Hidden Follower Detection: How Is the Gaze-Spacing Pattern Embodied in Frequency Domain?

Shu Li¹, Ruimin Hu¹*, Suhui Li¹, Liang Liao²

¹School of Cyber Engineering, Xidian University
²School of Computer Science and Engineering, Nanyang Technological University

Abstract

Spatiotemporal social behavior analysis is a technique that studies the social behavior patterns of objects and estimates their risks based on their trajectories. In social public scenarios such as train stations, hidden following behavior has become one of the most challenging issues due to its probability of evolving into violent events, which is more than 25%. In recent years, research on hidden following detection (HFD) has focused on differences in time series between hidden followers and normal pedestrians under two temporal characteristics: gaze and spatial distance. However, the time-domain representation for time series is irreversible and usually causes the loss of critical information. In this paper, we deeply study the expression efficiency of time/frequency domain features of time series, by exploring the recovery mechanism of features to source time series, we establish a fidelity estimation method for feature expression and a selection model for frequency-domain features based on the signal-to-distortion ratio (SDR). Experimental results demonstrate the feature fidelity of time series and HFD performance are positively correlated, and the fidelity of frequency-domain features and HFD performance are significantly better than the time-domain features. On both real and simulated datasets, the accuracy of the proposed method is increased by 3%, and the gaze-only module is improved by 10%. Related research has explored new methods for optimal feature selection based on fidelity, new patterns for efficient feature expression of hidden following behavior, and the mechanism of multimodal collaborative identification.

Introduction

Many achievements have been made in modeling abnormal behaviors with obvious visual features (Liu et al. 2022; Ionescu et al. 2019). As a prelude to many criminal crimes, hidden following is a behavior of secretly tracking and monitoring behind the target. Hidden follower detection (HFD) seeks to identify the hidden follower in all pedestrians from the surveillance video. However, methods based on computer vision (Li, Zhang, and Diao 2020; Jiang et al. 2020; Duan et al. 2020), end-to-end learning (Zhou et al. 2019), or sensors (Wang et al. 2017) are helpless for this task because there are no obvious posture characteristics; Then mainly studies the relative position of two trajectories (Andersson et al. 2008; Siqueira et al. 2011) and temporal-spatial trajectory (Kjærgaard et al. 2013; Li et al. 2013; Xie, Ren, and Liu 2020; Jiang et al. 2018), nevertheless, the following behavior is ubiquitous, making it difficult to distinguish whether the pedestrians behind with similar trajectories are due to coincidence or hidden following intention.

The study of “hidden” behavior has seen new light in recent works (Xu et al. 2022, 2021). They consider behavior patterns can reflect human intentions: first found there are significant pattern differences in gaze-spatial behavior between hidden followers and normal pedestrians. In the time domain, they extract the gaze state series and distance series from the surveillance video and generate gaze-spacing-flow features to represent the gaze-spatial low of walking. Then, a hidden follower detection framework embedded with gaze-spacing-flow (HFDF-GS) is proposed to improve the accuracy of HFD. However, the time domain representation of time series often leads to information loss, which is mainly attributed to 1) Irreversibility: time-domain features are irreversible, for example, we cannot reconstruct from the gazing frequency to the gazing state series. This one-direction feature extraction will inevitably lose some original information; 2) One-sidedness: the subjective time-domain features such as gaze frequency only describe the hidden following behavior in limited aspects.

Nevertheless, early research has proved the vulnerability of time-domain parameters (S.B. and Rao 2016). Signal changes not only with time but also with frequency and phase, etc. Any movement signal in frequency domain can be decomposed into different sine waves, which makes it reflect the essence of things or phenomena from an objective view (Yadav and Rai 2020): 1) Better representation: many studies have demonstrated the frequency-domain features can characterize more accurate and comprehensive original information (Van Segbroeck, Tsiartas, and Narayanan 2013), such as Mel-frequency cepstrum coefficient (MFCC) (Lai et al. 2022), wavelet feature (Lee et al. 2022), spectral entropy (Yu et al. 2022), and Constant-Q Cepstral Coefficients (CQCC) (Bhattacharjee et al. 2020), etc; 2) Reversibility: reversible frequency-domain features can be lossily reconstructed, which maximizes the preservation of original information.

In this paper, in order to prove the information loss of
time-domain features, we define fidelity using signal-to-distortion ratio (SDR) (Boeddeker et al. 2021) to estimate the expression efficiency of time/frequency features and provide a universal selection method for frequency-domain features. Based on frequency-domain analysis, we then propose a decision-fusion hidden follower detection framework based on reversible time-frequency transform (HFDF-TF). HFDF-TF detects the gaze direction and trajectory for all pedestrians in the surveillance to generate the gaze state series and distance series (Xu et al. 2022), and then our framework acts on it to mine the frequency-domain features that distinguish hidden followers from location followers. The contributions of this paper are as follows:

- We demonstrate there is information loss in time-domain features by exploring the recovery of features to the original time series. We establish a fidelity estimation method for feature expression and a selection model for frequency-domain features based on SDR.
- We explore the expression mechanism of human behavior in frequency domain, there are also significant differences in gaze-spacing patterns between the hidden follower and normal pedestrians.
- In HFDF-TF, we first redefine the gaze state in gaze state series; then introduce the decision-fusion training to fully integrate the characteristics of both gaze features and spacing features. Compared to the baselines, HFDF-TF achieves a considerable improvement.

**Definitions and Preliminaries**

Xu et al. (2022) defines two kinds of following pedestrians: Position follower: pedestrians with following characteristics in temporal-spatial position and Hidden follower: position followers with real hidden following intention.

**Definition 1 (Position follower)** Given two moving pedestrians $F = \{f_1, \cdots, f_k, \cdots\}$ and $T = \{t_1, \cdots, t_k, \cdots\}$, where $f_k = (u_{f_k}, v_{f_k})$ and $t_k = (u_{t_k}, v_{t_k})$ are 2D locations at the timestamp $k$. Given a distance threshold $\tau$, if 1): $\Delta(k) = \| f_k - t_k \|_2 < \tau$; 2): $\langle t_k - f_k \rangle \cdot (f_{k+1} - f_k) > 0$ (the angle of $F$ and $T$ in moving direction is less than 90°); and 3): $\langle t_{k+1} - t_k \rangle \cdot (f_{k+1} - f_k) > 0$ ($\mathcal{T}$ is in front of $\mathcal{F}$ in $\mathcal{F}$'s moving direction). In a time interval, if $\mathcal{F}$ follows $\mathcal{T}$ more than $\epsilon$ (default=50%) frames and the distance is always less than $\tau$, we say $\mathcal{F}$ follows $\mathcal{T}$ in this time interval, and $\mathcal{F}$ is defined as a position follower.

**Definition 2 (Hidden follower)** If a pedestrian $\mathcal{F}$ is walking with the intention to “know the real-time position of $\mathcal{T}$” and to “not be found by $\mathcal{T}$”. We say $\mathcal{F}$ hides $\mathcal{T}$, and $\mathcal{F}$ is defined as a hidden follower $\mathcal{H}$.

We call who are not hidden followers $\mathcal{H}$ as normal pedestrians $\mathcal{H}$, there are generally two types: acquaintances ($\mathcal{H}_1$) and strangers ($\mathcal{H}_2$). $\mathcal{H}$ may be the position followers or not.

**Gaze-Spacing Flow.** Xu et al. (2022) define the gaze-spacing pattern to represent behavioral differences between the hidden follower $\mathcal{F}$ and the target $\mathcal{T}$:

- **Gaze pattern:** $\mathcal{H}$ need to gaze at $\mathcal{T}$ frequency to prevent being lost, nor too frequently to avoid being found. • **Spacing pattern:** $\mathcal{H}$ should not be too far away from $\mathcal{T}$ to prevent being lost, nor too close to avoid being found.

In the time domain, Xu et al. (2022) represents the spacing pattern of the distance series using the spacing flow features: distance range and average distance; and represents the gaze pattern of the gaze state series using the gaze flow features: gaze frequency and gaze density.

**Gaze State Series and Distance Series.** In HFDF-GS (Xu et al. 2022), the gaze state series only uses a coarse-grained representation (1 or 0) of threshold dichotomy to determine gaze or not, which is not enough to describe complex gaze behavior. In this paper, we introduce a dynamic score to represent the gaze degree at timestep $i$.

For the “short-time analysis” in the frequency domain, given a gaze state series and a distance series with $L$ frames for a video, the frame rate is $fps$, we first divide the series into multiple overlapping segments through a sliding window (the overlap between two frames is to maintain smooth transition); we set the window length as $w(s)$, each movement of the window (frameshift) is $inc(s)$, so the overlap is $overlap = w - inc$. Then, the number of segments $N$ is:

$$N = \frac{L - w}{inc} + 1. \quad (1)$$

Each segment contains $w \times fps$ frames (timestamps) of the gaze state or distance information. For each segment, the gaze state series is described by $G = \{g_1, g_2, \cdots, g_w \times fps\}$, where $-1 \leq g_i \leq 1$ that means the degree of $\mathcal{F}$ gaze at $\mathcal{T}$ who in front in timestamp $i$. Similarly, the distance series is recorded as $D = \{d_1, d_2, \cdots, d_w \times fps\}$, where $d_i$ denotes the following distance in timestamp $i$. Obviously, the segmentation of time series can also preserve temporal variability.

**Frequency-Domain Features.** To analyze the universality of frequency-domain features for hidden following behavior, we choose a classical feature and a complex feature: Mel-Frequency Cepstral Coefficient (MFCC) (Murty and Yegnanarayana 2006; Brown et al. 2020): MFCC is the most common and representative feature (Lee et al. 2019). The time-frequency transform is based on the Short-Time Fourier Transform (STFT) (Lu et al. 2009).

**Constant-Q Cepstral Coefficients (CQCC)** (Todisco et al. 2016; Bhattacharjee et al. 2020): CQCC has a better time-frequency resolution, but has high time and computational complexity. The time-frequency transform is based on constant-Q transform (CQT) (Shah et al. 2023).

Considering the CQCC is not conducive to the timeliness of HFD, the subsequent frequency-domain analysis is mainly based on the most representative MFCC feature.

**Expression Efficiency of Features**

In this section, we will evaluate the expression efficiency of time/frequency domain features from two aspects: fidelity and mode differentiation, while proving there is greater information loss in time-domain features.

**Fidelity.** The signal-to-distortion ratio (SDR) (Boeddeker et al. 2021) when recovering the initial time series can imply the information loss rate, the smaller the SDR, the more
Fidelity for Feature Selection in Frequency Domain. To explore the impact of the fidelity of frequency-domain features on the recognition performance of hidden followers, we obtain the fidelity of MFCCs under different dimension, $w(s)$, and overlap(s). The line graph is shown in Fig. 1, both in gaze-spacing pattern, we find HFD performance is positively correlated with feature fidelity, although there will be a slight disturbance in $F_1$ score when the fidelity of the two features is very similar. We can say when the feature fidelity increases by more than 0.5%, the HFD performance is also likely to show positive growth. So the optimal frequency-domain feature can be chosen based on fidelity without tedious training. This method is widely applicable to other reversible frequency-domain features or tasks.

Mode Differentiation of Source Information and Features. On both real and simulated datasets, by performing K-S tests (Bickel 1969) on the $k$-means clustering modes, we set up some control groups to compare the pattern differences of time-frequency features or source series (Table 2). In this section, we only analyze the mode differences between initial time series $\mathcal{L}$ and time/frequency domain features: gaze state series and gaze flow of $\mathcal{H}_F$ ($\mathcal{H}$ in time domain), distance series and spacing flow of $\mathcal{H}_T$, gaze state series and gaze MFCCs of $\mathcal{H}_F$ ($\mathcal{H}$ in frequency domain), distance series and spacing MFCCs of $\mathcal{H}_F$. The pattern differences between $\mathcal{L}$ and the MFCCs ($\mathcal{L}\mathcal{H}_F$) are more significant from the gaze-spacing flow ($\mathcal{L}\mathcal{H}_T$), this indicates frequency-domain features have strong expressive power for time series. Moreover, our gaze-only module (HFDF-GF, see Table 3) is greatly enhanced due to the significant difference in gaze pattern (0.076 in $\mathcal{L}\mathcal{H}_T$ vs 1.21E-39 in $\mathcal{L}\mathcal{H}_F$).

Gaze-Spacing Pattern in Frequency Domain

Based on the above analysis, we make and verify the following assumption:

**Assumption 1 (Gaze-Spacing pattern)** In frequency domain, the gaze-spacing pattern between the hidden followers and the normal pedestrians is significantly different.

MFCC Clustering Modes. For all hidden following pairs and normal walking pairs in real dataset, we extracted the $d$-dimensional MFCCs and then conducted the $k$-means clustering analysis, they are clustered into four modes: A, B, C, D. Fig. 2 shows the comparison of MFCC mode distribution with the radar chart. It indicates the gaze pattern of $\mathcal{H}$ is more likely concentrated in mode $C$, and the spacing pattern of $\mathcal{H}$ is more likely distributed in mode $A$ and $B$; yet the clustering modes of $\mathcal{H}$ is disorderly and different to $\mathcal{H}$.

### Table 1: The fidelity of time/frequency features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Time domain</th>
<th>Frequency domain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaze</td>
<td>Spacing</td>
</tr>
<tr>
<td>Fidelity</td>
<td>52%</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 2: Mode differentiation in control groups: p-value.

<table>
<thead>
<tr>
<th>Control group</th>
<th>Real-HFD</th>
<th>Sim-HFD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaze</td>
<td>Spacing</td>
</tr>
<tr>
<td>$L\mathcal{L}$</td>
<td>8.42E-03</td>
<td>1.51E-10</td>
</tr>
<tr>
<td>$L\mathcal{H}_T$</td>
<td>0.076</td>
<td>1.25E-04</td>
</tr>
<tr>
<td>$L\mathcal{H}_F$</td>
<td>1.21E-39</td>
<td>8.52E-23</td>
</tr>
<tr>
<td>$\mathcal{H}_T\mathcal{H}_F$</td>
<td>8.82E-03</td>
<td>5.50E-08</td>
</tr>
<tr>
<td>$\mathcal{H}_F\mathcal{H}_F$</td>
<td>0.250</td>
<td>0.497</td>
</tr>
<tr>
<td>$\mathcal{H}_F\mathcal{H}_T$</td>
<td>0.437</td>
<td>0.372</td>
</tr>
<tr>
<td>$\mathcal{H}_T\mathcal{H}_F$</td>
<td>1.15E-21</td>
<td>5.44E-29</td>
</tr>
</tbody>
</table>

Figure 1: Relationships between the fidelity of various gaze-spacing MFCCs and HFD performance ($F_1$ score).
At the micro level, features are completely different from the time domain. P that each pedestrian hidden following the whole moving process of ture. As shown in Fig. 4, for both gaze and spacing pattern, high-frequency components (see Fig. 3(a)(b)).pared to normal walking, it shows a significant increase in walking speed of Gment in G.

Figure 2: The MFCC clustering modes of gaze-spacing pattern (radar chart). (a) and (b) showed the mode distribution of gaze and spacing MFCCs respectively. Each radar chart displays the mode distribution of one video, the first and second line is the hidden following and normal walking pairs.

<table>
<thead>
<tr>
<th>Image 54x514 to 293x619</th>
</tr>
</thead>
</table>

**Figure 3:** Power spectrum of the gaze-spacing pattern when hidden following occurs. (a) and (b) depicted the frequency (power) structure of the gaze and spacing patterns of the hidden following pair and the normal walking pair respectively.

**Trajectory Tracking**

We first use the TraDeS (Wu et al. 2021) model (online multi-target tracker) to track the pixel coordinates of all pedestrians in the surveillance video and also convert the pixel coordinates into real ground coordinates by persppective transformation. The distance series of each pedestrian pair is calculated by their relative real ground distance.

**Gaze State Series Extraction**

The gaze state of the follower to the target needs to be determined by gaze direction.

**Gaze Direction Detection**

First, the HFDF-FT combines DensePose model (Güler, Neverova, and Kokkinos 2018) with the TraDeS tracking results for more accurate head-box tracking. Then, the Gaze360 model (Kellnhofer et al. 2019) outputs the 2D gaze direction for each pedestrian.

**Gaze State Series**

The gaze state series $\mathcal{G}$ consists of the gaze state of each frame. Note that we introduce “gaze degree (score)” to indicate the gaze state instead of using a simple 1 or 0 to denote gaze or not (1 when the gaze angle is less than 60, otherwise it is 0) (Xu et al. 2022). For i-th frame in the video, the gaze state $g_i (g_i \in \mathcal{G})$ from $F$ to $T$ is calculated by the ground coordinates of $F$: $(u_{fi}, v_{fi})$ and $T$: $(u_{ti}, v_{ti})$ and the 2D gaze directions of $F$: $(gd_x, gd_y)$.

$\Delta u_i = u_{fi} - u_{ti}, \Delta v_i = v_{fi} - v_{ti}, \quad (3)$

$\alpha = \arccos \frac{(gd_x, gd_y) \cdot (\Delta u_i, \Delta v_i)}{|(gd_x, gd_y)| \cdot |(\Delta u_i, \Delta v_i)|}, \quad (4)$

$g_i = 1 - 2 \times \angle(\alpha) / 180, \quad (5)$

where $\angle(\alpha) \in [0,180]$, and $-1 < g_i < 1$ represents the degree of $F$ gaze at $T$ in timestamp $i$. The closer the gaze angle is to 0, the $F$ is more likely to gaze at $T$.

**Frequency-Domain Features Extraction**

After segmentation by Eq. 1, the distance series and gaze state series are divided into $N$ segments. We extracted the $d$-dimensional MFCC of each gaze segment and spacing segment, and finally obtained the $(N \times d)$ MFCC feature matrix as the gaze feature and the spacing feature, and input them
(a) Gaze spectrogram of hidden following pair
(b) Gaze spectrogram of normal walking pair
(c) Space spectrogram of hidden following pair
(d) Space spectrogram of normal walking pair

Figure 4: Spectrograms of the gaze-spacing pattern in hidden following and normal walking pair (x label: Time, y label: Hz).

Figure 5: Architecture of the proposed HFDF-TF. HFDF-TF first detects the trajectory and gaze direction of each pedestrian in the surveillance video to acquire the distance series and gaze state series, then transforms them to frequency domain to obtain the spacing MFCCs and gaze MFCCs, they were each sent to a TSC model and outputs the probability of “being a hidden follower”, respectively. Finally, the final output of HFDF-TF comes from the decision fusion of the two TSC models.

We apply DS evidence theory (DS inference) (Martin, Zhang, and Liu 2010) as the decision fusion strategy (see Fig. 5). Regard the two TWIESN classifiers as $m_1$ and $m_2$, we defined the following identification domain:

$$\Psi = \{\theta_1, \theta_2\},$$

where $\theta_1$ and $\theta_2$ represents the proposition of: $F$ is a hidden follower and $\overline{F}$ is not a hidden follower, respectively. $\theta_1$ and $\theta_2$ output the prediction probability of $m_1$ and $m_2$, respectively, which is recorded as: $m_1 : \{m_1(\theta_1), m_1(\theta_2)\}$ and $m_2 : \{m_2(\theta_1), m_2(\theta_2)\}$. The final decision of proposition $\theta_1$ is calculated as follows.

$$P(\theta_1) = \frac{m_1(\theta_1)m_2(\theta_1)}{K},$$

where $K = \sum_{i=1}^{2} m_1(\theta_i)m_2(\theta_i)$ is the conflict coefficient.

**Decision Fusion Training**

We apply DS evidence theory (DS inference) (Martin, Zhang, and Liu 2010) as the decision fusion strategy (see Fig. 5). Regard the two TWIESN classifiers as $m_1$ and $m_2$, we defined the following identification domain:

$$\Psi = \{\theta_1, \theta_2\},$$

where $\theta_1$ and $\theta_2$ represents the proposition of: $F$ is a hidden follower and $\overline{F}$ is not a hidden follower, respectively. $\theta_1$ and $\theta_2$ output the prediction probability of $m_1$ and $m_2$, respectively, which is recorded as: $m_1 : \{m_1(\theta_1), m_1(\theta_2)\}$ and $m_2 : \{m_2(\theta_1), m_2(\theta_2)\}$. The final decision of proposition $\theta_1$ is calculated as follows.

$$P(\theta_1) = \frac{m_1(\theta_1)m_2(\theta_1)}{K},$$

where $K = \sum_{i=1}^{2} m_1(\theta_i)m_2(\theta_i)$ is the conflict coefficient.

**Experiments and Results**

**Datasets**

**Real-HFD.** The Real-HFD includes 20 pedestrians with a total video length of 160 minutes. Each video lasts one minute, in which the number of pedestrians varies from 5 to 12. In the hidden following videos, each contains 1 or 2 hidden following pairs, and others are normal pedestrians. The non-hidden following videos contain 4 to 6 pairs of acquaintances. If each sample refers to one behavior of a pedestrian
in a 1-minute video. Real-HFD includes 160 samples of hidden followers \( (H) \), 280 samples of acquaintances \( (T_1) \) and 370 samples of strangers \( (T_2) \). More details can be found in (Xu et al. 2021, 2022).

**Sim-HFD.** Due to the hidden following data in the real world is rare for training, Xu et al. (2022) established a simulation model to simulate the gaze behavior and movement behavior of four types of pedestrians: 1) The motion of \( (H) \) is adjusted according to the principle of following \( (T) \) and not being detected. 2) The motion of \( (T_1) \) is similar to that of \( (T) \). 3) The motion of \( (T_2) \) is basically independent of \( (T) \). After performing the simulation model 40 times, once putting 31 pedestrians into the scene: \( 1 \times T, 10 \times H, 10 \times T_1, \) and \( 10 \times T_2 \), there are 1200 series of spacing and gaze behaviors of normal pedestrians and hidden followers in the end. We trained with 4-fold cross-validation, of which the ratio of the training set and test set in both datasets is 7:3.

**Experimental Setup.** We evaluate the HFDF-TF model from the four aspects:

- How is the HFDF-TF model performance?
- How does the model perform with only gaze MFCCs or spacing MFCCs?
- Is the model applicable to different frequency-domain features?
- Will multimodal feature collaboration further improve HFD performance?

**Comparison methods.** In this paper, we compare HFDF-TF with the traditional following detection method based on trajectory (Li et al. 2013), the hidden follower detection model HFDF (Xu et al. 2021), and the SOTA HFD model HFDF-GS (Xu et al. 2022).

**Parameter settings.** The optimal gaze and spacing MFCC parameters in Fig. 1 are: gaze MFCCs with \textit{dimension} = 4, \textit{w} = 10s, \textit{overlap} = 2s and spacing MFCCs with \textit{dimension} = 4, \textit{w} = 8s, \textit{overlap} = 1s.

**Evaluate metrics.** Precision, Recall, \( F_1 \) score, Accuracy and AUC.

The Performance of HFDF-TF

On real-HFD and Sim-HFD, we evaluate the performance of the proposed HFDF-TF with other HFD methods.

**Results.** Table 3 reveals the performance comparison results on five evaluated metrics. Both on Real-HFD and Sim-HFD, the HFD performance of HFDF-TF is significantly better than the traditional trajectory-based method and the SOTA baseline method. Compared to HFDF-GS, on Sim-HFD, the proposed HFDF-TF is improved 2% to 3% in four evaluate metrics, and improved 3% to 4% on Sim-HFD. Besides, see Fig. 6 for the visual comparison results between HFDF-TF and HFDF-GS.

In order to test the effects of the gaze module, spacing module, and decision fusion strategy separately, the corresponding ablation model is defined as 1) HFDF-GF (gaze-only HFDF-TF), 2) HFDF-SF (spacing-only HFDF-TF), and 3) HFDF-ND (HFDF-TF without decision fusion, which based on the feature concatenation and only one TWIESN network). The ablation results are also displayed in Table 3. Compared to HFDF-G, HFDF-GF meets great improvement in gaze features, an increase of about 9% to 10%; Compared to HFDF-G, the improvement of HFDF-GF in spacing features is about 2% to 3%. Nevertheless, why does the spacing pattern in the frequency domain increase far less than the gaze pattern? The reasons are as follows:

1. \textit{Explanation from the behavior mode.} Relatively speaking, trajectory tracking is more accurate than the gaze angle. On the other hand, the distance law is a strong constraint, yet whether to gaze is a weak constraint: the target suddenly turns back is a small probability event, and the follower does not need to strictly control his gaze direction.

2. \textit{Explanation from benefits of time-frequency transform.} The more accurate the time-domain parameters express the behavior, the smaller the additional benefits we can get in the frequency domain. It is obvious that the time-domain gaze parameters in HFDF-GS are not enough to express the following behavior \( (F_1 \text{ score}: 71.5\%) \).

The comparison between HFDF-ND and HFDF-TF in Table 3 shows that simple feature concatenation will destroy the unique frequency structure of gaze mode and spacing.
In cases where all metrics are small differences, MFCCs domain features, such as FFT, spectrum, Fbank features, etc. process of MFCC involves many other typical frequency-its the practical application of CQCCs. Yet the extraction CQCCs is almost twice that of MFCCs, which greatly lim-

comparison of each iteration time that, the time complexity of MFCCs in all three models, it can be seen from the com-

and CQCCs on Real-HFD. Although CQCCs outperform Table 4 shows the comparison between MFCCs

Table 3: Comparison of five evaluate metrics between MFCCs and CQCCs on Real-HFD, and the comparison of each iteration time during training.

Table 4: Comparison of five evaluate metrics between MFCCs and CQCCs on Real-HFD, and the comparison of each iteration time during training.

Table 5: Performance of multimodal feature collaboration. Where \((G \text{mfcc} + G_{\text{flow}}, S_{\text{mfcc}})\) is a decision fusion model for gaze features (gaze MFCCs concatenate with gaze flow) and spacing features (spacing MFCCs concatenate with the spacing flow). “+” denotes the concatenation along specific dimensions, if there is no “+”, it means no feature concatenation. TF decision is a decision fusion model for HFDF-GS and HFDF-TF.

The Performance of Multimodal Collaboration

On Real-HFD, we also explored the performance of collabora-
ting time-frequency features. Moreover, we conducted decision fusion training on the HFDF-GS and HFDF-TF to in-
tegrate time-frequency domain information fully, the model is recorded as TF decision model.

Results. The gaze-spacing flow (Xu et al. 2022) combines with the gaze MFCCs and spacing MFCCs, respectively, or simultaneously (see Table 5). But compared to HFDF-TF, it doesn’t seem to meet expectations. The decision fusion model of HFDF-GS and HFDF-TF achieved the best HFD.

Conclusions

Our paper seeks to open the minds of hidden follower detec-
tion (HFD) for a new research agenda. By studying the fi-
delity of source information recovery of time/frequency do-
main features, we found the frequency-domain features have better expression efficiency for time series, and establish a selection model for frequency-domain features based on the fidelity. Furthermore, by analyzing the motion of hidden fol-

The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

Feature Model Time Precision Recall \(F_1\) Accuracy AUC

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Model</th>
<th>Time</th>
<th>Precision</th>
<th>Recall</th>
<th>(F_1)</th>
<th>Accuracy</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trajectory</td>
<td>HFDF-GF 0.75s</td>
<td>81.8</td>
<td>80.4</td>
<td>80.6</td>
<td>80.4</td>
<td>80.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HFDF-SF 0.75s</td>
<td>94.0</td>
<td>93.8</td>
<td>93.9</td>
<td>93.8</td>
<td>93.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HFDF-TF 1.22s</td>
<td>94.6</td>
<td>94.4</td>
<td>94.5</td>
<td>94.4</td>
<td>94.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HFDF-GF 1.39s</td>
<td>83.2</td>
<td>81.9</td>
<td>82.1</td>
<td>81.9</td>
<td>82.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CQCC</td>
<td>HFDF-SF 1.33s</td>
<td>95.2</td>
<td>94.6</td>
<td>94.9</td>
<td>94.6</td>
<td>94.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HFDF-TF 2.36s</td>
<td>96.1</td>
<td>95.4</td>
<td>94.9</td>
<td>95.4</td>
<td>95.2</td>
</tr>
</tbody>
</table>

Acknowledgements

This work was supported in part by the National Natural Sci-
ence Foundation of China under Grant U22A2035; in part by the Guangxi Natural Science Foundation Program under Grant 2021GXNSFDA075011. We thank Danni Xu for her guidance and help in our work.
References


Todisco; Massimiliano; Delgado; Héctor; Evans; and Nicholas. 2016. A New Feature for Automatic Speaker Verification Anti-Spoofing: Constant Q Cepstral Coefficients. In Odyssey 2016 - The Speaker and Language Recognition Workshop.


