Colour-based Image Retrieval from Video Sequences

D. Koubaroulis[1,2] J. Matas[1,2] J. Kittler[1]

[1]Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, GU2 5XH, UK
d.koubaroulis@ee.surrey.ac.uk
[2]Czech Technical University, Center for Machine Perception, Prague, Karlovo nám. 13, CZ 121 35
matas@waltz.felk.cvut.cz

Abstract

The multimodal neighbourhood signature (MNS) algorithm represents local object appearance by stable colour-based invariants efficiently computed from image neighbourhoods with multimodal colour density function. Local feature extraction facilitates region-based interactive query specification and computation of illumination invariant features. The method allows for fast signature matching and supports retrieval of objects covering only a fraction of the database image. MNS signatures are generally compact and storage requirements are typically a few hundred bytes per image.

In this paper, the proposed algorithm is tested on a challenging region-based image retrieval task, searching for objects in a sports video sequence. Image regions depicting objects of typical interest in a sports image database are delineated by the users and used as queries. Acceptable retrieval results are presented for a number of experiments. Relevant images were retrieved regardless of background clutter dominating the scene, partial occlusion and/or non-rigid object deformation. Finally, the suitability of the approach is investigated in the context of web-based applications where efficient signature creation, good matching speed and minimal storage requirements are required.

1 Introduction

In the past few years, research in colour-based image and video retrieval has presented acceptable results and commercial systems have already been published in various application domains which require retrieval of images from an image collection (see [20] for a survey). 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 appearance must take into account the factors that influence the formation of colour images. Recorded colours depend on the viewing geometry, the illumination conditions, the sensor spectral sensitivities and the object's surface reflectances. In particular, illumination colour, intensity as well as view point and background may change in many applications. Other not so well modelled or not readily measured effects like mutual illumination and individual camera characteristics (e.g. gamma correction settings) also result in variations of object appearance. Moreover, partial occlusion and deformation of non-rigid objects must also be taken into consideration. Consequently, invariance or at least robustness to the above diverse factors is highly desirable.

The proposed Multimodal Neighbourhood Signature algorithm [11] represents *local* colour structure by illumination invariant features computed from image neighbourhoods with multimodal colour density function. Flexible signatures are constructed from neighbourhood modes efficiently located with the mean shift algorithm [4]. A number of illumination invariant features can be computed from a signature depending on the illumination conditions in the application environment. Object appearance is efficiently described by compact MNS signatures – their typical size ranging between a few hundred to a few thousand bytes for complex data. Signature matching is then applied to

retrieve images based on a computationally simple and efficient strategy. In addition, due to local processing of the data, region-based query specification and retrieval is facilitated.

In the work reported in this paper, we test the suitability of the MNS algorithm to retrieve images from a sports video sequence. The task is challenging since images contained in the sequence were shot under arbitrary conditions, indoors and outdoors. In addition, region-based queries are more difficult to resolve than full image-based retrieval. A number of typical queries are chosen e.g. to retrieve frames containing presentation legends or those frames relevant to British events and athletes. In most cases, the objects of interest cover part of the database images. Retrieval results show the ability of the MNS method to perform well in this real-world task. Relevant images were retrieved regardless of object size, deformation of non-rigid objects and significant background clutter.

The efficiency of the proposed method for image retrieval is also discussed in terms of its storage requirements, signature creation time and matching speed. In image retrieval, speed is a very important performance characteristic and any method aspiring to challenge the dominance of the histogram-based approaches must have comparable run-time. Especially for web-based applications, where a comparatively large number of image signatures need to be computed on-line, efficiency is highly desired. 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. The MNS method is competitive in both properties [13].

In the next section, a number of colour-based retrieval methods is briefly reviewed and contrasted with the proposed method. An outline of the computation of a MNS signature including details about modality estimation is given in section 3. The MNS signature matching technique is discussed in section 4. Section 5 presents details of the experimental setup and the results obtained are presented in section 6. The efficiency of the method in terms of its speed and storage requirements are reported in section 7 and section 8 concludes the paper.

2 Related Work

Most current colour-based retrieval systems utilise various versions of the colour histogram [22, 9] which has proven useful for describing the colour content of the whole image. Histograms record areas (or relative areas if normalised) and have no problems discriminating between objects with almost identical colours but with different sizes of colour region. For some synthetic databases this property is beneficial, since most objects undergo only rotations and translations and have approximately the same scale. In the presence of occlusion, object deformation or general view point change (e.g. as in the database used in this work) reliance on a non-invariant and/or global property like area or relative area will negatively affect performance. 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 colour constancy itself poses a number of challenging problems [5].

As an alternative to histograms, a more flexible signature representation and matching of global colour distributions was proposed by Rubner et al.[19]. Colour signatures were computed from the global colour histogram of an image and the area (mass) corresponding to most frequent colours were recorded in the signature. In contrast, multimodal neighbourhood signatures consist of measurements originating from local distributions and area (or relative area) is not taken into account neither for representation nor for signature matching. Other methods have also been used to retrieve scenes with similar appearance e.g. using wavelets [10] and moments of the image colour distribution [8, 15]. Finally, graph representations of colour content (like the colour adjacency graph [12] and its extension to a hybrid graph [18]) have provided good recognition for scenes with fairly simple colour structure.

To retrieve images based on object properties rather than image information, localised invariant features can been computed in order to gain robustness to background changes, partial occlusion and varying illumination conditions. Histograms of local invariants computed from pairs of neighbouring pixels for every image pixel [6] or across detected edges [7] have been proposed. However, both methods are limited due to the global nature of the histogram representation. Localised invariant features have been extracted from nearby pixels across boundaries of segmented regions for object recognition and retrieval [17, 16, 21, 14]. However, reliable image segmentation is arguably a notoriously difficult task [20, 16]. As an alternative to segmentation, the image is covered by compact regions from where local colour features are computed. For example, in the FOCUS system [3] a graph of the modes of the local colour density function was constructed from every image block. Both MNS and the FOCUS approach compute features from

modes of local colour distributions over image neighbourhoods, however, not only extracting features from every image neighbourhood is inefficient, but also the features used in FOCUS do not account for illumination colour change. In addition, use of graph matching for image retrieval has often been criticised due to its relatively high complexity.

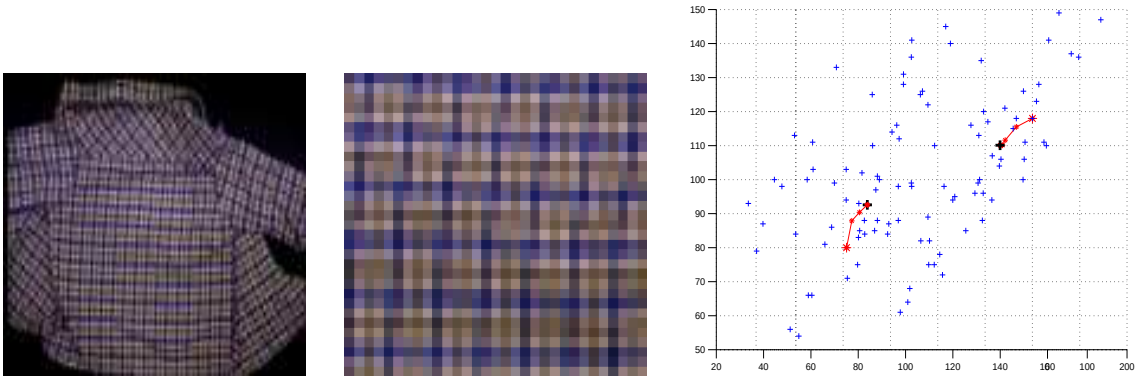
3 Computing the MNS Signature

The image plane is covered by a set of overlapping small compact regions. In the current implementation, rectangular neighbourhoods with dimensions (b_x, b_y) were chosen. Compact regions of arbitrary shape - or even non-contiguous compact sets of pixels - could have been used. Rectangular neighbourhoods were selected since they facilitate simple and fast processing of the data. To avoid aliasing each rectangle is perturbed with a displacement with uniform distribution in the range $[0, b_x/2)$, $[0, b_y/2)$. To improve coverage of an image (or image region), more than one randomised grids can be used, slightly perturbed from each other.

For every neighbourhood defined by such randomised grids, the modes of the colour distribution are computed with the mean shift algorithm described below. Modes with relatively small support are discarded as they usually represent noisy information. The neighbourhoods are then categorised according to their modality as unimodal, bimodal, trimodal etc.

For the computation of the colour signature only multimodal 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 establish the local maxima. The computed signature consists of a number of RGB^2 vectors depending on the colour complexity of the scene. The resulting structure is, generally, 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 seeking, 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 multiscale 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.



(a) original image (b) sample neighbourhood (c) colour values in the R-G domain

Figure 1: Mode detection for a multimodal neighbourhood using the mean shift algorithm

3.1 Computing Neighbourhood Modality Using the Mean Shift Algorithm

To establish the location of a mode of the colour density function the mean shift algorithm is applied in the RGB domain. The general kernel-based estimate of a true multivariate density function $f(\bar{x})$ at a point $\bar{x}_0$ in a d-dimensional data space is given by

$$\hat{f}(\bar{x}_0) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\bar{x}_i - \bar{x}_0}{h}\right) \tag{1}$$

where $\bar{x}_i,\ i = 1..n$ are the sample data points and K is the kernel function with kernel width h. In this work, we are not interested in the value of the density function at the point $\bar{x}_0$ but rather in the location of its maxima in the data space. A simple and efficient algorithm for locating the maximum density points was proposed by Fukunaga [4] for the case when the kernel function in Eq. 1 is the Epanechnikov kernel

$$K_E(\bar{x}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-\bar{x}^T\bar{x}) & \text{if } \bar{x}^T\bar{x} < 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where c_d is the volume of the unit d-dimensional sphere and $\bar{x}$ are the data points. The kernel has been shown to be robust to outliers and optimum in the sense of having minimum integrated square error [1].

The mechanism of the mean shift algorithm consists of iteratively shifting the kernel to the average of the data points within by the mean difference vector

$$M_h(\bar{x}) = \frac{1}{n_{\bar{x}}} \sum_{\bar{x}_i \in S_h(\bar{x})} (\bar{x}_i - \bar{x}) = \frac{h^2}{d+2} \frac{\hat{\nabla} f(\bar{x})}{\hat{f}(\bar{x})} \tag{3}$$

where $n_{\bar{x}}$ is the number of data points inside the hypersphere S of radius h centred at $\bar{x}$. Equation 3 is an estimate of the normalised gradient of the density function $f(\bar{x})$ in the d-dimensional space. As shown in [4], translation of the kernel centre towards the direction of the mean difference vector is equivalent to a gradient ascent to the local mode of the distribution. Convergence to the closest mode is guaranteed [2]. An example of the process is given in Fig. 1 for a multimodal neighbourhood with two prominent modes (shown on the diagram together with the convergence trajectories).

Due to the non-linearity of the kernel, the filtering preserves discontinuities, details and retains local image structure. This is particularly important for images containing small objects and fine colour texture. The speed of the algorithm was tested experimentally, and convergence was very fast (typically 4-5 iterations for complex data). Due to its advantageous properties the mean shift algorithm has been used in the past for image segmentation [2] and face tracking. For the MNS method, a computationally simple algorithm was implemented (see [13] for an efficient implementation).

4 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 query features assuming that the query signature contains only information about the object of interest. This assumption is realistic, since 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 query, a loss in retrieval performance is likely. On the other hand, database images may originate from scenes containing the query object only as a fraction of the picture. The matching procedure is therefore asymmetric. A mismatch of a query feature is penalised whereas a mismatch of a database image feature is not. In other words, the matching algorithm attempts to interpret the query signature as a distorted subset of the database image signature.

Let $I = 1..n$ and $J = 1..m$ be the indices of the query and test features (of the database image signature) respectively. We define a match association function $u(i) : I \rightarrow 0 \bigcup J,\ i \in I$, mapping each query feature i to the test feature it matched or to 0 if it did not match. Similarly, a test association function $v(j) : J \rightarrow 0 \bigcup I,\ j \in J$, maps

a database feature to a query feature or to 0 in case of no match. A single threshold T_h defines the maximum allowed distance between two matching features. The matching problem, i.e the problem of uniquely associating each feature $s_i^M, i = 1..n$ of the query 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 query and test features.
3. Set $u(i) = j,\ v(j) = i$ if $d_{ij} < d_{kl}$ and $d_{ij} < T_h \quad \forall\, k, l$ with $u(k) = 0$ and $v(l) = 0$.
4. Compute signature dissimilarity as

$$D(s^M, s^T) = \sum_{(\forall\, i\, : u(i) \neq 0)} d_{ij} + \sum_{(\forall\, i\, : u(i) = 0)} T_h \tag{4}$$

Computing overall image similarity, the quality of the query features that matched is taken into account and the score is penalised for any unmatched query features. Note that features are allowed to match only once. In general, the more query features matched, the lower the $D(s^M, s^T)$ value and the more similar the compared images. An efficient implementation of the matching algorithm is described in [13].

5 Experimental Setup

We tested the suitability of multimodal neighbourhood signature method for region-based image retrieval using a 30 minute video sequence of a BBC summary of the Atlanta Olympic games (1996). In total, 145 frames were randomly chosen from the sequence yielding a database of very different images, taken both indoors and outdoors. Object pose, scale as well as illumination were arbitrary (Fig. 2). No image was removed from the original selection and no image preprocessing was applied. The size of each frame was 144×176 pixels.

The objective was to test retrieval on typical queries relevant to a sports sequence. For example, one could ask for frames containing images of British athletes (thus searching for British national colours) or for frames introducing individual athletes or announcing events/results with an overlayed computer generated legend. In order to evaluate performance, sets of target images were manually selected from the original video sequence and added to the database. Frames for each target set were chosen so that they all contained objects with colour structure similar to the intended queries. The experimental setup was identical for all queries. The total number of the database images after adding the target images was 235.

Objects containing the sought colours in the target images were often occluded, non-rigidly deformed and/or of various sizes like objects with the Irish national colour structure in Fig. 3. Sometimes, the frames were taken at shot transitions where video editing effects were apparent. Finally, illumination conditions changed dramatically between some of the frames resulting in significantly different recorded colours. For example compare the query image image Fig. 4 with image Fig. 3(c) taken in the evening under very different light. For each experiment, the query image

Figure 2: Sample images from the BBC video summary of the 1996 Olympics

(a) (b) (c) (d)

Figure 3: Sample target images demonstrating possible cases of: (a) background clutter, (b) non-rigid deformation, (c) illumination change and (d) object size

was a rectangular fraction of a database image delineated by the user. For instance, a fraction of an Irish flag (Fig. 4) was used to retrieve images with similar colours.

The parameters involved in the computation of the signature and matching were not especially tuned for the task. Current implementation of the MNS algorithm uses only bimodal neighbourhoods for retrieval although incorporating information from neighbourhoods with more than 2 modes is straightforward (e.g. by considering pairs of modes). Colour object recognition results reported in [13] were obtained with identical MNS settings.

The MNS matching algorithm was implemented in C++ and tested on a SUN Ultra Enterprise 450 with quad 400MHz UltraSPARC-II CPUs. Signatures for both the database images and the query were computed in real-time (0.1 sec). Two consecutive searches with a randomised grid were performed with the same neighbourhood size (8×8 pixels). For the mean shift algorithm, a fixed kernel width of 25 units was used for the detection in the RGB space and 20 for the joint 6D space. Modes with coverage less than 10% of the neighbourhood were considered noise and were ignored. Low intensity modes (less than 5 percent of the luminance scale) were also not taken into account to improve stability especially in the case of relative colour feature matching. Although relative colour features (ratios) from pixels with saturated colours cannot not be expected to be stable under illumination SPD change, we did not remove saturated colours for the reported experiments. The matching threshold was also fixed and was dependent only on the nature of the features used. For example, for RGB^2 feature matching, matching threshold was fixed to 100 using the L1 metric for measuring distance in the 6-dimensional space. No sub-sampling or smoothing was applied before signature computation. All internal parameters (mean shift kernel width, neighbourhood size etc.) were set to default values, that is, they were not tuned for this image database.

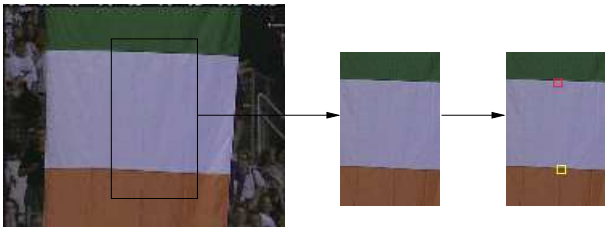


Figure 4: Query selection and representative multimodal neighbourhoods

(a) stripes on jumper (b) deformed flag in crowd (c) small flag on cap

Figure 5: Neighbourhoods that matched with the Irish flag query

6 Results

All database images were matched to each query image and sorted by decreasing similarity (increasing dissimilarity) to the query. For all reported experiments, results are displayed as the set of 20 images with top ranks. The rank order increases from left to right and from top to bottom. Therefore, the image at the upper left hand corner of the picture is ranked 1 and the one at bottom right is ranked 20. Performance was evaluated qualitatively according to the number of relevant images that were retrieved in the top 20 ranks of the retrieved list. The retrieval algorithm was based on the implementation described in [13]). A single signature match score was computed in approximately 1.5 milliseconds on average i.e. the retrieval proceeded at approximately 600 matches/sec.

British events. At first, we searched the database for images containing the British national colours. The query was part of a British flag from one database frame. Results are shown in Fig. 9 together with the selected query region. 14 images in the top 20 ranks were relevant to British events. The ones that were irrelevant, still contained combinations of the wanted colour structure but in different spatial arrangement.

Irish events. The next experiment, was identical to the one described in [11] only with a larger database. The objective was to retrieve frames related to Irish events or athletes. The query used is shown in Fig. 4. 6 out of 20 images were indeed relevant to Irish events. Another 6 relevant images were highly ranked in the first 50 positions. Only one image (Fig. 3(c)) was poorly ranked (113) because of significant illumination SPD change. The result is slightly worse than the one previously reported since more images with similar colours are added in the database this time.

Information legends. Another experiment was conducted searching for frames introducing athletes or displaying results and events. For this video, all these frames contained an informational strip with essentially identical colours for all images. At the same time, this experiment demonstrates MNS performance for synthetic objects like computer graphics which do not typically adhere to the rules of natural image formation process. 18 images in the top 20 ranks contained the object of interest and clearly the rest included similarly coloured parts. The retrieved images and the query region can be found in 7.

Commentary breaks. Apart from sport event coverage, the video sequence of the Olympic games contained parts of daily studio breaks where the events of each day were summarised. A search was performed for frames belonging to such summary studio sessions. All wanted frames had the same background colours and no significant illumination change was assumed. However, the camera position, the displayed day and the speaker were changing between sessions. 7 relevant images were among the first 20 (Fig 8). A whole frame (the best matched image) was used as the query. Like before, most irrelevant images in the top 20 set contained similar colours (in this case British national colours) in different spatial configuration.

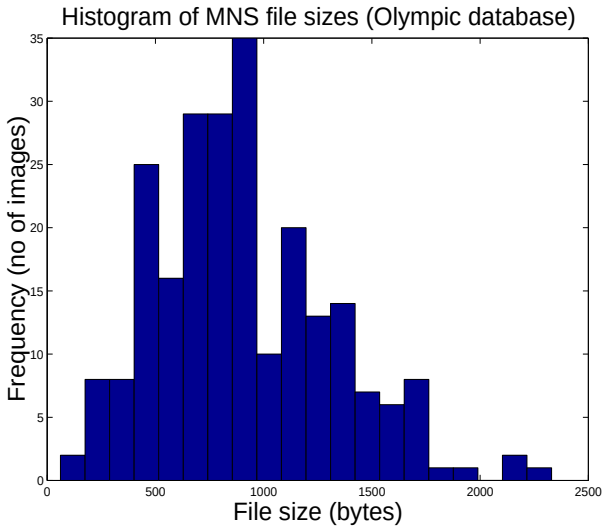


Figure 6: Distribution of file sizes of the 235 images of the Olympic database

Applying the MNS method for image retrieval generally gives acceptable results. Among the top ranked images are frames that contain very different instances of the query colour structure even as a small fraction of the database frame (see Fig 5 for some examples). A number of the frames that were ranked poorly, although they contained the object of interest, were images shot under very different illumination than that of the query image. The MNS method has already used illumination invariant features for retrieval [11]. However, experiments in this work were performed using only raw RGB values assuming that the illumination conditions were approximately constant for all the scenes in the database. Despite this strict assumption, the MNS method successfully retrieved a large number of relevant images.

7 Efficiency Considerations

Apart from signature creation and matching speed, a very 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 multimodal 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

$$S_n = n \times 2 \times 3 \times 4 = 24n \tag{5}$$

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 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. The distribution of signature sizes (Fig. 6) of the images of the Olympic database showed that 65% of signatures were smaller than 1Kbyte; the average size was 900 bytes (using fixed point arithmetic). For images with most complicated colour structure signature size did not exceed 2.5 Kb. Therefore,

(a) (b)

Figure 7: (a) Information legend query image and (b) top 20 retrieved images

storing MNS signatures of millions of images is non-prohibitive. The storage requirement of the MNS is certainly competitive with the colour histogram method, even if storage space saving techniques (e.g. eigenhistograms) are used.

Another advantageous property of the MNS method is efficient query specification. A rectangular area containing the object(s) of interest is specified by the user. Once the multimodal neighbourhoods are detected, the users can select those neighbourhoods they are indeed interested in or even define neighbourhoods they do not wish to match within the same rectangle. In fact, interactivity in query selection can be used from the application to *learn* from user's selections thus increasing the accuracy of the search. Implementing such an interactive query specification framework and a learning/training scheme will be subject of our future work.

8 Conclusions

In this paper, we demonstrated the potential of the Multimodal Neighbourhood Signature (MNS) method for image and video retrieval. Typical region-based queries were constructed from a selection of frames from a sports video sequence of the Olympic games and retrieval results were reported. The algorithm performed well and relevant images were successfully retrieved regardless of background clutter, partial occlusion and/or non-rigid deformation. In particular, very small regions were successfully matched like the small Irish flag on the swimmer's cap (Fig. 5(c)). MNS signatures were computed in real-time (0.1.sec) on a Sun UltraEnterprise 450 with quad CPUs at 400 MHz and search speed was 600 image matches per second. In addition, signature size was generally small (average 900 bytes) which, combined with fast signature computation and retrieval, seems promising for demanding web-based retrieval applications.

Although the MNS method supports search with illumination invariant features and use of spatial information for retrieval (e.g. for query localisation), these features were not tested in this work. Future improvements to the algorithm include introducing a training/learning stage to efficiently exploit discriminative colour characteristics inherent to the database at hand and a multi scale approach to compensate for scale changes. Finally, we intend to investigate the potential of multimodal neighbourhoods with more than two modes for recognition and retrieval.

9 Acknowledgements

The video sequence was provided by the British Broadcasting Corporation (BBC). The first author was supported by the EPSRC, the Digital VCE and the I. Latsis Foundation. The second author acknowledges support by the Czech Ministry of Education under the grant VS96049.

References

[1] Silverman B.W. *Density Estimation for Statistics and Data Analysis*. New York:Chapman and Hall, 1986.

[2] D. Comaniciu and P. Meer. Mean Shift Analysis and Applications . In *Proceedings of the International Conf. On Computer Vision*, pages 1197–1203, 1999.

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

[4] 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.

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

[6] B. Funt and G. Finlayson. Color Constant Color Indexing . *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.

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

[8] 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.

[9] J. Huang, R.S. Kumar, M. Mitra, J.-W. Zhu, and R. Zabih. Image Indexing Using Color Correlograms . In *IEEE Computer Vision and Pattern Recognition, San Juan, Puerto Rico*, pages 762–768, 1997.

[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, December 1999. (available at http://www.ee.surrey.ac.uk/Personal/D.Koubaroulis/thesis/pub/koubaroulis-tr699.ps.gz).

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

[13] J. Matas, D. Koubaroulis, and J. Kittler. Performance Evaluation of the Multi-modal Neighbourhood Signature Method for Colour Object Recognition. In *Proceedings of the Czech Pattern Recognition Workshop, Perslac, Czech Republic*, pages 27–34, 2000. (available at http://www.ee.surrey.ac.uk/Personal/D.Koubaroulis/thesis/-pub/matas-cprw00.ps.gz).

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

[15] F. Mindru, T. Moons, and L. Van Gool. Recognizing Color Patterns Irrespective of Viewpoint and Illumination. In *Proceedings of the Computer Vision and Pattern Recognition,Fort Collins, Colorado*, pages 368–373, 1999.

[16] K. Nagao and W. Grimson. Using Photometric Invariants for 3D Object Recognition. *Computer Vision and Image Understanding*, 71(1):74–93, July 1998.

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

[18] K. Park, Il-Dong Yun, and Sang Uk Lee. Color Image Retrieval Using a Hybrid Graph Representation. *Journal of Image and Vision Computing*, 17(7):465–474, 1999.

[19] Y. Rubner, C. Tomasi, and L. Guibas. A Metric for Distributions with Applications to Image Databases . In *IEEE Proceedings of the International Conference On Computer Vision*, pages 59–66, 1998.

[20] Y. Rui, T.S. 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.

[21] R. J. Smith and S.-F. Chang. Integrated Spatial and Feature Image Query . *Multimedia Systems*, 7(2):129–140, 1999.

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

(a) (b)

Figure 8: (a) The studio break query and (b) top 20 retrieved images

(a) (b)

Figure 9: (a) British national colours query and (b) top 20 retrieved images