Image Indexing and Retrieval Using Image-Derived, Geometrically and Illumination Invariant Features [*]

Ronald Alferez and Yuan-Fang Wang
Department of Computer Science
University of California at Santa Barbara
E-mail: {ronald,yfwang}@cs.ucsb.edu

Abstract

In this paper, we propose novel image-derived features for image indexing and retrieval in digital library applications. The new features capture the intrinsic geometry and color properties of an imaged object. That is, these features are insensitive to the change of an object's appearance due to incidental environmental factors such as rigid motion, affine shape deformation, changes of parameterization and scene illumination, and view point change. Recording both intrinsic shape and color information of an imaged object, these features allow the similarity between two objects to be accurately measured. Hence, they can be used to index objects of the same class with similar appearance (e.g., different airplanes) instead of objects of different classes (e.g., airplanes vs. automobiles).

Keywords: Digital Libraries, Image Content Analysis, Content-based Indexing and Retrieval, Invariant Features

1 Introduction

Searching an image library can be difficult and the main difficulty lies in designing powerful features to represent images in the library. Manual annotation works only for small library collections and do not scale to large image ensembles. Furthermore, it is not clear what aspects of an image would be of interest to a particular user and the possible image descriptions can vary. Many studies are geared toward solving this problem and a few are even considered to be performance standards [4, 5, 6]. But the difficulty in designing powerful features are further aggravated if (1) query images and images in the database can appear different due to non-essential environment changes, such as rigid motion, affine shape deformation, change in illumination and perspective, and (2) objects in the database are similar in appearance, re-

quiring highly discriminating descriptors to tell them apart. Many current systems for image indexing do not address the above two difficulties satisfactorily.

In this paper, we propose novel image-derived features for image indexing and retrieval to address the two difficulties mentioned above. More specifically, we design features that capture the essential traits of an image and are insensitive to environmental changes. Invariant features form a compact, intrinsic description of an imaged object and possess high discriminative power. Hence, they can be used in an image retrieval system that require high accuracy and insensitivity to environmental factors.

Although segmentation (contour extraction) is not addressed, our strategy still has many practical applications, particularly when there is absolute control of the image database (*e.g.*, when the database is a collection of imaged objects photographed with an uncluttered background, such as catalogs), and the object of interest in the query image is pinpointed (or drawn) by a human. There are many image databases that fit this type of scenario, such as FishBase [1] which contains a large collection of fish images photographed under standard pose and lighting against plain background. (However, only query by fish name is currently supported). Other applications include indexing leaves and flowers in a botanic image database, and machine tools in a retail catalog. Note that all these applications involve database objects of similar appearance and query images which may differ from those in the database due to non-essential environmental changes. Hence, a highly discriminative, invariant descriptor will be ideal for such applications.

The new invariant features analyze the shape of the object's contour as well as the color characteristics of the enclosed area. The analysis involves projecting the shape or color information onto one of many basis functions of finite, local support (e.g., splines, short-

[1]http://ibs.uel.ac.uk/ibs/sp2000/fishbase

time Fourier analysis and wavelets). The projection coefficients, in general, are sensitive to changes induced by rigid motion, shape deformation, and change in illumination and perspective. Our analysis, as presented in the next section, derives expressions by massaging these sets of projection coefficients to cancel out these environmental factors to achieve invariance of the descriptors.

Furthermore, the invariant features enable a quasi-localized, hierarchical shape and color analysis, which allows for the examination of information at multiple resolution scales. The result is an invariant framework which is more flexible and tolerant to a relatively large degree of noise.

We also introduce the use of rational basis functions to facilitate the analysis of invariants under perspective projection. Rational basis functions, such as NURBS, have been widely used in the computer graphics community. However, their usage in perspective invariants is novel.

2 Technical Rationale

We will illustrate the design of invariant image features using a specific scenario where invariants for curves are sought. These directly apply to the silhouette (contour) of objects in the image database. For illumination invariants, the same technique applies by linearizing internal regions by a characteristic sampling curve and computing invariant color signatures along the characteristic curve. The particular basis functions we use in the illustration are the wavelet bases and spline functions. However, the same framework can be easily extended to other bases such as the short-time Fourier analysis, and for 3D surfaces.

Affine Invariant Parameterization When defining parameterized curves $\mathbf{c}(t) = [x(t), y(t)]^T$, most prefer to use the intrinsic arc length parameter, t, because of its simplicity. Intrinsic arc length transforms linearly (or it is invariant) under rigid-body motion, but does not do so under an affine deformation. A more suitable parameterization is thus required. Two common parameterizations are described in [2] which are transformed linearly under an affine deformation. One of them is the *enclosed area parameter*, $\sigma = \frac{1}{2} \int_a^b |x\dot{y} - y\dot{x}| \, dt$, which is the area of the triangular region enclosed by the two line segments from the centroid to two contour points a and b. The other, called *affine arc length*, is defined as: $\tau = \int_a^b \sqrt[3]{\dot{x}\ddot{y} - \ddot{x}\dot{y}} \, dt$ where $\dot{x}, \dot{y}$ are the first and $\ddot{x}, \ddot{y}$ are the second derivatives with respect to any parameter t (possibly the intrinsic arc length).

To properly align two contours for comparison under affine transform, it is also important to choose the same traversal starting point and the direction of traversal. It might appear that a one-point correspondence (*i.e.*, the starting point), and knowledge of the direction of the contour (*i.e.*, clockwise or counterclockwise) is required. However, it can easily be shown that the invariant signatures of two contours, differing only by the starting point, is just a phase-shifted version of each other. Similarly, the invariant signature of a contour parameterized in opposing directions are just flipped, mirror images of each other. Hence, a match can be chosen that maximizes the cross correlation between the two signatures.

Allowing an arbitrary change of origin and traversal direction, together with the use of an affine invariant parameterization, implies that *no point correspondence is required* when computing the affine invariants of an object's contour.

Rigid-Body Motion and Affine Transform Consider a 2D curve, $\mathbf{c}(t) = [x(t), y(t)]^T$ where t denotes a parameterization which is invariant under affine transform (as described above), and its expansion onto the wavelet basis $\psi_{a,b} = \frac{1}{\sqrt{a}} g(\frac{t-b}{a})$ (where $g(t)$ is the mother wavelet [3]) as $\mathbf{u}_{a,b} = \int \mathbf{c}(t) \psi_{a,b} dt$. If the curve is allowed a general affine transform with the possibility of being traversed from a different starting point and along an opposite direction, then the transformed curve is denoted by: $\mathbf{c}'(t) = \mathbf{mc}(t') + \mathbf{t} = \mathbf{mc}(\pm t + t_0) + \mathbf{t}$, where $\mathbf{m}$ is any nonsingular 2×2 matrix, $\mathbf{t}$ represents the translational motion, t_0 represents a change of the origin in traversal, and $\pm$ represents the possibility of traversing the curve either counterclockwise or clockwise. It follows that:

$$
\begin{aligned}
\mathbf{u}'_{a,b} &= \int \mathbf{c}' \psi_{a,b} dt \\
&= \int (\mathbf{mc}(\pm t + t_0) + \mathbf{t}) \psi_{a,b} dt \\
&= \mathbf{m} \int \mathbf{c}(t') \frac{1}{\sqrt{a}} g(\frac{\mp(t'-t_0)-b}{a}) dt' + \int \mathbf{t} \psi_{a,b} dt \\
&= \mathbf{m} \int \mathbf{c}(t') \frac{1}{\sqrt{a}} g(\frac{t'-(\pm b + t_0)}{a}) dt' \\
&= \mathbf{m} \int \mathbf{c}(t') \psi(t')_{a,\pm b + t_0} dt' \\
&= \mathbf{mu}_{a,\pm b + t_0} \, .
\end{aligned}
\tag{1}
$$

Note that we use the wavelet property $\int \psi_{a,b} dt = 0$ to simplify the second term in Eq. 1. If $\mathbf{m}$ represents a rotation (or the affine transform is a rigid-body motion of a translation plus a rotation), it is easily seen that an **invariant expression** (this is just one of many possibilities) can be derived using the ratio expression

$$
\frac{|\mathbf{u}'_{a,b}|}{|\mathbf{u}'_{c,d}|} = \frac{|\mathbf{mu}_{a,\pm b + t_0}|}{|\mathbf{mu}_{c,\pm d + t_0}|} = \frac{|\mathbf{u}_{a,\pm b + t_0}|}{|\mathbf{u}_{c,\pm d + t_0}|} \, ,
\tag{2}
$$

which is a function of the scale a and the displacement b. If we fix the scale a, by taking the same number of sample points along each curve, we can construct a function $f_a(x)$ which we call the **invariant signature** of an object as:

$$f_a(x) = \frac{|\mathbf{u}_{a,x}|}{|\mathbf{u}_{a,x+x_0}|} \qquad \text{and} \qquad (3)$$

$$f'_a(x) = \frac{|\mathbf{u}'_{a,x}|}{|\mathbf{u}'_{a,x+x_0}|} = \frac{|\mathbf{m}\mathbf{u}_{a,\pm x+t_0}|}{|\mathbf{m}\mathbf{u}_{a,\pm(x+x_0)+t_0}|} = \frac{|\mathbf{u}_{a,\pm x+t_0}|}{|\mathbf{u}_{a,\pm(x+x_0)+t_0}|},$$

where x_0 represents a constant value separating the two indices. Then it is easily verified that when the direction of traversal is the same for both contours, $f'_a(x) = \frac{|\mathbf{u}_{a,x+t_0}|}{|\mathbf{u}_{a,x+x_0+t_0}|} = f_a(x+t_0)$. If the directions are opposite, then $f'_a(x) = \frac{|\mathbf{u}_{a,-x+t_0}|}{|\mathbf{u}_{a,-x-x_0+t_0}|} = \frac{1}{f_a(-x-x_0+t_0)}$. As the correlation coefficient of two signals is defined as

$$R_{f(x)g(x)}(\tau) = \frac{\int f(x)g(x+\tau)dx}{\|f\| \cdot \|g\|},$$

we define the **invariant measure** $I_a(f, f')$ (or the similarity measure) between two objects as

$$I_a(f, f') = max_{\tau,\tau'}\{R_{f_a(x)f'_a(x)}(\tau), R_{f_a(x)\frac{1}{f'_a(-x)}}(\tau')\}. \quad (4)$$

It can be shown [1] that the invariant measure in Eq. 4 attains the maximum of 1 if two objects are identical, but differ in position, orientation, scale, and traversal direction and starting point.

Other Invariant Expressions Due to the page limit, we will only summarize other invariant expressions (similar to Eq. 2) without derivation. The process of deriving the *invariant signatures* (similar to Eq. 3) and *invariant measures* (similar to Eq. 4) are done similarly and will not be repeated here. Interested readers are referred to [1] for more details. Here, we provide a summary of the scenarios and their corresponding invariant expression:

Rigid-body motion (using spline basis):

$$\frac{|\mathbf{u}_{a,b} - \mathbf{u}_{c,d}|}{|\mathbf{u}_{e,f} - \mathbf{u}_{g,h}|} \quad (5)$$

Affine transform (using wavelet basis):

$$\frac{\left| \begin{array}{cc} \mathbf{u}_{\mathbf{a},\mathbf{b}} & \mathbf{u}_{\mathbf{c},\mathbf{d}} \end{array} \right|}{\left| \begin{array}{cc} \mathbf{u}_{\mathbf{e},\mathbf{f}} & \mathbf{u}_{\mathbf{g},\mathbf{h}} \end{array} \right|} \quad (6)$$

Affine transform (using spline basis):

$$\frac{\left| \begin{array}{ccc} \mathbf{u}_{\mathbf{a},\mathbf{b}} & \mathbf{u}_{\mathbf{c},\mathbf{d}} & \mathbf{u}_{\mathbf{e},\mathbf{f}} \\ 1 & 1 & 1 \end{array} \right|}{\left| \begin{array}{ccc} \mathbf{u}_{\mathbf{g},\mathbf{h}} & \mathbf{u}_{\mathbf{i},\mathbf{j}} & \mathbf{u}_{\mathbf{k},\mathbf{l}} \\ 1 & 1 & 1 \end{array} \right|} \quad (7)$$

Figure 1: A database of airplane images, numbered 1-16 in raster scan order.

Perspective transform (using *rational spline basis R*):

$$\int_t (\mathbf{d}(t) - \sum_i \mathbf{p}_i R_{i,k}(t))^2 dt \quad (8)$$

where $\mathbf{d}(t)$ is the observed image curve, and $\sum_i \mathbf{p}_i R_{i,k}(t)$ is the model curve in rational spline form.

Change of illumination:

$$\frac{\left| \left[\mathbf{u}_{a_1,b_1} \mathbf{u}_{a_2,b_2} \cdots \mathbf{u}_{a_k,b_k} \right]^T \left[\mathbf{u}_{a_1,b_1} \mathbf{u}_{a_2,b_2} \cdots \mathbf{u}_{a_k,b_k} \right] \right|}{\left| \left[\mathbf{u}_{c_1,d_1} \mathbf{u}_{c_2,d_2} \cdots \mathbf{u}_{c_k,d_k} \right]^T \left[\mathbf{u}_{c_1,d_1} \mathbf{u}_{c_2,d_2} \cdots \mathbf{u}_{c_k,d_k} \right] \right|} \quad (9)$$

3   Experimental Results

To illustrate the use of these invariant expressions as image features for image indexing and retrieval, we conducted an experiment on a database of 16 airplane images, shown in Fig. 1. Each airplane is posed in a canonic view and under standard lighting. It should be noted that many of the airplane images have approximately the same shape (e.g., images 5 and 6, and images 3, 7 and 14), making the matching process difficult. Furthermore, query images can undergo large perspective distortion and illumination changes. Eleven query images of the same airplanes are shown in Fig. 2, each in a different pose. Moreover, colored filters were placed in front of the light sources (and the number of light sources varied) to simulate changes in illumination. From these images, we extracted the silhouettes of the airplanes and used both affine and perspective invariants as indices (done offline for the

Figure 2: Query images: the same airplanes from Figure 1 in varying poses and illumination, labeled A to K in raster scan order.

	Rank (using affine invariants)				
Image	1st	2nd	3rd	4th	5th
A	(5) 0.7919	(14) 0.0222	(3) 0.0077	(12) -0.0054	(9) -0.0081
B	**(1) 0.6772**	(8) 0.2487	(9) 0.1879	(2) 0.1556	(6) 0.1512
C	(16) 0.3473	(14) 0.2944	(6) 0.1457	(2) 0.1046	(13) 0.0781
D	**(2) 0.7236**	(14) 0.4480	(11) 0.3666	(6) 0.3650	(3) 0.2065
E	(6) 0.3014	(12) 0.2680	(9) 0.2555	**(2) 0.2541**	(14) 0.2262
F	(15) 0.3350	(16) 0.2513	(9) 0.2492	(14) 0.2490	(8) 0.1634
G	(12) 0.6979	(14) 0.4114	(2) 0.2093	(3) 0.1452	(5) 0.0531
H	(14) 0.5241	(13) 0.4595	(6) 0.4225	(2) 0.3987	(10) 0.2601
I	(15) 0.3693	(2) 0.3445	(7) 0.1674	(3) 0.1644	(1) 0.0487
J	**(14) 0.4709**	(6) 0.3821	(2) 0.3416	(11) 0.3314	(8) 0.2884
K	(7) 0.7407	(4) 0.5313	(11) 0.3247	(9) 0.0994	(8) 0.0372

Table 1: Top matches between affine invariant signatures of query images and database images. Numbers in parentheses indicate the airplane image selected. The value beneath it is the correlation coefficient of the two signatures. The expected (correct) airplane image is in boldface. Each row corresponds to a query image. The columns are arranged left to right, from the best match to worse.

database images). Without highly discriminative and invariant features, it could be difficult to locate the right matches from the database.

Hence, we exploit both geometric properties and color for a robust matching strategy. Integrating geometric and illumination invariants is a novel approach which has achieved only limited success in object recognition [7]. Our integration of geometric and illumination invariants has produced promising results.

We first demonstrate the performance of affine invariants. It will be shown that affine invariants work well with objects that are relatively far from the camera, but fails miserably with objects that are near the camera or otherwise showing large perspective distortion. We then present the results of using perspective invariants, which correctly matched all sample images. Finally, we show the results of applying illumination invariants to further verify a correct match.

Table 1 shows both the effectiveness and limitations of using affine invariants, when dealing with objects under perspective transformations. Affine invariant signatures of the airplane images were stored in a database (performed offline using Eq. 7). The resolution scale used for recognition was pre-determined by computing the energy of the image signatures at different scales. The scale where energy appeared to be concentrated, was selected. (Large energy values imply more information; in this case, scale 3). For each query image (A through K), the affine invariant signature was computed (Eq. 7), and compared with the signature indices of all images in the database. Correlation coefficients (invariant measure of Eq. 4) were used to determine the similarity between each pair of signatures. Each row in Table 1 refers to a query image. Each of the five columns represents the rank given to each airplane image from the database (shown in parentheses). The columns are ordered from left to right, with the leftmost column being the best match found. Only the top five matches are shown. The values (not in parentheses) are the correlation coefficients. Entries printed in boldface are the expected (correct) matches.

It is clear from Table 1 that the affine invariant works well in cases where the object is far from the camera relative to its size. Query images B and D are consistent with this scenario. As expected, the correlation between their signatures and that of the correct image in the database is significantly higher (both rank the 1st), relative to that for the rest of the images. Fig. 3 shows the affine invariant signature of airplane B (solid) correctly matched with that of image 1 (dashed) (best match found). The next best match (image 8) is shown in Fig. 3(b).

However, the performance with other images was unsatisfactory. Hence, the affine invariant formulation can be used to approximate perspective invariance only when the object is relatively far from the camera. For our test images, the affine invariant correctly identified only three of them. For most of the
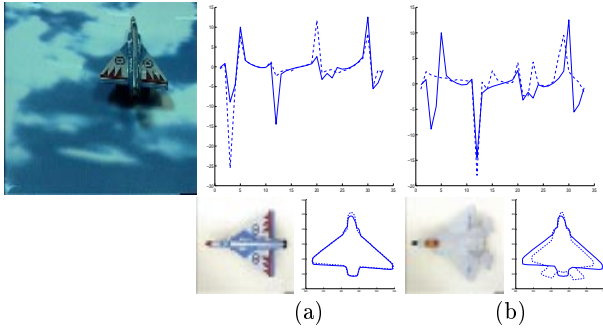
Figure 3: (a) Affine invariant signature of airplane B (solid) correctly matched with that of image 1 (dashed) (best match found), and (b) the next best match.

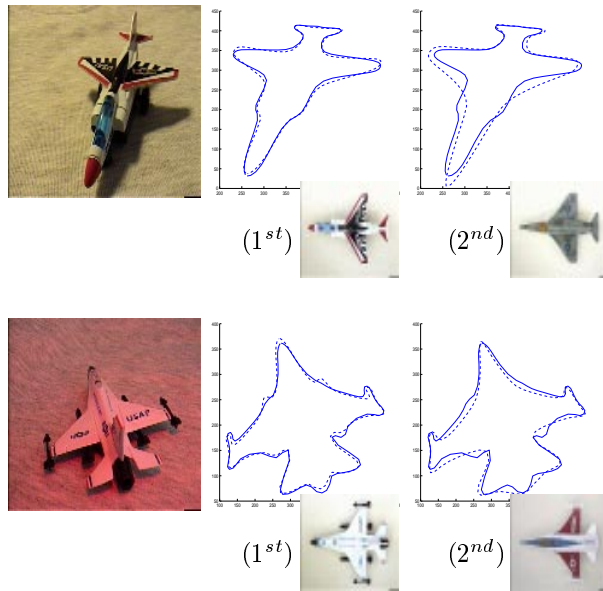(1st)   (2nd)

(1st)   (2nd)

Figure 4: Query images E and H, with their corresponding top two matches from the database, using perspective invariants. (Solid for the query image, dashed for the estimated image.)

Image	Rank (using perspective invariants)				
	1st	2nd	3rd	4th	5th
A	(**1**) **120.82**	(9) 279.01	(4) 383.86	(6) 503.34	(5) 533.66
B	(**1**) **47.30**	(9) 146.81	(10) 233.45	(4) 252.14	(6) 337.03
C	(**1**) **146.16**	(4) 319.36	(2) 347.87	(9) 398.37	(6) 437.71
D	(**2**) **71.67**	(6) 99.78	(5) 103.76	(4) 182.25	(13) 190.34
E	(**2**) **77.18**	(5) 225.33	(6) 237.78	(14) 302.50	(12) 383.28
F	(**4**) **363.12**	(1) 399.78	(9) 418.96	(6) 470.93	(10) 479.52
G	(**6**) **174.62**	(13) 270.70	(5) 297.42	(4) 338.39	(2) 354.01
H	(**7**) **125.27**	(14) 144.79	(3) 160.16	(11) 177.42	(13) 215.16
I	(**13**) **139.15**	(6) 310.96	(3) 343.69	(14) 353.16	(12) 365.75
J	(**14**) **118.53**	(3) 198.34	(12) 243.56	(13) 248.79	(7) 294.51
K	(**14**) **122.10**	(3) 144.23	(7) 237.71	(13) 272.76	(12) 272.96

Table 2: Top five matches between each query image and images in the database, using perspective invariants. Numbers in parentheses indicate the airplane image selected. The value beneath it is the relative shape deviation (error) between the estimated image and query image. The expected (correct) airplane image is in boldface. Each row corresponds to an query image. The columns are arranged left to right, from the best match to worse.

query images, the correct database image were not even among the top five candidates.

In cases of large perspective distortion, the affine invariant performs poorly as expected. Fortunately, a more suitable invariant can be used. Table 2 shows the results of using the perspective invariants on the same query images (Eq. 8). A polynomial degree of order 2 was used since we want to treat the query image points as control points. (*I.e.*, The curve passes through each of the query image points exactly.) Only the top five matches are shown, with the leftmost column as the top match. The values shown are the relative shape deviations (error) between the query image and the best estimate using a particular database image. The numbers in parentheses correspond to the image number from the database. Entries printed in boldface

are the expected (correct) matches. As can be seen, all query images retrieved the correct image as the best match. Fig. 4 shows some results in more detail. The leftmost image is the query image. The top two matches are in the next two columns − the query image (solid) and estimated image (dashed) with the corresponding database image are shown. See [1] for more result examples.

For this experiment, *all* query images retrieved the correct database image as the best match, using perspective invariants. However, the system is far from perfect. The results for query image F in Table 2 show that it correctly retrieved image 4 from the database. Nonetheless, the relatively large error (possibly due to noise) between the query image and estimated image may be unacceptable. Furthermore, the user may want to discriminate between images that appear similar geometrically, but are in fact different in terms of color patterns within the image region. For instance, the error values of the top two matches for airplane K were very close to each other, indicating that both are geometrically similar to the query image. If the color pattern of the airplanes are put into consideration, however, image 14 should be considered to be far

(a)   (d)

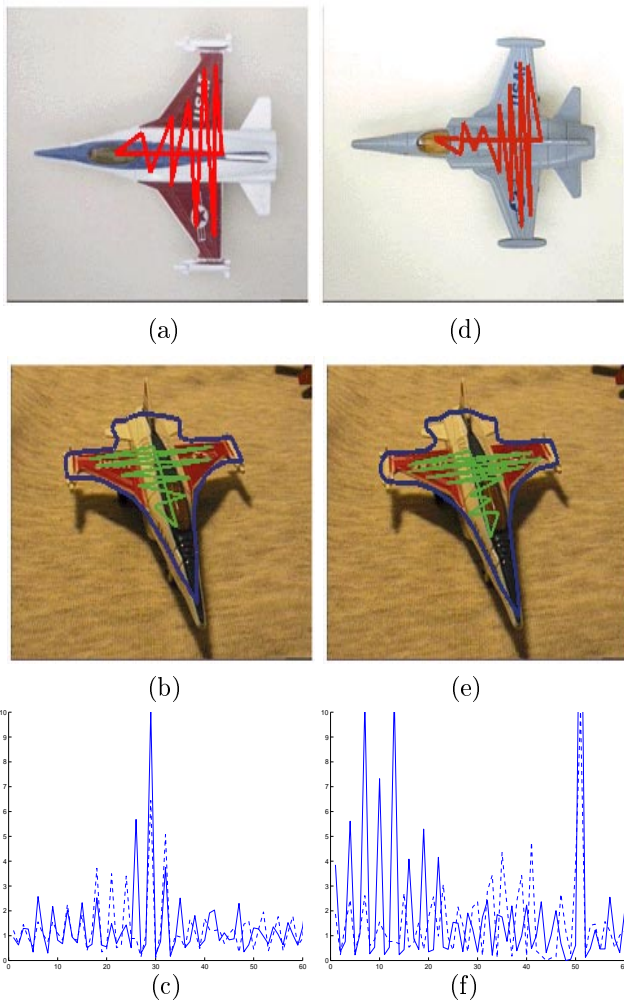
(b)   (e)

(c)   (f)

Figure 5: (a),(d) Airplane images with one of the defined curves (in red) superimposed, (b),(e) query image with the transformed curve superimposed (in green). The transformed contour of the image is in blue, and (c),(f) illumination invariant signatures for query image K (solid) and for database images 14 and 3 (dashed).

more similar to the query image. Illumination invariants readily applies here.

For illumination invariants, curves were uniquely defined on the surface of each airplane image in the database (performed off-line), so that its superimposition over the image emphasizes important (or interesting) color patterns in the image. Figs. 5(a) and (d) show one of many such curves defined for images 14 and 3, respectively. (In this case, a zig-zagging curve over the different colors of the airplane). A different set of curves for other images in the database were also defined. Using the same strategy as that for affine invariants, we determined a resolution level (possibly many) to be used for recognition; in this case, level 7.

Computation of the illumination invariant signatures were done off-line (using Eq. 9).

After computing the perspective invariants, the transformation parameters were easily obtained. The same parameters were used to transform the curve defined for each image, to its correct pose in the query image. Hence, the colors defined by the curve in the image should match the colors defined by the transformed curve in the query image (except for changes due to illumination). Illumination invariant signatures for the query images were then computed, and compared with the signatures stored in the database. Again, correlation coefficients (Eq. 4) were used to measure the similarity between the pairs of signatures.

Illumination invariants were computed only for those query images where it was desirable to discriminate between the object's regions. Here, we only show results for query image K.

In Fig. 5, (a) and (d) show one of many curves superimposed (in red) over the images of 14 and 3. The transformed curves, shown in (b) and (e), is superimposed (in green) over the query image K. The estimated image (from perspective invariants) is outlined in blue. Finally, (c) and (f) show the illumination invariant signatures. Clearly, the signatures in (c) are more consistent.

4 Conclusion

We present a new framework for computing invariant image-derived features for image indexing and retrieval, providing high discriminative power and insensitivity to many environmental changes. Preliminary results are very promising, demonstrating the efficacy for shape retrieval in a database containing similar models and with query images with large perspective distortion and illumination change.

References

[1] R. Alferez and Y.F. Wang. Geometric and Illumination Invariants for Object Recognition. *IEEE Trans. Patt. Analy. Mach. Intell.* To appear as a regular paper.

[2] K. Arbter, W. E. Snyder, H. Burkhardt, and G. Hirzinger. Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects. *IEEE Trans. Patt. Analy. Mach. Intell.*, 12:640–647, 1990.

[3] I. Daubechies. Orthonormal Bases of Compactly Supported Wavelets. *Comm. Pure Appl. Math.*, 41:909–960, 1988.

[4] Hampapur et.al. Virage Video Engine. *Proc. of SPIE, Storage and Retrieval for Image and Video Databases V*, 3022:188–200, 1997.

[5] M. Flickner et.al. Query by Image and Video Content: The QBIC System. *IEEE Comput.*, pages 23–32, Sept 1995.

[6] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Tools for Content-Base Manipulation of Image Databases. *Int. J. Comput. Vision*, 18(3):233–254, 1996.

[7] D. Slater and G. Healey. The Illumination-Invariant Matching of Deterministic Local Structure in Color Images. *IEEE Trans. Patt. Analy. Mach. Intell.*, 19(10):1146–1151, 1997.