, if all schemes produce qualitatively comparable results to the human eye, overlap still appears necessary to obtain precision, and hierarchy reveals symmetry better in a systematic way. As for the computing complexity, K and N being respectively the kernel and image sizes, hierarchy is $O(N^2)$ and hierarchy with overlap is $O(N^2K^2/3)$, to be compared to $O(N^2K^2)$ for the initial Symmetry Transform.

To conclude the study, the two reduction processes were compared. On faces as in Figure 15 the global symmetry is well exhibited by both techniques relying on the axis stability, whether the head is straight or tilted. In the pyramid case the stability is generally lost at the fifth layer, subject to resolution that sub-samples directions as well. The case of multiple symmetries and the robustness to ambiguities is even more interesting. The pyramid is more consistent in its results, and so less precise than Erosion. A case of texture is displayed Figure 16. If the maximum elongation remains constant, the corresponding angle switches from horizontal to vertical. It can be noticed that the maximum and second maximum responses of IOT at bifurcation are very close. Erosion is more sensitive and precise, while pyramid is more robust. That is illustrated on group photos in Figure 17. A significant difference between the two methods is in the computing time, which is twice as long for Erosion *ceteris paribus*. More details and a comprehensive bibliography about the topic discussed in this section can be found in [8] and [9].

3. Capturing Symmetry

In order to introduce symmetry as a visual feature of image processing, in the previous section we have shown how it can serve as an onset of attention, in the same vein as shape (edges or regions) or displacement (motion or depth). The constancy of the detector response over pattern reduction is the main basis of the feature extraction, for this feature to be proven dynamic. In the case of pattern warping by erosion, an implicit assumption is made: in order to evaluate symmetry, objects are isolated (e.g., well contrasted or quasi-binary images) or were previously conjectured through another feature. In the case of multi resolution, blind symmetry detection is more plausible thanks to the fact that the technique is inherently block-based, though the usual phase and frequency trimming considerations still apply. Moreover, as for the other conventional features, symmetry does not serve only as an attention trigger. In a complete vision process, features support comparisons and various inductions that require a measure of the phenomenon intensity: for contours this indicator is the global contrast or the edginess, for regions the uniformity and various moments, for motion the velocity or disparity fields, their density and their fit to a given model of geometric transform, all variables are computed during or after detection to measure the confidence in the results or to support comparison between patterns. Likewise patterns are symmetric to a certain degree, and several patterns can present more or less the same symmetries or the same symmetry variation with the angle or the conjectured axis or centre position. That leads to think of some sort of "symmetry measure" and, due to the very nature of the symmetry phenomenon, brings back to the common standard of "the closest maximum model"; in this instance, the most symmetric pattern included in a given pattern will be searched for. That offers three basic advantages:

- a model is made explicit to support optimality claims;
- this model puts forward:
    i.   the invariance to the transform again in relying on explicit comparison between the pattern and a transformed version of it;
    ii. the distance that could evolve into approximate comparison for similarity;
- pattern inclusion introduces set operations (such as Minkowski's), associating as a result logic and geometry.

Figure 15. Respective face symmetry constancies with erosion and multiresolution.
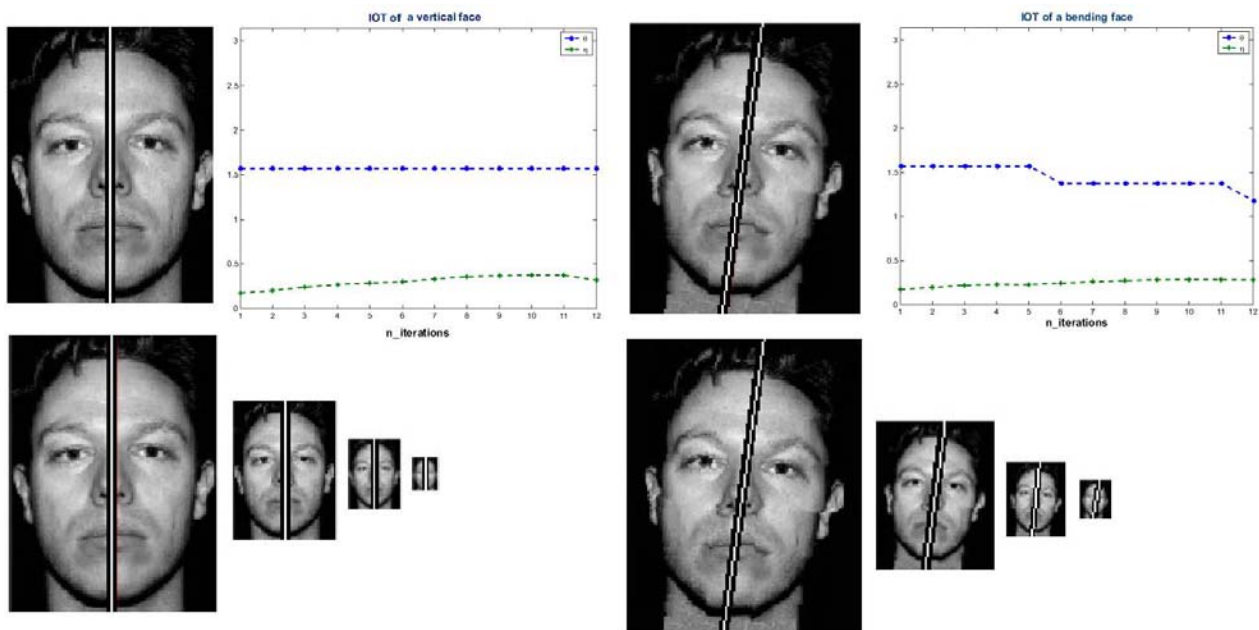
Figure 16. Comparing symmetry evolutions via the maximum elongation under erosion and multiresolution in the ambiguous case of a multiple symmetry texture.
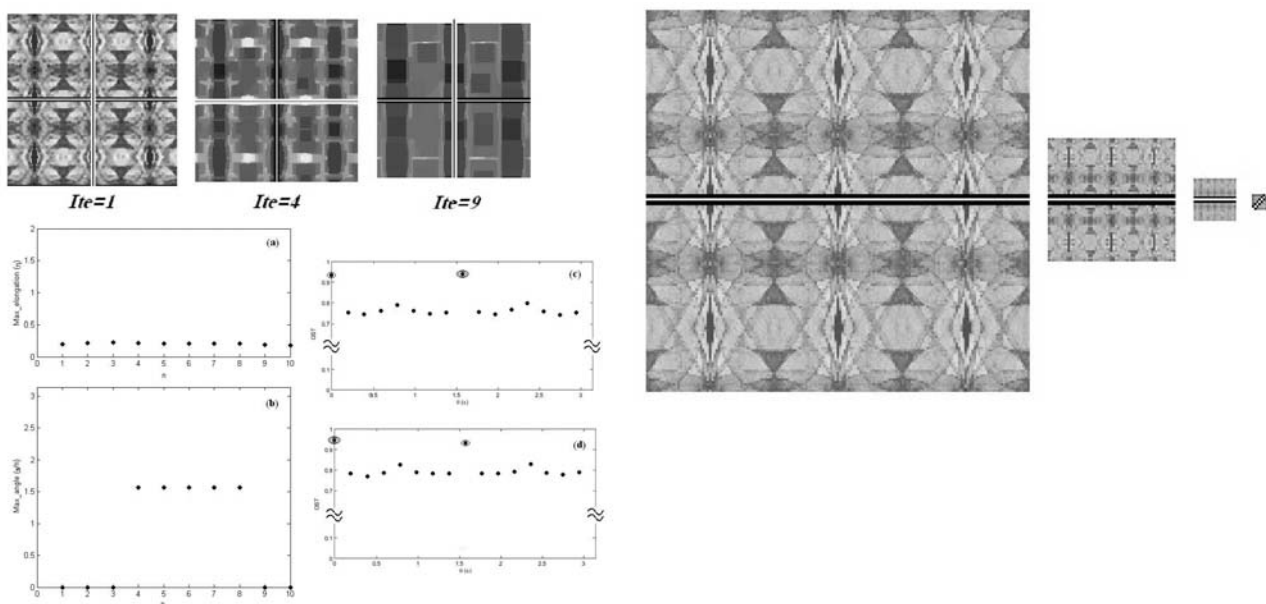

3.1. Optimal Symmetry Detection

With reference to the erosion in *IOT*, the symmetric kernel *SK* of a given pattern P is first defined. *SK(*P*)* is the *maximal* included *symmetric* pattern in P. Let us underline that the transposition to the pyramidal scheme is straightforward, considering the corresponding symmetric collection of blocks. In this case, taking in account the discussion outlined in the previous section, precision is preferred to robustness here, as stemming from the measure itself. The sketch of such an algorithm is not difficult to describe (see Figure 18).

Figure 17. Results of face extraction from a crowd by their symmetry constancy along resolution decrease and iterated erosion respectively.


Aiming to exhibit the kernel wrt a given axis, either directly or by prior rotation, symmetric bands are progressively added, their width related to the computing precision.

Figure 18. Graphic representation of a plausible kernel finding procedure: Union of couples of maximal bands apart the chosen axis.
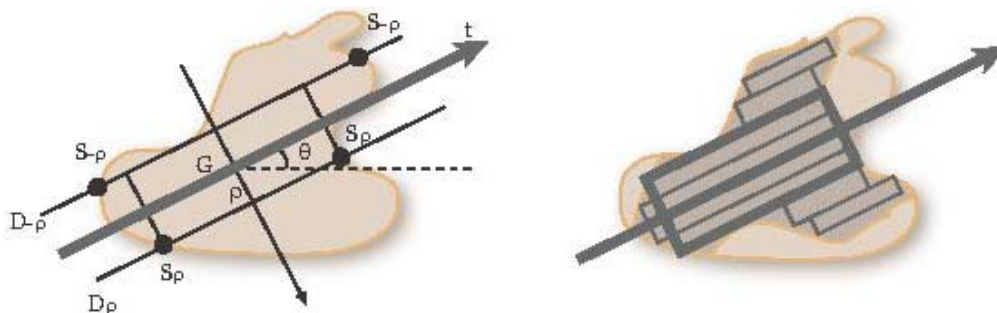
The measure is thus introduced using the "closer to the pattern equals to more symmetric" paradigm. The drawback of being subjected to an arbitrary centre position through the axis still remains, echoing the assumption of an isolated pattern. This is even more problematic as the axis corresponding to the maximum response of the symmetry detector on pattern P does not necessarily coincide with the kernel axis, except in the simplest cases; neither do their centers of mass, as illustrated by Figure 19.

Figure 19. The axis of the maximally symmetric related pattern may not be the maximal symmetry axis, depending on the detector.
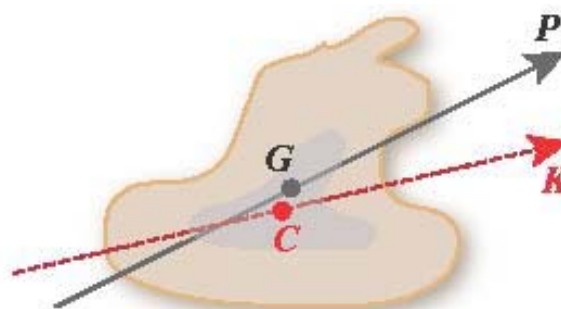
A bit of modeling and of elementary analysis allows straightening the difficulty up. We outline the model and the associated proof in considering the one dimensional case displayed in Figure 20.

Figure 20. Symmetrized version of a function in the L_2-norm sense.
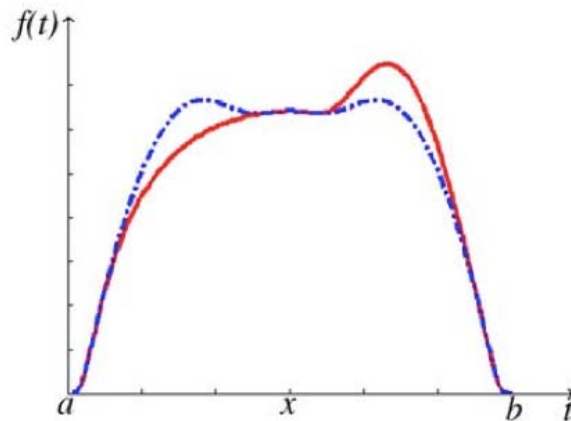
Indeed, within a L_2-frame the symmetric version of a function f with respect to an abscissa x is given by:

$$S_f^x(t) = \frac{f(x+t) + f(x-t)}{2}$$

(11)

And then the best L2-norm "axis" x^* following our idea of a kernel is the one that minimises the difference between S and f. Thus:

$$S_{x^*}(f) = \min_x \int_a^b \left[f(x+t) - f(x-t) \right]^2 dt$$

(12)

Assuming that $f(x) = 0$ for $x < 0$ and $x > b$, we have:

$$\frac{d}{dx} S_x(f) = = \int_{x-b}^{x+b} f(2x-u) \times f'(u) du = f \otimes f'(2x) = \frac{d}{dx}(f \otimes f)$$

(13)

So the best axes coincide with local maxima of $(f \otimes f)$ and that gives a computing process by the same token. It is based on convolution *i.e.*, the map product of symmetry and correlation.

Let us remark that, in the same representation, the centre of mass is given by:

$$X_G / \int_0^b (X-u) f(u) du = \frac{1}{2} \int_0^b u^2 f'(X-u) du = 0$$

(14)

which is likely different from:

$$x^* / \int_{sup(0,2x-b)}^{inf(2x,b)} f(u) \times f'(2x-u) du = 0$$

(15)

Actually comparing X_G and x^* expressions shows that picking the centre of mass for the locus of symmetry axes amounts to assimilate any pattern with their paraboloid approximation. Designing algorithms still requires to specify which parts of the image are to be correlated with their transformed version. That leads again to multi resolution, possibly combined with erosion. The technique is to tile the image again using blocks of a given shape, size and periodicity, indicative of the size of targeted symmetric patterns. Note that inside blocks or on isolated patterns a process of sub-axis alignment

from multiple bands makes sense too. See Figure 21 for an illustration, and the following figures for more detail.

Figure 21. Tiling a pattern with blocks to be correlated or with bands for their axes to be pieced together (medialaxis style).
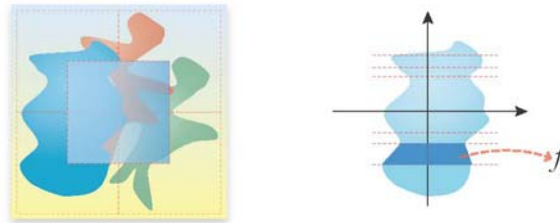
3.2. Symmetry Measures

After the hint of a systematic comparison of sets of pixels to exhibit their symmetry, measures are logically built from areas. This is made even easier in considering that symmetric objects are de facto binarized by their extraction. As a consequence, with *A* the object and *K* its kernel the first measure to be considered can be:

$$\lambda_1 = 1 - \frac{area(A-K)}{area(A)}$$

(16)

subject to the tiling (cropping) and binarizing processes bound to symmetry detection. But this measure does not translate the symmetry of an object with sufficient accuracy. Indeed the set difference:

$$A-K = D = \{n_c \text{ components } D_i\}$$

(17)

most often happens to be a collection of n_c components as already illustrated in Figure 19, where n_c ranges between *3* and *8* depending on the kernel precision, and by first kernel examples in the results shown in Figure 23. The distribution of D_i's is likely uneven. Consequently, three correcting factors λ_2 were tried:

$$1 - n_c \frac{\min_{1 \le i \le n_c}(Area(D_i))}{Area(A)}$$

(18)

$$1 - n_c \frac{median_{1 \le i \le n_c}(Area(D_i))}{Area(A)}$$

(19)

$$\frac{card\left\{D_i \mid Area(D_i) \le (Area(D)/n_c)\right\}}{n_c}$$

(20)

leading to the experimentally best tradeoff:

$$\lambda_2 = 1 - \left[\frac{\min_{1 \le i \le n_c}(Area(D_i))}{\max_{1 \le i \le n_c}(Area(D_i))} \times \frac{Area(D)}{Area(A)} \right]$$

(21)

For the measures, again, the question arises whether the internal kernel–*max_included*–or the external one–*min_including*–would matter. It is easy to prove that the residual D gets the same area and corresponds to the same direction in both cases. Then using external or internal kernels would

provide expressions of λ's respectively of the form 1-ε and $(1+\varepsilon)^{-1}$ so very close in case of symmetry when ε is small. Indeed, let $\tilde{P}_x$ be a given symmetric version of P when the invariant (axis or centre) is in x. The inner and outer kernels are respectively:

$$\tilde{P}_{x^*} \cap P \; / \; x^* \mapsto \max_x area\left(\tilde{P}_x \cap P\right) \text{ and } \tilde{P}_{y^*} \cup P \; / \; y^* \mapsto \min_x area\left(\tilde{P}_x \cup P\right) \tag{22}$$

The following statements hold:

$$\max_x area\left(\tilde{P}_x \cap P\right) \equiv \min_x area\left[P - \left(\tilde{P}_x \cap P\right)\right]$$

$$p \in \left[P - \left(\tilde{P}_x \cap P\right)\right] \Leftrightarrow \tilde{p} \in \left[\left(\tilde{P}_x \cup P\right) - P\right] \tag{23}$$

$$\text{because } p \in \left[P - \left(\tilde{P}_x \cap P\right)\right] \Leftrightarrow p \in P \wedge p \notin \left(P_x \cap P\right) \text{ and } p \notin \left(P_x \cap P\right) \Leftrightarrow \tilde{p} \notin \left(P_x \cap P\right) \tag{24}$$

$$\text{since } p \in \left(\tilde{P}_x \cap P\right) \Leftrightarrow \left(p \in P\right) \wedge \left(p \in \tilde{P}_x\right) \Leftrightarrow \left(\tilde{p} \in \tilde{P}_x\right) \wedge \left(\tilde{p} \in P\right) \Leftrightarrow \left[\tilde{p} \in \left(\tilde{P}_x \cap P\right)\right]$$

$$\text{and } p \in P \Leftrightarrow \left(\tilde{p} \in \tilde{P}_x\right) \tag{25}$$

$$\text{then } \left(p \in P\right) \wedge \left[p \notin \left(\tilde{P}_x \cap P\right)\right] \Leftrightarrow \left(\tilde{p} \in \tilde{P}_x\right) \wedge \left[\tilde{p} \notin \left(\tilde{P}_x \cap P\right)\right] \Leftrightarrow \tilde{p} \in \left[\left(\tilde{P}_x \cup P\right) - P\right] \tag{26}$$

$$\text{which implies that } x^* / \min_x area\left[P - \left(\tilde{P}_x \cap P\right)\right] = x^* / \min_x area\left[\left(\tilde{P}_x \cup P\right) - P\right] \tag{27}$$

As a conclusion of this subsection let us remark that all measure definitions introduced above range between *0* and *1*. According to [10] that is enough for them to be extendable to a measure of similarity between objects (see also [11] for the concept of similarity).

3.3. Results

To conclude this section, some results from an experiment on different kinds of pictures, both artificial and real and both binary and grey level, are reported. These results are useful to assess four parameters:

- sensitivity of the symmetry detection to the *centre position*
- validity of λ the measure (degree) of symmetry in comparing patterns to their kernel through the elongation η and then the kernel evolutions with *IOT*
- quality of the correlation kernel wrt the kernel
- validity of the symmetry axis from correlation wrt the best axis over shifts

Figure 23 displays some results obtained from the eight images shown in Figure 22: the respective kernels of the patterns (a)1, (b)1 and (b)2 in two directions–the maxima by inclusion and by correlation. The associated curves sketch the variations of the symmetry measure with the direction (quasi identical for (b)1). Images in Figure 22 are a sample of binary artificial images (a), real images (b) and textured images (c).

Figure 22. A sample of binary artificial images (a), real images (b) and textured images (c).
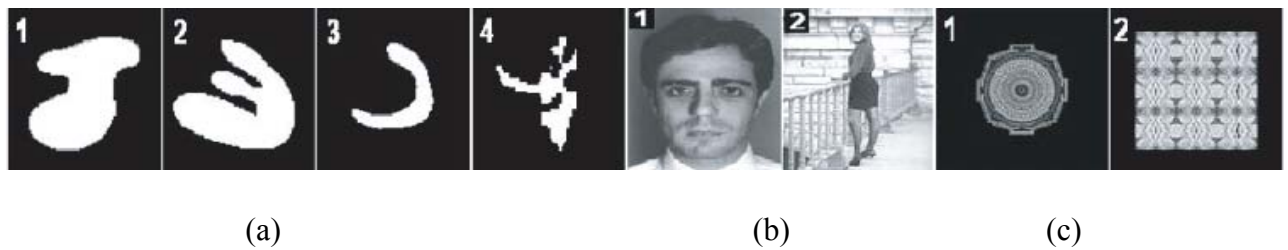

(a)   (b)   (c)

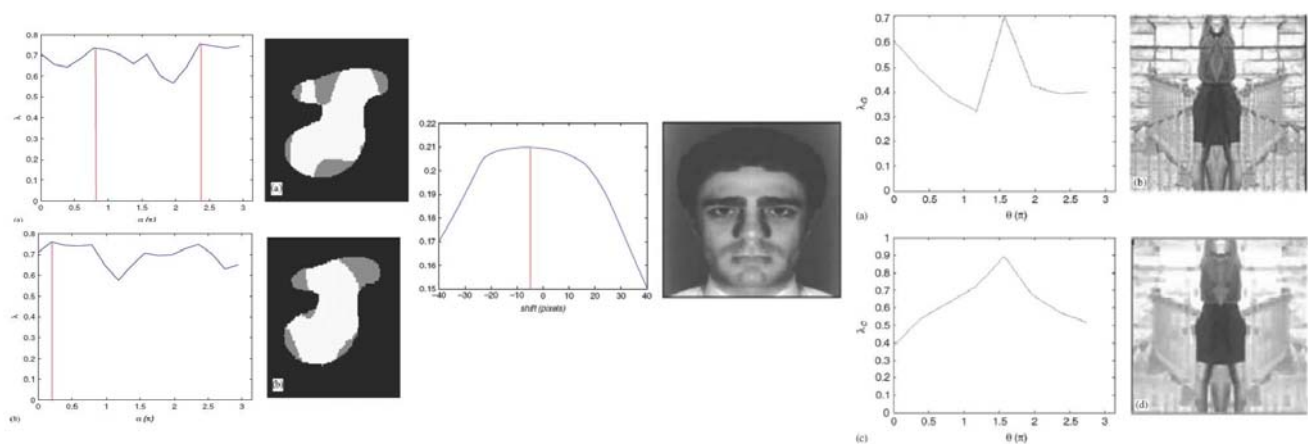**Figure 23.** Kernels and maximal symmetry of patterns (a) 1, (b) 1 and (b) 2 in Figure 22.

Table 2 compares the measures and directions of their maximum in the cases of axes through the centre of mass and axes by correlation and of the OST response at maximum stability.

Table 2. Results from the eight patterns shown in Figure 22: symmetry measures with the corresponding maximum-symmetry angles.

Image	λ_G	α_G	λ_C	α_C	OST	α_{OST}
1a	0.76	135.00°	0.76	11.25°	0.86	112.50°
2a	0.74	90.00°	0.79	33.75°	0.93	101.00°
3a	0.82	157.50°	0.76	22.50°	0.87	56.25°
4a	0.76	0.00°	0.80	0.00°	0.80	0.00°
1b	0.80	90.00°	0.80	90.00°	0.72	90.00°
2b	0.70	90.00°	0.89	90.00°	0.92	45.00°
1c	0.99	90.00°	0.99	135.00°	0.90	90.00°
2c	0.99	0.00°	0.99	90.00°	0.96	0.00°

Figure 24 displays on the same frames the values of the correlation ρ at is maximum and of the corresponding measure λ for pictures (c) 1 and 2. For sake of immediate comparison, in the cases of multiple symmetry (c), we give again the same curves λ_G and λ_C as in Figure 23.

Table 3 summarizes the values of the maximum correlation and corresponding angles to be compared to Table 2. Figure 25 displays the symmetry axes exhibited by correlation on a famous painting and on a group photo, the latter to be compared with Figure 17.

Figure 24. First row: symmetry indexes *vs*. the direction for image (c) 1. Second row: as above for image (c) 2.
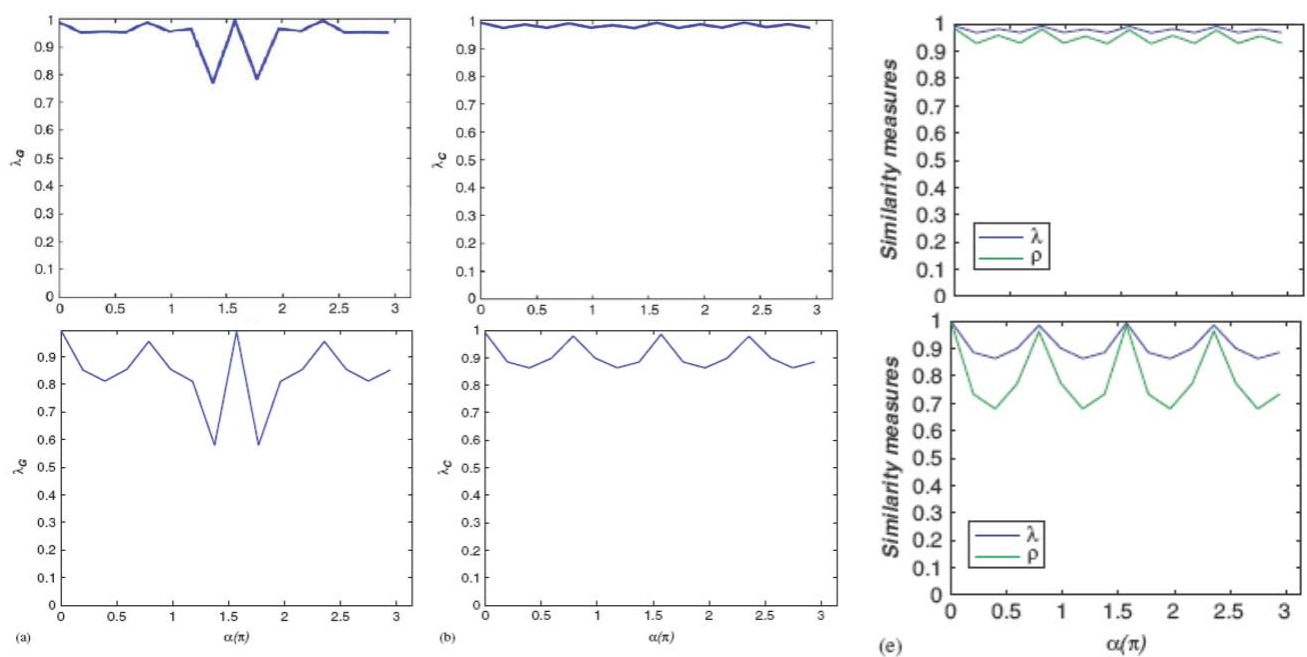
The analysis of *200* faces in groups (*276×410* pixels) or crowd (*768×1024*) with windows of size *20×20* and angle accuracy of $\pi/8$, and focus on λ greater than the mean provided the following results: the detection of *170* true faces and *42* false ones, a localization accuracy of ± 3 pixels and *30* wrong angles among the *170*.

More details and a comprehensive bibliography about the topic discussed in this section can be found in [4] and [12].

Table 3. Results from the eight patterns shown in Figure 22: the maximum correlation *vs*. the angle (to be compared with the mid column of Table 2 – *i.e.*, symmetry measure and angle for the correlation kernel).

Image	ρ	α
1a	0.67	101.25°
2a	0.67	112.50°
3a	0.58	112.50°
4a	0.55	157.50°
1b	0.80	90.00°
2b	0.94	90.00°
1c	0.99	90.00°
2c	0.98	0.00°

Figure 25. The symmetry axes exhibited by correlation on a famous painting and on a group photo.

4. Symmetry Detection and Face Expressions

The basic principles we have stated in previous sections are now employed in a practical task; if symmetry is a real, tangible feature in artificial vision, then should be able to use it in carrying or aiding in visual discrimination tasks usually spotlessly performed by humans - the best symmetry detection machine currently available.

Face expression recognition is one of the main problems of contemporary artificial vision. This is indeed an activity in which humans and animals alike excel at the finest degree: our social interaction mainly depends on our ability to spot even the slightest alteration in the facial expression of our counterpart (the sound and tone of voice accounting for the rest), as surviving used to depend on classifying our opponents based on their feelings, and so ultimately on their face expression.

The interest for face expression recognition has been steadily growing in the last decade [13-15], with a wide spectrum of applications in biometrics, ranging from security systems to criminology [16]. But by far the most interesting foreseen application for this task has to be found in human-machine interaction [17]: a computer that can understand our frustration and eases our workload, or tailors its interface complexity based on the happiness derived from the applications' usage. These types of system would require interaction in realtime, and should be based on some strong features, in order to overcome all the usual noise-over-signal problems in feature extraction (e.g., slightest differences in orientation, non uniform lighting, obstructions).

State of the art methods for face expression recognition, with stated recognition rates surpassing 90%, are generally based on feature extraction, using neural networks and Gabor wavelets [16], or Locality Preserving Projections in Linear Subspace [18]. Such methods, although ingenuous in execution, tell us nothing about the real phenomenon going on behind the scenes. Furthermore, having not reconciled the apparent dualism between local and global detection, results tend to be quite data-specific, and validation becomes a matter of choosing the right dataset; and barring humongous computing power, realtime processing is also not a possibility.

4.1. An Approach to Face Expression Recognition Based on Broken Symmetry Detection

We strongly support the idea that a simple global face signature, based on symmetry detection of local components, may be usefully incorporated as a robust component in a face expression recognition system. The idea goes as follows:

• Human face is by nature mostly symmetrical. More so in the so-called Neutral expression (see FACS [19]).

• Any expression different from Neutral is obtained by stretching a different subset of face muscles (the so-called action units [20]).

• Such stretching is rarely completely symmetrical; as such, the more marked those changes in expression are, the more breakage of symmetry is introduced in different parts of the face.

• Collecting and measuring those differences in symmetry from different portions of the face allows us to compile a typical signature for each expression. These signatures are then vectorized and feed to a classifier.

Figure 26. Face Expressions in FACS. A sample of face expressions from the JAFFE database [21]: (a) neutral, (b) sadness, (c) disgust, (d) happiness, (e) fear, (f) anger, (g) surprise. Expressions are obtained by self-attribution-the subject is asked to produce a specific expression, then a snapshot is taken.
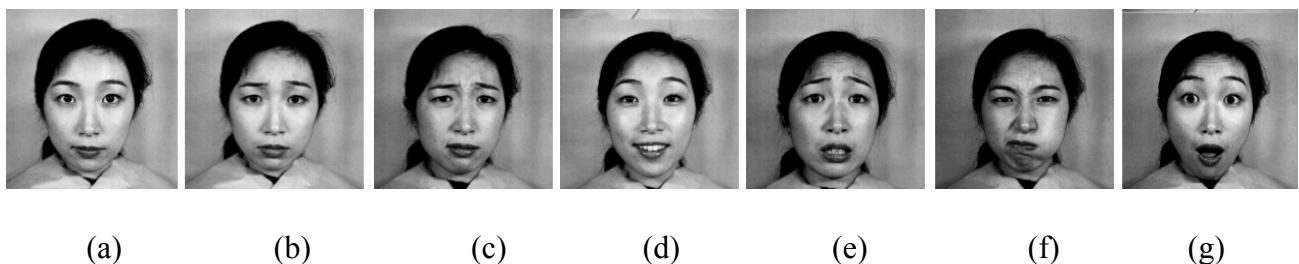

(a)   (b)   (c)   (d)   (e)   (f)   (g)

Figure 27. Are you happy to see me? Some instances of an expression are not easily reconciled with their self attribution. This problem encompasses geographical, gender and social entities. (a) Self–attributed as happiness, easily misclassified; (b) not self–attributed as happiness, often misclassified as happiness; (c) one is self–attributed as happiness, the other is not: can you spot which is which?

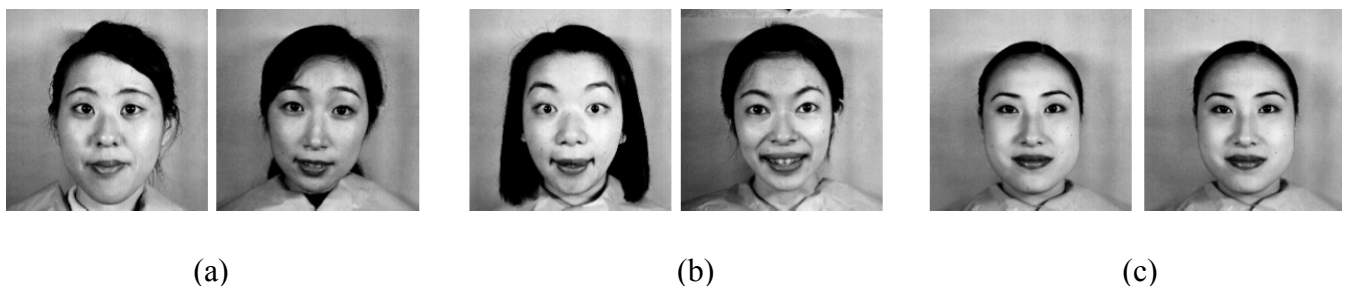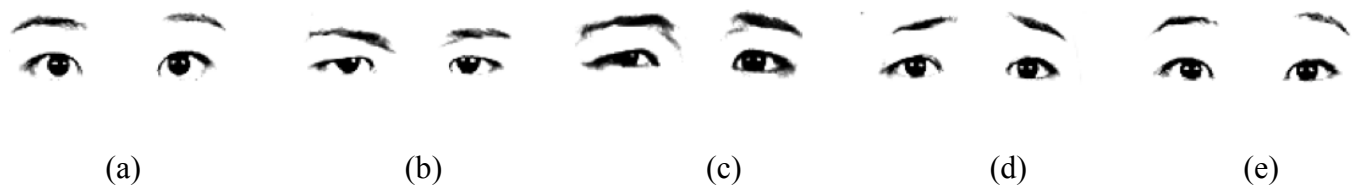(a)   (b)   (c)

Figure 28. Broken Symmetries in expressions. This is a detail from the eyes of the same subject in five different expressions: (a) neutral, (b) anger, (c) disgust, (d) fear, (e) happiness. Symmetry is clearly broken in different ways through these expressions.

(a) (b) (c) (d) (e)

The method used to is based on the same principle outlined in the previous Section: symmetry can be quantified by application of the measurement of symmetry kernel, using a suitable combination of both *max_included* and *min_including*. This time we will not use the multi-resolution pyramid, but will calculate our symmetry using an ordered covering with overlapping of the original image, an example of which is given in Figure 29 (a more rigorous definition of covering is given in the next subsection). The use of a covering allows us to consider local changes in the global symmetry, while the overlapping is an important feature to preserve contributions of structures that would split on the borders of two elements (as evident in Figure 29). The covering strategy cannot be else than problem driven, and in our application horizontal covering accounts for most anatomic components that characterize faces, in particular eyes and mouth, the spatial geometry of which is affected by emotions. A vertical covering could be added to be more sensitive to nose changes or to alignments like nose-mouth; however such measurements are not as precise as horizontal ones.

Figure 29. Covering of a face. The usefulness of overlapping is self-evident: eyes are captured in slice 2 and mouth in slice 6, while eyebrows are captured in slice 1 and cheeks in slice 5.
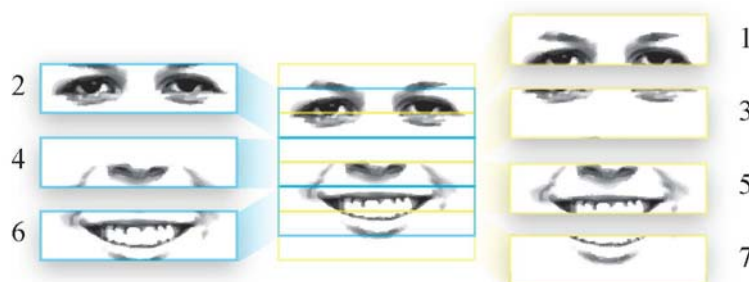

4.2. Method

4.2.1. Dataset

In order to test our procedure we have used the JAFFE database. Faces in the JAFFE database are grey level bitmap images, each categorized in one of the seven canonical FACS expressions. The covering procedure has been carried out on *256×256* greyscale images. Five of such expressions (neutral, anger, disgust, fear, and happy) have been used, regardless of the human rating attached to the database. In order to filter out effects due to bad lightning, all the images have been preprocessed using a completely automated cleaning algorithm which uses only information gathered from the image itself, based on equalization of intensity levels and a cutting threshold derived from the mean

luminance value. An example of the results obtained from the clearing algorithm is given in Figure 30. The images have been divided in two sets (data and training, *N:50*), carefully avoiding subject overlapping.

Figure 30. Cleaning algorithm. The effect of the preprocessor used on JAFFE: (a) original image; (b) preprocessed image.


(a)   (b)
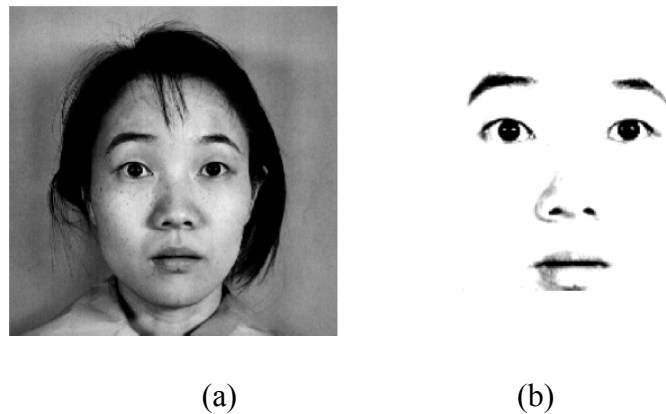
4.2.2. Procedure

We first define a covering *CF* of an image *F* as a collection, *{F₁, F₂, ..., Fₕ, ..., F_L}* of sub-images of F ($F_h \subseteq F$) such that $\cup F_h \equiv F$. The ordered covering, *OCF*, of *F* is a covering, *CF*, elements of which have been ordered according to a given exploration rule of *F*: $OCF = (F_{i1}, F_{i2}, ..., F_{ih}, ..., F_{iL})$.

Then we define our measure of symmetry based on *max_included* and *min_including*. In a pattern with perfect symmetry, *max_included* and *min_including* are clearly the same. As such, we can consider the difference between them as a good approximation of the symmetry of a pattern.

Figure 31. Covering samples: (a) vertical; (b) horizontal; (c) quadrant.

(a)   (b)   (c)
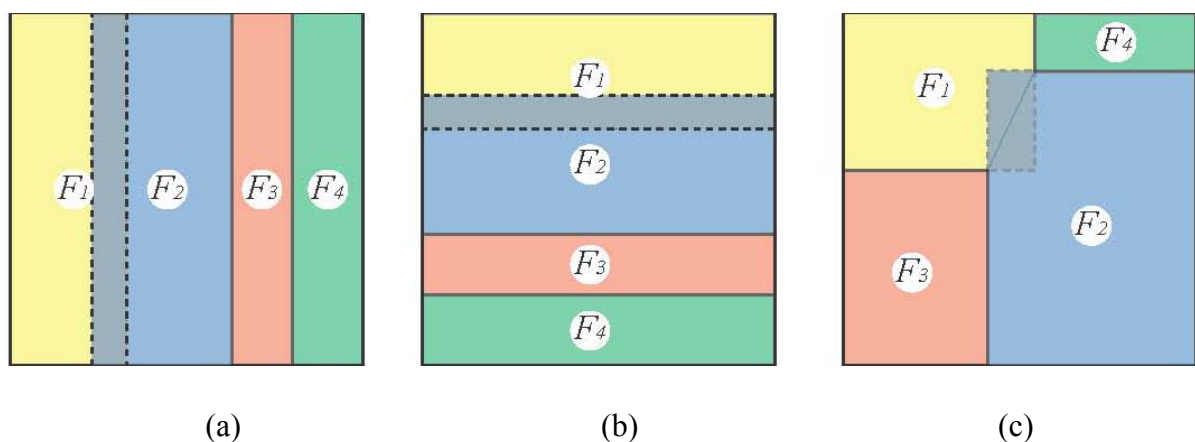
Since we are working with grayscale images, we should better use some measurement in a compact set, the simplest choice of which is area. As such, our measure of symmetry is given by:

$$sym(A) = \arg\min_{0 < \theta < \pi} \frac{Area(\text{min\_including}(A, \theta)) - Area(\text{max\_included}(A, \theta))}{Area(\text{min\_including}(A, \theta))}$$

where Area(S) indicates the measure of the set S (in a greyscale image, the sum of the value of the pixels).

It is easily shown [9] that:

1) parameters (e.g., axis position and angle) for which *max_included* and *min_including* are obtained are the same, and

2) the measure given above is invariant for translation, rotation, and scaling.

However, in the case of digital images, scaling and rotation operators cause perturbation due to the discreteness of the space possibly resulting into a loss of precision.

Figure 32. *Max_included* and *Min_including*: (a) original object, (b) min_including; (c) max_included; (d) *min_including–max_included.*

(a)   (b)   (c)   (d)
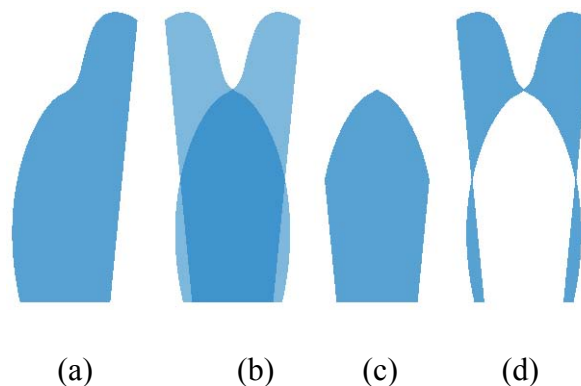
The symmetry signature is obtained by measuring *sym(A)* for each element of *OCF*.

Choice of parameters for *OCF* is done heuristically, cycling among all sensible values of stripes number (*2,4,8,16,32*) and overlapping (*10%, 20%, 30%, 40%, 50%*).

Classification is carried out by a nearest-neighbor algorithm that assigns each individual expression signature in the test set to the nearest symmetry signature in the training set. One main issue is the definition of a suitable distance between two signatures; we tested the Euclidean, the Manhattan, the correlation, the *min_difference* and the *max_difference*.

For each strip value and overlap percentage the confusion matrix, *CM*, for the five classes (neutral, anger, disgust, fear, happy) was derived and the mean recognition rate was computed as follows:

$$\text{meanRR} = \sum_{i=1}^{5} CM_{i,i}$$

4.3. Experimental Results and Discussion

From these experiments, it comes out that a reasonable *meanRR* is obtained for a number of stripes equal to *16* and *32* (*meanRR = 78%*), but in the case of *32* stripes and *25%* overlap we had the best discrimination between neutral and other expressions (*82%*) that compares well to more sophisticated and complex methods. More generally, a covering with 16 and 32 horizontal stripes provides the best mean recognition rate. Intuitively these numbers of stripes are perfectly tuned with natural face

segments (eyebrows, eyes, nose, lips) that are currently used in expressions analysis. Figure 33 shows the trend of the recognition rate for different numbers of stripes and increasing overlaps (up to *50%*), while in Table 4, we show the confusion matrix for the case "*32* stripes and *20%* overlap" that corresponds to the best recognition rate for the classes neutral and disgust; the mean recognition rate is *75%*.

Figure 33. MeanRR for different stripe widths and overlaps. The graph shows *meanRR* value for stripe widths of *8,16,32,64,128* versus overlap of *10%, 20%, 30%, 40%, 50%*.
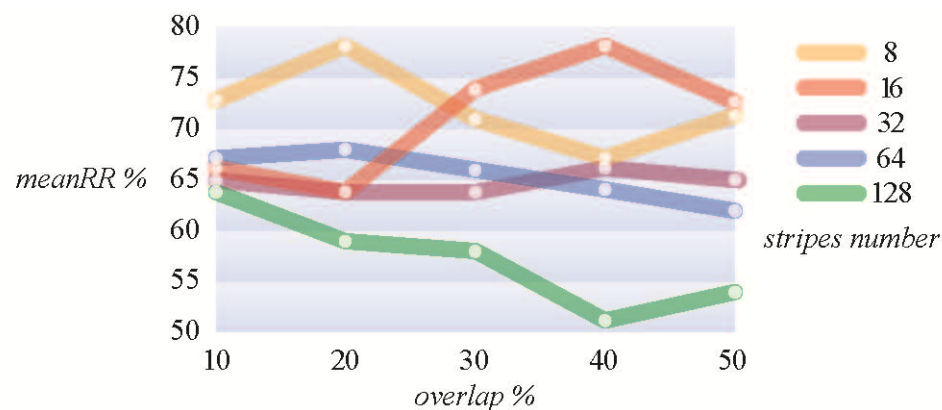

Table 4. Confusion Matrix for five expressions.

	Neutral	*Anger*	*Disgust*	*Fear*	*Happiness*
Neutral	0,88	0,00	0,02	0,05	0,05
Anger	0,07	0,61	0,12	0,13	0,07
Disgust	0,02	0,00	0,93	0,00	0,05
Fear	0,07	0,07	0,05	0,76	0,05
Happiness	0,15	0,07	0,12	0,05	0,61

These results compare favorably, if we consider the cardinality of the database used, to other more complicated and time intensive systems: the current state-of-the-art methods often offer recognition beyond the *90%* mean, but not on every expression, and at the expenses of the cardinality of the training set and the resources used, and as such of the possibility to implement the system in real time. In [18], which at the moment of this writing is the best performer currently available, a mean recognition rate of *93%* is reached using a much larger database with repeated images from single subjects, a practice that easily leads to skewed results; when the same method is applied to the very same JAFFE database used in this paper, results are much weaker and the authors chose not to publish them. The method hereby presented strikes a good balance between the stark simplicity of the idea, the speed of implementation and execution, and the results obtained even when database cardinality is low, a situation often encountered in practice.

5. Conclusions

This paper summarizes some of our research on Symmetry Detection by Machines. We first postulated that Symmetry is an image feature similar to contours, regions or motion: it can serve to trigger attention but also to recognize patterns. Main parameters of symmetry sensing have then been

exhibited, briefly modeled and discussed, such as the position of the center or axis, the transform relating to the symmetry, noise over signal factors. The search for optimality, always underlying feature extraction, favors the use of iterative procedures. Two algorithmic lines have been thus proposed and investigated along the way, based on the so-called symmetry kernel and its evolution with pattern warping, and by correlation of blocks with varying sizes and positions. Eventually, on top of a review of various experiments on both synthetic and natural images of many kinds made in order to check the algorithmic validity and tuning associated model parameters, a complete application to face expression clustering has been detailed, and its results are given in conclusion. The whole theoretical and experimental process, up to the full application example, definitely confirms that symmetry is a dynamic image processing feature suitable to machine vision.

References

1. Kanizsa, G. The role of regularity in perceptual organization. In *Studies in perception: Festschrift for Fabio Metelli*; Martello-Giunti Editore: Firenze, Italia, 1975.
2. Palmer, S.E. The Role of Symmetry in Shape Perception. *Acta Psychol.* **1985**, *59*, 67–90.
3. Zavidovique, B.; DiGesù, V. The S-kernel: A measure of symmetry of objects. *Pattern Recogn. Lett.* **2007**, *40*, 839–852.
4. DiGesú, V.; Zavidovique, B. Iterative symmetry detection: Shrinking *vs.* decimating patterns. *Integr. Comput. Aided Eng.* **2005**, *12*, 319–332.
5. Palmer, S. *Vision Science: Photons to Phenomenology*; Bradford Books/MIT Press: Cambridge, MA, USA, 1999; pp. 4–15.
6. DiGesù, V.; Valenti, C. Symmetry operators in computer vision. *Vistas Astron.* **1996**, *40*, 461–468.
7. DiGesù, V.; Zavidovique, B. A note on the iterative object symmetry transform. *Pattern Recogn. Lett.* **2004**, *25*, 1533–1545.
8. DiGesú, V.; Lo Bosco, G.; Zavidovique, B. Classification Based on Iterative Object Symmetry Transform. In *ICIAP '03: Proceedings of the 12th International Conference on Image Analysis and Processing*; IEEE Computer Society: Washington, DC, USA, 2003.
9. Zavidovique, B.; Di Gesú, V. Pyramid symmetry transforms: From local to global symmetry. *Image Vision Comput.* **2007**, *25*, 220–229.
10. Oliva, D.; Samengo, I.; Leutgeb, S.; Mizumori, S.A Subjective Distance Between Stimuli: Quantifying the Metric Structure of Representations. *Neural Comput.* **2005**, *17*, 969–990.
11. Tversky, A. Features of similarity. *Psychol. Rev.* **1977**, *84*, 327–352.
12. Di Gesu`, V.; Zavidovique, B. S-Kernel: A New Symmetry Measure. In *Pattern Recognition and Machine Intelligence*; Pal, S., Ed.; Springer Verlag: Berlin-Heidelberg, Germany, 2005.
13. Moghaddam, B.; Pentland, A.P. Face recognition using view-based and modular eigenspaces. In *Proc. SPIE*; SPIE: San Diego, CA, USA, 1994; Volume 2277.
14. Zhao, W.; Chellappa, R.; Phillips, P.J.; Rosenfeld, A. Face recognition: A literature survey. *ACM Comput. Surv.* **2003**, *35*, 399–458.

15. Shakhnarovich, G.; Moghaddam, B. Face Recognition in Subspaces. In *Handbook of Face Recognition*; Stan, Z.L., Anil, K.J., Eds.; Springer-Verlag: Secaucus, NJ, USA, 2004; Volume I; Chapter 7; pp. 141–168.

16. Jain, A.K.; Ross, A.; Prabhakar, S. An introduction to biometric recognition. *IEEE Trans. Circ. Syst. Video Technol.* **2004**, *14*, 4–20.

17. Pelachaud, C.; Poggi, I. Multimodal communication between synthetic agents. In *Proceedings of the Working Conference on Advanced Visual Interfaces*; ACM: L'Aquila, Italy, 1998.

18. Shan, C.; Gong, S.; McOwan, P.W. A Comprehensive Empirical Study on Linear Subspace Methods for Facial Expression Analysis. In *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*; IEEE Computer Society: Washington, DC, USA, 2006.

19. Ekman, P.; Friesen, W. *Facial Action Coding System: A Technique For The Measurement Of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978.

20. Essa, I.A.; Pentland, A.P. Coding, Analysis, Interpretation, and Recognition of Facial Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 757–763.

21. Lyons, M.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding Facial Expressions with Gabor Wavelets. In *IEEE International Conference on Automatic Face and Gesture Recognition*; IEEE Computer Society: Los Alamitos, CA, USA, 1998.