Object descriptors invariant to affine distortions

A Kadyrov, M Petrou,
School of Electronics, Computing and Mathematics,
University of Surrey, Guildford GU2 7XH, United Kingdom

Abstract

The Trace transform is a generalisation of the Radon transform that allows us to construct image features that are invariant to a chosen group of image transformations. In this paper we propose functionals which can be computed from the image function, and which can be used to describe objects in a way that is invariant to the group of affine transforms. We demonstrate the usefulness of the constructed image descriptors in retrieving images from an image database and compare it with the moments based method.

1 Introduction

Image analysis requires the use of image features that capture the characteristics of the objects depicted so that they are invariant to the way the objects are presented in the image. Historically, the process of extracting image features has been anthropocentric: the features calculated are defined in a way that captures the attributes the human vision system would recognise in the image. Thus, features like compactness, elongatedness, brightness etc are features which have some physical and perceptual meaning. It is not however necessary for the features to have a meaning to the human perception in order to characterise well an object. Indeed, features which broaden the human perception may prove to be more appropriate for the characterisation of complex structures, like the objects often one wishes to identify in an image. In the next section we shall present the background to a new theory that has been proposed recently for constructing image features that have desired properties. In section 3 we shall present the modification of this theory for the construction of features invariant to affine distortions. In section 4 we shall present some experimental results and we shall conclude in section 5.

2 Feature construction from the Trace Transform

Let us imagine an image criss-crossed by all possible lines one can draw on it. Each line is characterised by parameters ϕ and p defined in figure 1. The Radon transform of the image is a 2D representation of the image in coordinates ϕ and p, with the value of the integral of the image computed along the corresponding line placed at cell (ϕ, p). It is well known that an image can be fully reconstructed from its Radon transform. This is the basis of computerised tomography. The Trace transform calculates a functional T over parameter t along line (ϕ, p), which is not necessarily the integral of the image. One then calculates another functional, P, along the columns of the Trace transform, ie over parameter p, and finally a functional Φ over the string of numbers created this way, ie over parameter ϕ.

BMVC 2001 doi:10.5244/C.15.41

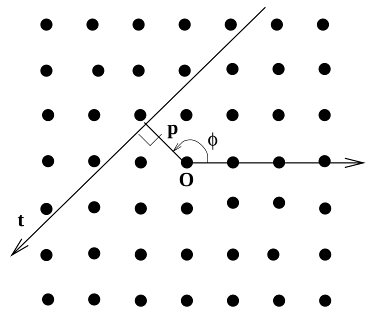

Figure 1: Definition of the parameters of an image tracing line

The result is a single number. With an appropriate choice of the three functionals T, P and Φ, one can make this number to be invariant to a certain group of transformations. In [4] we presented the functionals one should use in order to produce features invariant to rotation, translation and scaling. In [5] we presented some functionals one should use, so that the produced features are invariant to affine distortions. In this paper we are not going to attempt to characterise an object by a single number produced by the cascaded application of the three carefully chosen functional T, P and Φ, but instead we are going to use only the first two functionals. Using only functionals T and P one after the other will allow us to characterise an object by a string of numbers, something like the object "signature". We call this "signature" the "circus" function, because it is a function of the angle parameter ϕ. In the next section we shall show how these signatures can be produced, so that if they correspond to two objects that are affine distortions of each other, they differ only by a shift.

3 Affine distortions

The basic idea behind the Trace transform-based feature construction is the observation that lines remain unchanged when the image undergoes linear or affine distortions. What changes however, are the parameters in terms of which each line is defined and the parameter defined along each line. If the image has undergone an affine distortion defined by matrix A and translation vector $\mathbf{s}$, then the parameters of a line in the distorted image t_{new}, p_{new} and ϕ_{new}, are given in terms of the parameters of the same line in the undistorted image by the following equations:

$$\begin{aligned} t_{new} &= k(\phi)t + b_1(\phi)p + b_2(\phi) \\ p_{new} &= c(\phi)p + d(\phi) \\ \phi_{new} &= \rho(\phi) \end{aligned}$$

where $k(\phi)$, $b_1(\phi)$, $b_2(\phi)$, $c(\phi)$, $d(\phi)$ and $\rho(\phi)$ are some complicated functions that can be expressed in terms of the elements of matrix A and vector $\mathbf{s}$.

Before we proceed, we must define the type of functional we shall use, namely *invariant to displacement* functionals.

A functional Ξ of a function $\xi(x)$ is *invariant* if

$$\Xi(\xi(x+b)) = \Xi(\xi(x)), \quad \forall b \in \Re \qquad (I_1)$$

An invariant functional may also have the following additional properties:

- Scaling the independent variable by a, scales the result by a^{κ_Ξ}:

$$\Xi(\xi(ax)) = a^{\kappa_\Xi}\Xi(\xi(x)), \quad \forall a > 0 \qquad (i_1)$$

- Scaling the function by c, scales the result by c^{λ_Ξ}:

$$\Xi(c\xi(x)) = c^{\lambda_\Xi}\Xi(\xi(x)), \quad \forall c > 0 \qquad (i_2)$$

In the above expressions κ_Ξ and λ_Ξ are some numbers which characterise functional Ξ.

If we choose functionals T and P that obey the above properties, and apply them to the image function, we shall derive the "circus" function $h(\phi)$ that characterises each image. It is to this "circus" function that in [5] we apply the final functional Φ in order to produce the invariant feature which is just a number. Here, however, we shall use function $h(\phi)$ itself as the signature of the object.

It can be shown that the two "circus" functions associated with the two versions of the same image that are affine distortions of each other, are related by the following expression:

$$h_1(\phi) = k(\phi)^{\lambda_T\kappa_T-\kappa_P}|A^{-1}|^{\kappa_P}h_2(\rho(\phi)) \qquad (1)$$

Next, we defined the concept of the "associated circus" $h_a(\phi)$ to a given circus $h(\phi)$:

$$h_a(\phi) \equiv |h(\phi)|^{-1/(\lambda_T\kappa_T-\kappa_P)} \qquad (2)$$

Note that as we assume that we have already chosen functionals T and P, the values of λ_T, κ_T and κ_P are known.

If $\lambda_T\kappa_T - \kappa_P = 0$, we define the "associated circus" by

$$h_a(\phi) \equiv \sqrt{\left|\frac{dh(\phi)}{d\phi}\right|} \qquad (3)$$

These associated circuses can be further normalised in a way described in detail in [5] to define the "normalised associated circus functions". Let us consider two of them, $h_{n1}(\phi)$ and $h_{n2}(\phi)$, which correspond to the original and the affinely distorted image respectively. It can be shown that they are associated by the following equation:

$$h_{n1}(\phi) = |A^{-1}|^{(\kappa_P-1)/2}h_{n2}(\phi+\psi) \qquad (4)$$

If $\lambda_T\kappa_T - \kappa_P = 0$, this expression takes the form $h_{n1}(\phi) = h_{n2}(\phi+\psi)$.

Here ψ is some angle we are not interested in. It can be used to extract the parameters of the affine transform, but we are not going to do this here.

Now we have defined these normalised associated circus functions, the problem of identifying an object is reduced to that of comparing two strings of numbers that are shifted and possibly scaled versions of each other.

4 Experimental results

To demonstrate our ideas we consider a database consisting of 60 images of fish. The most commonly used method for affine invariant description of images is that of Flusser and Suk [1, 2, 3]. It is based on the calculation of moments of the object to be described and the combination of these moments to form affine invariant descriptors. Affine invariant moments have been widely used in many applications, eg [6]. Throughout the experiments we shall present, we shall compare our method with the moments method. The proposers of the moments method have suggested the use of either 4 or 6 invariant moments for the description of objects [3]. We run all our experiments for both cases. The use of 6 moments produced better results than the use of 4 moments, so we only present those results here. As a measure of similarity between two images we use the sum of the absolute differences between the six corresponding invariant moments. Before this sum is taken, each moment is normalised by being divided with an appropriate number so that all invariant features combined are of the same order of magnitude. The six affine invariants used are defined below:

$$
\begin{aligned}
I_1 &= \frac{1}{\mu_{00}^4}(\mu_{20}\mu_{02} - \mu_{11}^2) \\
I_2 &= \frac{1}{\mu_{00}^{10}}(\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} + 4\mu_{30}\mu_{12}^3 + 4\mu_{03}\mu_{21}^3 - 3\mu_{21}^2\mu_{12}^2) \\
I_3 &= \frac{1}{\mu_{00}^7}(\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)) \\
I_4 &= \frac{1}{\mu_{00}^{11}}(\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} - 6\mu_{20}^2\mu_{21}\mu_{02}\mu_{03} + 9\mu_{20}^2\mu_{02}\mu_{12}^2 + 12\mu_{20}\mu_{11}^2\mu_{03}\mu_{21} \\
&\quad +6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} - 8\mu_{11}^3\mu_{03}\mu_{30} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} \\
&\quad +9\mu_{20}\mu_{02}^2\mu_{21}^2 + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} - 6\mu_{11}^2\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2) \\
I_5 &= \frac{1}{\mu_{00}^6}(\mu_{40}\mu_{04} - 4\mu_{31}\mu_{13} + 3\mu_{22}^2) \\
I_6 &= \frac{1}{\mu_{00}^9}(\mu_{40}\mu_{04}\mu_{22} + 2\mu_{31}\mu_{22}\mu_{13} - \mu_{40}\mu_{13}^2 - \mu_{04}\mu_{31}^2 - \mu_{22}^3)
\end{aligned}
$$

where μ_{pq} is defined by

$$\mu_{pq} = \int\int_{object} f(x,y)(x-x_t)^p(y-y_t)^q dxdy \qquad (5)$$

with $f(x,y)$ being the grey level image function, and (x_t, y_t) the centre of mass of the object.

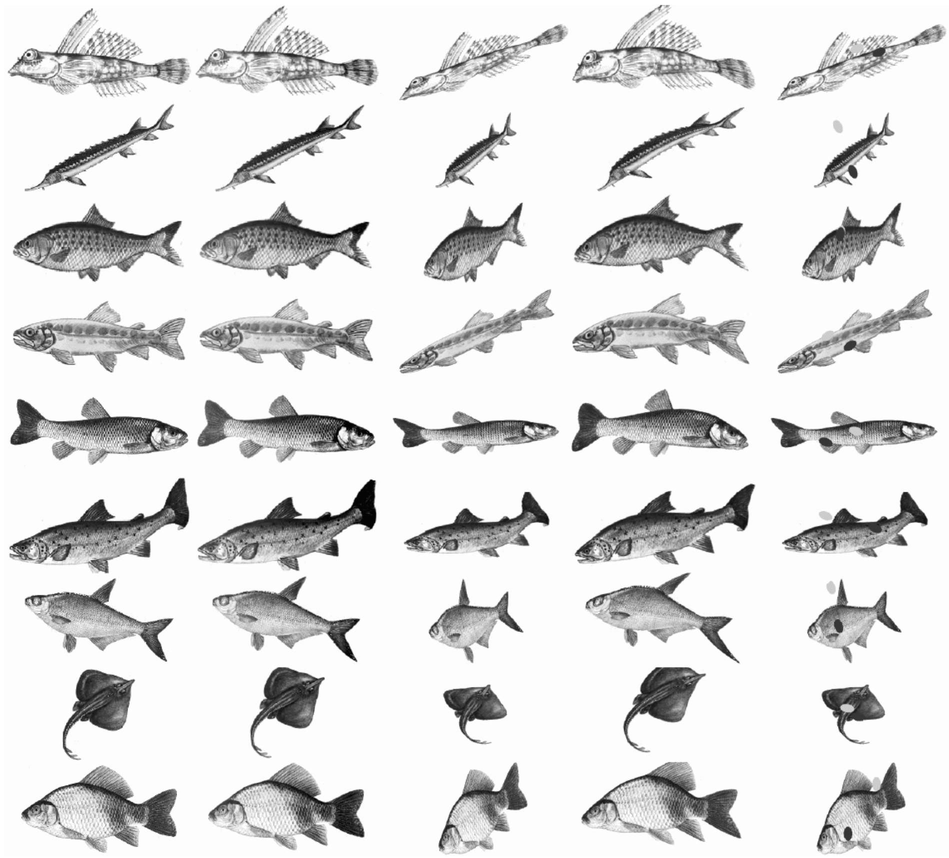

Figure 2: First column: example images in the database. Second column: images distorted by spatially variant illumination; the illumination varies linearly from left to right, being a multiplicative factor of 0.8 at one end and 1.2 at the other. Third column: images distorted by affine distortion. Fourth column: Images distorted by nonlinear distortion with parameter (maximum shift along the middle horizontal line and along the middle vertical line) 35 pixels; shifting decreases to 0 as the inverse square distance from the centre as the edges are approached. Last column: images distorted by affine distortion and two blobs of radius 15 pixels and randomly chosen grey value added at random positions in the image.

For the trace transform method, each image, being of size 200×400, was traced by lines two pixels apart, ie the values of parameter p for two successive parallel lines used differed by 2. For each value of p 80 different orientations were used, ie the orientations of the lines with the same p differed by 4.5^o Each line was sampled with points two pixels apart, ie parameter t took discrete values with step equal to 2 inter-pixel distances.

We used four different trace functionals T computed along each tracing line:

- T_1: $\int f(t)dt$, where $f(t)$ is the value of the image function along the tracing line.
- T_2: $\left[\int f(t)^4 dt\right]^{1/4}$

- T_3: The weighted median value of the sample points $f(t_n)$ along the line[1].
- T_4: $\left[\int |FT(f)|^4 d\omega\right]^{1/4}$, where $FT(f)$ means the Fourier Transform of $f(t)$.

These four functionals were combined with three different P functionals to produce a total of 12 circus functions to characterise an image:

- P_1: $\left[\int f(p)^{0.5} dp\right]^2$. Here $f(p)$ is the function made up from the values of the trace tranform along a column, ie each $f(p)$ is characterised by a different value of angle ϕ which is the parameter measured along the horizontal axis of the trace transform.
- P_2: $\int \left|\frac{df(p)}{dp}\right| dt$
- P_3: The weighted median value of the sample points $f(p_n)$ along each column of the trace transform. The weights used are $\left|f(p_{n+1}) - f(p_{n-1})\right|$.

To compare two circuses, one from a reference image and one from the query image, we compute their correlation coefficient for all possible shifts via the fast Fourier transform. We choose the maximum value over all shifts. Two circuses are most similar when their correlation is maximum. To express this as a distance that is smallest for most similarity, we take the inverse cosine of the maximum value of the correlation coefficient. This is equivalent to measuring the distance between two vectors by measuring the angle between them, and it is 0 when we have maximum correlation. This way we produce 12 different numbers when we compare two images. We use as measure of similarity the sum of the smallest 9 of these numbers.

The two methods were compared on the way they respond to different types of distortion present in the test image when we are trying to identify it in the database. Figure 2 shows the image deformations used. In the first column we present some entries in the database. In the second column we have imposed a small illumination distortion to the images by multiplying their gray values with a linear illumination function that varies from left to right and it has value 0.8 on the left and 1.2 on the right. This is a very small non-uniform illumination effect, nevertheless we shall show that the moments method breaks down in this case, while the proposed method does not. In the third column of figure 2 we have applied just affine distortions to the images. In the fourth column we have applied a non-linear distortion, as follows: All pixels were shifted first from left to right by a shift that had a parabolic profile which was maximum in the middle and 0 along the top and bottom edges. Then all pixels were shifted from top to bottom, also by a parabolic profile and it was maximum in the middle and 0 along the left and right edges. The maximum shift in the middle is the parameter of this distortion, and in this particular example it was 35 pixels. In the last column of figure 2 we present the deformation consisting of an affine distortion and the addition of two random blobs of uniform grayness in the image.

[1] If a sample has weight w_n, its value is repeated w_n times when computing the median. The weights used here were $w_n = f(t_n)$.

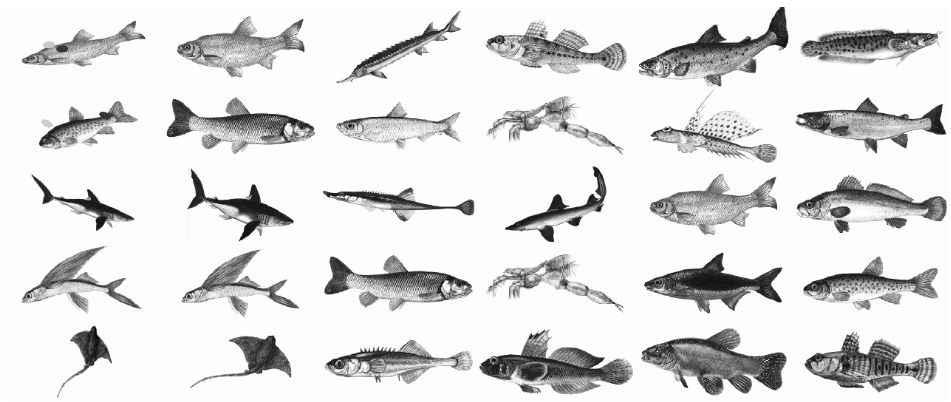

Figure 3: First column: example query images. Subsequent columns, retrieved images in order of similarity, when the trace transform method was used.

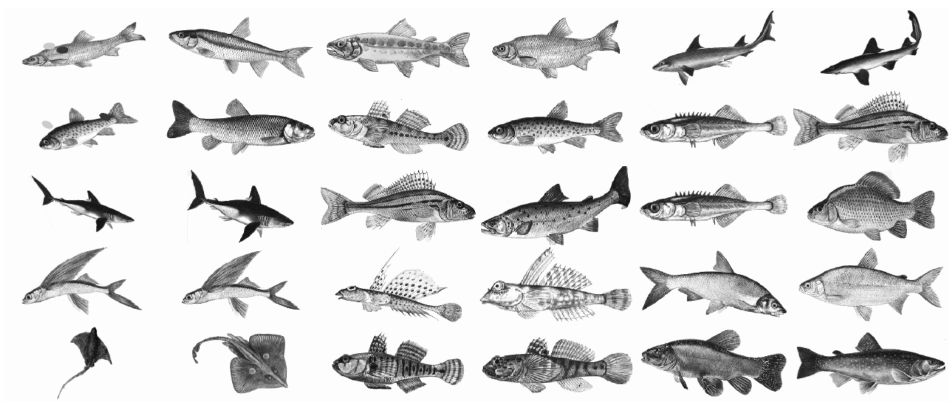

Figure 4: First column: example query images. Subsequent columns, retrieved images in order of similarity, when the moments-based method was used.

Figure 3 shows a few examples of answers received when querying the database: The first column is the query image. The subsequent columns show the five most similar fish retrieved from the database in order of similarity attached to them. Figure 4 shows the result obtained for the same query images when the moments method was used.

The summary of the results of all our experiments are presented in tables 1–7. In all these tables, we put in the first column the number of times the query fish was retrieved as the first choice. In columns 2–5 we put the number of times it was retrieved in the 2nd–5th position. In the last column we put the number of times the query fish was retrieved in the 6th position and beyond.

Table 1 shows the results when only affine distortion was present, ie the case when the assumptions on which both approaches are based are ideally satisfied.

Table 2 shows the results for added Gaussian noise to the affinely distorted image (the whole image, not just the object) with various standard deviations.

Method	Affine					
Trace	57	0	0	0	1	2
Moments	51	4	2	0	0	3

Table 1: Affine distortion with matrix $(random\_rotation) \times diag(1, 0.6) \times (random\_rotation)$

Method	Affine+Gaussian noise					
	$\sigma = 2$					
Trace	27	5	6	2	1	19
Moments	3	2	2	3	2	48
	$\sigma = 5$					
Trace	21	5	3	2	2	27
Moments	2	1	3	1	2	51
	$\sigma = 10$					
Trace	15	3	2	4	2	34
Moments	2	1	2	3	0	52

Table 2: Affine distortion with matrix $(random\_rotation) \times diag(1, 0.6) \times (random\_rotation)$ plus Gaussian noise added to the whole image

Table 3 shows the results when the distortion of the query images is not affine but non-linear with various parameter values.

Table 4 shows the results when the distortion of the query images is affine plus non-linear with parameter 20.

Table 5 shows the results when the distortion of the query images is non-linear with parameter 10 and there is non-uniform illumination with factor 0.8 on the left and 1.2 on the right.

Table 6 shows the results when the distortion of the query images is non-linear with parameter 10, there is non-uniform illumination with factor 0.8 on the left and 1.2 on the right, plus affine distortion.

Finally table 7 shows the results when the distortion of the query images is affine plus 2 blobs of radius 15 pixels each have been added at random positions in the image.

5 Conclusions

From all the results presented in tables 1–7, we can see that the method based on invariant moments and the method based on the trace transform have comparative performance when the assumptions on which they are based are fulfilled, ie when the query images are only affinely distorted. However, if the images are distorted in a different way, the trace transform method is much more robust. The moments based method is particularly sensitive to distortions that affect the “centre of gravity” of the image, like non-uniform illumination and the addition of blobs which could represent clutter or local damage of the image.

Method	Non-linear					
	parameter=10					
Trace	58	0	1	0	0	1
Moments	57	1	1	0	0	1
	parameter=20					
Trace	35	2	2	1	2	18
Moments	22	10	9	5	5	9
	parameter=35					
Trace	9	4	2	1	3	41
Moments	11	2	11	3	3	30

Table 3: Non-linear image distortion for various parameter values indicating maximum shift

Method	Affine+non-linear					
Trace	27	2	3	0	1	27
Moments	24	11	3	5	4	13

Table 4: Affine distortion with matrix $(random\_rotation) \times diag(1, 0.6) \times (random\_rotation)$ plus non-linear distortion with parameter 20

Method	Non-linear +Illum					
Trace	57	0	2	1	0	0
Moments	35	10	4	1	3	7

Table 5: Non-linear distortion with parameter value 10 and non-uniform illumination with value 0.8 on the left side of the image and 1.2 on the right side

Method	Non-linear+Illum + affine					
Trace	40	7	4	3	1	5
Moments	27	11	4	3	2	13

Table 6: Non-linear distortion with parameter value 10 plus non-uniform illumination with multiplicative value 0.8 on the left side of the image and 1.2 on the right plus affine distortion

Method	Affine+Two blobs					
Trace	52	1	0	0	2	5
Moments	30	9	9	1	3	8

Table 7: Affine distortion plus two random blobs of radius 15 pixels

Acknowledgements: This work was supported by an EPSRC grant GR/M88600.

References

[1] J Flusser and T Suk, 1993. "Pattern Recognition by Affine Moment Invariants". Pattern Recognition, Vol 26, pp 167–174.

[2] J Flusser and T Suk, 1994. "Affine moment invariants: a new tool for character recognition". Pattern Recognition Letters 15, pp 433–436.

[3] J Flusser and T Suk, 1994. "A Moment-Based Approach to Registration of Images with Affine Geometric Distortion". IEEE Transactions on Geoscience and Remote Sensing, Vol 32, pp 382–387.

[4] A Kadyrov and M Petrou, 2001. "The Trace transform and its applications", accepted by PAMI.

[5] M Petrou and A Kadyrov, 2001. "Features invariant to affine distortions from the trace transform". ICIP 2001, Thessaloniki, Greece, October 2001, accepted.

[6] T Tuytelaars and L Van Gool, 2000. "Wide base line stereo matching based on local affine invariant regions". BMVC2000, pp 412–425.