A Reliable, Self-Adaptive Face Identification Framework via Lyapunov Optimization

Dohyeon Kim

Chung-Ang University, Seoul, Korea
dohyeonkim0412@gmail.com

Joongheon Kim

Chung-Ang University, Seoul, Korea
joongheon@cau.ac.kr

Jae young Bang

Kakao Corp., Seongnam, Korea
jae.bang@kakaocorp.com

Abstract

Realtime face identification (FID) from a video feed is highly computation-intensive, and may exhaust computation resources if performed on a device with a limited amount of resources (e.g., a mobile device). In general, FID performs better when images are sampled at a higher rate, minimizing false negatives. However, performing it at an overwhelmingly high rate exposes the system to the risk of a queue overflow that hampers the system's reliability. This paper proposes a novel, queue-aware FID framework that adapts the sampling rate to maximize the FID performance while avoiding a queue overflow by implementing the Lyapunov optimization. A preliminary evaluation via a trace-based simulation confirms the effectiveness of the framework.

1. Introduction

Realtime face identification (FID) is an active area of research in the artificial intelligence domain that has widely been adopted by the industry. As it matures, today, we are witnessing it spreading into our daily lives and being implemented even in mobile applications that often have limitations in the amount of computation resources available.

The computation resource limitation puts the system reliability and the FID performance in a trade-off relationship. An FID system samples images from a video feed and inserts them into a queue. The system concurrently retrieves the images and executes an FID algorithm to detect faces that appear in those images. With a higher sampling rate, the system would less likely to miss faces that do appear in the feed but not in the sampled images. However, an overwhelmingly high sampling rate could also cause a queue overflow, resulting in unexpected system behaviors. On the other hand,
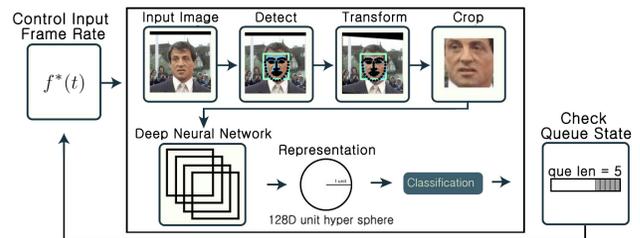
Figure 1: The proposed FID framework using OpenFace

while sampling at a lower rate would help prevent an overflow, the system would more likely to fail to identify faces.

In this paper, we present a queue-aware FID framework that automatically adapts the sampling rate to the computation resource availability via the Lyapunov optimization in order to achieve maximum FID performance (fewest false negatives) while avoiding a queue overflow. We do not develop new FID algorithms; we exploit the existing ones and provide a self-adaptive framework on which the algorithms can be implemented and executed in a reliable fashion.

2. Stochastic Frame Rate Adaptation

2.1 Lyapunov optimization framework

The theory of stochastic optimization (Neely 2010) aims at optimizing a time-average utility subject to queue stability when the objective function and the queue stability constraints are in a trade-off relationship. Stochastic optimization models the queue stability using the Lyapunov drift. It takes actions that minimizes the Lyapunov drift while pursuing the minimization of a time-average objective function with the gap of $O(1/V)$ under queue stability with the bound of $O(V)$ where V is defined as a trade-off between utility and stability. Neely discusses the details of the theory in his book (Neely 2010), and there are a number of applications that implement the theory (Kim et al. 2016; Koo et al. 2017).

2.2 Reference face identification (FID) platform

Fig. 1 depicts our reference FID framework. A system that implements the framework will have a video source (e.g., a camera) that constantly feeds a stream of images (frames) into the framework, and the framework stores the images in

Algorithm 1 Frame rate control via Lyapunov optimization

1: $Q(t) \leftarrow 0$ and $t \leftarrow 0$
2: **while** $t \leq T$ **do** // T: operation time
3:     Observe $Q(t)$ and $\mathcal{T}^* \leftarrow -\infty$
4:     **for** $f(t) \in \mathcal{F}$ **do** $\mathcal{T} \leftarrow V \cdot S(f(t)) - Q(t) \cdot \lambda(f(t))$
5:         **if** $\mathcal{T} \geq \mathcal{T}^*$ **then** $\mathcal{T}^* \leftarrow \mathcal{T}$ and $f^*(t) \leftarrow f(t)$

---

a queue in which they wait until the framework processes them. The framework concurrently conducts face identification on the images in the queue using the OpenFace library in the following four steps that correspond to those illustrated in Fig. 1: (1) input image loading (*Input Image*), (2) face recognition and preprocessing (*Detect*, *Transform*, and *Crop*), (3) neural network forwarding (*Deep Neural Network*, *Representation*), and (4) classification.

The FID framework automatically controls the frame rate of the video feed in order to maximize FID performance while avoiding a queue overflow. In this paper, we define the *FID performance* as following: the performance with a frame rate $f(t)$ at time t is defined as $S(f(t)) \triangleq \frac{\alpha(f(t))}{\beta(t)}$ where $\alpha(f(t))$ and $\beta(t)$ respectively denote the number of identified faces with given $f(t)$ and the total number of faces appeared in the original video feed, at time t. $S(f(t)) = 1$ if all faces are identified in $f(t)$ whereas $S(f(t)) = 0$ if no face is identified at all in $f(t)$. Higher $f(t)$ generally leads to higher $S(f(t))$; that lowers the chance of having faces that quickly pass by between two frames, unidentified (false negatives). However, in reality, because it is likely for the FID system to have a limited amount of queue storage and finite computation power, an overwhelmingly high $f(t)$ could cause a queue overflow and result in unreliable system behaviors. The framework continuously monitors the queue status and adapts the rate at which the frames from the video feed are inserted into the queue to the resource availability.

2.3 Frame rate control via Lyapunov optimization

This section describes how we model the time-average optimal frame rate using the Lyapunov optimization (Neely 2010) on which our FID framework controls $f(t)$ based. An FID system that implements our framework continuously adjusts $f(t)$ to the outputs of the optimal frame rate model.

We first model the arrival queue dynamics as $Q(t + 1) = \max\{Q(t) - \mu(t), 0\} + \lambda(f(t))$ where $Q(t)$, $\mu(t)$, and $\lambda(f(t))$ respectively denote the queue-backlog size, the number of images departing from the queue, and the number of images arriving in the queue with $f(t)$. We then formulate the mathematical program for maximizing the time-average system performance $S(f(t))$ as $\max$: $\lim_{t \to \infty} \sum_{\tau=0}^{t-1} S(f(\tau))$ and the queue stability constraint (to avoid a queue overflow) as $\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} Q(t) < \infty$. According to the Lyapunov optimization theory (Neely 2010), this program can be re-formulated as following where $f^*(t)$ is time-average optimal frame rate for observed $Q(t)$:

$$f^*(t) \leftarrow \arg\max_{f(t) \in \mathcal{F}} \{V \cdot S(f(t)) - Q(t) \cdot \lambda(f(t))\}$$
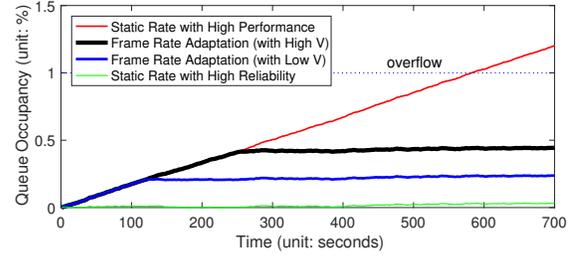
Figure 2: Evaluation results: queue dynamics.

3. Implementation and Evaluation

We evaluated our framework via a trace-based simulation, which showed that the framework maximizes the FID performance while not having any queue overflow. The evaluation was preliminary, and we made an assumption that maximizing the number of frames that the framework processes would also maximize the FID performance. We designed the simulation to mimic an FID system with a threshold of 10 frames/sec at which a queue-divergence would occur. We then varied the frame rate from 1 to 10, with and without our framework. Fig. 2 depicts the four simulation results: (1–red) the queue eventually overflows with a fixed frame rate of 10, (2–black / 3–blue) the queue stabilizes at certain points depending on the given V, and (4–green) the queue is stable but the FID performance is the lowest at 1 frames/sec. Our framework does not require a predetermined frame rate since it self-adapts the rate to the queue status on-the-fly.

4. Concluding Remarks

This paper proposes a reliable FID framework that achieves maximum FID performance by self-adapting the frame rate to the queue-backlog status using a unique Lyapunov optimization model. There are other system objectives such as time to identify a face or energy consumption of an FID system that would also be crucial for an FID system in practice, and we see an interesting research opportunity in creating new optimization models that target those objectives.

References

J. Kim, G. Caire, and A. F. Molisch. 2016. Quality-Aware Streaming and Scheduling for Device-to-Device Video Delivery. *IEEE/ACM Transactions on Networking* (2016).

J. Koo, J. Yi, J. Kim, M. A. Hoque, and S. Choi. 2017. REQUEST: Seamless Dynamic Adaptive Streaming over HTTP for Multi-homed Smartphone under Resource Constraints. In *Proc. ACM Multimedia*.

M.J. Neely. 2010. *Stochastic network optimization with application to communication and queueing systems*. Morgan & Claypool.