Improving Zernike Moments Comparison for Optimal Similarity and Rotation Angle Retrieval

Jérôme Revaud, Guillaume Lavoué and Atilla Baskurt

Abstract—Zernike moments constitute a powerful shape descriptor in terms of robustness and description capability. However the classical way of comparing two Zernike descriptors only takes into account the magnitude of the moments and loses the phase information. The novelty of our approach is to take advantage of the phase information in the comparison process while still preserving the invariance to rotation. This new Zernike comparator provides a more accurate similarity measure together with the optimal rotation angle between the patterns, while keeping the same complexity as the classical approach. This angle information is particularly of interest for many applications, including 3D scene understanding through images. Experiments demonstrate that our comparator outperforms the classical one in terms of similarity measure. In particular the robustness of the retrieval against noise and geometric deformation is greatly improved. Moreover, the rotation angle estimation is also more accurate than state of the art algorithms.

Index Terms—Zernike moments, scene analysis, 3D object recognition, shape

1 Introduction

Zernike moments are widely used to capture global features of an image in pattern recognition and image analysis. Firstly introduced in computer vision by Teague [1], this shape descriptor has proved its superiority over other moment functions [2], [3] regarding to its description capability and robustness to noise or deformations. Hence rotation invariant pattern recognition using Zernike moments has been extensively studied [4], [5]. Even very recently, a lot of authors have been working on these moments, particularly to improve their computation time [6], [7], [8], [9] or their accuracy [10].

Practically one Zernike moment is a complex number that contains two different values: *magnitude* and *phase*, however, the usual way (i.e. used in all existing algorithms) of comparing two Zernike descriptors only considers the moments magnitudes (as it brings invariance to rotation). In the context of 2D and 2D-3D indexing and recognition, this loss of information is not harmless when comparing two different patterns, and can induce erroneous results and impreciseness, as it will be further illustrated.

Using the phase information of Zernike moments (together with the magnitude) in the comparison process seems a natural way to improve the similarity measure in terms of robustness against geometric deformation or noise particularly. However in that case the resulting comparator is not invariant anymore to rotation, unless the in-plane rotation angle between the two patterns is known. Fortunately in this paper we show that the moment phases can also be used to retrieve this rotation angle in an optimal way. Finding both information (i.e. a robust rotation-invariant similarity measure together with the optimal angle of rotation) can be of great interest for many applications including image registration [11], motion estimation in video and particularly scene understanding: indeed recognizing the objects composing the image and then extracting their in-plane orientation angles may help to compute their precise 3D pose and thus to understand accurately the corresponding 3D environment. A lot of work has been done for angle/similarity recognition using keypoint-based local descriptors like SIFT [12], however, this kind of tools works only on textured objects and fails to describe smooth shapes or drawings (i.e. sketch) for instance. In such hard description/recognition cases, global shape descriptors like Zernike moments are particularly robust, that is the reason why they have recently been used for rotation invariant 2D/3D object recognition through sketches [13], [14]. It appears quite relevant to compute the in-

Emails: {jerome.revaud, glavoue, abaskurt} @insa-lyon.fr
The authors are with LIRIS, UMR 5205 CNRS, INSA-Lyon, F-69621 Villeurbanne, France.

plane rotation together with the similarity distance in such 2D/3D indexing scenarios, particularly for some emerging applications like sketch-based modeling [15].

Apart from us, one approach have focused on the sub-problem of the rotation angle estimation using Zernike moments. The method was brought by Kim and Kim [16] and it proved to be very robust with respect to noise even for circular symmetric patterns. Nevertheless, the probabilistic model used to recover the rotation angle has no concrete interpretation and does not correspond to any geometrical reality, so it does not return any similarity distance. Besides, this method is based on the hypothesis that the two patterns are the *same* (that is, except some noise and the rotation), which does not always hold in practice.

In this context, we have developed a new general and rigorously founded approach for comparing two Zernike descriptors that takes use of both magnitude and phase information. Our approach keeps the same complexity as the standard technique (Euclidean distance between magnitude values) but provides a more accurate rotation invariant similarity measure while retrieving the optimal in-plane rotation angle. Thanks to the adaptability of Zernike description, our approach is suited to compare any kinds of images/patterns: Binary, gray level or sketch images (i.e. drawings). We also compared our results with two state-of-the-art approaches for sketch and object recognition: the geometric hashing [17] and the deformation tolerant generalized Hough Transform from Anelli et al. [18].

The following section concisely presents Zernike moments. In section 3, we lean upon drawbacks of the conventional approach to build our method. The computational efficiency is also a constraint because the resulting algorithm will be used within a matching process; we thus detail a fast implementation in section 4. Finally, we present experimental results in section 5 and an application to a real 2D/3D indexing scenario in section 6.

2 Zernike moments

Complex Zernike functions constitute a set of orthogonal basis functions mapped over the unit circle. Zernike moments of a pattern are constructed by projecting it onto those functions. They share three main properties:

- The orthogonality: this property ensures that the contribution of each moment is unique and independent.
- The rotation invariance: the magnitude of Zernike moments is independent of any planar rotation of the pattern around its center of mass.
- The information compaction: low frequencies of a pattern are mostly coded into the low order moments. As a result, relatively small descriptors are robust to noise or deformations.

Mathematically, Zernike basis functions are defined with an order p and a repetition q over $D = \{(p,q)|0 \leq p \leq \infty, |q| \leq p, |p-q| = \text{even}\}$.

$$Z_{pq} = \frac{p+1}{\pi} \int\int_{x^2+y^2 \leq 1} V_{pq}^*(x,y) f(x,y) \partial x \partial y \quad (1)$$

where V_{pq}^* denotes the complex conjugate of V_{pq}, itself defined as:

$$V_{pq}(\rho,\theta) = R_{pq}(\rho).e^{iq\theta} \quad (2)$$

$$\text{and} \quad R_{pq}(\rho) = \sum_{\substack{k=|q| \\ |p-k| \ even}}^{p} \frac{(-1)^{\frac{p-k}{2}} \frac{p+k}{2}!}{\frac{p-k}{2}!\frac{k-q}{2}!\frac{k+q}{2}!} \rho^k$$

From eq. (1) and (2), Zernike moments of a pattern rotated by an angle α around its origin are given in polar coordinates as :

$$Z_{pq}^{\alpha} = Z_{pq} e^{iq\alpha} \quad (3)$$

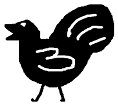
Eq (3) proves the invariance of the magnitude of Zernike moments to rotation since $|Z_{pq}e^{iq\alpha}| = |Z_{pq}|$. Thanks to the property of orthogonality, the reconstruction of the pattern can be simply expressed as the sum of every Zernike basis functions weighted by the corresponding moments:

$$\tilde{f}(x,y) = \sum\sum_{(p,q)\in D} Z_{pq} V_{pq}(x,y) \quad (4)$$

3 Similarity measure and rotation angle retrieval

3.1 The classical approach

The usual way of comparing two patterns in the Zernike space only considers the magnitudes of the moments [4]. In the reminder of this paper, we will denote the usual comparator of Zernike descriptors as the *classical* one. Formally, this comparator is

(a) the pattern to retrieve: 

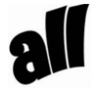
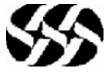
rank	1st	2nd	3rd	4th
classical comparison (without phase)				
distance to (a)	7130.2	7345.5	7515.0	8200.9
proposed comparison (using phase)				
distance to (a)	2110.3	2339.0	2406.4	2683.1

Fig. 1. The best four retrievals for the drawing of a rooster (a) from a database of 200 binary logos. Results are ordered by similarity measure (a distance of zero means a complete similarity) for both comparators.

nothing else than an Euclidean distance between the magnitudes:

$$d^2 = \sum\sum_{(p,q)\in D} (|Z_{pq}| - |Z'_{pq}|)^2 \qquad (5)$$

However, the advantage of losing the phase information - this allows the invariance to rotation - also brings some drawbacks: The first consequence is that the classical comparator is unable to retrieve the rotation angle between two similar patterns as this information is encoded onto the moments phases. A corollary is that two *symmetrical* patterns will be classified as *identical* since their moments magnitudes are the same. Consequently, for an application which has to differentiate symmetric patterns (e.g. 2D-3D recognition that aims to retrieve viewpoint of potentially symmetric 3D objects like cars or people), the classical comparator is inoperative.

More generally, one can assume that this missing information has a negative influence on the retrieval efficiency, especially in hard cases like noisy, deformed or occluded patterns. Our approach is based on the assumption that using the phase information may result in a more robust description. Experiments have confirmed that this hypothesis is exact whatever the number of moments and for a wide type of distortions (see section 5). Figure 1 presents a short example of image retrieval from sketch which illustrates the superiority of the proposed approach upon the classical one.

3.2 The proposed Zernike comparator

In this subsection, we present a new way of comparing Zernike moments that takes both magnitude and phase information into account. Our new comparator provides a similarity score more robust than the classical one and retrieves for free an accurate angle of rotation between the two images. The angle is considered to be optimal since the Euclidean distance between the first image and the rotated second one is minimized. Let I and J be two different images, and $(J * \Re_\phi)$ be the J image rotated by ϕ. The Euclidean distance between I and $(J * \Re_\phi)$ can be expressed as a function of the rotation angle ϕ as follows:

$$d^2_{I,J}(\phi) = \sum\sum_{x^2+y^2\leq 1} |I(x,y) - (J * \Re_\phi)(x,y)|^2 \qquad (6)$$

Thus our objective is to minimize this expression so as to determine the corresponding angle ϕ. Eq. (3) has shown that if the set of moments $\left\{Z^J_{pq}|(p,q)\in D\right\}$ represents the J image, then $\left\{Z^J_{pq}e^{iq\phi}|(p,q)\in D\right\}$ represents $J * \Re_\phi$. By replacing I and $(J*\Re_\phi)$ in equation (6) by their exact Zernike reconstruction (4), we obtain eq. (7) (see next page) where Z^I_{pq} and Z^J_{pq} represent Zernike moments of images I and J, respectively, and $\langle V_{pq}, V^*_{uv}\rangle$ denotes the scalar product of two Zernike basis functions over the unit disc. Thanks to the orthogonality of the basis, this product is null except for the case where $(p,q) = (u,v)$. In that case, it can be simplified into:

$$\left\langle V_{pq}, V^*_{pq}\right\rangle = \frac{\pi}{p+1}$$

At first glance eq. (7) is not trivial to minimize, but it can be rewritten into eq. (8) (see next page, with $|Z_{pq}|$ and $[Z_{pq}]$ respectively the modulus and the argument of Z_{pq}). Formula (8) points out that the only parameter whose the distance depends is ϕ, which in addition is only present into the cosine functions. As a consequence, the search for optimal distance and angle will result in minimizing a sum of cosines.

Even if in real applications only a subset of Zernike moments is used (as it brings robustness to the description), the proposed method remains valid: simply, it can be seen as a fast way of retrieving the Euclidean distance between the two *blurred* patterns (that is, their reconstruction from the subset of moments) using their projection in Zernike space. The next section focuses on resolving the minimization so as to insure a low complexity and a fast computing time.

$$
\begin{aligned}
d^2_{I,J}(\phi) &= \sum\sum_{x^2+y^2\leq 1} \left| \sum\sum_{(p,q)\in D} Z^I_{pq}.V_{pq}(x,y) - \sum\sum_{(p,q)\in D} Z^J_{pq}e^{iq\phi}.V_{pq}(x,y) \right|^2 \\
&= \sum\sum_{x^2+y^2\leq 1} \left| \sum\sum_{(p,q)\in D} (Z^I_{pq} - Z^J_{pq}e^{iq\phi})V_{pq}(x,y) \right|^2 \\
&= \sum\sum_{x^2+y^2\leq 1} \left\{ \left[\sum\sum_{(p,q)\in D} (Z^I_{pq} - Z^J_{pq}e^{iq\phi})V_{pq}(x,y) \right] \left[\sum\sum_{(u,v)\in D} (Z^I_{uv} - Z^J_{uv}e^{iv\phi})V_{uv}(x,y) \right]^* \right\} \\
&= \sum\sum_{x^2+y^2\leq 1} \left\{ \sum\sum_{(p,q),(u,v)\in D^2} (Z^I_{pq} - Z^J_{pq}e^{iq\phi})V_{pq}(x,y)\,(Z^{I*}_{uv} - Z^{J*}_{uv}e^{-iv\phi})V^*_{uv}(x,y) \right\} \\
&= \sum\sum_{(p,q),(u,v)\in D^2} \left\{ (Z^I_{pq} - Z^J_{pq}e^{iq\phi})(Z^{I*}_{uv} - Z^{J*}_{uv}e^{-iv\phi}) \left[\sum\sum_{x^2+y^2\leq 1} V_{pq}(x,y)V^*_{uv}(x,y) \right] \right\} \\
&= \sum\sum_{(p,q),(u,v)\in D^2} (Z^I_{pq} - Z^J_{pq}e^{iq\phi})(Z^{I*}_{uv} - Z^{J*}_{uv}e^{-iv\phi}). \begin{cases} \langle V_{pq}, V^*_{uv}\rangle & if\ (p,q) = (u,v), \\ 0 & else \end{cases} \\
&= \sum\sum_{(p,q)\in D} \left| Z^I_{pq} - Z^J_{pq}e^{iq\phi} \right|^2 \langle V_{pq}, V^*_{pq}\rangle
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
d^2_{I,J}(\phi) &= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left| Z^I_{pq} - Z^J_{pq}e^{iq\phi} \right|^2 \\
&= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left| \left|Z^I_{pq}\right| e^{i[Z^I_{pq}]} - \left|Z^J_{pq}\right| e^{i(q\phi+[Z^J_{pq}])} \right|^2 \\
&= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left| e^{i[Z^I_{pq}]} \left(\left|Z^I_{pq}\right| - \left|Z^J_{pq}\right| e^{i(q\phi+[Z^J_{pq}]-[Z^I_{pq}])} \right) \right|^2 \\
&= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left| e^{i[Z^I_{pq}]} \right|^2 . \left| \left|Z^I_{pq}\right| - \left|Z^J_{pq}\right| e^{i(q\phi+[Z^J_{pq}]-[Z^I_{pq}])} \right|^2 \\
&= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left| \left(\left|Z^I_{pq}\right| - \left|Z^J_{pq}\right| cos(q\phi + [Z^J_{pq}] - [Z^I_{pq}]) \right) - i \left(\left|Z^J_{pq}\right| sin(q\phi + [Z^J_{pq}] - [Z^I_{pq}]) \right) \right|^2 \\
&= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left[\left(\left|Z^I_{pq}\right| - \left|Z^J_{pq}\right| cos(q\phi + [Z^J_{pq}] - [Z^I_{pq}]) \right)^2 + \left(\left|Z^J_{pq}\right| sin(q\phi + [Z^J_{pq}] - [Z^I_{pq}]) \right)^2 \right] \\
&= \sum\sum_{(p,q)\in D} \frac{\pi}{p+1} \left[\left|Z^I_{pq}\right|^2 + \left|Z^J_{pq}\right|^2 - 2\left|Z^I_{pq}Z^J_{pq}\right| cos(q\phi + [Z^J_{pq}] - [Z^I_{pq}]) \right]
\end{aligned} \tag{8}
$$

4 Efficient minimum search

In this section, we describe how to efficiently extract the global minimum of eq. (8) by restricting the search using Nyquist-Shannon sampling theorem. Assuming that the Zernike moments of each image are known until the N^{th} order, the sum initially comprises $O(N^2)$ cosine terms. Nonetheless, the sum can be simplified by removing the constant terms and by aggregating the cosine terms that own the same frequency:

$$
\begin{aligned}
&A_1\, cos(q\phi + B_1) + A_2\, cos(q\phi + B_2) \\
&= |C|\, cos(q\phi + [C])
\end{aligned} \tag{9}
$$

where C is a complex that worth $A_1 e^{iB_1} + A_2 e^{iB_2}$. Then, eq. (8) can be equivalently expressed as a sum of N cosines :

$$f_N(\phi) = \sum_{q=1}^{N} A_q \, cos(q\phi + B_q) \tag{10}$$

with $A_q \in \mathbb{R}^+$ and $B_q \in [-\pi, \pi[$.

4.1 Restricting the search of a global minimum

One can notice that $f_N(\phi)$ is a 2π-periodic function. Usually, the general technique for finding the minimum of a periodic function is a gradient descent. Indeed, $f_N(\phi)$'s first and second derivatives are easy to compute. However, such functions generally have many local minima whereas our approach requires to find the global one. One expensive solution is then to find every local minima and maxima with the gradient descent method by following the function from $\phi = 0$ to 2π.

However, $f_N(\phi)$ owns a discrete Fourier spectrum bounded by a maximal frequency $f_N^{MAX} = N/2\pi$. Hence, $f_N(\phi)$ has at most N local maxima and N local minima in $[0, 2\pi[$. Moreover, the Nyquist–Shannon sampling theorem teaches us that the function can not change substantially between two consecutive sampling points taken at the Nyquist frequency $F = 1/T = N/\pi$. The minimal distance between two consecutive minima is thus bounded by π/N. The initial starting points for finding every possible minima with a gradient descent can thus be equally scattered in $2N$ points. Moreover, by cutting f_N into $4N$ intervals, we ensure that only one minimum *or* one maximum is present in each interval (see fig. 2).

4.2 Optimized minima retrieval

Our approach for optimizing the gradient descent takes advantage of the previously formulated properties. f_N is sampled by $4N$ points equally spread between $[0, 2\pi[$: $\{x_n = n\pi/2N \mid 0 \leq n < 4N\}$. We compute f_N's differential, denoted as f'_N, for each of those points. Section 4.1 ensures that if and only if a minimum is present between two consecutive points $[x_n, x_{n+1}]$, then $f'_N(x_n)$ is negative and $f'_N(x_{n+1})$ is positive (see fig. 2). Moreover, those differential values enable to approximate the abscissa of the local minimum. Indeed, by approximating f_N between $[x_n, x_{n+1}]$ as a second degree polynomial, then the minimum abscissa can be evaluated at:

$$\begin{aligned} x_{minimum} &= \frac{x_{n+1} f'_N(x_n) - x_n f'_N(x_{n+1})}{f'_N(x_n) - f'_N(x_{n+1})} \\ &= x_n + \frac{\pi}{2N} \frac{f'_N(x_n)}{f'_N(x_n) - f'_N(x_{n+1})} \end{aligned} \tag{11}$$

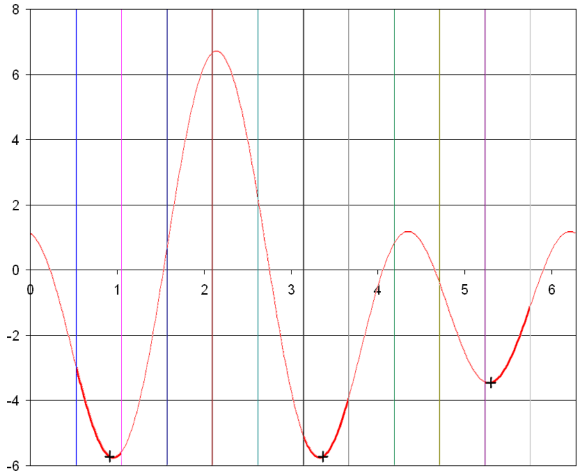


Fig. 2. A random function f_3 cut into $4N = 12$ intervals. Each interval contains at most one local minimum. When a minimum exists in the interval, the left derivative is negative and the right derivative is positive. For each minimum, a black cross figures the minimum position approximated with second order polynomial.

For our application, the gradient descent algorithm does not need to be iterated to reach a high precision since this simple approximation (represented as crosses in fig. 2) is precise enough for our purpose (cf. experimental results in 5.1). Finally, the computational complexity of our approach for the distance minimization using the approximation is $O(N^2)$. Our approach has then the same computational complexity than the classical Zernike comparator (eq. (5)).

5 Experimental Results

This section describes several experiments that illustrates:

- The efficiency of our minimum search (section 5.1).
- A comparison of our approach with the classical one in terms of similarity accuracy, and retrieval performance (section 5.2.2).
- A comparison of our approach with two state-of-the-art methods: geometric hashing and generalized Hough Transform (section 5.2.3).

		$\phi \approx \tilde{\phi}$	$\phi \neq \tilde{\phi}$
$N = 6$	Occurrences	9,986	14
	RMS error on $\tilde{\phi}$	0.81°	107.6°
	RMS error on $\tilde{f_N}$	0.197%	0.537%
$N = 12$	Occurrences	9,991	9
	RMS error on $\tilde{\phi}$	0.36°	119.4°
	RMS error on $\tilde{f_N}$	0.122%	0.472%
$N = 24$	Occurrences	9,986	14
	RMS error on $\tilde{\phi}$	0.16°	98.8°
	RMS error on $\tilde{f_N}$	0.083%	0.215%

TABLE 1
The average RMS errors corresponding to the approximation of the minimum position for various N.

- A comparison of our approach with the state of the art angle estimation algorithm (section 5.3).

5.1 Efficient minimum search

In order to demonstrate the efficiency of our minimum search algorithm (see section 4), a set of 30,000 random functions $f_N(\phi)$ (see eq. (10)) has been created. For each of three different bandwidths $N = \{6, 12, 24\}$, there are 10,000 functions with random A_p and B_p. Firstly, we have computed for each function the exact solution ϕ and the approximation $\tilde{\phi}$ using respectively an exhaustive gradient descent and the proposed optimization. Secondly, we have compared both results and distinguished two types of situation:

1) The global minimum is correctly found by our approach: $\phi \approx \tilde{\phi}$.
2) Another minimum is found.

The second case derives from situations where the function admits more than one solution: In term of Zernike description, when there are $s-1$ secondary minima as small as the global one, that concretely means that the pattern is s-fold symmetric. For instance, a circular symmetric (2-fold symmetric) pattern can be rotated in an equivalent way by two different angles $\{\alpha, \alpha + 180°\}$. In reality, what truly determines the rightness of the retrieved angle is the vertical error on the depth of the approximated minimum.

Table 1 details the RMS errors on the angle $\tilde{\phi}$ and on the depth of the minimum $f_N(\tilde{\phi}) = \tilde{f_N}$ for both situations. The proportion of occurrences of the second case (the retrieved minimum is not the global one) is comprised below 0.15%. In the first case, the RMS errors of the angle and the minimum depth do not run over 0.81° and 0.2%, respectively, even for $N = 6$ (when the approximation is the coarser). In the second case, the RMS error of the angle is high because a different minimum is retrieved. However, the depth of this minimum is similar to that of the global one since the maximum RMS error of $\tilde{f_N}$ is comprised below 1% in the worst case. That means that the minimum found may not be the global one but is really close in terms of depth.

As a conclusion, our minimum search seems precise enough (less than 1° of error for most of the cases) even for the case $N = 6$ for which the approximation is the coarser; this correspond to 16 moments. However the appropriate number of moment depends on the application, since it will also influence the robustness of the similarity measure.

5.2 Comparative study of similarity accuracy

We have conducted experiments to test the efficiency of the proposed comparator with respect to the state of the art, in terms of similarity accuracy and robustness. To that aim, we have gathered 502 logo images (about 25% are binary images and others are gray level images). Some of them are shown in figure 3. These images were cropped and re-sized in order to fit the size of a square of 100×100 pixels. Then, synthetic distorted images were generated for each of these original patterns. What is denoted as *distortions* include additive uniform noise, non-affine geometric deformation[1], occlusion and translation. Examples of such distorted patterns at various levels are displayed in figure 4. These four types of distortion, applied at 19 different degrees, have been used to create four corresponding databases. In each database, the set of one original pattern and its 19 gradually distorted versions constitutes one class. Hence, one can see each test database as a collection of 10,040 different patterns divided into 502 classes of 20 elements.

5.2.1 Databases construction

For each test database, the distortion is applied gradually among 19 levels. The four types of distortion are:

1) An additive uniform noise added to the patterns such that the SNR varies from 30dB

[1] the case of affine deformation has already been investigated in [19]

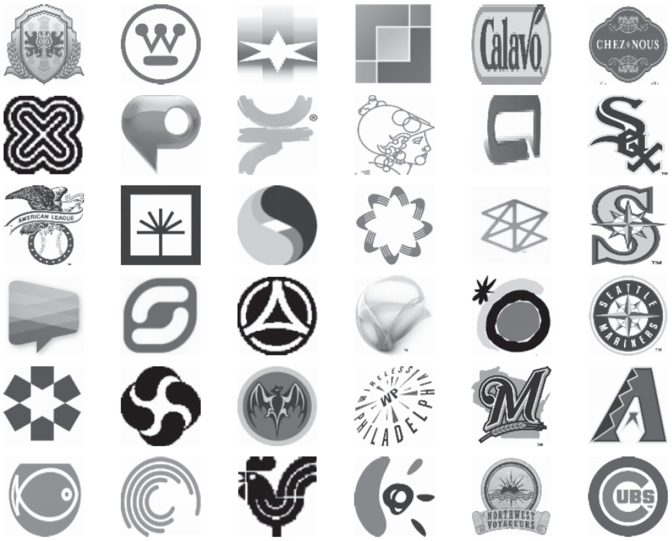


Fig. 3. Some of the 502 patterns used in the experiments.

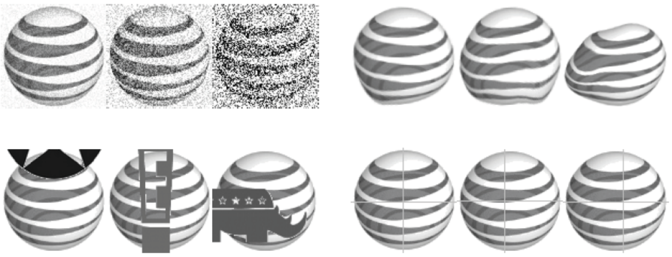


Fig. 4. Illustration of the four types of distortion. From left to right the four groups of distortions: additive noise, deformation, occlusion, translation. In each group the 6th, 13th and 19th distortion levels are presented. The images are inverted (black becomes white) for visualization purposes.

(level 0) to 1.5dB (level 19). The resulting values are bounded in [0,255].

2) A non-affine smooth deformation (i.e. a geometric deformation) whose amplitude varies locally from 0 (level 0) to 25 pixels (level 19) at most. This deformation is generated by two 100x100 displacement maps Dx and Dy that are initialized to 0, except for 8 random cells where we create vectors of the given amplitude and random directions. These 8 deformation vectors are then diffused until convergence (principle of the heat diffusion). The deformation function consists in $DeformedPattern(x,y) = Pattern(x + Dx(x,y), y + Dy(x,y))$. We used bi-linear interpolation for non-integer values of displacement.
3) A partial occlusion: we chose another random pattern and we paste it on the original one such that the occluded part varies from 0 (level 0) to 47.5 percents (level 19). The white parts of the pasted pattern are set transparent.
4) A translation. Its direction is random and its amplitude is comprised between 0 (level 0) and 7 pixels (level 19).

5.2.2 Comparison with the classical Zernike comparator

In order to conduct an exhaustive evaluation of our approach with respect to the classical way of comparing two Zernike descriptors, we have considered the similarity measure evaluation as a classification problem like in [4]. We have made two types of measurement using both comparators on the four databases:

- The recall-precision graph considering the previously defined classes. This measure accounts for the classification capability of our comparator and allows its complete evaluation for indexing issues.
- The error rate on first retrieval: knowing a distorted pattern, we try to retrieve the original one. Contrary to the recall-precision measurement, this one corresponds to a recognition scenario where only the first retrieval is important.

The Zernike moments of the patterns are extracted from the center of mass of the patterns in the case of noise and deformation, and from the center of the pictures for the cases of occlusion and translation since it has no sense to compute a center of mass when the pattern is occluded. All experiments were performed for three different numbers of moments : 25 (up to 8thorder), 49 (12thorder) and 81 (16thorder), which are representative of the numbers of Zernike moments used in the literature. In all cases, results were about the same so we decided to only illustrate the 12thorder case.

5.2.2.1 Recall Vs Precision: A good way of measuring the efficiency of our approach in a global indexing framework is to evaluate the recall vs. precision of both comparators (see fig. 5). The experiment consists in comparing a pattern P of the database to every others, then to order the results by similarity, and finally to count the number A of relevant patterns (i.e. from the same class than P) in the first N retrieved patterns. The precision is

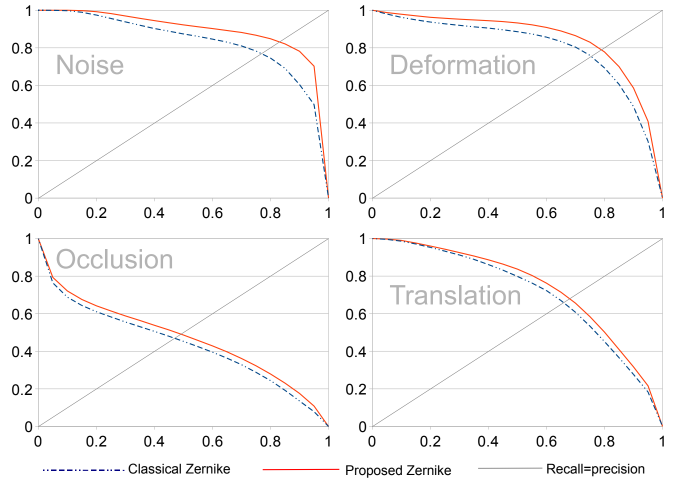


Fig. 5. Recall-Precision curves for 49 moments (i.e. up to 12thorder) on the four databases: additive random noise, non-affine deformation, occlusion, and translation. Each database contains 10,040 patterns divided into 502 classes of one original and 19 distorted versions.

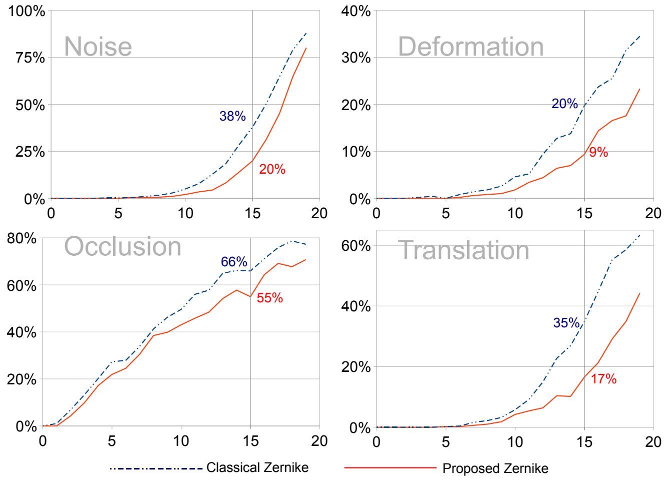


Fig. 6. The proportion of recognition errors when trying to retrieve the original pattern (among the 502 ones) from a distorted version using the classical comparator and the proposed comparator. The percentage of errors is highlighted for each method at the 15th distortion level.

then equal to A/N and the recall to A/A_{max}, with A_{max} the size of the corresponding class. The results (averaged over all patterns from all the classes of the corresponding database) are illustrated in fig. 5 for each database (noise, deformation, occlusion, translation); the precision is presented for each value of recall. Practically, the higher is the curve, the better is the similarity measure. The first tier (i.e. recall value for $N = A_{max}$) and second tier (i.e. recall value for $N = 2A_{max}$) measures have also been computed for each graph (see table 2).

As shown by the curves, the proposed comparator performs always better than the classical one. The gain is particularly high for additive noise and deformation, where the proposed method performs up to respectively 6.8% and 5.6% better in terms of the first tier measure.

5.2.2.2 Recognition performances: We have also conducted recognition experiments where the objective, starting from each of the 10,040 images of each database, is to retrieve the correct corresponding original pattern among the 502 original ones. The figure 6 displays the percentage of recognition errors (an erroneous pattern is returned in first position) for the classical and proposed comparator as a function of the distortion level. Globally, the proposed comparator is making about twice less errors than the classical one for medium distortions, except for the case of occlusion where our method makes about 15% less errors only.

In conclusion, the use of both magnitude and phase during comparison much improves the efficiency of the descriptor in terms of classification and recognition. Indeed Recall-Precision measures have shown that the proposed method better classifies the set of patterns than the classical comparator in spite of huge distortions. Besides there were about twice less recognition errors, in our second experiment, than with the classical method for medium distortion strengths.

5.2.3 *Comparison with geometric hashing and generalized Hough transform*

We have also compared our approach with two state-of-the-art methods for sketch and object recognition: the geometric hashing [17] and the recent deformation tolerant generalized Hough Transform from Anelli et al. [18]. The geometric hashing is a well-known indexing technique that is used to quickly find matches between two sets of features (e.g. points). On the contrary, the generalized Hough transform is an object identification technique suitable for matching a shape contour model with unsegmented images: edges are extracted in the image and each edge point casts a vote in the space of the parameters (x and y position of the object to find). Anelli et al. have added two extra steps after the voting phase: 1) the votes are clustered in order to deal with small local deformations and then 2) the shape segmentation is further verified

	Noise		*Deformation*		*Occlusion*		*Translation*	
	1st tier	2nd tier	1st tier	2nd tier	1st tier	2nd tier	1st tier	2nd tier
Classical comparator	82.9%	86.2%	77.9%	84.8%	44.2%	50.8%	68.7%	75.9%
Proposed comparator	89.7%	92.0%	83.5%	88.9%	47.2%	53.3%	71.6%	78.3%
Difference	**+6.8%**	**+5.9%**	**+5.6%**	**+4.1%**	**+3.0%**	**+2.5%**	**+2.9%**	**+2.4%**

TABLE 2
Comparison of the first tier and second tier measures for each database.

by back-projecting the image segments on the shape model and computing the model coverage score.

- For the geometric hashing, we need feature points, hence we have extracted Harris [20] and DoG keypoints [21], [12] in each pattern. An example of pattern with its keypoints is presented on figure 7. The descriptor for a given pattern consists of a list of its 15 strongest keypoints in terms of response value. Then, we have trained a hash table for each type of distortion. In the retrieval step, we begin by extracting the 25 strongest keypoints of the pattern to identify (we take more than 15 because the presence of noise can add new keypoints), and we successively project them on the hash tables of the learned patterns for different random basis (50 times). The best scores are stored for each learned pattern and it constitutes at the end the final matching scores. We deliberately chose small values for the number of keypoints and iterations so that a comparison is tractable in terms of time (see table 3).
- In the case of the deformation tolerant GHT (DT-GHT) from Anelli et al. [18], each pattern was first indexed (i.e. canny edges extraction, segmentation of the edges, building of a R-table). Then to compare two patterns, we simply search the first one (the model) into the second one. This returns a matching score between 0 and 1 (1 means perfect match).

In order to properly compare our Zernike comparator with these methods, we have evaluated the recall vs. precision scores. However the comparison steps for geometric hashing and Hough transform are both quite slow so it was not possible to process the $10,000^2$ comparisons required for the recall-precision measures. As a consequence, we deliberately sampled each database into 5 deformations levels (3rd, 7th, 11th, 15th and 19th) × 100 patterns instead of 502, with each type of pattern (gray level / black & white) equally sampled. Each of the four databases thus contains 500 different elements.

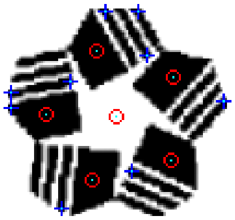


Fig. 7. 15 strongest keypoints of a pattern detected by the Harris corner detector (blue crosses) and the extrema in scale-space of Dog (red circles)

Method	*time (s)*
Zernike classic	0.002
Zernike proposed	0.057
DT-GHT	37.2
Geometric Hashing	26.9

TABLE 3
Average processing times for 1000 comparisons on a 2 GHz computer for similarity calculation with each method

5.2.3.1 Timing Analysis: Table 3 gives a comparison of the similarity calculation processing time for each method. Even if our method is slower than the classical Euclidean comparison, its complexity is the same ($O(order^2)$) and both processing times are very low (less than 60 milliseconds for 1000 comparisons for our approach). However the DT-GHT and the geometric hashing are several orders of magnitude slower (37 and 27 seconds respectively for 1000 comparisons). With a number of comparisons of 500^2 per database for the recall-precision experiment, it still required about 10 hours to process the four databases for the DT-GHT or the geometric hashing against less than one minute for Zernike experiments.

5.2.3.2 Recall Vs Precision: Figure 8 and table 4 illustrate the results.

In the case of translation, the geometric hashing

	Noise		*Deformation*		*Occlusion*		*Translation*	
	1st tier	2nd tier	1st tier	2nd tier	1st tier	2nd tier	1st tier	2nd tier
Zernike proposed	**93.4%**	**95.6%**	**88.6%**	**93.6%**	**52.9%**	**60.6%**	78.0%	85.2%
DT-GHT	75.5%	79.4%	88.4%	**93.6%**	43.5%	49.4%	**96.5%**	**98.4%**
Hashing	62.0%	63.6%	57.2%	60.2%	47.8%	49.9%	86.9%	87.4%

TABLE 4
Comparison of the first tier and second tier measures for each comparators and for each distortion.

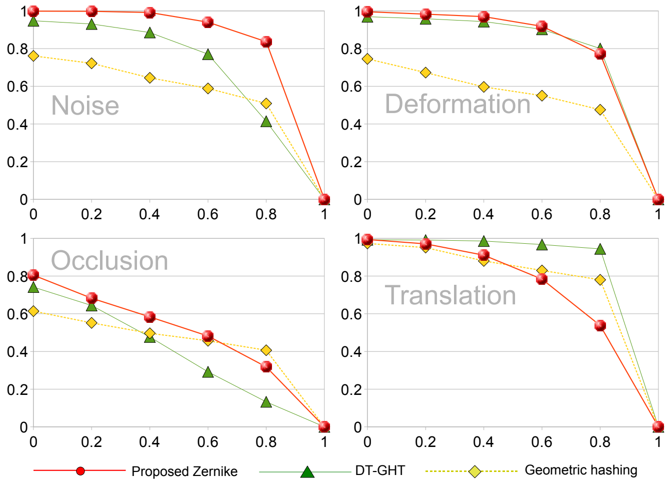


Fig. 8. Recall-Precision curves on the four reduced databases for the following methods: proposed Zernike, DT-GHT, geometric hashing. Each database contains 500 patterns divided into 100 classes of one original and 4 distorted versions.

and DT-GHT approaches perform uppermost due to their invariance in position (however the geometric hashing is not so good because of the interpolation noise), but it can be noticed that the proposed Zernike method does not perform so bad (almost 80% on the first tier measure). For every other distortions, our approach performs better than DT-GHT and geometric hashing in terms of the first and second tier measures ; in the case of noise, the proposed method works largely better since the noise is localized in the high frequencies and hence has a small effect on the low order Zernike moments. The geometric hashing achieves low performance in all cases except for translation since the keypoint response values and locations are strongly disturbed when the images are distorted.

5.3 Rotation angle retrieval

We have conducted experiments to demonstrate the accuracy of our approach in the specific case of rotation angle estimation between similar but distorted patterns. We have performed the experiment on the previous set of 502 logo images. The images were rotated by random angles. Then, various distortions (the same as in section 5.2) were applied to each rotated pattern. The distortion levels, however, are twice smaller than in the previous section otherwise images would have been too much different to be put in correspondence by a rotation. Then, we have estimated the rotation angle with respect to the original pattern using the proposed algorithm and the most robust and acknowledged method in the state of the art: the Kim and Kim robust estimator [16]. The total number of test images is thus 200,800: 502 (the number of original patterns) × 10 (the number of rotations) × 10 (the number of distortion levels) × 4 (the number of types of distortion).

To compare to the estimator from Kim and Kim [16], we have used the same number of Zernike moments than them: 25 moments were computed up to order 8. For each of the four databases (noise, deformation, occlusion, translation), we have computed the root mean square (RMS) error of the retrieved angles using each algorithm; results are presented in table 5. The RMS error using our method is systematically lower than for the estimator from Kim and Kim, for all the types of distortion. In particular, in the case of additive noise and deformation, the proposed method provides results whose accuracy is about twice better than with the estimator from Kim and Kim. The occlusion and translation cases yields an important RMS error: in the first case, this comes from the fact that sometimes some crucial parts of the pattern in terms of rotation evidences are occluded ; in the second case this is caused by the displacement of the rotation center.

6 2D-3D OBJECT RECOGNITION FROM HAND-DRAWN SKETCH

The proposed algorithm has been tested within a real industrial application: the automatic transcription of sketched storyboards into reconstructed

	No distortion	*Additive noise*	*Deformation*	*Occlusion*	*Translation*
Kim and Kim method	0.61°	1.13°	2.96°	27.3°	17.1°
Proposed method	0.57°	0.42°	1.62°	23.4°	12.8°

TABLE 5
The average RMS error of rotation angle using the estimator from Kim and Kim [16]and the proposed method.

3D scenes. An example of such scenario is illustrated in figure 9: a 3D scene has been generated from a hand-drawn storyboard. Formally, an object database contains various 3D models (as polygonal meshes); on the other side, the cartoonist draws a 2D storyboard by digitally sketching the objects of the scene he has in mind. Then the objective is to recognize each 3D model from the corresponding piece of sketch, along with its 3D viewing angle, its scale and its rotation angle in the drawing plane, so as to automatically and correctly place it in the 3D scene. To obtain this result, an important number of views of each 3D model is indexed using Zernike moments like in [13] (about 50 views per object). We do not compute Zernike moments on the whole view, but on a bounding circle containing at least 70% of the pattern. We limit the indexed surface for the following reasons:

1) we try to reduce the surface potentially disturbed by the background during retrieval
2) such approach may bring some kind of robustness regarding small occlusions. A complete robustness to occlusion could be reached if each model view was described by several Zernike circles of various scales and positions, however, this development is out of the scope of this paper.

Once we have described all the 3D model views, we search them in the sketched story-board with a two-pass process to speed up the recognition. The first pass aims at finding a set of raw correspondences and the second pass refines this set. Without loss of generality, we will now describe the search of one given model into the sketch: in the first pass, we scan the whole sketched story-board with a circular window at different scales and positions. For each position, we describe the window using Zernike moments and we store the model view which achieves the smallest distance with the local descriptor according to our comparator. We use a standard non-maxima suppression technique in the resulting scale-space of distances and we apply a first threshold to the obtained maxima: we obtain a set of *potential matches.* This first search is processed for a reduced set of positions and scales in order to reduce the processing time: 8 scales, starting from 400×400 pixels until 1400×1400 pixels, and for each scale the window slides over the image with a step width of $windowSize/15$ (since that corresponds to a translation whose amplitude is correctly retrieved in section 5). In our experiment, the story-board dimension is 3350×2260 and this greedy recognition takes about 2 minutes on a 2Ghz machine. The computation of Zernike moments is made faster by precomputing a set of $100{\times}100$ Zernike filters and by applying a fast smoothing approach along scales (like pyramids of Gaussian in [12]) to quickly sample each $100{\times}100$ window. As we saw in section 5.2, Zernike distance is not so invariant to translation, so we refine the search during a second pass: for each potential match we search locally around its position in the scale-space (the steps widths are decreased both in position and scale). This allows to find the optimal position for each potential match and thus to make the difference between true and false positive after a second thresholding. This second pass takes about 30s per potential match (there is typically 4-6 potential matches per storyboard image).

Moreover, we assume that the 3D models are approximately vertically positioned on an horizontal flat ground, which means that their 3D vertical axis are probably also vertical in the picture. Since our method provides not only the similarity but also the the rotation angle, we are able to eliminate from the first pass further false positives which could not be eliminated by the classical Zernike comparator alone. Recognition results are presented in the figure 10. Note that all the bushes are not detected: this is caused by the fact that sometimes their sketch deviates too much from their 3D model. Even if there are slight errors in the placement estimation, the recognition results are still acceptable.

For such an application, the estimations are significantly better and more stable with the proposed comparator than with the classical comparator coupled with the in-plane rotation estimator from [16], which is less reliable when it has to estimate a rotation angle between two different patterns (see 5.3).

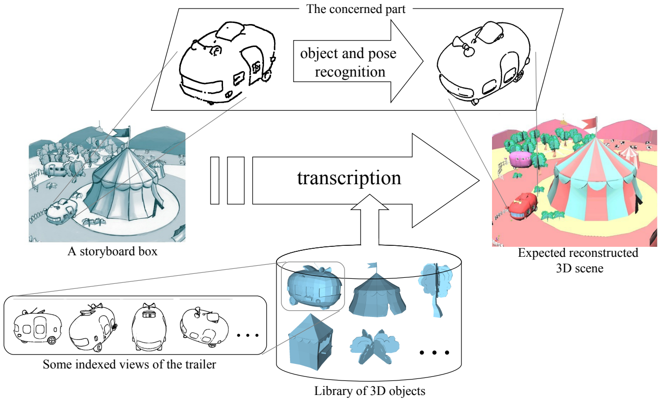

The concerned part
object and pose recognition
transcription
A storyboard box
Expected reconstructed 3D scene
Some indexed views of the trailer
Library of 3D objects

Fig. 9. Example of 2D/3D object recognition application: reconstruction of a 3D scene from a hand-drawn 2D storyboard associated with a 3D model database. On the top, the recognized 3D model (on the right) is the best retrieved result between the drawing (on the left) and the whole set of indexed views, using the proposed Zernike comparator. Doing so, the 3D pose and the in-plane rotation angle are retrieved in the same time and thereafter allows the 3D reconstruction.

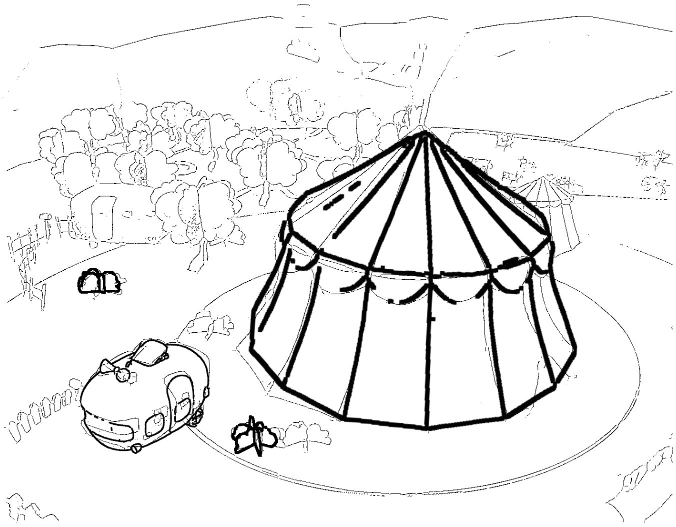


Fig. 10. Recognition results for the circus tent, the trailer and the bushes pasted on the sketch image after edge extraction. No human intervention was needed to obtain this result.

7 Summary and Discussion

We have presented an efficient comparator of Zernike descriptors whose novelty is to take advantage of the phase information in the comparison process while still preserving the invariance to rotation. The provided similarity measure is more robust to distortions (especially geometrical deformation and noise). The recognition errors are also about twice less important for medium distortions than with the classical comparator (i.e. the Euclidean distance between Zernike magnitudes). Moreover, our approach has the same $O(order^2)$ complexity than the classical one. Finally, it provides for free an estimation of the rotation angle that outperforms the robust estimator from Kim and Kim [16]. To conclude with, this novel theoretical contribution to the Zernike framework can apply to any application already using Zernike moments. It is worth noting that it can also apply to the pseudo-Zernike moments [22] as both theories share almost the same background.

References

[1] M. Teague, “Image analysis via the general theory of moments,” *J. Optical Soc. Am.*, vol. 70, no. 8, pp. 920–930, August 1980.

[2] S. Belkasim, M. Shridhar, and M. Ahmadi, “Pattern recognition with moment invariants: A comparative study and new results,” vol. 24, no. 12, pp. 1117–1138, 1991.

[3] S. Liao and M. Pawlak, “On image-analysis by moments,” vol. 18, no. 3, pp. 254–266, March 1996.

[4] A. Khotanzad and Y. H. Hong, “Invariant image recognition by zernike moments,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, 1990.

[5] C. Kan and M. Srinath, “Invariant character recognition with zernike and orthogonal fourier-mellin moments,” vol. 35, no. 1, pp. 143–154, January 2002.

[6] C.-W. Chong, P. Raveendran, and R. Mukundan, “A comparative analysis of algorithms for fast computation of zernike moments.” *Pattern Recognition*, vol. 36, no. 3, pp. 731–742, 2003.

[7] C.-Y. Wee and R. Paramesran, “Efficient computation of radial moment functions using symmetrical property,” *Pattern Recogn.*, vol. 39, no. 11, pp. 2036–2046, 2006.

[8] ——, “On the computational aspects of zernike moments,” *Image Vision Comput.*, vol. 25, no. 6, pp. 967–980, 2007.

[9] L. Kotoulas and I. Andreadis, “Real-time computation of zernike moments,” vol. 15, no. 6, pp. 801–809, June 2005.

[10] G. Amayeh, A. Erol, G. Bebis, and M. Nicolescu, “Accurate and efficient computation of high order zernike moments.” in *ISVC*, 2005, pp. 462–469.

[11] B. Reddy and B. Chatterji, “An fft-based technique for translation, rotation, and scale-invariant image registration,” vol. 5, no. 8, pp. 1266–1271, August 1996.

[12] D. G. Lowe, “Object recognition from local scale-invariant features,” in *International Conference on Computer Vision*, 1999, pp. 1150–1157.

[13] T. Filali Ansary, M. Daoudi, and J.-P. Vandeborre, "A bayesian 3D search engine using adaptive views clustering," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 78–88, January 2007.

[14] S. Hou and K. Ramani, "Calligraphic interfaces: Classifier combination for sketch-based 3d part retrieval," *Comput. Graph.*, vol. 31, no. 4, pp. 598–609, 2007.

[15] L. B. Kara and K. Shimada, "Sketch-based 3d-shape creation for industrial styling design," *IEEE Comput. Graph. Appl.*, vol. 27, no. 1, pp. 60–71, 2007.

[16] W.-Y. Kim and Y.-S. Kim, "Robust rotation angle estimator," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 768–773, 1999.

[17] H. Wolfson and I. Rigoutsos, "Geometric hashing: an overview," *Computational Science & Engineering, IEEE*, vol. 4, no. 4, pp. 10–21, 1997.

[18] M. Anelli, L. Cinque, and E. Sangineto, "Deformation tolerant generalized hough transform for sketch-based image retrieval in complex scenes," *Image Vision Comput.*, vol. 25, no. 11, pp. 1802–1813, 2007.

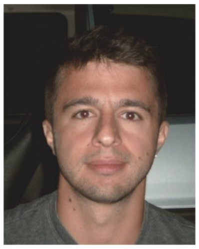
[19] S. K. G.R. Amayeh and A. Tavakkoli, "Improvement of zernike moment descriptors on affine transformed shapes," in *IEEE International Symposium on Signal Processing and its Applications*, 2007.

[20] C. Harris, "Geometry from visual motion," pp. 263–284, 1993.

[21] J. L. Crowley and A. C. Parker, "Representation for shape based on peaks and ridges in the difference of low-pass transform," Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-83-04, May 1983.

[22] C.-H. Teh and R. T. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 4, pp. 496–513, 1988.

Jérôme Revaud received a diplome d'Ingénieur in Computer Science (2006) and a master's degree in Image Processing (2007) from the Institut National des Sciences Appliquées de Lyon (INSA, Lyon, France). He is currently a PhD student in computer graphics supervised by Prof. Atilla Baskurt from INSA of Lyon and Prof. Yasuo Ariki from Department of Computer and System Engineering of Kobe University, Japan. His main research focuses on 3D object recognition applicable to robots.



Guillaume Lavoué has an Engineering Degree in Electronic, Signal Processing and Computer Science from CPE-Lyon (France), an MSc in Image Processing from the University Jean Monnet of St-Etienne (France) and a PhD in Computer Science from the University Claude Bernard of Lyon (France). From February to April 2006, he was a postdoctoral fellow at the Signal Processing Institute (EPFL) in Switzerland. Since September 2006, he has been Associate Professor at the French engineering university INSA of Lyon. He is involved in the M2Disco team from the LIRIS Laboratory (UMR 5205 CNRS). His research interests focus on 3D model analysis and processing, including compression, watermarking, geometric modeling and 2D/3D recognition.

Pr. Atilla Baskurt leads his research activities in two teams of LIRIS: the IMAGINE team and the M2DisCo team. These teams works on image and 3D data analysis and segmentation for image compression, image retrieval, shape detection and identification. His technical research and experience include digital image processing, 2D-3D data analysis, compression, retrieval and watermarking, especially for multimedia applications. He has published over 150 publications in some of the most distinguished scientific journals and international conferences and 7 book chapters. He is the co-author of the recent book "3D Object Processing: Compression, Indexing and Watermarking", edited by John Wiley & Sons, in april 2008. He served as associate editor for the EURASIP Journal on Applied Signal Processing (JASP) between 2005-2008. Pr. A. Baskurt is "Chargé de mission" on Information and Communication Technologies (ICT) at the French Research Ministry.