Distinct multicolored region descriptors for object recognition

*Pattern Analysis and Machine Intelligence*

vol. 29, no. 7, 2007

Sarif Kumar Naik, and C.A. Murthy

Presented by Dong Chang Lee

*School of Electrical Engineering and Computer Science*

*Kyungpook National Univ.*
Abstract

◆ Proposed method
  – Color descriptions from distinct regions covering multiple segments in the problem of object recognition
    • Detection of distinct multicolored regions using edge maps and clustering
Object recognition

– The efficient representation and comparison of two objects through their representation
– Two types of approaches to object representation
  • Utilizing the knowledge gained from the spatial arrangements of the “shape features”
    – Edge elements, boundaries, corners and junctions
  • Using the brightness or color features obtained more directly from the object images
◆ Proposed method
  – A scheme to describe on the object surface
    • Containing the color information and the patterns of colors
      on the object surface

◆ Global representation schemes
  – Histogram-based
    • Swain and Ballard
      – Using color as a primal cue for object recognition and image
        retrieval
    • Stricker
      – An indexing technique based on the boundary histogram of
        multicolored objects
- Advantages of histogram-based approach for object recognition
  - Simplicity, speed, and robustness

- Drawbacks
  - Inability to encode shape and structural information of the objects
  - Using of only color information for distinguishing the objects

  - Eigenspace-based
    - Representing an object by considering the whole image as a vector
    - Projecting it over a set of eigenvectors to achieve data compression and reduction of redundant information
    - Using principal component analysis
• Murase and Nayar, Turk and Pentland
• Effective methods
  – When eigenspace captures the characteristics of the whole database

  – Graph-based representation
• Regions with their corresponding feature vector and the geometric relationship
  – Encoding the form of a graph
• Tu et al.
  – Segmenting the image into regions of approximately constant color
  – Representing the geometric relationship of the segmented colored regions by an attributed graph
  – Formulating as an approximate graph-matching problem
• Drawbacks
  – Complicated process to match two such representation

  – Support vector machine-based
    • Classifying both globally and locally obtained feature vectors of object

◆ Local representation schemes
  – Local Affine Frames
  – Scale Invariant Feature Transform
    • Gray-scale images
  – Shape Context
    • Gray-scale images
  – Multimodal Neighborhood Signature
Motivation for the proposed method

◆ Two important cues to distinguish between two objects
  – The overall shape and structure of the object
  – The occurrence of different colors

◆ Color information
  – Difficult to visualize or reconstruct the actual shape of an object
  – Important cue to represent the objects for classification
◆ A way of preserving the positional information of adjacent segments
  – Storing their representing color vectors as a unit
  – Connected set is the region of interest (ROI)
    • Multicolored neighborhood (MCN)

Fig. 1. Examples of three types of junctions where multiple regions merge and three examples of the presence of parts of different image segments in an image neighborhood. A rectangular window, in each case, shows the ROI.
Similar to the proposed method

- Multimodal neighborhood signature (MNS)
  - Proposed Mata et al.
  - Neighborhoods having multimodal color distribution in RGB color space
    - Locating in the object image using a simplified mean-shift algorithm
    - The number of modes found in a neighborhood → \( n(n \geq 2) \)
    - The set of modes → \( U = \{\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_n\} \)
      where \( \tilde{\mu}_i \) three-dimensional RGB vector of the \( i \)th mode
    - \( \binom{n}{2} \) color pairs → \( \{(\tilde{\mu}_i, \tilde{\mu}_j), i, j = 1, 2, \ldots, n \text{ and } i \neq j\} \)
• Finding all possible pairs of vectors from each multimodal neighborhood of the image

• The set of all such distinct pair of vectors \((\mu_i, \mu_j)\)
  – Defining as the MNS of the object
  – Containing the color information in the object

  – Drawback
  • The neighborhoods having more than two-modal color distributions
    – Not represented efficiently
    – Lacking the crucial discrimination power regarding the neighborhoods

◆ Motivation
  – Neighborhoods having more than two modes
    • Representing efficiently
Object representation

◆ An MCN
  – A unit consisting of the representatives of the color present in it

◆ Representative color
  – The center of the cluster formed by the color vector present in the MCN
    • Average color values corresponding to the different segments of MCN

◆ Multicolored region descriptor (M-CORD)
  – Representing object
    • By the distinct sets of units of the cluster centers of the constituent MCNs
- Color values found from the cluster centers of an MCN
  - Storing as a unit for each MCN
- Suppose N distinct MCNs are selected by the proposed algorithm
  - The signature of the object
    - Containing N units of cluster centers and each unit of cluster center represents a single MCN
- Color distribution of each MCN $\rightarrow$ multimodal
  - Employing a clustering technique to find the number of colors present in a region
  - Dividing by the edge pixels present in the region
– A special property of MCN
  • Containing either a junction of a part of boundary of the object or simply an edge which divides the region into several parts
Detection of MCNs using clustering
  – Proposed method to obtain the different colors present in a neighborhood
    • A simple and fast clustering algorithm to find the cluster centers
    • Three parameters
      – A set of color vectors \( V = \{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_n\} \)
      – Dissimilarity parameter for two colors \( r \)
      – Parameter to check the validity of a cluster \( \text{min}_{-}\text{clst\_size} \)
• Algorithm 1

Algorithm1Cluster(V, r, min_clst_size)

1: c = 0, i_max = 1 /*V = {v_1, v_2, ..., v_n}*/
2: while n > min_clst_size do
3:     for all i = 1:n do
4:         if c = 0 then
5:             Find \( d_{ji} = d_{ij} = \| \vec{v}_i - \vec{v}_j \| \forall j > i \)
6:         \( V_i = \{ \vec{v}_i \} \cup \{ \vec{v}_j \in V : d_{ij} < r \forall j \neq i \} \)
7:         \( U_i = V / V_i \)
8:         if \( |U_i| < \min_clst_size \) then
9:             return \{c + 1\} /*V is from a neighborhood of uniform color*/
10:         end if
11:     else
12:         \( V_i = \{ \vec{v}_i \} \cup \{ \vec{v}_j \in V : d_{ij} < r \forall j \neq i \} \)
13:         \( U_i = V / V_i \)
The number of elements in $V_i$

14: \[\text{end if}\]
15: \[\text{if } |V_i| > |V_{i_{max}}| \text{ then}\]
16: \[i_{max} = i\]
17: \[\text{end if}\]
18: \[\text{end for}\]
19: \[\text{if } |V_{i_{max}}| > \text{min \_ clst \_ size} \text{ then}\]
20: \[c \leftarrow c + 1\]
21: \[
\hat{\mu}_c = \frac{1}{|V_{i_{max}}|} \sum_{\tilde{v}_i \in V_{i_{max}}} \tilde{v}_i
\]
22: \[V = U_{i_{max}}, n = |U_{i_{max}}|\]
23: \[\text{else}\]
24: \[\text{return } \{c, \hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_c\}\]
25: \[\text{else if}\]
26: \[\text{end while}\]
• Algorithm 1

A MCN
(w x w)

Clustering algorithm

Color values found from the cluster centers

Eliminating windows in regions

\{v_1, v_2\}
– Eliminating region
  • Not needed for the construction of the descriptor
    – All of the color vectors $\tilde{v}_i$ are within a small disc of radius $r$
– To detect the M-CORD
  • Performing clustering at every considered neighborhood
  • Selecting overlapping windows of size $w \times w$ as neighborhoods
Detection of MCNs using edge map

- Finding the edge maps of the object images with the default set of parameter values
  - Considering regions of size $w \times w$ around the edge pixels
- The edges in a region divide it into disjoint smaller regions
  - The number of connected components in the region is at least two $\rightarrow$ Declaring as an MCN
  - Finding average color values of each of the smaller regions in the MCN
- Similar to representation using clustering
Fig. 2. MCNs detected in obj39A of the SOIL-47A data set using the edge map of the image. (a) Only 10 percent of the MCNs are shown among all of the MCNs detected in the image to avoid cluttering. (b) Selected MCNs.
Matching

◆ Two types of matching operations
  – Finding the distinct MCNs to construct M-CORD
  – Comparing two objects through their M-CORDs

◆ Matching two MCNs of an object image
  – Detecting all of the MCNs in an object image
    • Using clustering or the edge map of the object image
  – To representation the object
    • All of the MCNs → Not need
– The dissimilarity between two MCNs $\rightarrow \delta$

- Let $U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, v_2, \ldots, v_n\}$

where $u_i = (u^1_i, u^2_i, u^3_i)$ and $v_j = (v^1_j, v^2_j, v^3_j)$ three-dimensional RGB vector

\[
\delta = \max \left( \max \left\{ \min \left\{ \| u_i - v_j \| \right\} \right\}, \max \left\{ \min \left\{ \| u_i - v_j \| \right\} \right\} \right)
\]

(1)

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

- Using Hausdorff distance
– Hausdorff distance
  • Let Set A, Set B
    – Definition of Hausdorff distance from set A to set B
      \[ h(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \} \]
    – Example
Matching two objects

- Matching by comparing their M-CORDs
  - Based on the dissimilarity between the MCNs from the M-CORDs of the images under consideration
- Let $P = \{U_1, U_2, \ldots, U_M\}$ and $Q = \{V_1, V_2, \ldots, V_N\}$
  - Two M-CORDs from two different images
  - Here, $U_i = \{\tilde{u}_{i1}, \tilde{u}_{i2}, \ldots, \tilde{u}_{ia_i}\}$ and $V_j = \{\tilde{v}_{j1}, \tilde{v}_{j2}, \ldots, \tilde{v}_{j\beta_j}\}$
    - MCNs of P and Q
    - $\tilde{u}_{ik}$ and $\tilde{v}_{jk}$ three-dimensional color vectors
- Two dissimilarity measures

\[
\delta_{MCN}(U_i, V_j) = \frac{1}{\alpha_i} \sum_{k=1}^{\alpha_i} \min_{v_{jl} \in V_j} \| u_{ik} - v_{jl} \| + \frac{1}{\beta_j} \sum_{l=1}^{\beta_j} \min_{u_{ik} \in U_i} \| u_{ik} - v_{jl} \| \quad (2)
\]

\[
\delta_{M-CORD}(P, Q) = \frac{1}{M} \sum_{i=1}^{M} \min_{V_j \in Q} \delta_{MCN}(U_i, V_j) + \frac{1}{N} \sum_{U_i \in P} \min_{V_j \in Q} \delta_{MCN}(U_i, V_j) \quad (3)
\]
Results and comparisons

Evaluation of proposed methods
- Columbia Object Image Library (COIL-100)
- Surrey Object Image Library (SOIL-47)
- Amsterdam Library of Object Images (ALOI)
Consideration the SOIL-47A data set – Comparison with MNS

Table 1. Recognition performance on SOIL-47A.

<table>
<thead>
<tr>
<th>View Ang.</th>
<th>M-CORD</th>
<th>MNS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Report from [28]</td>
<td>L₁ metric</td>
</tr>
<tr>
<td>Time¹</td>
<td>Edge</td>
<td>Cluster</td>
</tr>
<tr>
<td>± 90 deg.</td>
<td>49</td>
<td>16</td>
</tr>
<tr>
<td>± 60 deg.</td>
<td>88.0</td>
<td>87.0</td>
</tr>
<tr>
<td>± 45 deg.</td>
<td>95.7</td>
<td>95.0</td>
</tr>
<tr>
<td>± 20 deg.</td>
<td>96.0</td>
<td>95.1</td>
</tr>
</tbody>
</table>

Average recognition with rank 1-3

<table>
<thead>
<tr>
<th>±90 deg.</th>
<th>±60 deg.</th>
<th>±45 deg.</th>
<th>±20 deg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.2</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>95.21</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>78.0</td>
<td>78.7</td>
<td>–</td>
<td>78.7</td>
</tr>
<tr>
<td>82.9</td>
<td>84.2</td>
<td>–</td>
<td>83.0</td>
</tr>
<tr>
<td>96.9</td>
<td>97.6</td>
<td>96.7</td>
<td>85.6</td>
</tr>
</tbody>
</table>
• The number of correct matches

**Fig. 3.** Improvement in objectwise correct matches by M-CORD over MNS.
Consideration the COIL-100 data set
– Consideration of the number training views

Table 2. Rank 1 recognition performance.

<table>
<thead>
<tr>
<th>No. of Tr. Views/Obj.</th>
<th>36</th>
<th>18</th>
<th>8</th>
<th>4</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Tr. images†</td>
<td>3600</td>
<td>1800</td>
<td>800</td>
<td>400</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>Tr. Views in ‡</td>
<td>0+k10</td>
<td>0+k20</td>
<td>0+k45</td>
<td>45+k90</td>
<td>0.90</td>
<td>0</td>
</tr>
<tr>
<td>M-CORD-Edge</td>
<td>100</td>
<td>99.9</td>
<td>99.0</td>
<td>96.5</td>
<td>93.4</td>
<td>86.6</td>
</tr>
<tr>
<td>M-CORD-Cluster</td>
<td>99.9</td>
<td>99.9</td>
<td>98.6</td>
<td>96.5</td>
<td>92.7</td>
<td>86.9</td>
</tr>
<tr>
<td>Extra-Trees+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Sub-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Windows RGB [19]</td>
<td>99.9</td>
<td>99.5</td>
<td>97.7</td>
<td>92.4</td>
<td>88.4</td>
<td>79.6</td>
</tr>
<tr>
<td>LAFs [16]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-windows [18]</td>
<td>99.9</td>
<td>99.6</td>
<td>98.5</td>
<td>95.1</td>
<td>88.0</td>
<td>75.2</td>
</tr>
<tr>
<td>Extra Trees [18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNoW/Edge [15]</td>
<td>99.7</td>
<td>98.0</td>
<td>92.5</td>
<td>87.6</td>
<td>75.1</td>
<td>63.9</td>
</tr>
<tr>
<td>SNoW/intensity [15]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear SVM [15]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN [15]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Consideration the ALOI data set
– Using 25 percent the total number of images

Table 3. Recognition performance.

<table>
<thead>
<tr>
<th>No. of Tr. Views/Obj.</th>
<th>8</th>
<th>4</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total No. Tr. Images†</td>
<td>2000</td>
<td>1000</td>
<td>500</td>
<td>250</td>
</tr>
<tr>
<td>View Angles in °</td>
<td>0 + k45</td>
<td>45+k90</td>
<td>0, 90</td>
<td>0</td>
</tr>
<tr>
<td>Using M-CORD-Cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 1</td>
<td>98.9</td>
<td>95.3</td>
<td>86.7</td>
<td>75.2</td>
</tr>
<tr>
<td>Rank 2</td>
<td>99.7</td>
<td>97.6</td>
<td>90.5</td>
<td>81.1</td>
</tr>
<tr>
<td>Rank 3</td>
<td>99.8</td>
<td>98.1</td>
<td>92.3</td>
<td>83.5</td>
</tr>
<tr>
<td>Using M-CORD-Edge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank 1</td>
<td>98.7</td>
<td>93.9</td>
<td>82.6</td>
<td>69.7</td>
</tr>
<tr>
<td>Rank 2</td>
<td>99.4</td>
<td>96.3</td>
<td>87.3</td>
<td>76.3</td>
</tr>
<tr>
<td>Rank 3</td>
<td>99.6</td>
<td>97.1</td>
<td>89.6</td>
<td>79.7</td>
</tr>
</tbody>
</table>
Conclusions

◆ Proposed method
  – Representation of object
    • Multicolored Region Descriptor (M-CORD)
  – Two dissimilarity measures
    • Comparison of between two MCNs
    • Comparison of between two M-CORD