Distinct Multi-Colored Region Descriptors for Object Recognition

Sarif Kumar Naik and C. A. Murthyˇ

Abstract— **The problem of object recognition has been considered here. Color descriptions from distinct regions covering multiple segments are considered for object representation. Distinct multi-colored regions are detected using edge maps and clustering. Performance of the proposed methodologies has been evaluated on three datasets and the results are found to be better than existing methods when a small number of training views is considered.**

Index Terms— **Object representation, Object descriptor, Object recognition, Object matching, Image representation**

I. Introduction

The challenges involved in object recognition are mainly the efficient representation and then the comparison of two objects through their representations. Broadly speaking, there are two types of approaches to object representation. While the first utilizes the knowledge gained from the spatial arrangements of the "shape features" such as the edge elements, boundaries, corners and junctions, the other uses the brightness or color features obtained more directly from the object images [1]. But, there are limitations to any algorithm which uses only either shape features or color features. The representation scheme should carry the color information and its pattern of appearance on the object surface. This study proposes a scheme to describe an object in such a way that the description contains the color information as well as the patterns of colors on the object surface. Note that, in most of the cases, wherever there is a shape or structural information in the object, the corresponding patterns in the image possess discontinuities in colors. Thus extraction of information regarding patterns of colors automatically leads to extracting shape and structural information of the object.

There are several approaches to object representation such as histogram based, eigenspace based, edge and corner based, graph based representations etc. Among histogram based methods the work by Swain and Ballard [2] is one of the earliest works which used color as a primal cue for object recognition and image retrieval. Stricker [3] introduced indexing technique based on boundary histogram of multicolored objects. Histogram based approach is an attractive method for object recognition because of its simplicity, speed and robustness [4]. Although it is simple the main drawbacks of this approach are its inability to encode shape and structural information of the objects, and the usage of only color information for distinguishing the objects.

The standard procedure in eigenspace based methods is to represent an object by considering the whole image as a vector and projecting it over a set of eigenvectors to achieve data compression and reduction of redundant information. Generally, the eigen vectors corresponding to the dominant eigenvalues are found using Principal Component Analysis (PCA). Some of the earliest works on object recognition using eigenspace based representation are by Murase and Nayar in [5], [6] and Truk and Pentland [7]. These methods are effective when eigen space captures the characteristics of the whole database. For example, when all the object images have uniform known background. If there is a large variation in the images, performance of the methods can deteriorate. Such methods are best suited for recognition of an object that constitutes a complete image [8].

In graph based representation, generally, regions with their corresponding feature vector and the geometric relationship between these regions are encoded in the form of a graph. Tu *et al*. [8] proposed a method which segments the image into regions of approximately constant color and the geometrical relationship of the segmented colored regions is represented by an attributed graph. Object matching, then, is formulated as an approximate graph-matching problem. The methods such as Color Adjacency Graph (CAG) [9], Attributed Relational Graph (ARG) [10], Shock graph [11] are prominent in this approach. Kostin *et al*. [12] proposed an object recognition scheme using graph matching. One advantage in graph based representation is that the geometric relationship can be used to encode certain shape information of the object and any sub-graph matching algorithm can be used to identify a single as well as multiple objects in query images. However, matching two such representations becomes a complicated process. Some of the issues in this regard are discussed in [12].

Support Vector Machine(SVM) based methods are used to classify both globally and locally obtained feature vectors of the objects [13], [14]. Roth *et al*. [15] proposed a view based algorithm for 3D object recognition using a network of linear units. Sparse Network of Winnows(SNoW) learning architecture is used to learn the representations of objects. Two experiments are carried out by them using pixel-based representation and edge-based representation of the objects separately.

In general above discussed methods use representation schemes, which are global in nature. The global representation schemes have certain short comings. These short comings can be overcome using local representations. In local representation schemes, generally, information from several regions of the images are encoded. Some of the local representation schemes are Local Affine Frames(LAF) [16], "Scale Invariant Feature Transform (SIFT)" [17], "Shape Context" [1], "Multi-modal Neighborhood Signature(MNS)". However, SIFT and Shape Context are designed for gray scale images. Maree *et al*. [18] have proposed a generic approach to image classification based on decision tree ensembles and local sub-windows and improvements upon this method is reported in [19].

Two methods are proposed in this article for object recognition. Section II describes the motivation of the work. Section III contains a representation scheme to represent an object image using the descriptions of different regions of interest. Two different schemes are proposed to extract the regions of interest. Section IV has two dissimilarity measures, one is to compare two regions of interest and the other is to compare two object images through their descriptors. Section V contains a brief description of the datasets used, and comparisons with the existing methods. The article is concluded with a discussion on the proposed methodology in Section VI.

II. Motivation for Proposed Method

Two important cues to distinguish between two objects are the overall shape and structure of the object, and the occurrence of different colors with respect to their spatial arrangements. Generally, human beings use both the cues for distinguishing objects in different stages. Although, it is difficult to imagine the actual shape of an object from the spatial arrangements of the different segments on its surface, it can be used as an important cue to represent the objects for classification. Several psychological studies regarding the representation of shape have been discussed in [20] and a survey of literature in this regard is provided in [21]. A way of preserving the positional information of adjacent segments is to store their representing color vectors as a unit. This connected set which covers pixels from all the adjacent segments and contains the color information from these segments is the region of interest. Let us call such a region as a "Multi-Colored Neighborhood (MCN)". Six examples of such MCNs are shown in Fig. 1.

Authors are with the Machine Intelligence Unit of Indian Statistical Institute, 203 B. T. Road, Kolkata - 700108, India. E-mail {sarif_r, murthy}@isical.ac.in

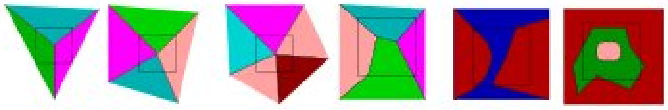


Fig. 1. Examples of three types of junctions where multiple region merges and three examples of presence of parts of different image segments in an image neighborhood. A rectangular window in each case shows the region of interest.

An object representation namely Multi-modal Neighborhood Signature(MNS), similar to the proposed one, was carried out by Matas *et al.* [22]. Neighborhoods having multi-modal color distribution in RGB color space are located in the object image using a simplified mean-shift algorithm. Let the number of modes found in a neighborhood be $n(n \geq 2)$, and the set of modes be U={$\bar{\mu}_1$, $\bar{\mu}_2$, . . . , $\bar{\mu}_n$}, where $\bar{\mu}_i$ is three dimensional RGB vector of the i^{th} mode. Then $\binom{n}{2}$ color pairs $\{(\bar{\mu}_i, \bar{\mu}_j), i, j = 1, 2, \ldots, n$ and $i \neq j\}$ are formed from U. In this way all possible pairs of vectors are found from each multi-modal neighborhood of the image. Set of all such distinct pair of vectors $(\bar{\mu}_i, \bar{\mu}_j)$ are defined as the MNS of the object. MNS contains the color information in the object. It also preserves the adjacency information of the colors on an object more or less. Each pair of vectors $(\bar{\mu}_i, \bar{\mu}_j)$ in the MNS contains the information that there is a region in the image where two segments of colors $\bar{\mu}_i$ and $\bar{\mu}_j$ are adjacent. However, this signature can not specify whether there are only two such adjacent segments or more than two adjacent segments in a region (Fig. 1). Thus, the neighborhoods having more than two-modal color distributions are not represented efficiently. Thus it lacks the crucial discrimination power regarding the neighborhoods having more than two segments. The discrimination power of the signature can be greatly enhanced if the properties of such type of neighborhoods could be preserved efficiently by the signature. This has been the motivation to propose one such representation which is capable of representing the neighborhoods having more than two modes. The proposed representation is described below.

III. Object Representation

An MCN is represented here as a unit consisting of the centers of the clusters. This unit of cluster centers contains the average color values corresponding to the different segments of MCN. Ultimately, the object is represented by the distinct sets of units of the cluster centers of the constituent MCNs. Let us call it as the "Multi-Colored Region Descriptor(M-CORD)" of the object.

A. Multi-Colored Region Descriptor

The color values found from the cluster centers of an MCN are stored as a unit for each MCN, thereby keeping track of the structural information, especially when there are more than two clusters. Suppose N distinct MCNs are selected by the proposed algorithm. Then the signature of the object contains N units of cluster centers, and each unit of cluster center represents a single MCN.

This descriptor contains the information regarding each MCN of the image and the MCNs are either from the boundaries or junctions present in the image. Thus, it contains the information that if there is a unit of k_i clusters present in the descriptor then there is a patch of pixels which covers parts of k_i segments present in the image. This greatly enhances the discrimination power of the recognition system when same colors are present in two objects but in different alignments.

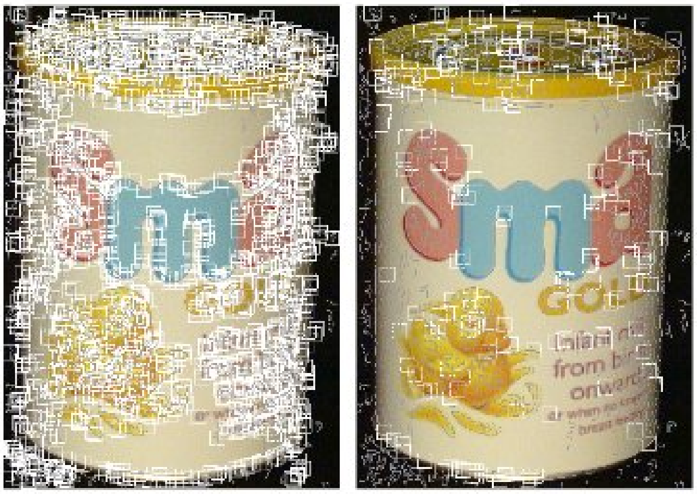
The color distribution of each MCN is multi-modal [22]. Thus a clustering technique can be employed to find the number of colors present in a region. Another way of detecting these regions is to see in how many parts it is divided into by the edge pixels present in the region. A special property of MCN is that it contains either a junction, or a part of boundary of the object, or simply an edge which divides the region into several parts. Each part of the region belongs to a different segment of the image. Thus edge maps of the images of the objects can also be used to locate such regions. These two approaches are explained separately below.

Algorithm 1 Cluster(V, r, min_clst_size)

```
$c = 0, i_{max} = 1$ and $V = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$
while $n > min\_clst\_size$ do
  for all $i = 1 : n$ do
    if $c = 0$ then
      Find $d_{ji} = d_{ij} = \|\bar{v}_i - \bar{v}_j\| \; \forall j > i$
      $V_i = \{\bar{v}_i\} \cup \{\bar{v}_j \in V : d_{ij} < r \; \forall j \neq i\}$
      $U_i = V \setminus V_i$
      if $|U_i| < min\_clst\_size$ then
        return $\{c + 1\}$ {$V$ is from a neighborhood of uniform color}
      end if
    else
      $V_i = \{\bar{v}_i\} \cup \{\bar{v}_j \subset V : d_{ij} < r \; \forall j \neq i\}$
      $U_i = V \setminus V_i$
    end if
    if $|V_i| > |V_{i_{max}}|$ then
      $i_{max} = i$
    end if
  end for
  if $|V_{i_{max}}| > min\_clst\_size$ then
    $c \leftarrow c + 1$
    $\bar{\mu}_c = \frac{1}{|V_{i_{max}}|} \sum_{\bar{v}_j \in V_{i_{max}}} \bar{v}_j$
    $V = U_{i_{max}}, n = |U_{i_{max}}|$
  else
    return $\{c, \bar{\mu}_1, \bar{\mu}_2, \ldots, \bar{\mu}_c\}$
  end if
end while
```

B. Detection of MCNs using Clustering

To obtain the different colors present in a neighborhood we propose a simple and fast clustering algorithm to find the cluster centers. It takes three parameters V, r and min_clst_size. Let $V = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$ be a set of color vectors. We call two color vectors $\bar{v}_i$ and $\bar{v}_j$ as similar if $\|\bar{v}_i - \bar{v}_j\| < r$. Thus r is the dissimilarity parameter for two colors. min_clst_size is the parameter to check the validity of a cluster. The steps of the algorithm are shown in Algorithm 1. To detect the M-CORD, $Cluster()$ is performed at every considered neighborhood in the image. For simplicity of pixel manipulation, overlapping windows of size $w \times w$ are selected as neighborhoods. All the detected MCNs of an object image from SOIL-47A dataset are shown in Fig. 2(a). It can be seen that most of the edges are covered by number of MCNs. To construct the descriptor of the object, all the MCNs are not needed because several MCNs detected over a stretch of boundary will have similar color distributions Thus, after finding an MCN, it is matched with all the previously considered MCNs and is included in the descriptor, if it is significantly different from all the previously considered MCNs. The amount of difference between two MCNs is determined by the dissimilarity value found using the Hausdorff distance (1) between the two sets of colors corresponding to two different MCNs that is



(a) Only 10% MCNs are shown among all the MCNs detected in the image to avoid cluttering

(b) Selected MCNs

Fig. 2. MCNs detected in obj39A of SOIL-47A dataset using edge map of the image.

explained in detail later in Section IV-A of the article. In this way all the distinct MCNs are extracted from the object image to construct the M-CORD of the object. Let us call this representation as M-CORD-Cluster. The matching algorithm for two different MCNs is described in Section IV-A. Finally, the M-CORD of each of the objects is stored in a separate file.

C. Detection of MCNs using Edge Map

Edge maps give crucial shape information of an object because connected edges are detected between every pair of adjacent segments. Thus every stretch of edge pixels gives information about two neighboring regions and any junction of more than two edges (as shown in Fig. 1) indicates the presence of multiple regions neighboring the surrounding point. The main problem in this approach is to use the "right" edge map for all the images. Most of the edge detectors fail to detect the correct edges with a fixed set of parameter values for all the images. Thus, it necessitates manual tuning of the parameters to obtain satisfactory results. This becomes an extremely difficult task while dealing with thousands of images.

Recently, the authors have suggested an efficient edge detection technique, which sets the same parameter values for color images [23]. This method is used here to find the edge maps of the object images with the default set of parameter values which the algorithm uses. Regions of size $w \times w$ are considered around the edge pixels. If the edges in a region divide it into disjoint smaller regions then the considered region is covering pixels from multiple image segments. In general, such regions are found over the boundaries where multiple image segments are present. If the number of connected components in the region is at least two, it is declared as an MCN. The average color values of each of the smaller regions in the MCN are found. Finally, all such distinct MCNs are clubbed together as described in Section III-A to construct the M-CORD of the object image and let us call it as M-CORD-Edge. Fig. 2 shows the MCNs detected using M-CORD-Edge in an object image. Each of the white rectangular windows is an MCN. Fig.2(a) shows only 10% of the MCNs among all the detected MCNs. Rest of the 90% MCNs are not shown to avoid cluttering. All the distinct MCNs considered for the formation of the M-CORD of the object are shown in Fig. 2(b).

Although, the idea of the representation using edge map of the object is same as the representation using clustering, the descriptors due to them are different because of the principles involved in them. For instance, in the third MCN from left in Fig. 1, M-CORD-Cluster will use 4 mean values whereas M-CORD-Edge will use 5 mean values to describe the neighborhood (the colors of the two smaller regions inside the MCN are same), because the edge map in this neighborhood divides it into 5 smaller regions due to five different segments of the image. In general, if there are n different colors present in the neighborhood then M-CORD-Cluster uses n means to represent the MCN irrespective of the spatial arrangements of the colors. But, in the case of M-CORD-Edge, if there are n types of color pixels and are divided into m(m greater than n) smaller regions then the descriptor uses m mean values to represent the MCN. Thus, M-CORD-Edge representation is richer than M-CORD-Cluster.

IV. Matching

Two types of matching operations are performed here. In one type, the matching is done at the time of finding the distinct MCNs to construct M-CORD. In the second case, matching is performed while comparing two objects through their M-CORDs. These procedures are described in the following two sections.

A. Matching two MCNs of an object image

All the MCNs in an object image are detected either using clustering or the edge map of the object image as described in previous section. It can be observed from Fig. 2(a) that several MCNs are detected over a stretch of boundary of the object and most of them have similar color distribution. To represent the object, information from all of these MCNs generally is not needed. Only a few MCNs from a stretch of boundary having significantly different color distributions are enough for this purpose. Two MCNs are said to be significantly different if the dissimilarity between them is greater than a value δ_{max}. The dissimilarity between two MCNs, δ, is defined as follows.

Let $U = \{\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_m\}$ and $V = \{\bar{v}_1, \bar{v}_2, \cdots, \bar{v}_n\}$ represent two different MCNs in an object image, where $\bar{u}_i = (u_i^1, u_i^2, u_i^3)$ and $\bar{v}_j = (v_j^1, v_j^2, v_j^3)$ are 3-dimensional color vectors from U and V respectively. Then the dissimilarity between U and V is defined as

$$\delta = \max\left(\max_{\bar{u}_i \in U}\{\min_{\bar{v}_j \in V}\{\|\bar{u}_i - \bar{v}_j\|\}\}, \max_{\bar{v}_j \in V}\{\min_{\bar{u}_i \in U}\{\|\bar{u}_i - \bar{v}_j\|\}\} \right), \tag{1}$$

where $\|\bar{u}_i - \bar{v}_j\| = \sqrt{(u_i^1 - v_j^1)^2 + (u_i^2 - v_j^2)^2 + (u_i^3 - v_j^3)^2}$.

Note that, in order that U and V are similar, each element in each set should have a similar element in the other set. If there is an element which does not have a similar element in the other set then these two sets are not similar. The expression (1) is the Hausdorff distance [24] between U and V.

B. Matching two Objects

Two objects are matched by comparing their M-CORDs. Under ideal conditions, if the images under consideration for comparison are from the same object and same view then the procedures described in Section III should produce identical MCNs. But, in practice, when the images are taken under different conditions (i.e, different lighting conditions or from different views) the MCNs are not identical even if the two images are of the same object. Thus, the matching is not exact as it is in the case of matching two MCNs from same object image. Here, the matching operation assigns a score i.e., a dissimilarity value, based on the dissimilarity between the MCNs from the M-CORDs of the images under consideration.

Let $P = \{U_1, U_2, \cdots, U_M\}$ *and* $Q = \{V_1, V_2, \cdots, V_N\}$ be two M-CORDs from two different images. Here, $U_i = \{\bar{u}_{i1}, \bar{u}_{i2}, \cdots, \bar{u}_{i\alpha_i}\}$ and $V_j = \{\bar{v}_{j1}, \bar{v}_{j2}, \cdots, \bar{v}_{j\beta_j}\}$ are MCNs of P and Q respectively and, $\bar{u}_{ik}$ and $\bar{v}_{jk}$ are 3-dimensional color vectors. It can be seen that the number of elements in P and Q may not be same. Wallraven *et al.* [14] have proposed a distance function to find the distance between two feature vectors of different sizes which is used in a SVM classifier for image classification. The entities of their feature vectors are vectors of same size. But, in our case the entities U_i and V_j of the feature vectors P and Q respectively are also sets and they can have different number of elements. Thus, two dissimilarity measures are proposed, one is between two MCNs (2) and the other is between two M-CORDs (3).

$$\begin{aligned}\delta_{\text{MCN}}(U_i, V_j) = & \ \tfrac{1}{\alpha_i}\sum_{k=1}^{\alpha_i} \min_{\bar{v}_{jl} \in V_j} \{ |\bar{u}_{ik} - \bar{v}_{jl}| \} \\ & + \tfrac{1}{\beta_j}\sum_{l=1}^{\beta_j} \min_{\bar{u}_{ik} \in U_i} \{ |\bar{u}_{ik} - \bar{v}_{jl}| \} \qquad (2)\end{aligned}$$

$$\begin{aligned}\delta_{\text{M-CORD}}(P, Q) = & \ \tfrac{1}{M}\sum_{i=1}^{M} \min_{V_j \in Q} \{ \delta_{MCN}(U_i, V_j) \} \\ & + \tfrac{1}{N}\sum_{j=1}^{N} \min_{U_i \in P} \{ \delta_{MCN}(U_i, V_j) \} \qquad (3)\end{aligned}$$

V. Results and Comparisons

The performance of the proposed methods has been evaluated on two well-known datasets namely, Columbia University Object Image Library (COIL-100)[1] [25] and Surrey Object Image Library (SOIL-47)[2] [26], corresponding to object recognition problem. In addition to these two datasets we have reported the experiments on a subset of Amsterdam Library of Object Images (ALOI) [27]. The image collection relevant for our purpose is the collection of 72,000 images of the 1000 objects under in-plane rotation aim to describe object view (ALOI-VIEW) [3] with a quarter resolution. We have selected only 18000 images from 250 objects from among the 1000 objects for evaluating the performance of our methods. We believe a collection of 18000 images from 250 different objects provides sufficient variety to test the performance of the proposed methods.

The performance of M-CORD with the two different approaches using edge map of the objects (M-CORD-Edge) and using clustering (M-CORD-Cluster) to find the descriptors has been compared with MNS method proposed by Matas *et al.* [22] using SOIL-47A dataset and some of the recently published methods on COIL-100 dataset. To compare the performance of the methods on SOIL-47A dataset, the results reported in [16], [28] as well as the results obtained using our implementation of MNS method have been considered for comparison. Similarly, results reported in [15], [16], [18] have been considered for the comparison of the performance of the methods on COIL-100 dataset. Additionally, recognition performances on ALOI using proposed methods have been included for better judgment.

TABLE I
RECOGNITION PERFORMANCE ON SOIL-47A

Average recognition with rank 1					
	M-CORD		MNS		
View Ang.	Edge	Cluster	Reported from [28]	Our Implementation	
				L_1 metric	L_2 metric
Time†	49	16		54	58
⊥ 90 deg.	87.98	87.02	52.5	54.89	57.45
⊥ 60 deg.	95.74	95.03	52.5	56.21	59.93
⊥ 45 deg.	95.95	95.10	–	57.02	61.06
⊥ 20 deg.	95.74	93.08	55.3	57.98	60.63
Average recognition with rank 1-3					
+90 deg.	96.17	95.21	78.0	82.87	96.88
+60 deg.	100.00	100.00	78.7	84.22	97.62
+45 deg.	100.00	100.00	–	83.62	96.67
+20 deg.	100.00	100.00	78.7	82.98	85.63

† Average time is provided per match in terms of milliseconds(ms)
– indicates that results for the corresponding boxes are not available

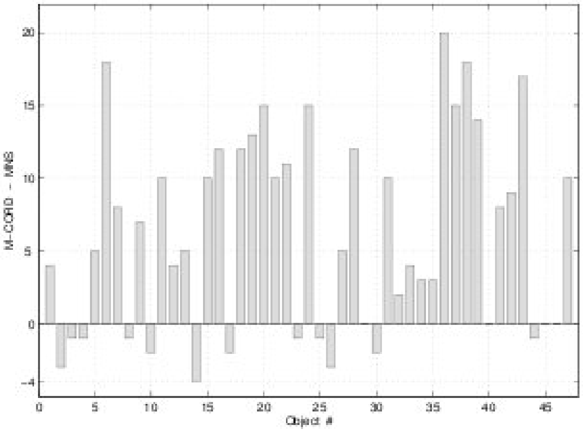


Fig. 3. Improvement in object wise correct matches by M-CORD over MNS

Let us consider the SOIL-47A dataset for evaluation of performance of the methods. The experiments have been done in the same way as described by Koubaroulis *et al.* [28]. To have a better understanding, we implemented the MNS method and parameters are tuned to obtain better results. The parameters used for MNS method are size of the window (w), Mean Shift kernel width (K), minimum mode size (C), threshold for the distance used at the time of suppression (d), and the threshold value used at the time of MNS matching (T). Irrespective of the parameter values used by Matas *et al.* [22] we have tuned the parameters independently for L_1 and L_2 metrics. The chosen parameter values, after several experiments, are $w = 16$, $K = 17.3$, $C = 20$, $d = 50$, and $T = 57$ for L_2 and 80 for L_1. The obtained results are reported in fourth and fifth columns of Table I. It can be seen that both for L_1 and L_2 metrics, we obtained better results than the ones reported in [28]. Additionally, the proposed methods outperform MNS method in terms of recognition rates.

To better understand the advantage of M-CORD over MNS we looked for the objects for which the methods have performed poorly. As the results of M-CORD-Edge and M-CORD-Cluster for SOIL-47A are similar, we have considered M-CORD-Edge for comparison. To observe the object-wise performance of the proposed method, number of correct matches by MNS method for each object is subtracted from the no. of correct matches by the proposed method and the difference table is plotted in Fig. 3. It can be seen in Fig. 3 that M-CORD-Edge performed better than MNS for most of the

[1] http://www.cs.columbia.edu/CAVE/

[2] http://www.ee.surrey.ac.uk/EE/VSSP/demos/colour/soil47/

[3] http://www.science.uva.nl/~mark/aloi/aloi_red4_view.tar

objects. It is observed that none of the 20 views of three objects (# 20, 36 and 21) in SOIL-47A dataset are matched correctly by MNS method, whereas at least some views of every object are matched correctly by the proposed method. Only for object # 34 proposed method has performed poorly for which the correct no. of matches is just 4 still it is better than the performance of MNS. For all other objects M-CORD-Edge has obtained at least 10 correct matches. Proposed method mismatched object # 34 with object # 35 for sixteen test views. The two objects are very similar to each other and it is extremely difficult to distinguish one from other even for a human being. The above mentioned results are evidence of the superior performance of M-CORD over MNS.

COIL-100 is a widely used dataset for object recognition. Five different experiments have been conducted to evaluate the performance of the M-CORD methods and they are compared to other methods found in the literature for this dataset. The experiments are classified according to the number of training views considered for the experiments. The average values of rank 1 recognition are listed in Table II. The results for other methods are taken from the cited papers except for the method Extra Tree + Random Sub-windows proposed by Maree *et al*. [19]. These recognition rates are obtained using the software PiXiT [4] provided by Maree and PEPITe. Maree *et al*. in their paper reported results using HSV color space. However, the results reported in row 3 of Table II are generated using RGB coding because proposed method too uses RGB values. It can be seen that the recognition performance increases with the increase in the number of training views of the objects. However, it is not always possible to have different training views available to obtain the model descriptor and the increase in the number of training views also increases the computational cost. Thus, a method should be judged better when it produces better results with less number of training views. Additionally, decreasing the number of training views increases demands on the method's generalization ability, and on the insensitivity to image deformations [16]. It can be seen from Table II that the rank 1 recognition rate obtained using proposed methods is better than other methods when one, two or four training views are considered. If eight training views are considered then the best result (99.40%) is reported for LAF [16] compared to 99.00% and 98.92% by M-CORD-Edge and M-CORD-Cluster respectively. In the case of 18 training views per object proposed method, M-CORD-Edge, achieved recognition rate (99.91%) which is equivalent to the best result reported in the literature for LAF. But M-CORD-Edge produces the best result with perfect (100%) recognition on this dataset. Overall, M-CORD-Edge produces uniformly better results compared to other methods except for the case of eight training views per object and M-CORD-Cluster produces significantly better results compared to other methods when one, two or four training views are considered.

It is to be noted that, in the proposed approach no object modeling is done from the available training views to obtain a single M-CORD. The different training views are selected as in [16] and [18], and are mentioned in Table II.

TABLE II
COIL-100: RANK 1 RECOGNITION PERFORMANCE

No. of Tr. Views/Obj.	36	18	8	4	2	1
No. of Tr. images†	3600	1800	800	400	200	100
Tr. Views in °	0+k10	0+k20	0+k45	45+k90	0,90	0
M-CORD-Edge	100	99.91	99.00	96.50	93.36	86.56
M-CORD-Cluster	99.92	99.87	98.64	96.46	92.74	86.93
Extra-Trees+ Random Sub-Windows RGB [19]	99.86	99.50	97.67	92.43	88.36	79.58
LAFs [16]	–	99.90	99.40	94.70	87.80	76.00
Sub-windows [18]	99.94	99.61	98.47	95.06	88.00	75.17
Extra Trees [18]	99.67	97.96	92.45	87.64	75.09	63.90
SNoW/Edge [15]	–	94.13	89.23	88.28	–	–
SNoW/intensity [15]	–	92.31	85.13	81.46	–	–
Linear SVM [15]	–	91.30	84.80	78.50	–	–
NN [15]	–	87.50	79.50	74.60	–	–

† No. of Test images in each case is 7200 - # of Training images
– indicates that results for the corresponding boxes are not available

TABLE III
ALOI: RECOGNITION PERFORMANCE

No. of Tr. Views/Obj.	8	4	2	1
Total No. Tr. Images†	2000	1000	500	250
View Angles in °	0 + k45	45+k90	0, 90	0
Using M-CORD-Cluster				
Rank 1	98.89	95.28	86.67	75.15
Rank 2	99.67	97.55	90.46	81.10
Rank 3	99.84	98.12	92.27	83.48
Using M-CORD-Edge				
Rank 1	98.68	93.94	82.63	69.77
Rank 2	99.39	96.34	87.30	76.28
Rank 3	99.62	97.08	89.56	79.68

† No. of Test images in each case is 18000 - No. of Training images

The last dataset considered is the ALOI-VIEW dataset. we have conducted the experiments on 250 objects only (i.e, 25% of the total no of images). The recognition performance is summarized in Table III. It can be seen that proposed method M-CORD-Edge and M-CORD-Cluster obtained moderately good recognition of 69.77% and 75.15% respectively, using one training view per object. Performance is not as good as in the case of COIL-100 dataset because the increase in the number of object classes increases the level of confusion between the objects and decreases the recognition performance. But, both the methods achieved good recognition when more number of views per objects are considered in the training set. No results of other methods are available on this dataset for comparison.

VI. CONCLUSIONS, DISCUSSION AND FUTURE WORKS

Performance of the proposed methodology depends mainly on the number of MCNs selected for each of the M-CORD and the size of the region selected. Here, region is a rectangular window of size $w \times w$. The number of MCNs detected for each of the M-CORD depends on the size of the window and the dissimilarity threshold δ_{max}. While the size of the window is crucial for obtaining better regional description, δ_{max} controls the number of MCNs selected. The more is the value of δ_{max}, less is the number of MCNs selected. The bigger is the size of the window, better is the representation. Larger windows increase the computational cost of the method. Similarly, if too many MCNs are selected, the methods suffer from the problem of over fitting and it also increases the recognition time.

The values of all the parameters are selected on the basis of several experiments. Windows of three different sizes ($w = 10, 16$ and 25) are selected for experiments in COIL-100 dataset. Although, the results obtained are not significantly different, best results are obtained using $w = 16$ and is reported in Table II. The other parameter values such as δ_{max}-the dissimilarity parameter between two MCNs, min_clst_size-the minimum number of pixels needed for a cluster to be valid and r-the parameter to check the dissimilarity between two color vectors are selected by varying them over different intervals. The final values are selected by observing the MCNs selected on a number of images. The values of r and min_clst_size are varied between 10 and 60 and best results are obtained using $r = 30$ and $min\_clst\_size = 20$. Similarly, the value of δ_{max} is varied between 20 and 80 and the best results are obtained using $\delta_{max} = 40$.

[4] http://www.montefiore.ulg.ac.be/~maree/pixit.html

The same parameter values are used for SOIL-47A dataset except for the window size, to generate the results shown in Table I. Initially, performance of the proposed methods is tested on SOIL-47A using windows of size $w = 10$ for both training and test views. In the SOIL-47A dataset, the size of the frontal(training) view of the object is twice the size of the other(test) views. Thus $w = 20$ is a reasonable size to be selected for the frontal views when the window size for the test views is taken to be 10. This produces better results compared to the case of $w = 10$ for all the views.

Comparison between the two proposed ways of representations : It is to be noted that M-CORD-Edge representation is richer than M-CORD-Cluster because it can also be sensible to the spatial distribution of the colors. Hence, M-CORD-Edge produces better results than M-CORD-Cluster in the case of SOIL-47A and COIL-100. But, rich representation in M-CORD-Edge is obtained at the cost of extra storage space to store the M-CORDs and more comparison time between two M-CORDs. In case of ALOI-VIEW dataset, although, the M-CORD-Edge descriptors of the objects are rich, the performance in terms of recognition rate is not good compared to M-CORD-Cluster. The possible reason may be the problem of over representation. Sometimes, the representation richer than needed is not helpful. Comparatively poor but reasonably good recognition is obtained using M-CORD-Cluster with smaller window sizes (10×10) for all the datasets. Thus to choose between two methods one can opt for M-CORD-Cluster anticipating a reasonably good recognition with a small window size such as $w = 10$. But, to get rich representation of the objects and better results, M-CORD-Edge with bigger window size should be considered.

The main contribution of this article are the proposed representation (M-CORD) of an object. Two dissimilarity measures, one is to compare between two M-CORDs and the other one is to compare between two MCNs, have also been proposed. The strength of the proposed methodology is the efficient representation of the colors appearing on the object surface which preserves the local shape information. Proposed methods would perform well when objects are multi-colored and, rich and colorful patterns appear on the object surface.

Acknowledgments

The authors thank the creators of SOIL-47, COIL & ALOI datasets for giving us the permission to use their datasets in the research work.

References

[1] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape context," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[2] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. of Computer Vision*, vol. 7, no. 1, pp. 11–32, Nov. 1991.

[3] M. A. Stricker, "Color and geometry as cues for indexing,," Department of Computer Science, University of Chicago, Tech. Rep. TR-92-22, 1992. [Online]. Available: http://citeseer.lcs.mit.edu/stricker92color.html

[4] B. Schiele and J. L. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *Int. J. of Computer Vision*, vol. 36, no. 1, pp. 31 – 50, Jan. 2000.

[5] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D object from appearance," *Int. J. of Computer Vision*, vol. 14, no. 1, pp. 5–24, Jan. 1995.

[6] S. K. Nayar and R. M. Bolle, "Reflectance based object recognition," *Int. J. of Computer Vision*, vol. 17, no. 3, pp. 219–240, 1996.

[7] M. A. Turk and A. P. Pentaland, "Face recognition using eigenfaces," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[8] P. Tu, T. Saxena, and R. Hartley, "Recognizing objects using color-annotated adjacency graphs," in *Shape, Contour and Grouping in Computer Vision*, ser. Lecture Notes in Computer Science, D. A. Forsyth, J. L. Mundy, V. D. Gesù, and R. Cipolla, Eds. Springer, 1999, pp. 246–263, ISBN 3-540-66722-9. [Online]. Available: http://users.rsise.anu.edu.au/~hartley/Papers/sicily/sicily.pdf

[9] J. Matas, R. Marik, and J. Kittler, "On representation and matching of multi-coloured objects," in *Proc. of Int. Conf. on Computer Vision*, Cambridge, MA, USA, June 1995, pp. 726–732.

[10] A. R. Ahmadyfard and J. Kittler, "Using relaxasation technique for region-based object recognition," *Image and Vision Computing*, vol. 20, no. 11, pp. 769–781, Sept. 2002.

[11] K. Siddiqi and B. B. Kimia, "A shock grammar for recognition," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, 1996, pp. 507–513.

[12] A. Kostin, J. Kittler, and W. Christmas, "Object recognition by symmetrised graph matching using relaxation labelling with an inhibitory mechanism," *Pattern Recognition Letters*, vol. 26, no. 3, pp. 381–393, Feb. 2005.

[13] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 6, pp. 637–646, June 1998.

[14] C. Wallraven, B. Caputo, and A. Graf, "Recognition with local features: the kernel recipe," in *Proc. of Int. Conf. on Computer Vision*, vol. 1, Nice, France, October 13 - 16 2003, pp. 257– 264.

[15] D. Roth, M.-H. Yang, and N. Ahuja, "Learning to recognize three-dimensional objects," *Neural Comp.*, vol. 14, no. 5, pp. 1071–1103, May 2002.

[16] S. Obdrzalek and J. Matas, "Object recognition using local affine frames on distinguished regions," in *Proc. of the British Machine Vision Conference*, Cardiff, 2002. [Online]. Available: http://www.bmva.ac.uk/bmvc/2002/papers/134/full_134.pdf

[17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[18] R. Marée, P. Geurts, J. Piater, and L. Wehenkel, "A generic approach for image classification based on decision tree ensembles and local sub-windows," in *Proc. of the Asian Conf. on Computer Vision*, vol. 2, 2004, pp. 860–865.

[19] ——, "Random subwindows for robust image classification," in *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, vol. 1, June 2005, pp. 34–40.

[20] D. Marr, *Vision*. W. H. Freeman and Company, 1982.

[21] R. Ramanath, "A framework for object characterization and matching in multi-and hyperspectral imaging systems," Ph.D. dissertation, Dept. of Elec. and Comp. Engg., North Carolina State University, August 13 2003. [Online]. Available: http://www.lib.ncsu.edu/theses/available/etd-08132003-223814/unrestricted/etd.pdf

[22] J. Matas, D. Koubaroulis, and J. Kittler, "The multimodal neighborhood signature for modeling object color appearance and aplications in object recognition and image retrieval," *Computer Vision and Image Understanding*, vol. 88, no. 1-3, pp. 1–23, Oct. 2002.

[23] S. K. Naik and C. A. Murthy, "Standardization of edge magnitude in color images," *IEEE Trans. Image Processing*, 2005, (Accepted for publication).

[24] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 9, pp. 850–863, Sept. 1993. [Online]. Available: citeseer.ist.psu.edu/huttenlocher93comparing.html

[25] S. A. Nene, S. K. Nayar, and H. Murase, "Colombia object image library (COIL-100)," Technical Report, CUCS-006-96, Dept. Of Computer Science, Colombia Uiversity, Tech. Rep., Feb. 1996. [Online]. Available: http://www1.cs.columbia.edu/CAVE/publications/pdfs/Nene_TR96_2.pdf

[26] J. Burianek, A. Ahmadyfard, and J. Kittler, "SOIL-47: The Surrey Object Image Library," Centre for Vision, Speech and Signal processing, Univerisity of Surrey. [Online]. Available: http://www.ee.surrey.ac.uk/Research/VSSP/demos/colour/soil47/

[27] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam Library of Object Images," *Int. J. of Computer Vision*, vol. 61, no. 1, pp. 103–112, Jan. 2005. [Online]. Available: http://www.science.uva.nl/~mark/pub/2005/geusebroekIJCV05a.pdf

[28] D. Koubaroulis, J. Matas, and J. Kittler, "Evaluating colour-based object recognition algorithms using the SOIL-47 database," in *Proc. of the Asian Conf. on Computer Vision*, Jan. 2002.