IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 12, 2020, accepted May 17, 2020, date of publication June 1, 2020, date of current version July 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2999059

A Spatio-Temporal Hierarchical Model for Crowd Formation Planning in Large-Scale Performance

YIHAO LI, TIANYU HUANG, YIFAN LIU, AND GANGYI DING
Beijing Institute of Technology, Beijing 100081, China
Key Laboratory of Digital Performance and Simulation Technology, Beijing 100081, China

Corresponding author: Tianyu Huang (huangtianyu@bit.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1403905, and in part by the Key Laboratory of Digital Performance and Simulation Technology, Beijing Institute of Technology.

ABSTRACT Crowd formation of aesthetic transformation is considered to have extremely high artistic value and is widely applied in large-scale performances. In this paper, a spatio-temporal hierarchical model that parts the crowd formation transform into multiple granularities is proposed. Its core idea is to add spatio-temporal constraints created by directors into transformation process after multi-level division. In this model, average hash value and energy optimization are used to achieve reasonable crowd formation arrangement, while smooth and collision-free formation transformations are presented by constrained region growth and Kuhn–Munkres algorithm. We have also proposed a framework to achieve the generation of visually pleasing crowd formation transform performance based on the constraints. Besides, a virtual crowd formation transformation simulation was built to verify the effect of the proposed model. Through simulation experiments and comparisons, it was demonstrated that this hierarchical model can generate aesthetic crowd formation transformation with a satisfactory process.

INDEX TERMS Crowd formation, crowd simulation, motion control, performance modeling.

I. INTRODUCTION

In recent years, crowd formation transformation has been increasingly used in mass performance rehearsal, unmanned aerial vehicles (UAV) formation control [1], tactical standing training in sports [2], etc. In addition to the rationality in the transformation process (smooth process, collision and obstacle avoidance, etc.), the artistic aesthetics process created by the complex crowd formation transformation has received more attention [3]. Especially in the field of large-scale outdoor gymnastic performance, the expending scale of the performance has become a common form in many occasions such as Olympic opening ceremony, National Day activities and so on. However, group planning of thousands of performers makes the traditional manual approaches of composing and directing no longer sufficient [4] and the performance with a massive crowd is difficult to model and decompose. In sum, the problems can be divided into two parts, respectively expressed in time and space. The first problem is how to organize static crowds with a large number of agents, which is related to space. The second problem, for time, is how to plan an efficient crowd formation transformation. These two problems are interdependent.

Existing researches related to crowd formation have described the crowd formation transform as matching and path planning process between two generated static crowd formation [4-8]. However, in a performance composed of crowd formation transformation, the goal is not only to achieve the transform from one pattern to another, but also to make sure the smoothness of the movement route and the synchronization of the overall pattern, highlighting the artistic rhythm in performance. Figure 1 shows an example of aesthetic crowd formation transformation with a large number of individuals. Moreover, the excessive segmentation of the process in order to achieve meticulous process was also limited by the calculation and storage bottlenecks. To our knowledge, there is a lack of research on the process of transformation and more considerable efforts are needed.

This paper focuses on crowd formation transformation in performance, which pays more attention to the delicate transformation process and performance aesthetics. In the proposed method, the average hash value of sequence frame is used in the data preparation and the generation of key formations is automated through energy optimization.

The associate editor coordinating the review of this manuscript and approving it for publication was Vladimir M. Mladenovic.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

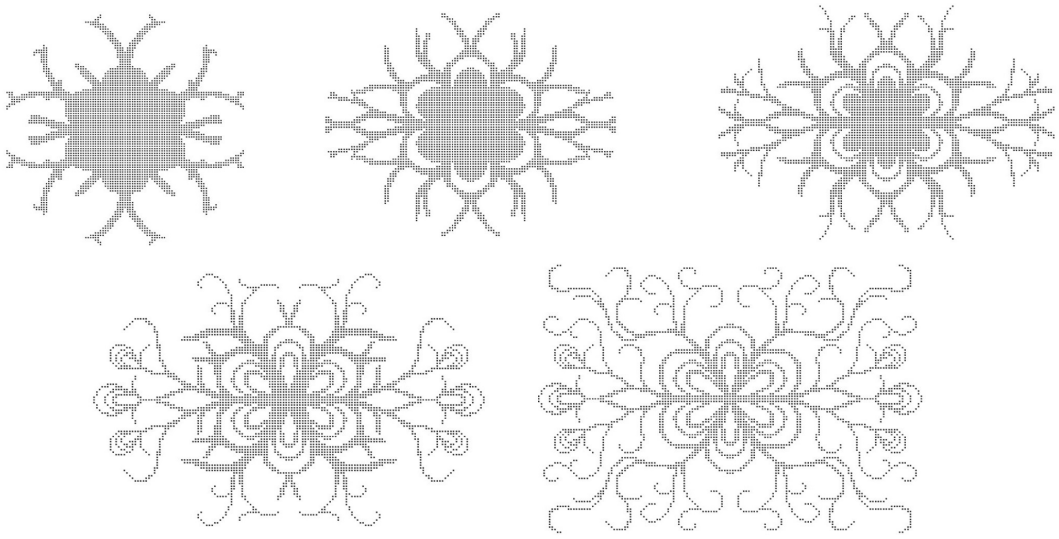

FIGURE 1. A virtual crowd formation transformation performance with thousands of agents.

Especially, the method implements a fine formation transformation through a hierarchical matching mechanism based on constrained region growth and Kuhn–Munkres algorithm. In addition, process simulation is used to fine-tune the crowd formation process, making the results more reasonable and satisfactory. This paper also proposed a fitness formula to evaluate the effect of crowd formation performance on directors' creativity. By comparing multiple spatial-temporal hierarchical parameters, the influence of different parameters on the model was also discussed.

The main contribution of this work can be summarized as follows:

- A process-based crowd planning model, focusing on the fine process of formation transformation, including an accurate and effective description of crowd formation planning;
- A multi-granularity spatial-temporal partitioning method that innovatively utilizes process association and spatial block partitioning to implement fine-grained crowd formation planning;
- A crowd formation planning framework which is used in a mass performance rehearsal.

The remainder of this paper is organized as follows: in Section II, we review the related work. In Section III, the structure of our model and implementation are introduced in details. Section IV presents the simulation environment and the method of evaluation. Section V is a validation of the proposed crowd formation planning model. Section VI is the conclusion of this work and a brief discussion of the future work.

II. RELATED WORK

In this section, we give a brief overview of prior work related to crowd modeling and formation transformation. The crowd formation transformation was developed based on crowd modeling and simulation.

Classic research about crowd simulation mainly focuses on the efficiency of the free crowd to reach the destination [9]–[11], in various scenarios such as sparse or dense crowds. Most of them divided the process of crowd simulation into two parts, global path planning and local obstacle avoiding [12]–[14], which is regardless of the artistic effect of crowd transform process. However, the crowd formation transformation in the performance focuses more on the aesthetic effect created by formation transform. It may need to impose certain behavioral constraints on the agents, which requires that while each agent travels through a reasonable route (no collision, as short as possible) the overall formation must always present an artistic aesthetic effect in the process. The imposing work either need nontrivial manual involvements or are focused on intuitive user interfaces for formation control and interaction [2], [15].

Many different approaches have been proposed for modeling the transform of crowd formation. Kwon *et al.* [16] proposed a framework to edit crowd motion as a whole and the user can deform a group by fixing or moving agents. A spectral-based approach [7] was proposed to control the group formation as well. These approaches required a lot of manual work to specify agents' distributions at a number of key frames, which is unable to adapt to a large performance scale. Gu and Deng [6] presented an effective freestyle group formation generation framework, with intuitive user interfaces. Xu *et al.* [3] proposed a method that can automatically generate visually pleasing crowd formation transformation. They discovered clusters of agent subgroups to extend their overall formation control.

In order to achieve aesthetic crowd formation, behavior constraints need to be imposed on all agents. Krontiris *et al.* [17] studied the physical constraints in formation simulation and their approach supports dynamic formations and transitions. Chang and Li [18] controlled the

shape template with a global motion planner and the shape constraints with a fuzzy controller. Xu *et al.* [19] proposed a novel shape-constrained flock animation system. They imposed hard constraints on agents' positions at specific time, retaining visual characteristics of flocking. Zheng *et al.* [20] proposed a geometry-constrained framework for smooth formation animation of regulated crowds. They employed the morphing method to generate a series of in-between constrained shapes, providing us with a reliable reference.

The study of group planning based on biomechanics is of heuristic significance to the formation planning of crowd as well. Guy *et al.* [21], [22] proposed the PLE principle, which minimized the effort of each agent by optimizing its energy-efficient, providing important ideas for the theory of crowd formation planning. For dense crowd simulation, an aggregate dynamics approach [10] was proposed to enable crowd simulation in effective computation time. Zhang *et al.* [23] proposed a method based on bionic intelligence algorithm and self-adaptive evaluation to generate special formation and implement crowd motion with concentric circle models and artificial bee colony algorithms, whose transformation process was scattered.

According to the existing research, we have found that most of the previous approaches simplified the formation transformation to the process from source formation to target formation, lacking of research on the fine-grained transformation process. In addition, in order to implement aesthetic transformation of the crowd, more clear and strict behavior rules should be applied to the agents, which is what we are mainly focusing on.

III. METHOD

In this section, we describe our approach of modeling crowd formation planning, computing the generation of crowd formation and planning the transform path.

A. CROWD FORMATION PLANNING MODEL

A complete crowd formation transformation includes both the overall formation pattern presentation and the position flow of each individual. Refer to traditional methods, the problem is firstly broken down into a series of continuous transformations between adjacent key formations(KF): $\{KF_0, KF_1, KF_2 \ldots\}$. The role of KF in the formation transformation is similar to the key frame in the animation [24]. This way of decomposition is very common in other methods.

In order to achieve a more visually pleasing transformation, the process between KFs in the proposed model is further divided in time and space:

$$P_i = \{p_t^s | p_t^s \in p_t, p_t \in (KF_i, KF_{i+1})\} \tag{1}$$

P_i represents the set of subdivisions in time and space in the intermedia process between KF_i to KF_{i+1}. Subscript t represents the time dimension and superscript s is the spatial dimension. In the proposed model, the time dimension is divided at first, and then is the spatial dimension, which means the planning priority of agents in space is higher than time (illustrated in Figure 2).

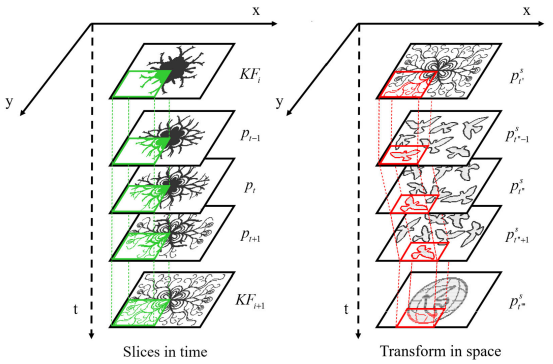

FIGURE 2. The spatial-temporal division process of the crowd formation.

Intermedia process p_t is defined as follows:

$$p_t = f_t(GRID, CSTR, NUM, PAT, MAP) \tag{2}$$

GRID represents the grid where agents stand on, which could be squares grid, triangles grid, hexagons grid and so on. For the performer in practice, the grid is a basic condition for them to move in a performance.

CSTR is the constraints that actor needs to observe during the transformation process, such as moving within the grid area; at any p_t, the actor must remain in the middle of the grid and so on.

NUM represents the number of agents and is always a constant.

PAT is a set,containing the distribution pattern composed of all agents, which form a static crowd formation. *PAT* is order independent, so it only contains the static formation information.

MAP is a perfect match set, containing matches for all agents in $PAT(p_t)$ and $PAT(p_{t+1})$.

In this model, all agents move according to a serial of p_t, with the rhythm of the performance to achieve an aesthetic crowd formation transform performance. In the proposed method, the collision problem was naturally avoided and the additional collision avoidance method was not focused like other scenarios.

The input data in our approach is monochrome video, which includes the complete transformation element. While the output data is *PAT* and *MAP*, where *PAT* includes the overall arrangement of the crowd, and *MAP* guides each agent to make smooth and collision-free formation transformation. *GRID*, *CRST* and *NUM* as the basis of the model is constant for every p_t in a certain application case.

B. DATA PREPARATION

As the input data is video and it is not reasonable to generate crowd formation for each frame of the video. Therefore, video content needs to be sampled as the basis for generating crowd formation. The sample interval (SI) is dynamically changing during sampling according to different art design or rhythm.

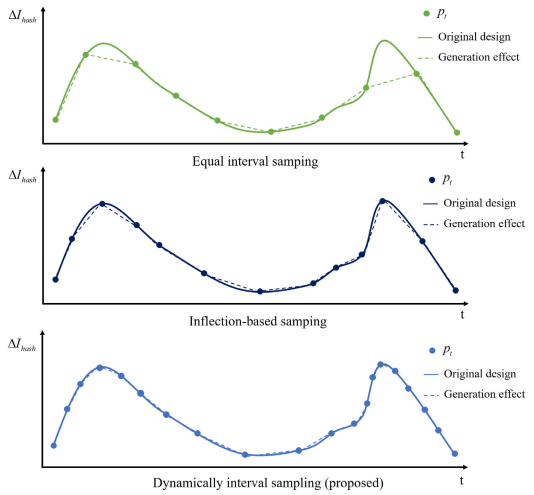

FIGURE 3. Comparison of original design and generated result with different sampling strategies.

It is defined as

$$SI = min(I_{hash}, c) \tag{3}$$

where constant c is the minimum sampling interval. I_{hash} is calculated based on the average hash value of each frame:

$$I_{hash} = |Num_{hash} - Num_{last}| \tag{4}$$

In Equation (4), Num_{last} is the number of last chosen frame. Num_{AHV} is the number of first frame after Num_{last} that the deviation of the average hash value between that frame and last chosen frame reaches the threshold T_{hash}. AHash algorithm provided in [25] was used in this approach.

In the proposed method, c and T_{hash} are two main parameters to control the selection of key frames, among which T_{hash} is used to control the high frequency part, while c is used to control the low frequency part. For T_{hash}, it mainly measures the degree of change in overall formation, and is sensitive to the data that changes rapidly. In contrast, c restricts the maximum sampling interval to ensure that slow and small changes can be captured. Continuously computing dynamic SI and sampling the input video can provide efficient data preparation. For the input data, One key frame corresponds to one p_t. Figure 3 shows the difference between three different sampling strategies. Dynamical interval sampling method could provide higher fitness between original design and generation result, compared with the other two traditional methods.

C. CROWD FORMATION GENERATION

The next step is to generate the crowd distribution pattern PAT based on key frame, mentioned in the previous subsection. The generation of the static formation needs to be under constraint $CRST(p_t)$. In addition to all agents being distributed on the grid, the number of agents also needs to be equal to $NUM(p_t)$, which is the target benchmark of the energy optimization algorithm. The generation of target formation is described as follows:

$$(PAT, NUM) = f(\delta_{offset}, T_g) \tag{5}$$

In Equation (5), δ_{offset} is defined as the position offset between $GRID(p_t)$ and the key frame. For convenience, the side length of the grid is set to 1. In the proposed model, as the discussion of ranges is in two dimensions, δ_{offset} is a two-dimensional vector and its range in two dimensions are both in $[-0.5, 0.5]$. T_g is the threshold of the ratio of the positive samples of the key frame to the corresponding grid. $T_g \in (0, 1]$. The generation function $f(\delta_{offset}, T_g)$ is mainly a mask operation. For any grid, if the positive area rate r_{pa} is more than T_g, the grid will be selected to place an agent.

As the output number of indivisuals is required to equal to $NUM(p_t)$, the generation of target PAT is converted to solve the following optimization problem:

$$min|NUM(f(\delta_{offset}, T_g)) - NUM(p_t)| \tag{6}$$

Method of steepest descent [26] is used to implement the energy optimization. The optimized search space is the value range of δ_{offset} and T_g. As the mask operation is computational inexpensive and δ_{offset} and T_g are in a relatively wide interval, the steepest descent algorithm can quickly converge.

In the two independent variables, δ_{offset} controls the overall large-scale adjustments, while T_g adjusts the results in details. Furthermore, δ_{offset} determines the overall shape of the formation and T_g influences the arrangement of the border agents. As a result, during iteration of gradient descent, mainly iterate the value of δ_{offset} to make the result close to the target value quickly in the early stage, and iterate the value of T_g to make the result converge to the target value in the later stage.

D. HIERARCHICAL FORMATION PLANNING

As PAT is disordered, the next step is to establish mapping relationship in order to connect adjacent key formations. Firstly, define all agents adjacent in the grid as neighbors to each other. To ensure the rationality of the partition, select the seed agent from the boundary of p_t, and grow the connected domain by searching the neighborhood to obtain the agent set p_t^s:

$$p_t^s = \{agt_i | agt_i \in PAT(p_t), i < T_n, \delta < T_v\} \tag{7}$$

For any set p_t^s, there is anchor $A = \frac{1}{S}\sum_{agt \in p_t^s} pos(agt)$. In Equation (7), $\delta_a = \frac{1}{S}\sum_{agt \in p_t^s} |pos(agt) - A|^2$, agent number threshold T_n and variance threshold T_v are constraints to control the growth of p_t^s. Figure 4 shows an example of dividing p_t^s and matching it. In the figure, the left part and the the right part are all divided into 7 p_t^ss based on region growing.

For any set p_t^s, the above method could be used to divide it again and again, which means that the hierarchy can be infinitely divided in theory. For different sets within the same division level between p_t and p_{t+1}, they compose a complete bipartite graph with anchors of the sets $\{p_t^1, p_t^2, p_t^3 \ldots\}$ and $\{p_{t+1}^1, p_{t+1}^2, p_{t+1}^3 \ldots\}$. Use Eclidean distance between

FIGURE 4. Division of p_t^s in crowds and matching result.

vertices as edge weights in the graph and apply the Kuhn-Munkres (KM) algorithm [27], [28] to obtain the matching results between the various level sets. As presented in the PLE theory [21], the minimum weight results obtained by the KM algorithm are reasonable to apply in group planning. Due to the characteristics of the algorithm, most of the invalid situations (illustrated in Figure 5) are naturally avoided.

Algorithm 1 Multi-Layered Optimal Match Searching

Input: two pattern PAT_i and PAT_{i+1}
Output: optimal match MAP_i

```
while matched agents < total agents do
    divide p_t into a list of p_t^s in both PAT_i and PAT_{i+1}
    calculate anchor in each p_t^s, get anchor sets s_a1, s_a2
    calculate optimal match between s_a1 and s_a2
    if length(s_a1) = length(s_a2) = 1 then
        MAP_i ← MAP_i + match result
        end recursion
    else
        continue to recursively divide and match
    end if
end while
```

The model is multi-layered, and through layer-by-layer matching, the agent obtains an optimal match with certain constraints, which similarly is mentioned in [20] and is the biggest difference between our method and the others. The constraints from hierarchical matching make the path planning of whole crowd not as short as other methods that apply KM algorithm directly on all agents, but the sacrifice in overall optimization is in exchange for higher transformation coherence and the aesthetic in formation transform.

However, the optimal matching is not global and not covering the agents in different sets in the same level. As a result, there are still a lot of conflict at the junction between sets in the same level. Lightweight simulation of the process by linearly interpolating $PAT(p_i)$ and $PAT(p_{i+1})$ according to $MAP(p_i)$ is used to solve this problem. Reverse the mapping relationship in $MAP(p_i)$ corresponding to all agents that have path conflicts in the simulation process (illustrated in Figure 6).

Different from other methods, the proposed method focuses on the formation transformation research on two dimensions ground without obstacles. The meticulous process created by constrained region growth and KM algorithm

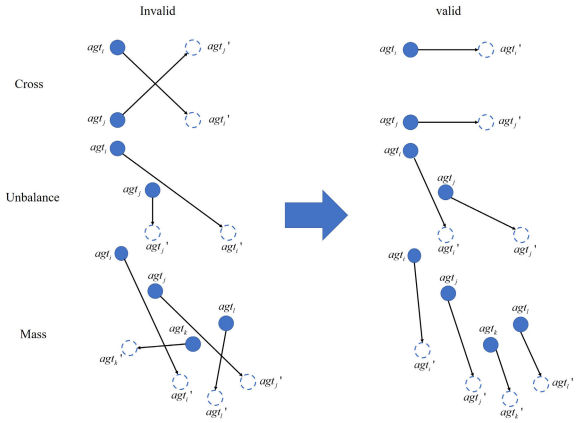

FIGURE 5. Invalid situations in formation transformation planning.

Algorithm 2 Optimal Matching Process

Input: two sets of agents, s_1 and s_2
Output: match result

```
create a bipartite graph with two sets of agents
calculate Euclidean distance as edge cost in the graph
set label = max(edgecost) to every agent in s_1 and
label = 0 to every agent in s_2
while current match != perfect match do
    find the largest match by the Hungarian algorithm
    if current match != perfect match then
        adjust the value of labels
    end if
end while
```

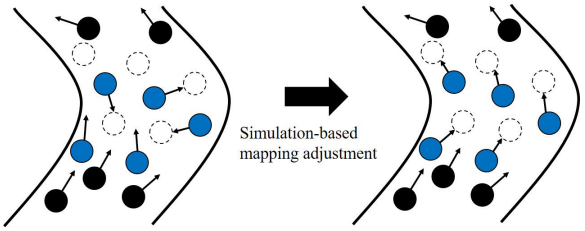

FIGURE 6. The process of mapping adjustment at the junction.

avoid collision between agents, which means there is no need to calculate through collision avoiding algorithms with high computational consumption such as RVO [29] and so on.

IV. EVALUATION

A. DATA-DRIVEN QUANTIFICATION

Since the computational cost in the simulation process is saved by giving a large number of intermediate processes, the simulation result can be real-time (more than 60 FPS) with the number of simulated agents under 1 million. The main calculation cost is in the generation process of P_i, including the generation of PAT and MAP for each p_t. Table 1 shows the calculation time and memory usage for generating unit results. As it shows, the computation consumption increases as NUM increases. As the generation process is automated,

TABLE 1. The calculation time and memory usage of our method for the same sample.

NUM	100	330	1000	3290	5000
Calculation time(s)	<0.1	0.41	2.75	18.92	64.09
Memory(MB)	<1	2	4	41	97

the generation time of a complete formation transform performance of 3290 actors for 8 minutes in this framework is less than 2 hours, which may cost more than 3 months in the past in manual work, according to an experienced crowd performing director.

B. FITNESS

In order to quantify the effect of our hierarchical model, the fitness of crowd formation transformation to the director's intentions is introduced as ρ_{fit}. ρ_{fit} reflects how well an actor's movement in the overall pattern fits the original idea from the directors. The calculation method of ρ_{fit} is similar to r_{pa} in Section III, part C, and the difference is that for ρ_{fit}, the fitness is not between single grid and key frame, but a certain area presented by the agent and a frame in the video. ρ_{fit} represents the fitness of single agent to a certain frame in the input video and $\rho_{fit} \in [0, 1]$.

$$\theta_{dev} = 1 - \frac{1}{NUM} \sum_{j=1}^{NUM} \rho_{fit}(j) \tag{8}$$

In Equation (8), θ_{dev} is the deviation set of rho_{fit} and it represents the deviation between $DIST(p_t)$ and the frame in the input video. $\theta_{dev} \in [0, 1]$ as well.

C. FINE ADJUSTMENT

Although we have achieved a highly automated workflow, manual inspection and modification of the results is still necessary, because it is difficult to completely present the creative details from the arranger based on the flowing crowd formation generated by the video content, although the workload is very small. With minor manual modification, the results may be further improved. As a result, we take the late manual modification and feedback of the performance arrangers as an important part to qualify the effect of our method, which is shown in the next section.

V. RESULT

Our approach was implemented to a software framework in C++. A complete simulation and rehearsal of a complete gymnastic performance was realized through the framework. In this work, 3290 actors performed smooth and artistically aesthetics formation transformation through synchronized position flow during the performance. Actually, in order to make the flow path of all actors consisting of a synchronous changing pattern, the movement path of actors during the performance needed to be well planned and reasonably rehearsed, which is a good application scenario for our framework.

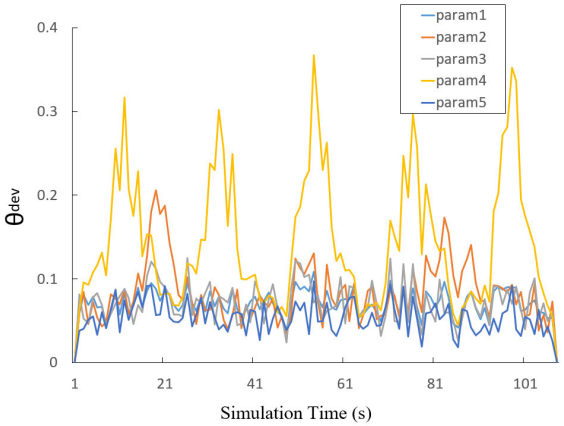

FIGURE 7. Comparison of results in different c and T_{hash}.

In order to verify the effectiveness of our approach, the result comparisons were performed under different granularities both in time and space. All our experiments were performed on a computer with Intel Core i7-7700 3.6GHz CPU and the software environment is Windows10 operating system.

The presentation of the results is mainly in three aspects: results in different time interval division strategy, results in different spatial layering strategy and results in interview and feedback. Besides, a crowd formation transformation simulation was conducted to verify the running effect of the hierarchical model in the case of multiple virtual characters.

A. RESULTS IN DIFFERENT TIME INTERVAL DIVISION STRATEGIES

Recalling the contents in Section III, part A, the key frames are selected in the video by two parameters, c and T_{hash}, of which c is for minor change and T_{hash} is for drastic change. In the experiment, the same spatial layering strategy were applied and the only difference was the values of c and T_{hash}. All results correspond to the same piece of performance, so they have the same simulation time. The results are shown in Figure 7.

θ_{dev} introduced in Section IV, part B is used to evaluate the results of the two parameters. The parameter correspondence in Figure 7 is shown in Table 2. As the results shows, both for c and T_{hash}, smaller intervals provide better results. For parameter 2, there are two intervals with larger θ_{dev}, which corresponds to the parts where morphing of overall pattern is relatively gentle in the performance. For parameter 4, it clearly shows a periodic increase of deviation. The fewest key frames are also selected in that group. Too few key frames lead to loss of process details, thus leading to the high deviation rate. The reason for periodicity is that the trough corresponds to the generated key formation, so the value of θ_{dev} is low. To the peak, the interval between two adjacent p_t is too long, and the deviation is high in the middle. For all the experiment groups, the zigzags in Figure 7 are because the crowd formation transformation cannot totally

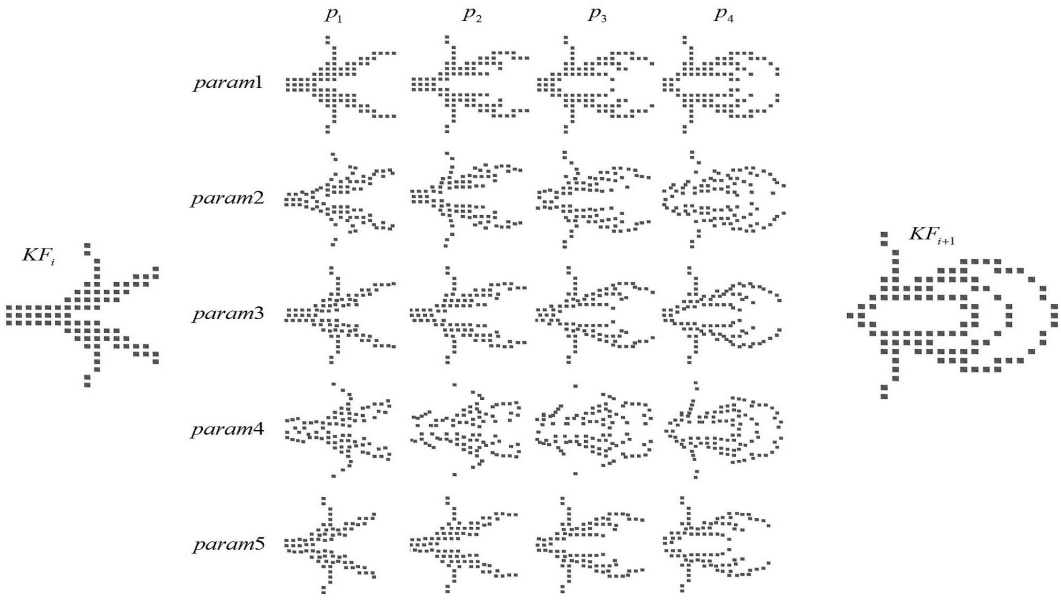

FIGURE 8. Comparison of formation transform processes under 5 different sampling parameters.

fit the complex changing pattern and the drastic change led to numerical changes of larger lead to more obvious zigzag.

Actually, in extreme case, if $c = 1$ or $T_{hash} = 0$, there will always be $SI = 1$, and the number of p_t will be as many as the number of frames in the input video. In this case, though the result may be better, the computational consumption would be unacceptable. In practice, by testing and judging over 120000 frames, it was found when c is between 80 and 90, T_{hash} is between 5% and 15%, the key frame sampled by SI can ensure both effective formation transformation and unnecessary calculation avoidance.

Figure 8 shows the formation transform processes under 5 different sampling parameters. It extracts the sub-process of grass-shaped formation changing into a peacock pattern in the complete experiment, including around 100 agents. As the figure shows, group 1, group 3 and group 5 produced great transformation effect, corresponding to more process segmentation in temporal dimension. In contrast, group 2 and group 4 show the results of rough process segmentation, which is similar to the method of directly matching [3]. The lack of details in the formation transform process showed in these groups is what the method was trying to avoid.

B. RESULTS IN DIFFERENT SPATIAL LAYERING STRATEGIES

In order to control the formation transformation process accurately and effectively, the hierarchical model was proposed. In the experiment, the same time interval division strategies were applied to test the effect of different spatial layering strategies, including a non-hierarchical model(global matching based on KM algorithm [3], [5]) and the proposed hierarchical model with two layers and three layers. Same as Section V, part A, all results of this experiment correspond to the same performance. In order to obtain accurate results,

TABLE 2. The values of c and T_{hash} for every parameter pair in Figure 7 and the average value of θ_{dev} for each results.

	c	T_{hash}	Key frames	Average $\theta_{dev}(\%)$
param 1	80	0.1	110	7.039
param 2	160	0.1	84	8.240
param 3	40	0.1	132	7.095
param 4	80	0.2	47	14.26
param 5	80	0.05	266	6.377

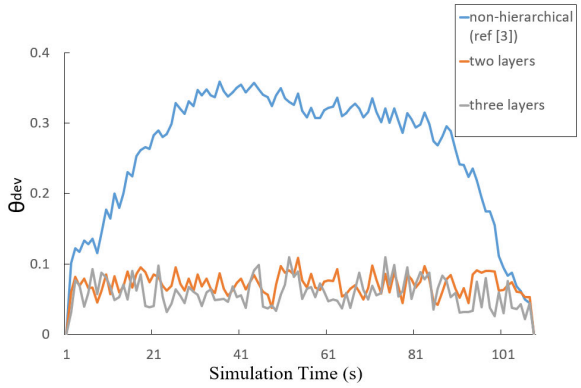

FIGURE 9. Comparison of results in different hierarchical strategies.

the experiment performance was composed of 3290 agents, which was a big number.

Results are shown in Figure 9. θ_{dev} is also used to evaluate the results. Obviously, compared with non-hierarchical model, the proposed method has a great advantage in the process presentation, reflected as stable low deviation. Like Figure 7, there are also zigzags in Figure 9 and the cause is the same. Although our method takes more upfront calculation time, which mainly cost on adding process details, effective reduction of θ_{dev} is obtained and the calculation is acceptable.

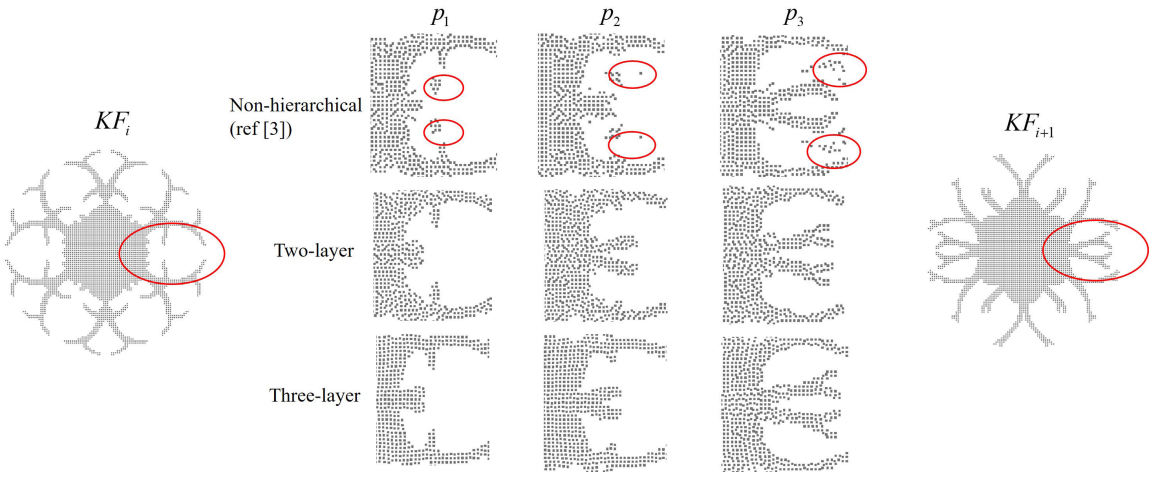

FIGURE 10. Comparison of three spatial hierarchical strategy results.

TABLE 3. Comparison of calculation time, average and variance of θ_{dev} in different hierarchical strategies.

	Calculation time(s)	Average of $\theta_{dev}(\%)$	Variance of θ_{dev}
Non-hierarchical(ref [3])	13.26	26.12	0.0084
Two-layer	29.98	7.039	0.00029
Three-layer	69.65	5.224	0.00081

Meanwhile, it also shows when the layer is over 2, the improvement of effect is limited, and it will cause much longer calculation time consumption, as shown in Table 3. It is worth noting that the result of two-layer model has less variance of θ_{dev} than three-layer model, which explains that excessive hierarchy may reduce the stability of the results due to overfitting.

In Figure 10, a special example was given to show the difference between different hierarchical strategies. As the figure shows, there is a sub-process of vine growth and it shows the details during the formation transformation. For non-hierarchical group, each agent chooses the path with the lowest total distance. As a result, some of the agents cross the outline and destroy the overall pattern, as shown by the red circled agents. This is a major problem for other methods. For the instances with two layers or three layers, because of reasonable temporal slicing and spatial partitioning, the crowd maintains the overall pattern and each agent performs collision-free moving during the formation transformation.

C. RESULTS IN INTERVIEW AND FEEDBACK

Recall in Section IV, part C, manual work and feedback as the next step of our work can provide an effective evaluation for our work. As already mentioned, we have worked on a mass gymnastics performance with more than 50 experienced professional performance directors and we get feedback data from the working with them.

To analysis the influence of directors' work experience on the results, the feedback results from directors were divided according to the number of years of work. Meanwhile, the evaluation index were set as the following five items:

TABLE 4. Feedback from directors of different work experience on the results of the framework.

item \ experience	0-5	5-10	10-15	15-20	>20
$ESKF$	8.4	9.2	8.0	8.2	8.4
$EPFT$	9.6	9.5	10.0	9.4	9.8
EOP	9.3	9.0	8.8	9.0	9.1
$MWSKF(\%)$	0.6	2.1	3.4	4.9	4.4
$MWPFT(\%)$	0.0	0.0	0.1	0.4	0.4

- evaluation of static key formations ($ESKF$)
- evaluation of the process of formation transform ($EPFT$)
- evaluation of the overall performance (EOP)
- the manual work on static key formations ($MWSKF$)
- the manual work on the process of formation transform ($MWPFT$)

The first three items are scored by directors, and the score ranges from 0 to 10. The last two items are calculated from the proportion of the modified part, whose value ranges from 0% to 100%.

From the feedback results, our work has received a relatively high evaluation, where the results of $EPFT$ and $MWPFT$ shows that to the generation of formation transform, our work can basically replace manual work. The results of $ESKF$ and MWSKF shows that our approach needs to be improved in the generation of static key formation, which is mainly related to the abstraction and expression of directors' ideas. In addition, especially, for $MWSKF$ and $MWPFT$, results show a clear trend of growth with work experience. According to the interview with the director, experienced directors are more likely to be influenced by traditional manual work methods and it has a negative impact on the use of new automated work models.

VI. DISCUSSION

In this paper, we have presented a novel crowd formation transform approach that mainly includes crowd arrangement generation method, hierarchical formation planning method,

fitness measure method and so on. This approach results in process-smooth and aesthetics crowd formation transformation and can be widely used in performance simulation and arrangement assistant. In addition, we propose a new formulation to evaluate the fitness of crowd formation transformation to the creator's intention.

Through a serial of experiments, we evaluate the performance of our approach under different parameters both in time and space. It turns out that our approach is effective in transforming crowd formations, especially in the field of performances containing a large number of actors and complex crowd planning.

Our present work focuses on the hierarchical model that the time dimension has higher priority. In the future, our search will mainly extend on the decoupling of the spatiotemporal correlation of the crowd formation transformation, which will lead to a more flexible formation and transform generation.

ACKNOWLEDGEMENTS

The authors thank the designers who provided them with beautiful videos.

REFERENCES

[1] X. Dong, B. Yu, Z. Shi, and Y. Zhong, "Time-varying formation control for unmanned aerial vehicles: Theories and applications," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 1, pp. 340–348, Jan. 2015.

[2] J. Ren, X. Jin, and Z. Deng, "Crowd formation generation and control," in *Handbook of Human Motion*. Cham, Switzerland: Springer, 2017, pp. 1–12.

[3] M. Xu, Y. Wu, Y. Ye, I. Farkas, H. Jiang, and Z. Deng, "Collective crowd formation transform with mutual information-based runtime feedback," *Comput. Graph. Forum*, vol. 34, no. 1, pp. 60–73, Feb. 2015.

[4] G. Ding, H. Li, T. Huang, L. Zhang, and G. Jing, "A prototype system for virtual composing and directing large-scale outdoor artistic performance," *J. Syst. Simul.*, vol. 20, no. S1, pp. 136–140, Sep. 2008.

[5] M. Xu, Y. Wu, and Y. Ye, "Smooth and efficient crowd transformation," in *Proc. 20th ACM Int. Conf. Multimedia MM*, 2012, pp. 1189–1192, doi: 10.1145/2393347.2396415.

[6] Q. Gu and Z. Deng, "Generating freestyle group formations in agent-based crowd simulations," *IEEE Comput. Graph. Appl.*, vol. 32, no. 1, pp. 20–31, Jan./Feb. 2013.

[7] S. Takahashi, K. Yoshida, T. Kwon, K. H. Lee, J. Lee, and S. Y. Shin, "Spectral-based group formation control," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 639–648, Apr. 2009.

[8] Q. Gu and Z. Deng, "Formation sketching: An approach to stylize groups in crowd simulation," in *Proc. Graph. Interface Conf.*, 2011, pp. 1–8.

[9] A. Golas, R. Narain, S. Curtis, and M. C. Lin, "Hybrid long-range collision avoidance for crowd simulation," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 7, pp. 1022–1034, Jul. 2014.

[10] R. Narain, A. Golas, S. Curtis, and M. C. Lin, "Aggregate dynamics for dense crowd simulation," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–8, Dec. 2009.

[11] M. Zhao, J. Zhong, and W. Cai, "A role-dependent data-driven approach for high density crowd behavior modeling," in *Proc. Annu. ACM Conf. SIGSIM Princ. Adv. Discrete Simulation SIGSIM-PADS*, 2016, pp. 89–97, doi: 10.1145/2901378.2901382.

[12] L. Yan, Z. Jing, and L. V. Min, "Motion simulation method for virtual crowd based on video," *Comput. Eng.*, vol. 41, no. 4, pp. 222–227 and 231, Apr. 2015.

[13] L. Sun, X. Sun, W. Qin, L. Sun, X. Sun, and W. Qin, "Dense crowd simulation based on continuum model and aggregate dynamics model," *Chin. J. Comput.*, vol. 39, no. 7, pp. 1275–1392, 2016.

[14] H. Seo, L. Yahiacherif, T. Goto, and N. Magnenat-Thalmann, "Genesis: Generation of e-population based on statistical information," in *Proc. Comput. Animation (CA)*,Jun. 2002, pp. 81–85, doi: 10.1109/CA.2002.1017511.

[15] M.-L. Xu, H. Jiang, X.-G. Jin, and Z. Deng, "Crowd simulation and its applications: Recent advances," *J. Comput. Sci. Technol.*, vol. 29, no. 5, pp. 799–811, Sep. 2014.

[16] T. Kwon, K. H. Lee, J. Lee, and S. Takahashi, "Group motion editing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–8, Aug. 2008, doi: 10.1145/1360612.1360679.

[17] A. Krontiris, S. Louis, and K. Bekris, "Simulating formations of nonholonomic systems with control limits along curvilinear coordinates," in *Proc. Int. Conf. Motion Games*, Nov. 2010, pp. 121–133.

[18] J.-Y. Chang and T.-Y. Li, "Simulating virtual crowd with fuzzy logics and motion planning for shape template," in *Proc. IEEE Conf. Cybern. Intell. Syst.*, Sep. 2008, pp. 131–136, doi: 10.1109/ICCIS.2008.4670865.

[19] J. Xu, X. Jin, Y. Yu, T. Shen, and M. Zhou, "Shape-constrained flock animation," *Comput. Animation Virtual Worlds*, vol. 19, nos. 3–4, pp. 319–330, 2008.

[20] L. Zheng, J. Zhao, Y. Cheng, H. Chen, X. Liu, and W. Wang, "Geometry-constrained crowd formation animation," *Comput. Graph.*, vol. 38, pp. 268–276, Feb. 2014.

[21] S. Guy, J. Chhugani, S. Curtis, P. Dubey, and D. Manocha, "PLEdestrians: A least-effort approach to crowd simulation," *Proc. Eurographics/ACM SIGGRAPH Symp. Comput. Animation*, 2010, pp. 119–128, doi: 10.2312/SCA/SCA10/119-128.

[22] S. J. Guy, S. Curtis, M. C. Lin, and D. Manocha, "Least-effort trajectories lead to emergent crowd behaviors," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 85, no. 1, Jan. 2012, doi: 10.1103/PhysRevE.85.016110.

[23] P. Zhang, H. Liu, and Y.-H. Ding, "Crowd simulation based on constrained and controlled group formation," *Vis. Comput.*, vol. 31, no. 1, pp. 5–18, 2013.

[24] N. Burtnyk and M. Wein, "Computer-generated key-frame animation," *J. SMPTE*, vol. 80, no. 3, pp. 149–153, Mar. 1971.

[25] J. Huang, X. Li, B. Chen, and D. Yang, "A comparative study on image similarity algorithms based on hash," *J. Dali Univ.*, vol. 2, no. 12, pp. 32–37, Dec. 2017.

[26] S. T. Alexander, "The method of steepest descent," in *Adaptive Signal Processing*. New York, NY, USA: Springer, 1986, doi: 10.1007/978-1-4612-4978-8_4.

[27] H. W. Kuhn, "The hungarian method for the assignment problem," *Nav. Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.

[28] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, Mar. 1957.

[29] J. van den Berg, M. Lin, and D. Manocha, "Reciprocal velocity obstacles for real-time multi-agent navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2008, pp. 1928–1935.

YIHAO LI received the B.E. degree from the Beijing Institute of Technology, China, in 2017, where he is currently pursuing the master's degree with the School of Computer Science and Technology.

His research interests include crowd simulation and digital performance.

TIANYU HUANG received the B.E. degree in computer science from Jilin University, China, in 2002, and the Ph.D. degree in computer science from the Beijing Institute of Technology, China, in 2007.

She joined the Faculty with the Beijing Institute of Technology, in April 2007. From 2012 to 2013, she worked as a Visiting Scholar with the Department of Computer and Information Science, University of Pennsylvania, USA. She is currently an Associate Professor with the School of Computer Science and Technology, Beijing Institute of Technology. Her research interests include human motion modeling and simulation, crowds simulation, and virtual reality.

YIFAN LIU received the B.E. degree from Shandong University, China, in 2019. She is currently pursuing the master's degree with the School of Computer Science and Technology, Beijing Institute of Technology, China.

Her research interests include computer simulation and virtual reality.

GANGYI DING received the B.E. degree from Peking University, China, in 1988, and the Ph.D. degree from the Beijing Institute of Technology, China, in 1993.

He joined the Faculty with the Beijing Institute of Technology, in 1993, where he is currently a Professor with the School of Computer Science and Technology. His research interests include computer simulation, software engineering, and digital performance.