Feature Learning by Multidimensional Scaling and its Applications in Object Recognition

2013 26th SIBGRAPI Conference on Graphics, Patterns and Images

Presented by:
Kim L. Boyer  kim@ecse.rpi.edu

Authors:
Quan Wang, Kim L. Boyer
Signal Analysis and Machine Perception Laboratory
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
Review on Image Features

• How do we represent an image by a fixed-length feature vector?
  ▫ Hand-designed features
Review on Image Features

- How do we represent an image by a fixed-length feature vector?
  - Hand-designed features

- Color histograms
- Wavelet coefficients
- SIFT
- Color-SIFT
- SURF
- HOG
- LBP
Review on Image Features

• How do we represent an image by a fixed-length feature vector?
  ▫ Hand-designed features

These features can be extracted without looking at the entire dataset
Review on Image Features

• How do we represent an image by a fixed-length feature vector?
  ▫ Learning from a large dataset
Review on Image Features

• How do we represent an image by a fixed-length feature vector?
  ▫ Learning from a large dataset
    • Raw pixels based
      - PCA, kernel PCA
      - auto-encoders
      - restricted Boltzmann machine (RBM)
Review on Image Features

• How do we represent an image by a fixed-length feature vector?
  ▫ Learning from a large dataset
    • Raw pixels based
      - PCA, kernel PCA
    • Hand-designed feature based
      - bag-of-visual-words (BOV)
      - Fisher vector
      - spatial pyramid matching (SPM)
      - restricted Boltzmann machine (RBM)
      - auto-encoders
      - PCA, kernel PCA
      - Fisher vector
      - spatial pyramid matching (SPM)
Review on Image Features

• How do we represent an image by a fixed-length feature vector?
  ▫ Learning from a large dataset
    • Raw pixels based
    • Hand-designed feature based
      - PCA, kernel PCA
      - Autoencoders
      - Restricted Boltzmann machine (RBM)
      - Fisher vector
      - Bag-of-visual-words (BOV)
      - Spatial pyramid matching (SPM)

Learn statistics/patterns from the entire dataset

\[ G^\Lambda = \frac{1}{T} \sum_{t=1}^{T} \varphi FV(x_t) \]
Other Possibilities?

- We already have:
  - Hand-designed features
  - Feature learning from raw pixels
  - Feature learning from hand-designed features
Other Possibilities?

• We already have:
  ▫ Hand-designed features
  ▫ Feature learning from raw pixels
  ▫ Feature learning from hand-designed features

• Now we propose a new group of features:
  ▫ Feature learning from semantics-sensitive image distances
Basic Idea

1. Given a set of images, measure pair-wise semantics-sensitive image distances
Basic Idea

1. Given a set of images, measure pair-wise semantics-sensitive image distances

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.5</th>
<th>1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>1.3</td>
<td>1.2</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Basic Idea

2. Apply Multidimensional Scaling (MDS) on these distances, and encode all images in a low dimensional \((m-d)\) Euclidean space
Basic Idea

2. Apply Multidimensional Scaling (MDS) on these distances, and encode all images in a low dimensional \((m-d)\) Euclidean space.
Basic Idea

3. We use the low dimensional representation of an image as its features, and call it *MDS codes*
Basic Idea

3. We use the low dimensional representation of an image as its features, and call it *MDS codes*

4. For an image classification application, MDS codes of training images are used to train classifiers
Basic Idea

3. We use the low dimensional representation of an image as its features, and call it *MDS codes*.

4. For an image classification application, MDS codes of training images are used to train classifiers.

image1: \((-0.005, 0.000, \ldots)\)

image2: \((-0.001, 0.013, \ldots)\)

\[\cdots\]
Basic Idea

5. Given a new testing image, measure the distances from this image to training images to encode it, then apply trained classifiers.
Basic Idea

5. Given a new testing image, measure the distances from this image to training images to encode it, then apply trained classifiers.
Basic Idea

5. Given a new testing image, measure the distances from this image to training images to encode it, then apply trained classifiers.
Contributions of Our Work

1. Proposing the MDS feature learning scheme
Contributions of Our Work

1. Proposing the MDS feature learning scheme

2. The iterated Levenberg-Marquardt algorithm for efficient encoding
Contributions of Our Work

1. Proposing the MDS feature learning scheme

2. The iterated Levenberg-Marquardt algorithm for efficient encoding

3. Exploring MDS with different image distances:
   - IMage Euclidean Distance (IMED) [38]
   - Spatial Pyramid Matching (SPM) distance [24]
   - Integrated Region Matching (IRM) distance [48]
MDS Model Training: Concepts

• Given $N$ images $\Omega = \{I_1, I_2, ..., I_N\}$
MDS Model Training: Concepts

• Given $N$ images $\Omega = \{I_1, I_2, ..., I_N\}$

• Measure pair-wise image distances

$$d(I_i, I_j) : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$$
MDS Model Training: Concepts

- Given $N$ images $\Omega = \{I_1, I_2, ..., I_N\}$
- Measure pair-wise image distances $d(I_i, I_j) : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$
- Encode each image $I_i$ to $x_i \in \mathbb{R}^m$
MDS Model Training: Concepts

- Given $N$ images $\Omega = \{I_1, I_2, ..., I_N\}$

- Measure pair-wise image distances
  
  $$d(I_i, I_j) : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$$

- Encode each image $I_i$ to $x_i \in \mathbb{R}^m$

- Representation error:
  
  $$e_{ij} = d(I_i, I_j) - \|x_i - x_j\|$$
MDS Model Training: Concepts

• Given $N$ images $\Omega = \{I_1, I_2, ..., I_N\}$

• Measure pair-wise image distances
  $d(I_i, I_j) : \Omega \times \Omega \rightarrow \mathbb{R}_{\geq 0}$

• Encode each image $I_i$ to $x_i \in \mathbb{R}^m$

• Representation error:
  $e_{ij} = d(I_i, I_j) - \|x_i - x_j\|$

• Raw stress (total cost to be minimized) [26]:
  \[
  \text{Stress}^* = \sum_{1 \leq i < j \leq N} e_{ij}^2
  \]
MDS Model Training: Concepts

• Problem:

\[
X^* = \arg \min_X \sum_{1 \leq i < j \leq N} (d(I_i, I_j) - \|x_i - x_j\|)^2
\]

where \( X = (x_1, \ldots, x_N)^T \)

• Existing methods:
  ▫ Iterative majorization algorithm (SMACOF) [30]
  ▫ Variants of SMACOF [40]-[43]

• Our solution:
  ▫ Iterated Levenberg-Marquardt algorithm (ILMA)
MDS Model Training: Concepts

• Problem:

\[ X^* = \arg \min_x \sum_{1 \leq i < j \leq N} (d(I_i, I_j) - \|x_i - x_j\|)^2 \]

where \( X = (x_1, \ldots, x_N)^T \)

• Existing methods:
  ▫ Iterative majorization algorithm (SMACOF) [30]
  ▫ Variants of SMACOF [40]-[43]

• Our solution:
  ▫ Iterated Levenberg-Marquardt algorithm (ILMA)
MDS Model Training: Algorithm

- Basic idea of ILMA:
  - Each time fix all $x_i$'s except for one, and apply the standard Levenberg-Marquardt algorithm [33][34]
MDS Model Training: Algorithm

- Basic idea of ILMA:
  - Each time fix all $x_i$’s except for one, and apply the standard Levenberg-Marquardt algorithm [33][34]
MDS Model Training: Algorithm

- Basic idea of ILMA:
  - Each time fix all $x_i$’s except for one, and apply the standard Levenberg-Marquardt algorithm [33][34]

New point added
MDS Model Training: Algorithm

- Basic idea of ILMA:
  - Each time fix all $x_i$'s except for one, and apply the standard Levenberg-Marquardt algorithm [33][34]
MDS Model Training: Algorithm

- Basic idea of ILMA:
  - Each time fix all $x_i$’s except for one, and apply the standard Levenberg-Marquardt algorithm [33][34]
MDS Model Training: Algorithm

• Basic idea of ILMA:
  ▫ Each time fix all $x_i$’s except for one, and apply the standard Levenberg-Marquardt algorithm [33][34]
MDS Model Training: Algorithm

• Please see our paper for algorithm details
• ILMA runs faster and converges to smaller raw stress than many other solutions
  ▫ Tested on the well-known Swiss roll flattening experiment [39][40]
MDS Model Training: Algorithm

- Please see our paper for algorithm details
- ILMA runs faster and converges to smaller raw stress than many other solutions
  - Tested on the well-known Swiss roll flattening experiment [39][40]

![Running time comparison with other solutions](image1)

![Swiss roll flattening results](image2)

This red curve is our method.
New Image Encoding

• After MDS model training, we need to be able to encode a new image $\tilde{I}$ to $\tilde{x}$:

$$\min_{\tilde{x}} \sum_{I_i \in \Omega_{\text{train}}} (||\tilde{x} - x_i|| - d(\tilde{I}, I_i))^2$$
New Image Encoding

• After MDS model training, we need to be able to encode a new image $\tilde{I}$ to $\tilde{x}$:

$$\min_{\tilde{x}} \sum_{I_i \in \Omega_{\text{train}}} \left( ||\tilde{x} - x_i|| - d(\tilde{I}, I_i) \right)^2$$

• This can be solved by the standard Levenberg-Marquardt algorithm [33][34]
Semantics-Sensitive Image Distances

• We have talked about:
  ▫ Feature learning
  ▫ MDS model training
  ▫ Encoding new image

• One thing left – what are we talking about when we say “distance”?
  ▫ A metric on set $\Omega=(I_1,I_2,\ldots,I_N)$ in the strict sense?
Semantics-Sensitive Image Distances

- We have talked about:
  - Feature learning
  - MDS model training
  - Encoding new image

- One thing left – what are we talking about when we say “distance”?
  - A metric on set \( \Omega = (I_1, I_2, \ldots, I_N) \) in the strict sense?
    - **No**, subadditivity triangle inequality does not necessarily hold.

\[
d(x, z) \leq d(x, y) + d(y, z)
\]
Semantics-Sensitive Image Distances

• We have talked about:
  ▫ Feature learning
  ▫ MDS model training
  ▫ Encoding new image

• One thing left – what are we talking about when we say “distance”?
  ▫ A metric on set $\Omega=(I_1,I_2,\ldots,I_N)$ in the strict sense?
    • No, subadditivity triangle inequality does not necessarily hold.
    $$d(x, z) \leq d(x, y) + d(y, z)$$
  ▫ Just a dissimilarity measurement!
Semantics-Sensitive Image Distances

- **IMage Euclidean Distance (IMED)** [38]
  - Traditional pixel-wise Euclidean distance on a smoothed version of two images
  - Low level, not much semantics information

- **Spatial pyramid matching (SPM) distance** [24]
  - Based on pyramid matching kernel [25]
  - Well applied to image classification
  - Highly semantics-sensitive

- **Integrated region matching (IRM) distance** [48]
  - Well applied to content-based image retrieval (CBIR)
  - Highly semantics-sensitive

We will skip the descriptions of these image distances
Please refer to original work [38][24][48]
Experiment 1: Car Recognition

- **Data:** UIUC car dataset [44]
  - (550 car, 500 non-car)

- **Task:** 2-way (binary) classification on fixed-length feature vectors

- **Classifier:** RBF kernel SVM [46][47]

- **Validation:** 5-fold cross validation
Experiment 1: Car Recognition

- **Methods employed:**
  1. PCA/kernel PCA on raw intensities
  2. MDS on IMage Euclidean Distances (IMED-MDS)
  3. MDS on Spatial Pyramid Matching distances (SPM-MDS)
  4. PCA on spatial pyramid vectors
Experiment 1: Car Recognition

• Methods employed:
  1. PCA/kernel PCA on raw intensities
  2. MDS on IMage Euclidean Distances (IMED-MDS)
  3. MDS on Spatial Pyramid Matching distances (SPM-MDS)
  4. PCA on spatial pyramid vectors

Make this comparison to see if MDS brings new information beyond pyramid kernels
Experiment 1: Car Recognition

- SPM-MDS outperforms all other methods
- SPM-MDS wins pyramid PCA, thus MDS encodes semantics beyond pyramid kernels
- IMED-MDS wins PCA / kernel PCA
- (Note: SPM1-MDS and SPM2-MDS use different scaling functions)
Experiment 1: Car Recognition

- 2-d feature scattering plot
Experiment 1: Car Recognition

- 2-d feature scattering plot
Experiment 2: Multi-Class Object Recognition

- **Data:** 12 categories from COREL dataset [48][50]

- **Task:** 12-way classification on fixed-length feature vectors

- **Classifier:** RBF kernel SVM [46][47]

- **Validation:** 5-fold cross validation
Experiment 2: Multi-Class Object Recognition

• Methods employed:
  1. PCA on color+HOG+LBP [5][6]
  2. Bag-of-visual-words (BOV) [19]-[21] on color+HOG+LBP
  3. MDS on SPM distances (SPM-MDS)
  4. MDS on IRM distances (IRM-MDS)
  5. Combined features
     • E.g. 12-d PCA + 12-d BOV = 24-d combined

• For method details, see our paper and references
Experiment 2: Multi-Class Object Recognition

COREL multi-class object recognition: precision

COREL multi-class object recognition: recall

feature dimension

precision

recall

feature dimension
Experiment 2: Multi-Class Object Recognition

Winner is SPM-MDS+BOV
Experiment 2: Multi-Class Object Recognition

• Conclusions:
  ▫ Without combining, PCA and bag-of-visual-words on color+HOG+LBP outperform MDS features 😞
Experiment 2: Multi-Class Object Recognition

• Conclusions:
  ▫ Without combining, PCA and bag-of-visual-words on color+HOG+LBP outperform MDS features 😞
  ▫ SPM-MDS combined with bag-of-visual-words significantly outperforms all other methods 😀
Experiment 2: Multi-Class Object Recognition

• Conclusions:
  ▫ Without combining, PCA and bag-of-visual-words on color+HOG+LBP outperform MDS features 😞
  
  ▫ SPM-MDS combined with bag-of-visual-words significantly outperforms all other methods 😊
  
  ▫ We conclude: SPM-MDS features capture semantics information from images that are not captured by other methods, such as PCA or bag-of-visual-words 😊
Experiment 2: Multi-Class Object Recognition

- Classification confusion matrix of SPM-MDS + BOV

- SPM-MDS running time:
  - Training MDS model on 960 images takes **20 min**
  - Encoding one new image takes **0.3 s** (including feature extraction)
Final Discussions

• We proposed a new feature learning scheme:
  ▫ Feature learning from semantics-sensitive image distances
Final Discussions

• We proposed a new feature learning scheme:
  ▫ Feature learning from semantics-sensitive image distances

• It works well for object recognition tasks:
  ▫ MDS on IMED, SPM distances, IRM distances
  ▫ Experiments with UIUC car data, COREL images
Final Discussions

• We proposed a new feature learning scheme:
  ▫ Feature learning from semantics-sensitive image distances

• It works well for object recognition tasks:
  ▫ MDS on IMED, SPM distances, IRM distances
  ▫ Experiments with UIUC car data, COREL images

• We also introduced an efficient MDS algorithm:
  ▫ Iterated Levenberg-Marquardt algorithm (ILMA)
  ▫ Code can be downloaded at: https://sites.google.com/site/mdsfeature/
Final Discussions

• There is more to do:
  ▫ MDS model training on large dataset is still slow. Can we make it parallel? (on multi-core, GPU, …)
  ▫ Are there other semantics-sensitive image distances that can be employed?
  ▫ Other applications apart from object recognition?
    • Style classification
    • Affective image classification
    • Aesthetics analysis
    • Emotion/sentiment detection
    • Face beautification rating
    • ……
Final Discussions

• There is more to do:
  ▫ MDS model training on large dataset is still slow. Can we make it parallel? (on multi-core, GPU, …)
  ▫ Are there other semantics-sensitive image distances that can be employed?
  ▫ Other applications apart from object recognition?
    • Style classification
    • Affective image classification
    • Aesthetics analysis
    • Emotion/sentiment detection
    • Face beautification rating
    • ……

All depends on how you can define a distance!
Thank you for your interest!

Welcome to our project wiki site:

https://sites.google.com/site/mdsfeature/
References


References

References