The Complete Rank Transform: A Tool for Accurate and Morphologically Invariant Matching of Structures

Oliver Demetz, David Hafner, and Joachim Weickert

Mathematical Image Analysis Group
Saarland University
Saarbrücken, Germany

Partly funded by the German Research Foundation (DFG)
Introduction

Illumination changes in reality

source: KITTI benchmark
Introduction

Illumination changes in reality

source: KITTI benchmark
Introduction

**Illumination changes in reality**

![Illumination change example](source: KITTI benchmark)

- Important for optic flow
- Assumption that intensity of objects stays constant violated
- In this work: only assume invariance under *monotonically* increasing greyvalue rescalings (e.g., additive, multiplicative)

→ develop morphologically invariant descriptor
Outline

- Introduction
- Complete Rank Transform
- Variational Optic Flow Model
- Experiments
- Conclusions
Complete Rank Transform

Rank Transform (RT)  (Zabih and Woodfill 1994)

- **Idea:**
  
  *How many pixels are smaller than me?*

- Invariant under monotonically increasing transformations
Complete Rank Transform

Rank Transform (RT) (Zabih and Woodfill 1994)

- Idea:
  
  How many pixels are smaller than me?

- Invariant under monotonically increasing transformations
Complete Rank Transform

**Rank Transform (RT)**  (Zabih and Woodfill 1994)

- **Idea:**
  
  *How many pixels are smaller than me?*

- **Invariant under monotonically increasing transformations**

- **Resulting signature:**

\[
 s_{RT} = 5
\]
Complete Rank Transform

**Census Transform (CT)**  (Zabih and Woodfill 1994)

- **Idea:**
  
  *Which pixels are smaller than me?*

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>14</td>
<td>83</td>
</tr>
<tr>
<td>88</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>65</td>
</tr>
</tbody>
</table>

Intensity values

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Census
Census Transform (CT) (Zabih and Woodfill 1994)

- **Idea:**
  
  *Which pixels are smaller than me?*

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

<table>
<thead>
<tr>
<th>4</th>
<th>14</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>65</td>
</tr>
</tbody>
</table>

Intensity values

<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Census
**Complete Rank Transform**

**Census Transform (CT)**  (Zabih and Woodfill 1994)

- **Idea:**
  
  Which pixels are smaller than me?

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

- Resulting signature:

  \[ s_{CT} = (1, \quad )^T \]
Complete Rank Transform

Census Transform (CT)  (Zabih and Woodfill 1994)

- Idea:
  
  *Which pixels are smaller than me?*

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

- Resulting signature:

\[ s_{CT} = (1, 1, \ldots)^\top \]
Complete Rank Transform

Census Transform (CT)  (Zabih and Woodfill 1994)

- Idea:
  - Which pixels are smaller than me?

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

- Resulting signature:

\[ s_{CT} = (1, 1, 0, )^T \]
Complete Rank Transform

Census Transform (CT)  (Zabih and Woodfill 1994)

- Idea:
  
  *Which pixels are smaller than me?*

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

- Resulting signature:

\[ \mathbf{s}_{CT} = (1, 1, 0, 0, 1, \ldots) \]
Complete Rank Transform

Census Transform (CT) (Zabih and Woodfill 1994)

- Idea:
  - Which pixels are smaller than me?
- Invariant under monotonically increasing transformations
- Census signatures carry spatial information
- Resulting signature:

\[ s_{CT} = (1, 1, 0, 0, 1, 1, 1, 0)^T \]

Intensity values:

<table>
<thead>
<tr>
<th>4</th>
<th>14</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>65</td>
</tr>
</tbody>
</table>

Census:

<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Complete Rank Transform

Census Transform (CT)  (Zabih and Woodfill 1994)

- Idea:
  
  *Which pixels are smaller than me?*

- Invariant under monotonically increasing transformations

- Census signatures carry spatial information

- Resulting signature:

  \[ s_{CT} = (1, 1, 0, 0, 1, 1, 1, 0)^\top \]
Complete Rank Transform (CRT)

- Capture the rank of every pixel in the local patch
- Vector-valued signature in every pixel
- Invariant under monotonically increasing transformations
- Carries as much local image information as possible

<table>
<thead>
<tr>
<th>4</th>
<th>14</th>
<th>83</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>65</td>
</tr>
</tbody>
</table>

Intensity values

<table>
<thead>
<tr>
<th>1</th>
<th>3</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

Census
Complete Rank Transform (CRT)

- Capture the rank of every pixel in the local patch
- Vector-valued signature in every pixel
- Invariant under monotonically increasing transformations
- Carries as much local image information as possible

Resulting signature:

$$s_{\text{CRT}} = (1, 3, 7, \ldots)^\top$$

<table>
<thead>
<tr>
<th>Intensity values</th>
<th>Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>88</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
Complete Rank Transform (CRT)

- Capture the rank of every pixel in the local patch
- Vector-valued signature in every pixel
- Invariant under monotonically increasing transformations
- Carries as much local image information as possible

Resulting signature:

\[ s_{CRT} = (1, 3, 7, 8, 5, 1, \ldots)^\top \]

<table>
<thead>
<tr>
<th>Intensity values</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
</tr>
<tr>
<td>88</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
Complete Rank Transform (CRT)

- Capture the rank of every pixel in the local patch
- Vector-valued signature in every pixel
- Invariant under monotonically increasing transformations
- Carries as much local image information as possible

Resulting signature:

\[ s_{\text{CRT}} = (1, 3, 7, 8, 5, 1, 0, 4, 6)^\top \]
**Complete Rank Transform (CRT)**

- Capture the rank of every pixel in the local patch
- Vector-valued signature in every pixel
- Invariant under monotonically increasing transformations
- Carries as much local image information as possible
- Resulting signature:

\[
s_{\text{CRT}} = (1, 3, 7, 8, 5, 1, 0, 4, 6)^\top
\]

<table>
<thead>
<tr>
<th>Intensity values</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
</tr>
<tr>
<td>88</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Census</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Complete Rank Transform
- Variational Optic Flow Model
- Experiments
- Conclusions
Variational Optic Flow Model

- Modify model of Brox et al. (2004)
- Minimise functional for optic flow field $\begin{bmatrix} u \v v \end{bmatrix}^T : \Omega \to \mathbb{R}^2$

$$E(u, v) = \int_\Omega (D + \alpha R) \, dx \, dy$$

- Data term $D$ models constancy of Complete Rank signatures of corresponding positions

$$\left\| \mathbf{s}_{\text{CRT}}(x + u, y + v, t + 1) - \mathbf{s}_{\text{CRT}}(x, y, t) \right\|^2$$

- Regularisation term $R$ leads to piecewise smooth solution
- Both terms equipped with robust estimator functions
- Typical warping-based coarse-to-fine minimisation (Brox et al. 2004)
Outline

- Introduction
- Complete Rank Transform
- Variational Optic Flow Model
- Experiments
- Conclusions
Experiments - $\gamma$ Changes

- Adjust second frame with an exponential function $f_\gamma = f_{\text{max}} \cdot \left(\frac{f}{f_{\text{max}}}\right)^\gamma$

$\gamma = 0.1$  $\gamma = 1$  $\gamma = 3$

$\rightarrow$ Unconditional morphological invariance
$\rightarrow$ Outperform Rank and Census
Experiments - KITTI Vision Benchmark

- GCPR Special Session Imagery (all scenes with lighting changes)

<table>
<thead>
<tr>
<th>KITTI image sequence:</th>
<th>#11</th>
<th>#15</th>
<th>#44</th>
<th>#74</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zimmer et al. 2011</td>
<td>37.3</td>
<td>32.3</td>
<td>23.2</td>
<td>62.9</td>
<td>38.9</td>
</tr>
<tr>
<td>Bruhn/Weickert 2005</td>
<td>33.9</td>
<td>47.7</td>
<td>32.4</td>
<td>71.4</td>
<td>46.7</td>
</tr>
<tr>
<td>Census Transform</td>
<td>36.5</td>
<td>28.6</td>
<td>28.5</td>
<td>63.8</td>
<td>39.4</td>
</tr>
<tr>
<td>Complete Rank Transform</td>
<td><strong>29.8</strong></td>
<td><strong>22.8</strong></td>
<td><strong>22.6</strong></td>
<td><strong>61.5</strong></td>
<td><strong>34.2</strong></td>
</tr>
</tbody>
</table>

[\%] bad pixel error measure bp3

→ CRT outperforms established methods in difficult lighting conditions

- Whole KITTI benchmark (195 testing image sequences)
  - Our method ranks 10th of 33
  - Top 4 methods use stereo information
  - Census-based method by Ranftl et al. (2012) ranks 14th

→ CRT carries enough image information for real-world applications
Experiments

Experiments - Middlebury Benchmark

Middlebury Training Set

\[
\begin{array}{l|ccccccccc}
\text{rw} & \text{dimetr.} & \text{grove2} & \text{grove3} & \text{hydr.} & \text{urban2} & \text{urban3} & \text{yos} & \text{avg} \\
\hline
\text{RT} & .111 & .092 & .191 & .764 & .191 & .457 & 1.03 & .211 & .381 \\
\text{CT} & .102 & .090 & .169 & .646 & .147 & .378 & .819 & .169 & .316 \\
\text{CRT} & .100 & .076 & .154 & .585 & .138 & .324 & .529 & .150 & .260 \\
\end{array}
\]

[pixel] average endpoint error

→ CRT clearly preferable over RT and CT

Middlebury Benchmark (July 31st 2013)

<table>
<thead>
<tr>
<th>Method</th>
<th>Average rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zimmer et al. (2009)</td>
<td>40.5</td>
</tr>
<tr>
<td>ours</td>
<td>45.8</td>
</tr>
<tr>
<td>Brox et al. (2004)</td>
<td>52.1</td>
</tr>
</tbody>
</table>

→ even without illumination changes, CRT gives acceptable quality
Conclusions

- Morphological invariance handles illumination changes
- Problem: Invariance discards information
- Our solution: CRT that carries as much local image information as possible
- Intentionally kept variational optic flow model simple
- Our proposed CRT clearly preferable over Census and Rank transforms
Conclusions

- Morphological invariance handles illumination changes
- Problem: Invariance discards information
- Our solution: CRT that carries as much local image information as possible
- Intentionally kept variational optic flow model simple
- Our proposed CRT clearly preferable over Census and Rank transforms

Thank You!