# Video based Children's Social Behavior Classification in Peer-play Scenarios

Lu Tian, Dingrui Duan, Jinshi Cui, Li Wang, and Hongbin Zha
Peking University, China
{cjs,zha}@cis.pku.edu.cn,{ddr,liwang}@pku.edu.cn

Hamid Aghajan
Stanford University, USA
hamid@icdsc.org

## 1. Introduction

In the studies of developmental psychology, labeling children's play behavior is an important process. Children's behaviors are usually classified into 3 categories in peer-play scenarios ("Solitary Play", "Parallel Play" and "Group Play") based on sounds and two main visual cues: the proximity of children and the attention distribution of children.

Since the proximity of children can be calculated by the locations, the computation of attention distribution and how to extract meaningful feature descriptions are two important challenges when analyzing children's social behavior by computer vision methods. In order to solve these challenges, this extended abstract presents a novel method to label children's social behavior automatically in peer-play scenarios based on visual attention computation.

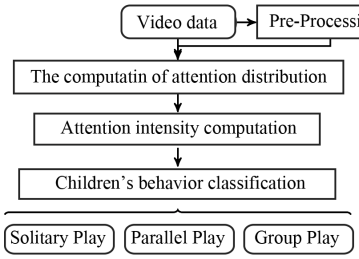


Figure 1. The overview of our system.

## 2. Visual attention computation

### 2.1. The computation of attention distribution

In our method, a two-stage method is used to estimate children's face poses, which are basic cues for the further attention computation. After that, the attention distribution could be expressed as $\{P(Z|X)\}$: $X$ is the face pose estimation result, and $Z$ is a attention target. Inspired by the descriptions of visual saliency, $P(Z|X)$ can be calculated:

$$P(Z=T'|X)=\frac{c}{d^2}\cdot\exp(-\frac{\theta^2}{2\sigma^2})\cdot\phi(T') \quad (1)$$

where $\phi(T')$ is the given weight of the target $T'$, $d$ is the distance between the target and the observer, $\theta$ is the angle between the orientation of face pose and the orientation from the observer to the target, and $c$ and $\sigma$ are constants.

### 2.2. Attention intensity computation

Based on Section 2.1, "Attention Process" $(AP_n)$ is defined to describe the continuous periods when a child pays attention to the same target (Figure 2). Thus, the attention intensity $(I^t_{C_i\to T_j})$ that child $C_i$ pays to target $T_j$ in an attention process $AP_n$ can be calculated:

$$I^{AP_n}_{C_i\to T_j}=\frac{(\sum_{t\in AP_n}A^t_{C_i\to T_j})\cdot\Delta^{AP_n}_{C_i,T_j}}{T_{AP_n}} \quad (2)$$

where $A^t_{C_i\to T_j}$ denotes the probability that the child $C_k$ pays attention to the target $T_j$, $T_{AP_n}$ is the duration of the attention process $AP_n$, and $\Delta^{AP_n}_{C_i,T_j}$ denotes the spatial proximity between $C_i$ and $T_j$ during $AP_n$:

$$\Delta^{AP_n}_{C_i,T_j}=\exp\{-\frac{1}{2\alpha^2}(\min_{t\in AP_n}\|l^t_{C_i}-l^t_{T_j}\|)^2\}(\alpha>0) \quad (3)$$

where $\alpha$ is a given constant, and $l^t_{C_i}$ and $l^t_{T_j}$ are the locations of the child $C_i$ and the target $T_j$ at time $t$.

The total attention intensity $M^t_{C_i\to C_k}$ that child $C_i$ pays to child $C_k$ at time $t$ can be regarded as the summation of the attention intensity that $C_i$ pays to the body parts of $C_k$:

$$M^t_{C_i\to C_k}=\sum_{t\in AP_n}(\sum_{T_j\in T^{C_k}}I^{AP_n}_{C_i\to T_j}) \quad (4)$$

where $T_j$ is a body part of the child $C_k$, and $T^{C_k}$ is the set of all body parts of the child $C_k$.

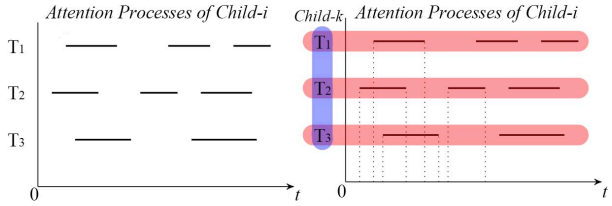


Figure 2. The "Attention Processes" that the child $C_i$ pays attention to the body parts $(T_1, T_2, T_3)$ of the child $C_k$.

## 3. Children's behavior classification

The play types $(P_t)$ of children can be classified into 3 categories: $\{S, P, G\}$, corresponding to "Solitary Play", "Parallel Play" and "Group Play", and it can be assumed that $P(P^t) = 1/N$, where $N$ is the amount of play types.

### 3.1. Feature extraction

Based on the attention computation result, there are two kinds of features extracted to describe children's social behavior: "Solitary Feature" and "Group Feature".

#### 3.1.1 Solitary Feature

"Solitary Feature" $(F_s^t)$ denotes the total attention intensity that the child $C_i$ pays to all the other children:

$$f_s^t = \sum_{C_k \neq C_i} M^t_{C_i \to C_k} \tag{5}$$

Thus, $P(F_s^t|P_t = S)$ and $P(F_s^t|P_t \neq S)$ can be calculated:

$$\begin{aligned} P(F_s^t|P_t = S) &= \lambda_1 N^{-\frac{f_s^t}{\alpha_1}} \\ P(F_s^t|P_t \neq S) &= \frac{\lambda_1}{N-1}(1 - N^{-\frac{f_s^t}{\alpha_1}}) \end{aligned} \tag{6}$$

where $\lambda_1$ is a constant and $\alpha_1$ is a threshold.

#### 3.1.2 Group Feature

"Group Feature" $(F_g^t)$ represents the summation of all the group interaction including one child at time $t$. According to Section 2.2, the group interaction among children can be expressed as a directed graph (Figure 3). In this directed

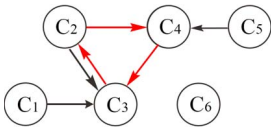


Figure 3. The directed graph of "Group Feature".

graph, two kinds of loops are considered to describe "Group Feature": two-children-loop ($\{C_2, C_3\}$) and three-children-loop ($\{C_2, C_3, C_4\}$). All these loops could be recorded as a set $\{L_i^t\}$ and $f_g^t$ of child $C_n$ can be calculated:

$$f_g^t = \sum_{L_i^t \text{ contains } C_n} (\min_{e_j^t \in L_i^t} M^t_{e_j^t}) \tag{7}$$

where $e_j^t$ is an edge of $L_i^t$. Hence, $P(F_g^t|P_t = G)$ and $P(F_g^t|P_t \neq G)$ can be calculated:

$$\begin{aligned} P(F_g^t|P_t = G) &= \lambda_2(1 - N^{-\frac{f_g^t}{\alpha_2}}) \\ P(F_g^t|P_t \neq G) &= \frac{\lambda_2}{N-1} N^{-\frac{f_g^t}{\alpha_2}} \end{aligned} \tag{8}$$

where $\lambda_2$ and $\alpha_2$ are two constants similar to $\lambda_1$ and $\alpha_1$.

### 3.2. Play behavior classification

For a certain child, the most possible play type $(P_t^*)$ can be computed by maximum a posterior estimation and simplified by the definition of surprisal $S(x) = -\log P(x)$:

$$P_t^* = \arg\max_{P_t} P(V_t, P_t) = \arg\min_{P_t} S(V_t|P_t) \tag{9}$$

where $V^t$ is a short video clip lasting several seconds.

Based on the two kinds of features mentioned above, the surprisal of three play types can be expressed as follows:

$$\begin{aligned} S(V_t|P_t = S) &= S(F_s^t|P_t = S) + S(F_g^t|P_t \neq G) \\ S(V_t|P_t = G) &= S(F_s^t|P_t \neq S) + S(F_g^t|P_t = G) \\ S(V_t|P_t = P) &= S(F_s^t|P_t \neq S) + S(F_g^t|P_t \neq G) \end{aligned} \tag{10}$$

## 4. Experimental result

The whole test video for behavior classification lasts 2 minutes (3600 frames), and it is divided into 24 clips. 1 minute of this 2-minute-video is used to evaluate the visual attention computation. Some examples and the results of experiments are shown below.

Figure 4. Some examples of experimental dataset.

Table 1. The result of visual attention computation.

| All | Accuracy | | | |
|---|---|---|---|---|
| | Child1 | Child2 | Child3 | Average |
| 5400 | 1428 | 1452 | 1395 | 1425 |
| 100% | 79.33% | 80.67% | 77.5% | 79.17% |

Table 2. The result of behavior classification.

| | Solitary | Parallel | Group |
|---|---|---|---|
| Solitary | 67.86%(19) | 32.14%(9) | 0%(0) |
| Parallel | 0%(0) | 77.78%(7) | 22.22%(2) |
| Group | 0%(0) | 6.25%(2) | 93.75%(30) |

## 5. Conclusions

In this extended abstract, a novel method is proposed to classify children's social behavior in peer-play scenarios based on attention computation. According to the visual attention computation result, "Solitary Feature" and "Group Feature" are extracted and used for the classification of children's play types. At last, this method is evaluated by a test video of developmental psychology and the results show this method has a good performance.