Sparse coding with shapelet dictionaries

Jeroen Chua, Brendan Frey
University of Toronto
Snowbird Learning Workshop 2012
Motivation

- In this work, we look at object recognition
Motivation

- In this work, we look at object recognition
- Shape important, colour less important
Motivation

- In this work, we look at object recognition

  - Shape important, colour less important
  - Shape less important, colour important
Motivation

- In this work, we look at object recognition

- Shape important, colour less important
- Shape less important, colour important
- Shape important, colour important
Patch-based object recognition

- Typical approach in vision for patch-based object recognition:

  - Our goal is to learn an image representation (image features) that captures local shape and colour information separately.
Patch-based object recognition

- Typical approach in vision for patch-based object recognition:

- Our goal is to learn an image representation (image features) that captures local shape and colour information separately.
Forming image features: sparse coding

- **Sparse coding** on image patches is a popular approach in vision, but often conflates shape and colour
Forming image features: sparse coding

- **Sparse coding** on image patches is a popular approach in vision, but often conflates shape and colour.
Forming image features: sparse coding

- **Sparse coding** on image patches is a popular approach in vision, but often conflates shape and colour
Forming image features: shapelet models

- Shapelet models are probabilistic generative models that factorize local structure and colour
  - Patch-based model
  - Define a dictionary (visual codewords) of probabilistic groupings of pixels that tend to co-occur in colour
Forming image features: shapelet models

- Shapelet models are probabilistic generative models that factorize local structure and colour
  - Patch-based model
  - Define a dictionary (visual codewords) of probabilistic groupings of pixels that tend to co-occur in colour
Forming image features: shapelet models

- Shapelet models are probabilistic generative models that factorize local structure and colour
  - Patch-based model
  - Define a dictionary (visual codewords) of probabilistic groupings of pixels that tend to co-occur in colour

Visualization of dictionary:

\[ \text{P(group = 1)} \times \text{\color{red}x} + \text{P(group = 2)} \times \text{\color{green}x} + \text{P(group = 3)} \times \text{\color{blue}x} \]
Generative shapelet model
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

Dictionary of shapelets

![Images of shapelets with RGB colors](image)

k=1  ...  k=K
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

Dictionary of shapelets

\[ k=1 \quad \ldots \quad \ldots \quad k=K \]
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

Dictionary of shapelets

$\cdots$ $\cdots$

$k=1$ $k=K$

Image
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

\[ p(k) \]

Sample a shapelet index

\[ k=1 \]

Dictionary of shapelets

\[ k=1 \]

\[ k=K \]
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

$p(k)$

Sample a shapelet index

$\mathbf{k}=1$

Sample a group index for each pixel

Dictionary of shapelets

$k=1$  $\cdots$  $k=K$

$k=1$

$1$

$2$

$3$

Image
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

\[ p(k) \]

Sample a shapelet index

Sample a group index for each pixel

Dictionary of shapelets

\[ k=1 \quad \ldots \quad \ldots \quad k=K \]

Patch-specific palette

Color 1
Color 2
Color 3

Image

\[ \text{Note: RGB colors denote different groups, not pixel colors.} \]
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

$p(k)$

Sample a shapelet index

$k=1$

Sample a group index for each pixel

Dictionary of shapelets

1

2

3

$k=1$

$k=K$

Patch-specific palette

Color 1

Color 2

Color 3

Paint patch

Image
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

$p(k)$

Sample a shapelet index

$k=3$

Dictionary of shapelets

$k=1$

$k=K$

Color 1

Color 2

Color 3

Patch-specific palette

Image
Generative shapelet model

Note: RGB colors denote different groups, not pixel colors.

Dictionary of shapelets

\[ p(k) \]

\[ k = 1 \quad \cdots \quad k = K \]
Shapelet dictionary

Dictionary elements (codewords) define groups of pixels that co-occur in colour without specifying what that colour should be.
Shapelet dictionary

Dictionary elements (codewords) define groups of pixels that co-occur in colour without specifying what that colour should be.
Shapelet dictionary

Dictionary elements (codewords) define groups of pixels that co-occur in colour without specifying what that colour should be.
Shapelet dictionary

Dictionary elements (codewords) define groups of pixels that co-occur in colour without specifying what that colour should be.
Related shapelet work

   - Patch-based shapelets, non-sparse image representation

This work: patch-based shapelet models with sparse image representations
Combining sparse coding and shapelets

- To encode an image patch given a dictionary \( \mathbf{D} \), sparse coding using the lasso solves:

\[
\arg \min_{\tilde{\alpha}_j} \| \tilde{x}_j - \mathbf{D} \tilde{\alpha}_j \|_2^2 + \lambda \| \tilde{\alpha}_j \|_1
\]

- \( \mathbf{D} \): visual dictionary
- \( \tilde{\alpha}_j \): sparse representation for patch \( j \)
- \( \lambda \): codeword penalty
Combining sparse coding and shapelets

- Idea: Before encoding an image patch, first allow the dictionary to be transformed to account for colouring. Find:

\[
\arg \min_{\tilde{x}_j, T_j \in \tau} \| \tilde{x}_j - T_j(D, \tilde{x}_j)\tilde{\alpha}_j \|_2^2 + \lambda |\tilde{\alpha}_j|_1
\]

- \(D\): shapelet dictionary
- \(\tilde{\alpha}_j\): sparse shape representation for patch \(j\)
- \(\lambda\): codeword penalty
- \(T_j(D, \tilde{x}_j)\): “coloured in” shapelet dictionary for patch \(j\)
Combining sparse coding and shapelets

- Idea: Before encoding an image patch, first allow the dictionary to be transformed to account for colouring. Find:

\[ \arg \min_{\tilde{\alpha}_j, T_j \in \tau} \| \tilde{x}_j - T_j(D, \tilde{x}_j)\tilde{\alpha}_j \|_2^2 + \sum_k \lambda_k |\alpha_{j,k}| \]

- \( D \): shapelet dictionary
- \( \tilde{\alpha}_j \): sparse shape representation for patch \( j \)
- \( \lambda_k \): penalty for shapelet \( k \)
- \( T_j(D, \tilde{x}_j) \): “coloured in” shapelet dictionary for patch \( j \)
Patch encoding

\[
\arg\min_{\tilde{\alpha}_j, T_j \in \tau} \| \tilde{x}_j - T_j(D, \tilde{x}_j)\tilde{\alpha}_j \|^2_2 + \sum_k \lambda_k |\alpha_{j,k}| 
\]

- Patch encoding is done by first estimating \( T_j \), then fixing \( T_j \) and finding \( \tilde{\alpha}_j \).
- Given \( T_j \), estimating \( \tilde{\alpha}_j \) is a standard sparse coding problem.
Inferring patch-specific appearance

- $T_j(D, \vec{x}_j)$ found by minimizing reconstruction error between each “coloured in” shapelet, and the image patch, $\vec{x}_j$. 
Inferring patch-specific appearance

- $T_j(D, \vec{x}_j)$ found by minimizing reconstruction error between each “coloured in” shapelet, and the image patch, $\vec{x}_j$. 

![Image of shapelets and image patches]
Inferring patch-specific appearance

- $T_j(\mathbf{D}, \mathbf{x}_j)$ found by minimizing reconstruction error between each “coloured in” shapelet, and the image patch, $\mathbf{x}_j$.
Inferring patch-specific appearance

- $T_j(D, \vec{x}_j)$ found by minimizing reconstruction error between each “coloured in” shapelet, and the image patch, $\vec{x}_j$. 

![Image of colored patches and a grayscale image](image-url)
Inferring patch-specific appearance

$T_j(D, \overrightarrow{x}_j)$ found by minimizing reconstruction error between each “coloured in” shapelet, and the image patch, $\overrightarrow{x}_j$. 
Inferring patch-specific appearance

$T_j(D, \vec{x}_j)$ found by minimizing reconstruction error between each “coloured in” shapelet, and the image patch, $\vec{x}_j$. 

$D$

![Image of shapelets]

$\vec{x}_j$

$T_j(D, \vec{x}_j)$

![Image of reconstructed patch]
Image representation

- After inferring coloured in dictionary, $T_j(D, \vec{x}_j)$, sparse coefficients found by solving:

$$\arg \min_{\vec{\alpha}_j} \left\| \vec{x}_j - T_j(D, \vec{x}_j)\vec{\alpha}_j \right\|_2^2 + \sum_k \lambda_k |\alpha_{j,k}|$$

- $\vec{\alpha}_j$ represents **local shape** information

- For **local colour** information, we compute a histogram of colours, $\vec{c}_j$, over the patch.
Classification

- Average pooling over three levels of spatial pyramid

- SVM classifier with weighted similarity of shape and color:

$$K(\vec{x}_1, \vec{x}_2) = w \cdot K_{\text{shape}}(\vec{x}_1^{\text{shape}}, \vec{x}_2^{\text{shape}}) + (1 - w) \cdot K_{\text{colour}}(\vec{x}_1^{\text{colour}}, \vec{x}_2^{\text{colour}})$$

- We use the intersection kernel, and $w = 0.5$

In short: infer shape and colour descriptors for images, compute similarity score, pass to SVM
Experiments

- Datasets:
  - Caltech101 [1]

- Dictionary learning:
  - Learn a dictionary of shapelets unsupervised using EM

- Feature extraction:
  - For each image patch, infer:
    - $\vec{a}_j$: local **structure**
    - $\vec{c}_j$: local **colour**

## Results: Caltech 101

<table>
<thead>
<tr>
<th>Method</th>
<th>Descriptor, no Colour</th>
<th>Descriptor + Colour Histogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapelet model [1]</td>
<td>56.7(0.2)</td>
<td>59.1(1.8)</td>
</tr>
<tr>
<td>Sparse coding</td>
<td>59.1(1.6)</td>
<td>54.1(1.7)</td>
</tr>
<tr>
<td>Shapelet model + Sparse coding</td>
<td>62.6(0.9)</td>
<td><strong>65.5(1.0)</strong></td>
</tr>
</tbody>
</table>

- Colour images resized to 100 x 100
- 8x8 patches, stride of 2 pixels
- 201 shapelet dictionary, 125-bin colour descriptor
- 30 training examples

Results: Caltech 101
Effect of # of codewords

Our methods
## Results: 15-scenes

<table>
<thead>
<tr>
<th>Method</th>
<th>Descriptor, no Colour</th>
<th>Descriptor + Colour Histogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapelet model [1]</td>
<td>62.2(1.3)</td>
<td>63.4(0.2)</td>
</tr>
<tr>
<td>Sparse coding</td>
<td>71.2(1.1)</td>
<td>68.78(0.83)</td>
</tr>
<tr>
<td>Shapelet model + Sparse coding</td>
<td>66.8(0.9)</td>
<td>70.2(0.5)</td>
</tr>
</tbody>
</table>

- **Grayscale** images resized to 100 x 100
- 8x8 patches, stride of 2 pixels
- 201 shapelet dictionary, 125-bin colour descriptor
- 100 training examples

Future Work

- Factorization of other appearance factors
  - Material type, texture
- For a particular object class, which is more important (and by how much): shape or colour? How should we measure similarity in shape and colour?
Conclusion

- Introduced shape-colour factorization for sparse coding on image patches, using shapelet models
- Encouraging results on Caltech101 (where colour information is available)
The End

Thanks! Questions?

Acknowledgements: Inmar Givoni (Toronto) and Prof. Ryan Adams (Harvard) for helpful discussions