MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

8—21

Local Zernike Moments Vector for Content-Based Queries in Image Data Base

Erwan Bigorgne*, Catherine Achard, Jean Devars
Laboratoire des Instruments et Systèmes d'Ile de France
Université Pierre et Marie Curie

Abstract

The definition of reliable local signal characterizations is of great importance for many computer vision tasks as mosaïcing, 3D-scene reconstruction or more recently in applications like content-based image retrieval systems. The following study concerns this last general pattern. Aiming at this, we present the use of Full-Zernike moments as a local characterization of the image signal. Their computation allows us to construct an invariant vector, of which the projection in an index table (feature space) provides a vote for some model-images. This approach is based on the quasi-invariant theory applied to perspective transformations and is an extension of a standard point to point matching between two images. It addresses the problem of similarity search in high dimensional space ($d > 20$).

1 Introduction

At the present time, there is a growing need of effective schemes in order to well manage and navigate through large collections of images or videos. A useful searching method is the query by examples. Similarity searches are then carried out between features extracted from a query image and those contained in the database. Swain and Ballard [1] have pioneered with success the use of global color signatures, nevertheless systems based on local features are very effective for inserted or occluded object recognition [2]. According to this last paradigm, the projection of the extracted features in the feature space produce votes for some indexed images. To be useful, local features have to show specific invariances:

- The characteric vectors has to be invariant to similarity transformations (composition of rotation, translation and scaling). Such characterizations are then quasi invariant to narrow bounded perspective transformations [3].
- Invariance to photometric changes is implicitly required for a view-based object recognition system.

In this article we propose the use of Zernike moments as a local description of feature points. We describe the so-computed quasi-invariant vector in section 2. A particular attention will be devoted to the invariance against rotation that is achieved without loss of the completeness properties of the set. In section 3 we present an adapted treatment in order to obtain the invariance against large scale changes (> 20%) regarding the scale-space theory. Furthermore, a normalization of the signal carries out an invariance against locally affine photometric changes. We have evaluated the capabilities of the proposed description for a simple matching task, and for image/object retrieval. In the last section, we describe the first results obtained with the use of an original clustering sheme [14] in order to avoid an exhaustive scanning of the database.

2 Feature Vector

Since their introduction by Hu [4] in 1961, the use of moments or functions of moments is widespread in pattern recognition domain. Moments allow us to form some invariant signatures to a large collection of image transformations. Teh and Chin [5] provide a study, both practical and theoretical, of characteristics of different moments in terms of noise sensitivity and redundancy. More particularly, they emphasize that the moments set formed by a decomposition on an orthogonal basis set is uncorrelated and then provides a more efficient decomposition of the signal. Finally, they show that in terms of overall performance, Zernike and Pseudo-Zernike moments outperform other descriptions.

Moments are regularly use for the characterization of large region or closed -well segmented- shape in the image. We show here that Zernike moments are able to provide an efficient local characterization around feature points. Interest points are usually defined as maxima of a "cornerness" measurement

*Address: UPMC, 4 place jussieu, 75252 Paris Cedex 05, boite 164, France. E-mail: erwan.bigorgne@lis.jussieu.fr

[10]. In the case of small perspective transformation hypothesis, these points are considered as particularly reliable features: they are highly repeatable on multiple views of the same scene and they focus relevant informations.

2.1 Zernike moments

The use of Zernike moments is quite usual in the characters recognition domain. Actually, we are interested here in their efficiency for local feature matching. Details about their construction may be found in [6] and [7] (for series expansions).The basis set of Zernike moments is a modified form of Jacobi polynomials. Zernike moments $\{A_{nl}\}$ are computed in a discrete form as:

$$\hat{A}_{nl} = \frac{n+1}{\pi} \tag{1}$$

$$\cdot \sum \sum_{x_i^2+y_j^2 \leq 1} V_{nl}\left[\sqrt{x_i^2+y_j^2}, \arctan\left(\frac{y}{x}\right)\right]^* \cdot f(x_i, y_j)$$

where

$$\begin{aligned} V_{nl}(\rho,\theta) &= R_n^{|l|}(\rho)e^{il\theta} \qquad (2) \\ &= [\sum_{s=0}^{(n-|l|)/2} (-1)^s \\ &\quad \cdot \frac{(n-s)!}{s!(\frac{n+|l|}{2}-s)!(\frac{n-|l|}{2}-s)!}\rho^{n-2s}]e^{il\theta} \end{aligned}$$

with the condition on l and m: $n = 1, 2, ...\infty$, $|l| \leq n$ and $n - |l|$ is even. (x_i, y_j) are the pixel coordinates in a local circular area centered on the feature point.

Tests have shown that up to the 9th order, noise does not interfere with the matching. Then we use typically 45 dimensionnal vectors computed on a circular area which the radius varies from 6 to 12 pixels.

2.2 Invariance to rotation

Zernike moments are rotation-variant feature but it is easy to show that for a rotation of the image by an angle α, the resulting moment is given by $\hat{A}'_{nl} = e^{-jl\alpha}\hat{A}_{nl}$. Several searchers as Khotanzad and Hong [8] use the magnitude of Zernike moments as features. It implies an important loss of information since different polar structures can produce moments with the same magnitude. An other way to avoid this loss consists in considering combination of power of moments [9]. However, this method produces an highly dimensional and correlated characterization.

So to use the complete Zernike moments, a preprocessing step is required. Since computations are applied to gray-level images, a brutal re-orientation of pixel area along a main direction should involve numerical errors (in addition to be costly). A more efficient and stable solution is to introduce ϕ as a forced dephasing in the definition of Zernike polynomials: $V_{nl}(\rho,\theta) = R_n^{|l|}(\rho)e^{il(\theta-\phi)}$. ϕ is chosen as the mean direction of gradient over the pixel area. Figure 1 reports the false matching ratios in the case of a rotating planar scene. Four images of the sequence are displayed in figure 2. In spite of the use of a standard webcam for the acquisition process, none particular processing have been realized except a gaussian filtering. The number of matched points varies from 50 to 100. The few mismatches originate often in locally repetitive patterns.

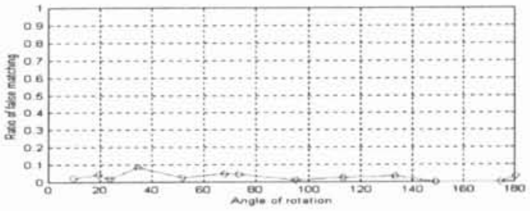

Figure 1: Ratios of false matching

Figure 2: Four shots of sequence "rotation"

3 Scale invariance

The already designed characterization is intrinsically invariant to image translations and rotations. In this section, we describe the complementary ways to take into account large scale changes.

Experiments show that matching based on Zernike moment vector is well carried out until a 20% scale change. In order to manage greater changes, we apply a multi-resolution technique for the matching process. In fact, no prior knowledge on the "pattern" scale is available, and the patterns are not closed shapes and then can't be framed. This method can be considered as a hypothesis generation-validation process for which the ratio of matching over the feature points population is the likelihood measurement.

As processed features are gray-leveled, techniques based only on a change of the computation domain size and an adequate normalization can not be reliable. In consequence, we use a scale-space representation of image structures. This approach introduced by Koenderink [11] justifies formally the use of the well-known gaussian "blurring". We briefly display the bases of this theory.

For a N-dimensional continuous signal $f : \mathbb{R}^N \rightarrow \mathbb{R}$, its scale-space representation $L : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined as the solution to the diffusion equation:

$$\partial_t L = \frac{1}{2}\nabla^2 L \qquad (3)$$

with initial condition $L(\cdot;0) = f(\cdot)$. Several works (e.g. [12]) prove that within the class of linear transformations the gaussian kernel is the unique kernel for generating a scale-space. This family is then defined by the particular solutions of the equation 3:

$$L(\cdot;t) = g(\cdot;t) * f(\cdot)$$

where $g : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is given by:

$$g(\vec{x};t) = \frac{1}{(2\pi t)^{N/2}} e^{-(\sum_i^N x_z^2)/2t}$$

In the following we pose $t = \sigma^2$, the variance of the gaussian.

Assuming this, let us consider two signals f and f' related by a scale change s, then $f(\vec{x}) = f'(\vec{x}') = f'(s\vec{x})$. The scale-space representation of f and f' in the two domains is defined by:

$$L(\vec{x};\sigma^2) = g(\vec{x};\sigma^2) * f(\vec{x})$$

$$L'(\vec{x}';\sigma'^2) = g(\vec{x}';\sigma'^2) * f(\vec{x}')$$

A simple change of spatial variables and scale parameter according to $\vec{x}' = s\vec{x}$ and $\sigma'^2 = s^2\sigma^2$ yields:

$$L'(\vec{x}';\sigma'^2) = L(\vec{x};\sigma^2)$$

This formal approach is applied naturally to a scale-oriented computation of Zernike moments. They are now calculated on the scale-space representation of f and f' and yield to an invariant characterization if the (implied-) hypothesis on the scale factor s is true.

$$A_{nl} = \frac{1}{\sigma^2}\frac{n+1}{\pi}\iint_{x^2+y^2\leq 1} V_{nl}(x,y)^* L((x,y);\sigma^2)dxdy$$

$$= \frac{1}{\sigma'^2}\frac{n+1}{\pi}\iint_{x'^2+y'^2\leq s^2} V_{nl}(\frac{x'}{s},\frac{y'}{s})^* L'((x',y');\sigma'^2)dxdy$$

$$\equiv A'_{nl}$$

where A'_{nl} is computed over an oversized domain (in pixel) by a factor s.

It is stated that a characterization is reliable for scale change up to 20%. In practice, in order to be robust to a scaling factor varying from 0.5 to 2, moments are computed only at three scales e.g. $\sigma \simeq (1, \sqrt{2}, 2)$ - radii of computation area in pixel have to follow the same scale law -. Matching is then attempted for five hypothesis of scale change : $s = \sigma_1/\sigma_2 = 1/2, 1/\sqrt{2}, 1, \sqrt{2}, 2$. Figure 3 shows the overlapping of the scale-characterization for matching of scene at different scales (figure 4).

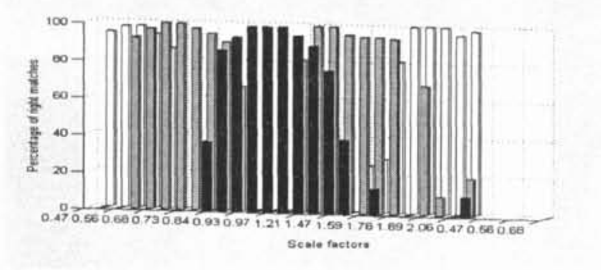

Figure 3: Multi-scale matching. Whatever the scaling factor, the pourcentages of right matches is superior to 95% for a combination of scale-space representation

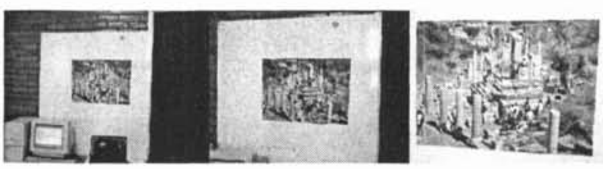

Figure 4: The scene at three different scales

4 Invariance to illuminance condition

Let $I(x,y)$ be the luminance function. We consider here locally affine photometric changes as:

$$I'(x_i,y_i) = a_{Ni}I(x_i,y_i) + b_{Ni}$$

where a_{Ni} and b_{Ni} are supposed to remain constant over a larger neighborhood than the area D over which Zernike moments are computed. Then a function based on local standard deviation normalized by the local gradient magnitude is invariant to such illuminance modifications. Let this function $f_D(x_i,y_i)$ be:

$$f_D(x_i,y_i) = \frac{I(x_i,y_i) - 1/card(D)\sum_D I(x_j,y_j)}{1/card(D)\sum_D \sqrt{I_x^2(x_j,y_j) + I_y^2(x_j,y_j)}}$$

Hereafter, the whole of moments is computed on this function $f_D(x_i,y_i)$ -The scale-space representation too-. Matching's results of an image with three others taken under different illuminations are reported in the table below.

image	mean	% of true matching
1	intensity (102)	with image 1
2	76	100
3	60	100
4	35	100

5 Matching and content-based image retrieval

We have evaluated the capabilities of the proposed characterization for simple matching tasks. The matching is essentially based on the computation and the sorting of mahalanobis distances between characteristic vectors. The required covariance matrix is computed over an adequate set of feature points. The complete matching process includes a locality condition: two points are matched if a given part of their nearest geometrical neighbours are sufficiently close in the feature space. Very good results are obtained. A similar approach is employed for object recognition in the database, but in this case, we make use of a K-nearest neighbours search in a 45-dimensional space. Similarity search in high dimensional spaces raises a problem known as "dimensional curse" : space- and data-partitioning approaches tend to scan the whole of feature space[13]. In order to avoid an exhaustive scanning of the data base, we have tested an original clustering scheme [14]. Meta-clusters are formed by the iteration of the algorithm. A tree-like structure is then employed as a way to calculate the relevant nearest distances. We have realized some tests on the very popular Columbia Object Image Library. In order to test the robustness of this approach, only 25 degrees incremented around-views for each 3D object are retained in the database (14 images per object). In these first experiments only 13 objects are considered. The base is then composed by 182 images. For simple queries (a given view obviously not contained in the base) the recognition rate is 1. Figure 5 shows the response of the system for an inserted and occluded object in a scene (image "house" from INRIA). The two most similar images are those of the correct object.

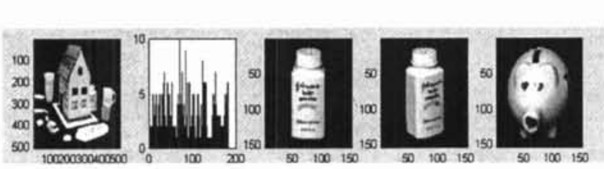

Figure 5: The three most similar reponses to a complex querie

6 Summary

In this paper, we have focused our attention on the design of a feature vector being quasi-invariant for perspective transformations. The use of complete Zernike moments provide an efficient local characterization of signal, with no need of high order moments. Furthermore, considering the proposed built-in invariance to illuminance condition and the capability of the process in managing important scale change, we investigate the integration of this characteristic vector to an efficient content-based image retrieval system for large databases. So obtained results are very promising.

References

[1] M.J. Swain and D.H. Ballard. *Color indexing.* International Journal of Computer Vision, 7(1):11-32, 1991.

[2] C. Schmid, R. Mohr. *Local grayvalue invariants for image retrieval,* IEEE Trans. on PAMI, Vol. 19(5), pp 530-535, may 1997.

[3] T.O. Binford and T.S. Levitt. *Quasi-invariants : theory and exploitation.* In Proceedings of DARPA Image Understanding Workshop, pp 819-829, 1993.

[4] M.K. Hu *Visual pattern recognition by moment invariants.* IEEE trans. on Inf. Th., 8:179-187, 1962.

[5] C.H. Teh and R. T. Chin, *On image analysis by the methods of moments,* IEEE Trans. on PAMI, Vol. 10, pp. 496-513, July 1988.

[6] M. Born and E. Wolf, *Principles of optics,* New York: Pergamon, 1975.

[7] R. Courant and D. Hilbert, *Methods of mathematical physics,* Vol. 1, New York : Interscience, 1953.

[8] A. Khotanzad and Y. H. Hong, *Invariant image recognition by zernike moments,* IEEE Trans. on PAMI, Vol. 12, pp. 489-498, May 1990.

[9] E. P. Simoncelli, *A rotation-invariant pattern signature,* Proc. of third Int. Conf. on Image processing, Vol. 3 pp185-188, Sept. 1996

[10] C. Achard, E. Bigorgne and J. Devars, *A Sub-Pixel and Multispectral Corner Detector,* Proceedings of ICPR00, Sept. 2000.

[11] J. J. Koenderink, *The structures of images,* Biol. Cybern., Vol. 50, pp. 363-370, 1984.

[12] M. Nielsen, L. Florack, and R. Deriche, *Regularization and Scale Space,* Research report n. 2352 INRIA Sept. 1994

[13] R. Weber, H-J.Schek, S. Blott, *A Quantitative analisys and performance study for similarity-search methods in high-dimensional spaces,* Proc. of the 24th VLDB Conference, New York, USA,1998.

[14] L. Prevost, M. Milgram, *Modelizing character allographs in omni-scriptor frame: a new non supervised clustering algorithm,* Pattern Recognition Letters 21 (2000) 295-302.