Gender profiling of pedestrian dyads*

Zeynep Yücel[1,2] and Francesco Zanlungo[2] and Takayuki Kanda[2,3]

[1] Okayama University, Okayama, Japan, zeynep@okayama-u.ac.jp
[2] ATR, Kyoto, Japan zanlungo@atr.jp
[3] Kyoto University, Kyoto, Japan kanda@i.kyoto-u.ac.jp

Abstract. In traffic safety community, behavioral differences between genders have been attracting considerable attention in recent decades. Various empirical studies have proven that gender has a significant relation to drivers', cyclists' or pedestrians' decision making, route choice, rule compliance, as well as risk taking/perception. However, most studies examine behavior of individuals, and only very few consider (pedestrian) groups with different gender profiles. Therefore, this study investigates effect of gender composition of pedestrian dyads on the tangible dynamics, which may potentially help in automatically understanding and interpreting higher level behaviors such as decision making. We first propose a set of variables to represent dyads's physical/dynamical state. Observing empirical distributions, we comment on the effect of gender interplay on locomotion preferences. In order to verify our inferences quantitatively, we propose a gender profile recognition algorithm. Removing one variable at a time, contribution of each variable to recognition is evaluated. Our findings indicate that height related variables have a more strict relation to gender, followed by group velocity and inter-personal distance. Moreover, the "male" effect on dyad motion is found to somehow diminish when the male is paired with a female.

Keywords: Social groups, behavioral variation, intrinsic factors

1 Introduction and related work

A deep understanding of pedestrians' movement patterns is crucial for various applications including designing of urban spaces [1], planning of transportation facilities, building more realistic crowd simulation systems [2], improving traffic safety [3] and integrating artificial intelligence in smart vehicles etc. [4, 5]. To achieve this understanding, it is important to analyze the significant elements that act on pedestrian locomotion, which can be categorized roughly into two as extrinsic and intrinsic factors.

Extrinsic factors involve location, built environment, time of the day and like [6], whereas intrinsic factors appertain to age, gender, level of mobility [7] etc. Obviously, an interplay of the extrinsic and intrinsic factors shape the behavior of all constituents of traffic (e.g. drivers or cyclists), but here we contain ourselves to only one of the intrinsic factors, gender, and only of the constituents of traffic, pedestrians [8, 9]. Therefore, this study focuses particularly on the effect of gender on pedestrian motion, which is shown to be an important element in decision making, rule compliance and risk perception/taking [10–12]. Although the effect of gender on individual's motion is investigated

*This work was supported by JSPS KAKENHI Grant Number JP18K18168.

in many studies, gender composition of pedestrian groups is not yet treated in detail apart from very few works [13, 1]. Therefore, this study aims to fill the void in literature by addressing diversity of locomotion patterns of pedestrian groups with various gender profiles (i.e. gender composition).

To that end, we focus initially on dyads (i.e. groups of two people), which are straightforward to categorize in terms of gender composition, as all-female, all-male or mixed gender. We propose a few variables derived from trajectory and physical features, which are first evaluated qualitatively concerning their relation to gender profile. Subsequently, we propose a method to automatically recognize gender composition based on the approach of [14]. In addition, by removing one variable at a time, we point out to the contribution of each variable to recognition of gender composition.

2 Data set and variables

We used the data set introduced by [15], which is collected in the atrium of a business center in Osaka, Japan. The atrium covers roughly 900 m^2; connects the business center to a ferry terminal, a train station and a shopping mall; and thus is visited by a diverse demographic profile of pedestrians. Specifically, the data is recorded on two week days and one weekend over a one year time period. The recordings involve readings of 3D range sensors and video footage. The range sensors are densely populated so as to track pedestrians over their entire trajectory [16]. In addition to the trajectory, thanks to their overhead configuration, depth sensors provide estimations of pedestrian height as well. By watching the video footage, a set of coders annotate groups (i.e. pedestrians constituting a social group), and gender composition of those groups (i.e. all-female F, all-male M or mixed gender X). In this manner, the distribution of dyad gender profiles in our data set is found to be as follows: 227 F, 424 M, 311 X[4].

It is known that individuals have significantly different velocities depending on their gender and age [17–19]. In addition, studies on pedestrian dyads and triads show that all-male, all-female and mixed gender dyads present different velocity patterns [13, 1]. Using the data set and the associated ground truth, we examine such motion attributes, in addition to a set of features relating the height of peers.

Inter-personal distance δ relating various gender profiles is found as in Fig. 1-(a), which ascertains that M dyads have a significantly larger inter-personal distance, whereas F and X dyads stay in closer proximity. Nevertheless, the tails of δ distributions indicate that X dyads may have a heavier tail beyond 1 m than F dyads.

Relating velocity, we consider two variables, namely, group velocity $\bar{v}$ and velocity difference ω between peers. Group velocity $\bar{v}$ is described as the average velocity of the peers and is found as in Fig. 1-(b). It is clear that the typical attribute of -individual-males to have a higher velocity than females is sustained, as long as they move as part of an all-male dyad, whereas it diminishes as they move as part of a mixed gender dyad. In that case, the dyad adapts to the velocity of the slower peer, which is the female. Therefore, the velocity distribution of F and X dyads do not present any significant difference, whereas the velocity distribution of M dyads is in line with the velocity pattern

[4] A detailed analysis on inter-rater agreement confirms that the coders have significant agreement rates [13].

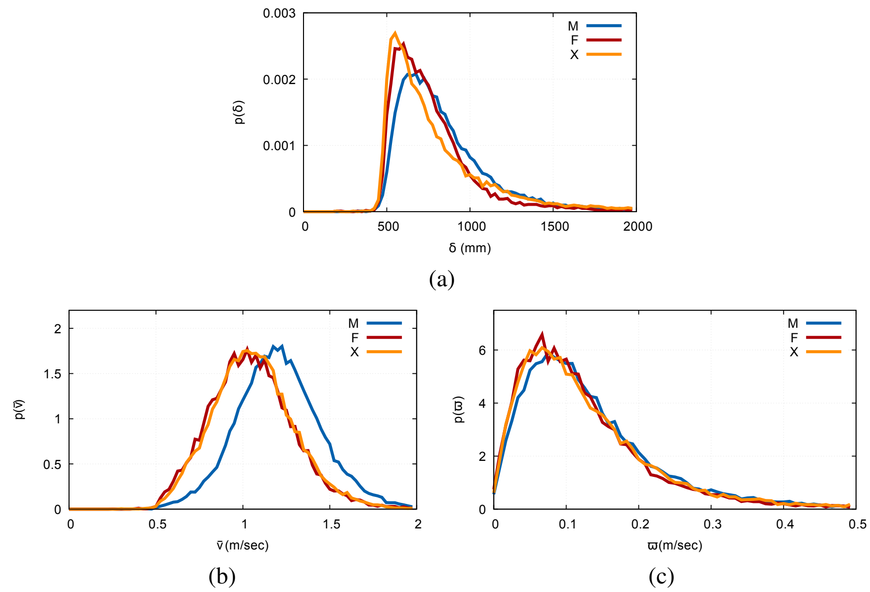


Fig. 1. (a) Interpersonal distance, (b) group velocity and (c) velocity difference.

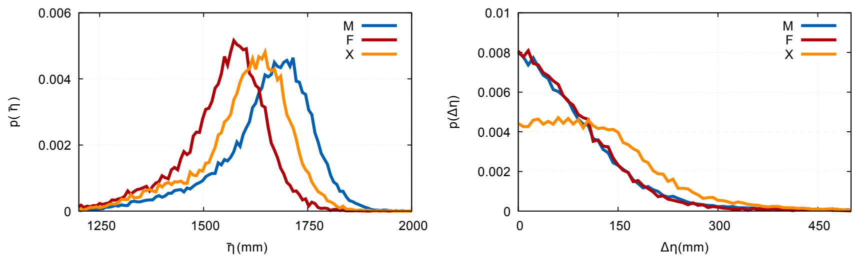
of individual males. Velocity difference ω is defined as the norm of the difference vector between peers' velocities. It can be seen in Fig. 1-(c) that M dyads have a slightly higher velocity difference, whereas F and X dyads have virtually the same ω distribution.

Average height of peers denoted with $\bar{\eta}$ is found as in Fig. 2-(a)[5]. As expected M dyads are significantly taller than F and X, where F dyads have the shortest average height. Although locations of the peaks may depend on the geographical location (thus, indirectly on genetic factors), we consider these findings to be persistent among countries. Despite the dependency of height on demographic factors, the difference of height, denoted by $\Delta\eta$, between males and females in different countries is found to be rather similar [20, 21], which is supported by Fig. 2-(b). Namely, height difference between two males and two females follow very similar distributions around 0, whereas $\Delta\eta$ concerning X has a distribution with a much larger deviation and a heavier tail.

3 Automatic recognition of gender profiles

Inspired by the Bayesian approach of [14], we propose a method to assess the probability at each time instant that a dyad belongs to one of the three gender compositions M, F, or X. Suppose that at each time instant we collect an observation set Σ composed of the five variables listed in Section 2, $\Sigma = \{\delta,\ \bar{v},\ \omega,\ \bar{\eta},\ \Delta\eta\}$. Let $g \in \{M,\ F,\ X\}$ denote the gender composition. Given Σ, the (posterior) probability at time t that a dyad

[5] Since height depends to a great extent on age particularly until the end of adolescence, we consider pedestrians whose apparent age are labeled to be more than 20 years old.



(a) (b)

Fig. 2. (a) Average height and (b) height difference for dyads of various gender composition.

belongs to a gender composition g, is updated in the Bayesian sense,

$$P_t(g|\Sigma) = \frac{P_t(\Sigma|g)P_t(g)}{P_t(\Sigma)}. \tag{1}$$

Although probability density function (pdf) of Σ can be computed in a five dimensional space, it is very hard to overcome sparsity. Thus, by verifying that any two pair of variables can be considered to be independent [14], we break down $P_t(\Sigma|g)$ into a product of individual pdfs.

In order to avoid any bias, we consider equal probabilities for the priors, $P_0(g)$. Furthermore, so as to avoid computing the exact value of the marginal likelihood $P_t(\Sigma)$, we make use of the fact that the sum of $P_t(g)$ for all g must be 1. In order to propagate the information from past observations, while also preventing any drift due to an accumulation of error; we update the prior at every instant as a weighted average of the first value of the prior $P_0(g)$ and the previous value of the posterior $P_{t-1}(g)$. In this manner, we obtain a probability of belonging to any one of the gender combinations at every instant.

Subsequently, probability of belonging to a particular gender profile is computed as the average of concerning instantaneous probabilities. Suppose that the "ground truth" gender profile assigned to a dyad, d, is represented with r_d^{GT}. Probability of belonging to g can be defined as the average of all posterior probabilities along its trajectory,

$$\overline{P_d}(r) = \frac{\sum_t P_t(r|\Sigma)}{n_d}, \tag{2}$$

where n_d is the number of time instants (i.e. number of trajectory data points). We then define the estimated gender composition as the one with the maximum probability[6]

$$r_d^p = \operatorname*{argmax}_r(\overline{P}_d(r)). \tag{3}$$

In this sense, a single estimation result of gender composition is yielded for each dyad, which implies that the length of their trajectory is not paid regard.

[6]If the maximum occurs multiple times, one of the relating gender profiles is picked randomly.

4 Results and discussion

First, we run the method proposed in Section 3 using all five variables depicted in Figures 1 and 2. Subsequently, we remove one variable at a time and repeat the same procedure for judging the contribution of each variable[7]. The performance rates are found as in Table 1. Using the set of all five variables yields the best performance. Moreover, removal of δ degrades particularly the detection of X dyads. Although the location of the peak regarding δ distribution of X dyads is positioned not far from those of M and F dyads, its lighter tail and smaller variation help distinguishing of X from other gender profiles, in particular from M dyads. Removal of $\bar{v}$ has roughly the same effect on M and X dyads, whereas it slightly works in favor of F dyads. From Fig. 1-(b) $\bar{v}$ of M dyads distinguishes from that of F and X. However, since M dyads have other prominent characteristics, removal of $\bar{v}$ results in a limited degradation in identification of M dyads. Additionally, removal of ω has the least effect on overall performance.

On the other hand, variables relating height have obviously a larger impact on gender profiling. Among the five variables, average height $\bar{\eta}$ emerges as the most influential one. Although height (and in turn the overall body volume) has a direct impact on how large is the step size -a plain rationale of velocity- and how far people position themselves from others (i.e. δ), it is certainly much more self-evident to observe directly the height of individuals than its indirect implications on δ and $\bar{v}$. Considering also that the behavior of X dyads converges to that of F dyads in terms of particularly $\bar{v}$ and ω but also δ, we can say that “male” effect on velocity and inter-personal distance is somehow diminished when a male is paired with a female. Interestingly, this does not reflect as a degradation in the identification of M dyads but F dyads, which is probably due to the fact there are almost twice as many M samples in our dataset (see Section 2) as F dyads. In addition, height difference $\Delta\eta$ is another influential variable, yet not as much as $\bar{\eta}$. As expected, removal of $\Delta\eta$ lowers the detection rates of particularly X dyads to a large extent. In addition, in Table 1, diagonals have almost always the largest entry, i.e. we most often correctly identify the gender composition as the true class.

5 Conclusion

This study makes an effort to demonstrate the effect gender at a granular scale by presenting the locomotion distinctions regarding 5 variables, which reflect tangible and quantified dynamics/state of dyads. Moreover, we propose a method to rate the impact of gender profile on each of those, through an automatic recognition method for gender profile, which to the best of our knowledge is the first method for automatic recognition of gender composition of social groups. By feeding the recognition algorithm a different set of inputs -a set which lacks exactly one variable out of the five- we demonstrate that average height is the feature that reflects the effect of gender profile the most, which is followed by height difference. Inter-personal distance and group velocity have yet a limited effect, though they contribute recognition of different gender compositions.

[7] We consider 30% of the data set as training set and the remaining 70% as test set; and repeat this procedure 20 times.

Table 1. Estimation results based on (a) Σ, (b) $\Sigma \setminus \delta$, (c) $\Sigma \setminus \bar{v}$, (d) $\Sigma \setminus \omega$, (e) $\Sigma \setminus \bar{\eta}$ and (f) $\Sigma \setminus \Delta\eta$.

(a)

	M	F	X
M	83.50	8.64	7.86
F	10.97	76.01	13.02
X	23.51	28.58	47.91
Tot		70.22	

(b)

	M	F	X
M	82.38	9.87	7.76
F	11.78	73.10	15.12
X	28.61	31.61	39.78
Tot		66.19	

(c)

	M	F	X
M	79.14	11.69	9.17
F	9.97	78.19	11.84
X	26.19	32.03	41.78
Tot		66.64	

(d)

	M	F	X
M	82.84	8.62	8.54
F	12.73	73.07	14.20
X	26.48	28.41	45.11
Tot		68.14	

(e)

	M	F	X
M	74.40	15.03	10.56
F	41.63	36.66	21.72
X	29.87	25.26	44.87
Tot		55.80	

(f)

	M	F	X
M	82.94	8.16	8.90
F	15.74	70.15	14.11
X	32.20	33.61	34.19
Tot		63.91	

References

1. A. Willis, N. Gjersoe, C. Havard, J. Kerridge, R. Kukla, Environ. Plan. B **31**(6), 805 (2004)
2. C. Holland, R. Hill, Accid Anal Prev **42**(4), 1097 (2010)
3. M.A. Granié, Accid Anal Prev **42**(2), 726 (2010)
4. I.Y. Noy, D. Shinar, W.J. Horrey, Safety science **102**, 68 (2018)
5. J. Son, M. Park, B.B. Park, Transp Res Part F Traffic Psychol Behav **31**, 12 (2015)
6. P. Stoker, A. Garfinkel-Castro, M. Khayesi, W. Odero, M.N. Mwangi, M. Peden, R. Ewing, Journal of Planning Literature **30**(4), 377 (2015)
7. E. Papadimitriou, S. Lassarre, G. Yannis, Transp. Res. Part F **36**, 69 (2016)
8. K.J. Anstey, R. Eramudugolla, L.A. Ross, N.T. Lautenschlager, J. Wood, International psychogeriatrics **28**(3), 349 (2016)
9. I.M. Bernhoft, G. Carstensen, Transp Res Part F Traffic Psychol Behav **11**(2), 83 (2008)
10. M.A. Granié, Transp Res Part F Traffic Psychol Behav **10**(5), 371 (2007)
11. M.A. Granié, Safety science **47**(9), 1277 (2009)
12. A. Tom, M.A. Granié, Accid Anal Prev **43**(5), 1794 (2011)
13. F. Zanlungo, Z. Yücel, D. Bršćić, T. Kanda, N. Hagita, PLoS one **12**(11), e0187253 (2017)
14. Z. Yucel, F. Zanlungo, C. Feliciani, A. Gregorj, T. Kanda, to appear in Pedestrian and Evacuation Dynamics (2018)
15. F. Zanlungo, D. Bršćić, T. Kanda, Physical Review E **91**(6), 062810 (2015)
16. D. Bršcic, T. Kanda, T. Ikeda, T. Miyashita, IEEE Tran. Human-Machine Systems **43**(6), 522 (2013)
17. R. Lobjois, N. Benguigui, V. Cavallo, Accid Anal Prev **53**, 166 (2013)
18. A. Dommes, V. Cavallo, J.B. Dubuisson, I. Tournier, F. Vienne, J. Saf. Res. **50**, 27 (2014)
19. N.N. Ferenchak, J. Traffic Transp. Eng. **3**(4), 345 (2016)
20. MEXT. Official Health Statistics Survey. http://www.mext.go.jp/component/b_menu/other/__icsFiles/afieldfile/2014/03/28/1345147_1.pdf. Accessed: 01-2019
21. OECD iLibrary. OECD social indicators. https://www.oecd-ilibrary.org/social-issues-migration-health/society-at-a-glance-2009/height_soc_glance-2008-26-en. Accessed: 01-2019