Generalization of Angular Radical Transform

Julien Ricard, David Coeurjolly, Atilla Baskurt

GENERALIZATION OF ANGULAR RADIAL TRANSFORM

Julien Ricard, David Coeurjolly, Atilla Baskurt

LIRIS, Laboratoire d'InfoRmatique en Image et Systèmes d'information
FRE 2672 CNRS, INSA Lyon, UCB Lyon 1, EC Lyon, Univ. Lyon 2
jricard@liris.cnrs.fr dcoeurjo@liris.cnrs.fr abaskurt@liris.cnrs.fr

ABSTRACT

Content based shape image retrieval is an important problem which gained the attention of the community. The challenge is to map the shape into compact and robust descriptor. This study presents a generalization of the Angular Radial Transform (ART). The ART, recommended by the MPEG-7 standard, is only limited to binary images and is not robust to perspective deformations. We propose two generalizations of the ART allowing to apply it to color images and to make it robust to all possible rotations and to perspective deformations.

1. INTRODUCTION

Content-based image retrieval has been a topic of intensive research in recent years, and particularly the development of efficient shape descriptors (SD). The MPEG-7 standard comity proposes a new region based shape descriptor, the Angular Radial Transform (ART) [4, 2]. This SD has many properties: compact size, robust to noise and scaling, invariant to rotation, ability to describe complex objects. In practical application, two main drawbacks of the ART have to be underlined. First, if we considered segmented planar objects from real images, we have to take into account unspecified rotations. As the basis functions are symmetrical in the angular direction, the invariance is inherent for planar rotations. Unspecified rotations induce a real deformation of the original shape due to the perspective projection into the image plan. In this study, we generalize the basis functions in order to ensure the robustness to all rotations and to perspective deformations. Secondly, the ART recommended by the MPEG-7 standard is limited to binary images. In order to index an object or a segmented region in a color or a grey level image, we propose an extension which combines the shape information and the spatial distribution of the dominant colors of the region and thus we define the Color Angular Radial Transform (CART). This paper is organized as follows: the ART is presented in section 2, the generalizations are explained in section 3 and the experiments and results are presented and discussed in the last section.

2. ANGULAR RADIAL TRANSFORM

Angular Radial Transform (ART) is a moment-based image description method adopted in MPEG-7 as a region-based shape descriptor [2]. It gives a compact and efficient way to express pixel distribution within a 2-D object region; it can describe both connected and disconnected region shapes. The ART is a complex orthogonal unitary transform defined on a unit disk that consists of the complete orthogonal sinusoidal basis functions in polar coordinates [4, 2]. The ART coefficients, F_{nm} of order n and m, are defined by:

$$F_{nm} = \int_0^{2\pi} \int_0^1 V_{nm}(\rho,\theta) f(\rho,\theta) \rho d\rho d\theta \qquad (1)$$

where $f(\rho,\theta)$ is an image function in polar coordinates and $v_{nm}(\rho,\theta)$ is the ART basis function that are separable along the angular and radial directions, that is,

$$V_{nm}(\rho,\theta) = A_m(\theta) R_n(\rho). \qquad (2)$$

In order to achieve rotation invariance, an exponential function is used for the angular basis function. The radial basis function is defined by a cosine function:

$$\left\{ \begin{array}{l} A_m(\theta) = \frac{1}{2\pi} \exp{(jm\theta)} \\ R_n(\rho) = \left\{ \begin{array}{ll} 1 & n = 0 \\ 2\cos(\pi n \rho) & n \neq 0 \end{array} \right. \end{array} \right. \qquad (3)$$

Real parts of basis functions are shown in Figure 1.

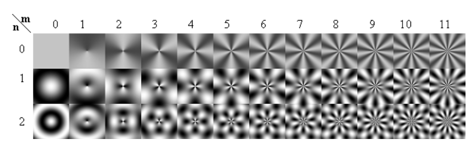

Fig. 1. Real parts of the ART basis functions.

The ART descriptor is defined as a set of normalized magnitudes of the set of ART coefficients. Rotational invariance is obtained by using the magnitude of the coefficients. In MPEG-7, twelve angular and three radial functions are used $(n < 3, m < 12)$ [4], these values will be used the rest of the work. For scale normalization, ART coefficients are divided by the magnitude of ART coefficient of order $n = 0$, $m = 0$. To decrease the descriptor size, quantification is applied to each coefficient using four bits per coefficient [4].As proposed in MPEG-7, the distance between two shapes described by the ART descriptor is calculated using L_1 norm:

$$d_{ART}(Q, I) = \sum_{i=0}^{n \cdot m} \|ART_Q[i] - ART_I[i]\| \qquad (4)$$

The subscript Q and I represent query image and image in the database, respectively and ART_I is the array of ART descriptor of the image I. Since the MPEG-7 standardization process and experiments which showed the efficiently of the method, a 2D indexing process used it in various domains.

3. GENERALIZATIONS OF ART

Two generalizations of the ART on 2D color image are proposed: first, we consider all rotations and perspective deformations. Then a generalization to color images is given.

3.1. Robust to rigid deformations

The goal of this generalization is to make the ART robust to every rotation and to perspective projections. A planar object in a natural scene can be viewed according to all orientations and carried by an unspecified plan. This highly probable situation will disturb the shape in the image and will prevent the identification. In Fig. 2, a plane object is seen with three angles of acquisition and corresponds to three different shapes projected on the same image plane. To make the ART descriptor robust to all possible rotations and to perspective projection it is necessary to generalize the ART transform with new basis functions (Fig. 2).

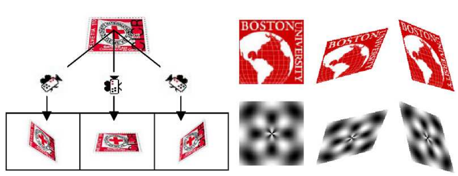

Fig. 2. Object seen according to various angles and example of projected basic functions on the support plane of the object to be identified.

In order to define the transformations undergone by an object during rotations and projection onto the image plane, we consider the transformation space given by the radial direction $\vec{\varsigma}$, the rotation angle ϕ and the perspective coefficient p. The first two parameters define the orientation of the object plan and the perspective coefficient defined the perspective deformation. Figure 3 shows the transformation.

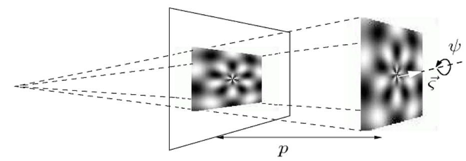

Fig. 3. The basis functions are projected on the image plan I according to k_ς, k_ϕ and k_p to obtain the projected basis functions.

This transformation space is sampled for each parameter according to k_ς, k_ϕ and k_p values. Hence we obtain a sampling of $K = k_\varsigma * k_\phi * k_p$ transformations. The basis functions are deformed in same way according to the transformations K. Each object is indexed with these K sets of projected basis functions. The number of projections is limited to have a reasonable computational cost. The values, $k_\varsigma = 12$, $k_\phi = 3$ and $k_p = 3$, are chosen in our experiments, because these values given the better ratio of cost to efficiency. In other words, we have $K = 108$ sets of coefficients to describe a shape. Hence we have to compute 108 similarity measure between a query object and a given object.
Indeed, the classical ART complexity is $\theta(n * m * N^2)$ because we compute $n*m$ basis functions values for the $N*N$ pixels of the image. The generalized ART creates K set of basis functions with a complexity $\theta(K * n * m * N^2)$. To make the retrieval process faster, we choose to inverse the indexation and retrieval processes. Without optimization, the indexation process computes the ART descriptor between the original object and the original basis function whereas the retrieval process computes the descriptor between the extracted object from a natural image and all the projected basis functions. In fact, the indexation process has computation cost K times less than the retrieval process. The retrieval is an online process and it is the longest phase. It is possible to inverse the two processes and to index the object of origin on the inverse projected basis functions and an extracted object only on the origin basis functions. This increases the cost of the offline indexing process but decreases the online retrieval process without modification of the description (Fig. 4 and Table 1). We can easily show that :

$$F_1(i', j')V_0(i', j') = F_0(i, j)V_{-1}(i, j) \ \begin{pmatrix} i \\ j \end{pmatrix} = T \begin{pmatrix} i \\ j \end{pmatrix} \qquad (5)$$

where $F_k(i, j)$ and $V_k(i, j)$ are the image pixel and the basis function pixel (i, j) by the T^k transform. We obtain the same descriptor by indexing the object of origin on the inverse transform basis functions and an extracted object only on the origin basis functions.
The shape similarity distance, knowing that each object is described by $K = k_\varsigma * k_\phi * k_p$ series of ART coefficients

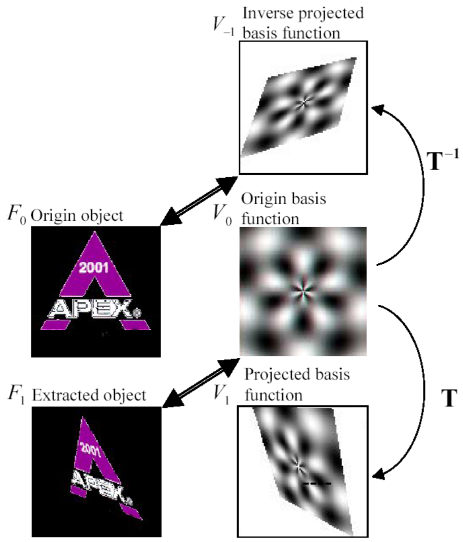

Fig. 4. Diagram of inverse indexation process.

	Original process	Optimized process
Online	K	1
Offline	1	K

Table 1. Number of online and offline ART descriptor computation during the original process and the optimized one.

created from the basis functions projected on K planes of projections, is achieved by computing a set of distances $d_{ART}(Q, I_j)$. For each value of j, the ART coefficients of Q computed on the original basis functions and those of I computed on the j^{th} projection of the basis functions are compared using (4). Then the shape distance between Q and I is given by:

$$d_{shape}(Q, I) = \min_{j \in K} \sum_{i=0}^{n \cdot m} \left\| ART_Q[i] - ART_I^j[i] \right\| \quad (6)$$

where Q is the ART coefficients of the key object and I_j is the coefficients of the I object, calculated on the j^{th} projection of the basis functions. The minimum is considered in order to take into account all the possible perspective views of the object. Note that other norms can also be considered.

3.2. Color ART

In the second generalization, we make the transformation tractable to color images. The value and the position of the dominant colors which compose the object must be taken into account in the shape retrieval system. The shape and the color of the objects are treated by two parallel studies: a study of the luminance and a study of the chrominance. We obtain two classifications of the image database which are combined to have a single one.

Study of the luminance: The basis functions of the Color ART are the same as those of the ART transform. The colored object is first represented in the perceptually uniform (L^*, a^*, b^*) color space [7]. The chrominance part of the information is not projected on the basis functions. Only the luminance component is considered to compute the ART coefficients. Note that MPEG-7 suggests the ART transform must be applied on binary objects but many systems [8, 6] used the luminance to compute the descriptor. The ART transform, applied to the luminance image of the object, gives a better result than the applied to the binary image [1]. The application of ART on the luminance component allows taking into account the internal variations of the objects (contours, holes, texture...).

Study of the chrominance: This study has the aim to classify the image database according to a color criterion. The color descriptions of the object are made using the dominant color analysis [3]. The object colors are described by their dominant colors, (DC_i, p_i, σ_i) in a Lab color space, where DC_i is the color vector (L_i, a_i, b_i), pi and σ_i are the percentage and the variance corresponding to the distribution of the i^{th} dominant color. The value of DC_i are supplied directly by the segmentation process [3], and the corresponding variance and percentage are computed on the object. We define a color histogram as the sum of the dominant color contributions, as follows:

$$H(x) = \sum_i \frac{p_i}{\sigma_i \sqrt{2\pi}} \exp - \frac{(x - DC_i)^2}{2\sigma_i^2} \quad (7)$$

where the value x corresponds to the bin of the histogram. The color similarity measure is computed between the histograms of all dominant colors [3]. A Kullback distance is thus performed in its symmetric form [5] to measure the similarity between two generated distributions H_Q and H_I. The color distance between the query images Q and a database image I is then given by:

$$d_{color}(Q, I) = \sum_{n=1}^{N} \sum_{m=1}^{3} (q_{nm} - i_{nm}) log_2 \left(\frac{q_{nm} + 1}{i_{nm} + 1} \right) \quad (8)$$

where N is the number of histogram bins (256), M is the number of color components ($M = 3$ for Lab space), q_{nm} is the percentage of the m^{th} component of the $n^t h$ color in Q and i^{nm} is the percentage of the m^{th} component of the n^{th} color in I.

3.3. Combining features for matching

To estimate the similarity between two images, we have to evaluate the similarities between their descriptors. The color distribution and the generalized ART to projection and to rotation distribution are mixed. This transformation is called the generalized color ART (GCART). A global similarity function D is computed as a weighted sum of the similarities:

$$D = \alpha.d_{color} + (1 - \alpha)d_{shape} \qquad (9)$$

where α is the weight controlling the sum. It is fixed iteratively by the user according to his request or evaluated automatically by the system when the image database classes are known.

4. EXPERIMENTS

First test compares the ART on the luminance and the generalization to projection. A test database was created, it contains 1813 images of 37 trademark images disturbed according to 49 random perspective projection with illuminating variations. The figure 6.a shows the Recall and Precision values. To evaluate the color ART, we have set up an application allowing to identify an object extracted from an image. The application can be split into two successive stages: the indexation and the retrieval steps (fig. 5).

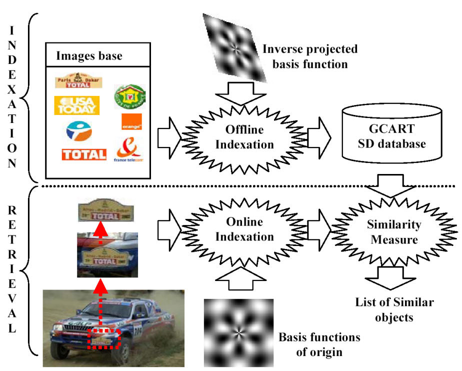

Fig. 5. General diagram of the application.

To evaluate the properties of the GCART and the retrieval process, 50 objects was extracted of the images and we evaluate the rank where one finds the origin trademark. The figure 6.b show the Recall values for the luminance, the color and the GCART studies. The GCART study gives the origin trademark at the first rank in 55%, against 38% for the luminance and 6 % for the color. At the rank 10, the origin trademark is found in 95% of the cases, whereas the luminance and the color study have found the origin object only in 65%, respectively 41%, of the cases.

5. CONCLUSION

The generalizations of the ART, to perspective projection and to color, increase the numbers of ART uses and the definition domains but in keeping the discriminating capacities. The optimization of the process made possible to keep a light online process and a quick answer.

(a) (b)

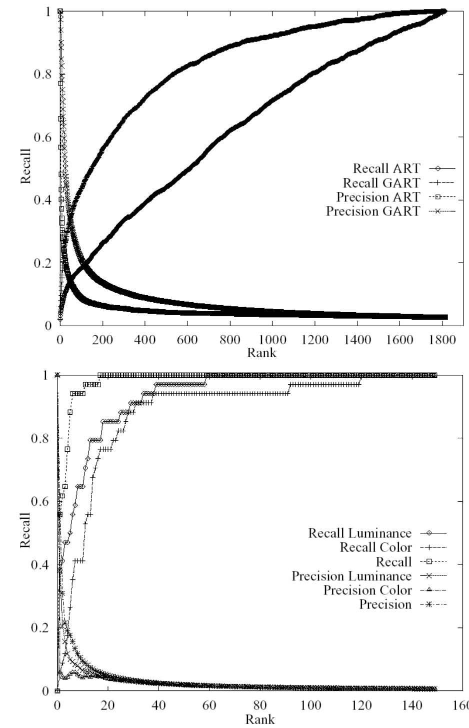

Fig. 6. Recall and precision values: (a) ART and generalized to perspective projection ART, (b) luminance, color and CART approach.

6. REFERENCES

[1] M. Akcay, A. Baskurt, and B. Sankur. Measuring similarity between color image regions. In *EUSIPCO*, volume 1, pages 115–118, Toulouse, France, Sept. 2002.

[2] M. Bober. Mpeg-7 visual shape descriptors. *IEEE Trans. Circuits Syst. Video Technol*, 1(6), June 2001.

[3] K. Idrissi, G. Lavoue, J. Ricard, and A. Baskurt. Object of interest based visual navigation, retrieval and semantic content identification system. *Computer Vision On Image Understanding*, 94(1-3), 2004.

[4] W.-Y. Kim and Y.-S. Kim. A new region-based shape descriptor. Technical report, Pisa, Dec. 1999.

[5] S. Kullback. *Information theory and statistics*. John Wiley and Sons, New York, 1959.

[6] J. Laaksonen, J. Koskela, S. P. Laakso, and E. Oja. PicSOM: Content-based image retrieval with self-organizing maps. *Pattern Recognition Letters*, 21(13-14), Dec. 2000.

[7] K. Nassau. *Color for Science, Art, and Technology*. Elsevier Science, Amsterdam, 1998.

[8] J. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 2001.