Color Invariant Object Recognition Using Entropic Graphs

Jan C. van Gemert, Gertjan J. Burghouts, Frank J. Seinstra, Jan-Mark Geusebroek

Intelligent Systems Lab Amsterdam, Informatics Institute, University of Amsterdam, Kruislaan 403, Amsterdam, The Netherlands

ABSTRACT: We present an object recognition approach using higher-order color invariant features with an entropy-based similarity measure. Entropic graphs offer an unparameterized alternative to common entropy estimation techniques, such as a histogram or assuming a probability distribution. An entropic graph estimates entropy from a spanning graph structure of sample data. We extract color invariant features from object images invariant to illumination changes in intensity, viewpoint, and shading. The Henze–Penrose similarity measure is used to estimate the similarity of two images. Our method is evaluated on the ALOI collection, a large collection of object images. This object image collection consists of 1000 objects recorded under various imaging circumstances. The proposed method is shown to be effective under a wide variety of imaging conditions. © 2007 Wiley Periodicals, Inc. Int J Imaging Syst Technol, 16, 146–153, 2006; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ima.20082

1. INTRODUCTION

Humans are capable of distinguishing the same object from millions of different images. Machines on the other hand have significant difficulty with this seemingly trivial task. One of the reasons that computational object recognition is such a hard problem is that machines take sensory information very literally, making object recognition vulnerable to accidental scene information. Such accidental variations include scale, illumination color, viewing angle, background, occlusion, shadows, shading, light intensity, highlights, and many more (Smeulders et al., 2001).

One approach to dealing with such photometric variations is found in the use of invariant features. Invariant features remain unchanged under certain operations or transformations and are used for various object recognition approaches. For example, the physical laws of image formation can be used to factor out accidental scene effects. The dichromatic reflection model by Shafer (1985) integrates body and surface reflection properties. This model may be extended upon, to obtain color invariant measurements (Forsyth, 1990; Funt and Finlayson, 1995; Geusebroek et al., 2001; Gevers and Stokman, 2004).

To compare object images, a similarity measure between image features is required. Often, similarity measures that require some parameter tuning are used in order to be applicable to other datasets or features. An example of such a parameter is the bin-size for histogram matching. A generic alternative is found in the use of unparametric similarity measures. We use entropic graphs (Hero et al., 2002) to compute an unparametric similarity between image features.

Color invariant features for object recognition are discussed is this paper. We employ an unparametric entropic similarity measure to match object images. Furthermore, the object recognition scheme is evaluated on a large dataset with real-world imaging conditions.

A. Related Work.
A popular method for object recognition is to apply salient point detectors. This method deals with problems, such as partial matching and occluded images. Specifically, Schmid and Mohr (1997) use salient point detection for indexing gray images. The detected points are subsequently made robust for scale changes and transformed to be rotationally invariant. In a similar approach, rotational and scale invariant keypoints allows for robust object detection (Lowe, 2004). Scale Invariant Feature Transform (SIFT) features are extracted and matched against a database. A Hough transform gives high probability to multiple features matched in one image. One problem with the interest point approach is the repeatability of the salient point detection. For example, detection may vary depending on pose, illumination, and background changes. Thus, salient points are not guaranteed to be the same over various imaging conditions. Moreover, for images without high curvature the method might not detect any salient points at all.

An alternative approach is given by Schiele and Crowley (2000) who take multiscale histograms of local gray value structure in an image. Translation invariance is given by the use of histograms. Rotational invariance is achieved by using several rotated versions

of a steerable filter in steps of 20°. This technique proves robust for rotated, occluded, and cluttered scenes. Grayscale images, however, lack a significant amount of information compared with color images. In our opinion, using color features in an object recognition approach is favorable, as color is a highly discriminative property of objects.

A biologically inspired object recognition method is presented with SEEMORE (Mel, 1997). Object matching is achieved with histograms of 102 different filters. Each filter responds to different image features like contour, texture, and color. Experiments are performed over a collection of 100 images. The highest experimental recognition rate of 97% is achieved with color and shape features. By using only color features, 87% recognition is achieved, as opposed to 79% without color. Thus color information may significantly improve object recognition.

Funt and Finlayson (1995) propose color invariant histograms for illumination-independent object recognition. Under the assumption of a slowly varying illumination, computed color ratios of neighboring pixels are color invariant. The color ratio is computed by taking derivatives of the logarithm of the color channels. Object recognition experiments were conducted for differing illuminations. Results show that histograms of color ratios outperform color histograms. Histogram bin size is usually set in an ad hoc manner, where the best bin size for a specific application is experimentally determined. Kernel density estimation tries to overcome the problem of selecting a suitable bin size for a histogram.

Color invariant histograms may be improved upon by using variable kernel density estimation (Gevers and Stokman, 2004). Here, an error propagation method is introduced to estimate the uncertainty of a color invariant channel. This associated uncertainty is used to derive the optimal parameterization of the variable kernel used during histogram construction. In this way, a robust estimator of invariant density is constructed. However, noise characteristics of the camera system are often not available.

A solution to image matching without the use of histograms is found in assuming prior knowledge about the probability distributions. A popular approach is mixture of Gaussian estimation (Westerveld and de Vries, 2004). However, not all processes can be described with a fixed parameterized model. Furthermore, assuming one distribution might severely over-simplify the complexity of the data.

Entropic graphs (Hero et al., 2002) offer an unparameterized alternative to histograms, circumventing choosing and fine tuning parameters such as histogram bin size or density kernel width. Alternatively, classifiers such as support vector machines may be used for object recognition (Pontil and Verri, 1998). A support vector machine (Vapnik, 1995) finds the best separating hyperplane between two classes. In contrast to support vector machines, entropic graphs allow to estimate information theoretic measures, like entropy, divergence, mutual information, and affinities.

In our approach, we extend the work of Schiele and Crowley (2000), Funt and Finlayson (1995), Gevers and Stokman (2004), and Hero et al. (2002), combining higher-order color invariant features with an entropy graph-based similarity measure. We extract color invariant features from object images, invariant to viewpoint, shadow and shading. As opposed to using a histograms or kernel density estimations, we employ entropic graphs. The Henze–Penrose similarity measure is then used to compute the similarity of two images. Finally, we evaluate our method on a large collection of object images. The object image collection consists of 1000 objects recorded under various imaging circumstances.

The paper is organized as follows. The next section discusses the color invariant model, Section 3 introduces entropic graphs and the Henze–Penrose (HP) similarity measure. Section 4 presents experimental results, and the Conclusions are given in Section 5.

2. COLOR INVARIANT FEATURES

Color is defined in terms of human observation. There is no one-to-one mapping of the spectrum of a light source to the perceived color. The Gaussian color model described in Geusebroek et al. (2000) approximates the spectrum with a smoothed Taylor series. In accordance with the human visual system, the Gaussian color model uses second-order spectral information. The zeroth-order derivative measures the luminance, the first-order derivative the "blue-yellowness," and the second-order the "red-greenness" of a spectrum.

A RGB image is measured in the Red, Green, and Blue sensitivity components of the light. The RGB sensitivities have to be transformed to the Gaussian spectral derivatives. In Geusebroek et al. (2000) an optimal transformation matrix with the Taylor expansion in the point $\lambda_0 = 520$ nm and with a Gaussian spectral scale of $\sigma_\lambda = 55$ nm is derived under the assumption of standard REC 709 CIE RGB sensitivities:

$$\begin{bmatrix} E \\ \partial_\lambda E \\ \partial_{\lambda\lambda} E \end{bmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \qquad (1)$$

When comparing images of the same object, differences in measurement due to the scene environment pose a problem. Taking two pictures of an object yields two different representations of the same scene. Differences in lighting conditions and in camera rotation change the recorded measurements of the scene. Image invariants deal with the problem to measure the information in a scene, independent of properties not inherent to the recorded object. Color invariance aims at keeping the measurements constant under varying intensity, viewpoint, and shading. In Geusebroek et al. (2001), several of these color invariants for the Gaussian color model are described. A property $\mathcal{C}$ invariant for viewpoint, shadow and shading invariance, is given by

$$\mathcal{C}_{\lambda^m x^n} = \frac{\partial^n}{\partial_{x^n}} \frac{1}{E(\lambda, x)} \frac{\partial^m}{\partial_{\lambda^m}} E(\lambda, x), \qquad m \geq 1, n \geq 0, \qquad (2)$$

where E is the energy. The $\mathcal{C}$ invariant normalizes the spectral information with the energy E and computes the spatial derivatives independent of the spectral energy. Note that the derivatives on the right-hand side of the equation represent measurements in the Gaussian color model. This makes the local spatial neighborhood invariant for intensity changes like shadow and shading.

Each pixel can be described with a color invariant feature vector. For example a second-order spatial representation of a pixel E yields the invariant counterparts of

$$\begin{aligned} \{ C_\lambda, C_{\lambda x}, C_{\lambda y}, C_{\lambda xx}, C_{\lambda xy}, C_{\lambda yy}, \\ C_{\lambda\lambda}, C_{\lambda\lambda x}, C_{\lambda\lambda y}, C_{\lambda\lambda xx}, C_{\lambda\lambda xy}, C_{\lambda\lambda yy} \}. \end{aligned} \qquad (3)$$

Note that only color information is used as all luminance information is discarded.

The invariant expressions up to second order are given by,

$$C_\lambda = \frac{E_\lambda}{E}, \quad C_{\lambda x} = \frac{E_{\lambda x}E - E_\lambda E_x}{E^2}, \quad C_{\lambda y} = \frac{E_{\lambda y}E - E_\lambda E_y}{E^2},$$

$$C_{\lambda\lambda} = \frac{E_{\lambda\lambda}}{E}, \quad C_{\lambda\lambda x} = \frac{E_{\lambda\lambda x}E - E_{\lambda\lambda}E_x}{E^2}, \quad C_{\lambda\lambda y} = \frac{E_{\lambda\lambda y}E - E_{\lambda\lambda}E_y}{E^2},$$

$$C_{\lambda xx} = \frac{E_{\lambda xx}E^2 - E_\lambda E_{xx}E - 2E_{\lambda x}E_x E + 2E_\lambda E_x^2}{E^3},$$

$$C_{\lambda yy} = \frac{E_{\lambda yy}E^2 - E_\lambda E_{yy}E - 2E_{\lambda y}E_x E + 2E_\lambda E_y^2}{E^3},$$

$$C_{\lambda xy} = \frac{E_{\lambda xy}E^2 + E_{\lambda x}E_y E - E_{\lambda y}E_x E - E_\lambda E_{xy}E - 2E_{\lambda x}E_y E + 2E_\lambda E_x E_y}{E^3},$$

$$C_{\lambda\lambda xx} = \frac{E_{\lambda\lambda xx}E^2 - E_{\lambda\lambda}E_{xx}E - 2E_{\lambda\lambda x}E_x E + 2E_{\lambda\lambda}E_x^2}{E^3},$$

$$C_{\lambda\lambda yy} = \frac{E_{\lambda\lambda yy}E^2 - E_{\lambda\lambda}E_{yy}E - 2E_{\lambda\lambda y}E_y E + 2E_{\lambda\lambda}E_y^2}{E^3},$$

$$C_{\lambda\lambda xy} = \frac{E_{\lambda\lambda xy}E^2 + E_{\lambda\lambda x}E_y E - E_{\lambda\lambda y}E_x E}{E^3}$$
$$- \frac{E_{\lambda\lambda}E_{xy}E - 2E_{\lambda\lambda x}E_y E + 2E_{\lambda\lambda}E_x E_y}{E^3}.$$

Indices denote differentiation by Gaussian convolution.

3. ENTROPIC GRAPHS

This section advocates entropic difference measures as an alternative to commonly used difference measures. The entropy measures the information content of a random variable. The information in one variable may be used to describe another, by using the mutual information between the two variables. High mutual information implies a high similarity between the two random processes. The difference between two probability distributions is given by the Kullback–Leibler (KL) divergence. The KL divergence between $p(x)$ and $q(x)$ may be seen as the average error by describing distribution $p(x)$ with a distribution $q(x)$. Entropic distance measures are theoretically sound and can capture nonlinear relations between probability distributions. Several applications of entropy can be found, for example, in image registration (Studholme et al., 1999), image retrieval (Vasconcelos, 2004), video modeling (Brand and Kettnaker, 2000), and saliency detection (Kadir and Brady, 2001).

The entropy of high-dimensional features is hard to estimate. Two methods to compare images are (1) histogram matching and (2) assuming a fixed probability density function. Entropy may be estimated from a histogram. A histogram is a fast and easy to compute method, making no assumptions on the underlying probability distribution. However, the problem of selecting a suitable histogram bin size is more of an art than science. Moreover, for a fixed resolution per dimension, the number of bins increases exponentially in the number of dimensions. Kernel density estimators are a general case of histogram methods (Wand and Jones, 1995). Nevertheless, the problems of selecting the size of the kernel and the curse of dimensionality also apply to kernel density estimation. Another solution to estimating entropy is by assuming prior knowledge about the probability distributions. When the probability distributions can be described with a parameterized model the computation of the entropy becomes feasible. However, not all processes can be described with a fixed parameterized model. Furthermore, assuming one distribution might severely over-simplify the complexity of the data.

Entropic graphs provide an unparameterized, efficient way to estimate the entropy of high-dimensional data (Hero et al., 2002).

An entropic graph is any graph whose normalized total weight (sum of the edge lengths) is a consistent estimator of Rényi's α-entropy. Examples of entropic graphs are the Minimum Spanning Tree and the k-nearest neighbor graph. One advantage of combinatorial methods is that the computation and storage complexity increase linearly in feature dimension. Additionally, graph-based estimators have fast asymptotic convergence rates and bypass the complication of choosing and fine tuning parameters such as histogram bin size or density kernel width.

Rényi's α-entropy (Rényi, 1961) is a generalization of the Shannon entropy and is defined by

$$H_\alpha(f) = \frac{1}{1-\alpha} \log \int_X f^\alpha(x) \, dx. \tag{4}$$

The α-entropy converges to the Shannon entropy $H(f) = -\int f(x) \log f(x) \, dx$, as $\alpha \to 1$. For α lesser than 1, the tails in the distribution are heavily weighted in the entropy.

The α-entropy can be estimated by the length of a minimal graph through sample points. Given a set $X_n = \{x_1, x_2, \dots, x_n\}$ of n i.i.d vectors in a d-dimensional feature space, the length of a graph is given by

$$L_\gamma(X_n) = \sum_{e \in G(X_n)} |e|^\gamma. \tag{5}$$

The graph G is over a suitable substructure, e.g. k-nearest neighbor graphs (see Fig. 1). Furthermore, e are edges in a graph connecting pairs of X_i's and $|e|$ denotes the Euclidean distance. The weighting $\gamma \in (0,d)$ relates to the value of α in the α-entropy as $\alpha = (d-\gamma)/d$, where d is the dimensionality of the feature space.

The entropic graph estimator

$$\hat{H}_\alpha(X_n) = \frac{1}{1-\alpha} \log L(X_n)/n^\alpha - \log c \tag{6}$$

is an asymptotically unbiased and consistent estimator of the α-entropy, where c is a constant independent of the data.

Entropic graphs can be used to estimate several similarity measures. These similarity measures include the α-mutual information, α-Jensen difference divergence, the HP affinity, and the α-
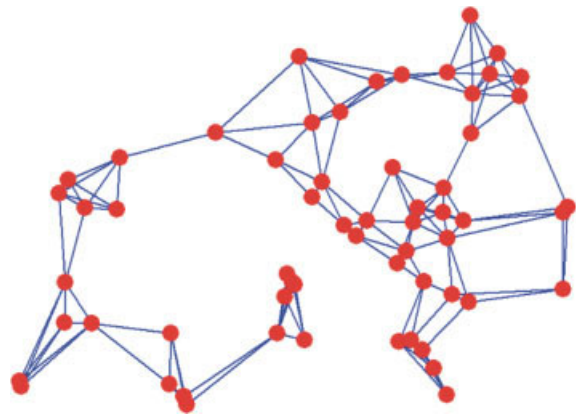
Figure 1. An example of a 4-nearest neighbors graph. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

geometric–arithmetic mean divergence. For $\alpha \to 1$, the α-divergence reduces to the KL divergence, and the α-mutual information to the Shannon mutual information. When α approaches 1, central differences between the two densities become highly pronounced. When α approaches 0, tail differences between two densities f and g become most influential. Therefore, if the feature densities differ in regions where there is a lot of mass one should choose α close to 1 to ensure locally optimum discrimination. Alternatively, if the tails or extreme values of the distribution describe the important events, α should be chosen close to 0.

One measure of similarity between probability distributions f and g is the Henze-Penroze (HP) (1999) affinity,

$$D_{\mathcal{HP}} = 2pq \int \frac{f(x)g(x)}{pf(x) + qg(x)} dx, \tag{7}$$

where $p \in [0,1]$ and $q = (1 - p)$.

In Neemuchwala and Hero (2005), an entropic graph algorithm for the Henze-Penroze affinity is introduced for given sample points $\{X_i\}_{i=1}^m$ of f, and $\{Y_i\}_{i=1}^n$ of g. For given samples, the value for p in Eq (7) is directly related to the number of samples: $p = \frac{m}{m+n}$. The entropic graph algorithm to estimate the HP affinity is given by

1. Construct the k-nearest neighbor graph on the sample points $\{X\} \cup \{Y\}$;
2. Keep only the edges that connect an X-labeled point to an Y-labeled point;
3. The HP-test affinity is given by the number of edges retained, divided by $(m + n)k$ for normalization.

This algorithm constructs an entropic graph on the edges that connect classes $\{X\}$ and $\{Y\}$. Counting the connecting edges implies a power weighting with 0. Therefore, the value for α in the estimated α-entropy is 1, emphasizing central differences between the two classes.

Figures 2 and 3 show two-dimensional examples of the HP affinity. The examples show sample points $\{X\}$ and $\{Y\}$ drawn from the same uniform distribution, and from a slightly different distribution, respectively. The affinity between the points drawn from the same distribution is significantly higher.
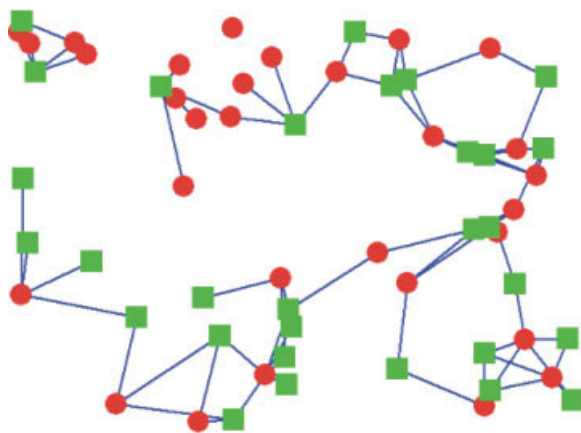
Figure 2. Example of the HP affinity in two dimensions. The sample points $\{X\}$ (circle) and $\{Y\}$ (square) are drawn from the same uniform distribution. The calculated affinity is 0.85. [Color figure can be viewed in the online issue, which is available at www.interscience. wiley.com.]
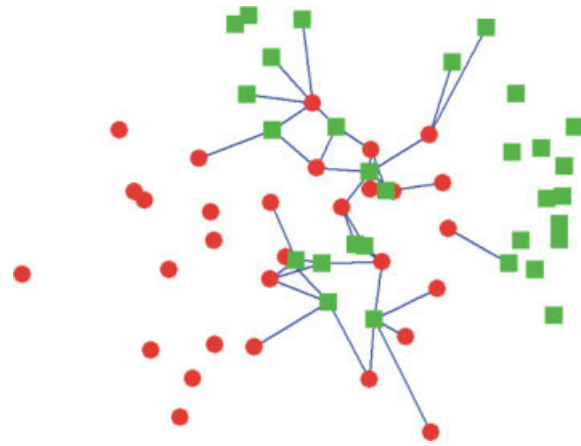
Figure 3. Example of the HP affinity in two dimensions. The sample points $\{X\}$ (circle) and $\{Y\}$ (square) are drawn from slightly different uniform distributions. The calculated affinity is 0.41. [Color figure can be viewed in the online issue, which is available at www.interscience. wiley.com.]

4. EXPERIMENTS

Performance is evaluated with an object recognition task on the ALOI dataset (Geusebroek et al., 2005). The ALOI collection consists of 1000 objects recorded under various imaging circumstances. For each object the viewing angle, illumination angle, and illumination color are varied. See Figures 4–6 and 10 for examples of the collection.

The combination of a large image dataset with a considerable variety of appearance offers a formidable challenge for object recognition. Object recognition is the problem of matching one appearance of an object against a standardized version. One object may give rise to millions of different images, as camera conditions may be varied endlessly. In our recognition experiment, one prototypical version of each object in the ALOI dataset is indexed and the diversity of recorded object variations in the collection are used for querying. An object is perfectly recognized when for all different variations the correct indexed object is returned. In this case, one may assume that the object can be recognized under a wide variety of real-life imaging circumstances.

A. Implementation. Entropic graphs are constructed with k-nearest neighbor search. The nearest-neighbor search is implemented using the approximation algorithm by Nene and Nayar (1997). The nearest-neighbor search is simple to implement and efficient in high dimensions. The algorithm proposed in Nene and Nayar (1997) constrains possible nearest neighbors of a point p inside a high-dimensional hypercube around p. For each dimension i, the points outside the limits $i - \varepsilon$ and $i + \varepsilon$ are discarded where the value of ε is typically small. For given distributions, ε can be set to an optimal value. For unknown data, however, ε may be empirically estimated. An offline sorted data structure makes discarding the points outside the hypercube efficient. In the case of entropic graph construction, this data structure needs to be computed for each query.

We extend the nearest-neighbor algorithm specifically for entropic graph construction. Particularly, the approximate nearest-neighbor algorithm is transformed to an optimal, exact algorithm. An entropic graph computes the k-nearest neighbors for each query
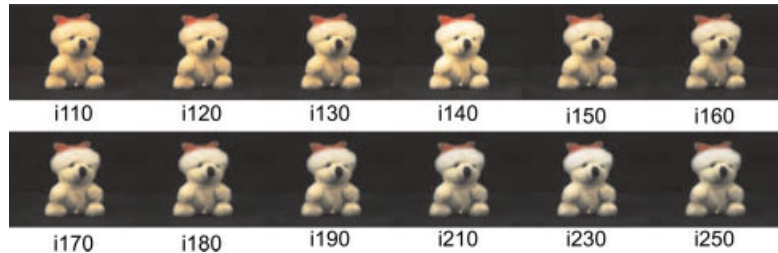
Figure 4. Example object from the ALOI collection, viewed under 12 different illumination color temperatures. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Figure 5. Example object from the ALOI collection, viewed from different viewing directions. [Color figure can be viewed in the online issue, which is available at www.interscience. wiley.com.]

image Q with every database image D. For each point p in the image D, the Euclidean distance to the kth nearest neighbor, which is furthest away, is stored. These distances are subsequently used as the ε values in computing the neighbors to p in Q. Because this ε value is the point furthest away in D, all points discarded can never be a k-nearest neighbor of $Q \cup D$. Hence, yielding an optimal value for ε, thus an exact, and more efficient entropic graph algorithm.

Before constructing the entropic graphs, we preprocess the images to extract features. The values of the color invariant N-jet are subsampled, thresholded, and whitened. We compute the second-order color invariant N-jet by convolution with a Gaussian of $\sigma = 2$. Because of Gaussian smoothing, there is a high correlation between neighboring pixel values. Therefore, we keep only 1 pixel in a block of 4 pixels. Subsampling will significantly increase the speed of the entropic graph construction. Color invariance is achieved by dividing by the intensity. Hence, the invariants are unstable when the intensity approaches zero. All pixels with intensity lower than 15 gray values are discarded. As the nearest-neighbor search uses a hypercube, whitened (or sphered) data are required. Whitening is achieved by dividing all data by a precomputed standard deviation for each invariant feature. About 1000 reference images are used for the calculation of the standard deviation. The extracted features are input for the entropic graph matching.

A single match on a standard PC takes 600 ms. Given the size of the dataset, all computations have been performed on the Distributed ASCI Supercomputer 2 (DAS-2), a wide-area distributed computer located at five different universities in The Netherlands (Bal et al., 2000). DAS-2 consists of five Beowulf-type clusters, one of

which contains 72 nodes, and four of which have 32 nodes (200 nodes in total). All nodes consist of two 1.0 GHz Pentium III CPUs, at least 1.0 GB of RAM, and are connected by a Myrinet-2000 network.

We used the parallel Horus framework introduced in Seinstra and Koelma (2004). The Parallel-Horus framework is a software architecture that allows nonexpert parallel programmers to develop fully sequential multimedia applications for efficient execution on homogeneous Beowulf-type commodity clusters. The core of the architecture consists of an extensive software library of data types and associated operations commonly applied in multimedia processing. To allow for fully sequential implementations, the library's application programming interface is made identical to that of Horus, an existing sequential library.

B. Results. We used the ALOI collection (Geusebroek et al., 2005) for evaluation of object recognition performance. For each object, 49 different appearance variations are evaluated. The 49 variations consist of 12 illumination color variations, 13 rotated views of the object, and 24 different illumination directions. Object recognition requires reference images and query images. The reference images are the ones recorded with white illumination and frontal camera with all lights turned on. The 49 query images per object are all matched against the 1000 reference images, making a total of 49,000 queries.

We compare our method with a standard work in object recognition (Gevers and Smeulders, 1999). This method uses histogram intersection on color invariant pixel values. The number of bins that
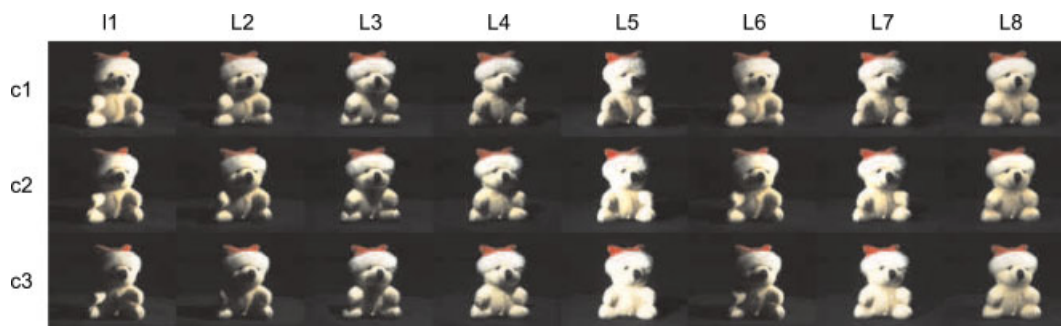
Figure 6. Example object from the ALOI collection, viewed under 24 different illumination directions. Each row shows the recorded view by one of the three cameras. The columns represent the different lighting conditions used to illuminate the object. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Correct Cumulative Object Recognition
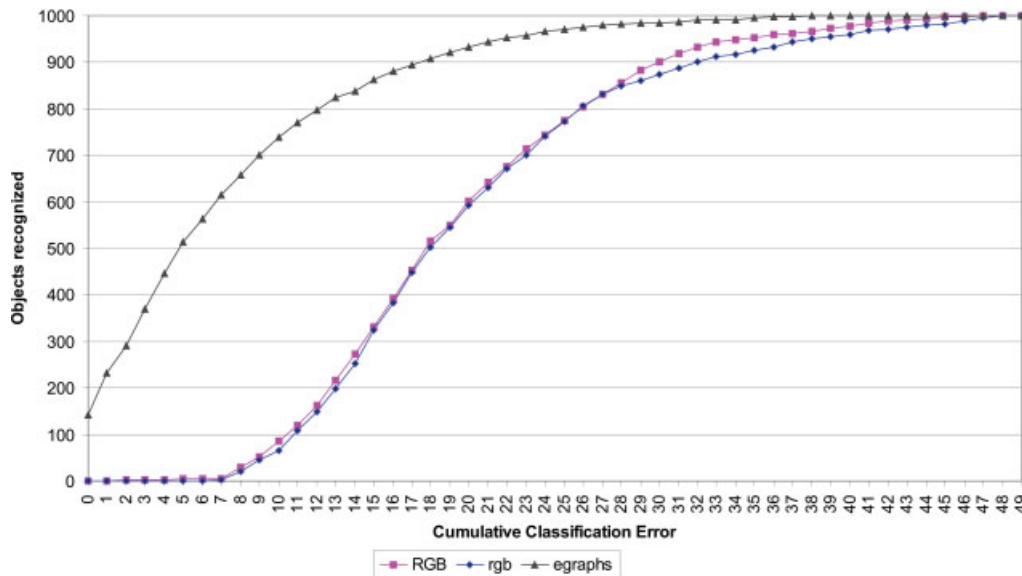
Figure 7. Correct cumulative object recognition. The number of object correctly recognized for an increasing error tolerance. The legend indicates different experiments with *RGB* for histogram intersection on RGB values, *rgb* for histogram intersection on normalized rgb values, and *egraphs* for our entropic graph algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

is used for the histograms is 32 per dimension, which is identical to value used in the original article. Object recognition results on the ALOI collection are computed for *RGB* histograms and for normalized *rgb* histograms.

Figure 7 shows the number of objects correctly recognized for an increasing error tolerance. Each of the 49 viewing condition gives rise to a possible mistake. Therefore, the graph displays the number of objects perfectly recognized if we allow 0 errors, to 1000 objects recognized if we allow all 49 mistakes. A desirable graph starts high and has a steep ascend. Our method starts at 141 objects and for a 5% error (2 errors) 291 objects are recognized. For histogram intersection no objects are recognized perfectly. Further-

more, it does not matter much if RGB or normalized rgb is used. However, the object recognition results based on entropic graphs significantly outperforms color histograms.

To acquire some insight in the results for both object recognition methods, we analyzed the recognition rate for each of the 49 viewing conditions. Figure 8 shows the object recognition performance of both methods grouped by color temperature and rotation direction. See Figures 4 and 5 for examples of these conditions. Note the considerable increase in recognition error for both methods under changes in illumination color (i250, ... , i110). Hence, both methods are not color constant, where the normalized color histograms suffer the most. Under different viewing angles (r30, ... , r330) our
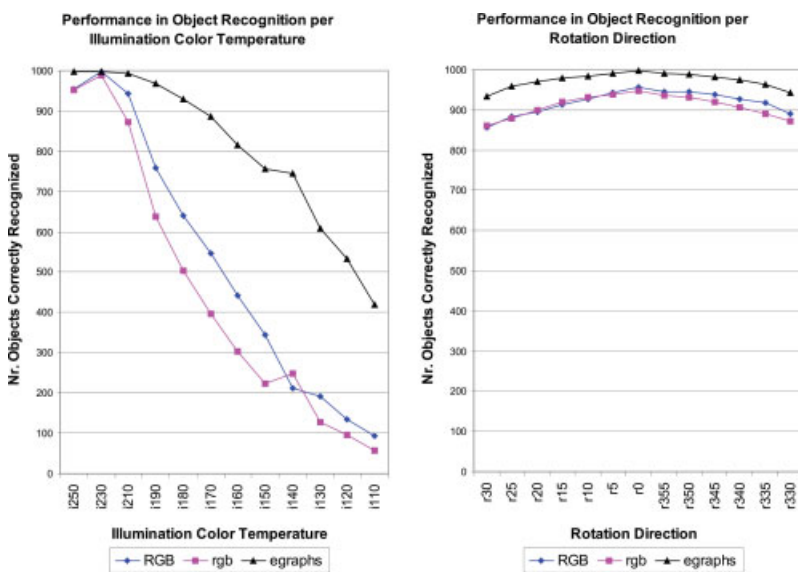
Figure 8. Number of objects recognized, grouped by color temperature and rotation direction. The conditions are abbreviated with letters. The prefix i indicates illumination color and r represents degrees of rotation. The legend indicates different experiments with *RGB* for histogram intersection on RGB values, *rgb* for histogram intersection on normalized rgb values, and *egraphs* for our entropic graph algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
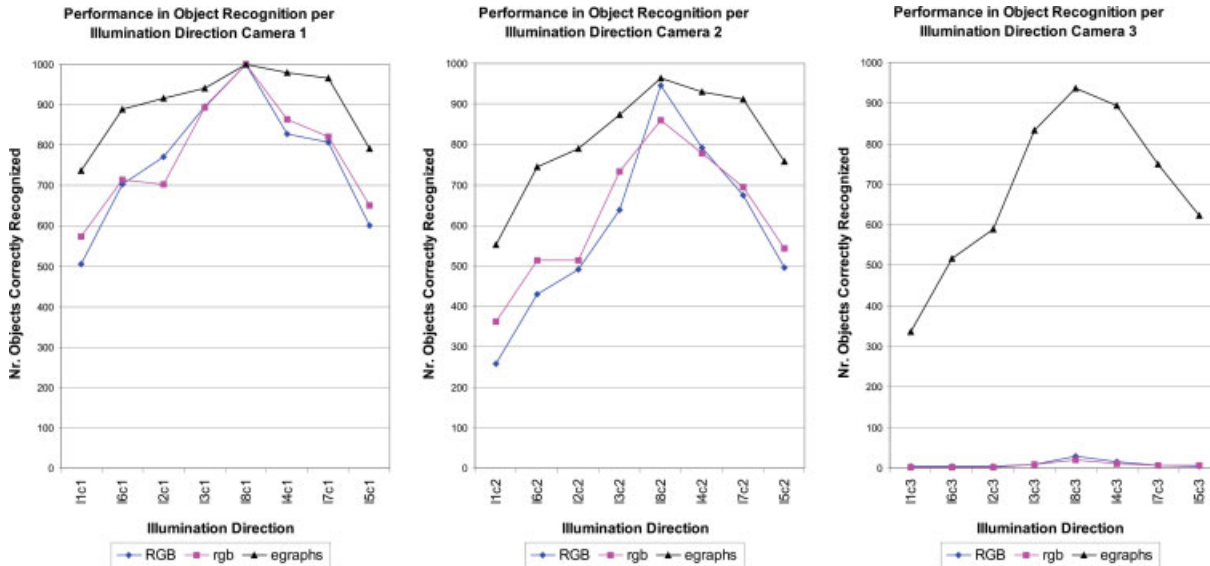
Figure 9. Number of objects recognized, grouped by camera and illumination direction. The conditions are abbreviated with letters. The prefix c conforms to camera position and l denotes the light source. The legend indicates different experiments with *RGB* for histogram intersection on RGB values, *rgb* for histogram intersection on normalized rgb values, and *egraphs* for our entropic graph algorithm. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

proposed method shows a high degree of robustness. The error for histogram intersection under different angles does not favor normalized or raw RGB values. Figure 9 shows the object recognition performance of both methods for each camera and illumination direc-

tion. See Figure 6 for examples of these conditions. For the lighting directions l1 and l5, performance degrades for both methods. This result is to be expected as the light shines only on a small part of the object. Performance further decreases as the position of the

Figure 10. 141 ALOI objects perfectly recognized by our method. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

camera (c1 vs c3) is farther away from the frontal position, where camera 3 is particulary difficult for the histogram-based method. The raw RGB histograms suffer most from changes in lighting directions, which is to be expected as no steps are taken to account for intensity changes. For all experiments, our method significantly outperforms the histogram-based methods.

For 1000 objects with 49 viewing conditions per object, we recognize 141 objects perfectly. That is, the number of objects that correctly match all different recordings. Given the diversity in recording circumstances, we may safely assume the objects will be recognized under a high variety of real-life imaging conditions. Figure 10 displays the perfectly recognized objects. These objects have no apparent visual similarity, indicating that our approach is not biased towards specific type of objects.

5. DISCUSSION AND CONCLUSIONS

In this paper, an unparameterized entropy estimator in combination with color invariant features are used for object recognition. We use color invariant features that keep image measurements constant under varying intensity, viewpoint, and shading. For similarity matching, we employ a measure based on entropic spanning graphs. Entropic graphs provide an alternative to traditional approaches of image matching such as assuming a fixed probability distribution or histogram binning. The parameters required are the number of nearest neighbors and the value for α in the α-entropy. The number k of the k-nearest neighbors is not critical; however, a higher k adds more robustness. The value of α is set through the power weighting γ, it determines the importance of the tails in a probability distribution. Therefore, α is an additional degree of freedom of the entropy, where $\alpha = 1$ is equivalent to the Shannon entropy. We introduce a new, efficient and exact entropic graph matching algorithm, based on an approximate nearest-neighbor algorithm. Despite an efficient algorithm, one drawback to entropic distance measures is that they are computationally more expensive than that of traditional approaches. Object recognition performance reported on a large dataset show that color invariant entropic graph matching significantly outperforms histogram-based methods.

ACKNOWLEDGMENTS

REFERENCES

H.E. Bal et al., The distributed ASCI supercomputer project, Operating Syst Rev, 34 (2000), 76–96.

M. Brand and V. Kettnaker, Discovery and segmentation of activities in video, IEEE Trans Pattern Anal Machine Intell, 22 (2000), 844–851.

D.A. Forsyth. A novel algorithm for color constancy, Int J Computer Vis, 5 (1990), 5–36.

B.V. Funt and G.D. Finlayson, Color constant color indexing, IEEE Trans Pattern Anal Machine Intell, 17 (1995), 522–529.

J.M. Geusebroek, G.J. Burghouts, and A.W.M. Smeulders, The Amsterdam library of object images. Int J Computer Vis, 61 (2005), 103–112.

J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, and A. Dev, Color and scale: The spatial structure of color images, In D. Vernon, editor, Sixth European Conference on Computer Vision (ECCV), Vol. 1, Springer

Verlag, Berlin, 2000, pp. 331–341. Lecture Notes in Computer Science, Vol. 1842.

J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, and H. Geerts. Color invariance, IEEE Trans Pattern Anal Machine Intell, 23 (2001), 1338–1350.

T. Gevers and A.W.M. Smeulders, Color-based object recognition, Pattern Recognit, 32 (1999), 453–464.

T. Gevers and H. Stokman, Robust histogram construction from color invariants for object recognition, IEEE Trans Pattern Anal Machine Intell, 26 (2004), 113–117.

N. Henze and M. Penrose, On the multivariate runs test, Ann Stat, 27 (1999), 290–298.

A. Hero, B. Ma, O. Michel, and J. Gorman, Applications of entropic spanning graphs, IEEE Signal Process Mag, 19 (2002), 85–95.

T. Kadir and M. Brady, Saliency, scale and image description, Int J Computer Vis, 45 (2001), 83–105.

D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int J Computer Vis, 60 (2004), 91–110.

B.W. Mel, SEEMORE: Combining color, shape and texture histogramming in a neurally-inspired approach to visual object recognition, Neural Comput, 9 (1997), 777–804.

H. Neemuchwala and A.O. Hero, Image registration in high dimensional feature space, In Proceedings of SPIE Conference on Electronic Imaging, San Jose, 2005.

S.A. Nene and S.K. Nayar, A simple algorithm for nearest neighbor search in high dimensions, IEEE Trans Pattern Anal Machine Intell, 19 (1997), 989–1003.

M. Pontil and A. Verri, Support vector machines for 3d object recognition, IEEE Trans Pattern Anal Machine Intell, 20 (1998), 637–646.

A. Rényi, On measures of entropy and information, In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, University of California Press, Berkeley, 1961, pp. 547–561.

B. Schiele and J. L. Crowley, Recognition without correspondence using multidimensional receptive field histograms, Int J Computer Vis, 36 (2000), 31–50.

C. Schmid and R. Mohr, Local gray value invariants for image retrieval, IEEE Trans Pattern Anal Machine Intell, 19 (1997), 530–534.

F.J. Seinstra and D. Koelma, User transparency: A fully sequential programming model for efficient data parallel image processing, Concurrency and Computat: Practice Experience, 16 (2004), 611–644.

S.A. Shafer, Using color to separate reflection components, Color Res Applic, 10 (1985), 210–218.

A.W.M. Smeulders, J.M. Geusebroek, and T. Gevers, Invariant representation in image processing, In IEEE International Conference on Image Processing, Vol. III, IEEE Computer Society, 2001, pp. 18–21.

C. Studholme, D.L.G. Hill, and D.J. Hawkes, An overlap invariant entropy measure of 3d medical image alignment, Pattern Recognit, 32 (1999), 71–86.

V.N. Vapnik, The nature of statistical learning theory, Springer-Verlag, New York, Inc., 1995.

N. Vasconcelos, On the efficient evaluation of probabilistic similarity functions for image retrieval, IEEE Trans Informat Theory, 50 (2004), 1482–1496.

M.P. Wand and M.C. Jones, Kernel smoothing, Chapman and Hall, London, 1995.

T. Westerveld and A.P. de Vries, Multimedia retrieval using multiple examples, In Proceedings of the International Conference on Image and Video Retrieval (CIVR2004), Dublin, Ireland, 2004.