Content-Based Emblem Retrieval Using Zernike Moments

Ezequiel Cura, Mariano Tepper, and Marta Mejail

Departamento de Computacion, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires, Argentina

Abstract. The problem of content-based image retrieval is becoming essential in many real-world applications, mostly due to the growth in size of modern image databases. In particular, this work addresses the retrieval of trademark emblems, which is key for detecting trademark infringement. A common approach that proved suitable for this task, is to encode emblems using shape descriptors and Zernike complex moments. This work focuses on their study, proposing a two-fold contribution. First, we present some modifications to Zernike complex moments and then we explore the use of different comparison metrics. Both have shown to report improvements in retrieved results and in execution time.

1   Introduction

With the growth in size of modern image databases, the problem of content-based image retrieval is becoming crucial in many real-world applications. For some applications, determining duplicates, or look-alike images, is an essential task. The detection of trademark infringement is one of them, with deep commercial and monetary implications. Already registered trademark logos, or emblems, have to be protected against hundreds of new ones that are created every day. Fig. 1 depicts some examples of similar emblems.

The trademark infringement problem can be very simply stated: given a database of registered emblems, does a new emblem look similar to any of them? Unfortunately this problem cannot be solved in a fully automatic way by any state-of-the-art method. In practice, the new emblem is used as a query which retrieves the most similar emblems from the database. Then a human expert performs the final recognition test by only observing these remaining emblems.

Direct image comparison methods such as correlation fail to account for the perceptual similarity between two emblems. A successful approach is to regard emblems as shapes. Each shape in the database is then encoded by using some descriptor and these descriptors are ultimately used for comparison. Specifically, a query consists on sorting the emblems from the most similar to the least similar to the query emblem and then returning the first ones among them.

In general, when requirements impose real-time queries, only global methods are considered at the expense of possibly missing similar emblem subparts. For example, Shape Context [1] is commonly used for this task.

Fig. 1. Examples of look-alike emblems. In the first example, NBC agreed to pay a compensation fee to Nebraska ETV Network for using a similar emblem.

An alternative and popular approach is derived from the use of image moments [2]. Zernike Moments (ZMs) are an interesting choice as they provide rotation and translation invariance, minimal redundancy and limited robustness to noise [3,4]. Although it used to be computationally prohibitive to compute higher order ZMs, nowadays it has become not only possible but also fast and numerically precise [5]. Our paper builds upon these approaches, improving the accuracy and the performance of ZMs-based image retrieval.

First, we propose a modification to ZMs base on an appropriate weighting and on PCA. The second contribution of this paper is to show that by a simle two-stage method for comparing ZMs, high efficiency and accuracy are achieved. The first stage, relies on a fast pruning technique, that allow to quickly retrieve a small set of likely candidates. The second stage performs more accurate and time demanding comparisons but, by applying it to the small ouput of the first stage, high speed is achieved.

This paper is organized as follows. Zernike Moments are reviewed in Section 1.1. In Sections 2 and 3 we propose a modification to Zernike Moments and a fast composite distance, respectively. Section 4 presents results that demonstrate the pertinence of our approach and provides final remarks.

1.1 Previous Work on Zernike Moments

Starting from a complete set of complex polynomials defined by Zernike, Khotanzad and Hong [3] introduce complex Zernike Moments.

Definition 1. *A Zernike polynom* $V_{nm} : \mathbb{R}^2 \to \mathbb{C}$ *is defined as*

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(im\theta) \qquad (1)$$

where $0 \leq n$, $|m| \leq n$, $n - |m|$ *is even,* $\rho = |x + iy|$, $\theta = \arg(x + iy)$ *and*

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \; \frac{(n-s)!}{s!(\frac{n+|m|}{2} - s)! \, (\frac{n-|m|}{2} - s)!} \; \rho^{n-2s}.$$

The input for computing ZMs consists of a binary image, represented by a function $I : \mathbb{R}^2 \to \{0, 1\}$. Of course not every emblem is binary and it must consequently be binarized. Binarization methods are outside the scope of this work and will not be covered: it is sufficient to state that any such method is suitable.

Definition 2. *Given a binary image I, a Zernike Moment A_{nm} is defined as*

$$A_{nm} = \frac{n+1}{\pi} \iint\limits_{x^2+y^2\leq 1} I(x,y)V_{nm}^*(x,y) \ dx \ dy \tag{2}$$

where $V_{nm}^*(x,y)$ is the complex conjugate value of $V_{nm}(x,y)$.

From the above definition it can be trivially deduced that ZMs are translation invariant. Additionally, the image can be reconstructed from its ZMs [5]. Li *et al.* [4] show how to modify ZMs to obtain fully rotation invariance.

Definition 3. *A rotation invariant Zernike Moment A'_{nm} is defined as*

$$A'_{nm} = A_{nm} \exp(-im\theta_{n_0 1}), \tag{3}$$

where $V_{nm}^*(x,y)$ is the complex conjugate value of $V_{nm}(x,y)$ and $\theta_{n_0 1} = \arg(A_{n_0 1})$.

The value of n_0 is global, in the sense that it must be remain the same for all emblems in the database. Otherwise, comparisons will not be consistent.

2 Zernike Moments as Image Descriptors

In this section we present methods for using Zernike complex moments as image descriptors. Following Def. 2, given $n, m \in \mathbb{N}_0$ we say A_{nm} is valid if $m \leq n$ and $n - |m|$ is even.

Definition 4. *Given $N, M \in \mathbb{N}_0$ such that A_{NM} is valid, we define #dmz as*

$$\#\mathrm{dmz}(N, M) = \# \{A_{nm} \text{ is valid} \mid 0 \leq n \leq N, \ 0 \leq m \leq M\}$$

$$= \frac{\frac{N(N+1)}{2} - \lceil \frac{N}{2} \rceil}{2} + N + 1 - \frac{N - M}{2}. \tag{4}$$

We also denote $\#\mathrm{dmz}(N) = \#\mathrm{dmz}(N, N)$.

Definition 5. *Given an image I and $N \in \mathbb{N}_0$, we define a Zernike Moment Descriptor (ZMD) as a vector $\mathbf{Z}_{I,N} \in \mathbb{C}^{\#\mathrm{dmz}(N)}$ where*

$$(\forall n, m \leq N, A_{nm} \text{ is valid}) \ \ \mathbf{Z}_{I,N}(\#\mathrm{dmz}(n, m)) = A_{nm}$$

2.1 First Modification: Higher Order Weighting

Using higher order ZMs allows to obtain a more precise description of the image. In fact, the reconstruction of an image from approximately 50 ZMs is very precise [5]. However, higher order ZMs are more sensitive to noise and prone to numerical errors [3].

The key idea behind this modification is that higher order ZMs are useful for encoding fine details, but, at the same time, more importance must be assigned
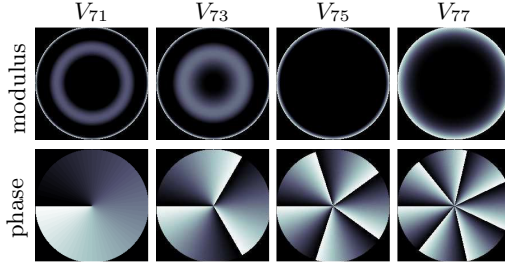
Fig. 2. Values taken by Zernike polynoms of order V_{7m}, $m = 1, 3, 5, 7$, on the unit circle

to lower order ZMs, which encode coarser details. In other words ZMs must be weighted, assigning more weight to lower order ZMs than to higher order ZMs.

Following Def. 4, valid ZMs describe a lower diagonal sparse matrix. By observing the values taken by Zernike polynoms (Def. 1) inside the unit circle in Fig. 2, we note that (1) the number of phase cycles depends only and is equal to m and (2) the number of weighted rings is equal to $\frac{n-m}{2} + 1$. In fact, given a fixed n, a larger value for m will result in a more sensitive ZMs. The direct implication is that when weighting ZMs we must not only take into account n but also m.

Definition 6. *We define a weighted Zernike Moment Descriptor (wZMD) as a vector* $\mathbf{Z}_{I,N}^w \in \mathbb{C}^{\#\mathrm{dmz}(N)}$ *such that*

$$(\forall n, m \leq N, A_{nm} \text{ is valid}) \quad \mathbf{Z}_{I,N}^w(k) = A_{nm} \cdot \exp(-k \cdot m \cdot |\sin(\theta_{n_01})|), \quad (5)$$

where $k = \#\mathrm{dmz}(n, m)$.

2.2   Second Modification: PCA ZMs

PCA is a classical technique that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables. PCA has been also extended to work with complex numbers [6]. A vector $v \sim \mathrm{N}(\mu, \Sigma)$, is transformed into a vector $\tilde{v} \sim \mathrm{N}(0, \mathrm{I})$ where I is the identity matrix, by using

$$\tilde{v} = (\Phi \, \Lambda^{-\frac{1}{2}})^T \, v \tag{6}$$

where Φ and Λ are the eigenvectors and eigenvalues matrices of Σ, respectively. The eigenvalues in Φ are sorted in decreasing order and the columns of Λ are arranged accordingly.

This process relies in an accurate estimation of Σ from the samples in our database. In our work, we compute Σ with the method proposed by Turk and Pentland [7]. Given a set of samples $\gamma_0, \gamma_2, \cdots, \gamma_n$, Σ is calculated as

$$\Sigma = \frac{1}{n} \sum_{i=0}^{n} \phi_i \, \phi_i^T \quad \text{where} \qquad \psi = \frac{1}{n} \sum_{i=0}^{n} \gamma_i \qquad \phi_i = \gamma_i - \psi$$

In our setting, the samples are simply the wZMD in the database.

Definition 7. *We define the PCA Zernike Moments Descriptors (PCA ZMD) as a vector* $\mathbf{Z}_{I,N}^{P} \in \mathbb{C}^{\#\mathrm{dmz}(N)}$ *where*

$$\mathbf{Z}_{I,N}^{wP} = (\Phi \, \Lambda^{-\frac{1}{2}})^{T} \, \mathbf{Z}_{I,N}^{w}. \tag{7}$$

We also note $\mathbf{Z}_{I,N}^{wP(n)}$ *the vector composed of the first n dimensions of* $\mathbf{Z}_{I,N}^{wP}$.

3   Metrics for Comparing Zernike Moments

Classically, two ZMD were compared by reconstructing both images and directly comparing them, i.e. using correlation or L_2 distance. Li *et al.* [4] proposed the use of a distance that takes into account phase and modulus.

Definition 8. *Let us define the* $D_\rho : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ *and* $D_\phi : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$

$$D_\rho(X,Y) = \left(\sum_{i=1}^{N} (|X_i| - |Y_i|)^2 \right)^{\frac{1}{2}} \quad D_\phi(X,Y) = \left(\sum_{i=1}^{N} (\arg X_i - \arg Y_i)^2 \right)^{\frac{1}{2}}$$

The distance $D_\alpha : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ *is defined as* $D_\alpha = \alpha D_\rho + (\alpha - 1)D_\phi$

Our tests with different values for α indicate, as suggested in Li *et al.* [4], that the best results are obtained with $\alpha = 0.5$. Although other distance functions have been proposed to compare ZMs [3,8], during experimentation we found better results using D_α and consequently we use it as our reference metric.

For the sake of computational efficiency, we define a combination of two simple metrics. This allow us to perform real-time and accurate queries. The first metric performs a fast pruning of the database, retrieving a small set of good look-alike candidate images

$$d_{\|\cdot\|}(u,v) = |(\|u\|_2 - \|v\|_2)|. \tag{8}$$

As the norm of all ZMD in the database can be precomputed, the k nearest ZMD can be retrieved in $\mathrm{O}(\max(\log(N),\ k))$. These neighboring ZMD can then be more carefully examined with

$$d_e(u,v) = \|u - v\|_2 \tag{9}$$

In synthesis, our algorithm will first prune the database using a threshold on $d_{\|\cdot\|}$ and then apply d_e on the smaller retrieved subset.

4   Results and Final Remarks

It is extremely hard to obtain real-world emblem datasets which are internally divided into similar classes. Therefore experimentation was performed on the following datasets:
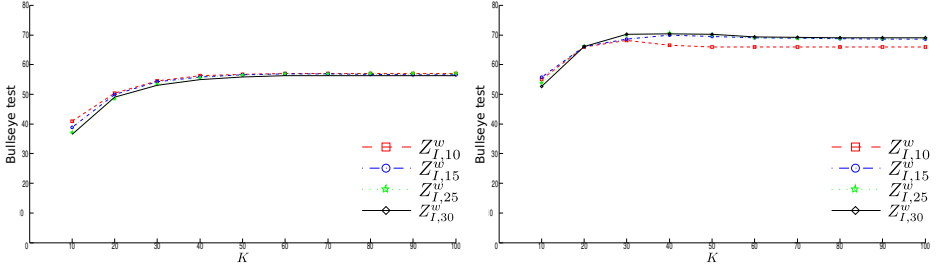
Fig. 3. Performance as the number of candidates K varies using wZMD of different dimensions. Left, on the `MPEG-7 CE Shape 1` dataset. Right, on the `ALOI` dataset.

MPEG-7 CE Shape 1: 70 conceptual classes of 20 silhouettes [9].

ALOI: 20 classes of 74 images [10]. Each class is a set of silhouettes of some object taken from different views.

MNIST: 10 classes of 1000 binarized handwritten digits [11].

The performance is measured using the so-called "bullseye test". Given a dataset S_n^c, where c is the number of classes in S and n the number of images per class, each image in S_n^c is used as a query and one counts the number of correct images (i.e. belonging to the same class than the query) in the top $2n$ matches. A perfect score is achieved when $n^2 c$ positive cases are found across all the dataset.

The main parameter of our method is the number of candidates K retrieved using $d_{\|\cdot\|}$ (Eq. 8). Defining the number of suitable look-alike image candidates will determine the retrieval performance. Fig. 3 depicts the change in score as K varies. The optimal value for K is tied to the characteristics of each dataset. However, in general, small datasets will require a larger K (around 50% of the dataset size), while K can be safely reduced to 10% on larger datasets.

Table 1 shows the performance of the methods dsicussed in this work. Following Li *et al.* [4], we use $D_{\alpha=0.5}$.

On the first part, when comparing performance using $D_{\alpha=0.5}$ versus $d_{\|\cdot\|} + d_e$, two phenomena can be observed. First, in all cases our approach outperforms the one by Li *et al.* [4]. Second, increasing the number dimensions in ZMD does not necessarily imply better results (e.g. on the MNIST dataset).

Table 1. Bulls-eye test scores. For each method best results are highlighted.

| Dataset	Metric	$\mathbf{Z}_{I,10}$	$\mathbf{Z}_{I,15}$	$\mathbf{Z}_{I,25}$	$\mathbf{Z}_{I,30}$	$\mathbf{Z}_{I,30}^{wP(20)}$	$\mathbf{Z}_{I,30}^{wP(15)}$	$\mathbf{Z}_{I,30}^{wP(10)}$	$\mathbf{Z}_{I,30}^{wP(5)}$
`MPEG-7`	$D_{\alpha=0.5}$	36.09	39.02	38.02	36.82	–	–	–	–
	$d_{\|\cdot\|} + d_e$	56.96	60.43	61.21	61.71	61.19	61.35	60.12	56.89
`ALOI`	$D_{\alpha=0.5}$	63.24	65.94	65.45	64.81	–	–	–	–
	$d_{\|\cdot\|} + d_e$	68.16	69.98	70.67	70.34	64.91	66.32	70.46	66.35
`MNIST`	$D_{\alpha=0.5}$	38.32	36.66	36.08	36.81	–	–	–	–
	$d_{\|\cdot\|} + d_e$	57.2	55.26	55.6	53.56	48.54	50.87	55.06	56.36

The second part of the table focuses on the result of using PCA. The performance of using PCA and $D_{\alpha=0.5}$ was very poor and results are not shown. Although dependent on the dataset, usually fixing $n = 15$ is sufficient. In our experiments with $\mathbf{Z}_{I,30}^{wP(n)}$, best results were obtained when n was chosen such that the cumulated standard desviation of the first n dimensions of $\mathbf{Z}_{I,N}^{wP}$ was around 90%. Note also that the best results obtained using PCA are almost as good as the best results obtained without it. However, comparing PCA ZMs is considerably faster because distance calculations involve fewer terms.

Fig. 4. Example images from the EMBLEMS dataset
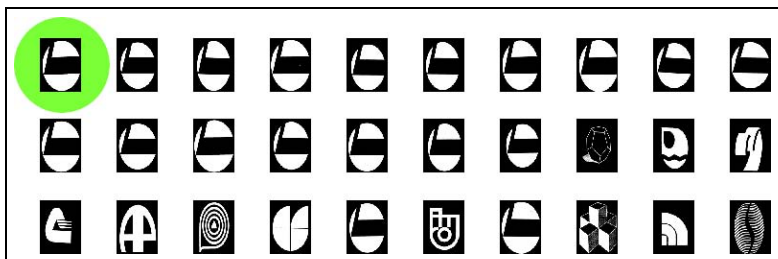
(a) using $D_{\alpha=0.5}$ and $\mathbf{Z}_{I,30}$
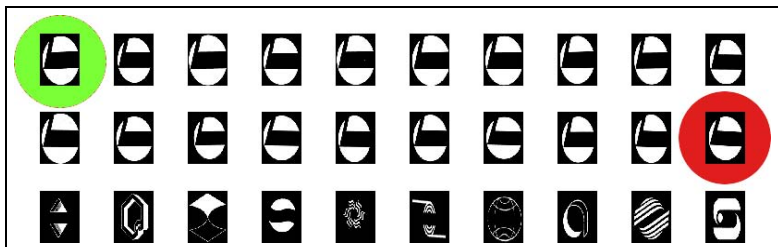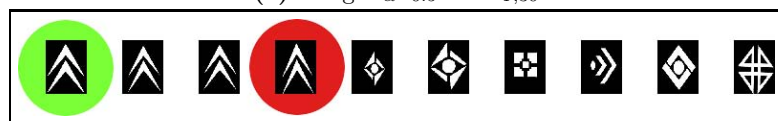
(b) using $d_{\|\cdot\|} + d_e$ and $\mathbf{Z}_{I,30}^{wP(15)}$

Fig. 5. EMBLEMS dataset results. In green (light grey) the query image, in red (dark grey) the farthest ranked "equal to the query" image.

Additionally, we tested the proposed method on a dataset of 8 thousand binary images of real trademark (Fig. 4) . This dataset includes several slightly distorted images for each emblem. Two query examples over the EMBLEMS dataset are shown in Fig. 5. Li *et al.*'s approach [4] retrieves the farthest "equal to the query" emblem on the 36th and 72th position in each test. All "equal to the query emblems" are ranked first by the proposed method. Note that some of the subsequent retrieved emblems are similar to the query.

To conclude, this paper introduces (1) a composite modification to Zernike Moments Descriptors and (2) a two-stage time efficient method for comparing ZMs. Experiments on diverse datasets show that the proposed approach significantly and consistently improves the accuracy of the content-based image retrieval based on ZMs. Being based on simpler distance calculations, considerable speed is gained. Such combination allows the proposed approach to outperform state-of-the-art content-based image retrieval methods based on ZMs.

References

1. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 509–522 (2002)
2. Kotoulas, L., Andreadis, I.: Image analysis using moments. In: Alexander Technological Educational Institute (ATEI) of Thessaloniki, Greece, pp. 360–364 (2005)
3. Khotanzad, A., Hong, Y.H.: Invariant image recognition by zernike moments. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 489–497 (1990)
4. Li, S., Lee, M.C., Pun, C.M.: Complex zernike moments features for shape-based image retrieval. Trans. Sys. Man Cyber. Part A 39, 227–237 (2009)
5. Amayeh, G.R., Erol, A., Bebis, G., Nicolescu, M.: Accurate and efficient computation of high order zernike moments. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS, vol. 3804, pp. 462–469. Springer, Heidelberg (2005)
6. Horel, J.D.: Complex principal component analysis: Theory and examples. Journal of Climate and Applied Meteorology 23, 1660–1673 (1984)
7. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586–591 (1991)
8. Revaud, J., Lavoué, G., Baskurt, A.: Improving zernike moments comparison for optimal similarity and rotation angle retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 627–636 (2009)
9. Latecki, L., Lakamper, R., Eckhardt, U.: Shape descriptors for non-rigid shapes with a single closed contour. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2000, pp. 424–429 (2000)
10. Geusebroek, J.M., Burghouts, G.J., Smeulders, A.W.M.: The amsterdam library of object images. Int. J. Comput. Vision 1, 103–112 (2005)
11. Lecun, Y., Cortes, C.: The mnist database of handwritten digits (2004), http://yann.lecun.com/exdb/mnist/