Discriminative Image Warping using Attribute Flow

Jianbo Shi

Joint work with Weiyu Zhang and Praveen Srinivasan

GRASP Laboratory
Attribute flow: idea 1

need to frame features right
Attribute flow: idea 2

Transformation that explains attribute change
Attribute Flow

Matching same object with deformation

Matching object with pose variation

Matching instances in the same object category
Challenges

Q1: How to compute the spatial transformation efficiently without explicit search?

Q2: How to constrain the computation such that it leads to valid spatial transformation?

Q3: How to deal with clutter environment?
Image Deformation as Attribute Flow

Input Model

Input Image

T
Transformation that explains attribute change

$$\min_T \int_{\mathbb{R}^2} \| A_I(x) - A_J(T(x)) \|_p dx$$
Image Deformation as Attribute Flow

Model Attribute

Image Attribute

\[ M \]
$$\min_M \int_{\mathbb{R}^{n+2}} |\delta(y - M(A^+_I)) - \delta(y - A^+_J)|_1 \, dy$$

Assuming the first elements of $A(I(x))$ are the spatial location $x$

$$M(A(I(x))) = (T(x), m(a_2), \ldots, m(a_n))$$

Attribute: edge orientation, or oriented filter
\[
\min_{\mathcal{M}} \int_{\mathbb{R}^{n+2}} |\delta(y - \mathcal{M}(\mathcal{A}^+_I)) - \delta(y - \mathcal{A}^+_J)|_1 \, dy
\]

Assuming the first elements of \( \mathcal{A}(\mathcal{I}(x)) \) are the spatial location \( x \)

\[\mathcal{M}(\mathcal{A}(\mathcal{I}(x))) = (T(x), m(a_2), \ldots, m(a_n)).\]

Matching instances in the same object category

Attribute: **edge orientation**, or oriented filter
Image Deformation as Attribute Flow
Solve Attribute Flow via Histogram Matching

$\mathcal{H}(A_{\mathcal{I}}^+) \ldots$ 

\[ \ldots \]

Model Mass Preserving 

$\mathcal{H}(A_{\mathcal{J}}^+)$
Image Deformation as Attribute Flow

Optical Flow

Marginalization

Attribute Flow

Interpolation

Histogram Flow
Histogram Flow

Input Model

$\mathcal{H}(A^+_I)$

$F$

$\mathcal{H}(A^+_F)$

Foreground w/o Clutters

How to differentiate two flows?
Histogram Flow Ground Distance Cost

**Input Model**

\[ \mathcal{H}(A^+_F) \]

**Foreground w/o Clutters**

\[ \mathcal{H}(A^+_F) \]

Earth Mover’s Distance:

\[
GD(F) = \sum_{k,l} F_{k,l}d_{k,l}
\]
Linear programming for solving EMD
But, Ground Distance fails in rotation!

Minimizing ground distance fails when the object rotates.
Affine Constraint on Spatial Transformation

Affine Constraint 1:
\[ T(x_p) = \sum_{i=1}^{3} \alpha_p^i T(x_i) \]
Affine Constraint on Spatial Transformation

Affine Constraint 2: \[ r(\mathbf{T}(x_p, \theta_p)) \sim \sum_{i=1}^{3} (\beta_p^i - \alpha_i^p) \mathbf{T}(x_i) \]

\[ r(\theta) = [\cos(\theta), \sin(\theta)]^T \]
$F$ can be viewed as a probability encoding of $T$. 

Model Mass Preserving
Geometrical constraint 1

\[ T(x_z) = \sum_{i=1}^{3} \alpha_i^z T(x_i) \]

\[ \text{AffCon}_x(F, k) : \| E_F(T(p_k)) - \sum_{i=1}^{k} \alpha_i^k E_F(T(p_{k_i})) \|_1 \]
Geometrical constraint 2

\[ T([\cos(\theta_z) \sin(\theta_z)]^T) \sim \sum_{i=1}^{3} (\beta_i^z - \alpha_i^z)T(x_i), \]

\[ E_F(T([\cos(\theta_k) \sin(\theta_k)]^T)) \sim \sum_{i=1}^{3} (\beta_i^k - \alpha_i^k)E_F(T(p_{ki})). \]
Constrained Attribute Histogram Flow

Input Model

\( \mathcal{H}(A^+_I) \)

AffCon\((F, k)\)

AffCon\(x\)(\(F, k\)):

\[ |E_F(T(x_k)) - \sum_{i=1}^3 \alpha_i^k E_F(T(x_{k_i}))|_1 \]

AffCon\(\theta\)(\(F, k\)):

\[ \min_{g} |g \cdot r(E_F(T(\theta_k))) - \sum_{i=1}^3 (\beta_i^k - \alpha_i^k) E_F(T(x_{k_i}))|_1 \]

Foreground w/o Clutters

Soft Affine Constraint
Q: How to efficiently select foreground region to avoid “cherry-picking”?
Match with Selected Region

“Shape Basis”

Foreground w/o Clutters

Input Image w/ Clutters
Match with Selected Region

Contour Selection

$H_J$ #bins #contours $\times$ $x_{sel}$

Foreground w/o Clutters

Input Image w/ Clutters
Application: Simultaneous Detection and Alignment

Input

Shape Model

Image with contours, or segments

Output

Detection

Shape Alignment

Foreground Segmentation
Algorithm Overview

Input Model

- Constrained Bin
- Anchor Bins

$\mathcal{H}(A^+_I)$

$F \text{ Hist. Flow}$

$F^T \mathcal{H}(A^+_I)$

$\mathcal{J}X^{\text{sel}}$

$X^{\text{sel}}$ Contour Selection

Input Image w/ Clutters
Algorithm Overview

Constrained Histogram Flow

\[
\min_{F, x^{\text{sel}}} \quad \text{GD}(F) + \lambda \sum_{k=1}^{m} \text{AffCon}(F, k) + \gamma |F^T \mathcal{H}(A_I^+) - \mathcal{H} \mathcal{J} x^{\text{sel}}|_1
\]
\[
s.t. 
F1 = 1, 
F \geq 0, 
x^{\text{sel}} \in [0, 1]^{|C|}
\]
Image Deformation as Attribute Flow

Attribute Optical Flow
\[
\min_T \int_{\mathbb{R}^2} \| A_I(x) - A_J(T(x)) \|_p dx
\]

Attribute Flow
\[
\min_M \int_{\mathbb{R}^{n+2}} |\delta(y - M(A_I^+)) - \delta(y - A_J^+)\|_1 dy
\]

Attribute Histogram Flow (LP)
\[
\min_F |F^T \mathcal{H}(A_I^+) - \mathcal{H}(A_J^+)\|_1
\]
\[
\text{s.t. } F1 = 1 \quad F \geq 0
\]
Model Mass Preserving
Challenges

Q1: How to compute the spatial transformation efficiently without explicit search?
A : Attribute Histogram Flow.

Q2: How to constrain the computation such that it leads to valid spatial transformation?
A : Soft Affine Constraint on expectation.

Q3: How to deal with clutter environment?
A : Contour Selection.
Detection Example

Shape Model

Image and Contours

Constrained Histogram Matching

Object Center Vote Map

For each detection candidate

Local Maxima + Rough Location
Detection Example

- Shape Model
- Image and Contours

Object Center Vote Map

Local Maxima + Rough Location

For each detection candidate

Ground distance only

Ground distance + Affine
Discrimination in Canonical Space

Object Center Vote Map

Local Maxima + Rough Location

For each detection candidate

Constrained Histogram Matching

Warp contour to canonical Model Space + Selection

Shape Model

Image and Contours

Selected Warped Contour

Model Space
Results on ETHZ
Quantitative Evaluation

Average Precision

<table>
<thead>
<tr>
<th>Attribute Flow</th>
<th>Applelogos</th>
<th>Bottles</th>
<th>Giraffes</th>
<th>Mugs</th>
<th>Swans</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ma et al. [6]</td>
<td>0.881</td>
<td>0.920</td>
<td>0.756</td>
<td>0.868</td>
<td>0.959</td>
<td>0.877</td>
</tr>
<tr>
<td>Srinidhi et al. [10]</td>
<td>0.845</td>
<td>0.916</td>
<td>0.787</td>
<td>0.888</td>
<td>0.922</td>
<td>0.872</td>
</tr>
<tr>
<td>Maji et al. [7]</td>
<td>0.869</td>
<td>0.724</td>
<td>0.742</td>
<td>0.806</td>
<td>0.716</td>
<td>0.771</td>
</tr>
<tr>
<td>Lu et al. [5]</td>
<td>0.844</td>
<td>0.641</td>
<td>0.617</td>
<td>0.643</td>
<td>0.798</td>
<td>0.709</td>
</tr>
<tr>
<td>Toshev et al. [12]</td>
<td>0.983</td>
<td>0.936</td>
<td>0.713</td>
<td>0.718</td>
<td>0.973</td>
<td>0.865</td>
</tr>
</tbody>
</table>
Results on ETHZ using Segments
Piecewise affine Attribute Flow:

\[ b \in B \]

\[ F^1 \quad F^2 \]

\[ E_{F^1}(T(b)) \neq E_{F^2}(T(b)) \]
<table>
<thead>
<tr>
<th>Attribute Flow</th>
<th>Head</th>
<th>Hips</th>
<th>Knees</th>
<th>Ankles</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhu et al. [15]</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>30</td>
</tr>
<tr>
<td>Mykhaylo et al. [1]</td>
<td>24</td>
<td>36.5</td>
<td>43</td>
<td>71</td>
<td>47</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>12</strong></td>
<td><strong>27</strong></td>
<td><strong>24</strong></td>
<td><strong>27</strong></td>
<td><strong>26.5</strong></td>
</tr>
</tbody>
</table>
$T: Spatial\ Flow$

$T = T(F)$

$F: Attribute\ Flow$