Image Database Indexing using a Combination of Invariant Shape and Color Descriptions *

Ronald Alferez and Yuan-Fang Wang
Computer Science Department
University of California
Santa Barbara, CA, 93106 U.S.A.

Abstract ***Image and video library applications are becoming increasingly popular. The increasing popularity calls for software tools to help the user query and retrieve database images efficiently and effectively. In this paper, we present a technique which combines shape and color descriptors for invariant, within-a-class retrieval of images from digital libraries. We demonstrate the technique on a real database containing airplane images of similar shape and query images that appear different from those in the database because of lighting and perspective. We were able to achieve a very high retrieval rate.***

Keywords: images, video, libraries, features

1 Introduction

Image and video library applications are becoming increasingly popular as witnessed by many national and international research initiatives in these areas, an exploding number of professional meetings devoted to image/video/multi-media, and the emergency of commercial companies and products. The advent of high-speed networks and inexpensive storage devices has enabled the construction of large electronic image and video archives and greatly facilitated their access on the Internet. In line with this, however, is the need for software tools to help the user query and retrieve database images efficiently and effectively.

Querying an image library can be difficult and one of the main difficulties lies in designing powerful features or descriptors to represent and organize images in a library. Many existing image database indexing and retrieval systems are only capable of *between-classes* retrieval (e.g., distinguishing fish from airplanes). However, these systems do not allow the user to retrieve images that are more specific. In other words, they are unable to perform *within-a-class* retrieval (e.g., distinguishing different types of airplanes or different species of fish). This is because the aggregate features adopted by many current systems (such as color histograms and low-ordered moments) capture only the general shape of a class and are not descriptive enough to distinguish objects within a particular class.

The within-a-class retrieval problem is further complicated if query images depicting objects, though belonging to the class of interest, may look different due to non-essential or incidental environment changes, such as rigid-body or articulated motion, shape deformation, and change in illumination and viewpoint. In this paper, we address the problem of *invariant, within-a-class* retrieval of images by using a combination of invariant shape and color descriptors. By analyzing the shape of the object's contour as well as the color and texture characteristics of the enclosed area, information from multiple sources is fused for a more robust description of an object's appearance. this places our technique at an advantage over most current approaches that exploit either geometric information or color information exclusively.

The analysis involves projecting the shape and color information onto basis functions of finite, local support (e.g., splines and wavelets). The projection coefficients, in general, are sensitive to changes induced by rigid motion, shape deformation, and change in illumination and perspective. We derive expressions by massaging these sets of projection coefficients to cancel out the environmental factors to achieve invariance of the descriptors. Based on these features, we have conducted preliminary experiments to recognize different types of airplanes (many of them having very similar shape) under varying illumination and

*Supported in part by a grant from the National Science Foundation, IRI-94-11330

viewing conditions and have achieved good recognition rates. We show that information fusion has helped to improve the accuracy in retrieval and shape discrimination.

2 Technical Description

In this section, we present the theoretical foundation of our image-derived, invariant shape and color features. Invariant features form a compact, intrinsic description of an object, and can be used to design retrieval and indexing algorithms that are potentially more efficient than, say, aspect-based approaches.

The search for invariant features (e.g., algebraic and projective invariants) is a classical problem in mathematics dating back to the 18th century. The need for invariant image descriptors has long been recognized in computer vision. Invariant features can be designed based on many different methods and made invariant to rigid-body motion, affine shape deformation, scene illumination, occlusion, and perspective projection. Invariants can be computed either globally, such is the case in invariants based on moments or Fourier transform coefficients, or based on local properties such as curvature and arc length. See [3, 4, 5] for survey and discussion on the subject of invariants.

As mentioned before, our invariant features are derived from a localized analysis of an object's shape and color. The basic idea is to project an object's exterior contour or interior region onto localized bases such as wavelets and splines. The coefficients are then normalized to eliminate changes induced by non-essential environmental factors such as viewpoint and illumination. We will illustrate the mathematical frameworks using a specific scenario where invariants for curves are sought. The particular basis functions used in the illustration will be the wavelet bases and spline functions. Interested readers are referred to [1] for more details.

Several implementation issues arise in this invariant framework which we will briefly discuss before describing the invariant expressions themselves. [1]

1. How are contours extracted?

Or stated slightly differently, how is the problem of segmentation (separating objects from background) addressed? Segmentation turns out to be an extremely difficult problem and, as fundamental a problem as segmentation is, there is no fail-proof solution. A "*perfect*" segmentation scheme is like the holy grail of low-level computer vision and a panacea to many high level vision problems as well.

We are not in search of this holy-grail, which, we believe, is untenable in the foreseeable future. In an image database application, the problem of object segmentation is simplified because

- Database images can usually be acquired under standard imaging conditions which allow the ingest and catalog operations to be automated or semi-automated. For example, to construct a database of airplane images, many books on civil and military aircrafts are available with standard front, side, and top views taken against a uniform or uncluttered background. (The above is also true for applications in botany and marine biology.) This allows the contours of the objects of interest to be extracted automatically or with the aid of standard tools such as the flood fill mask in Photoshop. Furthermore, the cataloging operations are usually done off-line and done only once. Hence, a semi-automated scheme will suffice.
- On the other hand, query images are usually taken under different lighting and viewing conditions. Objects of interest can be embedded deeply in cluttered background which makes their extraction difficult. However, we can enlist the help of the user to specify the object of interest instead of asking the system to attempt the impossible task of automated segmentation. A query-by-sketch or a "human-in-the-loop" type solution with an easy-to-use graphics interface and segmentation aids such as the flood fill mask is perfectly adequate here and does not impose undue burden on the user. This proved to be feasible in our experiments.

2. How are contours parameterized?

For a contour based description, a common frame of reference is usually needed that allows point correspondences to be established between two contours for comparison. The common frame of reference comprises a common starting point of traversal, a common direction of traversal, and a parameterization scheme that traverses to the corresponding points in the two contours at the same parameter setting. We will first discuss the parameter-

[1] A word on the notational convention: matrices and vectors will be represented by bold-face characters while scalar quantities by plain-face characters. 2D quantities will be in small letters while 3D and higher-dimensional quantities in capital letters. For example, coordinates (bold for vector quantities) of a 2D curve (small letter for 2D quantities) will be denoted by $\mathbf{c}$.

ization issue and then address the issues of point correspondence and traversal direction.

When defining a parameterized curve $\mathbf{c}(t) = [x(t), y(t)]^T$, most prefer the use of the intrinsic arc length parameter because of its simplicity and the fact that it is either invariant or transforms linearly in rigid-body motion and uniform scaling. However, under more general scenarios where shape deformation is allowed (e.g., deformation induced in an oblique view), intrinsic arc length parameter is no longer invariant. Such deformation can stretch and compress different portions of of an object's shape, and a parameterization based on intrinsic arc length will result in wrong point correspondence.

It is well known that many shape deformation and distortion resulting from imaging can be modeled as an affine transform, through which the intrinsic arc length is nonlinearly transformed [2]. An alternative parameterization is thus required. There are at least two possibilities. The first, called *affine arc length*, is defined [2] 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), and (a, b) is the path along a segment of the curve.

Another possibility [2] is to use the *enclosed area parameter*: $\sigma = \frac{1}{2}\int_a^b |x\dot{y} - y\dot{x}|\ dt$. One can interpret the enclosed area parameter as the area of the triangular region enclosed by the two line segments from the centroid of an object to two points a and b on the contour. It can be shown that both these parameters transform linearly under a general affine transform [2]. Hence, they can easily be made absolutely invariant by normalizing them with respect to the total affine arc length or the total enclosed area of the whole contour, respectively. We use these parameterizations in our experiments.

3. How are identical traversal direction and starting point guaranteed?

It will be shown that the *invariant signatures* (to be defined later) of two contours are phase-shifted versions of each other when only the starting point of traversal differs. Furthermore, the same contour parameterized in opposite directions produces invariant signatures that are flipped and inverted images of each other. Hence, a match can be chosen that maximizes certain cross-correlation relations between the two signatures.

Allowing an arbitrary change of origin and traversal direction, together with the use of an affine invariant parameterization, imply that **no point correspondence is required in computing our invariants.**

Now we are ready to introduce the invariant expressions themselves. Our invariants framework is very general and considers variation in an object's image induced by rigid-body motion, affine deformation, and changes in parameterization, scene illumination, and viewpoint. Each formulation can be used alone, or in conjunction with others. Due to the page limitation, we can only give a brief discussion of the invariants under rigid-body and affine transform and summarize the invariant expressions under change of illumination and viewpoint. Interested readers are referred to [1] for more details.

Invariants under 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) 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: $\mathbf{c}'(t) = \mathbf{m}\mathbf{c}(t') + \mathbf{t} = \mathbf{m}\mathbf{c}(\pm t + t_0) + \mathbf{t}$, where $\mathbf{m}$ is any nonsingular 2×2 matrix, **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:

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

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

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

which is a function of the scale a and the displacement b. If we fix the scale a, by taking the same

Scenarios	Invariant expressions
Rigid-body motion (using spline basis)	$\frac{\lvert \mathbf{u}_{a,b} - \mathbf{u}_{c,d} \rvert}{\lvert \mathbf{u}_{e,f} - \mathbf{u}_{g,h} \rvert}$
Affine transform (using wavelet basis)	$\frac{\begin{vmatrix} \mathbf{u_{a,b}} & \mathbf{u_{c,d}} \end{vmatrix}}{\begin{vmatrix} \mathbf{u_{e,f}} & \mathbf{u_{g,h}} \end{vmatrix}}$
Affine transform (using spline basis)	$\frac{\begin{vmatrix} \mathbf{u_{a,b}} & \mathbf{u_{c,d}} & \mathbf{u_{e,f}} \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} \mathbf{u_{g,h}} & \mathbf{u_{i,j}} & \mathbf{u_{k,l}} \\ 1 & 1 & 1 \end{vmatrix}}$
Perspective transform (using *rational spline basis* R)	$\int_t (\mathbf{d}(t) - \sum_i \mathbf{p}_i R_{i,k}(t))^2 dt$ where $\mathbf{d}(t)$ is the observed image curve, and $\sum_i \mathbf{p}_i R_{i,k}(t)$ is the database curve in rational spline form.
Change of illumination	$\frac{\left\lvert [\mathbf{u}_{a_1,b_1} \mathbf{u}_{a_2,b_2} \cdots \mathbf{u}_{a_k,b_k}]^T [\mathbf{u}_{a_1,b_1} \mathbf{u}_{a_2,b_2} \cdots \mathbf{u}_{a_k,b_k}] \right\rvert}{\left\lvert [\mathbf{u}_{c_1,d_1} \mathbf{u}_{c_2,d_2} \cdots \mathbf{u}_{c_k,d_k}]^T [\mathbf{u}_{c_1,d_1} \mathbf{u}_{c_2,d_2} \cdots \mathbf{u}_{c_k,d_k}] \right\rvert}$

Table 1: Other invariant measures

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:

$$\begin{aligned} f_a(x) &= \frac{\lvert \mathbf{u}_{a,x} \rvert}{\lvert \mathbf{u}_{a,x+x_0} \rvert} \quad \text{and} \\ f'_a(x) &= \frac{\lvert \mathbf{u}'_{a,x} \rvert}{\lvert \mathbf{u}'_{a,x+x_0} \rvert} = \frac{\lvert m\mathbf{u}_{a,\pm x+t_0} \rvert}{\lvert m\mathbf{u}_{a,\pm(x+x_0)+t_0} \rvert} \\ &= \frac{\lvert \mathbf{u}_{a,\pm x+t_0} \rvert}{\lvert \mathbf{u}_{a,\pm(x+x_0)+t_0} \rvert} , \end{aligned} \tag{3}$$

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, $f'_a(x) = \frac{\lvert \mathbf{u}_{a,x+t_0} \rvert}{\lvert \mathbf{u}_{a,x+x_0+t_0} \rvert} = f_a(x + t_0)$. If the directions are opposite, then $f'_a(x) = \frac{\lvert \mathbf{u}_{a,-x+t_0} \rvert}{\lvert \mathbf{u}_{a,-x-x_0+t_0} \rvert} = \frac{1}{f_a(-x-x_0+t_0)}$. 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

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

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. Due to the page limit, we will only summarize other invariant expressions in Table 1 without derivation. The entries shown in the table are the *invariant expressions* (similar to Eq. 2). The process of deriving *invariant signatures* (similar to Eq. 3) and *invariant measures* (similar to Eq. 4) are similar and will not be repeated here.

3 Experimental Results

In the following, we will present some preliminary results. The purpose is to provide a proof-of-concept demonstration and to discover research issues that need be addressed for a large-scale implementation and testing. Hence, the database used is of a relatively small size.

The scenario is that of a digital image database comprises a collection of sixteen airplanes in canonical (top) view (Fig. 1). The airplane contours were automatically extracted from the images and invariant shape and color signatures computed off-line. Eleven query images (Fig. 2) were photographed of these airplanes from different viewpoints and under varying illumination. The airplanes in the query images were extracted using a semi-automated process with user assistance. Even though the image database is relatively small, it contains objects of very similar appearance (e.g., models 5 and 6, and models 3, 7, and 14). Furthermore, the query images (Fig. 2) differ greatly from the database images due to large changes in perspective and illumination. This is in contrast with many digital image library retrieval schemes which can perform only between-classes (e.g., airplanes vs. cars) retrievals with small changes in imaging condition.

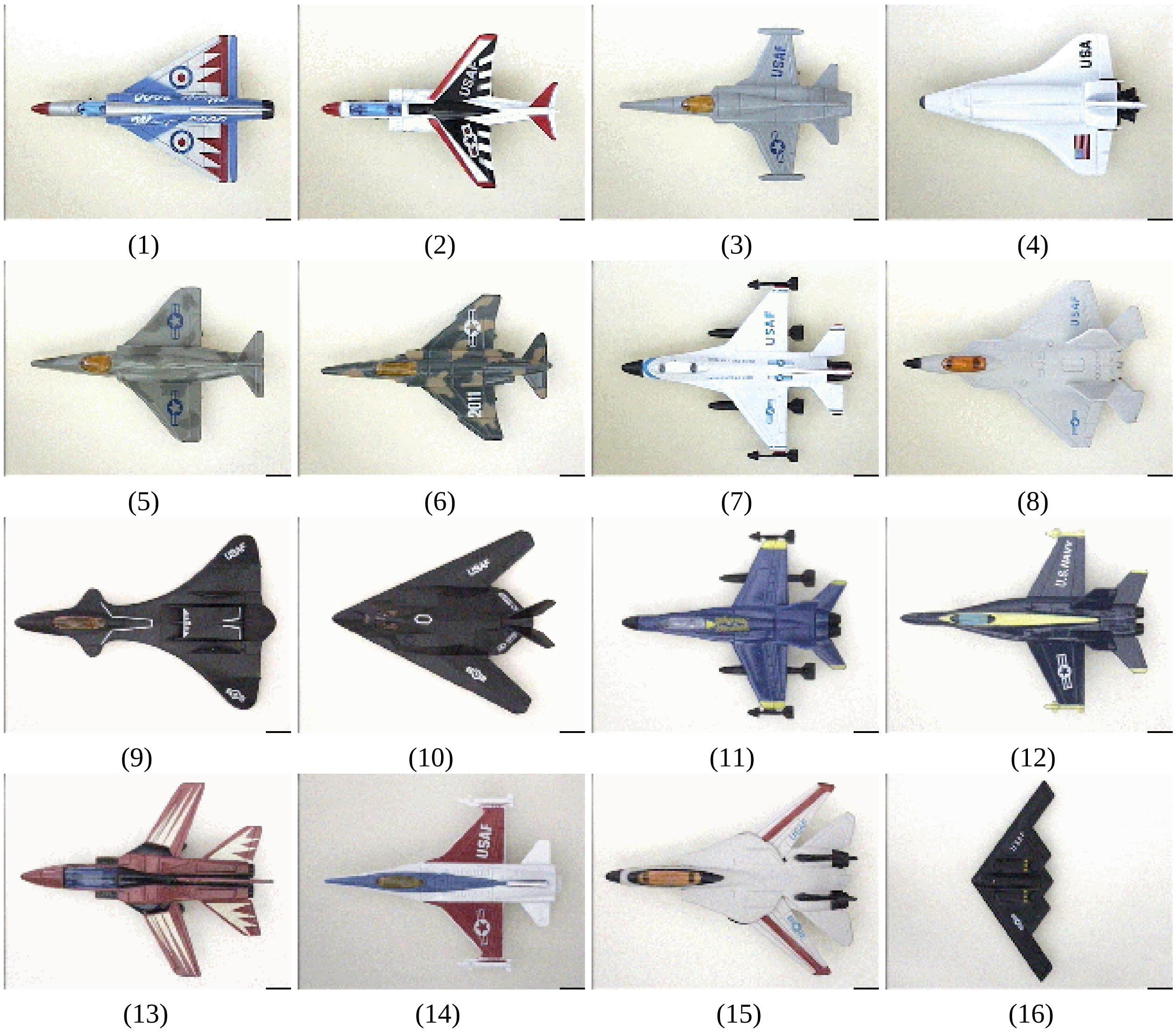

Figure 1: A database of airplane models.

We used a two-stage approach in information fusion. Features invariant to affine deformation and perspective projection were first used to match the silhouette of the query airplane with the silhouettes of those in the database. We then employed the illumination invariants computed on objects' interior to disambiguate among models with similar shape but different colors. **The results show that we were able to achieve 100% accuracy using our invariants formulation for a database comprising very similar models, presented with query images of large perspective shape distortion and change in illumination.**

Table 2 shows the performance of using affine and perspective invariants for shape matching under a large change of viewpoint. For each query image (A through K), the affine and perspective invariant signatures were computed, and compared with the signatures of all models in the database. Correlation coefficients as described in Sec. 2 were used to determine the similarity between each pair of signatures. Each row in Table 2 refers to a query image. Each of the ten columns represents the rank given to each airplane model 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 ten matches are shown. The values (not in parentheses) are the correlation coefficients. Entries printed in boldface are the expected (correct) matches.

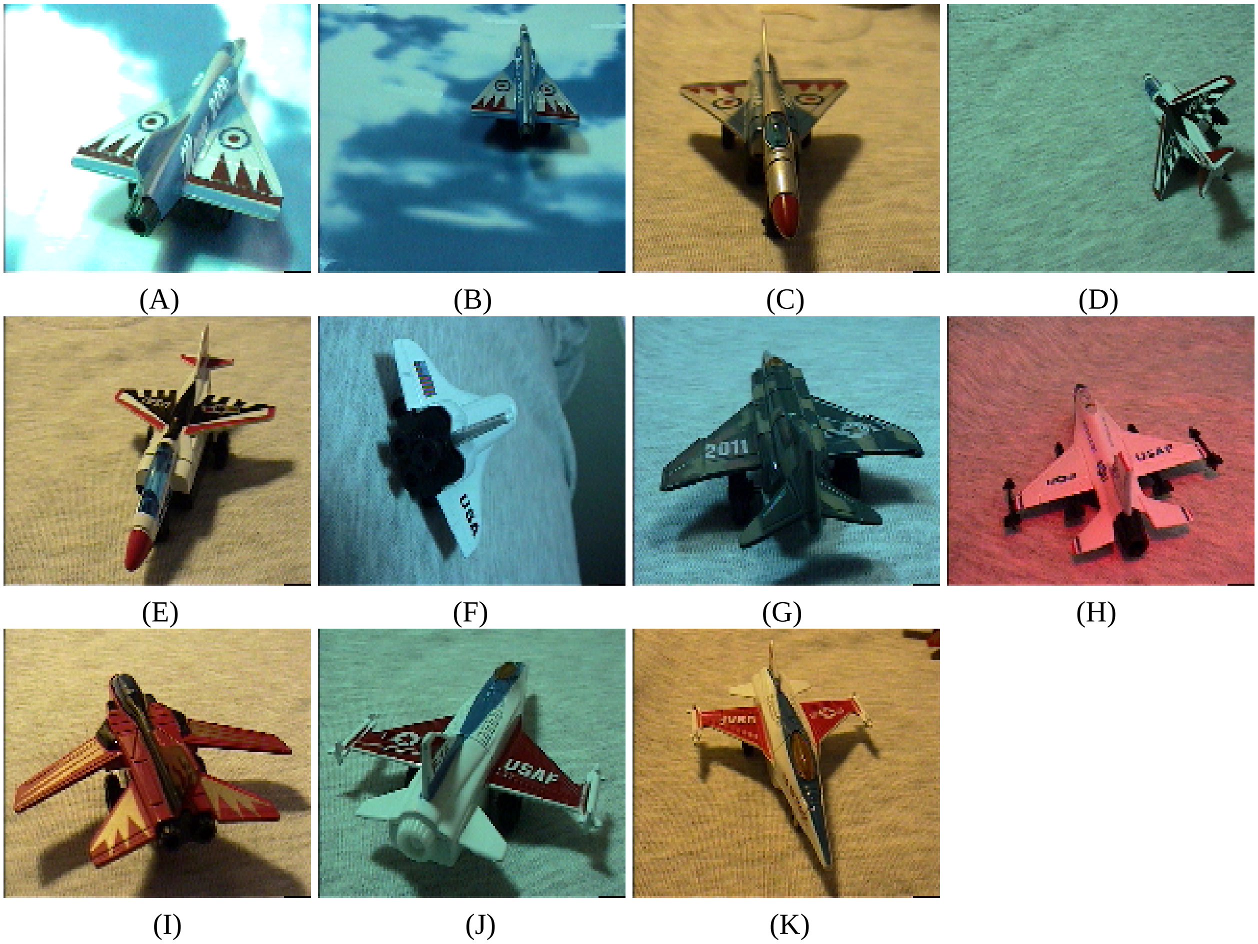

Figure 2: The same airplanes in varying poses and illumination.

As can be seen, all query images were identified correctly. Fig. 3 shows a sample result. The leftmost image is the query image. The top three matches are in the next three columns—the query image (solid) and estimated image (dashed, using perspective invariants) with the corresponding database model are shown.

For this experiment, *all* query images were correctly matched with the models from the database, using affine and perspective invariants. However, the error values of the top two matches for, say, airplane K were very close to each other. This is because the top two matches have similar shapes and both are similar to the query image. The confidence in the selected matches can be strengthened by testing whether the interior regions of the objects are also consistent. Illumination invariants readily applies here.

For illumination invariants, a characteristic curve was uniquely defined on the surface of each airplane model in the database (performed offline), so that its superimposition over the model emphasizes important (or interesting) color patterns in the image. Our perspective invariants scheme computed the transformation parameters that best match the two given contours. The same parameters were used to transform the characteristic curve defined for each model to its assumed pose in the query image. Hence, the colors defined by the characteristic curve in the model 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 using Eq. 4.

We show one result of illumination invariants where the (perspective invariant) errors of the 1^{st}

	Rank (using affine and perspective invariants)									
Image	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
A	**(1)** **0.8792**	(9) 0.7210	(4) 0.6161	(6) 0.4967	(5) 0.4663	(10) 0.4578	(2) 0.4030	(7) 0.3248	(11) 0.2443	(14) 0.2388
B	**(1)** **0.9527**	(9) 0.8532	(10) 0.7666	(4) 0.7479	(6) 0.6630	(2) 0.6103	(5) 0.5943	(15) 0.5364	(16) 0.4756	(7) 0.4576
C	**(1)** **0.8538**	(4) 0.6806	(2) 0.6521	(9) 0.6016	(6) 0.5623	(5) 0.5353	(10) 0.4446	(14) 0.3359	(7) 0.3095	(11) 0.2386
D	**(2)** **0.9283**	(6) 0.9002	(5) 0.8962	(4) 0.8177	(13) 0.8097	(14) 0.7801	(1) 0.7730	(7) 0.7663	(3) 0.7502	(12) 0.7439
E	**(2)** **0.9228**	(5) 0.7747	(6) 0.7622	(14) 0.6975	(12) 0.6167	(4) 0.6167	(3) 0.6146	(13) 0.5902	(7) 0.5704	(15) 0.4813
F	**(4)** **0.6369**	(1) 0.6002	(9) 0.5810	(6) 0.5291	(10) 0.5205	(14) 0.5056	(5) 0.4486	(11) 0.4283	(2) 0.4036	(7) 0.3946
G	**(6)** **0.8254**	(13) 0.7293	(5) 0.7026	(4) 0.6616	(2) 0.6460	(14) 0.6396	(12) 0.6287	(3) 0.6035	(1) 0.5930	(7) 0.5638
H	**(7)** **0.8747**	(14) 0.8552	(3) 0.8398	(11) 0.8226	(13) 0.7848	(6) 0.7668	(12) 0.7663	(5) 0.7282	(2) 0.7007	(4) 0.6980
I	**(13)** **0.8609**	(6) 0.6890	(3) 0.6563	(14) 0.6468	(12) 0.6343	(5) 0.6107	(7) 0.5916	(2) 0.5849	(15) 0.5775	(1) 0.5516
J	**(14)** **0.8815**	(3) 0.8017	(12) 0.7564	(13) 0.7512	(7) 0.7055	(11) 0.6805	(6) 0.6501	(4) 0.6346	(5) 0.5838	(15) 0.5711
K	**(14)** **0.8779**	(3) 0.8558	(7) 0.7623	(13) 0.7272	(12) 0.7270	(6) 0.7235	(11) 0.7209	(2) 0.6503	(5) 0.6191	(4) 0.5459

Table 2: Top ten matches between each query image and database models, using affine and perspective invariants. Numbers in parentheses indicate the airplane model selected. The value beneath it is the similarity measure between the selected image and query image. The correct airplane model is in boldface. Each row corresponds to a query image. The columns are arranged left to right, from the best match to worse.

and 2^{nd} best matches differ by a small amount (see Table 2); in this case, query image K. Figs. 4 (a) and (d) show the characteristic curves (the zigzag lines) superimposed over the images of models 14 and 3. The transformed characteristic curves, shown in (b) and (e), is superimposed over the query image K, using parameters estimated from perspective invariants. Finally, (c) and (f) show the illumination invariant signatures. Clearly, the signatures in (c) is much more consistent, which reinforces the results from shape invariants.

4 The Concluding Remarks

We present a technique where shape/color information from interior/contour points is used to describe an imaged object for database retrieval. The technique is superior in that it tolerates changes in appearance induced by incidental environmental factors and is powerful enough for within-a-class retrieval.

References

[1] Ronald Alferez and Y. F. Wang. Geometric and Illumination Invariants for Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999. to appear as a regular paper.

[2] H. W. Guggenheimer. *Differential Geometry*. McGraw-Hill, New York, 1963.

[3] J. Mundy and A. Zisserman (eds.). *Geometric Invariance in Computer Vision.* MIT Press, Cambridge, MA, 1992.

[4] T. H. Reiss. *Recognizing Planar Objects Using Invariant Image Features*. Springer-Verlag, Berlin, 1993.

[5] I. Weiss. Geometric Invariants and Object Recognition. *Int. J. Comput. Vision*, 10(3):207–231, 1993.

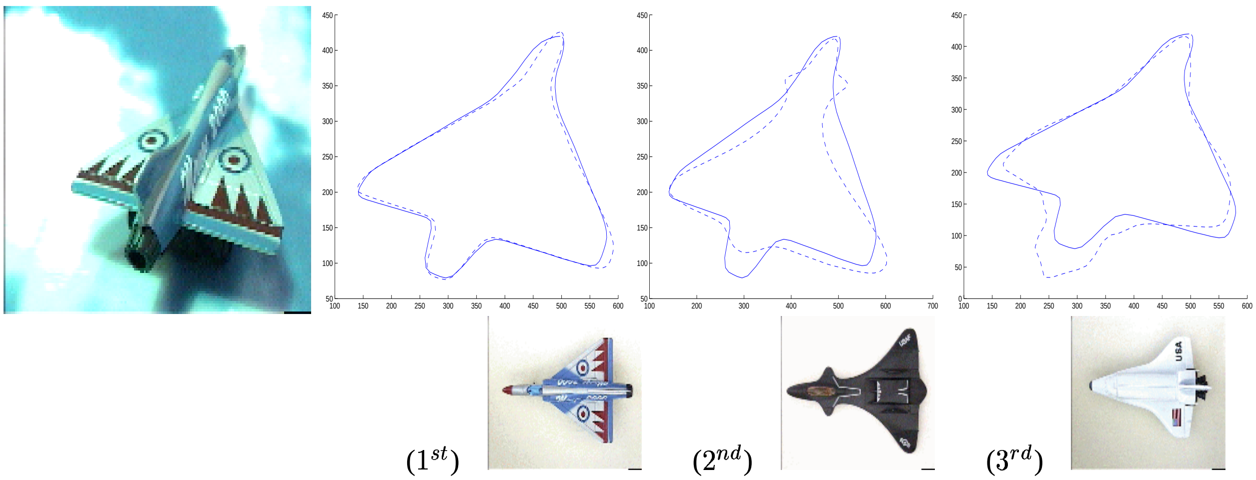

Figure 3: Query image A, with the top three matches from the database, using perspective invariants. (Solid for the query image, dashed for the estimated image using perspective invariants.)

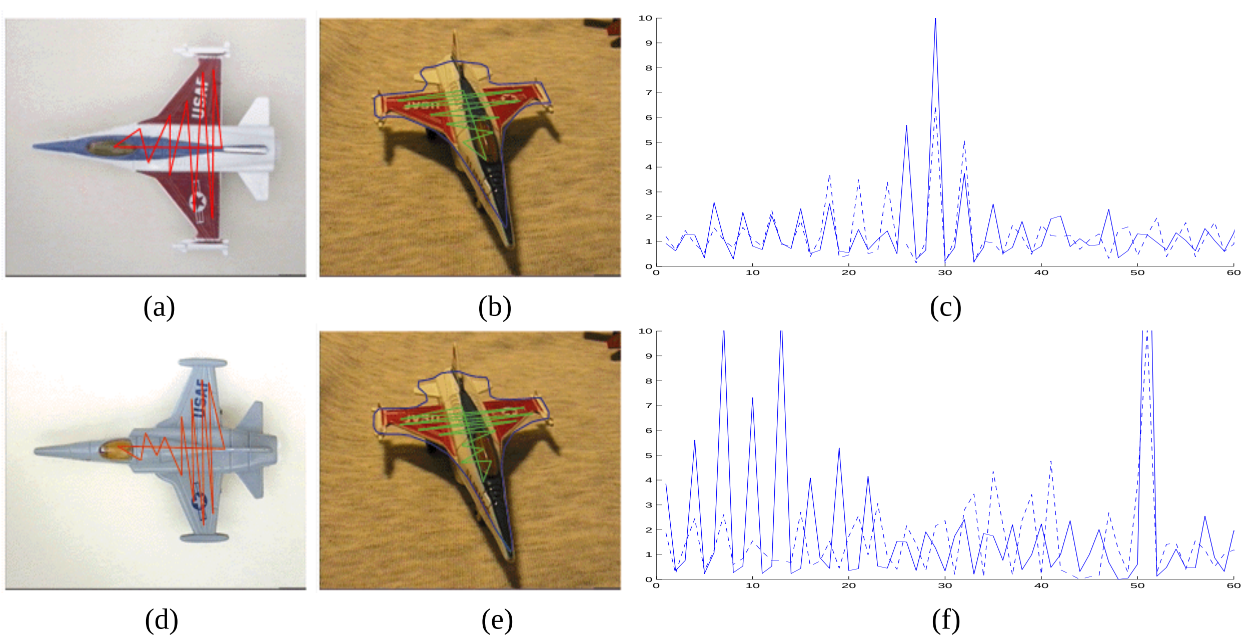

Figure 4: (a),(d) Airplane models with the characteristic curves superimposed, (b),(e) query image with the transformed characteristic curves superimposed, and (c),(f) illumination invariant signatures for query image K (solid) and for models 14 and 3 (dashed).