Visual Social Distancing and human proxemics

Pietro Morerio, Gian Luca Bailo, Matteo Bustreo, Alessio Del Bue
Patterna Analysis and Computer Vision (PAVIS) - Istituto Italiano di Tecnologia
{pietro.morerio, gian.bailo, matteo.bustreo, alessio.delbue}@iit.it

Abstract

The ongoing pandemics has raised interest in effective measures to contain viral outbreak, Social Distancing (SD) being one of the most effective. To comply with this constraint, governments are adopting restrictions over the minimum inter-personal distance between people along with the maximum room allowance in term of number of people.

Given this scenario, it is crucial to massively measure the compliance to physical constraint especially in indoor environments, in order to figure out the reasons of the possible breaks of such distance limitations, understand if this implies a potential threat and act consequently. To this end, we have recently investigated the Visual Social Distancing (VSD) problem, defined as the automatic estimation of the inter-personal distance from an image, and the characterization of people aggregation and proxemics. VSD is pivotal for a non-invasive analysis to whether people comply with the SD restriction, and to provide statistics. We discuss two use-cases and their challenges.

1 Introduction

Since when proxemics was introduced by Hall in 1966 [Hall, 1966] as a key observable clue to decode human behaviours, it has become an important proxy to understand living spaces resilience, especially in light of the recent severe worldwide pandemic outbreaks [Hynes *et al.*, 2020]. Visual assessment of the interpersonal distances with ambient cameras has became a preferred option thanks to its non-invasive nature that allows the analysis on human behaviours at distance.

The use of proxemics in general scenarios is restricted by the fact that it is hard to extract robust metric references to measure distance [Criminisi *et al.*, 2000] and most of the installed camera networks are not calibrated and have difficult to access intrinsic parameters.

In this abstract we investigate approaches to estimate interpersonal distances from a single uncalibrated image, as shown in Fig. 1. We propose to exploit the rich structural information encoded in the body joints, whose length and relative position can be used to train an end-to-end network to infer the metric distance among any arbitrary pair of people captured by any uncalibrated camera. We show that the 2D joints do contain enough information for a network to learn and solve such task, despite its intrinsic ambiguity. A key property of VSD algorithms in the wild should be their ability to generalise to novel camera viewpoints (that are not seen at training time). which attain by proposing a self-calibration loss coupled with a gradient reversal layer [Morerio *et al.*, 2021]. Remarkably, our pipeline encourages privacy-safe implementations since it removes any visual information about the people in the scene immediately after the pose detection step.

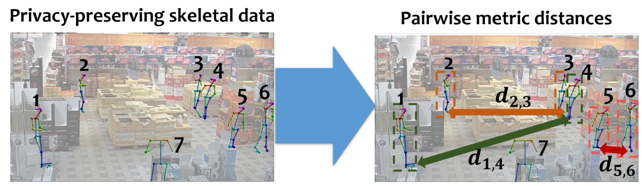

Figura 1: We estimate the pairwise metric distances using only 2D skeletal inputs of any pair of detected persons (each pair is indicated by a different color) without the need of pre-calibrating cameras.

We train and evaluate our pipelines ([Aghaei *et al.*, 2021] and [Morerio *et al.*, 2021]) against baselines on public dataset that are upgraded to provide metric distances among people in the scenes. In particular, we prepared four datasets, out of the publicly available Epfl-Mpv [Fleuret *et al.*, 2008], Epfl-Wildtrack [Chavdarova *et al.*, 2018], OxTown [Benfold e Reid, 2011] and KITTI [Geiger *et al.*, 2012]. Additionally, we benchmark the cross-view generalisation capabilities of the proposed method and compare its capabilities on monitoring *Social Distancing (SD)* against state of the art approaches [Bertoni *et al.*, 2019]. Different versions of the algorithm have been deployed in real scenarios within collaborations with industrial partners, namely 'Ipercoop' and 'Aereoporto di Genova'. For more information on the algorithms, their implementation and benchmark results the reader can refer to section 4.

2 Projects

A first release of the algorithm was deployed on the field during the early months of the pandemic. Its output is depicted in Figure 2. The algorithm estimates a safe area around in-

Figura 2: A first implementation of the VSD pipeline locally estimates a safe area around each detected individual and identifies overlaps, i.e. Social Distancing violations (red circles). The size of the safe area can be customized according to local regulations.

dividuals and projects it to the image plane. Intersections are identified as Social Distancing violations.

The proposed setup is currently able to monitor up to 18 cameras with 640x480 resolution, running on a single Alienware Aurora R8 PC equipped with an NVidia RTX2080 GPU at 1 fps. The system has been successfully applied for monitoring social distancing 24/7 in challenging scenarios such as corporate offices, airports, and shopping malls, where more than 100,000 observations per hour are being evaluated. The collected statistics are used for both generating real time alarms finalised to break apart big groups of persons and for generating statistics in order to achieve effective space redesigns (see Fig. 3).

The collection of this huge stream of data on real application scenarios allows us to highlight the most important open issues that future developments should focus on:

- System alarms for the violations of safe spaces should not be based only on single frame analysis, but rather encompass temporal information and people tracking.
- Currently many of the SD laws allow family members to stay close to each other. A robust system ideally should be able to recognise family member formations and discard their violations from the computed ones.
- Our approach is frame-based and it can be applied to any kind of camera under certain setting constraints. In the case of fixed cameras, a statistical approach can be used for inferring metric references. This statistical approach would allow the system to be more robust to outliers (such as kids on the scene).

Figura 3: The heatmap in the image represents the areas on which SD violations mostly occur. The heatmap is produced using observations collected in 1-hour from a monitored retail shop. The area with the most observed violations is due to an exhibit placed in the main entrance of the shop that causes increasing SD violations.

3 Real world Scenarios

The project received significant push thanks to the interest shown by important companies that had to deal with the pandemic spread. In particular, IIT started in Spring 2020 a collaboration with both Genova Airport and Coop Liguria.

3.1 Genova Airport

During 2020 lockdown, all the social activities and the movement of people suddenly stopped, without any hint about the duration of the imposed limitations and about the restrictions that would have been applied after the lockdown end. Because of this reasons, Genoa Airport decided to install the developed VSD System and test it during the lockdown, in order to be ready to start the operations with an increased safety standard at the end of the lockdown. Using the VSD System, Genova airport had the opportunity to identify the activities and situations with reduced interpersonal distance, consisting primarily in the waiting room and queues organizations and in the check-in phase, and reorganize them for reducing the personnel and customer risk.

3.2 Coop Liguria

Supermarkets have been one of the few commercial activities which did not stop during 2020 lockdown. The increased request in safety standards related to the pandemic risk, forced the supermarket managers to reorganize the available human resources and dedicate part of them to the monitoring of crowding situations and to the control of the access. Coop Liguria and IIT started a collaboration for real time monitoring of the crowding situations using the described VSD System. The generated alarms and the related anonymized camera view was shared via mobile phone interface to the chosen security persons, in order to break apart groups of persons too close each others. Thanks to this solution, most of the supermarket personnel had the possibility to move back to its original tasks and the supermarket manager collected extended statistics about the time and the location of the violations occurring in its store.

4 Publications and code

- A first position paper on the Visual Social Distancing Problem was published during the first wave of the pandemic [Cristani *et al.*, 2020], describing the problem and its challenges and also sketching possible solutions.
- An early release of the algorithm (deployed in 'Aereoporto di Genova' and 'IperCoop') was publishes at the "*Winter Conference on Applications of Computer Vision, 2021*" [Aghaei *et al.*, 2021]. The code is available at https://github.com/IIT-PAVIS/Social-Distancing.
- A more sophisticated,end-to-end Deep-Learning-based solution, was published in the "*IEEE International Conference on image Processing, 2021*" [Morerio *et al.*, 2021] and is currently deployed in the IIT labs for scientific purposes.

Riferimenti bibliografici

[Aghaei *et al.*, 2021] Maya Aghaei, Matteo Bustreo, Yiming Wang, Pietro Morerio, e Alessio Del Bue. Single image human proxemics estimation for visual social distancing. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2021.

[Benfold e Reid, 2011] Ben Benfold e Ian Reid. Stable multi-target tracking in real-time surveillance video. In *CVPR*, pages 3457–3464, June 2011.

[Bertoni *et al.*, 2019] Lorenzo Bertoni, Sven Kreiss, e Alexandre Alahi. Monoloco: Monocular 3d pedestrian localization and uncertainty estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6861–6871, 2019.

[Chavdarova *et al.*, 2018] Tatjana Chavdarova, Pierre Baqué, Andrii Maksai, Stéphane Bouquet, Cijo Jose, Louis Lettry, Francois Fleuret, Pascal Fua, e Luc Van Gool. Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection. In *Proc. of IEEE/CVF Conference On Computer Vision And Pattern Recognition*, pages 5030–5039, New York, 2018. IEEE.

[Criminisi *et al.*, 2000] Antonio Criminisi, Ian Reid, e Andrew Zisserman. Single view metrology. *International Journal of Computer Vision*, 40(2):123–148, 2000.

[Cristani *et al.*, 2020] M. Cristani, A. Del Bue, V. Murino, F. Setti, e A. Vinciarelli. The visual social distancing problem. *IEEE Access*, 8:126876–126886, 2020.

[Fleuret *et al.*, 2008] Francois Fleuret, Jerome Berclaz, Richard Lengagne, e Pascal Fua. Multicamera people tracking with a probabilistic occupancy map. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):267–282, February 2008.

[Geiger *et al.*, 2012] Andreas Geiger, Philip Lenz, e Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[Hall, 1966] Edward Twitchell Hall. *The hidden dimension*. Doubleday & Co, 1966.

[Hynes *et al.*, 2020] William Hynes, Benjamin Trump, Patrick Love, e Igor Linkov. Bouncing forward: a resilience approach to dealing with covid-19 and future systemic shocks. *Environment Systems and Decisions*, pages 1–11, 2020.

[Morerio *et al.*, 2021] Pietro Morerio, Matteo Bustreo, Yiming Wang, e Alessio Del Bue. End-to-end pairwise human proxemics from uncalibrated single images. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 3058–3062, 2021.