Data-driven characterization of pedestrian flows

Marija Nikolić, Michel Bierlaire

September 14, 2016
Outline

1. Introduction
2. Related research
3. Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances
4. Empirical analysis
5. Conclusion and future work
Outline

1. Introduction
2. Related research
3. Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances
4. Empirical analysis
5. Conclusion and future work
Motivation
Background

Importance

• Understanding, reproducing and forecasting phenomena that characterize pedestrian traffic is necessary in order to provide services related to pedestrian safety and convenience.

Indicators

• Density $k$ \( (\text{ped}/m^2) \), speed $v$ \( (m/s) \) and flow $q$ \( (\text{ped}/ms) \)
• Used to observe and to model the flows of pedestrians
• Little concern is dedicated to the nature of spatial and temporal discretization underlying the definitions
Outline

1 Introduction

2 Related research

3 Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances

4 Empirical analysis

5 Conclusion and future work
Methods

[Duives et al., 2015], [Helbing et al., 2007], [Steffen and Seyfried, 2010], [Saberi and Mahmassani, 2014], [van Wageningen-Kessels et al., 2014]
Methods

[Duives et al., 2015], [Helbing et al., 2007 ], [Steffen and Seyfried, 2010], [Saberi and Mahmassani, 2014],[van Wageningen-Kessels et al., 2014]
# Characteristics of methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale</th>
<th>Spatial aggregation</th>
<th>Temporal aggregation</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Area</td>
<td>Shape Size Location</td>
<td>Interval</td>
</tr>
<tr>
<td>XY-T</td>
<td>Macroscopic</td>
<td></td>
<td></td>
<td>Duration</td>
</tr>
<tr>
<td>Grid-based (GB)</td>
<td>Macroscopic</td>
<td>Cell</td>
<td>Size Location</td>
<td>Interval</td>
</tr>
<tr>
<td>Range-based (RB)</td>
<td>Macroscopic</td>
<td>Circle</td>
<td>Radius Location</td>
<td>Interval</td>
</tr>
<tr>
<td>Exponentially-weighted (EW)</td>
<td>Macroscopic</td>
<td>Range</td>
<td>Influence function</td>
<td>Interval</td>
</tr>
<tr>
<td>Voronoi-based (VB)</td>
<td>Microscopic</td>
<td>Voronoi cell</td>
<td>Boundary conditions</td>
<td>Interval</td>
</tr>
</tbody>
</table>
How to define the discretization...

...independent of arbitrary chosen values?
Outline

1 Introduction

2 Related research

3 Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances

4 Empirical analysis

5 Conclusion and future work
Data-driven approach

Keep calm and let data speak!
Preliminaries

- A space-time representation: $\Omega \subset \mathbb{R}^3$
- The distance along each of the two spatial axes is expressed in meters, and the unit for time is seconds
- $p = (p_x, p_y, p_t) = (x, y, t) \in \Omega$ represents a physical position $(x, y)$ in space at a specific time $t$
- Assumption: $\Omega$ is convex (obstacle-free and bounded)
- Generator set $\Gamma$: pedestrian trajectories

$$\Gamma_i : \{p_i(t)|p_i(t) = (x_i(t), y_i(t), t)\}$$

$$\Gamma_i : \{p_is|p_is = (x_is, y_is, t_s)\}, t_s = [t_0, t_1, ..., t_f]$$
Outline

1. Introduction

2. Related research

3. Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances

4. Empirical analysis

5. Conclusion and future work
Data-driven discretization framework

- 3D Voronoi diagrams associated with pedestrian trajectories
- Partitioning: the assignment of each point \( p \in \Omega \) to one generator from \( \Gamma \)

\[
\delta_{\Gamma}(p, \Gamma_i) = \begin{cases} 
1, & D(p, \Gamma_i) \leq D(p, \Gamma_j), \forall j \neq i \\
0, & \text{otherwise}
\end{cases}
\]

\[
D(p, \Gamma_i) = \min_{p_i} \{ d(p, p_i) | p_i \in \Gamma_i, \Gamma_i \in \Gamma, p \in \Omega \}
\]

- Discretization units: the set of points \( p \) assigned to the same generator

\[
V_i = \{ p | \delta_{\Gamma}(p, \Gamma_i) = 1, p \in \Omega, \Gamma_i \in \Gamma \}
\]
The plane through the point \( p_0 = (x_0, y_0, t_0) \) and with non-zero normal vector \( \vec{n} = (a, b, c) \)

\[
\mathcal{P}_{\vec{n}, p_0} : ax + by + ct + d = 0,
\]

where \( d = -ax_0 - by_0 - ct_0 \)

The intersection of \( V_i \) and the plane \( \mathcal{P}_{\vec{n}, p_0} \)

\[
A(V_i, \mathcal{P}_{\vec{n}, p_0}) = \{ p | p \in \{ V_i \cap \mathcal{P}_{\vec{n}, p_0} \} \} \]
Data-driven discretization framework (cont.)

\[ A(V_i, P_{(0,0,1),p_0}) = \{ p | p \in V_i \text{ and } p_t = t_0 \} \]

\[ A(V_i, P_{(a,b,0),p_0}) = \{ p | p \in V_i \text{ and } ap_x + bp_y = ax_0 + by_0 \} \]
Outline

1. Introduction

2. Related research

3. Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances

4. Empirical analysis

5. Conclusion and future work
Voronoï-based traffic indicators

**Density indicator**

\[ k(x, y, t) = \frac{1}{|A(V_i, P_{(0,0,1)},(x,y,t))|} \]

**Flow indicator**

\[ \bar{q}_{(a,b,0)}(x, y, t) = \frac{1}{|A(V_i, P_{(a,b,0)},(x,y,t))|} \]

**Velocity indicator**

\[ \bar{v}_{(a,b,0)}(x, y, t) = \frac{\bar{q}_{e}(x,y,t)}{k(x,y,t)} = \frac{|A(V_i, P_{(0,0,1)},(x,y,t))|}{|A(V_i, P_{(a,b,0)},(x,y,t))|} \]
Outline

1. Introduction
2. Related research
3. Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances
4. Empirical analysis
5. Conclusion and future work
Spatio-temporal distances

Spatial Euclidean distance (E-3D Voro)

\[ d_E(p, p_i) = \begin{cases} \sqrt{(p - p_i)^T(p - p_i)}, & \Delta t = 0 \\ \infty, & \text{otherwise} \end{cases} \]

Time-Transform distances (TT\{1,2,3\}-3D Voro)

\[ d_{TT_1}(p, p_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2 + \alpha^2(t - t_i)^2} \]
\[ d_{TT_2}(p, p_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2 + \alpha_i(t_i)(t - t_i)} \]
\[ d_{TT_3}(p, p_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2 + \alpha_i^2(t_i)(t - t_i)^2} \]

\(\alpha\) and \(\alpha_i\) - conversion constants expressed in meters per second
Spatio-temporal distances

Predictive distance (P-3D Voro)

\[ d_P(p, p_i) = \begin{cases} \sqrt{(x_i(t) - x)^2 + (y_i(t) - y)^2}, & t - t_i \geq 0 \\ \infty, & \text{otherwise} \end{cases} \]

The anticipated position of pedestrian \( i \) at time \( t \):
\[ x_i(t) = x_i(t_i) + (t - t_i)v_i^x(t_i), \quad y_i(t) = y_i(t_i) + (t - t_i)v_i^y(t_i) \]

The speed of pedestrian \( i \) at \( t_i \) in \( x \) and \( y \) directions: \( v_i^x(t_i), v_i^y(t_i) \)

Mahalanobis distance (M-3D Voro)

\[ d_M(p, p_i) = \sqrt{(p - p_i)^T M_i (p - p_i)} \]

\( M_i \) - symmetric, positive-definite matrix that defines how distances are measured from the perspective of pedestrian \( i \)
Outline

1. Introduction
2. Related research
3. Methodology
   - Discretization framework
   - Definitions of the indicators
   - Spatio-temporal distances
4. Empirical analysis
5. Conclusion and future work
Performance of the approach

Synthetic data - unidirectional flow

NOMAD simulation tool [Campanella, 2010]

Scenario I: low congestion, homogenous population

Scenario II: high congestion, heterogeneous population

Indicators

Robustness w.r.t. the simulation noise

Robustness w.r.t. the sampling frequency
Characterization based on trajectories

Robustness with respect to the simulation noise

- 100 sets of pedestrian trajectories synthesized per scenario
- $\theta_r^M(p) = (k_r^M(p), v_r^M(p), q_r^M(p))$ - a vector of indicators at point $p$ obtained by applying the method $M$ to the $r^{th}$ set of trajectories
- The standard deviation of the indicators at $p$ as

$$\sigma_R^M(p) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} (\theta_r^M(p) - \mu_R^M(p))^2}$$

$$\mu_R^M(p) = \frac{1}{R} \sum_{r=1}^{R} \theta_r^M(p)$$ and $R = 100$
Robustness with respect to the simulation noise

Standard deviation (1000 points) - Scenario I
Robustness with respect to the simulation noise

Standard deviation (1000 points) - Scenario II
Characterization based on sampled data

Robustness with respect to the sampling frequency

- Ability of tolerating missing data
- Benchmark: indicators calculated on the true synthetic trajectories
- Sampled data: different sampling frequencies ($3s^{-1} - 0.5s^{-1}$)
- Indicators calculated via
  1. 3D Voro applied to the interpolated trajectories
  2. 3D Voro applied directly to the samples
- Comparison of the indicators to the corresponding benchmark values at 1000 randomly selected points
Robustness w.r.t the sampling frequency - Scenario I

High sampling frequency: $3s^{-1}$

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Mode</th>
<th>Median</th>
<th>90% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IT</td>
<td>SoP</td>
<td>IT</td>
<td>SoP</td>
</tr>
<tr>
<td>XY-T</td>
<td>1.47E-02 / 1.25E-02</td>
<td>1.25E-02 / 6.25E-02</td>
<td>1.25E-02 / 6.25E-02</td>
<td>1.25E-02 / 6.25E-02</td>
</tr>
<tr>
<td>E-3DVoro</td>
<td>1.17E-02 / 0</td>
<td>0 / 4.48E-04</td>
<td>0 / 3.96E-02</td>
<td>0 / 3.96E-02</td>
</tr>
<tr>
<td>TT1-3DVoro</td>
<td>2.70E-03 / 6.70E-03</td>
<td>0 / 3.00E-04</td>
<td>2.30E-03 / 7.30E-03</td>
<td>1.02E-02 / 7.30E-03</td>
</tr>
<tr>
<td>TT2-3DVoro</td>
<td>5.80E-03 / 3.50E-02</td>
<td>0 / 6.00E-04</td>
<td>2.08E-02 / 1.50E-02</td>
<td>6.69E-02 / 1.50E-02</td>
</tr>
<tr>
<td>TT3-3DVoro</td>
<td>5.40E-03 / 4.34E-02</td>
<td>0 / 6.00E-04</td>
<td>2.83E-02 / 1.32E-02</td>
<td>9.22E-02 / 1.32E-02</td>
</tr>
<tr>
<td>P-3DVoro</td>
<td>8.20E-03 / 5.36E-02</td>
<td>0 / 2.40E-03</td>
<td>3.03E-02 / 1.30E-02</td>
<td>1.14E-01 / 1.30E-02</td>
</tr>
<tr>
<td>M-3DVoro</td>
<td>4.50E-03 / 5.65E-02</td>
<td>0 / 1.10E-03</td>
<td>4.55E-02 / 1.28E-02</td>
<td>1.04E-01 / 1.28E-02</td>
</tr>
</tbody>
</table>

Low sampling frequency: $0.5s^{-1}$

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Mode</th>
<th>Median</th>
<th>90% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IT</td>
<td>SoP</td>
<td>IT</td>
<td>SoP</td>
</tr>
<tr>
<td>XY-T</td>
<td>1.90E-01 / 1.00E-01</td>
<td>1.00E-01 / 3.38E-01</td>
<td>1.50E-01 / 3.38E-01</td>
<td>1.50E-01 / 3.38E-01</td>
</tr>
<tr>
<td>E-3DVoro</td>
<td>1.64E-01 / 1.12E-02</td>
<td>1.12E-02 / 3.02E-01</td>
<td>1.46E-01 / 3.02E-01</td>
<td>1.46E-01 / 3.02E-01</td>
</tr>
<tr>
<td>TT1-3DVoro</td>
<td>2.54E-01 / 1.27E-01</td>
<td>1.35E-02 / 8.97E-02</td>
<td>9.00E-03 / 3.41E-01</td>
<td>8.97E-02 / 3.41E-01</td>
</tr>
<tr>
<td>TT2-3DVoro</td>
<td>1.64E-01 / 1.12E-01</td>
<td>1.44E-02 / 3.52E-01</td>
<td>1.06E-02 / 3.52E-01</td>
<td>1.06E-02 / 3.52E-01</td>
</tr>
<tr>
<td>TT3-3DVoro</td>
<td>1.89E-01 / 1.24E-01</td>
<td>1.84E-02 / 3.40E-01</td>
<td>1.09E-02 / 3.40E-01</td>
<td>1.09E-02 / 3.40E-01</td>
</tr>
<tr>
<td>P-3DVoro</td>
<td>3.19E-01 / 1.21E-01</td>
<td>3.26E-02 / 3.36E-01</td>
<td>6.20E-03 / 2.10E-01</td>
<td>6.20E-03 / 2.10E-01</td>
</tr>
<tr>
<td>M-3DVoro</td>
<td>1.97E-01 / 1.24E-01</td>
<td>3.48E-02 / 3.21E-01</td>
<td>9.90E-03 / 2.31E-01</td>
<td>9.90E-03 / 2.31E-01</td>
</tr>
</tbody>
</table>
Robustness w.r.t the sampling frequency - Scenario II

High sampling frequency: $3s^{-1}$

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Mode</th>
<th>Median</th>
<th>90% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IT</td>
<td>SoP</td>
<td>IT</td>
<td>SoP</td>
</tr>
<tr>
<td>XY-T</td>
<td>2.05E-02</td>
<td>0</td>
<td>1.25E-02</td>
<td>5.00E-02</td>
</tr>
<tr>
<td>E-3DVoro</td>
<td>1.43E-02</td>
<td>0</td>
<td>2.67E-02</td>
<td>2.64E-02</td>
</tr>
<tr>
<td>$TT_1$-3DVoro</td>
<td>8.00E-03</td>
<td>4.55E-02</td>
<td>1.75E-02</td>
<td>2.36E-02</td>
</tr>
<tr>
<td>$TT_2$-3DVoro</td>
<td>1.49E-02</td>
<td>1.07E-01</td>
<td>5.72E-02</td>
<td>3.33E-02</td>
</tr>
<tr>
<td>$TT_3$-3DVoro</td>
<td>1.24E-02</td>
<td>1.60E-01</td>
<td>3.50E-03</td>
<td>2.98E-02</td>
</tr>
<tr>
<td>P-3DVoro</td>
<td>2.10E-02</td>
<td>1.66E-01</td>
<td>4.20E-03</td>
<td>5.27E-02</td>
</tr>
<tr>
<td>M-3DVoro</td>
<td>1.31E-02</td>
<td>2.40E-01</td>
<td>2.50E-03</td>
<td>2.91E-02</td>
</tr>
</tbody>
</table>

Low sampling frequency: $0.5s^{-1}$

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Mode</th>
<th>Median</th>
<th>90% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IT</td>
<td>SoP</td>
<td>IT</td>
<td>SoP</td>
</tr>
<tr>
<td>XY-T</td>
<td>5.29E-01</td>
<td>1.63E-01</td>
<td>4.75E-01</td>
<td>1.01E00</td>
</tr>
<tr>
<td>E-3DVoro</td>
<td>4.02E-01</td>
<td>0</td>
<td>2.49E-01</td>
<td>1.03E+00</td>
</tr>
<tr>
<td>$TT_1$-3DVoro</td>
<td>4.06E-01</td>
<td>2.90E-01</td>
<td>3.10E-01</td>
<td>2.64E-01</td>
</tr>
<tr>
<td>$TT_2$-3DVoro</td>
<td>3.92E-01</td>
<td>4.58E-01</td>
<td>2.85E-01</td>
<td>2.34E-01</td>
</tr>
<tr>
<td>$TT_3$-3DVoro</td>
<td>4.41E-01</td>
<td>5.07E-01</td>
<td>2.89E-01</td>
<td>5.89E-02</td>
</tr>
<tr>
<td>P-3DVoro</td>
<td>4.31E-01</td>
<td>3.71E-01</td>
<td>2.58E-01</td>
<td>9.43E-01</td>
</tr>
<tr>
<td>M-3DVoro</td>
<td>4.34E-01</td>
<td>5.01E-01</td>
<td>2.75E-01</td>
<td>9.96E-01</td>
</tr>
</tbody>
</table>
Outline

1 Introduction

2 Related research

3 Methodology
   • Discretization framework
   • Definitions of the indicators
   • Spatio-temporal distances

4 Empirical analysis

5 Conclusion and future work
Conclusion

• A novel approach to pedestrian traffic characterization: data-driven discretization via 3D Voronoi diagrams
• Discretization based on trajectories available either in the form of an analytical description or as a finite collection of points
• The exact characterization of the Voronoi diagrams can be adapted to specific situation
• Superior to existing methods w.r.t. robustness to the simulation noise
• Robustness to the sampling frequency
  – Higher sampling frequency: 3DVoro based on interpolated trajectories shows better results (Time-Transform 3D Voronoi)
  – Lower sampling frequency: 3DVoro based on sample of points exhibit better performance (anticipating distances)
Future work

- Analysis of the performance for other behavioral situations (bi-directional and multi-directional scenarios)
- The effectiveness of the approach using real data (railway station, Lausanne)
- Characterization in the presence of obstacles
- Weighted assignment rules to account for the anisotropy of pedestrian movements
Thank you

hEART 2016 - 5th Symposium of the European Association for Research in Transportation, Delft University of Technology:  
Data-driven characterization of pedestrian traffic  
Marija Nikolić, Michel Bierlaire

Help by S. S. Azadeh and F. Hänsele is appreciated.

- marija.nikolic@epfl.ch