

To appear in the Pattern Recognition Journal

Image Retrieval via Isotropic and Anisotropic Mappings¹

Qasim Iqbal,² J. K. Aggarwal³

Computer and Vision Research Center

Department of Electrical and Computer Engineering

The University of Texas at Austin

Austin, Texas 78712, USA

Abstract

This paper presents an approach for content-based image retrieval via isotropic and anisotropic mappings. Isotropic mappings are defined as mappings invariant to the action of the planar Euclidean group on the image space – invariant to the translation, rotation and reflection of image data, and hence, invariant to orientation and position. Anisotropic mappings, on the other hand, are defined as those mappings that are correspondingly variant. Structure extraction (via a perceptual grouping process) and color histogram are shown to be representations of isotropic mappings. Texture analysis using a channel energy model comprised of even-symmetric Gabor filters is considered to be a representation of anisotropic mapping. An integration framework for these mappings is developed. Results of retrieval of outdoor images by query and by classification using a nearest neighbor classifier are presented.

Key words: Image retrieval, Euclidean group, perceptual grouping, structure, texture, color histogram, Gabor filter, nearest neighbor classifier.

1 Introduction

The interest in the automatic analysis of images based upon their content has significantly increased with recent developments in digital image collections, World Wide Web (WWW), networking and multimedia. Active research in content-based image retrieval (CBIR) is geared towards the development of methodologies for analyzing, interpreting, cataloging and indexing image databases. In addition to their development, efforts are also being made to evaluate the performance of image retrieval systems [1].

Most of the previous work in image retrieval has focused on retrieval by image query [2–5]. However, retrieval by image classification has also gained attention [6–10]. Retrieval by image query refers to the retrieval of images similar to a given query image from an image database, whereas retrieval by classification refers to the classification of images into certain known classes for retrieval. In this paper we develop a methodology for retrieval of outdoor images using both image query and image classification by using a nearest neighbor classifier.

In image analysis, a desirable attribute is the notion of isotropy of computations in the sense of Euclidean invariance: any rotation, translation or reflection of the input should produce an identical result under these transformations, thus achieving orientation and position invariance. These image transformations are generated by the

¹ This work was supported in part by the Army Research Office under contracts DAAD19-00-1-0044, DAAG55-98-1-0230 and DAAD19-99-1-0012 (Johns Hopkins University sub-contract agreement 8905-48168).

² Email. qasim@mail.utexas.edu

³ Corresponding author. Tel. +1-512-471-1369. Fax. +1-512-471-5532. Email. aggarwaljk@mail.utexas.edu

action of the (planar) Euclidean group on the image space. An isometry is a transformation of a space which preserves (Euclidean) distance. The Euclidean group is the group of isometries of an Euclidean space and is (isomorphic to) the semi-direct product of the orthogonal group and the translation group. The orthogonal group is represented by rotations and reflections. The translation group is represented by shifts of the Euclidean space.

The action of the Euclidean group on the space of positions and directions $\mathfrak{R}^2 \times \mathcal{S}^1$, where positions are represented using \mathfrak{R}^2 and directions using the unit circle \mathcal{S}^1 , generates isometric geometrical objects. It has been argued that visual computations occur on $\mathfrak{R}^2 \times \mathcal{S}^1$, rather than on just \mathfrak{R}^2 [11]. Using this notion of isotropy, we present an approach for content-based image retrieval via isotropic and anisotropic mappings.

We define an *isotropic mapping* as a mapping that is *invariant* to the action of the Euclidean group on the image space – invariant to translation, rotation, and reflection of image data. Similarly, we define an *anisotropic mapping* as a mapping that is variant to the action of the Euclidean group. Isometries are important for developing a framework for isotropic mappings. It is shown later that all isometries of a plane can be represented using the product of a translation and either a rotation or a reflection. Isotropic mappings acting on perceptually salient image structures are useful in retrieval, as they illustrate the *similarity* of different structures in an image. On the other hand, anisotropic mappings indicate the *uniqueness* of certain attributes of different images. We show that structure extraction via perceptual grouping is a natural candidate for isotropic mappings, as are histograms of pixel color values. On the other hand, lower-level texture analysis via a Gabor filter bank (which possesses affinity for certain preferred directions) operating in a channel energy model is shown as an effective candidate for anisotropic mappings.

This paper discusses an integration framework for these mappings. The integrated framework takes advantage of the strength of structure, color histogram and texture in their respective domains for retrieval. The motivation is to develop a system that is able to retrieve images ranging from purely natural objects, such as images of vegetation, flowers, water and sky, to images containing conspicuous structure, such as images of building, towers, bridges and other architectural objects. Results of retrieval of outdoor images by query and by classification using a nearest neighbor classifier are presented.

The rest of the paper is organized as follows: section 2 explains the perceptual grouping process to extract structure. Section 3 provides a brief introduction to the Euclidean group. Section 4 establishes structure and the color histogram as representations of isotropic mapping. Section 5 describes the texture analysis via a channel energy model as a representation of anisotropic mapping. Section 6 outlines the integration of isotropic and anisotropic mappings. Section 7 describes the results obtained, and finally, section 8 provides the conclusions.

2 Perceptual grouping – Structure extraction and feature selection

The human visual system can detect many classes of patterns and statistically significant arrangements of image elements. Perceptual grouping refers to the human visual ability to extract significant image relations from lower-level primitive image features without any knowledge of the image content and hierarchically group them to obtain meaningful higher-level structure. It stresses the uniformity of psychological grouping for perception and recognition, as opposed to recognition by analysis of discrete primitive image features, and embodies such concepts as grouping by *proximity*, *similarity*, *continuation*, *closure*, and *symmetry* [8].

Manmade objects have sharp edges and straight boundaries. The presence of a manmade object in an image generates a large number of significant edges, junctions, parallel lines and groups, and closed structures, in comparison with an image with predominantly non-manmade (non-structural) objects. These features are generated by the presence of corners, windows, doors and boundaries of objects such as buildings, towers, bridges and other architectural objects. They exhibit regularity and relationship, and are strong evidence that *structure* is present in an image. The presence of these distinguishing features follows the “principle of non-accidentalness” [8]; therefore, these features are more likely to be generated by manmade objects. Hence, these discriminating features distinguish between an image containing manmade objects and an image containing none.

In our approach, segmentation and detailed object representation are not required. We extract the following features hierarchically in an unconstrained environment, i.e., with no constraints on the viewing angle and depth, using the approach detailed in [8,9]: *line (edge) segments, longer linear lines, retained lines, coterminations, “L” junctions, “U” junctions, parallel lines, parallel groups, “significant” parallel groups, and polygons*. The features are shown in figure 1. Perceptual grouping rules of similarity, continuity, parallelism and closure are used to extract these features.

In general, the extracted feature vector $\mathbf{X}_S = (\tilde{x}_{S_1}, \dots, \tilde{x}_{S_d})^t$, where d is the dimensionality of the feature space, and $\tilde{x}_{S_i} = (\sum_j \chi_{\omega_{S_i}}(l_j)) / (\sum_k \chi_{\omega_{\Theta_i}}(l_k))$, where $i \in [1, \dots, d]$. In addition, χ denotes the characteristic (indicator) function, l is a retained line, ω_{Θ_i} is the set of all retained lines, ω_{S_i} is a higher-level structure extracted, and $\tilde{x}_{S_i} \in [0, 1]$, i.e., the feature space is represented by a unit hypercube. For retrieval by both image query and image classification, we set $d = 3$, and let ω_{S_i} represent “L” junctions, “U” junctions, and “significant parallel groups and polygons” for $i \in \{1, 2, 3\}$, respectively. Thus, \tilde{x}_{S_i} represents the corresponding

normalized number of lines. Hence, the feature vector extracted is expressed as:

$$\mathbf{X}_S = (\tilde{\mathbf{x}}_{S_1}, \tilde{\mathbf{x}}_{S_2}, \tilde{\mathbf{x}}_{S_3})^t \quad (1)$$

where

$$\tilde{\mathbf{x}}_{S_1} = \frac{\text{\# of lines in "L" junctions}}{\text{Total \# of retained lines}}$$

$$\tilde{\mathbf{x}}_{S_2} = \frac{\text{\# of lines in "U" junctions}}{\text{Total \# of retained lines}} \quad (2)$$

$$\tilde{\mathbf{x}}_{S_3} = \frac{\text{\# of lines in (significant) parallel groups and polygons}}{\text{Total \# of retained lines}}$$

Detailed justification for using this feature vector is provided in [8]. In addition, elimination of *weak-edged* line segments and lines *shorter* than a given threshold helps to keep background clutter to a minimum [8,9].

3 Action of the Euclidean group – Action by translation, rotation, and reflection

An isometry is a mapping that preserves distances, i.e., the distance between any two points in a space remains invariant after the application of the mapping. It is well-known that the set of all isometries of \mathfrak{R}^2 forms a group, called the Euclidean group. To see this, let Γ be an isometry of \mathfrak{R}^2 , and let $\mathbf{b} = \Gamma(\mathbf{0})$, where $\mathbf{b}, \mathbf{0} \in \mathfrak{R}^2$. Let $\tau_{\mathbf{b}}$ represent a member of the translation group $T(2)$ of \mathfrak{R}^2 , such that $\tau_{\mathbf{b}}(\mathbf{r}) =$

$\mathbf{r} + \mathbf{b}, \mathbf{r} \in \mathfrak{R}^2$. It is easy to see that the translation group is isomorphic to the additive group \mathfrak{R}^2 , and that a translation is an isometry. Then, $\varrho = \tau_{-\mathbf{b}}\Gamma$ is an isometry of \mathfrak{R}^2 , satisfying $\varrho(\mathbf{0}) = \mathbf{0}$. It can be shown that if $\varrho(\mathbf{0}) = \mathbf{0}$, then ϱ is linear [12], and thus, $\Gamma = \tau_{-\mathbf{b}}^{-1}\varrho = \tau_{\mathbf{b}}\varrho$ is a product of a linear isometry and a translation. Further, it can also be shown that the linear isometries can be represented by the orthogonal group $O(2, \mathfrak{R})$ of 2×2 orthogonal matrices that represent rotations and reflections.

The matrix orthogonal group $O(2, \mathfrak{R})$ is the set of all orthogonal matrices. The determinant of any element (matrix) of $O(2, \mathfrak{R})$ is either 1 or -1. The set of orthogonal matrices with determinant equal to 1 forms a subgroup (called the special orthogonal group) that represents rotations. Reflections have a determinant equal to -1. It can be shown that the special orthogonal group is a normal subgroup of $O(2, \mathfrak{R})$.

The (semi-direct) product of the orthogonal group and the translation group is the group of isometries of \mathfrak{R}^2 (called Euclidean group $E(2)$). Semi-direct product is a mechanism by which two groups can be fitted together to form a larger group, where one of the two groups is a normal subgroup and the other is a subgroup of the larger group formed. The fact that the translation group is a normal subgroup of $E(2)$, and the intersection of the translation group and the orthogonal group is the identity, can be used to deduce that $E(2) \cong O(2, \mathfrak{R}) \rtimes T(2)$, where \cong denotes isomorphism and \rtimes denotes the semi-direct product [13]. The construction is straightforward. As a set, $O(2, \mathfrak{R}) \rtimes T(2)$ is $T(2) \times O(2, \mathfrak{R})$, but now the product is defined by $(\tau_{\mathbf{b}}, \varrho) (\tau_{\mathbf{b}'}, \varrho') = (\tau_{\mathbf{b}\varrho} \cdot \tau_{\mathbf{b}'}, \varrho\varrho')$, where $\tau_{\mathbf{b}\varrho} \cdot \tau_{\mathbf{b}'}$ is to be interpreted as a translation by an amount $\varrho\mathbf{b}' + \mathbf{b}$, i.e., the translation $\tau_{(\varrho\mathbf{b}'+\mathbf{b})}$. The semi-direct product contains isomorphic copies of the orthogonal group and the translation group as subgroups.

The elements of $E(2)$ are invertible. The distance-preserving mappings that are

elements of $E(2)$ are implicit in many computations in the Euclidean geometry. It may be noticed, however, that in agreeing not to distinguish between two congruent figures in a plane, we are in essence agreeing not to distinguish between the figures if there is an element of the Euclidean group that maps one of the figures onto the other [14]. In this sense, all of the symmetry groups of the two-dimensional figures are subgroups of the Euclidean group. In addition, the Euclidean group is a subgroup of the affine group. Further, it can be shown that the quotient group $E(2)/T(2) \cong O(2, \mathfrak{R})$.

4 Isotropic mapping

We consider features extracted from the structural analysis of an image via the perceptual grouping process and the color histogram, and show that they are representations of isotropic mappings.

4.1 Euclidean isotropy of \mathbf{X}_S

Let $\omega = \{\omega_i\}$ represent the set of objects of interest present in an image. Each object ω_i is a set of $\omega_{i_k} = \{\mathbf{r}, \phi\} \in \mathfrak{R}^2 \times \mathcal{S}^1$, where $\mathbf{r} = \{x, y\} \in \mathfrak{R}^2$ is a coordinate pair, \mathcal{S}^1 is the unit circle and $\phi \in \mathcal{S}^1$ represents an orientation. We treat \mathbf{r} and ϕ as independent variables, so that all possible orientations for ϕ exist at each corresponding position \mathbf{r} . The relation between ω_i and various image structures must be properly understood. At the lowest level of vision, ω_{i_k} are represented by points (pixels) on an edge segment ω_i (where each ω_i is obtained using the edge detection process described in [8,9]) and ϕ represents the orientation of ω_i . At this level, ω_i are identified with edge segments l_i as shown in figure 1. For example, in figure

1(a), ω_1 might be identified with the edge segment l_1 .

At the next level of perceptual grouping, certain ω_i will be combined to generate a higher-level structure. The structure obtained from the grouping of ω_i may be called ω_j for consistency of notation, although it should be understood that ω_j now represents a structure at a higher level than ω_i . (Refer to figure 2, where edge segments $\omega_3, \omega_4, \omega_5$ and ω_6 combine to form ω_7 .) As an example, at a higher level, ω_i might refer to higher-level structures such as polygons shown in figure 1(g). It may be noted that though the representations of objects ω_i change as they represent hierarchical higher-level structures, the representations of ω_{i_k} remain the same, since ω_{i_k} are points.

A *group action* of a group G on a set \mathcal{A} is a map from $G \times \mathcal{A} \rightarrow \mathcal{A}$ (written as $g \cdot a$ for all $g \in G$ and $a \in \mathcal{A}$) that satisfies [13]:

$$\begin{aligned} g_1 \cdot (g_2 \cdot a) &= (g_1 g_2) \cdot a, \quad \forall g_1, g_2 \in G, a \in \mathcal{A} \\ \mathbf{I}_e \cdot a &= a, \quad \forall a \in \mathcal{A} \end{aligned} \quad (3)$$

where \mathbf{I}_e is the identity element of G . As mentioned before, we do not distinguish between two congruent figures if an element of the Euclidean group maps a (planar) figure to another congruent figure. This representation includes all of the symmetry groups of the two-dimensional figures, since the symmetry groups are subgroups of the Euclidean group.

We define a mapping $\psi : \omega \rightarrow \mathfrak{R}^d$ (where d is the dimensionality of the feature space) to be isotropic if it is invariant to the action of the Euclidean group on the space of positions and directions $\mathfrak{R}^2 \times \mathcal{S}^1$:

$$\psi(E_j \cdot \omega) = \psi(\omega) \quad (4)$$

where E is the Euclidean group $E(2)$ – the semi-direct product of the orthogonal group and the translation group. The extraction of the feature vector \mathbf{X}_S is represented by ψ . Since the group action is originally defined on $\mathfrak{R}^2 \times \mathcal{S}^1$, the action $E_j \cdot \omega_i$, for all $E_j \in E$ and $\omega_i \in \omega$, transforms each $\omega_{i_k} \in \omega_i$ (refer to figure 3):

$$\begin{aligned}
\tau_{\mathbf{b}} \cdot (\mathbf{r}, \phi) &= (\mathbf{r} + \mathbf{b}, \phi), \quad \mathbf{r}, \mathbf{b} \in \mathfrak{R}^2, \phi \in \mathcal{S}^1 \\
R_\theta \cdot (\mathbf{r}, \phi) &= (R_\theta \mathbf{r}, \phi + \theta), \quad \theta \in \mathcal{S}^1 \\
\kappa_\theta \cdot (\mathbf{r}, \phi) &= \acute{\kappa} R_{-2\theta} \cdot (\mathbf{r}, \phi) = (\acute{\kappa} R_{-2\theta} \mathbf{r}, -(\phi - 2\theta))
\end{aligned} \tag{5}$$

where $\tau_{\mathbf{b}} \in T(2)$, ($\mathbf{b} \in \mathfrak{R}^2$) represents a member of the translation group of \mathfrak{R}^2 , $T(2)$, such that $\tau_{\mathbf{b}}(\mathbf{r}) = \mathbf{r} + \mathbf{b}$, $\mathbf{r} \in \mathfrak{R}^2$, $R_\theta \in O(2, \mathfrak{R})$ is a rotation by an angle θ (with center at the origin), $\kappa_\theta \in O(2, \mathfrak{R})$ is a reflection along an axis (through the origin) in \mathfrak{R}^2 , and $\acute{\kappa}$ is the reflection along the x-axis: $\{x, y\} \rightarrow \{x, -y\}$. The axis of reflection κ_θ is inclined at an angle θ with the x-axis and is spanned by the unit vector $(\cos \theta, \sin \theta)^t$. In general, $E_j = \tau_{\mathbf{b}} \varrho$, where ϱ is either R_θ or κ_θ .

The action $\kappa_\theta \cdot (\mathbf{r}, \phi) = R_\theta \acute{\kappa} R_{-\theta} \cdot (\mathbf{r}, \phi) = \acute{\kappa} R_{-2\theta} \cdot (\mathbf{r}, \phi)$ (by using the identity $R_\theta \acute{\kappa} = \acute{\kappa} R_{-\theta}$), because reflection along an arbitrary axis is equivalent to the rotation of \mathfrak{R}^2 by an angle $-\theta$ to align the axis of reflection along the direction of the original x-axis, followed by a reflection in the (new) x-axis, and then a (reverse) rotation by an angle θ .

The homomorphism implied in the upper equality in equation 3 may be seen as follows. The action of E_j on a function $\Upsilon : \mathfrak{R}^2 \times \mathcal{S}^1 \rightarrow \mathfrak{R}$ (e.g., an image function)

is given by $E_j \cdot \Upsilon(\mathbf{r}, \phi) = \Upsilon(E_j^{-1} \cdot (\mathbf{r}, \phi))$. Therefore,

$$\begin{aligned} (E_j \cdot E_k) \cdot \Upsilon(\mathbf{r}, \phi) &= \Upsilon((E_j \cdot E_k)^{-1} \cdot (\mathbf{r}, \phi)) = \Upsilon(E_k^{-1} \cdot E_j^{-1} \cdot (\mathbf{r}, \phi)) \\ &= E_k \cdot \Upsilon(E_j^{-1} \cdot (\mathbf{r}, \phi)) = E_j \cdot (E_k \cdot \Upsilon(\mathbf{r}, \phi)) \end{aligned} \quad (6)$$

It may be noted that if the center of rotation is not the origin, or if the axis of reflection does not pass through the origin, then the fact that the translation group is a normal subgroup of the Euclidean group may be used to reduce the resulting transformation into a product of a linear isometry and a translation. This assertion (regarding the generalized rotation and reflection) is proved in appendix A. If the translation invariance of the first equality in equation 5 is established, then the generalized rotation and reflection reduce to action by R_θ and $\acute{\kappa}R_{-2\theta}$, respectively. It is interesting to note that if the translation and rotation invariance of the first and second equality in equation 5, respectively, are established, it is sufficient to show the invariance of $\acute{\kappa}$ to show the invariance of κ_θ .

4.1.1 Linear feature modeling

The premise of linear feature modeling is to extract rich descriptions of lower-level local image primitives and use these descriptions for subsequent grouping into higher-level features (linear line segments). We develop a mathematical model of the perceptual grouping process described in [8,9] for the collection of edge segments ω_k to form a longer linear line ω_j (figure 1(a)). At this level of vision, ω_k are identified with lines l_k as shown in the figure. Let $\mathbf{r} = \{x, y\}$ denote the x- and y-coordinates of an end-point of an edge segment ω_k , and $\phi \in \mathcal{S}^1$ represent the orientation of the edge segment. We treat \mathbf{r} and ϕ as independent variables, so that all possible orientations for ω_k exist at each corresponding position \mathbf{r} .

A certain collection \mathcal{C}_i of ω_k is collected, which will be replaced by ω_j , that maximizes the energy λ_i given as:

$$\lambda_i^{(n)} = \lambda_i^{(n-1)} + \sum_{k \in \mathcal{K}, l \notin \mathcal{K}; \mathcal{K} = \{\tilde{k}: \omega_{\tilde{k}} \in \mathcal{C}_i\}} \xi_{kl}, \quad \lambda_i^{(0)} = 0 \quad (7)$$

where the superscript n is an iteration index and (omitting the subscript i) the energy functional $\xi_{kl} : (\omega_b, \omega_k, \omega_l) \rightarrow \Re$ is expressed as:

$$\xi_{kl}(\omega_b | \omega_k, \omega_l) = \Lambda(q) \Lambda(st) \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) \delta(\phi_b - \phi_l) \quad (8)$$

where ω_b (identified with l_b in figure 1(a)) is a certain *base* edge segment in the collection that is used to determine that all other edge segments are parallel to it. Further, Λ is a weighting function and q is the maximum length of the orthogonal distance of any point of ω_l from ω_b . In the above equation, \mathbf{r}_k and \mathbf{r}_l represent those end-points of two edge segments ω_k and ω_l , respectively, which are closer to each other (at the lower level), and ϕ_b and ϕ_l are the orientations of ω_b and ω_l , respectively. In addition, δ is the Dirac delta function, \mathbf{e}_{kl} is a unit vector in the direction of $\mathbf{r}_k - \mathbf{r}_l$ and s is a distance parameter along an axis parallel to the direction of $\mathbf{r}_k - \mathbf{r}_l$. The Boolean parameter t is such that $t = 0$ if the length of the orthogonal projection of ω_l on ω_k is greater than zero, otherwise $t = 1$.

We represent Λ by a constant function (not equal to zero) with compact support. Specifically, we have selected the constant as 1 and the support is equal to 5 units (pixels). Equation 7 indicates the iterative nature of the grouping. At the start \mathcal{C}_i consists of only one segment ω_b . At the end of each iteration those ω_l 's for which ξ_{kl} is non-zero are put into \mathcal{C}_i . The grouping is started again and continued until there is no increase in λ_i . The higher-level longer linear line ω_j is then obtained by a weighted average of the lengths and orientations of all edge segments in \mathcal{C}_i [8].

The energy functional expressed in equation 8 is similar to the one defined in [15], however, in their model \mathbf{r}_k represents the V1 image of the center of the receptive field of a neuron, and \mathbf{e}_{kl} represents the V1 image of the orientation preference of the neuron. Unlike their model, in our system \mathbf{e}_{kl} points in the direction of $\mathbf{r}_k - \mathbf{r}_l$ and incorporates the non-collinearity of two edge segments to an arbitrary extent (e.g., figure 1). (To further emphasize closer points, unequal weights, as opposed to constant weights in the support of Λ , can be obtained by replacing Λ with an appropriate weighting function such as a Gaussian function.) It may be noted that the energy functional given in equation 8 incorporates the Gestalt principles of proximity, collinearity, parallelism, and good continuation.

4.1.2 Euclidean invariance of ξ_{kl}

The energy functional expressed in equation 8 has a well-defined symmetry: it is invariant under the action of $E(2)$; invariant under translations $\{\mathbf{r}, \phi\} \rightarrow \{\mathbf{r} + \mathbf{b}, \phi\}$, rotations $\{\mathbf{r}, \phi\} \rightarrow \{R_\theta \mathbf{r}, \phi + \theta\}$ and reflections $\{\mathbf{r}, \phi\} \rightarrow \{\kappa R_{-2\theta} \mathbf{r}, -(\phi - 2\theta)\}$. In appendix B, it is verified that the invariance of $s = \|\mathbf{r}_k - \mathbf{r}_l\|$ can be established as $\|\tau_{\mathbf{b}} \varrho \mathbf{r}_k - \tau_{\mathbf{b}} \varrho \mathbf{r}_l\| = \|\mathbf{r}_k - \mathbf{r}_l\| = s$, where ϱ is either a rotation R_θ , or a reflection κ_θ . The invariance of q and t may also be established in a similar manner. Translation, rotation, and reflection invariance of equation 8 imply the following equalities, respectively:

$$\begin{aligned} \xi_{kl}(\tau_{\mathbf{b}} \cdot \omega_b \mid \tau_{\mathbf{b}} \cdot \omega_k, \tau_{\mathbf{b}} \cdot \omega_l) &= \xi_{kl}(\omega_b \mid \omega_k, \omega_l) \\ \xi_{kl}(R_\theta \cdot \omega_b \mid R_\theta \cdot \omega_k, R_\theta \cdot \omega_l) &= \xi_{kl}(\omega_b \mid \omega_k, \omega_l) \\ \xi_{kl}(\kappa_\theta \cdot \omega_b \mid \kappa_\theta \cdot \omega_k, \kappa_\theta \cdot \omega_l) &= \xi_{kl}(\omega_b \mid \omega_k, \omega_l) \end{aligned} \tag{9}$$

These relations are also proved in appendix B.

Equation 8 is at the heart of the perceptual grouping process. Its Euclidean invariance, as stated above in equation 9 and proved in appendix B, means that equation 7 remains invariant, and the perceptual grouping process will produce the *same* groupings – retained lines. All higher-level structures are extracted using the retained lines.

4.1.3 Higher-level structures

The fundamental perceptual grouping proposed in [8,9] for higher-level structures can be modeled as follows. The proximity of two objects ω_k and ω_l can be modeled by the relation $\Lambda(s) \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})$. Here, \mathbf{r}_k and \mathbf{r}_l refer to those end-points of ω_k and ω_l , respectively, that are closer to each other than any other pair of end-points. The variation in the orientations of ω_k and ω_l can be controlled by the relation $\tilde{\Lambda}(p) \delta(\phi_k - \phi_l - p)$, where the variable $p = \phi_k - \phi_l$, $p \in [0, 2\pi]$ and $\tilde{\Lambda}$ is a constant function (not equal to zero) with compact support (similar to Λ). At a higher-level, ϕ_k and ϕ_l represent the general orientations associated with the entire objects ω_k and ω_l , respectively. Using an argument similar to the one shown above, it can be verified that these relations are invariant under the action of $E(2)$. Hence, \mathbf{X}_S obtained by the mapping ψ remains invariant to the action of $E(2)$.

4.2 Color histogram

Color histogram measures are invariant to both $O(2, \mathfrak{R})$ and $T(2)$, and hence, $E(2)$, because histogram measures are only dependent on summations of identical pixel values and do not incorporate orientation and position. The extraction of the normalized histogram $\mathbf{X}_{\mathcal{H}} \in \mathfrak{R}^{512}$ is used as a representation of an isotropic mapping.

A color space is *perceptually uniform* if a small perturbation to a component value is approximately equally perceptible across the range of that value. The *RGB* color space does not exhibit perceptual uniformity. However, the CIE *LAB* space, conceived in 1976, improves the perceptual uniformity of *RGB* space considerably. In this space *L* defines lightness, *A* denotes red/green chrominance and *B* the yellow/blue chrominance. Given an image $I_{RGB}(x, y)$ in *RGB* space we generate $I_{LAB}(x, y)$, where the pair (x, y) denotes the coordinates in an image *I*. A 512-dimensional feature vector $\mathbf{X}_{\mathcal{H}}$, representing the 512-bin normalized histogram, is extracted from the image $I_{LAB}(x, y)$ by uniformly quantizing the *LAB* space, i.e.,

$$\mathbf{X}_{\mathcal{H}} = (\tilde{\mathbf{x}}_{\mathcal{H}_0}, \dots, \tilde{\mathbf{x}}_{\mathcal{H}_{511}})^t \quad (10)$$

where $\tilde{\mathbf{x}}_{\mathcal{H}_j}$ (where the index integer $j \in [0, 511]$) represents the normalized value of the j^{th} bin of the histogram such that $\sum_{j=0}^{511} \tilde{\mathbf{x}}_{\mathcal{H}_j} = 1$. This feature space represents a unit hypercube.

5 Anisotropic mapping

In most quantitative channel energy models of texture analysis, an image is processed by channel-selective filters along certain fundamental stimulus dimensions such as spatial-frequency and orientation. Texture analysis via a channel energy model employing a Gabor filter bank is considered to be a representation of anisotropic mapping. The representation is accomplished by the extraction of the feature vector $\mathbf{X}_{\mathcal{T}} \in \mathfrak{R}^{48}$, which measures the fractional energy in various spatial channels after treating the input image with the Gabor filter bank.

The fact that this representation is anisotropic may readily be verified from the fact

that the translation of an image $I(\mathbf{r}) \rightarrow I(\tau_{\mathbf{b}}(\mathbf{r}))$ transforms the Fourier transform of the image $\mathcal{I}(\nu) \rightarrow \mathcal{I}(\nu) e^{j2\pi\langle\mathbf{b},\nu\rangle}$, where $\mathbf{b}, \mathbf{r}, \nu \in \Re^2$, $\mathbf{r} = \{x, y\}$ are the space domain coordinates and $\nu = \{u, v\}$ are the Fourier domain co-ordinates. A rotation of the image $I(\mathbf{r}) \rightarrow I(R_{\theta}(\mathbf{r}))$, transforms the Fourier transform $\mathcal{I}(\nu) \rightarrow \mathcal{I}(R_{\theta}(\nu))$. Similarly, the reflection of an image $I(\mathbf{r}) \rightarrow I(\kappa_{\theta}(\mathbf{r}))$ transforms the Fourier transform $\mathcal{I}(\nu) \rightarrow \mathcal{I}(\kappa_{\theta}(\nu))$. (These relations are derived in appendix C.) Hence, texture analysis is not invariant after the action of $E(2)$ on an image.

The *LAB* space is used for multiresolution texture analysis by measuring the fractional energies in the lightness and the two chrominance channels. Given an image I , the convolved sequence $\{I * \hat{f}_{m,n}\}$ defines the multiresolution image texture characteristics, where $\hat{f}_{m,n}$ denotes the base texture extraction function f at scale m and orientation n . The filter energy ($\|\hat{f}_{m,n}\|^2$) is held constant. Even-symmetric, two-dimensional Gabor filters have been used to represent $\hat{f}_{m,n}$. The impulse response of the base filter f is given as:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)} \cos(2\pi u_0 x) \quad (11)$$

where $f(x, y)$ represents the response at spatial locations x and y , u_0 is the frequency of a sinusoidal plane wave along the x-axis (i.e., the 0^0 orientation), and σ_x and σ_y are the spreads of the Gaussian envelope along the x- and y-axis, respectively.

A set of self-similar Gabor filters is obtained by appropriate rotations and scalings of $f(x, y)$ through the generating function $\hat{f}_{m,n}(x, y) = k^{-m} f(k^{-m}\hat{x}, k^{-m}\hat{y})$, where $k \geq 1$, m and n are integers, $\hat{f}_{m,n}(x, y)$ is the rotated and scaled version of the original filter and k is the scale factor. In the above equation, $n = 0, 1, \dots, N - 1$ is the current orientation index, where N is the total number of

orientations, and $m = 0, 1, \dots, M - 1$ is the current scale index, where M is the total number of scales. In addition, \acute{x} and \acute{y} are the rotated coordinates $\acute{x} = x \cos \theta + y \sin \theta$, $\acute{y} = -x \sin \theta + y \cos \theta$, and $\theta = \frac{n\pi}{N}$ is the orientation. The scale factor k^{-m} ensures that the filter energy is independent of m . In order to make the filters zero-mean, we set $\acute{F}_{m,n}(0, 0) = 0$, where $\acute{F}_{m,n}$ is the Fourier transform of $\acute{f}_{m,n}$. A total of 16 Gabor filters (per channel) are selected, with four filters in equi-angular orientations at four different scales, i.e., $N = 4$, and $M = 4$, starting at the 0° orientation. Parameters σ_x , σ_y and k are calculated using a multiresolution filter design approach [16].

Channels L , A and B are treated with the Gabor filter bank described above. The 48-dimensional feature vector \mathbf{X}_T is constructed using the fractional energies in each of the 16 filters operating in the L , A and B channels, i.e.,

$$\mathbf{X}_T = (\tilde{\mathbf{x}}_{TL_{0,0}}, \dots, \tilde{\mathbf{x}}_{TL_{3,3}}; \tilde{\mathbf{x}}_{TA_{0,0}}, \dots, \tilde{\mathbf{x}}_{TA_{3,3}}; \tilde{\mathbf{x}}_{TB_{0,0}}, \dots, \tilde{\mathbf{x}}_{TB_{3,3}})^t \quad (12)$$

where $\tilde{\mathbf{x}}_{TL_{m,n}}$, $\tilde{\mathbf{x}}_{TA_{m,n}}$ and $\tilde{\mathbf{x}}_{TB_{m,n}}$ represent the fractional energy at the output of the filter in the n^{th} orientation and the m^{th} scale, for L , A and B channels, respectively. The fractional energy $\tilde{\mathbf{x}}_{TL_{m,n}}$ (in the discrete form) is given as:

$$\tilde{\mathbf{x}}_{TL_{m,n}} = \frac{\sum_{y=0}^{W_y-1} \sum_{x=0}^{W_x-1} \hat{L}_{m,n}^2(x, y)}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{y=0}^{W_y-1} \sum_{x=0}^{W_x-1} \hat{L}_{m,n}^2(x, y)} \quad (13)$$

where $\hat{L}_{m,n}$ is the L channel treated with the filter $\acute{f}_{m,n}$, W_x is the width of the image, W_y is the height, and $\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \tilde{\mathbf{x}}_{TL_{m,n}} = 1$. In a similar manner, we define

$$\tilde{\mathbf{x}}_{TA_{m,n}} = (\sum_{y=0}^{W_y-1} \sum_{x=0}^{W_x-1} \hat{A}_{m,n}^2(x, y)) / (\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{y=0}^{W_y-1} \sum_{x=0}^{W_x-1} \hat{A}_{m,n}^2(x, y))$$

and $\tilde{\mathbf{x}}_{TB_{m,n}} = (\sum_{y=0}^{W_y-1} \sum_{x=0}^{W_x-1} \hat{B}_{m,n}^2(x, y)) / (\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{y=0}^{W_y-1} \sum_{x=0}^{W_x-1} \hat{B}_{m,n}^2(x, y))$,

where $\hat{A}_{m,n}$ and $\hat{B}_{m,n}$ represent channels A and B treated with Gabor filters, respectively. This feature space is also represented by a unit hypercube.

6 Integration Framework

A retrieval process is described that integrates perceptual grouping, color histogram and texture. Both image query and image classification are discussed.

6.1 Image query

A two-level framework is employed for integrating lower-level and higher-level vision features. Given the isotropic feature vectors \mathbf{X}_S and \mathbf{X}_H and anisotropic feature \mathbf{X}_T extracted from a query image, and \mathbf{X}_{S_j} , \mathbf{X}_{H_j} and \mathbf{X}_{T_j} extracted from the j^{th} image in the database, the first level of the framework maps the feature vectors to a discriminant value within each of the three categories: structure, histogram and texture. The respective mappings $\Phi_S: \mathbb{R}^{N_S} \rightarrow \mathbb{R}$, $\Phi_H: \mathbb{R}^{N_H} \rightarrow \mathbb{R}$ and $\Phi_T: \mathbb{R}^{N_T} \rightarrow \mathbb{R}$, where $N_S = 3$, $N_H = 512$ and $N_T = 48$, are explained as follows. The mappings Φ_S and Φ_T are selected as ℓ_2 norms, $\Phi_S(\mathbf{X}_{S_j}, \mathbf{X}_S) = \|\mathbf{X}_{S_j} - \mathbf{X}_S\|$, $\Phi_T(\mathbf{X}_{T_j}, \mathbf{X}_T) = \|\mathbf{X}_{T_j} - \mathbf{X}_T\|$, and Φ_H is selected as the histogram intersection measure [2]: $\Phi_H(\mathbf{X}_{H_j}, \mathbf{X}_H) = 1 - \beta(\mathbf{X}_{H_j}, \mathbf{X}_H)$, where $\beta(\mathbf{X}_{H_j}, \mathbf{X}_H) = (\sum_{k=1}^{N_H} \min(\tilde{\mathbf{x}}_{H_{j_k}}, \tilde{\mathbf{x}}_{H_k})) / (\sum_{k=1}^{N_H} \tilde{\mathbf{x}}_{H_k})$. Since, $\sum_{k=1}^{N_H} \tilde{\mathbf{x}}_{H_{j_k}} = \sum_{k=1}^{N_H} \tilde{\mathbf{x}}_{H_k} = 1$, the difference in the size of images is incorporated. At the second level, a supra discriminant is generated by utilizing the mapping $\Psi_{SHT}: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}$, given by:

$$\Psi_{SHT}(\mathbf{X}_{S_j}, \mathbf{X}_{H_j}, \mathbf{X}_{T_j}, \mathbf{X}_S, \mathbf{X}_H, \mathbf{X}_T) = \tag{14}$$

$$\mathcal{W}^t \cdot \Phi_{SHT}(\mathbf{X}_{S_j}, \mathbf{X}_{H_j}, \mathbf{X}_{T_j}, \mathbf{X}_S, \mathbf{X}_H, \mathbf{X}_T)$$

where $\mathcal{W} = (w_1, w_2, w_3)^t$ is a weight vector such that $\sum_{i=1}^3 w_i = 1$, $\Psi_{SHT} \in [0, 1]$ and $\Phi_{SHT}: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$, such that $\Phi_{SHT} \in [0, 1] \times [0, 1] \times [0, 1]$, is given as:

$$\Phi_{SHT}(\mathbf{X}_{S_j}, \mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{T_j}, \mathbf{X}_S, \mathbf{X}_{\mathcal{H}}, \mathbf{X}_T) = \quad (15)$$

$$(\hat{\Phi}_S(\mathbf{X}_{S_j}, \mathbf{X}_S), \hat{\Phi}_{\mathcal{H}}(\mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{\mathcal{H}}), \hat{\Phi}_T(\mathbf{X}_{T_j}, \mathbf{X}_T))^t$$

where

$$\begin{aligned} \hat{\Phi}_S(\mathbf{X}_{S_j}, \mathbf{X}_S) &= \frac{\Phi_S(\mathbf{X}_{S_j}, \mathbf{X}_S)}{\max_j \Phi_S(\mathbf{X}_{S_j}, \mathbf{X}_S)} \\ \hat{\Phi}_{\mathcal{H}}(\mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{\mathcal{H}}) &= \frac{\Phi_{\mathcal{H}}(\mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{\mathcal{H}})}{\max_j \Phi_{\mathcal{H}}(\mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{\mathcal{H}})} \\ \hat{\Phi}_T(\mathbf{X}_{T_j}, \mathbf{X}_T) &= \frac{\Phi_T(\mathbf{X}_{T_j}, \mathbf{X}_T)}{\max_j \Phi_T(\mathbf{X}_{T_j}, \mathbf{X}_T)} \end{aligned} \quad (16)$$

These normalizations ensure that $\hat{\Phi}_S \in [0, 1]$, $\hat{\Phi}_{\mathcal{H}} \in [0, 1]$ and $\hat{\Phi}_T \in [0, 1]$. The index \hat{i} of the image most similar to a given query image is given by:

$$\hat{i} = \arg \min_i \Psi_{SHT}(\mathbf{X}_{S_i}, \mathbf{X}_{\mathcal{H}_i}, \mathbf{X}_{T_i}, \mathbf{X}_S, \mathbf{X}_{\mathcal{H}}, \mathbf{X}_T) \quad (17)$$

The above integration framework has the following advantages over a simple concatenation of vectors \mathbf{X}_{S_j} , $\mathbf{X}_{\mathcal{H}_j}$ and \mathbf{X}_{T_j} . First, the different lengths of these three vectors preclude the proper construction of a concatenated vector that is equally sensitive to all of its components. The three-dimensional vector output by Φ_{SHT} is equally sensitive to all of its three one-dimensional components. Second, the size of the corresponding weight vector for the concatenated vector will be large, making the selection of proper weights difficult and unfeasible. Third, in our proposed integration, weights are assigned at the *module level*, i.e., structure, histogram and texture, whereas weights in a concatenated vector are assigned at the vector component level without particular regard to the modular structure of the system. The weight vector plays an important role in controlling the content of images retrieved by assigning different weights to structure, histogram and texture.

6.2 Image classification

A nearest neighbor classifier [17] is used for image classification. The image space is partitioned into three classes, *Structure*, *Non-structure* and *Intermediate*, based upon the measure of manmade object structure present in an image. Here, a more intuitive approach is presented than the original approach presented in [10]. In that approach, for the combined retrieval methodology, a slight alteration of $\Phi_{S\mathcal{H}\mathcal{T}}$'s (equation 15) were taken as patterns. A new three-dimensional feature space was generated by using the three feature spaces (structure, histogram and texture) as $\{\mathbb{R}^{N_S}, \mathbb{R}^{N_{\mathcal{H}}}, \mathbb{R}^{N_{\mathcal{T}}}\} \rightarrow \mathbb{R}^3$, by using $\Phi_S(\mathbf{X}_{S_j}) = \|\mathbf{X}_{S_j}\|$ and $\hat{\Phi}_S(\mathbf{X}_{S_j}) = \frac{\Phi_S(\mathbf{X}_{S_j})}{\max_j \Phi_S(\mathbf{X}_{S_j})}$. The mappings $\Phi_{\mathcal{H}}$, $\Phi_{\mathcal{T}}$, $\hat{\Phi}_{\mathcal{H}}$ and $\hat{\Phi}_{\mathcal{T}}$ were defined similarly. Individual components of resultant three-dimensional vector were multiplied by a weight vector to yield a final discriminant value.

In the current approach, the distance function for the nearest neighbor classifier is redefined to incorporate distances of training feature vectors from the test vectors in the three pattern spaces (structure, histogram and texture) to generate a discriminant value. Specifically, it is defined as the weighted ℓ_1 norm on the product space $\mathbb{R}^{N_S} \times \mathbb{R}^{N_{\mathcal{H}}} \times \mathbb{R}^{N_{\mathcal{T}}}$. Let $d_{S\mathcal{H}\mathcal{T}} : \{\{\mathbb{R}^{N_S}, \mathbb{R}^{N_{\mathcal{H}}}, \mathbb{R}^{N_{\mathcal{T}}}\}, \{\mathbb{R}^{N_S}, \mathbb{R}^{N_{\mathcal{H}}}, \mathbb{R}^{N_{\mathcal{T}}}\}\} \rightarrow \mathbb{R}$ denote the distance function, and \mathbf{X}_S , $\mathbf{X}_{\mathcal{H}}$, and $\mathbf{X}_{\mathcal{T}}$, be the structure, histogram and texture feature vectors, respectively, for a given test image. Let \mathbf{X}_{S_j} , $\mathbf{X}_{\mathcal{H}_j}$ and $\mathbf{X}_{\mathcal{T}_j}$ denote the corresponding training feature vectors of the j^{th} training image. Then,

$$d_{S\mathcal{H}\mathcal{T}}(\mathbf{X}_S, \mathbf{X}_{\mathcal{H}}, \mathbf{X}_{\mathcal{T}}, \mathbf{X}_{S_j}, \mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{\mathcal{T}_j}) = \mathcal{W}^t \cdot (\hat{\Phi}_S(\mathbf{X}_{S_j}, \mathbf{X}_S), \hat{\Phi}_{\mathcal{H}}(\mathbf{X}_{\mathcal{H}_j}, \mathbf{X}_{\mathcal{H}}), \hat{\Phi}_{\mathcal{T}}(\mathbf{X}_{\mathcal{T}_j}, \mathbf{X}_{\mathcal{T}}))^t \quad (18)$$

where $\hat{\Phi}_S$, $\hat{\Phi}_{\mathcal{H}}$ and $\hat{\Phi}_{\mathcal{T}}$ are defined in equation 16.

7 Results obtained

We employ three different image databases consisting of a total of 4,329 24-bit color images. Database #1 consists of 2,139 images of size adjusted to 1024×1024 acquired from two CDs obtained from the Visual Delights, Inc. [18]: “Austin and Vicinity: The Human World” and “Austin and Vicinity: The World of Nature”. Database #2 consists of 521 images of size adjusted to 512×512 acquired from the ground level using the Sony Digital Mavica camera. Database #3 consists of 1,669 images of size adjusted to 512×512 downloaded from the internet (Free Nature Pictures, Info. for Travel Media, Dave’s Wall Paper and Photo Art of Nature [19]).

7.1 Image query

Results for image query were obtained using $\mathcal{W} = (1/3, 1/3, 1/3)^t$. The first 16 images retrieved are displayed in all the figures shown. Figures 4 - 5 show queries for the retrieval of images containing conspicuously natural objects: flowers, leaves and grass, and a duck in water, respectively. The retrieved images closely match the natural content of the supplied query images. Figure 6 shows an interesting query. An image containing an automobile was supplied as a query. This image contains a mixture of an intermediate-level structural object (an automobile) with road and vegetation. The system retrieved images containing automobiles of various colors, because an integral part of the system depends on perceptual grouping that examines the structure of an object regardless of the color. Figure 7 depicts a query for a purely structural object: a building facade. The retrieved images again match closely in structural content to the supplied query image.

7.2 Image classification

Results for image classification were also obtained using $\mathcal{W} = (1/3, 1/3, 1/3)^t$. Tables 1 - 4 display results for retrieval by image classification obtained using a nearest neighbor classifier. Based upon the measure of structure present in an image, the image space was partitioned into three classes, *Structure*, *Non-structure* and *Intermediate*. Each class was represented by 10 training samples. Table 1 shows the overall retrieval rate. Table 2 displays class-conditional retrieval performance measured in terms of *recall* and *precision*. Recall is defined as the fraction of the total number of images that are correctly retrieved for a particular class. Precision is defined as the fraction of images retrieved for a particular class that are actually correct. The detailed retrieval statistics are shown in the confusion matrix shown in Table 3. Table 4 shows the distribution of images that *actually* belong to a particular class within the “best matches” for that class, in intervals of 100 images, and the corresponding *efficiency* of the system. The best matches were obtained by sorting images in ascending order based upon their distances from the training samples of each class. Efficiency is defined as the ratio of the number of images that actually belong to a particular class in the block of closest best matches, to the size of the block. The block size is set equal to the number of images in that class.

8 Conclusions

This paper has presented an approach for content-based image retrieval via isotropic and anisotropic mappings. Isotropic mappings were defined as mappings invariant to the action of the planar Euclidean group on the image space – invariant to the translation, rotation and reflection of image data, and hence, invariant to ori-

entation and position. Anisotropic mappings, on the other hand were defined as those mappings that are correspondingly variant. Structure extraction (via a perceptual grouping process) and color histogram were shown to be representations of isotropic mappings. Texture analysis using a channel energy model comprised of even-symmetric Gabor filters was considered to be a representation of anisotropic mapping. Segmentation of an image and detailed object representation were not required.

An integration framework for these mappings was also described. The integrated framework took advantage of the strength of structure, color histogram and texture in their respective domains for retrieval. Results of retrieval of outdoor images by query and by classification using a nearest neighbor classifier were presented. The system was able to retrieve images ranging from purely natural objects, such as images of vegetation, flowers, water and sky, to images containing conspicuous structure, such as images of building, towers and bridges. In addition, the system gave good performance for retrieval of images containing intermediate-level structure such as images containing automobiles (even when they were of different colors). The judicious use of perceptual grouping to extract structure gives our system an edge over content-based image retrieval systems that retrieve images containing structural objects based purely upon color and texture. Results obtained show the efficacy of combining structure, histogram and texture for retrieval.

Acknowledgments

The authors wish to thank Prof. B. S. Manjunath (University of California at Santa Barbara) for his comments on Gabor filters. Thanks are also due to Sadia Sharif for her painstaking effort in locating and downloading images from the internet.

Appendices

A General rotation and reflection

For the case of a general rotation, the center of rotation can be shifted to the origin by a translation, followed by a rotation, and then a reverse translation of the same magnitude. The resulting transformation $\tau_{\mathbf{b}} R_{\theta} \tau_{\mathbf{b}}^{-1}$ is given as:

$$\tau_{\mathbf{b}} R_{\theta} \tau_{\mathbf{b}}^{-1} = \tau_{\mathbf{b}} R_{\theta} \tau_{-\mathbf{b}} = \tau_{\mathbf{b}} \tau_{(R_{\theta}(-\mathbf{b}))} R_{\theta} = \tau_{(R_{\theta}(-\mathbf{b})+\mathbf{b})} R_{\theta} \quad (\text{A.1})$$

where we have made use of the fact that, in general, $\varrho \tau_{\mathbf{b}} \varrho^{-1} = \tau_{\varrho(\mathbf{b})}$, such that ϱ represents either a rotation or a reflection.

For the case of a general reflection, after the rotation of the axis of reflection to align it along the original x-axis, the rotated axis can be translated to align it on the original x-axis. This is followed by the reflection $\acute{\kappa}$, then a reverse translation and a reverse rotation. The resulting transformation $R_{\theta} \tau_{\mathbf{b}} \acute{\kappa} \tau_{-\mathbf{b}} R_{-\theta}$ is given as:

$$\begin{aligned} R_{\theta} \tau_{\mathbf{b}} \acute{\kappa} \tau_{-\mathbf{b}} R_{-\theta} &= (R_{\theta} \tau_{\mathbf{b}}) \acute{\kappa} (R_{\theta} \tau_{\mathbf{b}})^{-1} = (\tau_{(R_{\theta}(\mathbf{b}))} R_{\theta}) \acute{\kappa} (\tau_{(R_{\theta}(\mathbf{b}))} R_{\theta})^{-1} \\ &= \tau_{(R_{\theta}(\mathbf{b}))} \acute{\kappa} R_{-2\theta} \tau_{(-R_{\theta}(\mathbf{b}))} = \tau_{(\acute{\kappa} R_{-2\theta}(-R_{\theta}(\mathbf{b})) + R_{\theta}(\mathbf{b}))} \acute{\kappa} R_{-2\theta} \end{aligned} \quad (\text{A.2})$$

B Euclidean invariance of ξ_{kl}

The energy functional expressed in equation 8 has a well-defined symmetry: it is invariant under the action of $E(2)$; invariant under translations $\{\mathbf{r}, \phi\} \rightarrow \{\mathbf{r} + \mathbf{b}, \phi\}$, rotations $\{\mathbf{r}, \phi\} \rightarrow \{R_{\theta} \mathbf{r}, \phi + \theta\}$ and reflections $\{\mathbf{r}, \phi\} \rightarrow \{\acute{\kappa} R_{-2\theta} \mathbf{r}, -(\phi - 2\theta)\}$.

The invariance of $s = \|\mathbf{r}_k - \mathbf{r}_l\|$ may be established as:

$$\begin{aligned}
\|\tau_{\mathbf{b}}\varrho\mathbf{r}_k - \tau_{\mathbf{b}}\varrho\mathbf{r}_l\|^2 &= \|(\varrho\mathbf{r}_k + \mathbf{b}) - (\varrho\mathbf{r}_l + \mathbf{b})\|^2 \\
&= \langle \varrho\mathbf{r}_k, \varrho\mathbf{r}_k \rangle + \langle \varrho\mathbf{r}_l, \varrho\mathbf{r}_l \rangle - 2 \langle \varrho\mathbf{r}_k, \varrho\mathbf{r}_l \rangle \quad (\text{B.1}) \\
&= \|\mathbf{r}_k\|^2 + \|\mathbf{r}_l\|^2 - 2 \langle \mathbf{r}_k, \mathbf{r}_l \rangle = \|\mathbf{r}_k - \mathbf{r}_l\|^2 = s^2
\end{aligned}$$

where \langle, \rangle denotes the dot product and ϱ is either a rotation or a reflection. The relation $\langle \varrho\mathbf{r}_k, \varrho\mathbf{r}_l \rangle = \langle \mathbf{r}_k, \varrho^{-1}\varrho\mathbf{r}_l \rangle = \langle \mathbf{r}_k, \mathbf{r}_l \rangle$ (where we consider ϱ as an operator) follows from the fact that, for example, in case $\varrho = R_\theta$, the adjoint operator of R_θ is given (in the matrix notation) as the conjugate-transpose of R_θ – which is real, and orthogonal, i.e., $R_\theta^t R_\theta = \mathbf{I}_e$, where \mathbf{I}_e is the identity. Hence, the adjoint operator of R_θ is $R_\theta^t = R_\theta^{-1}$. The adjoint operator of κ_θ is κ_θ^{-1} and the adjoint operator of ϱ is ϱ^{-1} . The first and second terms in the second line of equation B.1 may also be dealt with in a similar manner. Similarly, the invariance of q and t may also be established.

Translation invariance of equation 8 is easy to see, because:

$$\begin{aligned}
\xi_{kl}(\tau_{\mathbf{b}} \cdot \omega_b \mid \tau_{\mathbf{b}} \cdot \omega_k, \tau_{\mathbf{b}} \cdot \omega_l) &= \Lambda(q) \Lambda(st) \delta((\mathbf{r}_k + \mathbf{b}) - (\mathbf{r}_l + \mathbf{b}) - s\mathbf{e}_{kl}) \delta(\phi_b - \phi_l) \\
&= \Lambda(q) \Lambda(st) \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) \delta(\phi_b - \phi_l) \quad (\text{B.2}) \\
&= \xi_{kl}(\omega_b \mid \omega_k, \omega_l)
\end{aligned}$$

Invariance with respect to a rotation R_θ follows from:

$$\begin{aligned}
& \xi_{kl}(R_\theta \cdot \omega_b \mid R_\theta \cdot \omega_k, R_\theta \cdot \omega_l) = \\
& = \Lambda(q) \Lambda(st) \delta(R_\theta \mathbf{r}_k - R_\theta \mathbf{r}_l - sR_\theta \mathbf{e}_{kl}) \delta((\phi_b + \theta) - (\phi_l + \theta)) \\
& = \Lambda(q) \Lambda(st) \delta(R_\theta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})) \delta(\phi_b - \phi_l) \\
& = \Lambda(q) \Lambda(st) \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) \delta(\phi_b - \phi_l) = \xi_{kl}(\omega_b \mid \omega_k, \omega_l)
\end{aligned} \tag{B.3}$$

Invariance under a reflection κ_θ about the an axis holds, since:

$$\begin{aligned}
& \xi_{kl}(\kappa_\theta \cdot \omega_b \mid \kappa_\theta \cdot \omega_k, \kappa_\theta \cdot \omega_l) = \\
& = \Lambda(q) \Lambda(st) \delta(\acute{\kappa}R_{-2\theta} \mathbf{r}_k - \acute{\kappa}R_{-2\theta} \mathbf{r}_l - s\acute{\kappa}R_{-2\theta} \mathbf{e}_{kl}) \delta(-(\phi_b - 2\theta) + (\phi_l - 2\theta)) \\
& = \Lambda(q) \Lambda(st) \delta(\acute{\kappa}R_{-2\theta}(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})) \delta(-(\phi_b - \phi_l)) \\
& = \Lambda(q) \Lambda(st) \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) \delta(\phi_b - \phi_l) = \xi_{kl}(\omega_b \mid \omega_k, \omega_l)
\end{aligned} \tag{B.4}$$

where $\delta(R_\theta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})) = \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})$ and $\delta(\acute{\kappa}R_{-2\theta}(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})) = \delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})$, as explained in the following. Let ϱ represent either R_θ or $\acute{\kappa}R_{-2\theta}$, where R_θ , $\acute{\kappa}$ and $R_{-2\theta}$ are orthogonal matrices. Recall that the product of any number of orthogonal matrices is also orthogonal, and hence, has an inverse. (The set of orthogonal matrices forms a group – which is closed under multiplication of the elements of the group.) It can be shown that the linear operator ϱ (defined on the finite dimensional space \mathfrak{R}^2) has an inverse if and only if $\varrho(\mathbf{b}) = \mathbf{0} \Rightarrow \mathbf{b} = \mathbf{0}$, $\mathbf{0} \in \mathfrak{R}^2$ [20]. That is, $\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl} = \mathbf{0} \iff \varrho(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) = \mathbf{0}$. From which it follows that $\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl} \neq \mathbf{0} \iff \varrho(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) \neq \mathbf{0}$. Recall that $\delta(\mathbf{b}) = 0$ if $\mathbf{b} \neq \mathbf{0}$ (in the sense of distributions). Hence, the assertion follows, i.e., $\delta(R_\theta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})) = 0$, if and only if $\delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) = 0$, and similarly

$\delta(\kappa R_{-2\theta}(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl})) = 0$ if and only if $\delta(\mathbf{r}_k - \mathbf{r}_l - s\mathbf{e}_{kl}) = 0$. The case for $\delta(\mathbf{b})$ when $\mathbf{b} = \mathbf{0}$ may also be argued in a distributional sense.

C Fourier transform of translated, rotated and reflected image data

We shall derive a general case of the Fourier transform under the affine mapping $\mathbf{r} \mapsto \mathcal{M}\mathbf{r} + \mathbf{b}$, where $\mathcal{M} \in \mathbb{R}^{2 \times 2}$ is a real-valued invertible matrix, and $\mathbf{r}, \mathbf{b} \in \mathbb{R}^2$. Translations, rotations and reflections will be seen as special cases.

The two-dimensional Fourier transform of a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given as:

$$F(\nu) = \int_{\mathbb{R}^2} f(\mathbf{r}) e^{-j2\pi \langle \mathbf{r}, \nu \rangle} d\mathbf{r} \quad (\text{C.1})$$

where F denotes the Fourier transform of f , $\nu \in \mathbb{R}^2$ are the Fourier domain coordinates, and \langle, \rangle denotes the dot product. The Fourier transform of $f(\mathcal{M}\mathbf{r} + \mathbf{b})$ is given as:

$$\begin{aligned} \hat{F}(\nu) &= \int_{\mathbb{R}^2} f(\mathcal{M}\mathbf{r} + \mathbf{b}) e^{-j2\pi \langle \mathbf{r}, \nu \rangle} d\mathbf{r} \\ &= \frac{1}{|\det(\mathcal{M})|} \int_{\mathbb{R}^2} f(\hat{\mathbf{r}}) e^{-j2\pi \langle \mathcal{M}^{-1}\hat{\mathbf{r}} - \mathcal{M}^{-1}\mathbf{b}, \nu \rangle} d\hat{\mathbf{r}} \\ &= \frac{e^{j2\pi \langle \mathcal{M}^{-1}\mathbf{b}, \nu \rangle}}{|\det(\mathcal{M})|} \int_{\mathbb{R}^2} f(\hat{\mathbf{r}}) e^{-j2\pi \langle \mathcal{M}^{-1}\hat{\mathbf{r}}, \nu \rangle} d\hat{\mathbf{r}} \\ &= \frac{e^{j2\pi \langle \mathbf{b}, (\mathcal{M}^{-1})^t \nu \rangle}}{|\det(\mathcal{M})|} \int_{\mathbb{R}^2} f(\hat{\mathbf{r}}) e^{-j2\pi \langle \hat{\mathbf{r}}, (\mathcal{M}^{-1})^t \nu \rangle} d\hat{\mathbf{r}} \\ &= \frac{e^{j2\pi \langle \mathbf{b}, (\mathcal{M}^{-1})^t \nu \rangle}}{|\det(\mathcal{M})|} F((\mathcal{M}^{-1})^t \nu) = \frac{e^{j2\pi \langle \mathbf{b}, \hat{\nu} \rangle}}{|\det(\mathcal{M})|} F(\hat{\nu}) \end{aligned} \quad (\text{C.2})$$

where \hat{F} is the Fourier transform of $f(\mathcal{M}\mathbf{r} + \mathbf{b})$ and $\hat{\mathbf{r}} = \mathcal{M}\mathbf{r} + \mathbf{b}$. Hence, $d\hat{\mathbf{r}} = |\det(\mathcal{M})| d\mathbf{r}$, where $\det(\mathcal{M})$ denotes the determinant of the matrix \mathcal{M} , i.e., the Jacobian determinant of the transformation. In the above equation, $\hat{\nu} = (\mathcal{M}^{-1})^t \nu$

are the coordinates ν transformed due to the affine mapping. In addition, we have made use of the fact that $\langle \mathcal{M}^{-1}\mathbf{r}, \nu \rangle = \langle \mathbf{r}, (\mathcal{M}^{-1})^t \nu \rangle$. It may be noted that the result derived in equation C.2 remains valid for higher-dimensional (Euclidean) spaces other than \mathfrak{R}^2 .

For translation of an image, let \mathcal{M} be the identity. Hence, translation of an image $I(\mathbf{r}) \rightarrow I(\tau_{\mathbf{b}}(\mathbf{r}))$ transforms the Fourier transform of the image $\mathcal{I}(\nu) \rightarrow \mathcal{I}(\nu) e^{j2\pi\langle \mathbf{b}, \nu \rangle}$, where $\mathbf{r} = \{x, y\}$ are the space domain coordinates and $\nu = \{u, v\}$ are the Fourier domain co-ordinates.

For rotation of an image by an angle θ about the origin, let $\mathcal{M} = R_\theta$ and $\mathbf{b} = \mathbf{0} \in \mathfrak{R}^2$. It may be realized that $(R_\theta^{-1})^t = R_\theta$, since R_θ is an orthogonal matrix, i.e., the product $R_\theta^t R_\theta$ is the identity matrix. The determinant of R_θ is equal to 1 (because R_θ is a special orthogonal matrix). Hence, a rotation of $I(\mathbf{r}) \rightarrow I(R_\theta(\mathbf{r}))$, transforms the Fourier transform $\mathcal{I}(\nu) \rightarrow \mathcal{I}(R_\theta(\nu))$.

For reflection of an image in an axis inclined at an angle θ with the x-axis, let $\mathcal{M} = \kappa_\theta = \acute{\kappa} R_{-2\theta}$, where $\mathbf{b} = \mathbf{0} \in \mathfrak{R}^2$. The determinant of κ_θ is equal to -1, since $\det(\kappa_\theta) = \det(\acute{\kappa}) \det(R_{-2\theta})$. The determinant of $R_{-2\theta}$ is equal to 1 (since $R_{-2\theta}$ is a rotation and hence, a special orthogonal matrix), and the determinant of $\acute{\kappa}$ is equal to -1, because in matrix notation the mapping $\{x, y\} \rightarrow \{x, -y\}$ is obtained by:

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

where $\acute{\kappa}$ is the matrix on the left side of the above expression. Since κ_θ is an orthogonal matrix, $(\kappa_\theta^{-1})^t = \kappa_\theta$. Therefore, the reflection $I(\mathbf{r}) \rightarrow I(\kappa_\theta(\mathbf{r}))$ transforms the Fourier transform $\mathcal{I}(\nu) \rightarrow \mathcal{I}(\kappa_\theta(\nu))$.

References

- [1] Henning Muller, Wolfgang Muller, David McG. Squire, Stephane Marchand-Maillet, and Thierry Pun, “Performance evaluation in content-based image retrieval: overview and proposals,” *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, 2001.
- [2] Michael J. Swain and Dana H. Ballard, “Color indexing,” *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [3] Jonathan Ashley, Ron Barber, Myron Flickner, James Hafner, Denis Lee, Wayne Niblack, and Dragutin Petkovic, “Automatic and semi-automatic methods for image annotation and retrieval in QBIC,” in *Proc. SPIE: Storage and Retrieval for Image and Video Databases III, San Jose, California*, Feb. 1995, vol. 2420, pp. 24–35.
- [4] A. P. Pentland, R. Picard, and S. Sclaroff, “Photobook: Content-based manipulation of image databases,” *Int. Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1996.
- [5] J. R. Smith and S.-F. Chang, “VisualSEEk: a fully automated content-based image query system,” in *ACM Multimedia*, Nov. 1996, pp. 87–98.
- [6] Martin Szummer and Rosalind W. Picard, “Indoor-outdoor image classification,” in *IEEE International Workshop on Content-based Access of Image and Video Databases, Bombay, India*, 1998, pp. 42–51.
- [7] A. Vailaya, A. K. Jain, and H.-J. Zhang, “On image classification: City images vs. landscapes,” *Pattern Recognition*, vol. 31, no. 12, pp. 1921–1935, December 1998.
- [8] Qasim Iqbal and J. K. Aggarwal, “Retrieval by classification of images containing large manmade objects using perceptual grouping,” *Pattern Recognition, to appear*.
- [9] Qasim Iqbal and J. K. Aggarwal, “Applying perceptual grouping to content-based image retrieval: Building images,” in *IEEE International Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado*, June 1999, vol. 1, pp. 42–48.

- [10] Qasim Iqbal and J. K. Aggarwal, “Image retrieval via isotropic and anisotropic mappings,” in *IAPR Workshop on Pattern Recognition in Information Systems, Setubal, Portugal*, July 6-8, 2001, pp. 34–49.
- [11] J. Zweck and L.R. Williams, “Euclidean group invariant computation of stochastic completion fields using shifttable-twistable functions,” in *6th European Conference on Computer Vision (ECCV '00), Dublin, Ireland*, 2000, pp. 100–116.
- [12] Frederick M. Goodman, *Algebra, Abstract and Concrete*, Prentice Hall, Inc., 1998.
- [13] David S. Dummit and Richard M. Foote, *Abstract Algebra, Second Edition*, John Wiley & Sons, Inc., 1999.
- [14] John R. Durbin, *Modern Algebra: An introduction*, John Wiley & Sons, 1985.
- [15] P.C. Bressloff, J.D. Cowan, M. Golubitsky, P.J. Thomas, and M.C. Wiener, “Geometric visual hallucinations, euclidean symmetry, and the functional architecture of striate cortex,” *Phil. Trans. Royal Soc. London B. 356.*, pp. 299–330, 2001.
- [16] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.
- [17] Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification, second edition*, John Wiley and Sons, Inc., 2001.
- [18] “Visual Delights, Inc.,” <http://www.visualdelights.net>.
- [19] “Downloaded images,” Free Nature Pictures, <http://members.nbci.com/5555623/>; Info. for Travel Media, <http://www.sfvisitor.org/travelmedia/html/slides.html>; Dave’s Wall Paper, <http://davydicus.tripod.com/>; Photo Art of Nature, <http://www.photoartofnature.com>.
- [20] Erwin Kreyszig, *Introductory Functional Analysis with Applications*, New York, Wiley, 1978.

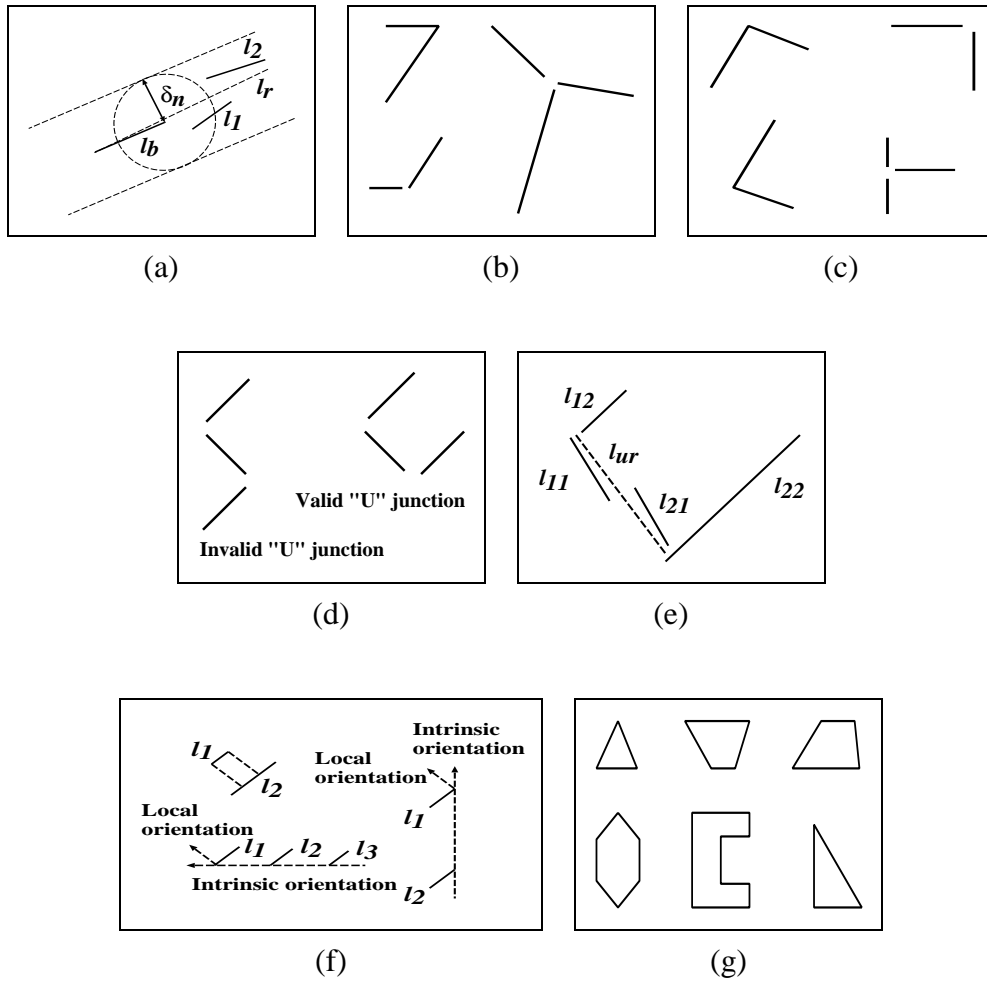


Fig. 1. Visualization of the groupings. (a) Longer linear line. (b) Coterminations. (c) "L" junctions. (d) "U" junction. (e) "U" junction. (f) Parallel groups. (g) Polygons.

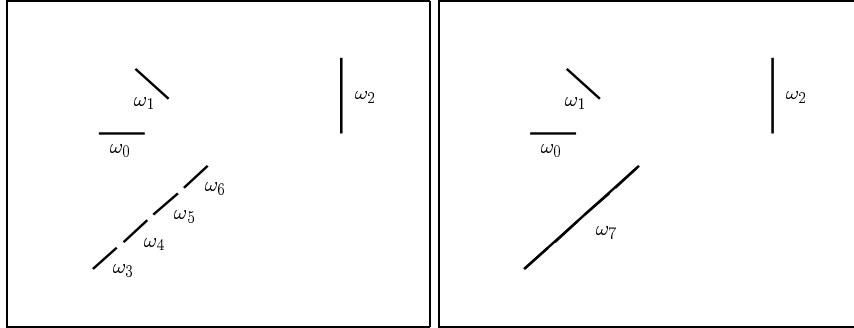
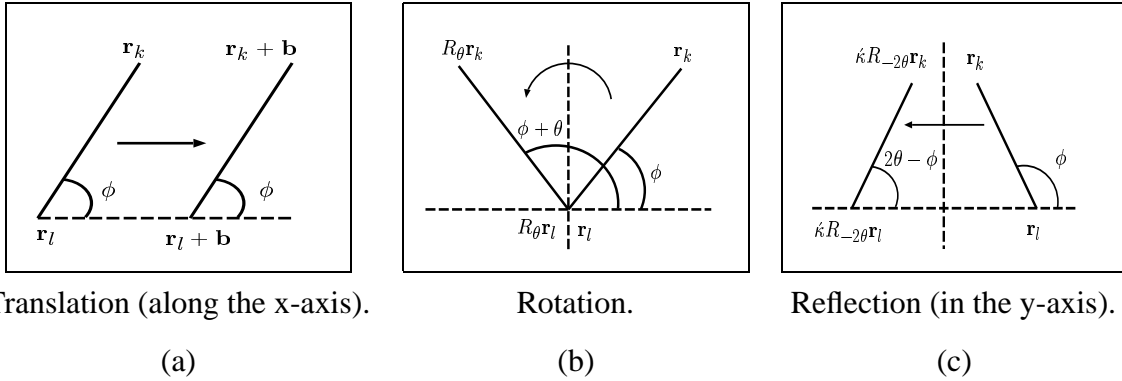


Fig. 2. $\omega_3, \omega_4, \omega_5$ and ω_6 combined to form ω_7 . At the lowest level of vision, ω_i are identified with edge segments.



Translation (along the x-axis).

(a)

Rotation.

(b)

Reflection (in the y-axis).

(c)

Fig. 3. Action of $E(2)$ on an edge segment ω_i . \mathbf{r}_k and \mathbf{r}_l represent the end-points of ω_i .



Query: Flowers, leaves and grass.

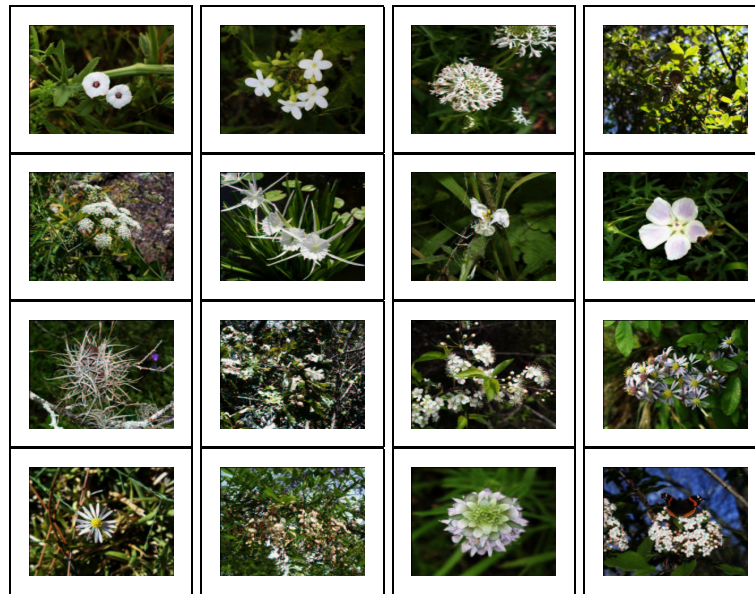
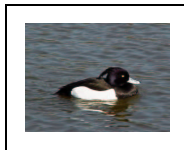


Fig. 4. Retrieval by image query (databases #1 and #2): Flowers, leaves and grass.



Query: Duck in water.

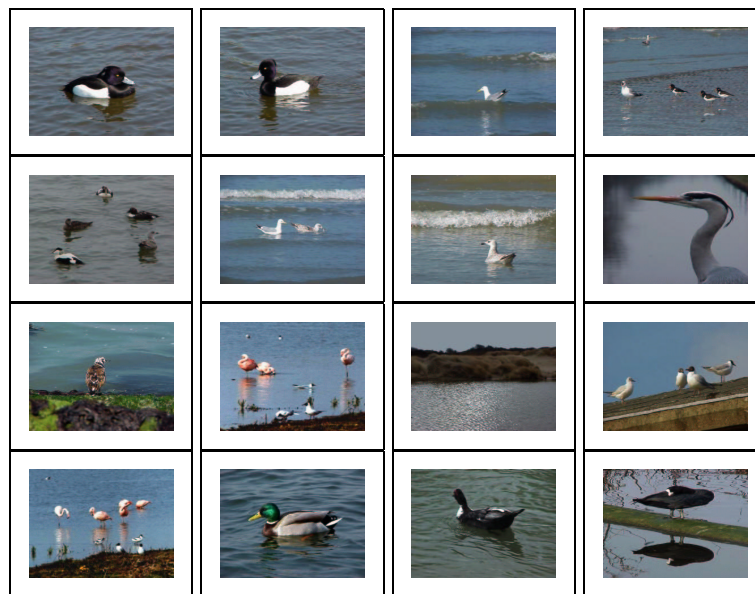


Fig. 5. Retrieval by image query (databases #2 and #3): Duck in water.



Query: An automobile.

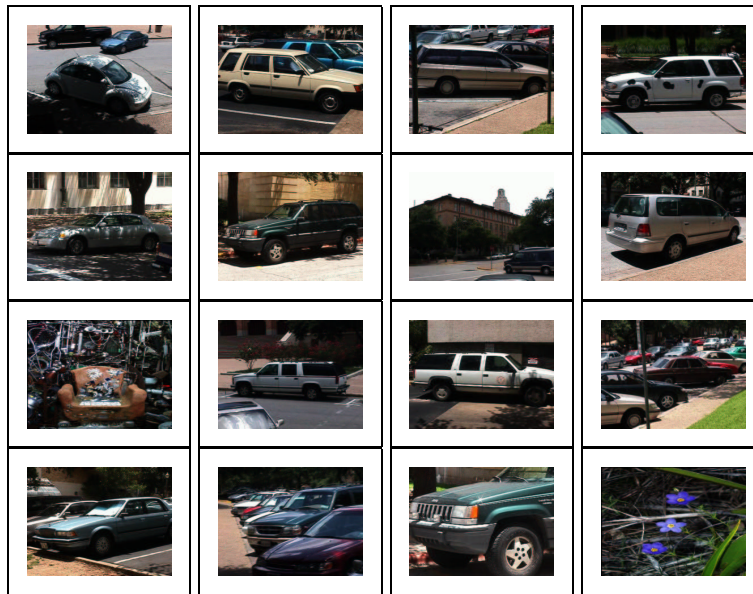
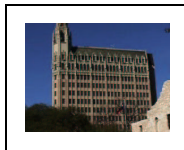


Fig. 6. Retrieval by image query (databases #1 and #2): An automobile.



Query: A building facade.



Fig. 7. Retrieval by image query (databases #1 and #2): A building facade.

Total	Training	Effective	Correct	RR
T		D	C	(C/D)
521	30	491	372	75.76%

Table 1

Retrieval by image classification (Database #2): Overall retrieval rate. T = Total # of images, D = Effective # of images, C = Correct and RR = Retrieval rate.

Class	T	R	C	Recall (C/T)	Precision (C/R)
Structure	255	203	188	73.73%	92.61%
Non-structure	140	151	119	85.00%	78.81%
Intermediate	96	137	65	67.71%	47.45%

Table 2

Retrieval by image classification (Database #2): Recall and precision. T = Total, R = Retrieved, C = Correct.

Class	Structure	Non-structure	Intermediate
Structure	188	13	54
Non-structure	3	119	18
Intermediate	12	19	65

Table 3

Retrieval by image classification (Database #2): Confusion matrix. Entries presented along rows, e.g., 188 Structure class images classified as Structure, 13 as Non-structure, and 54 as Intermediate.

Class	1-100	101-200	201-300	301-400	401-491	T	Q	Eff.=Q/T
Structure	85	70	57	32	11	255	189	74.12%
Non-structure	82	30	18	5	5	140	101	72.14%
Intermediate	56	23	8	3	6	96	55	57.29%

Table 4

Retrieval by image classification (Database # 2): Distribution of images *actually* belonging to a particular class in the “best matches” for that class, in intervals of 100 images, and the efficiency of the system. T = Total # of images belonging to a certain class, Q = # of images that actually belong to a certain class in the first T best matches for that class, and Eff. = Efficiency.