Czech Pattern Recognition Workshop 2000, Tomáš Svoboda (Ed.)
Peršlák, Czech Republic, February 2–4, 2000
Czech Pattern Recognition Society

Performance Evaluation of the Multi-modal Neighbourhood Signature Method for Colour Object Recognition

Jiří Matas[1,2], Dimitri Koubaroulis[2] and Josef Kittler[2]

[1]Czech Technical University
Center for Machine Perception
Karlovo nám. 13, CZ 121 35

[2]Centre for Vision Speech and Signal Processing
University of Surrey, Guildford, GU2 5XH
Surrey, UK
D.Koubaroulis@ee.surrey.ac.uk

Abstract

The proposed method represents object or image colour structure by features computed from neighbourhoods with multi-modal colour density function. Stable invariants are derived from modes of colour density that are robustly estimated by the mean shift algorithm. The problem of extracting local invariant colour features is addressed directly, without a need for prior segmentation or edge detection. The signature is concise — an image is typically represented by a few hundred bytes, a few thousands for very complex scenes.

We demonstrate the algorithm's performance on a standard colour object recognition task using a publicly available dataset. Very good recognition performance (average match percentile 99.5%) was achieved in real time (average 0.28 seconds per match) which compares favourably with results reported in the literature. The method has been shown to operate successfully under changing illumination, viewpoint, object pose, non-rigid deformation and partial occlusion.

1 Introduction

Colour-based image and video retrieval has many applications and acceptable results have been demonstrated by many research and commercial systems during the last decade [17]. Very often, applications require retrieval of images where the query object or region cover only a fractional part of the database image, a task essentially identical to appearance-based object recognition with unconstrained background. Retrieval and recognition based on object colours must take into account the factors that influence formation of colour images: viewing geometry, illumination conditions, sensor spectral sensitivities and the surface reflectances. In many applications, illumination colour, intensity as well as view point and background may change. Moreover, partial occlusion and deformation of non-rigid objects must also be taken into consideration. Consequently, invariance or at least robustness to these diverse factors is highly desirable.

Most current colour based retrieval systems utilise various versions of the colour histogram [19] which has proven useful for describing the colour content of the whole image. However, histogram matching cannot be directly applied to the problem of recognising objects that cover only a fraction of the scene. Moreover, histograms are not invariant to varying illumination and not generally robust to background changes. Applying colour constancy methods to achieve illumination invariance for histogram methods is possible but an effective technique has yet to be developed [6]. Other methods addressing image (as opposed to object) similarity like those using wavelets for retrieval require image of objects at fixed pose to achieve invariance to illumination intensity changes [10]. Approaches based on moments of the colour distribution [9, 14] have been shown to perform well, but only images of planar scenes were tested. Finally, graph representations of colour content (like the colour adjacency graph [12]) have provided good recognition for scenes with fairly simple colour structure.

Departing from global methods, localised invariant features have been proposed in order to gain robustness to background changes, partial occlusion and varying illumination conditions. Histograms of colour ratios computed locally from pairs of neighbouring pixels for every image pixel [7] or across detected edges [8] have been used. However, both methods are limited due to the global nature of histogram representation. In the same spirit, invariant ratio features have been extracted from nearby pixels across boundaries of segmented regions for object recognition [16, 15]. Similarly, absolute colour features have been extracted from segmented regions in [18, 13]. However, reliable image segmentation is arguably a notoriously difficult task [17, 15]. Other approaches, split the image into regions where local colour features are computed. The FOCUS system [3] constructs a graph of the modes of the colour distribution from every image block. However, not only extracting features from every

image neighbourhood is inefficient, but also the features used do not account for illumination change. In addition, use of graph matching for image retrieval has often been criticised due to its high complexity.

The Multi-modal Neighbourhood Signature (MNS) method[11] addresses the colour indexing task by computing colour features from local image neighbourhoods with multi-modal colour probability density function. A robust mode estimator, the mean shift algorithm [5], was used to detect the modes of the density function. From the mode colours a number of local invariant features were computed, depending on the adopted model of colour change. Under different assumptions, the resulting multi-modal neighbourhood signatures (MNS) consisted of colour ratios, chromaticities, raw colour values or combinations of the above.

The advantages of extracting colour information from multi-modal neighbourhoods are manifold. Local processing guarantees robustness to partial occlusion and deformation of non-rigid objects. Data reduction is achieved by extracting features only from a subset of all image neighbourhoods. The computation time needed to create the colour signature is small since the most common neighbourhood type - uni-modal neighbourhoods - are ignored after being detected very efficiently. Moreover, illumination invariant features can be computed from pairs of mode values in a robust way. In particular, multi-modal neighbourhoods with more than 2 modes provide good characterisation of objects like the ball in Fig 1(d) and can result in efficient recognition on the basis of only few features. A rich description of colour content is achieved since a single-coloured surface can contribute to more than one multi-modal neighbourhood. Regarding retrieval, partial similarity queries are efficiently handled and localisation of the query instance in the database images is possible. Finally, the proposed representation allows the users to select exactly the local features they are interested in from the set of the detected multi-modal neighbourhoods of the query image.

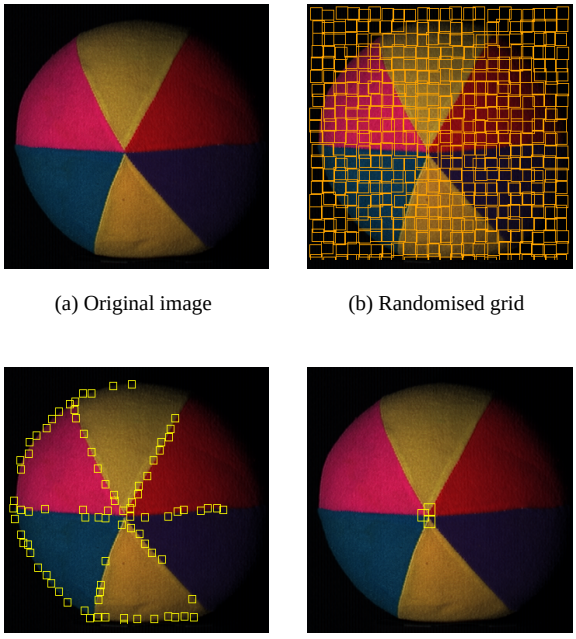
In our previous work, we reported on image retrieval experiments [11] where the multi-modal neighbourhood signature method performed successfully. However, in common image retrieval application, speed and storage requirements of the approach adopted are of similar importance as the relevance of the retrieved images. A method aspiring to challenge the dominance of the histogram-based approaches — the de-facto standard in the field — must have comparable run-time. Efficiency is highly desirable especially in web-based applications, where a comparatively large number of image signatures need be computed on-line. In addition, retrieval from large image databases or video sequences, as well as object recognition in real-time, require very fast signature matching and low storage requirements.

In this paper we evaluate the performance of the of the MNS method, with a focus on efficiency. Firstly, the computation speed for both signature creation and matching is measured and a number of modifications investigated. A randomised approach allowing trading off performance for speed is tested. Secondly, storage requirements for the MNS are measured. Finally, sensitivity of the algorithm to its control parameters was evaluated.

The rest of the paper is structured as follows. An outline of the computation of an MNS signature is given in section 2 and the matching technique is discussed in section 3. Section 4 presents details about the experimental setup and reports results on a baseline experiment previously reported in the literature. In section 5, the speed of MNS signature computation and matching is discussed. Brief information about storage requirements of the method is given in section 6. Sensitivity to a selected set of control parameters is analysed in section 7. Section 8 concludes the paper.

2 Computing the MNS signature

The image plane is split into rectangular regions of dimensions (b_x, b_y). To avoid aliasing each rectangle is perturbed with a displacement with uniform distribution in the range $[0, b_x/2)$, $[0, b_y/2)$, Fig. 1(b). For every neighbourhood defined by this randomised grid, the modes of the colour distribution are computed with the mean shift algorithm described in [5]. Modes with relatively small coverage are discarded as they usually represent noisy information. The neighbourhoods are then categorised according to their modality as uni-modal, bi-modal, tri-modal etc. (e.g. see Fig. 1)



(a) Original image

(b) Randomised grid

(c) Detected bimodal neighbourhoods

(d) Detected trimodal neighbourhoods

Figure 1: Multimodal neighbourhood detection

For the computation of the colour signature only multi-modal neighbourhoods are considered. For every pair of mode colours m_i and m_j in each neighbourhood, we construct a vector $v = (m_i, m_j)$ in a joint 6-dimensional domain denoted RGB^2.

In order to create an efficient image descriptor, we cluster

the computed colour pairs in the RGB^2 space and a representative vector for each cluster is stored. The colour signature we propose consists of the modes of the distribution in the RGB^2 space. For the clustering, the mean shift algorithm is applied once more to detect the mode values. The computed signature consists of a number of RGB^2 vectors depending on the colour complexity of the scene. The resulting structure is, thus, very concise and flexible.

Note that for the computation of the signature no assumption about the colour change model was needed. The parameters controlling mode detection, that is the kernel width and the neighbourhood size are dependent on the database images; the former being related to the amount of filtering (smoothing) associated with the mean shift and the latter depending on the scale of the scene. A multi-scale extension of the algorithm, though relatively straightforward to implement (e.g. by applying the MNS computation to an image pyramid), has not yet been tested.

Details about mode estimation using the mean shift algorithm are described in our earlier work [11]. In the same research, a number of different colour feature invariants was proposed to enable recognition under changing geometrical and illumination conditions.

3 Matching MNS signatures

A simple signature matching technique was applied to compute the dissimilarity between two MNS image signatures. The algorithm attempts to find a match for all model features assuming that the model signature contains only information about the object of interest. This assumption is realistic, since in object recognition applications a model database is typically built off-line in controlled conditions (e.g. with background allowing easy segmentation). In image retrieval applications, the query region is delineated by the user. Sometimes the full image is the object of interest and its MNS description is an appropriate model. However, if only part of the image is covered by the object of interest and the full image descriptor is stored as a model, a loss in recognition performance is likely.

On the other hand, test images may originate from scenes containing the model (query) object only as a fraction of the picture. The matching procedure is therefore asymmetric. A mismatch of a model feature is penalised whereas a mismatch of a test image feature is not. In other words the matching algorithm attempts to interpret the model signature as a distorted subset of the test image signature.

Let $I = 1..n$ and $J = 1..m$ be the indices of the model and test features respectively. We define a match association function $u(i) : I \rightarrow 0 \bigcup J$, $i \in I$, mapping each model feature I to the test feature it matched or to 0 if it didn't match. Similarly, a test association function $v(j) : J \rightarrow 0 \bigcup I$, $j \in J$, maps test to model features or null. A single threshold is introduced to determine a match between two features and reject outliers. The matching problem, i.e the problem of uniquely associating each feature $s_i^M, i = 1..n$ of the model signature with a test feature $s_j^T, j = 1..m$ and the computation of a match score is resolved in the following 4 steps:

1. Set $u(i) = 0 \quad \text{and} \quad v(j) = 0 \quad \forall i, j$. From each signature s compute the invariant features f_i^M, f_j^T according to the colour change model dictated by the application.
2. Compute all pairwise distances $d_{ij} = d(f_i^M, f_j^T)$ between the model and test features.
3. Set $u(i) = j, v(j) = i \quad \text{if} \quad d_{ij} < d_{kl} \quad \forall k, l$ with $u(k) = 0$ and $v(l) = 0$.
4. Compute signature dissimilarity as

$$D(s^M, s^T) = \sum_{(\forall\, i\,:u(i)\neq 0)} d_{ij} + \sum_{(\forall\, i\,:u(i)=0)} d_{max} \quad (1)$$

where d_{max} is the threshold value.

Computing overall image similarity, the quality of the model features that matched is taken into account and the score is penalised for any unmatched model features. Note that features are allowed to match only once. Apparently, the more model features matched, the lower the $D(s^M, s^T)$ value and the more similar the compared images.

4 Baseline Experiment

To compare MNS performance with results reported in the literature, we performed a well known colour object recognition experiment using a dataset collected by M. Swain. The database is publicly available [1] and has been used in a number of colour recognition experiments (e.g. [19, 7]). The model image set consists of 66 household objects imaged on black background under the same light (for a full colour image of the database see [19]) . The test set contains 32 images, a subset of model objects that are rotated, displaced or deformed (e.g. clothes). The test database and the corresponding model objects are shown in Fig. 2.

The MNS[1] evaluation followed the methodology adopted by Swain and Ballard [19] and consequently by Funt and Finlayson citeFunt-PAMI95 for recognition experiments using ratio histogram matching. However, in the experiments reported by Funt and Finlayson, 11 model and 8 test images with saturated pixels were removed from the database. Resolution of both model and test images is 128×90. No image preprocessing, sub-sampling or smoothing was applied before MNS signature computation. Default values of internal parameters (mean shift kernel width, neighbourhood size etc.) were used and the parameters were not specifically tuned for Swain's database. Retrieval results reported in [11] were obtained with identical MNS settings.

Assuming that illumination was kept approximately constant for all images in Swain's database the multi-modal neighbourhood signature was tested using 6D RGB^2 feature matching. For each test object, signature dissimilarity from 66 model signatures (as defined in section 3) was computed

[1]Current implementation of the MNS algorithm uses only bi-modal neighbourhoods for recognition although incorporating information from neighbourhoods with more than 2 modes is straightforward (e.g. by considering pairs of modes)

(a) Test objects used for the recognition experiment

(b) Model objects corresponding to the tests in (a)

Figure 2: Sample test and model images from Swain's database

and the rank of the correct pair stored. To allow comparison with previous experiments, recognition performance of the algorithm was assessed in terms of the average match percentile. The match percentile for each image matched is defined [19] as $\frac{N-r}{N-1}$ where N is the number of model images and r is the rank of the model image containing the test object.

Results are presented in Table 1. Recognition performance is compared with reported results for the colour indexing (CI) and colour constant colour indexing (CCCI) methods respectively. Recognition using the MNS compared favourably to the other two algorithms with an average match percentile of 99.5% using the default MNS parameters. The experiment was repeated for a range of mean shift kernel widths. Recognition performance reached 99.9%

Method	Rank				Average Match Percentile
	1	2	3	>3	
MNS (default)	27	2	2	1	0.995 (32 images)
CCCI	22	2	0	0	0.998 (24 images)
CI	29	3	0	0	0.999 (32 images)
MNS (Swain)	29	3	0	0	0.999 (32 images)

Table 1: Comparative colour object recognition results for Swain's database

(a) Clam chowder can

(b) Chicken soup can

(c) Mickey underwear

(d) Red-white jumper

Figure 3: Examples of Swain's model images with very similar red-white regions

which equals previously reported best performance for this dataset.

The objects that were not classified as rank 1 include mainly objects with red-white colour boundaries (e.g. Fig. 3). Such object are common in Swain's database and their MNS signature is similar.

5 Efficiency of the MNS method

In the previous section we argued that, besides recognition/retrieval performance, efficiency in terms of run-time and storage is a desirable characteristic of a retrieval system. For retrieval using the MNS method, the signature must be computed from the query object and then matched against a potentially large number of signatures of the database (model) images. The delay perceived by the user is the sum of the times required to complete the two stages. We first turn our attention to the speed of MNS computation and investigate the trade-off between performance and speed achieved via quasi-random sampling of the neighbourhood pixels.

5.1 Speed v. performance with quasi-random sampling

The computation of a MNS is dominated by repeated applications of the mean shift algorithm. Using a code profiler, it was established that data input, post and pre-processing operations account for a very small percentage of run time (in our case $\approx 5\%$). From the analysis of the mean shift implementation it is clear that its speed is approximately linearly related to the number of points used to initialise the search. The linear relationships is confirmed for Swain's database by results shown in Figure 4.

The mean shift algorithm is a gradient ascent search method. Since we are estimating only the *number* of modes (as opposed to associating every pixel with a mode which would effectively amount to local segmentation) it is not nec-

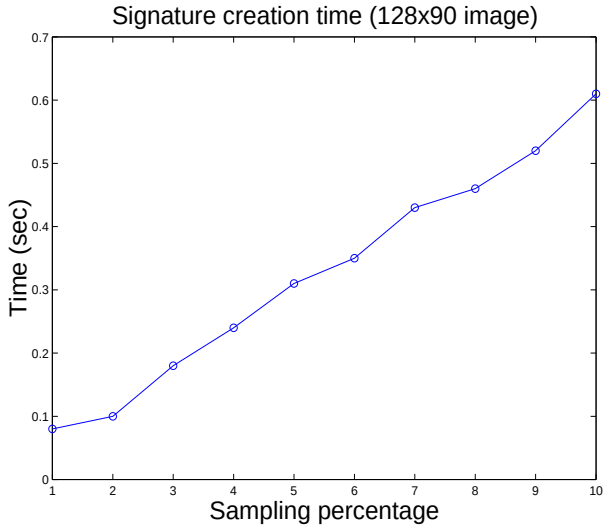


Figure 4: The time required to create an MNS signature depends linearly on the number of mean shift searches for local modes

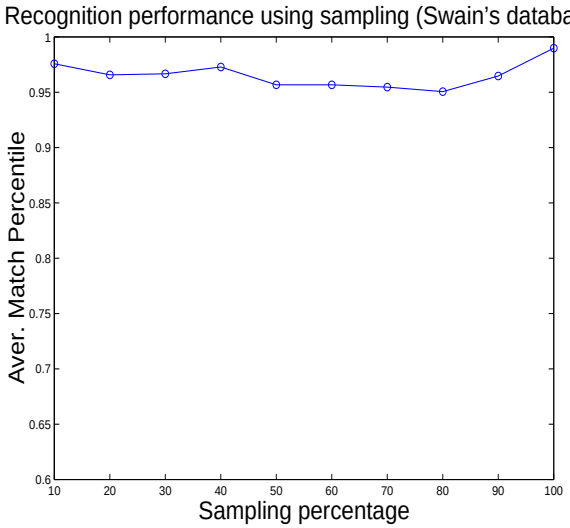


Figure 5: Recognition performance was generally insensitive to the percentage of the pixels initiating the mean shift

essary to use every pixel as a starting point of the search. Significant modes have support of at least $p\%$ of neighbourhood pixels (default $p = 10$), i.e. at least p percent of the pixel values lie within the kernel positioned at the mode. Initiating a search from every neighbourhood pixel is in most cases highly redundant, since it is sufficient if a search initiated from at least one of a random subset of the the $(N_x \times N_y)\frac{p}{100}$ pixels (where N_x, N_y denote neighbourhood size) converges to each mode.

Instead of selecting a random subset to initialise the search, we adopted a quasi-random procedure. Advantages of quasi-random sampling in comparison to random sampling are discussed in depth in chapter 7 of [20]. The subset of pixels selected from the neighbourhood was defined by a sampling array implemented as a dithering matrix [4]. This sampling ensures equal density of samples in the neighbourhood, prevents unfavourable spatial distribution of samples (e.g. all samples from one corner of the neighbourhood) and avoids using neighbouring pixels which are likely to have similar values.

The simple 2×2 dithering matrix is defined as

$$D^{(2)} = \begin{bmatrix} 0 & 2 \\ 3 & 1 \end{bmatrix} \tag{2}$$

and recursively a $2N \times 2N$ matrix is computed having the general form

$$D^{(2n)} = \begin{bmatrix} 4D^{(\frac{n}{2})} + D^{(2)}_{00}U^{(\frac{n}{2})} & 4D^{(\frac{n}{2})} + D^{(2)}_{01}U^{(\frac{n}{2})} \\ 4D^{(\frac{n}{2})} + D^{(2)}_{10}U^{(\frac{n}{2})} & 4D^{(\frac{n}{2})} + D^{(2)}_{11}U^{(\frac{n}{2})} \end{bmatrix} \tag{3}$$

where $U^{(n)}$ is the $N \times N$ matrix of ones.

In our case, the $M \times M$ dithering matrix was used where $M = 2^k,\ k \in \{1, 2, \ldots\}$ was the largest number for which $M < min(N_x, N_y)$. A subset of neighbourhood pixels was then specified by considering a percentage of the dithering matrix indices.

Swain's experiment was repeated using sampling as described above. Recognition performance (Fig. 5) was not significantly affected by a sampling rate change from 10 to 100% which was expected since mode estimation was shown not to be dependent on the number of pixels initiating the gradient ascent search of the mean shift algorithm. With the randomised approach, MNS can be computed in less than 0.1 seconds without loss of recognition performance.

5.2 MNS matching speed

Consider the $N_t \times N_m$ similarity matrix S where N_t is the number of the test features and N_m is the number of the model features respectively[2]. Profiling the code showed that besides the time spent on computing the similarity matrix, matching time was greatly affected by the computation of the signature dissimilarity $D(s^m, s^t)$ from S. Two implementations of the computation of D from S (described in section 3) were tested, both using a single parameter (the matching threshold T). Besides allowing to control the influence of outliers, the threshold on matching scores enables us to speed up the matching process since a test-model feature pair with dissimilarity greater than the threshold can never match (see section 7.2 for details on the role of the outlier rejection threshold).

In a first implementation, all pairwise dissimilarities which were below the rejection threshold T and the indices of the matching features were copied into a list. A fast merge-sort routine was then applied to sort the list by ascending dissimilarity score. Starting from minimum dissimilarity, features were marked as matched or not according to steps 2 to 4 of the matching algorithm. The process was terminated as soon as all model features were matched or the end of the list was reached.

In a different approach, the similarity matrix was pro-

[2] Matching methods which do not require full computation of S were not considered

Implementation	Average run-time	Speed
List	0.1 msec	236 matches/sec
Matrix	0.6 msec	207 matches/sec

Table 2: Comparison of the speed of two implementations of the MNS matching algorithm

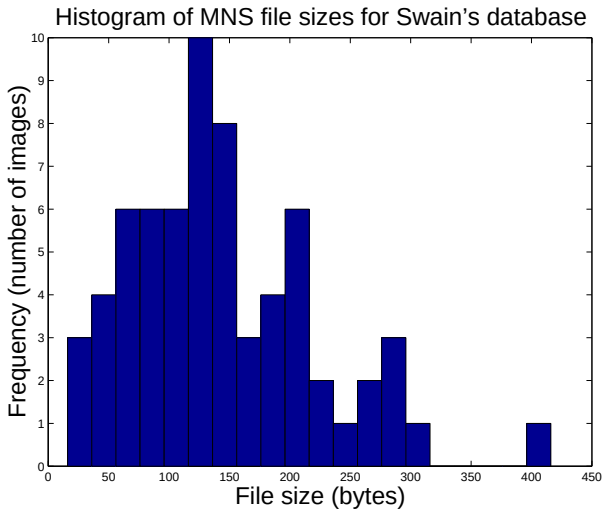


Figure 6: MNS signature were small in general with an average size of 150 bytes (8 bit mode value representation)

cessed directly. For each model feature (corresponding to a column) of S the minimum distance value was computed and the corresponding test feature (a row of S) was marked. The selected minima (one for each column of S) were inserted in a list which was then sorted by ascending distance score. For those model features that matched the same test feature, the second (or third or k-th when needed) minimum distance was considered for the one with the lower matching score and the minima list was resorted. The dissimilarity value was computed like before, using equation (1).

For Swain's database, cases where one test feature matches more than one model were rare and repeated sorting of the dissimilarity list was needed in very few cases. However, using a list as described above, results in a more compact implementation which for the reported experiments was faster than its matrix counterpart (Table 2).

Using the faster implementation of signature matching, an image retrieval experiment described in [11] was repeated. In the experiment, images containing Irish national colours were retrieved from a video archive of the Atlanta Olympic games provided by the BBC. Since the colour structure of the Irish flag is very simple compared to the structure of most objects in Swain's database, the matching was much faster. A speed of 600 image matches per second was achieved which is 50% faster than the previously reported result[11].

6 Storage requirements

Another important parameter of a retrieval system is the space needed to represent a single image. For many applications (especially those retrieving images from the World Wide Web), the number of images that will potentially be indexed is huge. The size of the signature determines the number of image descriptors that can be stored on a local disk by the retrieval system. A web-based search can thus be performed locally and only images similar to the query need to be downloaded. The MNS method stores pairs of RGB values originating from multi-modal neighbourhoods regardless of the representation used in matching. Since each colour component is stored as 4 byte floating point number (the mode RGB values are computed as averages and are not integers), the MNS signature file size is given by

$$S_n = n \times 2 \times 3 \times 4 = 24n \tag{4}$$

where n is the number of pairs of mode values in the signature and the numerical values correspond to the number of modes per RGB^2 vector (2), the number of colour components (3), and the number of bytes used for the representation of a floating point value (4).

The distribution of signature sizes for images in Swain's database showed that 70% of signatures were smaller than 1Kbyte; the average size was 0.88 Kbytes. The storage requirement of the MNS is certainly competitive with the colour histogram method, even if space saving techniques (e.g. eigenhistograms) are used. Nevertheless, the size of the signature can be reduced by a significant factor if stored using fixed point arithmetic. The range of mode values is $[0..255]$ and it is unlikely that more than a few bits after the decimal point are significant. Therefore, even 8, 10 or 12 bits per colour component, corresponding to 0, 2 and 4 binary digits after the decimal point may be considered. For Swain's dataset, 8 bit representation performed identically to the floating point representation. In this case, the average signature size was 150 bytes (see Table 6 for the distribution of signature sizes), which makes storing signatures of millions of images non-prohibitive.

7 Sensitivity to control parameters

Large number of control parameters and high sensitivity to their settings often prevent satisfactory use of computer vision programs by uninitiated users (or even anyone but the author). Ideally, parameter settings that reflect properties of the data should be learned. Such training procedures have not yet been incorporated in the proposed scheme. At least, results should not critically depend on ad hoc design choices. In this section we test the influence of three parameters on the performance of the MNS method.

7.1 Matching with different Minkowski metrics

Distance in the colour feature space is frequently computed using a Minkowski metric (also referred to as L-metric). The L-metric distance of order p between two n-dimensional fea-

Metric	Rank							Average Match Percentile
	1	2	3	4	5	6	>6	
L_1	27	2	2	0	1	0	0	0.995
L_2	27	2	2	0	0	1	0	0.994
L_3	27	2	2	0	0	0	1	0.993
L_∞	27	1	3	0	0	0	1	0.993

Table 3: Recognition results for different L-metrics

ture vectors x, y is defined as

$$L_p(x,y) = \left(\sum_{i=1}^{n} (x_i - y_i)^p \right)^{1/p} \quad (5)$$

Usually, colour based systems utilise L_1 (City-block distance), L_2 (Euclidean distance) or weighted versions of these. Some experiments in the literature have used the L_∞ metric defined as

$$L_\infty(x,y) = max(|x_i - y_i|), \;\; i = 1..n \quad (6)$$

The sensitivity of the MNS algorithm to the choice of p was tested experimentally. The results showed that recognition rate was not significantly affected by the selection of the L-metric. The number of test objects that were ranked up to rank 6 and above for a number of different metrics are presented in table 3.

The marginally better result for the L_1 metric was most probably due to its smaller sensitivity to large errors (outliers). Note that the time to compute the L_1 and L_∞ distance scores is minimal compared to other L-metrics requiring calculation of a p-th order root at each run.

7.2 Outlier rejection threshold

The MNS matching algorithm uses a threshold value T to increase robustness to outlier values in the computation of the dissimilarity value. Testing a large range of values of T we concluded that the MNS method is fairly insensitive to the threshold setting. Recognition performance deteriorated slowly (Fig. 7) even for extreme values of T and converged to a performance limit of 80% for the L_1 metric for large thresholds (practically infinite, i.e. bigger than any dissimilarity computed). The same test was repeated for the L_2 metric; the limit was 81%.

7.3 Exploiting modes with very dark colours

As a default, multi-modal neighbourhoods with very dark modes are not included in the MNS signature, since colour constant features (ratios) cannot be reliably computed from dark pixels. Dark colour (black) is used in both in Funt's [2] and Swain's [1] database to mark background – another reason for a default removal of very dark modes.

However, the removal prevents use of multi-modal neighbourhoods on the edge of objects,which carry discriminative information especially in the case of single-coloured objects. Since background is the same for test and model images, allowing the use of neighbourhoods containing background colours improved recognition. Table 4 presents comparative

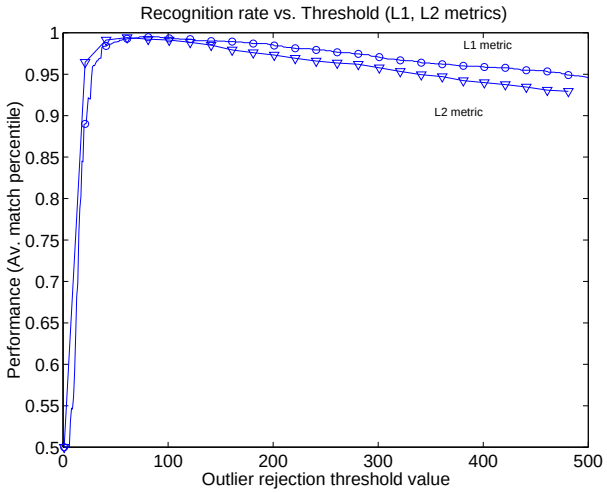


Figure 7: Recognition performance deteriorated slowly for increasing threshold values

results (using the L_∞ metric) for signatures created with and without taking dark (black) pixels into account. Only 2 objects out of 32 were misclassified of which one was ranked second and the other (the red-white jumper of Fig. 3d) ranked 17th (which affected the average rank performance score).

boundary	Rank							Average Match Percentile
	1	2	3	4	5	6	>6	
used	30	1	0	0	0	0	1	0.991
not used	27	2	2	0	0	1	0	0.992

Table 4: Recognition results with and without using neighbourhoods at object boundaries (L_∞ metric)

8 Conclusions

In the work reported in this paper we focused on efficiency related issues of the MNS method. The speed of both signature computation and matching was investigated. For signature computation, run-time was reduced by a factor of 90% resulting in an average MNS computation time of 0.1 seconds. MNS matching speed was also improved by a factor of 50% achieving a matching rate of 600 image matches per second for a sample image retrieval task on a SUN Ultra Enterprise 450 with quad 400MHz UltraSPARC-II CPUs. Signature size was measured and it supports the claim that the MNS methods is competitive for applications with fast matching and low storage requirements.

We tested the algorithm's performance on a standard colour object recognition task using a publicly available dataset. Very good recognition performance (average match percentile 99.5%) was achieved in real time (4 msec per match) for the default parameter settings of the MNS algorithm, which compares favourably with results reported in the literature. Recognition rate was fairly insensitive to large changes of the outlier rejection threshold and the dis-

tance function in the feature space for a selection of common Minkowski metrics (e.g. L_1, L_2, L_∞). The method has been shown to operate successfully under changing illumination, viewpoint, object pose, non-rigid deformation and partial occlusion.

Future improvements to the algorithm include introducing a training/learning stage to efficiently exploit discriminative colour characteristics inherent to the database at hand. For example, the distance used to compare colour features should be selected by learning the properties of the training database feature set. An extension to MNS involving a multi-scale approach to compensate for scale changes has not yet been studied. Finally, we intend to investigate the potential of multi-modal neighbourhoods with more than two modes for recognition and retrieval.

Acknowledgements

The ball image of Fig. 1 is from the image database of Simon Fraser University, available on-line [2]. The second author would like to thank EPSRC, Digital VCE and the Latsis Foundation for financially supporting this research. The first author acknowledges support under grant VS96049 of the Czech Ministry of Education.

References

[1] http://cs-www.uchicago.edu/users/swain/color-indexing/.

[2] http://www.cs.sfu.ca/ colour/image_db/.

[3] M. Das, E. Riseman, and B. Draper. FOCUS: Searching for Multi-coloured Objects in a Diverse Image Database. In *Computer Vision and Pattern Recognition*, pages 756–761, 1997.

[4] J. Foley, A. Dam, S. Feiner, and J. Hughes. *Computer Graphics: Principle and Practice*. Addison-Wesley, 1990.

[5] K. Fukunaga and L. Hostetler. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. In *IEEE Transactions in Information Theory*, pages 32–40, 1975.

[6] B. Funt, K. Barnard, and L. Martin. Is Machine Colour Constancy Good Enough? In *5th European Conference on Computer Vision*, pages 445–459, 1998.

[7] B. V. Funt and G. Finlayson. Color Constant Color Indexing. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 522–529, 1995.

[8] T. Gevers and W. M. Smeulders. Color-based Object Recognition. *Pattern Recognition*, 32(3):453–464, 1999.

[9] G. Healey and D. Slater. Global Color Constancy - Recognition of Objects by Use of Illumination Invariant properties of color distributions. *Journal Of The Optical Society Of America A-optics Image Science And Vision*, 11(11):3003–3010, 1994.

[10] D. Jacobs, P. Belhumeur, and R. Basri. Comparing Images Under Variable Illumination. In *IEEE Proceedings in Computer Vision and Pattern Recognition*, pages 610–616, 1998.

[11] D. Koubaroulis, J. Matas, and J. Kittler. MNS: A Novel Method for Colour Based Object Recognition and Image Retrieval. Technical Report VSSP-TR-6/99, University of Surrey, 12 1999.

[12] J. Matas. *Colour Object Recognition*. PhD thesis, University Of Surrey, 1995.

[13] K. Messer, J. Kittler, and M. Kraaijveld. Selecting Features for Neural Networks to Aid an Iconic Search Through an Image Database. In IEE, editor, *IEE 6th International Conference on Image Processing and Its Applications*, pages 428–432, 1997.

[14] F. Mindru, T. Moons, and L. V. Gool. Recognizing Color Patterns Irrespective of Viewpoint and Illumination. In *Proceedings of the Computer Vision and Pattern Recognition,Fort Collins, Colorado*, 1999.

[15] K. Nagao and W. Grimson. Recognizing 3d Objects Using Photometric Invariant. Technical report, Massachusetts Institute of Technology Artificial Intelligence Lab, 1995.

[16] S. Nayar and R. Bolle. Reflectance Based Object Recognition. *International Journal of Computer Vision*, 17(3):219–240, 1996.

[17] Y. Rui, T. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues. *Journal Of Visual Communication And Image Representation*, 10(1):39–62, 1999.

[18] R. J. Smith and S.-F. Chang. Integrated Spatial and Feature Image Query. *Multimedia Systems*, 7:129–140, 1999.

[19] M. J. Swain and D. H. Ballard. Color Indexing. *International Journal of Computer Vision*, 7:1:11–32, 1991.

[20] W. H. Press and B. P. Flannery and S. A. Teukolsky and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1992.