Projective and illumination invariant representation of disjoint shapes

Startchik, Sergei; Milanese, Ruggero; Pun, Thierry

How to cite

STARTCHIK, Sergei, MILANESE, Ruggero, PUN, Thierry. Projective and illumination invariant representation of disjoint shapes. In: Proceedings of the Fifth European Conference on Computer Vision, ECCV 98. Freiburg (Germany). [s.l.] : Springer, 1998. p. 264–280. (Lecture Notes in Computer Science) doi: 10.1007/BFb0055672

This publication URL: https://archive-ouverte.unige.ch//unige:47932

Publication DOI: 10.1007/BFb0055672

Projective and illumination invariant representation of disjoint shapes*

Sergei Startchik, Ruggero Milanese, Thierry Pun

Dept. of Computer Science (CUI)
University of Geneva
24, rue General Dufour
1211 Geneva 4, Switzerland

E-mail: <FirstName>.<LastName>@cui.unige.ch
Tel.: *+41 (22) 705-76-60 (secr.)*

Abstract

We describe a new projectively invariant representation of disjoint contour groups which is suitable for shape-based retrieval from an image database. It consists of simultaneous polar reparametrization of multiple curves where an invariant point is used as the origin. For each ray orientation, a cross-ratio of its intersections with other curves is taken as a value associated to the radius. With respect to other methods this representation is less reliant on single curve properties, both for the construction of the projective basis and for calculating the signature. It is therefore more robust to contour gaps and image noise and is better suited to describing complex planar shapes defined by multiple disjoint curves. The proposed representation has been originally developed for planar shapes, but an extension is proposed and validated for 3D faceted objects. Moreover, we show that illumination invariance fits well within the proposed framework and can easily be introduced in the representation in order to make it more appropriate for shape-based retrieval. Experiments are reported on a database of real trademarks.

Keywords: Projective invariance, Cross-ratio, Geometry, Illumination invariance, Shape-based retrieval, Object representation.

*. This work is supported by a grant from the Swiss National Fund for Scientific Research 20-40239.94

Projective and illumination invariant representation of disjoint shapes

Abstract

We describe a new projectively invariant representation of disjoint contour groups which is suitable for shape-based retrieval from an image database. It consists of simultaneous polar reparametrization of multiple curves where an invariant point is used as the origin. For each ray orientation, a cross-ratio of its intersections with other curves is taken as a value associated to the radius. With respect to other methods this representation is less reliant on single curve properties, both for the construction of the projective basis and for calculating the signature. It is therefore more robust to contour gaps and image noise and is better suited to describing complex planar shapes defined by multiple disjoint curves. The proposed representation has been originally developed for planar shapes, but an extension is proposed and validated for 3D faceted objects. Moreover, we show that illumination invariance fits well within the proposed framework and can easily be introduced in the representation in order to make it more appropriate for shape-based retrieval. Experiments are reported on a database of real trademarks.

Keywords: Projective invariance, Cross-ratio, Geometry, Illumination invariance, Shape-based retrieval, Object representation.

1. Introduction

The emerging field of image databases [6,10,22,30] has created a demand for new querying techniques. Such techniques must be able to cope with huge amounts of image data, without restrictions on the image content. Computer vision methods can be employed, provided that they are fast, reliable and not reliant on application-specific constraints [12].

A user needs to retrieve images according to some semantic description. We hypothesize that part of these descriptions can be represented in terms of image properties. Searching for images containing a particular object (e.g. a trademark in figure 1.(a)), amounts to searching for properties like colour [7] or shape [6,10]. However, all such properties are influenced by the viewing conditions. The reliability of a search method depends on how well it can separate information influenced by viewing conditions from object properties, and detect the latter. Separable object properties are called invariants and their application to computer vision was first broadly reviewed in [19]. The geometric changes introduced by varying viewing conditions can be modelled by either projective or affine transformations, while linear transformations can be used to model chromatic changes under different illuminants.

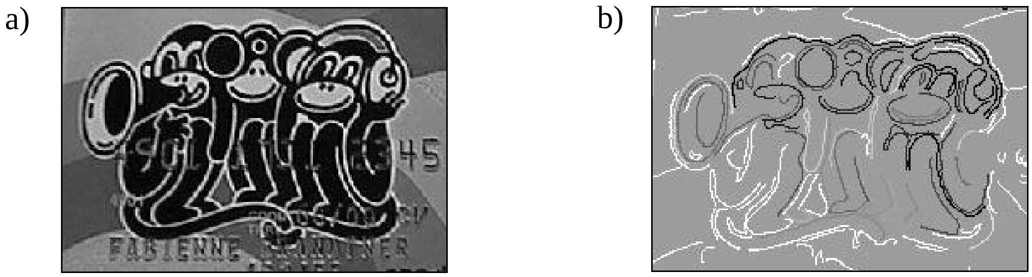

Figure 1 Example of a trademark image (a). The large number and form of the curves, extracted from the image (b), illustrates the highly distributed nature of a shape.

Several attempts to obtain invariant representations of geometric properties were summarized in [19]. For mathematical reasons, invariants are obtained more easily for planar geometric structures [24,3]. This has limited their use to object facets and trademark recognition. Some studies have proposed the use of algebraic invariants, i.e. measures that are obtained from *regular* geometric structures like a group of lines [29] or conics [20]. The difficulty of characterizing real objects with such structures has constrained the use of algebraic invariants to specific (mostly industrial) classes of objects. Also, as pointed out in [15], the small number of these invariants fails to provide sufficient discriminative capability when the amount of objects increases.

Differential invariants (based on derivatives) were designed to generalize previous approaches to a larger class of objects and consist of expressing the behaviour of a shape in a reference frame, defined by some regular invariant geometric structures [32,31]. Such structures could be invariant points [2], tangents [18,24], lines [9]. As an alternative to the projective camera model, an affine model can be used. This has the advantage of simplifying the invariant properties, and proves useful in a wide range of applications [21,28]. Advantages of differential invariants include locality, and the coverage of a relatively large class of shapes. These methods, however, represent one curve at a time and most curves (like those in figure 1.b) do not possess a sufficient number of invariant properties to characterize them. A lot of geometric information is thus lost. For completeness, two more shape representation methods should be mentioned, namely shape decomposition with ellipses [2] and deformable templates [27].

In practice, when invariant representations are used for shape-based retrieval, two major weaknesses can be observed. First, they focus on single-curve properties, thereby neglecting the fact that shapes are generally defined over a neighbourhood containing multiple curves. The second weakness is the tendency of purely geometric and local representations to produce a large number of false matches. A solution to this problem is to enrich the representation with a (possibly invariant) description of the chromatic properties of the shape. This paper proposes a technique that combines a geometric shape representation integrating multiple curves with illumination invariant information.

The rest of this paper is structured as follows: section 2 outlines the idea of the proposed representation and shows its projective invariance; Section 3 focuses on the problem of finding reference lines, necessary for invariant reparametrization; in Section 4 the method is extended with illumination invariance. Finally, in section 5 experimental results on invariance and database aspects such as shape comparison and indexing are reported.

2. Projectively Invariant Description of Disjoint Curves

In this section we express invariant relationships between multiple disjoint curves. A representation is derived and used for further experiments. Special attention is paid to guaranteeing projective invariance at each step of the representation construction.

2.1 Building multi-curve descriptors

In order to represent geometric arrangements of multiple planar curves, one needs to represent relations between points on those curves. Let us suppose that each curve c in an image has its associated length parameter t and each of its points is defined as a point vector $\boldsymbol{c}(t)$ for some value of $t \in [0,1]$. Let us take N_c such curves into account with one point per curve, so that each point is allowed to move freely along its corresponding curve. The dimensionality of the representation space for the relationship between those points will be . In the case of three such curves, the 3D parameter space is already too large to search for relationships between the curves and to extract invariants.

This dimensionality can be reduced by imposing some constraints on the free points taken from different curves. The simplest such relationship is collinearity of points. Any two points from two curves uniquely specify one line and therefore all other points are uniquely defined with respect to this line. So, with the collinearity condition, the dimensionality of the representation space is two, whatever the number of curves. It should be noted that collinearity is a projectively invariant condition.

As the number of curves in an image approaches a few hundred, two-dimensional descriptions for each pair of curves are still not a promising approach. We can reduce this description space to one dimension by constraining the line to pass through one point (e.g. C_0 in figure 2.a). By selecting this point as one extreme of the line, one obtains a one-parameter family of rays, uniquely defined by their angle of orientation θ

For each ray $r(\theta)$, we can detect its intersection points P_1, P_2, ... with all image curves (cf. figure 2.a). It is now possible to characterize this set of points with some function and plot this function against the parameter θ. This provides a "signature" $f(C_0)$ for any choice of the origin C_0, which is based on *multiple* curves and describes information about their spatial arrangement. The number of rays N_r cast from C_0 over the interval $[0, 2\pi]$ defines the signature's resolution, and can be defined a priori.

This representation scheme will be of interest only if it guarantees projective invariance of the signature. For this to be true, all stages of the signature construction method should be projectively invariant. Collinearity of intersection points is already so. The invariance of the position of the centre point C_0 is provided by construction methods addressed in Section 3. Also, the way rays are cast from C_0 should be invariant. This is equivalent to the invariance of the parametrization $r(\theta)$, and is studied in the following subsection. Subsection 2.3 then describes how to obtain an invariant value for a set of intersection points on the ray, and how to construct a shape signature.

2.2 Reparametrization of rays

Let N_r be a total number of rays originating from C_0. A point on the ray r_i is characterized by a homogeneous vector $P_{\alpha_i} = [l_i c_{\alpha_i} \;\; l_i s_{\alpha_i} \;\; 1]$ where l_i is the position along the ray, and c_{α_i} and s_{α_i} are the cosine and sine of the orientation parameter α_i. A configuration where C_0 is the centre of coordinates and the distribution of orientation parameter α_i is uniform over the 2π interval is referred to as *canonical*.

In the example of figure 2.b a family of $N_r = 30$ rays is in its canonical coordinate system, where the orientation parameter α is equal to the polar angle. In figure 2.c a projectively transformed version of these rays is presented. This configuration will be referred to as *image* coordinate frame and corresponds to the unknown projective transformation of the canonical frame. C_0 corresponds to the invariant point detected in the image.

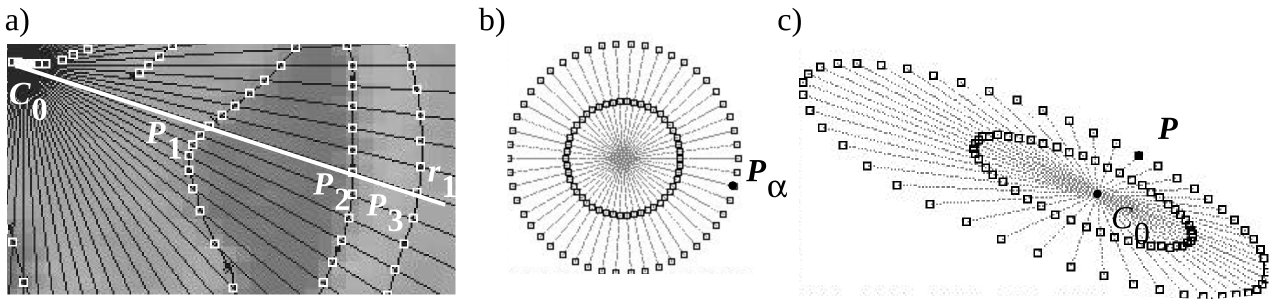

Figure 2 Intersection of image curves with rays originating from a point C_0. White squares represent the intersections points. Canonical coordinate frame and its projectively transformed version.

Let θ be a new orientation parameter which now describes the unknown, projectively distorted, distribution of rays in the image. This non-uniform distribution has to be compensated for projective transformation or, in other words, a correspondence between the canonical and the projected rays should be found.

Let $M = (m_{ij})$ denote the 3x3 matrix of the unknown 2D projective transformation from canonical to image frame. This matrix is expressed in homogeneous coordinates up to a scale factor which can be fixed by setting its m_{33} element to 1. The transformation thus has eight degrees of freedom (DOFs). As already pointed out, C_0 is by construction an invariant point, detected in the image and thus known. Denoting its homogeneous coordinates as $[c_x \ \ c_y \ \ 1]^T$, it implies that its pre-image in the canonical coordinate system is the centre or zero vector $\mathbf{0}^T$. Applying M to the zero vector and writing this correspondence of centres in a matrix form gives:

$$M\mathbf{0}^T = \begin{bmatrix} m_{13} & m_{23} & 1 \end{bmatrix} = \begin{bmatrix} c_x & c_y & 1 \end{bmatrix} = C_0 \tag{1}$$

This equation directly gives two elements of the matrix M which, after their substitution, leaves six DOFs (let us denote the new form by M').

Since ray orientations can be described by the tangents of the corresponding angles, we need to find a correspondence between tangents in canonical and projected frames. The question is what amount of information is necessary to establish such a correspondence. Although in the canonical frame the orientation is already given by the angle α, in the image system the parameter θ has to be determined. A point P_α in the canonical frame is transformed to the point P (cf. figure 2.b) by applying M' to P_α and taking the affine coordinates. The polar version of the point P with respect to the centre C_0 is thus the vector C_0P denoted by:

$$P_\theta = P - C_0 = (M'P_\alpha)_{aff} - C_0 \tag{2}$$

Polar coordinates with respect to C_0 are characterized by their orientation θ and position k along the ray, therefore P_θ can be written as $[kc_\theta \ \ ks_\theta \ \ 1]^T$ where $c_\theta = \cos\theta$ and $s_\theta = \sin\theta$. To obtain the tangent t_θ of the ray orientation, the ratio between its second and first coordinates should be taken. Doing this with the right side of eq.2 and rearranging terms gives the following expression for the tangent in image space, which does not depend on l and k:

$$t_\theta = \frac{s_\alpha(m_{22} - c_y m_{32}) + c_\alpha(m_{21} - c_y m_{31})}{s_\alpha(m_{12} - c_x m_{32}) + c_\alpha(m_{11} - c_x m_{31})} \tag{3}$$

Coefficients c_α and s_α are the unknowns of this equation. Dividing numerator and denominator by $c_\alpha(m_{21} - c_y m_{31})$, we obtain an expression for the tangent t_θ of the image ray:

$$t_\theta = \frac{t_\alpha + u_1}{t_\alpha u_2 + u_3} \tag{4}$$

where t_α is the tangent and u_i are the new unknown coefficients which depend on the elements of M and C_0. The consequence of this expression is that in order to obtain an orientation correspondence between rays in canonical frame and rays in the image we need to resolve three unknowns. With one expression (such as eq.4) per ray, this means that *three* reference rays are needed to establish the full correspondence.

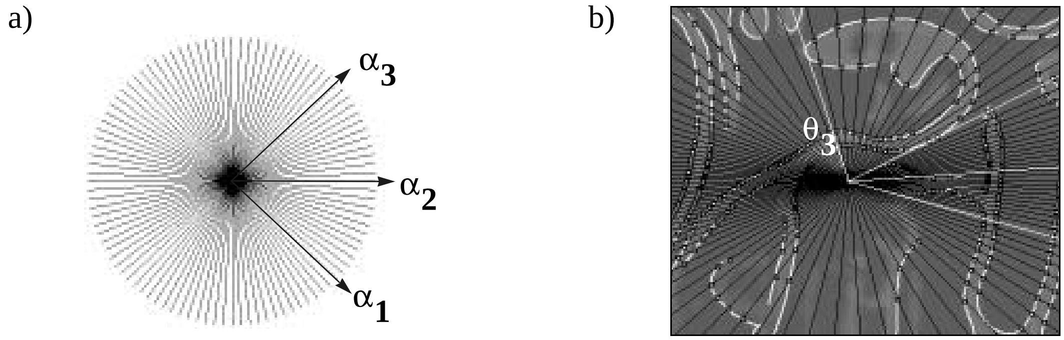

Figure 3 Canonical (a) and image (b) coordinate frames with reference and sampling rays. In (b), the fourth ray indicates the symmetry frontier. The sampling rays (black) follow the projective transformation determined by the reference rays θ_1, θ_2, θ_3.

Taking three rays with predefined orientations α_1, α_2, α_3 in the canonical frame and three corresponding rays invariantly identified in the image space with orientations θ_1, θ_2, θ_3 will give three equations of the type of eq.4. Solving them for the u_i and making substitutions in the general form of the equation, we obtain an expression relating any image orientation θ and canonical orientation α for any ray.

In practice, let us take the canonical reference orientations as $[\alpha_1, \alpha_2, \alpha_3] = [-\pi/2, 0, \pi/2]$ (cf. figure 3.a) which correspond to tangents $[t_{\alpha_1}, t_{\alpha_2}, t_{\alpha_3}] = [-1, 0, 1]$ (interval of highest tangent stability). Substituting them into the expressions of u_i, we obtain the mapping between the tangents t_α and t_θ in the canonical and image frames:

$$t_\theta = \frac{(t_{\theta_3} t_{\theta_2})(t_\alpha + 1) + (t_{\theta_1} t_{\theta_2})(t_\alpha - 1) - 2 t_\alpha t_{\theta_3} t_{\theta_1}}{(t_{\theta_3}(1 - t_\alpha) - t_{\theta_1}(t_\alpha + 1) + 2 t_\alpha t_{\theta_2})} \tag{5}$$

Once this correspondence is established, we construct rays with a uniform distribution in the canonical frame and transform them to the image frame with the above formula (cf. figure 3.b). All rays so defined in the image space are projectively invariant with respect to the reference rays. They are fully invariant provided that reference rays were invariantly identified. It should be noted that working with tangents in eq.5 provides correspondence only up to the central symmetry. This direction ambiguity is removed during ray construction. As will be shown later, at least one point is available on one reference ray, thereby allowing the selection of the positive direction.

It is interesting to note that exactly the same conclusions about reference rays could also be obtained in the framework of dual cross-ratio reasoning [13]. In fact, a cross-ratio of four concurrent lines is an absolute projective invariant (constant), so the orientation of the fourth line can be expressed as a one-parameter expression in the orientations of the other three.

To summarize, we now possess a method for projective normalization of ray orientations from an invariant point, given three reference rays. This normalization has removed five DOFs from the projective transformation matrix. In the next section, we will concentrate on how to resolve the three remaining DOFs by attributing an invariant value to each ray.

2.3 Calculating the signature

In this section we show how, given one ray and some points of intersection with image curves, it is possible to find a projectively invariant measure for a subset of such points. Each point on the ray is a one-dimensional entity. With three DOFs remaining, three points are needed to eliminate them, and one extra point to obtain an invariant value. Indeed, this is the case of an unknown projective line. A well-known projective invariant on such line is a cross-ratio based on four points [19,3,13]. By taking the centre point and the *first* three other points on one ray, one can compute their cross-ratio, providing an invariant value for *that* ray. Using the notation of figure 2.a the cross-ratio will be calculated as:

$$cr(r_1) = (|\boldsymbol{C_0P_2}||\boldsymbol{P_1P_3}|)/(|\boldsymbol{P_1P_2}||\boldsymbol{C_0P_3}|) \tag{6}$$

where $|\boldsymbol{xy}|$ denotes the distance $\|\boldsymbol{x}-\boldsymbol{y}\|$ or the determinant of corresponding homogeneous coordinate vectors. It is now clear that only the three closest curves to the point C_0 will determine the points P_1, P_2, P_3 selected for each ray. This is an attractive property because our signature will be based on multiple curves, expressing their relative position. At the same time, this signature will remain local, without going beyond the three closest curves.

Projective invariance is now achieved. To construct a signature, we take in the canonical frame N_r uniformly spaced rays and transform them, with the help of the reference lines, to the image domain. For each ray obtained, a cross ratio of three intersection points gives the signature value. In practice, cross-ratio values are bound. If curves can not be closer that d_{min} pixels due to edge detector properties and the image size does not exceed d_{max} pixels, then the upper bound for the cross-ratio is: $cr_{max} = (d_{max}-d_{min})^2/(4(d_{max}d_{min}))$. With $d_{min} = 3$ and $d_{max} = 600$ the cr_{max} we have $cr_{max} = 50$ which can be used as a normalization factor for signature comparison. When the number of intersections is less than three, the signature value is undefined and arbitrarily set to zero.

To compare the signatures of two patterns m and n, we need a matching measure. Let $s(\alpha)$ be the value of a signature for the orientation α (which corresponds to $cr(r_\alpha)$). In order to stress the suitability of the signature itself, we consider the simplest version of matching function between signatures s_m and s_n given by the normalized sum of Euclidean distances across all rays:

$$d(s_m, s_n) = \frac{1}{cr_{max}N_r}\sum_{\alpha=0}^{N_r-1}\|s_m(\alpha)-s_n(\alpha)\| \tag{7}$$

where $\|x\|$ denotes in this case the absolute value but could be extended to the L_2 norm for multidimensional signatures (cf. section 5).

The following example illustrates the invariance of these signatures. In figure 4 (a) and (b) the same group of curves is viewed from two different viewpoints (they are projectively equivalent). For both images, one invariant point and three reference rays (gray) are shown. In this case, $N_r = 100$ and the two corresponding signatures are shown in figure 4.c. Their normalized difference, according to eq.7, is 003. More extensive tests with variations of distance under projective transformation are presented in the experimental results section.

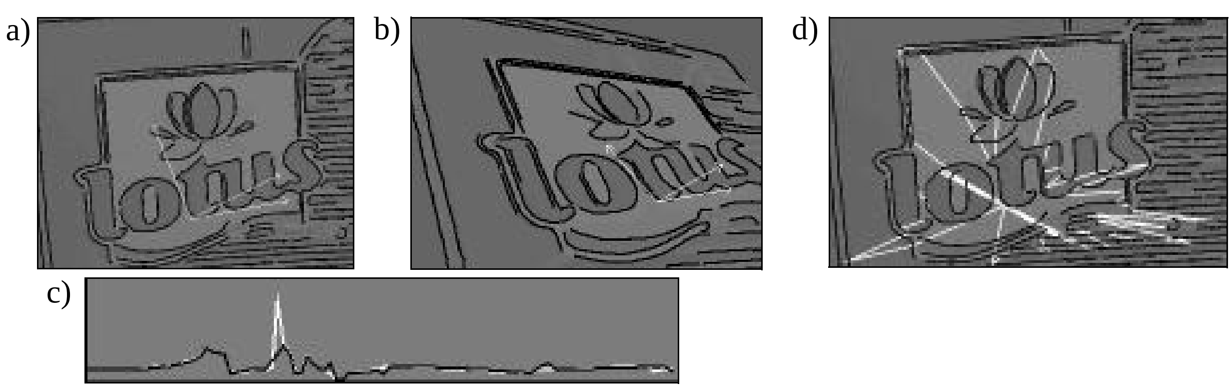

Figure 4 Projectively equivalent shapes (a), (b). Example of signatures (c) computed from these shapes. In (d) The “cover zone” of the signature in the original image is shown.

It should be noticed that for the simpler case of affine projection the number of DOFs is six, which is two degrees less than in the projective case. If we remove one DOF from rays and one from point cross ratio along each ray, we obtain that, for the affine case, two reference rays are sufficient together with only three points on the curve. This simpler case is not further studied in this paper.

Given a triplet of lines representing the reference frame, it is possible to define a region of the image whose perimeter is formed by the last curves participating in the signature (i.e. the third curve for each ray). Figure 4.d. contains one example of such region called "cover zone". Two interesting observations can be made from this example. First, the signature remains local while spanning multiple curves. Second, attention should be paid to gaps in some curves producing unpredictable variations in signature values. This also suggests a possible way to improve our definition of distance between signatures, namely the possibility to disregard *small* intervals of α where two signatures clearly diverge.

3. Construction of invariant reference frames

In the previous section we have shown that exactly three projectively invariant lines passing through one point are necessary to build an invariant signature for such a point. The present section addresses the issue of constructing these reference lines from such invariant curve properties as points and tangents.

3.1 Construction of new lines

Projectively invariant properties of a curve include points and straight lines. Points on the curve are projectively invariant if they are cusps, inflections or bitangent points of contact. A straight line, given either by a bitangent line, inflection tangent or by a piece of straight curve is also projectively invariant [24]. Due to the relatively high instability of cusps, we restricted our interest to bitangents, inflections and straight pieces of curves. Moreover, for bitangents and inflections either tangents or points can be used but tangents are used first whenever possible because of their higher stability [19, 32].

Unfortunately, none of these properties has a configuration where three lines meet in one point. At most we have one point and one line (bitangent, inflection). Therefore, different invariant properties should be combined to build a frame. If we start from one invariant point, we need to add the missing lines, whereas if we start with one existing tangent, one point and more lines need to be constructed. By taking the intersection of *two* tangents of invariant properties on the curve, one obtains an invariant point C_0 and two lines. The third line cannot be produced without extra information and so a third curve property should be considered. In this case linking one further invariant point and C_0 by a line would complete the construction.

In order to reduce the number of combinations, the grouping operation underlying the construction of invariant frames should respect the order of invariant components along the curve. All invariant lines are associated with some points on the curve. Bitangents have two points of contact with the curve and can be considered as two separate points with equal tangents (of course their intersection will in this case be avoided). The straight part of a curve can be approximated by a line segment. For grouping purposes, its two endpoints can be considered as points of contact for this line. All invariant properties of one curve are thus ordered and their successive triples can be used for frame construction.

It should be noted that the unknown direction of a curve still leads to an ambiguity about the global order of points, i.e. the same frame should be obtained if the order of points in the triple is reversed. To achieve this in a local fashion, we suggest to construct the centre point C_0 as the intersection of tangents of the two *external* points of the triple. The third line would then pass through C_0 and the *middle* invariant point.

Let us take, for example, a triple of invariant points, such as the bitangent point B_1 and the two inflections I_1, I_2 of figure 5.a with their respective tangents b_1, l_1, l_2. Taking the intersection of tangents from the first and third points (B_1 and I_2) gives the centre point C_0. The third line is passed through C_0 and the middle point in the triple which is I_1. The order between the three constructed rays is selected in correspondence with canonical rays and becomes the following: b_1, l_2, l_1.

a) b)

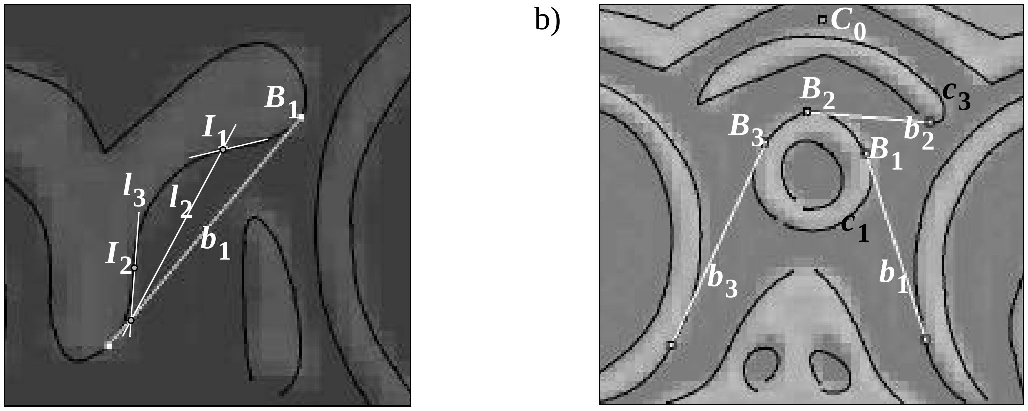

Figure 5 Three-line configuration constructed from a bitangent and two inflections (a). A bitangent b_1 and two inflections I_1, I_2 are used to construct a reference frame of three lines: b_1, l_2, l_3. Fitting common bitangents to multiple curves (b) and further construction of a reference frame.

The constructive approach described above is a general method for constructing reference frames by selecting successive triples of invariant properties. The only exception is the particular case when a straight line is the middle invariant property in the triple. Fortunately, this specific situation is compensated by one important advantage of the whole approach. Indeed, taking the intersection of *external* (rather than neighbouring) points in the triple places the centre point distinctly outside of the curve. This prevents points on the ray from being too close to the centre point and thus produces a more distinctive signature pattern in each configuration. Indeed, if the intersection of two *neighbouring* tangents was taken, the centre point would often lie almost on the curve. One of the distances in the cross-ratio calculation would then become zero and the signature value be identically one for a whole range of orientations.

3.2 Bitangents of multiple curves

As mentioned above, curves play the role of grouping operator for invariant points. However, practice shows that the topology of curves in the image is affected by perspective projection and image noise. In curve zones where a particular projective transformation increases the curvature, a potential gap can be expected because of the fixed geometry and finite resolution of edge detectors. Thus, curve topology depends on the transformation and can not be relied on for grouping remote invariant properties. To overcome this problem, more invariant properties are needed to increase their density along the curves. We make the assumption that within a local neighbourhood curved contours belong to the same object and so are coplanar. In the case of trademarks, curved contours rarely correspond to 3D edges and we expect this hypothesis to hold. Quantitatively, this assumption depends on the number of planar facets in the scene and on the number of curves belonging to each facet. To validate it experimentally we have found that for our database of trademarks (cf. section 5) approximately 4% of neighbouring curve pairs do not belong to the same object.

If neighbouring curves do belong to the same rigid object, their *joint* projectively invariant properties can be used. In this case, only bitangents are suitable since they have two-point contact and so can be fitted to a pair of curves. For each curve a subset of neighbouring curves is thus constructed and bitangents are fitted to them. We impose the condition that such bitangents do not intersect other curves so as to keep properties local.

Figure 5.b illustrates this stage. The curve c_1 does not have any invariant points of its own; therefore no invariant frame could be found for it. However, several invariant properties can be found in common with its neighbours c_2, c_3 and c_4, such as the three bitangents b_1, b_2, b_3. These lines are suf-

ficient to construct at least one reference frame for c_1. Taking the intersection of b_1 and b_3 produces the centre point C_0 and the third line would pass through B_2.

We statistically estimated the advantage of using multi-curve properties for signature construction with respect to methods based on a single curve. For our database we found an average of 0.19 bitangents, 0.62 lines, and 0.17 inflections per curve. This gives a total of 0.98 invariant properties per curve while the number of triplets of these properties, necessary for frame construction is on average below 0.24 for each individual curve. However, if we consider bitangents spanning two curves, their occurrence per curve is 1.7 and the average number of triples increases to 0.76. This affects the density of reference frames in the image making it high enough to not only cover the whole object with invariant descriptors, but also to provide sufficient level of redundancy to deal with noise and occlusion.

3.3 Extension to 3D faceted objects

The construction of the invariant signature described above has been defined for shapes such as trademarks located on a planar surface. However, trademarks are often placed on pack boxes that have orthogonal sides. If a box corner is visible from the camera, two or three facets are visible simultaneously. Trademarks located on each facet can be represented independently, but in this case the integration of the information from different facets would also be of considerable interest. In this section we address the issue of finding a reference frame of three rays for *each* facet using the assumption of facet orthogonality.

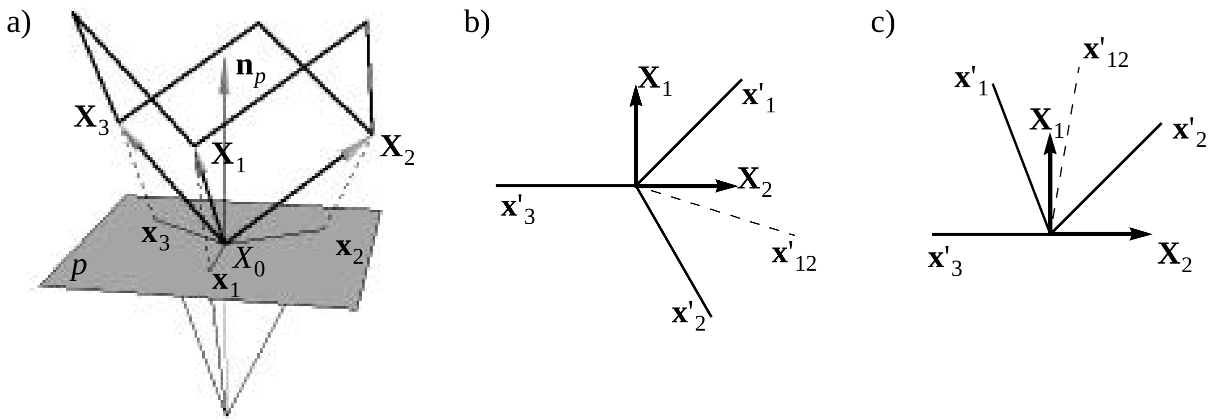

Figure 6 (a) Projective configuration for the case of an orthogonal corner with three visible facets (seen from C_0). (b) Three rays configuration recovered for the facet $\mathbf{X}_1\ \mathbf{X}_2$. (c) Three ray configuration reconstructed for facet $\mathbf{X}_1\ \mathbf{X}_2$ with correct definition of the midray

Figure 6.a illustrates a homogeneous projective configuration, corresponding to a visible orthogonal corner. In this case p is the projective plane (image plane) and C_0 is the optical centre. The corner point X_0 will be selected as coordinate centre for convenience. Let the basis unit vectors $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ lay on 3D corner edges. For now we will study the case when all the three facets are visible.

The vectors $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ will project onto vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ on the plane p. As we are interested only in *rays* corresponding to the edges of the corner, only the orientation of the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ is important, and not their length. This corresponds to an arbitrary depth of the corner in the scene or, equivalently, to arbitrary position of the projective plane along the line C_0X_0. Thus we can consider that it passes through X_0 and it is defined by its normal $\mathbf{n}_p = [a, b, c]$ in the $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ frame.

Let us take the facet spanned by $\mathbf{X}_1$ and $\mathbf{X}_2$ as an example (cf. figure 6.a). As we saw in section 3, ray normalization in the plane requires three reference rays. The two vectors spanning the facet already give two such rays and so we need to find the third one. One excellent candidate is the bisector of the angle between $\mathbf{X}_1$ and $\mathbf{X}_2$ because the triple of rays will then correspond to the canonical frame $[\alpha_1, \alpha_2, \alpha_3]$ defined in section 2.2. Since $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal and of unit length, their bisector is spanned by the vector $\mathbf{X}_1 + \mathbf{X}_2$. Let $\mathbf{X}_{12}$ denote this vector and $\mathbf{x}_{12}$ its projection on the plane p (not shown on the figure). In the following, we shall show that orthogonality of facets imposes a rigidity constraint on the orientation of three rays and exploit this to derive a closed-form expression for the orientation of $\mathbf{x}_{12}$.

The vector $\mathbf{x}_1$ which defines the ray of projection of $\mathbf{X}_1$ on p is defined by:

$$\mathbf{x}_1 = (\mathbf{n}_p \times \mathbf{X}_1) \times \mathbf{n}_p \tag{8}$$

and the same formula applies to three other vectors $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_{12}$. By definition, all these vectors lay on the p plane in space. Let x_i denote the *orientation* (tangent) of any such vector $\mathbf{x}_i$ in the projective plane with respect to some basis, we then need to find x_{12} from x_1, x_2, x_3. Together with x_1, x_2, the x_{12} will become the third reference ray for the facet $\mathbf{X}_1\mathbf{X}_2$ and will complete the construction of projectively invariant frame described in the previous section. The same reasoning applies to the other two facets.

The orientation x_i of a vector $\mathbf{x}_i$ lying in the projective plane, can be measured only with respect to a selected basis in this plane. The coordinates of vectors $\mathbf{x}_i$ are already expressed with respect to the three unit vectors $\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$. Keeping the same basis we rotate the projective plane p together with the four orientation vectors $\mathbf{x}_i$ around the centre X_0 so as to align it with one facet. Selecting, for example, the one spanned by vectors $\mathbf{X}_1$ and $\mathbf{X}_2$, these latter vectors become the basis of the transformed plane (cf. figure 6.b). So, we can use coordinates of the transformed vectors to calculate their orientations (tangents).

In practice, the rotation of the plane p can be sought as a rotation of its normal $\mathbf{n}_p$ so that after the transformation the normal is aligned with $\mathbf{X}_3$. This rotation can be decomposed into two rotations. The first, denoted by R_{X_3}, rotates $\mathbf{n}_p$ around the vector $\mathbf{X}_3$ and brings it into the plane spanned by $\mathbf{X}_2$ and $\mathbf{X}_3$. The second, denoted by R_{X_1}, rotates $\mathbf{n}_p$ around the vector $\mathbf{X}_1$ to finally put the normal onto $\mathbf{X}_3$. The matrices of the transformations R_{X_3} and R_{X_1} are easily expressed in coordinates of $\mathbf{n}_p$ only (since all rotations are defined by this vector position). Multiplying two matrices together give the final transformation matrix in unknown coordinates of $\mathbf{n}_p$:

$$\begin{bmatrix} b/l & -a/l & 0 \\ ca/l & cb/l & -l \\ a & b & c \end{bmatrix} \tag{9}$$

where $l = \sqrt{a^2 + b^2}$.

Applying this transformation (eq.9) to four vectors $\mathbf{x}_i$ in the projective plane defined by (eq.8) we obtain new vectors $\mathbf{x}'_i$ all belonging to the $\mathbf{X}_1\mathbf{X}_2$ plane (cf. figure 6.b). Taking the ratio of the *first* and *second* coordinates of each vector gives a tangent for each ray as follows:

$$x_1 = b/(ca) \qquad x_2 = -a/(cb) \qquad x_{12} = (b-a)/((b+a)c) \tag{10}$$

By construction, $\mathbf{x}_3$ is aligned with the $\mathbf{X}_2$ axis and so its orientation x_3 is zero. By rearranging terms and using the fact that $\mathbf{n}_p$ has unit length we obtain an expression for x_{12} in x_1 and x_2 only:

$$x_{12} = \frac{k(x_1 - k)}{x_1 + k} \text{ where } k = \sqrt{-x_1 x_2} \tag{11}$$

The first observation is that x_{12} depends only on two rays spanning the facet it belongs to. This is true as long as $\mathbf{x}_3$ is aligned with the horizontal axis and the order between x_1 and x_2 is correct. Selecting the correct order, however, can not be done in the absence of x_3 and should be done in the *clockwise* direction as illustrated in figure 6.b. A second remark is that x_1 and x_2 should be of different sign. This condition is a consequence of the rigidity imposed by the orthogonality of facets and it is always satisfied when $\mathbf{x}_3$ is aligned with the horizontal axis. So, if we find a Y-junction in the image, we align one ray with the horizontal axis and evaluate the midray for the other two according to the proposed formula.

Let us consider now the case when only two facets are visible (). Rays projected into image plane are shown in figure 6.c. This case differs from the previous one by the fact that all pairwise angles between rays in the image plane are less than $\pi/2$ and thus can be easily detected in the image. The same rotations are applied to align $\mathbf{x}_3$ with the horizontal axis. However, two other rays, due to the rigidity constraint, are now placed on the opposite sides of the $\mathbf{X}_1$. Though, the expression in eq.11 will give a correct midray orientation only when orientations of $\mathbf{x}_1, \mathbf{x}_2$ rays are measured with first and second coordinates reversed.

Let us take figure 7 for a practical example. A box corner can be detected in the image by searching for Y-junctions of lines. Three rays were detected, shown in the figure 7.a. For the S facet a bisector was detected according to the proposed method and a signature evaluated. This operation was also performed with a different view of the same box, shown in figure 7.b. Again, a signature was evaluated and a comparison between the two is provided in figure 7.c. It can be seen that except for few points, the signature profiles match rather well.

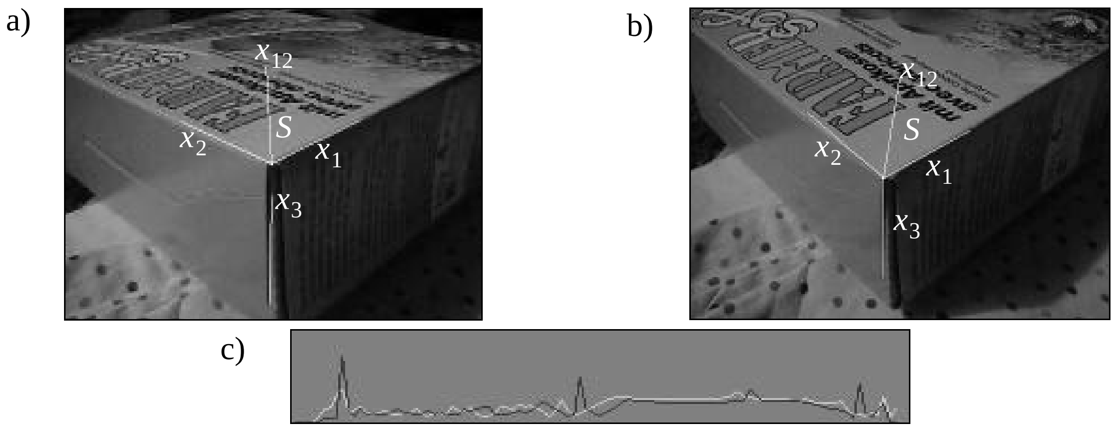

Figure 7 (a) A bisector ray found for the facet S. In (b) the same ray is found for an image of the same corner viewed from another viewpoint. In (c) the signatures constructed from the two corresponding facets are shown.

To summarize, we have shown that, even in the 3D case the projective normalisation with lines allows to represent trademarks with sufficient precision. It should be noticed that, unlike the planar case, in 3D signatures for all facets correspond to three $\pi/2$ intervals in the canonical frame. There is no circular order for these intervals. A comparison technique that takes the best distance over 3 circular permutations of intervals should thus be considered.

4. Illumination invariance for indexing

The method presented above for computing a pattern signature is purely geometric. In order to increase its discriminating capability we propose to add chromatic information to the signature. In line with the whole approach, we shall try to obtain invariance to illumination changes. Several models exist to describe chromatic changes under illuminant variations [5,7]. Invariance to illumination is then possible to seek as invariance to a specific transformation model.

One of the optimal approximations is the scaling model [5] where, under illuminant change, each colour channel changes its intensity according to a separate scale factor. In this case chromatic values measured one point under one illuminant $[R \;\; G \;\; B]$ change to $[R' \;\; G' \;\; B']$ according to the following expression:

$$\begin{bmatrix} R' \; G' \; B' \end{bmatrix} = \begin{bmatrix} s_R R \; s_G G \; s_B B \end{bmatrix} \tag{12}$$

Let us assume that two neighbouring pixels 1 and 2 belong to the same surface. Due to their proximity we consider them as subject to the same illuminant. In this case, the following relation [7] allow the scaling factors of the previous expression to be discarded: $(R'_1 G'_2)/(R'_2 G'_1) = (R_1 G_2)/(R_2 G_1)$.

Such ratios are therefore locally invariant to illumination. In chromatically uniform image areas, this ratio should be approximately constant. The disadvantage of this method is that local changes of colours occurring at the border of two surfaces result in large variations of this ratio, making recognition unstable.

In our case, we have an invariantly constructed ray with four points. It would be interesting if we could complement the geometric information represented by their cross-ratio with a more stable chromatic measure computed on *intervals* between such points. Since curves in an image correspond to chromatic variations, the areas they enclose tend to be more uniform or textured and can thus be well described by simple functions. By analysing the profiles of three colour channels along one of the rays we can see an example of typical global and local variations of intensity (cf. figure 8). Global variations come from the illuminant while local ones originate from object-specific chromatic variations.

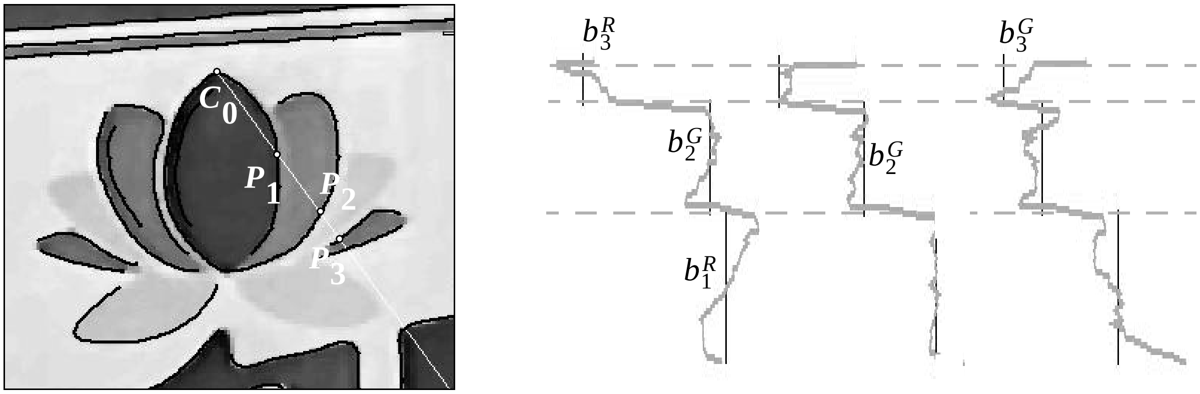

Figure 8 (Left) Example of radius in input image; (right) radius profiles for red, green, blue channels.

The simplest method to model their variations is to take averages over the profile between points of intersection and work with these values to find a possible invariant to illumination. Such an illumination invariant value can be stored with each geometric signature point and used as an additional dimension for discrimination.

Let $y = b_L^F$ denote the average value of the part of the chromatic profile under some canonical illuminant. Here b is simply the average over the interval, the index $L = \{1, 2, 3\}$ indicates the interval and the index $F = \{R, G, B\}$ indicates the chromatic channel. Under a change of illumination each point in the interval will be subject to a vertical scaling with a factor s_R, s_G, s_B corresponding to the chromatic channels R, G, B. The averages, which are in fact fitted horizontal lines will exactly follow this transformation. For instance, the equation of the first interval of the red profile $y = b_1^R$ will become $y = s_R b_1^R$.

Due to the projective transformation the above line equation is also subject to additional changes. The projective transformation along the ray modifies the *density* of points in the whole interval. For the operation of averaging this amounts to weighting differently chromatic values of each point over the interval. The net effect on the line position is a vertical displacement that can be modelled also by a scale factor a_L. So, the modified equation of, for instance, the first interval of the red profile finally becomes $y = s_R a_1 b_1^R$ whose right part can be denoted as b'^R_1. Given nine lines equations, three unknown chromatic factors, three unknown transformation parameters a_L and one line parameter y leaves three independent invariant values. Taking ratios so as to eliminate all parameters, one obtains, the following three expressions:

$$b'^R_2 b'^G_1/(b'^R_1 b'^G_2),\; b'^G_3 b'^B_2/(b'^B_3 b'^G_2),\; b'^R_1 b'^B_3/(b'^R_3 b'^B_1). \qquad (13)$$

These can be used to characterize the signature from a chromatic point of view in addition to the geometric invariant descriptors. Overall, the proposed invariant signature consists of N_c vectors containing the cross-ratio of points detected on each ray, plus three chromatic invariants.

5. Experimental Results

In this section, we first test the stability of the proposed invariant representation under different types of image noise. Second, we assess its usefulness for image database applications with standard performance measures used in information retrieval.

A database of 203 images of 41 planar objects (c.f. figure 9 for a few samples) was collected using different acquisition devices (camcoder and two digital cameras). Images were taken from different viewpoints under various illumination conditions (daylight, neon/bulb lamp). The signature extraction process was run fully automatic and produced an average 40 valid signatures for each image. For each image we computed the cover zone (cf. section 2.3) of all its signatures which, on average, amounts to 2.1 times the image surface. The average overlap is thus 50% of the cover zone.

Figure 9 Fifteen typical images from the database. The last row features faceted objects

Separate tests were conducted for faceted objects. A database of 170 corner views of 53 boxes was collected under the same conditions as described above. Retrieval tests are presented at the end of this section.

5.1 Stability of the invariant representation

The construction of the invariant representation can be divided into three steps: curve detection, extraction of invariant properties, grouping and signature evaluation. The stability of each step is estimated with respect to “image noise” produced from various sources. These include viewpoint transformation and resolution changes. The latter can be modelled by a scaling transformation while viewpoint change can be approximated by a general projective transformation (cf. section 2.2). Illumination changes are produced by different lamps, and their effects are only estimated with respect to database retrieval (cf. section 5.2).

In order to make curve extraction less sensitive to changes in resolution, scale and illumination, we use a multiscale edge detector [14] on the RGB colour planes. In this way we considerably reduce curve gaps. Furthermore a multiscale approach prevents from detecting spurious curves as the resolutions increases. Because of this multiscale analysis the edge detector cannot separate two curves if they are less than 5 pixels apart.

The scaling range that a shape can withstand depends on the smallest distance between its curves with respect to its full size. Let r be such a ratio and s be the image size (maximum camera resolution in pixels). It is straightforward to express the maximum resolution reduction after which the closest curves can still be discriminated, which is: $sr/5$. Given typical values, such as $r = 0.06$ and $s = 512$ the maximum scaling factor allowed for full-image objects amounts to 6.1 and will be used as a reference scale for resolution stability tests.

A similar reasoning can be made about the allowed range of the projective transformation (change in viewpoint). In this case, for the same viewpoint position, remote parts of the object are subject to stronger contraction. Thus, distance reduction depends not only on the transformation parameters, but also on the image position of the point to be transformed. To quantify this reduction, a value that combines both parameters and position should be used. For this purpose we use “homogeneous depth” i.e. the value of the third homogeneous coordinate after the transformation M (cf. section 2.2).

The depth for the frontal view of the object is equal to 1.0 and under any projective transformation the maximum reduction will occur at object corners. For the same values of r and s introduced above, the average maximum distance between closest curves at highest resolutions is 15 pixels. If two curves, separated by this distance are found in the corner of the image, the distance between their transformed versions can be expressed as a function of the “depth”. The upper bound for the range of allowed projective transformations is equal to 1.68 (homogeneous coordinates). This value is obtained by setting the obtained distance equal to the minimal allowed distance of 5 pixels and by solving for the "depth". For the real images of the database the viewpoint position with respect to the object was unknown and so the depth is estimated by recovering transformation with respect to a reference frontal image.

Next, we evaluate the robustness of the recovery of invariant properties. As long as curves are detected, the detection of bitangents, inflections and lines presents no major problems. However these properties exhibit different degree of numerical stability, as can be seen in figure 10.a for scaling and in figure 10.b for the projective transformation. Each graph represents the proportion of detected features (manually verified a posteriori) with respect to ideal situation, averaged over the database. It can be seen that up to 75% of the allowed transformation range we still obtain 80% of the same invariant properties. In figure 10.c we show an example image of an object taken from a viewpoint of the extreme of the allowed interval.

Finally, we consider the stability of the grouping and signature construction process. The grouping operation is clearly sensitive to curve gaps. The decreasing number of detected reference frames as a function of viewpoint transformation is also shown in figure 10.b (triples). This can be explained by the fact that extreme viewpoints increase the number of curve gaps at high curvature points.

By definition, the stage of signature construction itself is not sensitive to gaps in curves (cf. section 5). These might cause a change in the signature values only within limited intervals and their influence on the distance between signatures can be neutralized by the use of robust estimators [11]. Nevertheless, the presence of spurious curves can undermine a large part of the signature. That is why the *same* curves should be detected when viewed from different viewpoints. This is achieved by the use of multiscale detector, as illustrated by variation of the proportion of curves in figure 10.b.

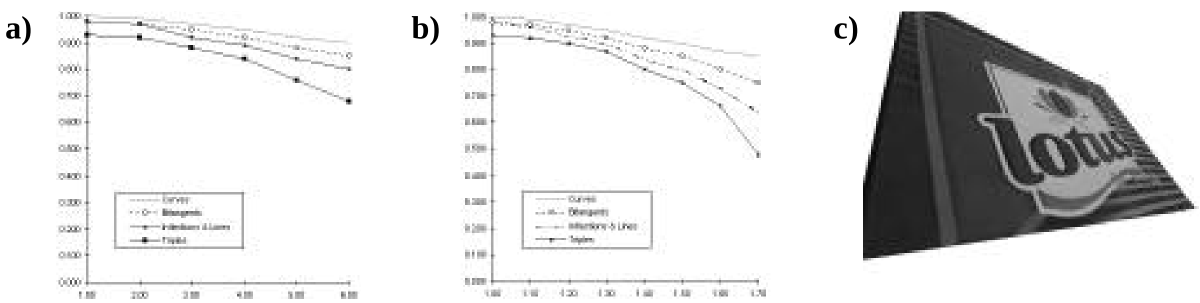

Figure 10 Robustness of the signature construction process. Proportion of detected features as a function of resolution changes (a) and projective transformations (b). (c) Example of the extreme projective image transformation withstood by the method, for a typical shape.

For the case of orthogonal facets, reference frame detection is greatly simplified. Detection of lines is facilitated by specular reflections on box edges, by different illumination conditions for each facet (higher contrast on the edge) and finally by the relatively long edges of the box that are hardly subject to projective distortion. The grouping operation is performed by selecting line triples and by verifying that three criteria are satisfied. First, three lines rarely intersect at one unique point, but rather form a "triangle of intersection". Therefore, the surface of this triangle should be small. The second condition is on the orientation of rays. Orthogonality of facets imposes a condition on the rays orientation that should be satisfied. Finally, lines often do not reach the "corner" point, introducing gaps between their endpoint and the corner point. Therefore the third constraint imposes that the sum of these gaps should not exceed a fixed percentage of the three lines total length.

5.2 Evaluation of the content-based retrieval capabilities

In this section we present two types of experiments in order to assess the suitability of the proposed invariant signature for content-based shape retrieval. Initially, an individual signature is used as a query to the database while in the second stage all signatures, automatically extracted from the same image are used as queries.

In both cases, separate tests with four different subsets of the database are performed. In case one, only one version of each image is included in the database (frontal view). In case two, close views of the same object but under different illumination conditions are considered. In the third case, object views from different viewpoints are included within the allowed range. The fourth case combines the images of the last two.

Shape-based retrieval was performed by pairwise comparison between signatures, using the Euclidean distance on the whole equally-weighted vector including geometric and chromatic information (cf. eq. 8). The search process is thus linear in the database size, although faster approaches can easily be introduced [1]. Given a query signature, the retrieval performance was assessed using standard information retrieval measures on the ranked hits, namely precision and recall. *Precision* is the proportion of correct hits in the set of retrieved items up to the last correct one. *Recall* is the proportion of target images that have been retrieved among the top N hits, $N \geq 1$.

In the first experiment, individual signatures were used for matching with the contents of four datasets. However, only the signatures for which a correct answer exist in the database were used. Table 1 shows the average values of the precision for all signatures of the frontal view, matched against all other signatures in the four separate datasets. Images from which the query was extracted were removed from their corresponding "front-view" dataset.

Table 1: Performance of image querying using individual signatures from frontal views

	Frontal views only	Illumination changes	Viewpoint variation	Both
Precision	69%	65%	58%	56%
Recall (N=15)	78%	73%	65%	61%

On average, most of the correct signatures are among the $N = 15$ top-ranked hits, for all types of allowed transformations. In general, if at least one signature of an object (complex objects may provide several signatures) is detected in the image its discriminative power is sufficient enough to perform shape-based retrieval.

In the second experiment, the same four datasets are used as the database contents, but the way to define a query is different. For all objects in the database, other frontal views (different from those already in the database) are used for signature extraction. All automatically extracted signatures are used for separately querying the database. The $N = 15$ top-ranked hits were retained for each case. By using a simple voting scheme for all signatures of the same object, the rank of each signature was accumulated into an *object* rank. Table 2 shows the precision results using this rank, averaged across all queries.

Table 2: Querying with frontal views

	One frontal view	Illumination changes	Viewpoint variation	Both
Precision	73%	70%	66%	62%
Recall (N=15)	82%	75%	69%	64%

Using the same procedure, we then performed retrieval test for faceted objects. Similar datasets were constructed. Views taken from different points in space but under the same illumination conditions were included in the first set. Views with all conditions allowed to vary were gathered in the second set ("Both"). Retrieval was performed by comparing three signatures of a test image with the whole collection of signatures.

Table 3: Performance of image querying using facet signatures

	Viewpointvariation	Both
Precision	78%	72%
Recall (N=10)	85%	78%

It can be seen, that retrieval performace is more stable and better than in a mere planar case. This stability can be explained by the more robust extraction of reference rays and by the fact that when one facet becomes hardly visible, the others automatically offers a good view to the camera.

6. Conclusions

In this article we have presented an approach for representing planar complex shapes. The proposed representation is projectively invariant and describes the local arrangement of neighbouring curves with respect to the invariant properties of one or move curves. This method presents two major advantages. First, less information is required with respect to previous approaches for one curve to produce a reference frame. Only three concurrent rays are necessary, against four points in a general projective case. Second, the local arrangement of neighbouring curves is incorporated into the description. This makes curves without any invariant properties at all also useful. Both advantages can be exploited to extend the application of invariant methods to a broad class of shapes found in the real-world situations and its extension to 3D faceted objects broadens its field of application to package boxes.

The proposed geometric construction is appropriate for the integration of chromatic information. Together with projective invariance, illumination invariant measures are associated with the shape description. This helps discriminate geometrically similar cases and leads to a more complete object representation.

The applicability of the proposed invariant representation to database retrieval has been validated with statistical tests. The representation maintains, within small variations, the property of projective invariance under reasonable viewpoint changes. At the same time, it allows discrimination among a few hundreds three-ray configurations selected from a database of real flat or faceted trademarks.

7. References

[1] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, A.Y. Wu, An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. Department of Computer Science, University of Maryland, College Park, Technical report CS-TR-3568, Dec. 1995.

[2] S. Carlsson, R. Mohr, T. Moons, L. Morin, C. Rothwell, M. Van Diest, L. Van Gool, F. Veillon, A. Zissermann, “Semi-local projective invariants for the recognition of smooth plane curve”, Intern. J. Comput. Vis. 19(3), 1996, p. 211-236.

[3] C. Coelho, A. Heller, J.L. Mundy, D. A. Forsyth, A. Zisserman, “An Experimental Evaluation of Projective Invariants”, In [56], p. 87-104.

[4] F. S. Cohen, J.-Y. Wang, “Part I: Modeling image curves using invariant 3-D object curve models - a path to 3-D recognition and shape estimation from image contours”, IEEE Trans. Patt. Anal. Mach. Intell., 1994, vol. 16, No. 1, p. 1-12.

[5] G. D. Finlayson, “Color constancy in diagonal chromaticity space”, Intern. Conf. Comp. Vis., MIT, 1995, p. 218-223.

[6] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker (1995). Query by image and video content: the QBIC system. *IEEE Computer*, September, 23-32.

[7] T. Gevers, A. W. M. Smeulders, “A comparative study of several color models for color image invariant retreival”, Proc. 1st Int. Workshop on Image Databases & Multimedia Search, Amsterdam, Netherlands, 1996, p. 17.

[8] A. Goshtasby, “Design and Recovery of 2-D and 3-D shapes using rational gaussian curves and surfaces”, Intern. J. Comput. Vis. 10:3, 1993, pp. 233-256.

[9] R. Hartley, “Projective reconstruction and invariants from multiple images”, IEEE Trans. Patt. Anal. Mach. Intell., 1994, vol. 16, No. 10.

[10] K. Hirata and T. Kato, *Query by Visual Example: Content-Based Image Retrieval*. A. Pirotte, C. Delobel and Gottlob (Eds.), Proc. E. D. B. T.'92 Conf. on Advances in Database Technology. Lecture Notes in Computer Science Vol. 580, Springer-Verlag, 1994, 56-71.

[11] P.J. Huber, “Robust statistics”, Wiley, 1981

[12] R. Jain, “Image databases and multimedia search”, Invited talk, Proc. 1st Int. Workshop on Image Databases & Multimedia Search, Amsterdam, Netherlands, 1996.

[13] K. Kanatani, “Computational cross-ratio for computer vision”, CVGIP: Image Understanding, Vol. 60, No. 3, 1994, p. 371-381.

[14] D. G. Lowe, “Perceptual Organization and Visual Recognition”, Kluwer Academic, Norwell, Ma, 1985.

[15] S. Maybank, “Probabilistic analysis of the application of the cross ratio to model based vision: misclassification”, Intern. J. Comput. Vis. 14, 1995, p. 199-210.

[16] S. Carlsson, R. Mohr, T. Moons, L. Morin, C. Rothwell, M. Van Diest, L. Van Gool, F. Veillon, A. Zissermann, “Semi-local projective invariants for the recognition of smooth plane curve”, Intern. J. Comput. Vis. 19(3), 1996, p. 211-236.

[17] R. Milanese, D. Squire and T. Pun, *Correspondence analysis and hierarchical indexing for content-based image retrieval*. IEEE Intl. Conference on Image Processing, Lausanne, Switzerland, Sept 16-19, 1996.

[18] T.Moons, E.J. Pauwels, L.J. Van Gool, A. Oosterlinck, “Foundations of semi-differential invariants”, Intern. J. Comput. Vis. 14, 1993, p. 25-47.

[19] J.L. Mundy and A. Zisserman (editors), “Geometric Invariance in Computer Vision”, MIT Press, Cambridge Ma, 1992.

[20] J.L. Mundy, D. Kapur, S. J. Maybank, P. Gros and L. Quan, “Geometric Interpretation of Joint Conic Invariants”, In [56], p. 77-86.

[21] E.J. Pauwels, T.Moons, L.J. Van Gool, P. Kempenaers, A. Oosterlinck, “Recognition of planar shapes under affine distortion”, Intern. J. Comput. Vis. 14, 1993, p. 49-65.

[22] A. Pentland, R.W. Picard, and S. Sclaroff (1994). Photobook: tools for content-based manipulation of image databases. (Storage and Retrieval for Image and Video Databases II, San Jose, CA, USA, 7-8 Feb. 1994). *Proceedings of the SPIE - The International Society for Optical Engineering*, 2185, 34-47.

[23] T. Pun, D. Squire, *Statistical structuring of pictorial databases for content-based image retrieval systems*. Pattern Recognition Letters, 17, 12, October 1996, 1299-1310.

[24] T. H. Reiss, “Recognizing planar objects using invariant image features”, Lecture Notes in Computer Science, Springer-Verlag, 676, 1993.

[25] E. Rivlin, I. Weiss, “Local invariants for recognition”, IEEE Trans. Patt. Anal. Mach. Intell., 1995, vol. 17, No. 3.

[26] Rogers D.F., Adams J.A., “Mathematical elements for computer graphics”,McGraw-Hill, New York, NY, 1990, p. 371-375.

[27] S. Sclaroff, “Encoding deformable shape categories for efficient content-based search”, Proc. 1st Int. Workshop Image Databases & Multimedia Search, Amsterdam, Netherlands, 1996, p. 107

[28] S. Startchik, R. Milanese, C. Rauber, T. Pun, “Planar shape databases with affine invariant search”, Proc. 1st Int. Workshop on Image Databases & Multimedia Search, Amsterdam, Netherlands, 1996, p. 202.

[29] S. Startchik, C. Rauber, T. Pun, “Recognition of planar objects over complex backgrounds using line invariants and relevance measures”, Workshop on Geometric Modeling & Invariants for Computer Vision, Xian, China , 1995, p. 301-307.

[30] VIRAGE web site, http://www.virage.com.

[31] I. Weiss, “Geometric Invariants and Object Recognition”, Intern. J. Comput. Vis. 10:3, 1993, pp. 209.

[32] A. Zissermann, D.A. Forsyth, J.L. Mundy, C.A. Rothwell, “Recognizing general curved objects efficiently”, In [19], p. 228-251