Resolving Collisions in Dense 3D Crowd Animations

Gonzalo Gomez-Nogales, Melania Prieto-Martin, Cristian Romero, Marc Comino-Trinidad, Pablo Ramon-Prieto, Anne-Hélène Olivier, Ludovic Hoyet, Miguel A Otaduy, Julien Pettre, Dan Casas

▶ **To cite this version:**

Gonzalo Gomez-Nogales, Melania Prieto-Martin, Cristian Romero, Marc Comino-Trinidad, Pablo Ramon-Prieto, et al.. Resolving Collisions in Dense 3D Crowd Animations. ACM Transactions on Graphics, 2024, pp.1-14. 10.1145/3687266 . hal-04681138

HAL Id: hal-04681138
https://inria.hal.science/hal-04681138v1

Submitted on 29 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Resolving Collisions in Dense 3D Crowd Animations

GONZALO GOMEZ-NOGALES*, Universidad Rey Juan Carlos, Spain
MELANIA PRIETO-MARTIN*, Universidad Rey Juan Carlos, Spain
CRISTIAN ROMERO, Universidad Rey Juan Carlos, Spain
MARC COMINO-TRINIDAD, Universidad Rey Juan Carlos, Spain
PABLO RAMON, Universidad Rey Juan Carlos, Spain
ANNE-HÉLÈNE OLIVIER, Inria, Univ Rennes, CNRS, IRISA, M2S, France
LUDOVIC HOYET, Inria, Univ Rennes, CNRS, IRISA, France
MIGUEL A. OTADUY, Universidad Rey Juan Carlos, Spain
JULIEN PETTRE, Inria, Univ Rennes, CNRS, IRISA, France
DAN CASAS, Universidad Rey Juan Carlos, Spain

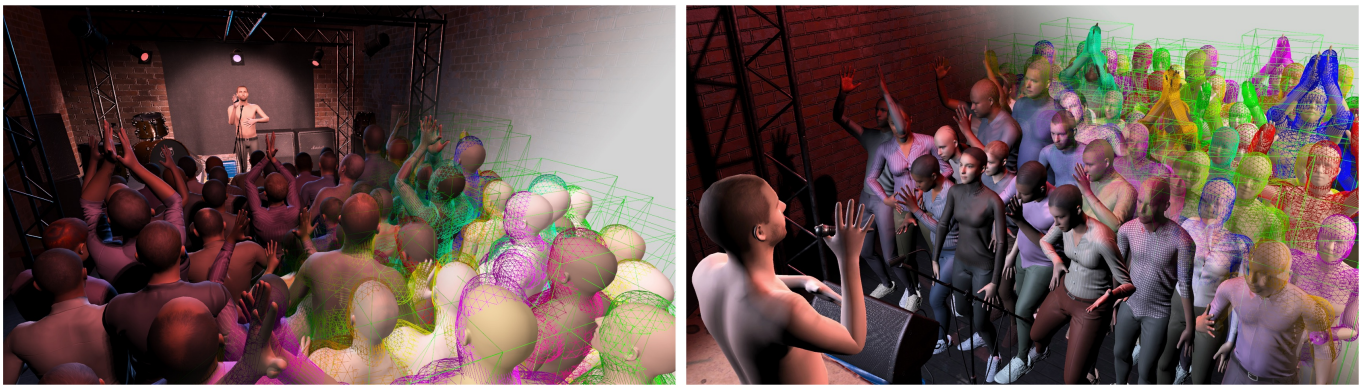

Fig. 1. Our physics-based method resolves the collisions in highly-dense 3D crowds, enabling the synthesis of 3D crowd animations where characters realistically interact and push each other, as shown in this indoor concert scene. Through a rigorous perceptual study, we demonstrate that resolving collisions is needed to generate 3D dense crowds that move in a natural and convincing way, while traditional animation methods do not model or correct such contacts between 3D characters.

We propose a novel contact-aware method to synthesize highly-dense 3D crowds of animated characters. Existing methods animate crowds by, first, computing the 2D global motion approximating subjects as 2D particles and, then, introducing individual character motions without considering their surroundings. This creates the illusion of a 3D crowd, but, with density, characters frequently intersect each other since character-to-character contact is not modeled. We tackle this issue and propose a general method that considers any crowd animation and resolves existing residual collisions. To this end, we take a physics-based approach to model contacts between articulated characters. This enables the real-time synthesis of 3D high-density crowds with dozens of individuals that do not intersect each other, producing an unprecedented level of physical correctness in animations. Under the hood, we model each individual using a parametric human body incorporating a set of 3D proxies to approximate their volume. We then build a large system of articulated rigid bodies, and use an efficient physics-based approach to solve for individual body poses that do not collide with each other while maintaining the overall motion of the crowd. We first validate our approach objectively and quantitatively. We then explore relations between physical correctness and perceived realism based on an extensive user study that evaluates the relevance of solving contacts in dense crowds. Results demonstrate that our approach outperforms existing methods for crowd animation in terms of geometric accuracy and overall realism.

CCS Concepts: • **Computing methodologies → Motion processing**.

Additional Key Words and Phrases: Crowd Animation, Physics-Based Animation

ACM Reference Format:
Gonzalo Gomez-Nogales, Melania Prieto-Martin, Cristian Romero, Marc Comino-Trinidad, Pablo Ramon, Anne-Hélène Olivier, Ludovic Hoyet, Miguel A. Otaduy, Julien Pettre, and Dan Casas. 2024. Resolving Collisions in Dense 3D Crowd Animations. *ACM Trans. Graph.* 1, 1, Article 1 (January 2024), 14 pages. https://doi.org/10.1145/3687266

*Equal contribution

Authors' addresses: Gonzalo Gomez-Nogales, Universidad Rey Juan Carlos, Spain, gonzalo.gomez@urjc.es; Melania Prieto-Martin, Universidad Rey Juan Carlos, Spain, melania.prieto@urjc.es; Cristian Romero, Universidad Rey Juan Carlos, Spain, cristian.romero@urjc.es; Marc Comino-Trinidad, Universidad Rey Juan Carlos, Spain, marc.comino@urjc.es; Pablo Ramon, Universidad Rey Juan Carlos, Spain, pablo.ramon@urjc.es; Anne-Hélène Olivier, Inria, Univ Rennes, CNRS, IRISA, M2S, France, anne-helene.olivier@univ-rennes2.fr; Ludovic Hoyet, Inria, Univ Rennes, CNRS, IRISA, France, ludovic.hoyet@inria.fr; Miguel A. Otaduy, Universidad Rey Juan Carlos, Spain, miguel.otaduy@urjc.es; Julien Pettre, Inria, Univ Rennes, CNRS, IRISA, France, julien.pettre@inria.fr; Dan Casas, Universidad Rey Juan Carlos, Spain, dan.casas@urjc.es.

© 2024 Copyright held by the owner/author(s).
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Transactions on Graphics*, https://doi.org/10.1145/3687266.

1 INTRODUCTION

A realistic crowd animation showing, for instance, many characters in a public place, a street, a stadium, or a concert, is achieved in practice by combining several techniques, each of which takes charge of a part of the animation, in a layered fashion. In this way, the animation of a crowd can be decomposed into computing its global features (i.e., positions and trajectories of characters relative to the environment) [Guy et al. 2010; Van den Berg et al. 2008] and into the detailed animation of the body pose of each character [Maïm et al. 2009; Pelechano et al. 2011; Sung et al. 2005] (generated independently afterwards, but driven by the evolution of global positions as well as some full-body motion assets).

Even though the field is recent, the animation process divided into successive passes is mainstream [Gustafson et al. 2016; Kanyuk 2016], independently of the application (e.g., video games or visual effects for movies), and it will probably remain as this for another decade, because it is a highly-efficient strategy to handle both the algorithmic and practical complexity of crowd animation. Nevertheless, as it was raised by Hoyet et al. [2016], even if such a decomposition simplifies the animation process, it also generates artifacts: because the body animation of crowd characters is computed separately and independently from the global motion features, some of the interactions between neighbors are neglected. In other words, there is a mismatch between each character's body motion and their immediate surroundings. Hoyet et al. demonstrated the benefit of reintroducing motion details (in their work, secondary shoulder motions) to give the illusion that physical interactions are simulated between characters where they are not. The value of reintroducing motion details –that would result from interactions– was proved, but no general principle was proposed to reintroduce them.

Our objective is to explore a more general solution to improve the quality of crowd animations by solving motion artifacts that result from the lack of interactions between crowd characters at the body animation stage. Like in previous work, we consider a two-step crowd animation process where global features (characters' positions for a stationary crowd, or their trajectories for non-stationary ones) are first computed, followed by a full body animation pass using a motion capture based technique. We consider as well the residual collisions that can result from this process which are known to be visually striking, especially at close distances [Kulpa et al. 2011]. Note that our approach works as a final additional animation layer and is therefore fully compatible with any traditional animation pipeline. Thus, our method considers a given animation of crowd characters, detects the presence of collisions (*i.e.*, geometry overlaps and inter-penetrations), and solves them by editing the characters' motion. Furthermore, our approach is based on physical simulation, computing and applying collision repulsion forces based on the configuration of the limbs and colliding bodies. This way, the motion is progressively edited so as to reach a negligible interpenetration volume between characters, while realistic surface contacts between them can remain. Because our approach works in a frame-by-frame fashion and can be both applied to an existing animation or included as a final step in an online animation loop, we can consider online (games, virtual reality) or offline (movies) applications.

In this work, we also set the objective of evaluating the effect of our technique on the overall crowd animation quality. This is crucial because we are facing a potentially contradictory objective. On the one hand, our technique removes residual collisions which are clearly physically wrong, and therefore reintroduces to the crowd a consistency that has been lost on the way. On the other hand, correcting residual collisions requires to edit character motions, which might alter visual quality by decreasing overall motion visual plausibility. Thus, as previously done to evaluate visual quality of crowd animations [Ennis et al. 2010; Kulpa et al. 2011; McDonnell et al. 2008], we propose a perceptual evaluation of our method applied to stationary and non-stationary crowds, changing densities and various points of view. Changing crowd density directly plays on the required amount of motion editing to solve collisions, enabling us to explore the trade-off set by our method between physical consistency and motion alteration.

Our contribution is thus twofold:

- We propose a new physically-based method to solve the residual collisions with inter-penetrations between virtual character bodies that remain in a given generated crowd animation. The method has real-time capabilities and is agnostic to the used crowd animation pipeline.
- We propose a perceptual study to evaluate the range of situations where this method is efficient. Our results suggest that crowd animations corrected with our method are perceived overall as being more realistic, with more natural contacts, and with characters more aware of their surroundings.

2 RELATED WORK

2.1 Crowd Simulation

Crowd simulation aims to reproduce the behavior of real crowds [Kapadia et al. 2016; Thalmann and Musse 2012]. In its applications to the graphics domain, for example for visual effects in cinema [Yang et al. 2020] or for Virtual Reality [Xu et al. 2014], the crowd is represented by detailed 3D agents, the limbs of which are articulated. However, to cope with the complexity raised by the synthesis of so many articular trajectories, crowd animation techniques decompose the problem into several sub-parts and, with few exceptions [Narang et al. 2018; Singh et al. 2011; Stüvel et al. 2016; Yao et al. 2023]. The algorithms for navigation or global positioning of the agents in the environment are thus decoupled from those used for character body animation [Lemonari et al. 2022; van Toll and Pettré 2021]. The former are based on simplistic 2D proxy geometries (e.g., discs) whilst the latter move poly-articulated bodies. Whatever the exact combination of navigation and animation techniques used in a crowd animation pipeline, this decoupling generates visual artifacts: the resulting animations do not reflect the effect of physical interactions between neighbor characters, and collisions can persist in the final result. It was shown that, when large crowds are displayed, spectators barely spot residual collisions [Daniel et al. 2021; Kulpa et al. 2011]. Nevertheless, correcting them by introducing secondary motions that capture the effect of physical interactions is beneficial to the overall realism of the crowd animation [Hoyet et al. 2016].

Since the visual realism of crowd animation is a common objective, but the respective contribution of each component of a crowd

animation pipeline is difficult to disentangle, perceptual studies have often been used in the past to better compose this pipeline or tune some parameters. Such studies include for example exploring the effects of motion variety [Adili et al. 2021; McDonnell et al. 2008], LOD clothing [McDonnell et al. 2006], character responsiveness [Kyriakou and Chrysanthou 2018], emotion expression [Carretero et al. 2014], gaze animation [Narang et al. 2016], visual vs. auditory channel dominance [Ennis et al. 2010], etc. Closer to the goal of this paper, Kulpa et al. [2011] investigated the perception of collisions in crowd animations. They showed that collisions at a distance from the point of view are not detected, but also that close collisions are detected and need to be solved, which justifies our goal.

We should also mention the various crowd animation software packages dedicated to visual effects in film (e.g. Golaem [Gol 2024], Miarmy [Mia 2024] or Massive [Mas 2024]). While the goal of achieving good quality visual results with limited computational time is common to the work mentioned above, a specific goal of the animation pipeline is to provide the user with the means to direct and edit crowd animations. In these software packages, physical animation is used to fully animate *ragdoll* characters (typically in explosion scenarios) rather than to edit an existing animation.

In conclusion, the separation between navigation layers and animation layers is still a reality in crowd simulation and animation, especially for real-time applications like virtual reality. In the same spirit as Hoyet et al. [2016], we propose a method to solve some resulting artifacts, but unlike previous work based on pre-recorded secondary motions, we propose a *general physically-based* method to achieve this goal. Moreover, we evaluate our method not only objectively and quantitatively, but also perceptually to validate that our method is also beneficial for the overall crowd animation realism.

2.2 Physics-Based Modelling of Human Contact

Physics-based character animation methods naturally incorporate contacts into their formulation, which potentially enables realistic interaction between kinematic characters and the environment. Early works [Fang and Pollard 2003; Liu and Popović 2002; Witkin and Kass 1988] use space-time optimization to compute a motion that satisfies physical constraints, but they are mostly limited to foot-ground contact. Forward dynamics methods compute joint torques and apply them to synthesize realistic movements [Coros et al. 2010; Hodgins et al. 1995; Yin et al. 2007], but they typically struggle with scenarios with multiple contacts. Adding reference trajectories into optimization-based methods helps [Macchietto et al. 2009; Muico et al. 2009], but it is difficult to generalize to unseen contact states. For better generalization, guided motions and randomized sampling of the controller [Liu et al. 2010; Sok et al. 2007] have been combined. More recently, many physics-based character animation works that use modern learning strategies have been proposed [Liu and Hodgins 2018; Peng et al. 2018; Starke et al. 2020] These methods combine physics and techniques such as reinforcement learning to train motion controllers that can interact with the environment. We also use physics in our approach but, while most of the literature in physics-based character animation aims at *synthesizing the kinematic motion* of a single character, we focus on *solving multi-character collisions* caused by the body volume overlap.

Another trend of physics-based methods closely related to us is the modeling of 3D volumetric characters as deformable objects [Capell et al. 2002; Galoppo et al. 2007]. These works typically use skinning-based reduced simulation methods [Gilles et al. 2011; Wang et al. 2015] to define subspace models for deformable objects, but fast and accurate contact is difficult to incorporate [Teng et al. 2015]. Closer to ours are the methods that propose volumetric models for parametric humans [Ramon et al. 2023] such as SMPL [Loper et al. 2015], which open the door to resolving collisions with external objects. For example, Kim et al. [2017] first compute a volumetric discretization of the SMPL template mesh, and then define and parameterize a mechanical model by fitting the deformed volumetric template to high-quality 4D scans. Similarly, Romero et al. [2020] enrich the pose-dependent static deformations from SMPL by using a volumetric deformable model that is also fitted into 4D scans. Additionally, they define an anisotropic nonlinear material that accurately represents skin dynamics to reproduce deformations due to external forces. Tapia et al. [2021] also propose a volumetric SMPL model by combining global data-driven static deformations with an efficient and local skinning-based subspace model, which enables real-time performance. Despite the impressive realism of contact deformations produced by these methods, the underlying physics-based models remain very computationally expensive and do not scale to dozens of individuals.

Very recently, data-driven methods have shown promising results to speed up the computation of deformations by analyzing contacts in physics-based methods. For example, Holden et al. [2019] learn the dynamic update of subspace deformable objects under external contact, even though their results are mostly limited to global deformations. Other learning-based works [Romero et al. 2022, 2021] enrich a model reduction strategy with a data-driven method to produce contact-driven deformations. Even if these methods succeed to produce detailed deformations, they require an exhaustive sampling of the subspace and are trained on a collider-specific dataset. Additionally, despite the speed-up provided by the machine learning component, they do not scale up to dozens of characters. In contrast, we opt for coarsely approximating the character volume, which enables real-time collision resolution for highly dense crowds.

Lastly, our work is also related to the more general methods for collision detection for deformable objects [Lan et al. 2020]. However, we argue that (expensive) nonrigid deformation is not needed to model accurate contacts in crowds and therefore opt for a model based on fast and responsive simulation of skeletal dynamics.

3 RESOLVING COLLISIONS IN DENSE CROWDS

Our objective is to adjust the individual pose of many characters in a dense crowd by solving the collisions that exist between their volumes while maintaining the overall crowd motion as faithfully as possible. To this end, we first define our underlying parametric human model (Section 3.1), which we then use to formulate a physics-based dynamic human model (Section 3.2) that is guided by motion capture data and the global crowd motion (Section 3.3). Finally, we show how we incorporate collision handling into our formulation by adding a set of parametric volumetric primitives to approximate our human model (Section 3.4).

3.1 Parametric Human Model

Starting from a human crowd consisting of a set of kinematic skeletons and their corresponding pose and global position over time, we first define a body representation that enables the computation of collisions between characters. To this end, we leverage the vast literature on 3D body models [Feng et al. 2015; Joo et al. 2018; Loper et al. 2015] that deform a rigged parametric human template

$$M(\beta, \theta) = W(T(\beta, \theta), J(\beta), \theta, \mathcal{W}) \quad (1)$$

where W is a skinning function (*e.g.*, linear blend skinning, or dual quaternion) with skinning weights $\mathcal{W}$, $J(\beta) \in \mathbb{R}^{3\times 24}$ the joint positions, and θ the pose parameters of the kinematic skeleton that deform an unposed parametric body mesh $T(\beta, \theta)$. More specifically, we use the nowadays standard SMPL model [Loper et al. 2015], which defines the unposed body mesh as

$$T(\beta, \theta) = \mathbf{T} + B_s(\beta) + B_p(\theta) \quad (2)$$

where $\mathbf{T} \in \mathbb{R}^{N_B\times 3}$ is a body mesh template with N_B vertices that is deformed using two blendshapes that output per-vertex 3D displacements: $B_s(\beta) \in \mathbb{R}^{N_B\times 3}$ models deformations to change the body shape; and $B_p(\theta) \in \mathbb{R}^{N_B\times 3}$ models deformations to correct skinning artifacts.

Several previous works have extended SMPL's kinematic representation of the body surface to support physics-based soft-tissue deformation [Kim et al. 2017; Romero et al. 2020; Tapia et al. 2021]. However, at the target scale of our work, the relevance of skeletal response to collisions is far more notorious, and soft-tissue modeling does not scale well to the dozens of individuals required for dense crowds. Therefore, we favor the fast, responsive simulation of skeletal dynamics, and defer soft-tissue modeling to future work.

3.2 Physics-Based 3D Humans

To endow the above parametric human model with physics-based motion, we compute the pose θ as the result of a physics-based simulation. Specifically, we formulate equations of motion that include the following mechanical terms: inertia, gravity, joint constraints, collisions, and a control term to follow the input body animation.

We formulate the motion of the articulated skeleton as a rigid body simulation with soft constraints to model joints. We denote as $\mathbf{q}$ the aggregate state of all individuals in a crowd, which includes the degrees of freedom of all rigid bones in each body. Given all mechanical terms for all bodies, we integrate the equations of motion using the popular optimization formulation of backward Euler [Gast et al. 2015; Kane et al. 2000]:

$$\mathbf{q} = \arg\min E_{\text{inertia}} + E_{\text{gravity}} + E_{\text{joints}} + E_{\text{collisions}} + E_{\text{control}}. \quad (3)$$

After each time-step solve, we project the rigid-body state $\mathbf{q}$ to each body's parametric pose θ, and we evaluate the parametric human model in Eq. 1 to obtain the posed body surfaces.

In our model, we use standard formulations for inertia, gravity and joints on rigid bones [Ferguson et al. 2021], and we discuss collisions and control below. We solve the optimization in Eq. 3 using Newton's method with analytical computation of gradients and Hessians. Just one Newton iteration per time step worked well for our dynamic simulations. Each Newton iteration yields a sparse linear system of size $6 \times 24 \times N$, with N the number of individuals in the crowd, 24 the number of bones per body, and 6 the number of degrees of freedom per bone. We solve this linear system using conjugate gradient.

Fig. 2. Three example poses of our underlying 3D body representation. We enrich the SMPL [Loper et al. 2015] parametric human model (in solid brown) with a set of parametric volumetric 3D primitives (in semitransparent colors), which enables fast and highly efficient evaluation of collisions.

3.3 Coupling to full-body animations

The skeletal motion of individuals is dictated by pre-recorded full-body animations, and we want bodies to respond naturally to collisions while still following the input animation. Animation control is a whole research topic in its own [Liu and Hodgins 2017; Peng et al. 2018], and we resort to a PD controller as it sufficed for our problem at hand. Our PD controller couples each body to its corresponding skeletal animation input in the following way. For the body root, we couple absolute translation and rotation. For the rest of the body, on the other hand, we couple relative joint rotations. Given a joint with current body and joint rotations $\mathbf{R}$ and $\hat{\mathbf{R}}$ respectively, and previous-step rotations $\mathbf{R}_{\text{old}}$ and $\hat{\mathbf{R}}_{\text{old}}$, we define its PD control term (i.e., spring and damping) as:

$$E_{\text{control}}(\mathbf{R}) = \frac{1}{2}\, k\, \left|\mathbf{R} - \hat{\mathbf{R}}\right|_F^2 + \frac{1}{2}\, \frac{d}{h^2}\, \left|\mathbf{R} - \mathbf{R}_{\text{old}} - \hat{\mathbf{R}} + \hat{\mathbf{R}}_{\text{old}}\right|_F^2, \quad (4)$$

where k and d are, respectively, spring and damping coefficients, h is the time step, and $|\cdot|_F$ denotes Frobenius norm. This joint control term penalizes deviations in joint rotations and joint velocities between the physics-based simulation and the animation input. As we show later in Figure 4, we found it was important to damp deviations in joint velocities, not absolute joint velocities, to ensure that bodies recover their trajectory smoothly after blocking collisions.

For the pelvis and the feet, we use higher spring and damping coefficients to ensure stronger alignment of the trunk with the skeletal animation, as well as to avoid footskate.

3.4 Modelling Physics-Based Collisions

Our physics-based 3D human simulation considers four types of collisions: body-ground, body-environment, inter-body, and intra-body collisions. An exact solution to collisions would require evaluating penetrations with respect to the exact surface model from Eq. 1, but pose blendshapes and skinning complicate this task. Instead, we propose to use a set of parametric geometric primitives to efficiently approximate the SMPL body volume and resolve all collision types.

To this end, for each bone $\mathbf{b}_j$ of the SMPL kinematic skeleton we define a capsule $\mathbf{k}_j$ computed as the intersection of two spheres of radius r_j and one cylinder of radius r_j and length l_j. These capsule parameters are computed as a function of the body shape β: l_j is set to match the length of the bone $\mathbf{b}_j$ given the joint positions $J(\beta)$, and r_j is optimized such that the volume of the capsule $\mathbf{k}_j$ is a best-fit to the body mesh $T(\beta, \theta)$ vertices influenced by bone $\mathbf{b}_j$. Capsules are transformed following the bone transformation given by pose parameter θ at each frame. Our parametric human model is illustrated in Figure 2, depicting the original surface SMPL mesh (in solid brown) and our volumetric primitives on top (in semi-transparent). The use of a set of simple geometric proxies significantly simplifies the computation of collisions, while closely approximating the true volume of the character.

We execute collision detection for all four collision types as follows. For each body, we build an AABB after each simulation time step. For body-ground collision, we simply test all capsules against the ground plane and compute the penetration depth for colliding capsules. For body-environment collisions, we point-sample static environment objects, build a static AABB-tree, and cull collisions against body AABBs. For colliding points, we compute the penetration depth with respect to the collision capsule. For inter-body collisions, we first test body-level AABBs to cull body pairs. Finally, for intra-body collisions, we discard adjacent capsule pairs, which we identify as those that collide in T-pose. For each potentially colliding capsule-capsule pair (either inter-body or intra-body), we first execute a faster sphere-sphere culling test, and we compute capsule-capsule interpenetration only for those pairs that survive all culling tests. Then, we formulate for each computed collision a penalty potential based on the penetration depth δ as $E_{\text{collisions}}(\delta) = \frac{1}{2} k_\delta \delta^2$.

4 EVALUATION

Our evaluation explores two types of crowd animations, walking and dancing, as well as various aspects of the method. We first describe how we generated the different types of crowd examples (Sec. 4.1), which were then used for our quantitative (Sec. 4.2) and qualitative (Sec. 4.3) evaluations, as well as stimuli for our user study (Sec. 4.4).

4.1 Crowd Animations

For our evaluations, we focus on two types of crowd animations: **stationary** crowds where characters move but their global position is fixed (*e.g.*, dancing crowds), and **non-stationary** crowds where characters perform navigation (*e.g.*, walking crowds). In both cases, our approach builds on top of a standard animation pipeline decoupling the simulation (computing the 2D global position or trajectory of characters) and animation (computing full body motions) of the characters in the crowds.

First, the original position of the N characters in each crowd animation is computed according to a desired density using a Poisson Disk Sampling algorithm, to ensure regular spacing between characters. In the case of non-stationary scenarios (*e.g.*, walking), any 2D simulator can then be used to compute the global movement of the crowd, which is commonly done by solving an optimization problem where each agent is approximated as a 2D particle, such as PLE [Guy et al. 2010] or RVO [Van den Berg et al. 2008]. For the non-stationary scenarios used in this paper, we use the RVO algorithm provided by van Toll et al. [2020] to generate the 2D crowd trajectories.

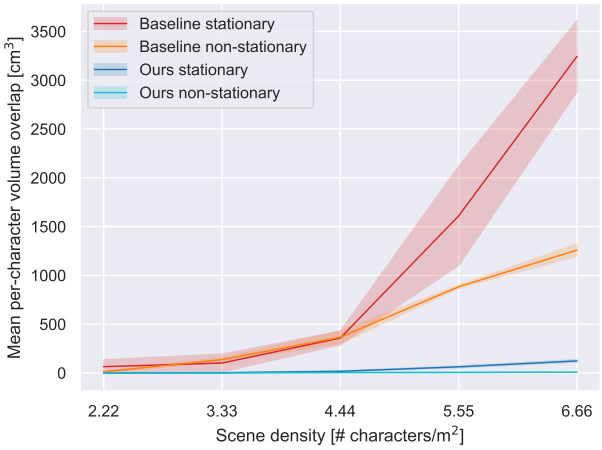

Fig. 3. Quantitative evaluation of the mean per-character volume overlap in animations of different crowd densities, for both stationary (*i.e.*, dancing) and non-stationary (*i.e.*, walking) scenes. Our approach resolves collisions even in highly dense crowds, while the baseline method based on state-of-the-art crowd simulation results in 3D characters that significantly intersect each other. See Figure 5 for qualitative visualization of this analysis.

Then, we created full-body animated characters using SMPL and animations from the Mixamo database [mix 2023]. More specifically, we selected 15 animations that evoked dancing characters for the stationary crowds, and 4 walking animations for the non-stationary crowds with different walking speeds which we combined in a controllable blend tree in Unity. Each animation was manually cut and looped, to ensure seamless continuous animation. These animations were displayed on N unique virtual characters created using SMPL models described in Section 3.1. To ensure a realistic distribution of body shapes, β parameters were selected randomly following a Gaussian distribution that matched men and women height distributions around the world, obtaining N unique bodies, half of them belonging to each gender. Regarding their appearance, 65 female and 57 male photorealistic synthetic textures were used, which cover a wide diversity of races and ethnicities.

The animations created in this manner served for the **Baseline** examples, i.e., they correspond to crowd animations created using standard state-of-the-art methods. These crowd animations were then processed using the approach described in Sections 3.2, 3.3 and 3.4, enabling the synthesis of collision-aware realistic dense 3D crowds (**Ours**), and used for the quantitive and qualitive evaluations, and as stimuli for the perceptual study presented in the following sections.

4.2 Quantitative Evaluation

To quantitatively validate the capability of resolving collisions of our approach, we evaluate the (undesired) per-character mean volumetric overlap in a set of 3D crowd animations of different character

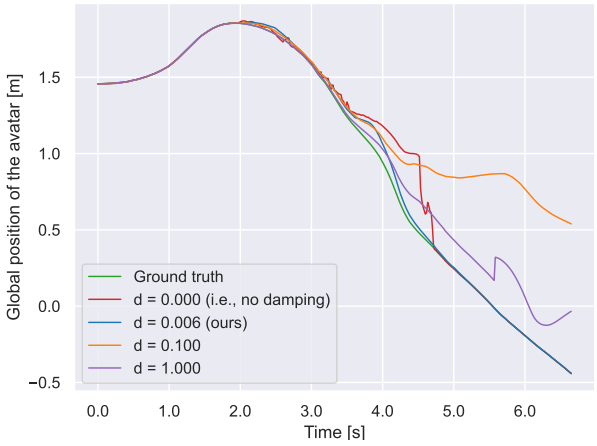

Fig. 4. Quantitative analysis of the damping parameter d of Equation 4. By damping the joint velocities with a small weight ($d = 0.006$, in blue) we are able to closely match the target position (in green) while recovering from a collision that occurred at $t = 3.5$ seconds. If no damping is applied or it is too high, the resulting animation is unstable (in red) or drifts too much from the target motion (in orange and purple).

density levels. Specifically, using the strategy described in Sec. 4.1, we generate stationary animations (*e.g.*, dancing) of 20, 30, 40, 50, and 60 characters on an area of 9 m^2, which results in crowded animations ranging from 2.2 persons/m^2 (*i.e.*, a density level common in a crowded street) to more than 6 persons per m^2. Note that we consider it to be a relevant upper density limit, since beyond this density level risk of crowd disasters becomes high and other kinds of crowd dynamics enter into play [Still 2000]. Analogously, to evaluate non-stationary crowds, we also generated walking animations for densities 2.2 to to 6.6 persons/m^2 going through a narrow corridor.

We then compute the average per-character volume overlap in each animation, and compare the results using our physics-based solution to the baseline approach (Figure 3). Using the baseline approach, the overlapping volume increases dramatically, leading to animations where most of the characters are *inside* others. In contrast, our approach is able to solve collisions even in highly-dense scenarios, and maintains a relatively low average overlapping volume despite the increase in density.

Figure 5 provides a visualization of this analysis, showing representative frames of the generated animations for different density level and approach. Pixels corresponding to 3D points with overlapping volumes are colored in red. It can be seen that the baseline method based on state-of-the-art crowd simulation suffers from significant 3D character-to-character collisions, even in sparse crowds, leading to heavily entangled and colliding animations in dense crowds. In contrast, our method is able to maintain a relatively low number of residual collisions, even in highly dense animations.

The visual quality of the animations produced by our system depends significantly on the damping parameter d from Equation 4. If no damping is applied, the joint velocities may suffer unrealistic changes in consecutive frames when recovering from a collision while trying to follow the target animation. To quantitatively evaluate the effect of the damping parameter d, Figure 4 shows the global position of a character that collides with another one, for a range of values d. We demonstrate that with a damping parameter $d = 0.006$ the character smoothly recovers from the collision, closely matching the target animation.

We also quantitatively evaluate the runtime cost of our method. We use a standard desktop PC with a CPU AMD Ryzen 7 2700 3.2 GHz, 32 GB of RAM, and a GPU NVIDIA GeForce GTX 1080 Ti. Our system can handle up to 200 characters (i.e., what fits into 32GB RAM). Since we first use an efficient body-body AABB collision check, we can quickly discard all the capsule-capsule collision checks (e.g., limb-level contacts) for all characters that are not in contact. More specifically, adding our physics-based model takes up to 5 ms/step in the sparse scenes (20 avatars), and up to 20 ms/step in the dense stationary scene. Hence, in the worst-case scenario, the computational cost scales linearly with the number of characters, but we can maintain real-time performance in all the demos showcased in the paper.

4.3 Qualitative Evaluation

To illustrate the overall quality of the crowd animations generated with our approach, we show representative frames of different scenes. Please, see the supplementary video for animated results.

Figure 1 shows a dense animation of 60 characters in a concert (9 m^2 dance floor). Despite the challenging setup (*i.e.,* 3D humans can barely move in a density of 6 persons/m^2 or more), our approach is capable of generating convincing results, where human bodies do not intersect with each other while performing dancing motions.

Our approach also generalizes to other types of motions, as can be seen in Figure 6. Columns 1 to 3 show representative frames of crowd animation sequences recorded specifically for illustrative purposes: an actor wearing a motion capture suit was immersed in a virtual crowd using a virtual reality Head Mounted Display, and we recorded his motions while interacting with the crowd. Despite being recorded on a real person, the animations display numerous collisions between the user and the virtual characters (top). In comparison, we show how our approach (bottom) is able to resolve the collisions between the walking character and the crowd, producing an animation that exhibits a natural behaviour commonly seen when people walk pass each other at a close distance, i.e., a rotation of the torso and shoulders with the arm passing last. Importantly, this result demonstrates that despite having access to motion captured sequences of interacting agents, residual collision can still appear in recorded data. Solving residual collisions produced when animating 3D virtual characters is therefore necessary to obtain a high-level of realism in the final animations. Similarly, column 4 shows a frame of a walking crowd generated using a more traditional animation pipeline (described in Section 4.1), using the RVO algorithm [Van den Berg et al. 2008] to compute global trajectories, combined with the previously mentioned walking-cycle blend tree to animate each character. It can be seen that such a standard approach leads to many undesired collisions. Our collision-aware method is able to resolve them, producing realistic interactions between the 3D agents.

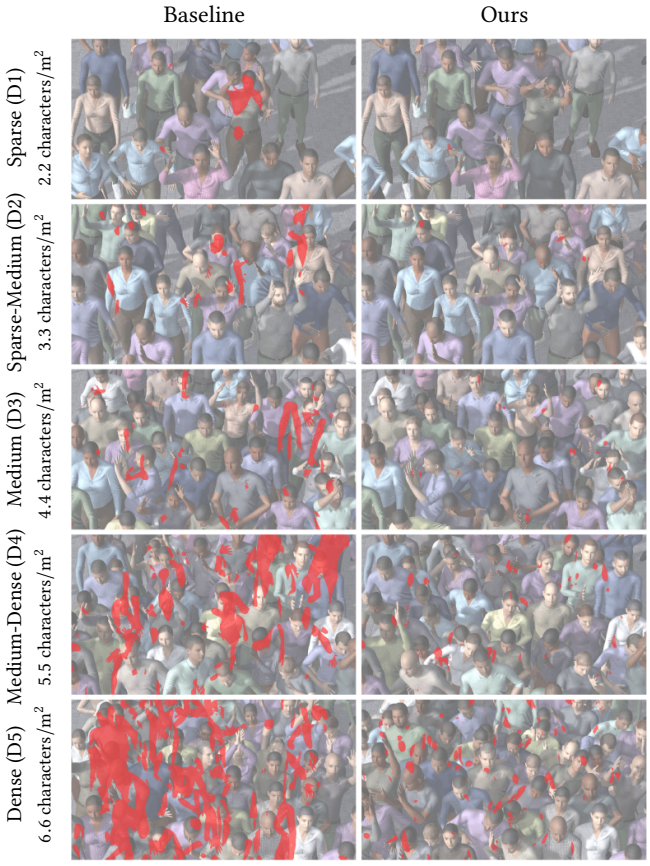

Fig. 5. Qualitative visualization of the character volume overlap in 3D crowds. In red, pixels of the image that contain overlapping volumes (i.e., colliding characters). Without solving collisions (left), collisions increase with the density of the crowd, leading to unrealistic animations where all characters intersect each other. Our approach resolves the collisions, resulting in realistic animations where characters push each other to fit into highly dense crowds. Due to our volumetric approximation of the body, residual collisions remain but do not affect the overall realism of the crowd.

4.4 Perceptual Evaluation

As our approach solves residual collisions by modifying the motion of crowd characters, the resulting increase in physical realism comes at the cost of altering the overall character motion. It is therefore essential for the evaluation of our method to investigate how users perceive such motion corrections. To this end, we designed a perceptual experiment where viewers were presented with a number of videos displaying stationary or non-stationary animated crowds of changing density, generated with or without applying our method. We asked them about the realism of character contacts (which is related to the presence of collisions with overlaps in animations), the motion quality (which is related to character motions being influenced by their neighbors), and the overall realism of the scene. The goal of this experiment is therefore to explore the following hypotheses.

H1 Crowd animations with residual collisions corrected with our approach will be preferred to the same animations without corrections. This will be observed in terms of higher perceived realism of contacts between characters (**H1-a**), higher perceived character awareness (**H1-b**), and higher overall realism (**H1-c**).

H2 These benefits will be higher at lower densities:
- **H2-a** We expect uncorrected residual collisions to be more visible at lower densities, as there is little mutual occlusion between characters, and therefore a higher benefit from correcting them with our approach.
- **H2-b** Due to the relatively higher free space around characters at lower densities, we expect that our method will require little motion editing to solve residual collisions, leading to animations corrected with our approach to be perceived as more realistic, and more aware of their surroundings.

H3 We expect that the perception of contact realism, character awareness and overall realism will be influenced by the point of view of the scene. In particular, we expect that bird's-eye viewpoint will ease the detection of collisions (**H3-a**), and will therefore increase both the perceived character awareness (**H3-b**) and overall realism (**H3-c**) of crowd animations, in contrast to eye-level or canonical viewpoints.

4.4.1 Experimental Design. To evaluate our hypotheses, we asked participants to watch a set of 3D crowd animation videos, populated at various levels of densities, with or without residual contact corrections using our method, rendered from several viewpoints, and displaying either stationary or non-stationary crowds. The stationary crowd scenarios illustrated a situation resembling a concert crowd, while non-stationary crowd scenarios displayed characters walking in a corridor. As in our quantitative evaluation from Section 4.2, we used 5 *Density* levels, from sparse to dense levels, corresponding respectively to 2.22, 3.33, 4.44, 5.55, and 6.66 characters/m^2 (labeled as D1 to D5). We also used 4 *Viewpoint* levels: front (V1), upper-front (V2), front-side (V3), and overhead (V4). Finally, for each condition, residual collisions were either left untouched (Baseline) or resolved by applying our method (Ours).

The experiment was then performed in two stages. In the first stage, we chose to demonstrate the benefits of using our method in the simplest stationary scenario, while simultaneously exploring the effects of density and viewpoint. Participants saw a total of 40 videos, presented in random order: 5 Density (D1 to D5) × 4 Viewpoint (V1 to V4) × 2 Method (Baseline vs. Ours). Following the positive results of the first stage, we then ran the second stage of the experiment to evaluate the benefits of our method on the more complex non-stationary scenario. Based on the analysis of the results of the first stage (described in Section 4.4.4), we however selected only two viewpoints for this stage (V3 and V4), as they were showing more varied results in the stationary Baseline condition. This subset eases that participants focus their attention on the goal of the study, despite the added complexity in the videos (i.e., non-stationary scenes are more dynamic). In the second stage, the same group of participants therefore saw a total of 20 videos, presented in random order: 5 Density (D1 to D5) × 2 Viewpoint (V3, V4) × 2

Fig. 6. Qualitative results of our approach for resolving collisions in dense crowds (bottom) compared to the baseline solution (top). Columns 1 to 3 depict representative frames of a sequence captured with an inertial Motion Capture suit. The character wearing a blue checkered shirt represents the motion-captured user who walked through a virtual crowd inside a virtual reality setup. Column 4 shows a representative frame of a sequence generated with a state-of-the-art crowd simulation algorithm [Van den Berg et al. 2008], used in our perceptual study for non-stationary scenes. The baseline method (top) produces an unrealistic animation, where characters intersect each other. In contrast, our physics-based approach (bottom) automatically adjusts the pose of each character such that they do not collide, producing a realistic animation despite the challenging scenario.

Method (Baseline vs. Ours). See Figure 8 for samples of the stimuli for different viewpoints, densities, and scenarios.

Participants were asked to answer the following assertions (using 6-point Likert-scales ranging from (1) "I strongly disagree" to (6) "I fully agree"), chosen to provide information about

(i) *contact realism*: **Q1** "Contacts made by characters between them seem natural to me", **Q2** "Character bodies overlap, it is not physically realistic to me",
(ii) *character awareness*: **Q3** "Characters seem to make contact and sometimes push one another", **Q4** "Characters move as if they were alone, their motion ignore neighbors",
(iii) and *overall realism*: **Q5** "A crowd moving like this one could exist in the real life", **Q6** "I feel this crowd as a whole does not move in a natural way".

Notice that each pair of questions was formulated in a way that there was one positive question, as well as one negative question (as for Control questions commonly used in evaluation questionnaires). Therefore, higher scores to **Q1**, **Q3** and **Q5**, and lower scores to **Q2**, **Q4** and **Q6**, mean better performance of our method.

The different crowd animations were generated using the procedure presented in Section 4.1. Each crowd animation was then rendered as a video at a 1920×1080 resolution. Each video has a duration of 20 secs, and looped until the participant had answered the 6 following assertions about the video (see Figure 7 for a screenshot of the UI design, which was displayed on a 24-inch display).

4.4.2 Participants. Thirty participants took part in the experiment (16♂, 14♀, age: 25.8 ± 7.4), which was carried out in the research facilities of Universidad Rey Juan Carlos (Spain). Participants were recruited through email lists among students, staff, and the general public. All participants were naive to the purpose of the experiment, had a normal or corrected-to-normal vision, and gave written and informed consent prior to the experiment.

4.4.3 Analysis. Questions with a similar but opposed meaning were first grouped by pair, by inverting the answer given to negative questions so that $\bar{\mathbf{Q}}i = 7 - \mathbf{Q}i, i \in (2, 4, 6)$, after checking the internal reliability between paired questions. Therefore, we analyze answers

Fig. 7. The GUI used for our perceptual user study. Participants were asked to watch videos of animated dense crowds (here presenting an example of the stationary dancing scenario) and to answer 6 questions for each video. See Section 4.4.1 for a detailed description of the questions and conditions.

about contact realism $\mathbf{Q}_{\text{CR}} = (\mathbf{Q1} + \bar{\mathbf{Q}}\mathbf{2})/2$, character awareness $\mathbf{Q}_{\text{CA}} = (\mathbf{Q3} + \bar{\mathbf{Q}}\mathbf{4})/2$, and overall realism $\mathbf{Q}_{\text{OR}} = (\mathbf{Q5} + \bar{\mathbf{Q}}\mathbf{6})/2$.

To assess the effect of our method, density and viewpoint, on participants' answers, we conducted for each crowd scenario 3-way repeated-measures ANOVAs with within-subject factors Density, Viewpoint, and Method. We set the level of significance to $\alpha = 0.05$ and used the notations * (p-value < 0.05), ** (< 0.01) and *** (< 0.001) to highlight significant differences in the figures. Results are reported as mean ± standard deviation. We assess the normality assumption using Q-Q plots, and sphericity using Mauchly's tests. Greenhouse-Geisser adjustments to the degrees of freedom were applied when appropriate to avoid any violation of the sphericity assumption. In case of a significant main or interaction effect, we performed pairwise comparisons using post-hoc Tukey tests.

4.4.4 Results. As we found similar effects across the three categories of questions, we present here the general results of our perceptual evaluation based on the studied factors. Also, we focus

Fig. 8. Sample frames from our user study. We show different viewpoints and scenarios (columns) for the set of densities (rows) used to generate the stimuli for the crowd animations of our perceptual study. For each viewpoint-density-scenario combination, we generate an animation with and without solving collisions. Here, for visualization purposes, we show representative frames of the stimuli generated with our method.

in this section on the principal results for the sake of clarity, while additional details are provided in the supplemental material.

Effect of Correcting Contacts. We found a strong main effect of Method in both scenarios for all the categories of questions: Contact Realism (stationary: $F_{1,29} = 154.81, p < 0.001, \eta_p^2 = 0.84$; non-stationary: $F_{1,29} = 326.63, p < 0.001, \eta_p^2 = 0.92$), Character Awareness (stationary: $F_{1,29} = 161.77, p < 0.001, \eta_p^2 = 0.85$; non-stationary: $F_{1,29} = 158.07, p < 0.001, \eta_p^2 = 0.85$) and Overall Realism (stationary: $F_{1,29} = 67.79, p < 0.001, \eta_p^2 = 0.70$; non-stationary: $F_{1,29} = 168.44, p < 0.001, \eta_p^2 = 0.85$).

The results show that *participants considered contacts to be inappropriate in the Baseline condition across scenarios and categories of questions.* More specifically, they suggest that when our method is applied i) contacts between characters are perceived to be more natural (validating **H1-a**), ii) characters are perceived as being adjusting their motion to the presence of neighbors (validating **H1-b**), iii) animated crowds appeared to be overall more realistic (validating **H1-c**). These results therefore completely validate **H1**.

Effect of Density. For the stationary scenario, we found a main effect of Density for all the categories of questions: Contact Realism ($F_{3.04,88.2} = 16.37, p < 0.001, \eta_p^2 = 0.36$), Character Awareness ($F_{3.08,89.30} = 5.57, p < 0.001, \eta_p^2 = 0.16$) and Overall Realism ($F_{4,116} = 4.38, p < 0.01, \eta_p^2 = 0.13$). For the non-stationary scenario, we only found a main effect of Density on Character Awareness ($F_{2.60,75.41} = 6.43, p < 0.001, \eta_p^2 = 0.18$). These main effects are illustrated in the supplemental material.

More interestingly, we also observed a Density × Method interaction in both scenarios for all the categories of questions (illustrated in Figure 9): Contact Realism (stationary: $F_{4,116} = 18.00, p < 0.001, \eta_p^2 = 0.38$; non-stationary: $F_{4,116} = 10.55, p < 0.001, \eta_p^2 = 0.27$), Character Awareness (stationary: $F_{4,116} = 22.60, p < 0.001, \eta_p^2 = 0.43$, non-stationary: $F_{4,116} = 4.29, p < 0.01, \eta_p^2 = 0.13$) and Overall Realism (stationary: $F_{4,116} = 15.09, p < 0.001, \eta_p^2 = 0.34$; non-stationary: $F_{4,116} = 4.97, p < 0.001, \eta_p^2 = 0.15$). For both Contact Realism and Overall Realism, post-hoc analyses showed an effect of density only when our method is not applied (Baseline), showing overall that in spite of larger occlusions between characters when density increases, *participants perceive body overlaps to be increasingly more unrealistic when they are not solved, which also negatively impacts the overall realism of the crowd.* For Character Awareness, the results even suggests that *characters were perceived to be more aware of their surroundings at the highest densities when contacts were corrected using our approach.* These results justify applying our method at both sparse and dense conditions, since scores about contact realism, overall realism, and character awareness are high and are not negatively affected by the density level with our method. This however contradicts **H2-a** and **H2-b**, as we expected our method to be more efficient at lower densities. This completely rejects **H2**,

while demonstrating that our method has a domain of validity in densities which is larger than we expected.

Effect of Viewpoint. First, we found a main effect of Viewpoint on Contact Realism ($F_{3,87} = 27.08, p < 0.001, \eta_p^2 = 0.48$) and Overall Realism ($F_{2.43,70.55} = 13.43, p < 0.001, \eta_p^2 = 0.32$), but only for the stationary scenario. More interestingly, we also observed a Viewpoint × Method interaction for all the categories of questions, again only for the stationary scenario: Contact Realism ($F_{3,87} = 13.69, p < 0.001, \eta_p^2 = 0.32$), Character Awareness ($F_{3,87} = 15.47, p < 0.001, \eta_p^2 = 0.35$) and Overall Realism ($F_{3,87} = 9.83, p < 0.001, \eta_p^2 = 0.25$). We also observed less significant Viewpoint × Density and Viewpoint × Density × Method interaction effects for all the categories of questions (reported in the supplementary material).

The Viewpoint × Method interaction results all reveal an unfavorable viewpoint (front-side **V3**) in the Baseline condition (illustrated in Figure 10-top), with lower scores on average than with any other viewpoint condition. This was the only viewpoint where both the whole body of foreground characters, as well as the upper body of background characters, can be observed. In contrast, the overhead bird's-eye viewpoint (V4) in the Baseline condition seems to be less favorable for perceiving unnatural contacts (which contradicts **H3-a**), leading to higher overall realism and perceived character awareness (which contradicts **H3-b** and **H3-c**). However, post-hoc analysis of the significant 3-way interaction effect for Contact Realism, Character Awareness, and Overall Realism suggest that these results only appeared for low densities (D1 and D2), as scores were not significantly different between viewpoints for the other densities (D3, D4 and D5). It also seems important to mention that these two viewpoints were not perceived to be significantly different in the non-stationary scenario across the categories of questions. Moreover, *realism and awareness scores were relatively high across viewpoints when our method was applied*, in both stationary and non-stationary scenarios. This suggests that corrected contacts are perceived to be quite realistic across viewpoints, with positive impacts on both character awareness and overall realism. These results therefore completely invalidate **H3**.

5 DISCUSSION

First of all, our results demonstrate that the proposed method is efficient. In Section 4.2, we show that our approach successfully resolves the initial collision volume, even though we still observe residual collision volumes as the method adjusts the pose of the characters while aiming to preserve surface contacts. Preserving contacts has two positive effects. First, it minimizes the changes with respect to the character's initial pose. Second, and maybe more interestingly, it gives the visual impression that characters push one another, as revealed by our perceptual study in relation to character awareness (Section 4.4.4, also cf. supp. video).

As explained in Section 3 and demonstrated in Section 4.3, collisions are solved to improve the physical correctness of crowd animations, nevertheless, editing the motion of each character can introduce other artifacts (*e.g.* unrealistic behaviors), or other violations of physical laws (*e.g.* unbalanced motions). In the range of examples we explored, especially the dancing (stationary) and walking (non-stationary) crowds of our perceptual study, the benefit of solving collisions is higher compared to potential drawbacks, as demonstrated by the results presented in Section 4.4.4. In particular, our results validate **H1**, showing that correcting residual collisions with our method leads to crowd animations perceived to display more natural contacts, higher character awareness, as well as perceived overall more realistic, both for stationary and non-stationary situations.

We can identify situations where our method would obviously fail to generate physically correct motions: for instance, if the initial collision volume is too large, or if the optimized pose parameters after solving Equation 3 significantly deviate from guiding full-body animations. Since our controller is very simple (*i.e.,* not designed to synthesize plausible motions), we need a good enough initial crowd animation to produce convincing results. Note that we could mitigate the possibly negative side effects of the method by incorporating more terms to the physics-based animation described in Section 3.2.

Performing the perceptual evaluation was important to demonstrate that the benefit of solving overlaps is higher than the risk of degrading motions in other aspects. Participants positively evaluated the overall realism of close-distance videos showing crowds of density up to 6.66 characters/m^2 which was not reported in any previous work we are aware of. Also, note that our perceptual study is based on pre-recorded videos for obvious practical reasons, but let us add that, technically, the same scene could have been generated in real-time after 3D characters model optimization given the performance of our method.

To the best of our knowledge, our method is the first to provide a general solution to remove large residual collisions between crowd characters. Note that our work does not aim at improving high-level strategy for collision avoidance, but it helps to achieve this goal because our characters react to local collisions, improving the overall realism of the crowd. The closest previous work is probably by Hoyet et al. [2016]. Our results are in line with their work: physically correcting motions, or re-introducing secondary motions as in previous work, has the same effect on the perceived behaviors of characters. They look more aware of their neighbors, and convey signs of interactions between them through their motion, even though inter-character interactions are not considered at the stage of animating their 3D bodies. However, compared to the previous work [Hoyet et al. 2016], we provide a more general solution to the long-standing problem of solving residual collisions in animated crowds, with no limitation to a specific kind of situation or to a specific type of motions to introduce (*e.g.* dedicated shoulder motions in walking crowds).

Moreover, we originally expected that our method would be limited in the density levels that could be handled, where higher levels of densities would lead to less natural contacts and overall realism. However, our perceptual study demonstrated that it is not the case (as **H2** was rejected): neither naturalness of contacts or overall realism were influenced by density when animations were corrected using our method, while both were perceived to be significantly higher than the same crowd animations where collisions were not solved. Our results therefore demonstrate that there is a positive

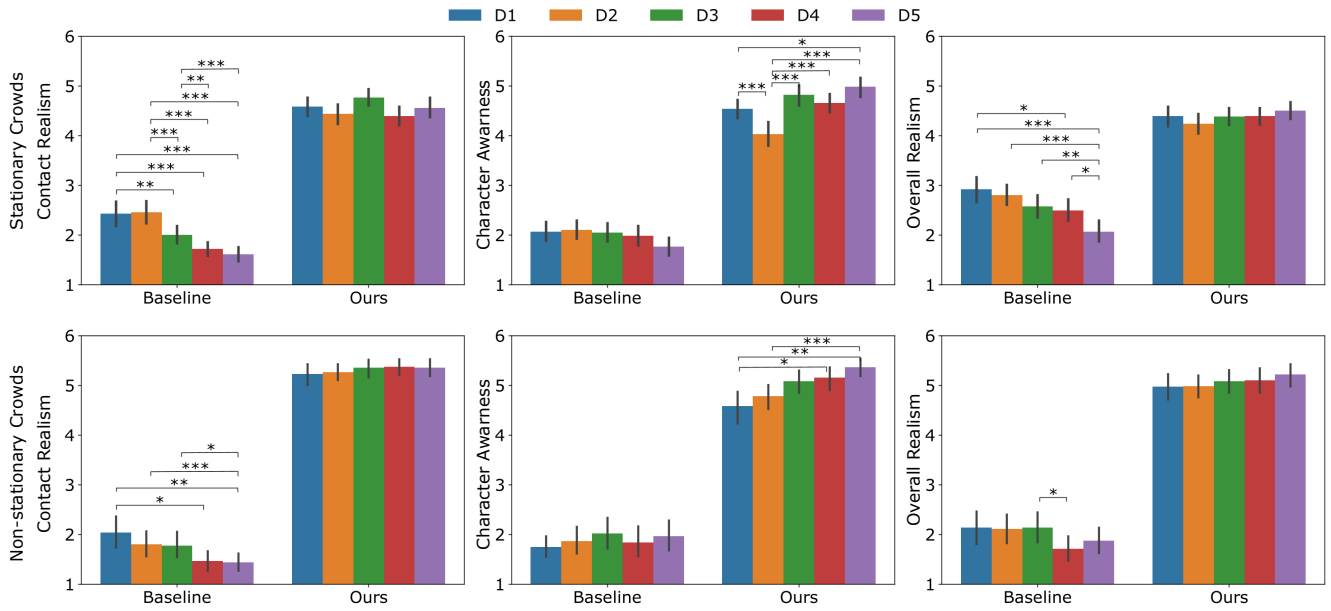

Fig. 9. Density × Method interaction effects for participant answers about Contact Realism (left), Character Awareness (center), and Overall Realism (right), without (Baseline) and with (Ours) correction of collisions in 3D crowd animations. Significant differences are displayed using * (p-value < 0.05), ** (< 0.01) and *** (< 0.001). For clarity purposes, differences between Baseline and Ours conditions are not displayed as they are all significant.

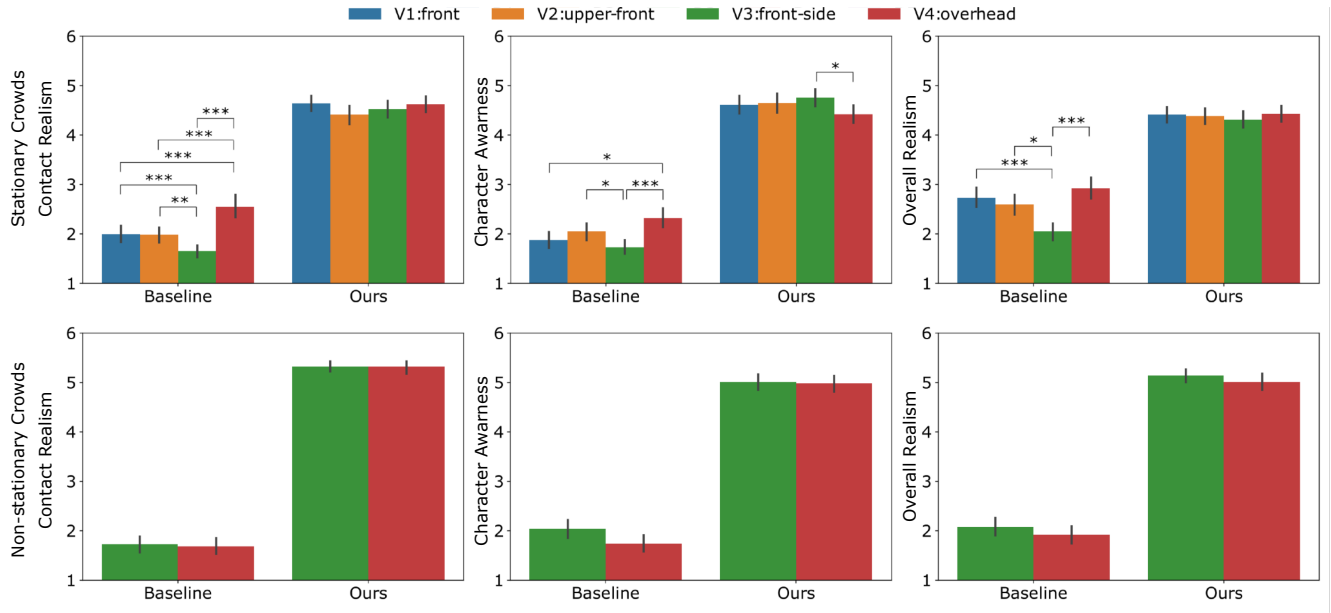

Fig. 10. Stationary Crowd Experiment (Top) Viewpoint × Method interaction effects for participant answers about Contact Realism (left), Character Awareness (center), and Overall Realism (right), without (Baseline) and with (Ours) correction of collisions in 3D crowd animations. While there was no such interaction effect for the Non-stationary Crowd Experiment (Bottom), the graphs are displayed for information. Significant differences are displayed using * (p-value < 0.05), ** (< 0.01) and *** (< 0.001). For clarity purposes, significant differences between Baseline and Ours conditions in the Stationary Crowd Experiment are not displayed as they do not contradict the found main effect of Method.

value in applying our method even at the high-density levels we considered. Nevertheless, it seems important to remind that the overall realism of a crowd animation as we explored it in our perceptual study comes from different components, *e.g.*, body shape, initial motion, appearance, etc. The main limitation of our study is that we did not explore the prevalence of large collisions over other sources of realism, which we leave for future work. Additionally, even though we demonstrated the benefits of our method in both stationary (dancing) and non-stationary (walking) situations, future evaluations would still be necessary to identify the specific differences related to other types of situations. For instance, it is possible that the results might not generalize as well in other types of walking crowds, for instance where characters come simultaneously from multiple directions. However, we believe that our method should generalize in such scenarios up to some extent and, more importantly, it paves the path to future research directions that are nowadays little explored due to the lack of efficient means for handling physical contact in these scenarios.

A final question left for future work is the possible coupling between different animation layers. We tested our method in a classical way where the navigation stage informs the characters' body animation stage, then proposed a novel additional step to finally solve residual collisions. We can easily imagine that the information we get about overlapping volumes could inform the former layers, for instance, to improve the navigation layer (navigation policy or parameters) or the body animation layer. This can be done, for example, in a reinforcement learning approach where the amount of interpenetration we calculate could be used as a metric to build a reward function. More generally, we facilitate the simulation of more complex behaviors in crowds: we lower the required quality of the animation resulting from the simulation of complex behaviors by letting our method improve its physical accuracy afterward.

6 CONCLUSION

We introduced a method for solving collisions in existing dense 3D crowd animations. We take a physics-based approach: intersecting volumes repulse one another to resolve overlapping parts, and body limbs are disentangled but left in contact to minimize the change with respect to the original motions. Beyond this technical contribution, our perceptual results contribute to the understanding of the perception of collisions in crowd animations and provide guidance for the application of our method.

Our results highlight a number of interesting properties of our method. Section 4.2, as well as the supplementary video, show that in different situations like a walking or dancing crowd, the method performs efficiently. Our approach is general not only because it can handle different situations, but also because it can consider animations generated in different ways: it acts as an additional motion editing layer that can be put last in any animation pipeline. In addition, the formulation of the solver coupled with the SMPL parametric body model makes the method ready for a large variety of characters. The method is fast and can handle dozens of characters in real-time. Also, Section 4.4 demonstrates the relevance of our method to significantly increase the visual quality of crowd animations. In a range from sparse to dense crowds, with few or many collisions, with little or large interpenetration volumes, the effect of the method on the visual quality of animations as well as on the realism of the behaviors of the characters is undeniable.

Our results also explore the limitations of the method, as the motion editing required to solve collisions can violate other physical rules or the plausibility of characters' behaviors. This is useful for us to sketch some directions for future work. First, we would like to explore new terms for the physically-driven animation approach proposed in Section 3.2, for instance, to avoid introducing violations of gravity or foot-sliding artifacts. Second, and equally important, we would like to explore virtual reality-based applications for our work. For instance, previous attempts to immerse people in dense crowds [Berton et al. 2020] used haptic feedback to limit the volume of collisions between a user and some moving characters. Users were made aware of collisions, and tended to limit them, however, since collisions were not propagated onto virtual characters they seemed to be unresponsive to contacts. The lack of methods like ours to solve such a problem was underlined by the authors, and given that our method runs in real-time we believe that it could be valuable for immersion in such virtual crowds. A final direction for future work would be to use the method in the context of a Reinforcement Learning approach. Information about overlapping volumes or the inability to correct motions could be used as a metric to design a reward function. This could be used to learn navigation and animation policies to generate high-density crowds, with plausible behaviors involving contacts between characters, which is clearly still an open challenge. Finally, we believe that sharing implementations of such methods is particularly important for our community, and will make our code available for the community upon acceptance, to be able to explore other applications or future research directions.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 899739 (H2020-FETOPEN-2018-2020 CrowdDNA project).

REFERENCES

2023. Mixamo. http://www.mixamo.com/. Accessed: 2024-06-23.
2024. Golaem. https://golaem.com/. Accessed: 2024-06-23.
2024. Massive. https://www.massivesoftware.com. Accessed: 2024-06-23.
2024. Miarmi – Crowd Simulation Solution for Maya. https://miarmy.com. Accessed: 2024-06-23.
Robin Adili, Benjamin Niay, Katja Zibrek, Anne-Hélène Olivier, Julien Pettré, and Ludovic Hoyet. 2021. Perception of Motion Variations in Large-Scale Virtual Human Crowds. In *Proceedings of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games*. 1–7.
Florian Berton, Fabien Grzeskowiak, Alexandre Bonneau, Alberto Jovane, Marco Aggravi, Ludovic Hoyet, Anne-Helene Olivier, Claudio Pacchierotti, and Julien Pettre. 2020. Crowd navigation in vr: exploring haptic rendering of collisions. *IEEE Transactions on Visualization and Computer Graphics* 28, 7 (2020), 2589–2601.
Steve Capell, Seth Green, Brian Curless, Tom Duchamp, and Zoran Popović. 2002. Interactive Skeleton-Driven Dynamic Deformations. *ACM transactions on graphics (TOG)* 21, 3 (2002), 586–593.
Miguel Ramos Carretero, Adam Qureshi, and Christopher Peters. 2014. Evaluating the Perception of Group Emotion from Full Body Movements in the Context of Virtual Crowds. In *Proceedings of the ACM Symposium on Applied Perception* (Vancouver, British Columbia, Canada) *(SAP '14)*. Association for Computing Machinery, New York, NY, USA, 7–14. https://doi.org/10.1145/2628257.2628266
Stelian Coros, Philippe Beaudoin, and Michiel Van de Panne. 2010. Generalized Biped Walking Control. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 29, 4 (2010).

Beatríz Cabrero Daniel, Ricardo Marques, Ludovic Hoyet, Julien Pettré, and Josep Blat. 2021. A perceptually-validated metric for crowd trajectory quality evaluation. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 4, 3 (2021), 1–18.

Cathy Ennis, Rachel McDonnell, and Carol O'Sullivan. 2010. Seeing is Believing: Body Motion Dominates in Multisensory Conversations. *ACM Transactions on Graphics* 29, 4 (2010). https://doi.org/10.1145/1778765.1778828

Anthony C Fang and Nancy S Pollard. 2003. Efficient synthesis of physically valid human motion. *Acm transactions on graphics (tog)* 22, 3 (2003), 417–426.

Andrew Feng, Dan Casas, and Ari Shapiro. 2015. Avatar Reshaping and Automatic Rigging Using a Deformable Model. In *Proc. of ACM SIGGRAPH Conference on Motion in Games (MIG)*. 57–64.

Zachary Ferguson, Minchen Li, Teseo Schneider, Francisca Gil-Ureta, Timothy Langlois, Chenfanfu Jiang, Denis Zorin, Danny M. Kaufman, and Daniele Panozzo. 2021. Intersection-Free Rigid Body Dynamics. *ACM Trans. Graph.* 40, 4, Article 183 (2021), 16 pages.

Nico Galoppo, Miguel A Otaduy, Serhat Tekin, Markus Gross, and Ming C Lin. 2007. Soft Articulated Characters with Fast Contact Handling. In *Computer Graphics Forum*, Vol. 26. 243–253.

Theodore F Gast, Craig Schroeder, Alexey Stomakhin, Chenfanfu Jiang, and Joseph M Teran. 2015. Optimization integrator for large time steps. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 21, 10 (2015), 1103–1115.

Benjamin Gilles, Guillaume Bousquet, Francois Faure, and Dinesh K Pai. 2011. Frame-based elastic models. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 30, 2 (2011).

Stephen Gustafson, Hemagiri Arumugam, Paul Kanyuk, and Michael Lorenzen. 2016. Mure: fast agent based crowd simulation for vfx and animation. In *ACM SIGGRAPH 2016 Talks*. 1–2.

Stephen J Guy, Jatin Chhugani, Sean Curtis, Pradeep Dubey, Ming C Lin, and Dinesh Manocha. 2010. PLEdestrians: A Least-Effort Approach to Crowd Simulation.. In *ACM/SIGGRAPH Symposium on Computer Animation*. 119–128.

Jessica K Hodgins, Wayne L Wooten, David C Brogan, and James F O'Brien. 1995. Animating human athletics. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 71–78.

Daniel Holden, Bang Chi Duong, Sayantan Datta, and Derek Nowrouzezahrai. 2019. Subspace neural physics: Fast data-driven interactive simulation. In *Proceedings of the 18th annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 1–12.

Ludovic Hoyet, Anne-Helene Olivier, Richard Kulpa, and Julien Pettré. 2016. Perceptual Effect of Shoulder Motions on Crowd Animations. *ACM Transaction on Graphics* 35, 4, Article 53 (2016). https://doi.org/10.1145/2897824.2925931

Hanbyul Joo, Tomas Simon, and Yaser Sheikh. 2018. Total Capture: A 3D Deformation Model for Tracking Faces, Hands, and Bodies. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*.

C. Kane, J. E. Marsden, M. Ortiz, and M. West. 2000. Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *Internat. J. Numer. Methods Engrg.* 49, 10 (2000), 1295–1325.

Paul Kanyuk. 2016. Virtual crowds in film and narrative media. *Simulating Heterogeneous Crowds with Interactive Behaviors* (2016), 217.

Mubbasir Kapadia, Nuria Pelechano, Jan M Allbeck, Norman I Badler, and Norm Badler. 2016. *Virtual crowds: Steps toward behavioral realism.* Springer.

Meekyoung Kim, Gerard Pons-Moll, Sergi Pujades, Seungbae Bang, Jinwook Kim, Michael J. Black, and Sung-Hee Lee. 2017. Data-Driven Physics for Human Soft Tissue Animation. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 36, 4 (2017).

Richard Kulpa, Anne-Hélène Olivierxs, Jan Ondřej, and Julien Pettré. 2011. Imperceptible Relaxation of Collision Avoidance Constraints in Virtual Crowds. *ACM Trans. Graph.* 30, 6 (dec 2011), 1–10. https://doi.org/10.1145/2070781.2024172

Marios Kyriakou and Yiorgos Chrysanthou. 2018. How Responsiveness, Group Membership and Gender Affect the Feeling of Presence in Immersive Virtual Environments Populated with Virtual Crowds. In *Proceedings of the 11th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Limassol, Cyprus) *(MIG '18)*. Association for Computing Machinery, New York, NY, USA, Article 12, 9 pages. https://doi.org/10.1145/3274247.3274509

Lei Lan, Ran Luo, Marco Fratarcangeli, Weiwei Xu, Huamin Wang, Xiaohu Guo, Junfeng Yao, and Yin Yang. 2020. Medial Elastics: Efficient and Collision-Ready Deformation via Medial Axis Transform. *ACM Trans. Graph.* 39, 3 (apr 2020). https://doi.org/10.1145/3384515

Marilena Lemonari, Rafael Blanco, Panayiotis Charalambous, Nuria Pelechano, Marios Avraamides, Julien Pettré, and Yiorgos Chrysanthou. 2022. Authoring Virtual Crowds: A Survey. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 677–701.

C Karen Liu and Zoran Popović. 2002. Synthesis of Complex Dynamic Character Motion from Simple Animations. *ACM Transactions on Graphics (TOG)* 21, 3 (2002), 408–416.

Libin Liu and Jessica Hodgins. 2017. Learning to Schedule Control Fragments for Physics-Based Characters Using Deep Q-Learning. *ACM Trans. Graph.* 36, 3 (2017).

Libin Liu and Jessica Hodgins. 2018. Learning Basketball Dribbling Skills Using Trajectory Optimization and Deep Reinforcement Learning. *ACM Trans. Graph.* 37, 4 (2018). https://doi.org/10.1145/3197517.3201315

Libin Liu, KangKang Yin, Michiel van de Panne, Tianjia Shao, and Weiwei Xu. 2010. Sampling-Based Contact-Rich Motion Control. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 29, 4 (2010). https://doi.org/10.1145/1778765.1778865

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

Adriano Macchietto, Victor Zordan, and Christian R. Shelton. 2009. Momentum Control for Balance. *ACM Trans. Graph.* 28, 3 (2009). https://doi.org/10.1145/1531326.1531386

Jonathan Maïm, Barbara Yersin, and Daniel Thalmann. 2009. Unique character instances for crowds. *IEEE Computer Graphics and Applications* 29, 6 (2009), 82–90.

Rachel McDonnell, Simon Dobbyn, Steven Collins, and Carol O'Sullivan. 2006. Perceptual evaluation of LOD clothing for virtual humans. In *Symposium on Computer Animation*. Citeseer, 117–126.

Rachel McDonnell, Michéal Larkin, Simon Dobbyn, Steven Collins, and Carol O'Sullivan. 2008. Clone Attack! Perception of Crowd Variety. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 27, 3 (2008). https://doi.org/10.1145/1360612.1360625

Uldarico Muico, Yongjoon Lee, Jovan Popović, and Zoran Popović. 2009. Contact-Aware Nonlinear Control of Dynamic Characters. *ACM Trans. Graph.* 28, 3 (2009). https://doi.org/10.1145/1531326.1531387

Sahil Narang, Andrew Best, and Dinesh Manocha. 2018. Simulating movement interactions between avatars & agents in virtual worlds using human motion constraints. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 9–16.

Sahil Narang, Andrew Best, Tanmay Randhavane, Ari Shapiro, and Dinesh Manocha. 2016. PedVR: Simulating Gaze-Based Interactions between a Real User and Virtual Crowds. In *Proc. of the ACM Conference on Virtual Reality Software and Technology*. 91–100. https://doi.org/10.1145/2993369.2993378

Nuria Pelechano, Bernhard Spanlang, and Alejandro Beacco. 2011. Avatar locomotion in crowd simulation. *International Journal of Virtual Reality* 10, 1 (2011), 13–19.

Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)* 37, 4 (2018), 1–14.

Pablo Ramon, Cristian Romero, Javier Tapia, and Miguel A. Otaduy. 2023. SFLSH: Shape-Dependent Soft-Flesh Avatars. In *SIGGRAPH Asia Conference Papers (SA Conference Papers '23)*. ACM. https://doi.org/10.1145/3610548.3618242

Cristian Romero, Dan Casas, Maurizio M Chiaramonte, and Miguel A Otaduy. 2022. Contact-Centric Deformation Learning. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–11.

Cristian Romero, Dan Casas, Jesús Pérez, and Miguel A. Otaduy. 2021. Learning Contact Corrections for Handle-Based Subspace Dynamics. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)* 40, 4 (2021).

Cristian Romero, Miguel A. Otaduy, Dan Casas, and Jesus Perez. 2020. Modeling and Estimation of Nonlinear Skin Mechanics for Animated Avatars. *Computer Graphics Forum (Proc. Eurographics)* 39, 2 (2020).

Shawn Singh, Mubbasir Kapadia, Glenn Reinman, and Petros Faloutsos. 2011. Footstep navigation for dynamic crowds. *Computer Animation and Virtual Worlds* 22, 2-3 (2011), 151–158.

Kwang Won Sok, Manmyung Kim, and Jehee Lee. 2007. Simulating Biped Behaviors from Human Motion Data. *ACM Trans. Graph.* 26, 3 (2007). https://doi.org/10.1145/1276377.1276511

Sebastian Starke, Yiwei Zhao, Taku Komura, and Kazi Zaman. 2020. Local Motion Phases for Learning Multi-Contact Character Movements. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 54–1.

G Keith Still. 2000. *Crowd Dynamics.* Ph. D. Dissertation. University of Warwick.

Sybren A Stüvel, Nadia Magnenat-Thalmann, Daniel Thalmann, A Frank Van Der Stappen, and Arjan Egges. 2016. Torso crowds. *IEEE transactions on visualization and computer graphics* 23, 7 (2016), 1823–1837.

Mankyu Sung, Lucas Kovar, and Michael Gleicher. 2005. Fast and accurate goal-directed motion synthesis for crowds. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*. 291–300.

Javier Tapia, Cristian Romero, Jesús Pérez, and Miguel A Otaduy. 2021. Parametric Skeletons with Reduced Soft-Tissue Deformations. *Computer Graphics Forum* 40, 6 (2021), 34–46.

Yun Teng, Mark Meyer, Tony DeRose, and Theodore Kim. 2015. Subspace Condensation: Full Space Adaptivity for Subspace Deformations. *ACM Trans. Graph.* 34, 4 (2015). https://doi.org/10.1145/2766904

Daniel Thalmann and Soraia Raupp Musse. 2012. *Crowd simulation.* Springer Science & Business Media.

Jur Van den Berg, Ming Lin, and Dinesh Manocha. 2008. Reciprocal velocity obstacles for real-time multi-agent navigation. In *IEEE International Conference on Robotics and Automation*. IEEEE, 1928–1935.

Wouter van Toll, Fabien Grzeskowiak, Axel López Gandía, Javad Amirian, Florian Berton, Julien Bruneau, Beatriz Cabrero Daniel, Alberto Jovane, and Julien Pettré. 2020. Generalized Microscropic Crowd Simulation Using Costs in Velocity Space. In

Symposium on Interactive 3D Graphics and Games. https://doi.org/10.1145/3384382.3384532

W. van Toll and J. Pettré. 2021. Algorithms for Microscopic Crowd Simulation: Advancements in the 2010s. *Computer Graphics Forum* 40, 2, 731–754.

Yu Wang, Alec Jacobson, Jernej Barbič, and Ladislav Kavan. 2015. Linear Subspace Design for Real-Time Shape Deformation. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 34, 4, Article 57 (2015). https://doi.org/10.1145/2766952

Andrew Witkin and Michael Kass. 1988. Spacetime Constraints. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '88)*. 159–168. https://doi.org/10.1145/54852.378507

Ming-Liang Xu, Hao Jiang, Xiao-Gang Jin, and Zhigang Deng. 2014. Crowd simulation and its applications: Recent advances. *Journal of Computer Science and Technology* 29, 5 (2014), 799–811.

Shanwen Yang, Tianrui Li, Xun Gong, Bo Peng, and Jie Hu. 2020. A review on crowd simulation and modeling. *Graphical Models* 111 (2020), 101081.

Xinran Yao, Shuning Wang, Wenxin Sun, He Wang, Yangjun Wang, and Xiaogang Jin. 2023. Crowd Simulation with Detailed Body Motion and Interaction. In *Advances in Computer Graphics: 39th Computer Graphics International Conference, CGI 2022, Virtual Event, September 12–16, 2022, Proceedings*. Springer, 227–238.

KangKang Yin, Kevin Loken, and Michiel van de Panne. 2007. SIMBICON: Simple Biped Locomotion Control. *ACM Trans. Graph.* 26, 3 (2007).