Image Information Distance Analysis and Applications

Nima Nikvand

April 3, 2014
Introduction

Outline

- Motivations
- Literature Review
- Generic Image Similarity based on Kolmogorov complexity
- Entropy Approximation of Kolmogorov complexity and its applications
- Information Distance based Feature extraction and its applications
- Conclusion and Future Works
Motivations

- Traditional Image Similarity methods have many problems in many scenarios.
- Attneave’s experiments and Barlow’s sensory coding principle: Human mind tends to describe changes in visual scenes using simplest explanation, Kolmogorov complexity provides a similar explanation.

Objective

- Developing novel frameworks and algorithms for practical applications of information distances in image similarity assessment
Motivations: where other similarity measures fail!
Motivations: Applications of Image Similarity Assessment
## Motivations: Traditional Image Quality Assessment Methods

### Traditional Methods
- **Mathematical Measures:** Minkowski error: Mean Squared Error (MSE) [Wang et al. 09]
- **Error Visibility Approaches** [Wang et al. 04]
- **Structural Similarity Index Measure (SSIM)** [Wang et al. 04]
- **Visual Information Fidelity (VIF)** [Sheikh et al. 06]

### Limitations
- Low Correlation to Visual Perception
- Constrained to Specific Applications
1. Introduction

2. Background

3. Contribution I: Generic Image Similarity


5. Contribution IV: Information distances as features

6. Conclusions and Future work
Introduction

Background

Contribution I: Generic Image Similarity

Contribution II & III: Entropy & Kolmogorov Complexity

Contribution IV: Information distances as features

Conclusions and Future work

Normalized Information Distance [Li et al. 04]
Kolmogorov Complexity

A turing machine, picture courtesy of www.aturingmachine.com

**Complexity**

- **Kolmogorov Complexity of a given object** \( x \):
  \[
  K(x) = \min\{l(p) : U(p) = x\}
  \]

- \( p \) is an instantaneously decodable prefix program (its length is encoded in \( p \))

- **Kolmogorov Complexity is non-computable**
Normalized Information Distance [Li et al. 04]

Information Distances

- Information Distance between objects x and y:
  \[ ID(x, y) = \max\{K(x|y), K(y|x)\} \]

- Normalized Information Distance (NID) [Li et al. 04]:
  \[ NID(x, y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}} \]

- Normalized Compression Distance (NCD) [Li et al. 04]:
  \[ NCD(x, y) = \frac{C(xy) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}} \]
1 Introduction

2 Background

3 Contribution I: Generic Image Similarity

4 Contribution II & III: Entropy & Kolmogorov Complexity

5 Contribution IV: Information distances as features

6 Conclusions and Future work
Normalized Conditional Compression Distance (NCCD)

Conditional Complexity by Compression

- Normalized Conditional Compression Distance (NCCD)
  \[
  \text{NCCD}(x, y) = \frac{\max\{C_T(x|y), C_T(y|x)\}}{\max\{C(x), C(y)\}}
  \]

- Estimating conditional complexity:
  \[
  C_T(y|x) = \min_i \{C[y - T_i(x)] + C_p^i[p(T_i, x)] + \log_2(N)\}
  \]

- **C**: Lossless Image Compressor (e.g. CALIC)
- **C_p^i**: General Transform Parameter Compressor
- **N**: Length of the list of Transforms
- **T_i**: i-th Transform
- **T_i(x)**: Transformed image of **x**
- **p(T_i, x)**: Parameters to encode the transform
List of Transforms

Transforms

- Global Contrast and Luminance Change: \( s = \alpha(r - \bar{r}) + \bar{r} + \beta \)
- Global Fourier Power Spectrum Scaling: \( Y(\omega) = p_1 X(\omega) + p_2 \)
- Global Affine Transform: Matching images by global affine registration
- Local Registration: Multi-scale local affine registration
List of Transforms

Transforms

- **Global Contrast and Luminance Change:** \( s = \alpha (r - \bar{r}) + \bar{r} + \beta \)
- **Global Fourier Power Spectrum Scaling:** \( Y(\omega) = p_1 X(\omega) + p_2 \)
- **Global Affine Transform:** Matching images by global affine registration
- **Local Registration:** Multi-scale local affine registration
List of Transforms

- Global Contrast and Luminance Change: $s = \alpha (r - \bar{r}) + \bar{r} + \beta$
- Global Fourier Power Spectrum Scaling: $Y(\omega) = p_1 X(\omega) + p_2$
- Global Affine Transform: Matching images by global affine registration
- Local Registration: Multi-scale local affine registration
List of Transforms

- **Global Contrast and Luminance Change**: \[ s = \alpha (r - \bar{r}) + \bar{r} + \beta \]
- **Global Fourier Power Spectrum Scaling**: \[ Y(\omega) = p_1 X(\omega) + p_2 \]
- **Global Affine Transform**: Matching images by global affine registration
- **Local Registration**: Multi-scale local affine registration
Showcase: Geometric distortions

NCCD = 0.0189
NCCD = 0.4901
NCCD = 0.0016
NCCD = 0.021
Showcase: Natural Scene Images

<table>
<thead>
<tr>
<th>Image</th>
<th>NCCD</th>
<th>SSIM</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8121</td>
<td>0.4649</td>
<td>73.4</td>
</tr>
<tr>
<td>2</td>
<td>0.2189</td>
<td>0.7059</td>
<td>49.8</td>
</tr>
<tr>
<td>3</td>
<td>0.3901</td>
<td>0.8425</td>
<td>78.4</td>
</tr>
<tr>
<td>4</td>
<td>0.0091</td>
<td>0.9094</td>
<td>105.5</td>
</tr>
</tbody>
</table>
Masking effect

Texture

Noisy Texture

Smooth

Noisy Smooth
Perceptual complexity

### Perceptual conditional image

#### Divisive Normalization Transform [Heeger et al. 92, Li et al. 09]
- Transform stimuli into a perceptually uniform space
- Degree of perceptual relevance of bits, reducing redundancy in natural scene similar to sensory systems using efficient coding transforms
Video Encoder as Conditional Image Compressor

- Motion Estimation in Video Coding, image courtesy of www.iptvdictionary.com

H.264 Video Encoder

- Video encoder is a conditional compressor
- The first frame is a reference, the second frame is compressed based on the first frame and some motion vectors
- It can be encoded in lossy mode to account for redundancy reduction
Results: Preliminary Tests

Shields & Butterflies

Digit Recognition

<table>
<thead>
<tr>
<th>Digit</th>
<th>NCCD (%)</th>
<th>MSE (%)</th>
<th>SSIM (%)</th>
<th>CW-SSIM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>84.0</td>
<td>76.1</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>96.3</td>
<td>65.4</td>
<td>45.3</td>
<td>98.4</td>
</tr>
<tr>
<td>3</td>
<td>90.9</td>
<td>49.4</td>
<td>47.7</td>
<td>97.1</td>
</tr>
<tr>
<td>4</td>
<td>93.5</td>
<td>63.8</td>
<td>41.6</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>91.2</td>
<td>47.7</td>
<td>18.5</td>
<td>96.3</td>
</tr>
<tr>
<td>6</td>
<td>96.6</td>
<td>56.4</td>
<td>42.0</td>
<td>97.9</td>
</tr>
<tr>
<td>7</td>
<td>92.9</td>
<td>68.3</td>
<td>60.9</td>
<td>94.2</td>
</tr>
<tr>
<td>8</td>
<td>97.4</td>
<td>49.8</td>
<td>39.1</td>
<td>99.6</td>
</tr>
<tr>
<td>9</td>
<td>91.9</td>
<td>59.3</td>
<td>51.4</td>
<td>100</td>
</tr>
<tr>
<td>0</td>
<td>90.2</td>
<td>51.4</td>
<td>46.5</td>
<td>93.0</td>
</tr>
<tr>
<td>All</td>
<td>94.1</td>
<td>59.6</td>
<td>46.9</td>
<td>97.7</td>
</tr>
</tbody>
</table>
## Texture Classification by NCCD

<table>
<thead>
<tr>
<th>Dataset</th>
<th>NCCD-I (%)</th>
<th>NCCD-II (%)</th>
<th>CK-I (%)</th>
<th>SBC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brodatz</td>
<td>73.2</td>
<td>70.1</td>
<td>54.0</td>
<td>76.2</td>
</tr>
<tr>
<td>Camouflage</td>
<td>81.9</td>
<td>73.0</td>
<td>87.5</td>
<td>87.0</td>
</tr>
<tr>
<td>Tire Tracks</td>
<td>93.8</td>
<td>89.5</td>
<td>79.2</td>
<td>85.2</td>
</tr>
<tr>
<td>VVT Wood</td>
<td>91.9</td>
<td>88.0</td>
<td>80.5</td>
<td>85.2</td>
</tr>
<tr>
<td>VisTex</td>
<td>39.3</td>
<td>42.0</td>
<td>32.9</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Face Recognition by NCCD

Face Retrieval in AT&T Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>NCCD-I(%)</th>
<th>NCCD-II (%)</th>
<th>CK-I (%)</th>
<th>SBC(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT&amp;T</td>
<td>82.8</td>
<td>70.1</td>
<td>76.5</td>
<td>81.6</td>
</tr>
<tr>
<td>Yale</td>
<td>73.1</td>
<td>73.9</td>
<td>64.1</td>
<td>65.9</td>
</tr>
</tbody>
</table>
Introduction

Background

Contribution I: Generic Image Similarity

Contribution II & III: Entropy & Kolmogorov Complexity

Contribution IV: Information distances as features

Conclusions and Future work

Nima Nikvand

Image Information Distance Analysis and Applications
Kolmogorov Complexity and Shannon Entropy

Estimating Kolmogorov Complexity

- If μ is a probability measure of source X and \( x^n \) is length n realization of the source [Daby et al. 03):

  \[
  \lim_{n \to \infty} \frac{K(x^n)}{n} = H(\mu)
  \]

- If source can be expressed using probability model \( f(x) \) [Cover 1991]:

  \[
  0 \leq \left( \sum_x f(x) K(x) - H(X) \right) \leq K(f) + O(1)
  \]
Perceptual NID for Image Quality Assessment

**Perceptual Information Distance Analysis**

- Transform domain to enhance independence:
  \[
  K(x|y) = \sum_{i=1}^{n} K(x_n|y_n)
  \]

- Shannon entropy version of NID:
  \[
  NIDs(x, y) = \frac{\max\{\sum_{i=1}^{n} H(x_n|y_n), \sum_{i=1}^{n} H(y_n|x_n)\}}{\max\{\sum_{i=1}^{n} H(x_n), \sum_{i=1}^{n} H(y_n)\}}
  \]
Generic System Model

Perceptual Information

- Gaussian Scale Mixture model [Simoncelli’ 97]:
  \[ C = SU = \{s_i \vec{U}_i, i \in I\} \]

- Generic Channel Distortion Model [Sheikh et al. 06]:
  \[ D = gC + \nu; \quad E = C + \mathcal{N}; \quad F = D + \mathcal{N}’ \]

- \( I(E; C) \) Info. in ideal image
- \( I(F; C) \) Info. in distorted image
Perceptual Similarity Measure

Framework

- Normalized Perceptual Information Distance (NPID):

\[ \text{NPID}(E, F) = 1 - \frac{\sum_j I(\vec{E}_j; \vec{F}_j)}{\max\{\sum_j H(\vec{E}_j), \sum_j H(\vec{F}_j)\}} \]

- Normalized Perceptual Information Similarity (NPIS):

\[ \text{NPIS}(E, F) = 1 - \text{NPID}(E, F) = \frac{\sum_j I(\vec{E}_j; \vec{F}_j)}{\max\{\sum_j I(\vec{E}_j, \vec{C}_j), \sum_j I(\vec{F}_j, \vec{C}_j)\}} \]
Perceptual Similarity Measure

Framework

- Local NPIS measure:

$$L\text{-NPIS}_i = \frac{I(\overrightarrow{E}_i; \overrightarrow{E}_i)}{\max\{I(\overrightarrow{E}_i; \overrightarrow{C}_i), I(\overrightarrow{F}_i; \overrightarrow{C}_i)\}}$$

- Information content-weighted NPIS [Wang et al. 11]

$$IW\text{-NPIS}_j = \sum_i \omega_{j,i} L\text{-NPIS}_i \sum_i \omega_{j,i}$$

- Incorporating fine-to-coarse scale weights [Wang et al. 03]:

$$IW\text{-NPIS} = \prod_{j=1}^{M} (IW\text{-NPIS}_j)^{\beta_j}$$
Performance Comparison: TID2008 database

Tampere Image Database 2008

- 25 Reference $\times$ 17 types of distortions $\times$ 4 levels.

<table>
<thead>
<tr>
<th>Model</th>
<th>PLCC</th>
<th>MAE</th>
<th>RMS</th>
<th>SRCC</th>
<th>KRCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>0.5223</td>
<td>0.8683</td>
<td>1.1435</td>
<td>0.5531</td>
<td>0.4027</td>
</tr>
<tr>
<td>IW-SSIM</td>
<td>0.8579</td>
<td>0.5276</td>
<td>0.6895</td>
<td>0.8559</td>
<td>0.6636</td>
</tr>
<tr>
<td>VIF</td>
<td>0.8090</td>
<td>0.5990</td>
<td>0.7888</td>
<td>0.7496</td>
<td>0.5863</td>
</tr>
<tr>
<td>NPIS</td>
<td>0.7855</td>
<td>0.6239</td>
<td>0.8111</td>
<td>0.7682</td>
<td>0.5013</td>
</tr>
<tr>
<td>IW-NPIS</td>
<td>0.8244</td>
<td>0.5846</td>
<td>0.7637</td>
<td>0.8167</td>
<td>0.5819</td>
</tr>
</tbody>
</table>
Applications to HDR Tone-Mapping

HDR Image

Operator 1

Operator 2
Tone-Mapping Operators

Operators

- Family of operators that can be expressed as linear combination of basis functions [Yeganeh et al. 12]:

\[ f(l) = \sum_{k=0}^{n-1} c_k \phi_k(l) \]

- Three segment Piecewise-linear operator:

\[ f(l) = \frac{l - W_l}{W} + c_1 \Delta \left( \frac{l - l_1}{W} \right) + c_2 \Delta \left( \frac{l - l_2}{W} \right) \]

- Three segment Sine-basis operator:

\[ f(l) = \frac{l - W_l}{W} + c_1 \sin \left( \frac{\pi (l - W_l)}{W} \right) + c_2 \sin \left( \frac{2\pi (l - W_l)}{W} \right) \]
Applications to HDR Tone-Mapping

Tone-mapping operator

$c_1$ & $c_2$ domain
Applications to HDR Tone-Mapping

Selecting Parameters

- Minimize information loss by minimizing NID:

\[ f_{opt-NID} = \arg \min_{f \in F_{[l_l, l_u]}} NID(x, T_f(x)) \]

- Normalized Shannon information distance:

\[ NID(x, y) \approx \frac{\max\{H(x|y), H(y|x)\}}{\max\{H(x), H(y)\}} \]

- \( H(y|x) = 0, \hat{x} = R_{f^{-1}}(y) = \text{round}\{f^{-1}(y)\}, H(x) > H(y): \)

\[ f_{opt-NID} = \arg \min_{f \in F_{[l_l, l_u]}} \frac{H(x - R_{f^{-1}}(T_f(x)))}{H(x)} \]
Applications to HDR Tone-Mapping

Top view of NID surface

Corresponding Operators

Nima Nikvand
Image Information Distance Analysis and Applications
Parameter Selection Scheme

- Linear Image
- MSE Image
- SSIM Image
- NID Image

Operators:
- MSE Operator
- NID Operator
- Linear Operator
- SSIM Operator
1 Introduction

2 Background

3 Contribution I: Generic Image Similarity

4 Contribution II & III: Entropy & Kolmogorov Complexity

5 Contribution IV: Information distances as features

6 Conclusions and Future work
Photoshop Similarity Assessment

Before

After
Photoshop Similarity Assessment

Re-implementation of Original Algorithm

<table>
<thead>
<tr>
<th>Method</th>
<th>PLCC</th>
<th>SRCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Re-implementation</td>
<td>0.5740</td>
<td>0.5932</td>
</tr>
<tr>
<td>NCCD-I</td>
<td>0.1811</td>
<td>0.1901</td>
</tr>
<tr>
<td>NCCD-II</td>
<td>0.3518</td>
<td>0.3621</td>
</tr>
</tbody>
</table>
Photoshop Similarity Assessment: Information distances

Features

- $L^*a^*b^*$ Color space difference attributes of image:

$$\Delta = \begin{pmatrix} \Delta L^T \\ \Delta a^*T \\ \Delta b^*T \end{pmatrix}^T$$

- Motion vector attributes of image:

$$M = \begin{pmatrix} M_x^T \\ M_y^T \end{pmatrix}^T$$

- Information distance features, based on image attributes:

$$\Delta \sim N(\mu_\Delta, \Sigma_\Delta), \ M \sim N(\mu_M, \Sigma_M)$$

$$h(\Delta) = \frac{1}{2} \log(2\pi e)^3 |\Sigma_\Delta|$$

$$h(M) = \frac{1}{2} \log(2\pi e)^2 |\Sigma_M|$$
Photoshop Similarity Assessment: Results

Train All / Test All (PLCC=78%)  
LOOCV (PLCC=68%)
Photoshop Similarity Assessment: Good results!

(MOS: 3.78 / P: 3.78) 
(MOS: 3.64 / P: 3.89) 
(MOS: 3.42 / P: 3.42)
Photoshop Similarity Assessment: Bad results!

(MOS: 4.62 / P: 3.31)

(MOS: 3.56 / P: 2.53)

(MOS: 4.08 / P: 2.89)
1. Introduction

2. Background

3. Contribution I: Generic Image Similarity


5. Contribution IV: Information distances as features

6. Conclusions and Future work
Conclusions and Future work

- Normalized Information distance provides a strong theoretical framework for image similarity and image quality applications.

- Information distances can be used as input features to Machine learning algorithms to predict similarity / quality of images and video.

- Information distances are proper tools for parameter-selection in many image processing algorithms.

- Information theoretic approaches are proper tools for modeling HVS.
Questions
Publications


