Design of Novel Benchmarking System for Power-Efficient Face Detection Algorithm (PE-FDA) in Artificial Intelligence (AI) based Security System

Minhee Jun[*] **Shreyas Venugopalan**[*]
Ajmal Thanikkal[*] **Marios Savvides**[†]

Department of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

October 2021

[*] CyLab HawXeye Inc, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
[†] CyLab Biometrics Center, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

This research is supported by HawXeye Inc.

Keywords: artificial intelligence, security, face recognition, tracking algorithm, power efficiency, metric, video frame rate, benchmark system

Abstract

Recently, there have been active studies of video surveillance using artificial intelligence (AI) and security that operates with face detection algorithms (FDA), such as face identity matching system (FIMS) and pedestrian tracking system (PTS). Benchmarking an optimal FDA is one of the important tasks for designing the AI-based security system. However, this AI-based security system suffers from enormous power consumption due to a high frame rate of multiple cameras. For this reason, the AI-based security system needs to find a power efficient face detection algorithm (PE-FDA). To the best of our knowledge, the conventional FDA benchmarking systems (such as using Iou metric) are not optimized with respect to the power efficiency of FDAs. In this paper, we propose a novel benchmarking system for PE-FDA, including power consumption in AI-based security system. We will define the design of benchmarking system and describe its spatial and temporal challenges. (1) In order to solve the spatial challenges, we propose a novel evaluation score, unitized-distance (UD) metric. (2) In order to improve the temporal challenge, we will introduce frame mapping algorithm. (3) our benchmarking system is designed for PE-FDA in AI-based security system. We validated our benchmarking system of PE-FDA using actual video data obtained from a state-of-the-art security system. Thus, this study of our benchmarking systems can allow FDA to be utilized in AI center monitoring system for the future security system.

Contents

1 Introduction

There has been much interest in upgrading a manual security system to an artificial intelligence (AI) and security system (i.e. security intelligence). This novel AI-based security system is designed to report/alarm security issues based on state-of-the-art face detection algorithms (FDA). FDA can process the video data from multiple cameras in the monitored public spaces such as an airport as given in Figure 1. In the AI-based security system, AI center monitoring system can monitor and report any security issues based on FDAs, such as face identity matching system (FIMS) and/or a pedestrian tracking system (PTS). In order to design an AI-based security system, it is important to design an appropriate benchmarking system of FDAs, that discerns an optimal algorithm of FIMS and/or PTS. This benchmarking system is important not only to optimize hyper parameters of FDAs but also to operate AI-based security system power-efficiently. Thus, the study of this paper focus on designing a novel benchmarking algorithm and its evaluation metric) of face detection algorithms for an AI-based security system.

In practical aspects, it is emphasized for this benchmarking system for AI-based security system to evaluate the performance of power-efficient FDA (PE-FDA). While there are studies of benchmarking algorithm for evaluating FDAs, the conventional benchmarking algorithms (or metrics) are not appropriate because AI-based security system is challenged by its power consumption issue of video surveillance. AI-based security system receives multiple videos from multiple cameras in multiple places in real-time. Every single moment, a huge data amount of videos are generated by each camera, transferred to, and analyzed in the center monitoring system. For example, when there are 50 cameras with 1080p and the frame rate of 30 fps, this security system handles 3 Giga pixels per second. Proportionally increased with the video data, the power consumption of a AI-based security system dramatically increases. In order to handle the size of video data and its power consumption, the frame rate of video data is considered for the choice of FDA in AI-based security system Thus, while adjusting a operating frame rate, the performance of FDAs should be evaluated with respect to its power consumption improvement.

Our study considered this power consumption aspect in benchmarking system design, while conventional benchmarking systems are not. Particularly, we designed a novel and practical benchmarking system in order to overcome the spatial and temporal issues that conventional benchmarking systems could not solve to be operated with the state-of-the-art security system.

The contribution of this paper is specified as follows:

1. We (mathematically) defined an optimization problem for a novel benchmarking system design of PE-FDA that considers the issue of huge power consumption. Also,

We analyzed spatial and temporal challenges in this benchmarking system for PE-FDA.

2. In order to improve spatial challenge, we proposed a novel metric, *Unitized-distance* (UD) base metric. Our UD metric achieves a better detection sensitivity than Iou metric.

3. In order to improve temporal challenge, we designed the Frame Mapping Algorithm that handles the frame rate mismatch of ground truth frame and system output frame.

4. We described our benchmarking algorithm of finding an optimal PE-FDA and its implementation for face identification matching system (FIMS) and pedestrian tracking system (PTS).

5. We validated our designed benchmarking system using the video data from an actual state-of-the-art security system. In this experiment, it is validated that our face detection algorithm improves its performance for being applied to the state-of-the-art security system.

Literature Review

For the last a few years, face detection algorithm (FDA) - including face identity matching system (FIMS) and pedestrian tracking system (PTS) - has been one of active research fields in computer vision. These redundant FDAs need to be evaluated and compared. Thue, benchmarking tool of FDA is used to calculate the value of FDA, and find an optimal FDA. Besides, benchmarking system is important to tune hyper-parameters in a chosen FDA algorithms.

There are studies of benchmarking systems that evaluate performance of FDA as presented in [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15].

In the benchmarking (evaluation) systems, we found that *evaluation metrics* - that numerically represent the performance of testing algorithms - are designed based on a few conventional metrics such as *distance* base metric and *overlap* base metric. In [15], the AUC metric is designed based on Euclidean distance metric. In [16], the OSPA metric measures multi-target tracking algorithm using distance base metric. In [9], a survey about performance evaluation of face detection algorithms is performed. In [6], the SFDA and SATA algorithms are designed using overlap metric with frame-based and object-based evaluation, respectively. In [7], the OMAT metric is presented in order to avoid the outlier problem with the order parameter p in Hausdorff metric.

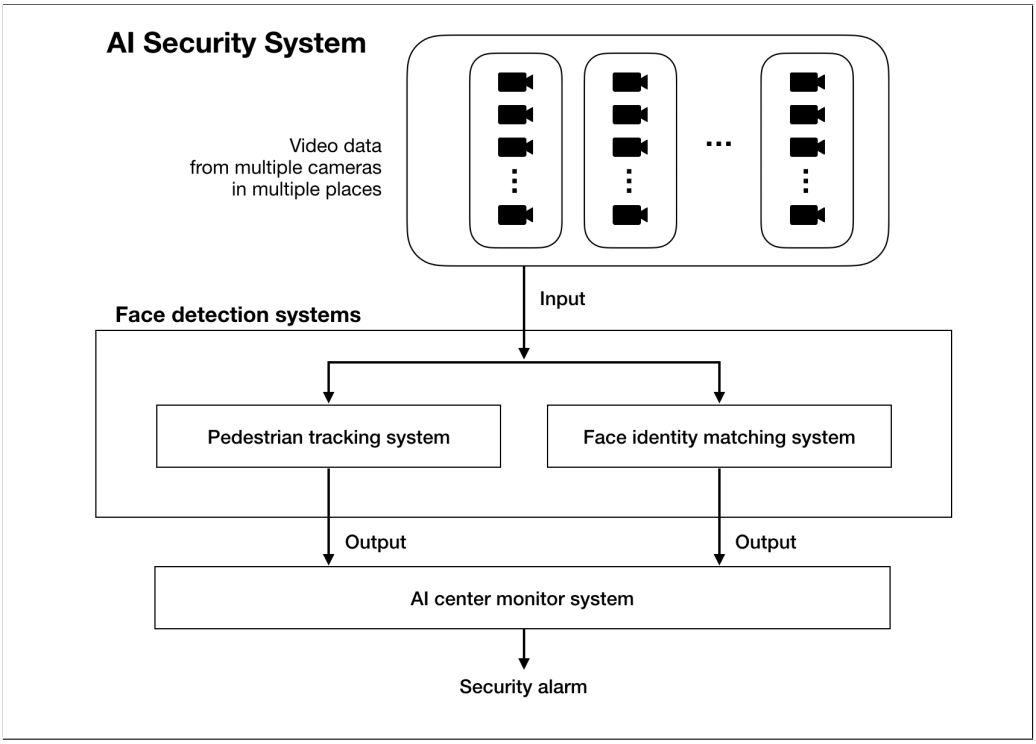


Figure 1: Design architecture of AI-based security system

However, these conventional metrics have limitations for being utilized in benchmarking system of PE-FDA as follows: (1) Distance base metric is limited to be applied to design the evaluation systems with general object detection. Although it is one of the most intuitively understandable and conventional metrics [7], [2], [17], [18], distance-based metrics are restricted to the annotation style of the position points (i.e. face landmarks). For example, the distance base metric is frequently used with face landmarks annotation as introduced in [4], [16], [15]. However, the distance base metric may not able to accommodate the 2D spatial information of detection box. The detection box is generally applicable even to object detection such as face, text, vehicle [6]. It is a more generalizable annotation style for building standardized evaluation system. (2) While overlap-based metrics (i.e., *Iou metric*) can accommodate this detection box annotation by handling the 2D special information, it has some issues with respect to the inconsistency of ground truth and the detection sensitivity.

Our evaluation system targets to evaluate FIMS and PTS. In most evaluation systems, the targeted system types are either FIMS or object tracking system. There are many studies on performance evaluation of multi-object tracking systems [9], [10], [8], [11], [12],

Figure 2: A pedestrian tracker algorithm [1] is applied on test videos. For its performance evaluation, the tracking information (object detection boxes and tracking IDs) on each video is recorded in a Json file.

[13], [6], [14] while performance evaluation for face identity matching system is considered in [6]. In our evaluation system of FIMS and PTS, the two types of face detection algorithms are evaluated in frame-based and object-based approaches, respectively.

Most of the previous evaluation systems obtain ground truth based on human-marked annotations [3], [13], [12], [19].

This paper is composed of the following six sections and appendix. We will introduce an optimization problem of finding a benchmarking system for PE-FDA in Section 2. We will explain our novel metric and its improvement with respect to the spatial accuracy in a benchmarking system in Section 3. We will propose a frame mapping algorithm that improves the temporal accuracy in Section 4. Also, we will describe the design structure of our benchmarking system. We will present the simulation setting and results when actual video data obtained from a security system are used in Section 5. Finally, the conclusion of our novel benchmarking system is given in Section 6. In Appendix, we present the description of three conventional metrics, the proof of Iou metric's property and the figures for UD metric's property with respect to order p (for Section 3), conventional design for implementation (for Section 4), and the scenarios used for sample test (for Section 5).

2 Novel Benchmarking System for Power-Efficient Face Detection Algorithm (PE-FDA)

2.1 Conventional Benchmarking System of Face Detection Algorithm (FDA)

Benchmarking systems have been developed in order to choose the most appropriate face detection algorithm. In other words, a benchmarking system of face detection algorithms(FDAs) is designed to solve the follows problem,

$$\widehat{\mathbf{f}} = \arg\max_{\mathbf{f}\in\mathbf{F}} s(\mathbf{f}), \tag{1}$$

where $\mathbf{f}$ is an element of the set $\mathbf{F}$ of FDAs, the subjects of evaluation. $\widehat{\mathbf{f}}$ is the optimal solution of FDA which is chosen to maximize the performance evaluation score $s(\cdot)$ that represents a designed benchmarking system.

Among all benchmarking systems in a set S, the optimal benchmarking system $\widehat{s}$ can be mathematically defined in the following equation,

$$\widehat{s} = \arg\min_{s\in\mathbf{S}} E\left[s(\mathbf{f_0}) - s(\widehat{\mathbf{f}})\right]^2, \tag{2}$$

where $\mathbf{f}_0$ is the optimal face detection algorithm, S is the set of available benchmarking systems.

The studies of conventional benchmarking systems of FDA are explained in Literature Review.

2.2 Novel Benchmarking System for Power-Efficient Face Detection Algorithm (PE-FDA)

While operating FDAs, the power consumption is one of the major issues for improving AI-based security system (as described in Section 1). In order to improve the power issue, we considered the constraint of power consumption in the optimization problem in (1) of evaluating FDAs as follows,

$$\widehat{\mathbf{f}} = \underset{\mathbf{f}\in F,\, \mathbf{P}<P_0}{\arg\max}\, s(\mathbf{f}), \tag{3}$$

where P_0 is a threshold for power consumption $\mathbf{P}$ in the security system, and $s(\mathbf{f})$ is the score of the FDA $\mathbf{f}$.

We will narrow down the focus of this optimization in (3) with respect to frame rate r because r significantly influence P.

$$\widehat{\mathbf{f}} = \underset{\mathbf{f}\in F,\, \mathbf{P}(r)<P_0}{\arg\max}\, s(\mathbf{f}, r), \tag{4}$$

where $\mathbf{P}(r)$ is the power consumption at a frame rate r, and $s(\mathbf{f}, r)$ is the score function with respect to FDA $\mathbf{f}$ and frame rate r. $\mathbf{f}$ is the optimal solution of PE-FDA.

(2) is written as follows,

$$\widehat{s} = \underset{s\in\mathbf{S},\, \mathbf{P}(r)<P_0}{\arg\min}\, E\left[s(\mathbf{f_0}, r) - s(\widehat{\mathbf{f}}, r)\right]^2. \tag{5}$$

Thus, the goal of benchmarking system for PE-FDA is described mathematically.

2.3 Challenges of the Novel Benchmarking System for PE-FDA

While the frame rate is considered for solving power consumption issue, there is a trade-off with respect to the accuracy of a benchmarking system. The score $s(\mathbf{f}, r)$ in (4) suffers from temporal error ϵ_t and spatial error ϵ_s due to the change of frame rate r and the property of Iou metric .

First, the temporal error is due to the frame rate mismatch of ground truth and (FDA) system output. After adjusting a frame rate of FDA, the benchmarking system may not find the frame in system output that can be mapped with any frame in ground truth, or vice versa. However, it is not recommended to generate ground truth that can be mapped to the new system output of the adjusted frame rate. While system output can be (frequently) regenerated for a purpose, it is costly to regenerate ground truth according to the adjusted frame rate. As a result, the ground truth and system output can not be compared (and/or evaluated) as shown in Figure 3.

Second, the spatial error is that the region-of-interest in ground truth (such as detection box) is not deployed where it should be. This is due to temporal error as explained above. After adjusting a frame rate of FDA, the mapped frame pairs of ground truth and system output are not generated at the same time. The offset of the region-of-interest in ground truth cause the spatial error. The effect of spatial error is critical to metric accuracy in *Iou metric*, the most popular evaluating metric. The score of Iou metric significantly decreases even with a slight offset of ground truth as shown in Figure 4, in which Iou metric is plotted on the y-axis when two unit detection boxes of ground truth and system output are deployed with the relative offsets $\Delta w = \Delta h = \Delta$ in the range of 0 to 1.

Thus, we need to consider these temporal and spatial challenges for a novel benchmarking system of PE-FDA. Improving spatial challenge, we introduce Unitized-distance (UD) metric in the next section (Section 3). Improving temporal error, we propose frame mapping algorithm in Section 4.1.

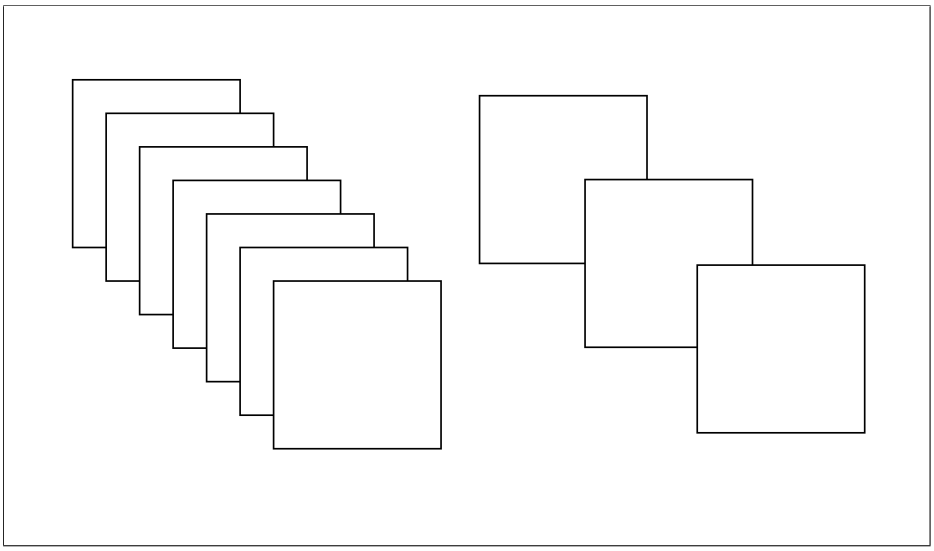


Figure 3: Temporal error: why the low frame rate does matter when designing benchmarking system of PE-FDA for AI-based security system?

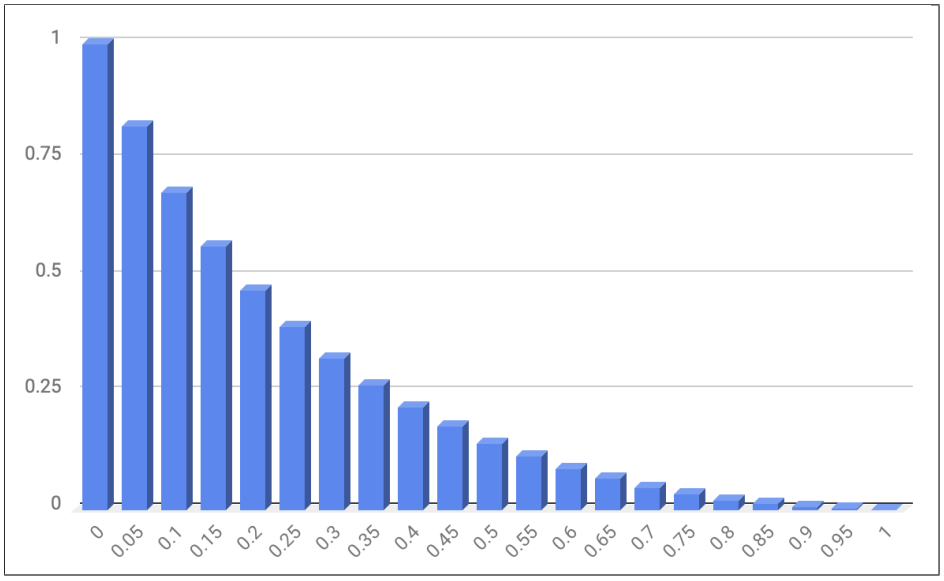


Figure 4: Spatial error: Iou metric is plotted on the y-axis when two unit detection boxes of ground truth and system output are deployed with the relative offsets Δ (=$\Delta w/w$ = $\Delta h/h$) in the range of 0 to 1 (increased by 0.05).

3 Unitized-distance (UD) Metric

In order to improve the spatial challenge mentioned above, we propose *Unitized-distance* (UD) metric that can discern the quality of FDA while reducing the spatial error ϵ_s for a benchmarking system for PE-FDA.

The conventional Iou metric decreases dramatically if the bounding box of ground truth is not perfectly matched with that of system output with as shown in Figure 4. This property of Iou metric has been issued that the Iou metric is vulnerable to ground truth boundary inconsistency as reported in [6]. For the same reason, the benchmarking system with Iou metric suffers significantly from spatial error. Due to the temporal adjustment after changing frame rate of FDA, there are slight misplacement of bounding box in all frames. This spatial error term of the score with Iou metric is accumulated over frames, and affects to the final score. In other words, if we use Iou metric for a benchmarking system for CE-FDA, this temporal error term of the score $s(\mathbf{f}, r)$ is a challenge to find $\widehat{\mathbf{f}}$ solving the optimization problem in (4). Thus, we propose a novel *Unitized-distance* (UD) base metric.

3.1 Definition of UD metric

Our UD metric m_{UD} is defined as follows,

$$m_{\text{UD}} = \tilde{d}(\mathbf{g}, \mathbf{s}) = \frac{1}{\sqrt[p]{2}} \sqrt[p]{\left(\frac{w_i}{w_u}\right)^p + \left(\frac{h_i}{h_u}\right)^p}, \tag{6}$$

where p is the base distance order. $p = 2$ is assumed for presenting numerically results in this paper. For other distance order p, this metric m_{UD} can be observed in Appendix B.

3.2 UD Metric Helps a Benchmarking System of PE-FDA Reduce Spatial Error

This UD metric reduce spatial error in a benchmarking system for PE-FDA. In order to help readers understand intuitively, the performance of this UD metric is compared with conventional Iou metrics in Figures 5 and 6. Iou metric is mathematically written as follows,

$$m_{Iou} = \frac{w_i \cdot h_i}{w_g \cdot h_g + w_s \cdot h_s - w_i \cdot h_i}, \tag{7}$$

where (w_i, h_i), (w_g and h_g), (w_s and h_s) are the width and height of bounding boxes of intersection, ground truth, system output, respectively.

In Figure 5, Iou metric is plotted on z-axis in Figure 5a, and it is found that the overlap metric linearly decreases if *either* Δx or Δy increases. UD metric is plotted on z-axis in Figure 5b when two unit detection boxes have the relative offsets $\Delta w/w$ and $\Delta h/h$ in the range of 0 to 1 along the x-axis and the y-axis, respectively. In this figure, it is found that the UD metric linearly decreases when *both* Δx and Δ increase (in the same amount). UD metric is larger than Iou metric in this Figure. It is mathematically provable.

Theorem 2 proves that UD metric is less affected by spatial error of a benchmarking system of PE-FDA. In order to prove Theorem 2, we need the following theorem,

Theorem 1 (the property of Iou metric)**.** *for any k s.t. $0 \leq k \leq 1$,*

$$m_{Iou} \leq k \cdot \frac{w_i}{w_u} + (1-k) \cdot \frac{h_i}{h_u}, \tag{8}$$

where $w_u = w_g + w_s - w_i$ and $h_u = h_g + h_s - h_i$.

We derived the proof of Theorem 1 in Appendix A. From Theorem 1, we can easily proof that the following inequality of UD metric m_{UD} and Iou metric m_{Iou},

Theorem 2 (the property between Iou metric and UD metric)**.** *the following inequality is true for m_{UD} in (6) and m_{Iou},*

$$m_{UD} \geq m_{Iou}, \tag{9}$$

the equality is true when $w_g = w_s = w_i$ and $h_g = h_s = h_i$.

Proof.

$$\begin{aligned} m_{\text{UD}} &= \frac{1}{\sqrt[p]{2}} \sqrt[p]{\left(\frac{w_i}{w_u}\right)^p + \left(\frac{h_i}{h_u}\right)^p} \\ &\geq \frac{1}{\sqrt[p]{2}} \sqrt[p]{(m_{Iou})^p + (m_{Iou})^p} = m_{Iou}. \end{aligned} \tag{10}$$

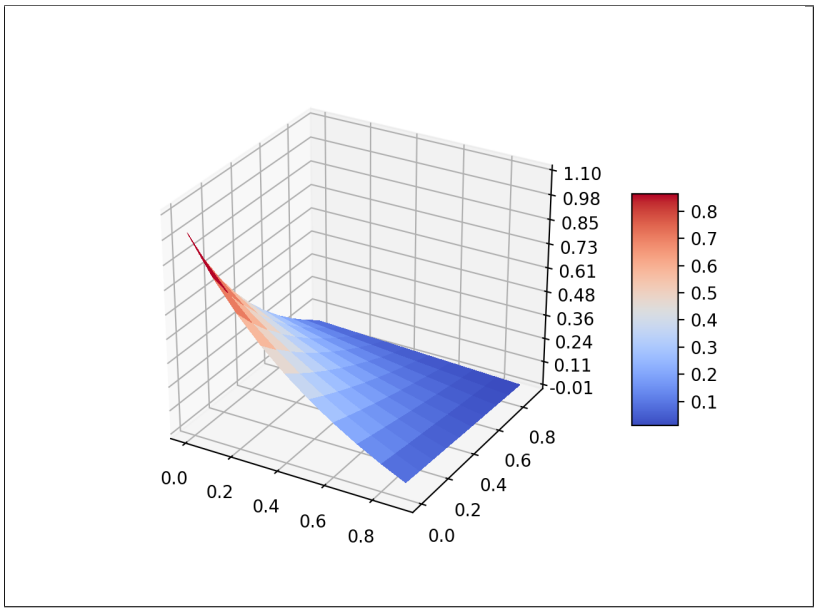
□

We can observe this property of Iou and UD metric in Theorem 2 is observed in Figure 6. Iou metric (blue-colored) and UD metric (red-colored) are observed on the y-axis when

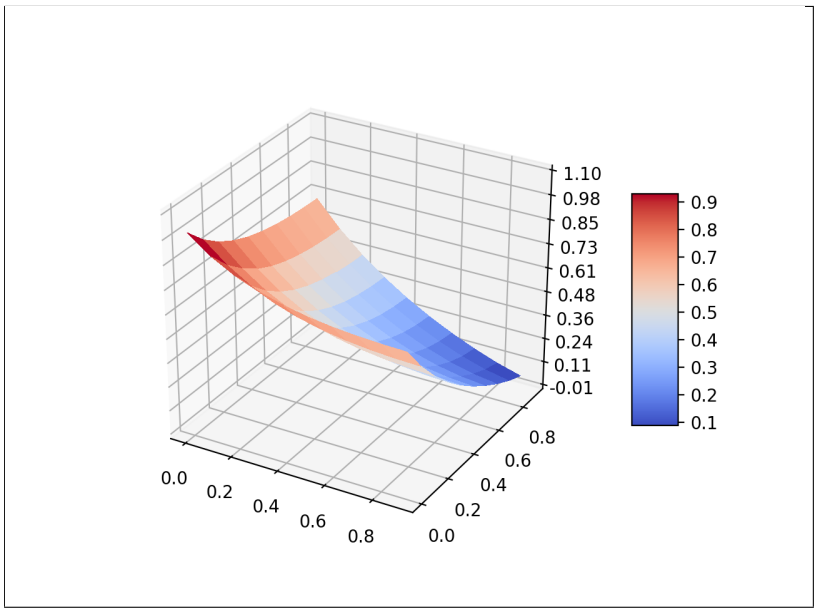
two unit detection boxes of ground truth and system output are deployed with the relative offsets Δ (=$\Delta w/w$ = $\Delta h/h$) in the range of 0 to 1 (increased by 0.05).

This figure tell us that UD metric is less affected by spatial error of a benchmarking system of PE-FDA than Iou metric. Thus, UD metric is a good option for designing a benchmarking system of PE-FDA in AI-based security system.

In this paper, $p = 2$ is chosen in order to provide readers the insight of this metric and numerical evidence when comparing with Iou metric. However, the order parameter p of UD metric can be chosen for even more minimizing the spatial error in the optimization problem of benchmarking system in (5). We leave the choice of order p for the future work. This propety of UD metric with respect to order p can be observed with Figures 9 and 10 in Appendix B.



(a) Iou base metric on the z-axis.



(b) UD base metric on the z-axis.

Figure 5: (a) Iou metric and (b) UD base metric are observed when two unit detection boxes have the relative offsets $\Delta w/w$ and $\Delta h/h$ in the range of 0 to 1 along the x-axis and the y-axis, respectively. It is found that while Iou metric is influenced by either changes of horizontal offset Δw or vertical offset Δh, UD metric is influenced when both Δw and Δh are changed.

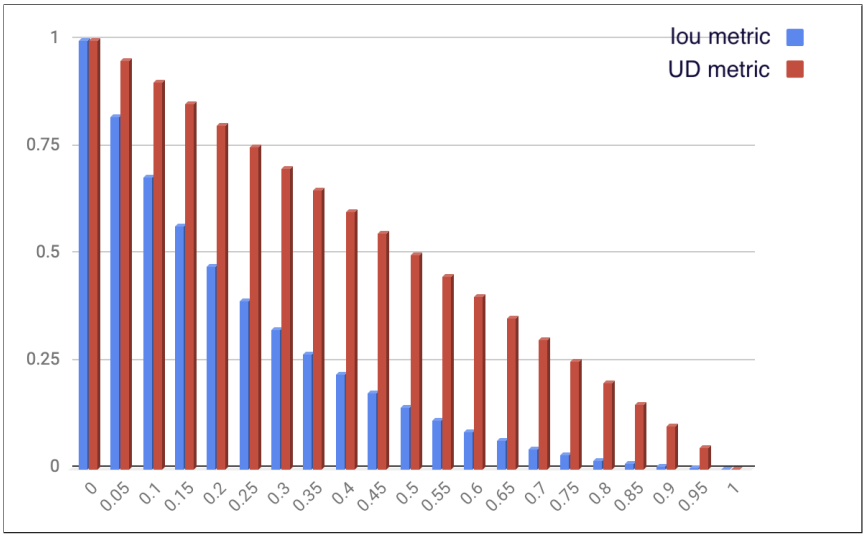


Figure 6: Iou metric (blue-colored) and UD metric (red-colored) are observed on the y-axis when two unit detection boxes of ground truth and system output are deployed with the relative offsets Δ (=$\Delta w/w$ = $\Delta h/h$) in the range of 0 to 1 (increased by 0.05).

4 Our Design and Implementation of Benchmarking System for PE-FDA

In this section, we will explain the framework of a benchmarking system of PE-FDA using our UD metric. First, we will discuss how this algorithm handle the frame rate mismatch of ground truth and system output in order to reduce temporal error when frame rate is changed. Second, we will present our implementation of a benchmarking system of PE-FDA for face identity matching system (FIMS) and pedestrian tracking system (PTS), respectively.

4.1 Frame Mapping Algorithm: Handling the Frame Rate Mismatch

As explained in the previous research, adjusting the frame rate for PE-FDA causes the design challenges of designing a benchmarking system for PE-FDA. Because the frame rates of ground truth and system output are not identical, we should solve the temporal error between the frames in ground truth and system output by handling frame fate mismatch. Frame Mapping Algorithm is a strategy for paring two comparable frames in ground truth and system output. Thus, we designed the frame mapping procedure which maps a frame

in ground truth corresponding to each frame of system output.

Then, for the frame number ${n_{\mathbf{s}}}^{(t)}$ of system $\mathbf{s}$ at the time stamp t, we aim to find the corresponding frame ${n_{\mathbf{g}}}^{(t)}$ of ground truth $\mathbf{g}$. The frame numbers ${n_{\mathbf{g}}}^{(t)}$ and ${n_{\mathbf{s}}}^{(t)}$ are calculated from time reference t and the frame rates $r_{\mathbf{g}}$ and $r_{\mathbf{s}}$ of $\mathbf{g}$ and $\mathbf{s}$, respectively, as follows,

$$n_{\mathbf{g}}^{(t)} = t \times r_{\mathbf{g}} \tag{11}$$

$$n_{\mathbf{s}}^{(t)} = t \times r_{\mathbf{s}} \tag{12}$$

Thus, we can find the ground truth frame number $n_{\mathbf{g}}^{(t)}$ corresponding to $n_{\mathbf{s}}^{(t)}$ as follows,

$${n_{\mathbf{g}}}^{(t)} = \left[\left(\frac{r_{\mathbf{g}}}{r_{\mathbf{s}}} \right) \cdot {n_{\mathbf{s}}}^{(t)} \right], \tag{13}$$

where $[\,\bullet\,]$ is the rounding notation (we need to round to calculate the frame number ${n_{\mathbf{g}}}^{(t)}$ because $n_{\mathbf{g}}$ is natural number).

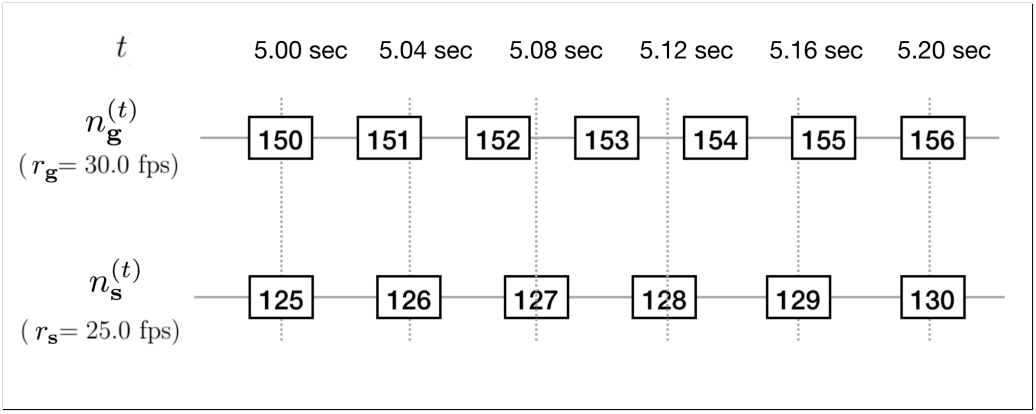


Figure 7: Our frame mapping algorithm handles the frame rate mismatch by finding the corresponding frame pairs of ground truth $\mathbf{g}$ and system $\mathbf{s}$.

4.2 Implementation of Our Benchmarking System of PE-FDA

In this section, we explain the implementation of our benchmarking (evaluation) system for face identity matching systems (FIMS) and pedestrian tracking systems (PTS). These designed evaluation algorithms corresponds to the *grading program* block of Figure 8.

Frame-based and object-based approaches are applied for implementing our evaluation systems for face identification matching algorithm and pedestrian tracking algorithm, respectively.

4.2.1 Our Benchmarking System for Face Identity Matching System (FIMS)

It is assumed that the face feature and identity of subjects enrolled in a database system are provided to this system evaluation algorithm. The pseudo-code of this evaluation system using UD metric for FIMS is given in Figure 1.

First, we obtain the matching frame n_g in ground truth that is pair-able for a given frame n_s in system $\mathbf{s}$ as described in 4.1. Second, we calculate $\tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right)$ in (6) where $\mathbf{g}_i[n_g]$ is ground truth $\mathbf{g}$'s the detection box of the i-th object at frame number n_g, and $\mathbf{s}_j[n_s]$ is system output $\mathbf{s}$'s detection box of the j-th object at frame number n_s. If no matching pair is found, its returned output value is zero. Third, after calculating our UD metric $\tilde{d}(\mathbf{g}, \mathbf{s})$ on multiple frames in a video, the metric $m_{\text{UD, FIMS}}$ is defined as the average of $\tilde{d}(\mathbf{g}, \mathbf{s})$ over frames as follows,

$$\begin{aligned} m_{\text{UD, FIMS}} &= Avg\left[\tilde{d}(\mathbf{g}, \mathbf{s})\right] \\ &= \frac{1}{|\mathbf{M}|} \sum_{(n_g, n_s) \in \mathbf{M}} \frac{1}{|\mathbf{O}_{n_g, n_s}|} \cdot \\ &\quad \left(\sum_{(i,j) \in \mathbf{O}_{n_g, n_s}} \tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right) \right), \end{aligned}$$

where the frame pair set $\mathbf{M}$ is obtained using the frame mapping algorithm as shown in (13),

$$\mathbf{M} = \{(n_g, n_s) | \quad n_g = \left[\left(\frac{r_{\mathbf{g}}}{r_{\mathbf{s}}} \right) \cdot n_s \right] \text{ for } {}^{\forall} n_s\}, \tag{14}$$

while the object pair set $\mathbf{O}_{n_g, n_s}$ is defined as follows,

$$\mathbf{O}_{n_g, n_s} = \{(i, j) | \quad i \in \mathbf{O}^g_{n_g} \text{ and } j \in \mathbf{O}^s_{n_s}\}, \tag{15}$$

where the set $\mathbf{O}^g_{n_g}$ is composed of the objects at the frame n_g in $\mathbf{g}$, and the set $\mathbf{O}^s_{n_s}$ is composed of the objects at the frame n_s in $\mathbf{s}$. Finally, we can calculate the mean of $m_{\text{UD, FIMS}}$ obtained over multiple videos.

4.2.2 Our Benchmarking System for Pedestrian Tracking System (PTS)

For implementing a benchmarking system for PTS, we will combine the values when using object-based approach. In this object-based approach, there is the ambiguity of matching pairs of tracking object IDs in ground truth and system output, and it is necessary to find

a matching pairs. In order to find matching pairs of tracking IDs in ground truth and system output, mapping list and correspondence table are attempted in [2] and [6]. For the simplicity, we build a correspondence table $\mathbf{T}$ in our implementation: we first can need to figure out the lists of available tracking object IDs in ground truth and system output, respectively. Then, the tracking object IDs in ground truth are listed on the row of a correspondence table while those in system output are listed on the column. It is assumed that pedestrian tracking systems assign (arbitrary) tracking object IDs for tracking all objects over video frames.

First, the object set $\mathbf{O}$ is defined as follows,

$$\mathbf{O} = \{(i,j) | \quad i \in \mathbf{O}^g \text{ and } j \in \mathbf{O}^s\}, \tag{16}$$

where the set $\mathbf{O}^g$ is composed of the objects in $\mathbf{g}$, and the set $\mathbf{O}^s$ is composed of the objects pair in $\mathbf{s}$.

Second, the set $\mathbf{M}_{i,j}$ is a set of the frame pairs (n_g, n_s) that contain the given object pair (i,j) in $(\mathbf{g}, \mathbf{s})$ as follows,

$$\begin{aligned}\mathbf{M}_{i,j} \quad = \{(n_g, n_s) \quad | \quad & {}^{\exists}\mathbf{g}_i[n_g] \in \mathbf{g} \text{ and } {}^{\exists}\mathbf{s}_j[n_s] \in \mathbf{s} \\ & \text{for a given object pair } (i,j)\}.\end{aligned}$$

Our metric $m_{\text{UD, PTS}}$ is defined as follows,

$$m_{\text{UD, PTS}} = \frac{1}{N} \cdot \sum_{(i,j)\in\mathbf{O}'} \sum_{(n_g,n_s)\in\mathbf{M}_{i,j}} \tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right),$$

where $\mathbf{g}_i[n_g]$ is the i-th object at the frame n_g in $\mathbf{g}$, $\mathbf{s}_j[n_s]$ is the j-th object at the frame n_s in $\mathbf{s}$, and the normalization factor N is defined as $N = \sum_{(i,j)\in\mathbf{O}'} \sum_{(n_g,n_s)\in\mathbf{M}_{i,j}} 1 = \sum_{(i,j)\in\mathbf{O}'} |\mathbf{M}_{i,j}|$.

This object pair set $\mathbf{O}'$ is the subset of $\mathbf{O}$ and it is defined as follows,

$$\mathbf{O}' = \{(i,j') \mid \quad j' = \arg\max_{j\in O^s} \tilde{D}\left(\mathbf{g}_i, \mathbf{s}_j\right), \text{ for } {}^{\forall} i \in \mathbf{O}^g\}. \tag{17}$$

where $\mathbf{O}^{\mathbf{g}}$ and $\mathbf{O}^{\mathbf{s}}$ are object pairs for $\mathbf{g}$ and $\mathbf{s}$, respectively, and $\tilde{D}\left(\mathbf{g}_i, \mathbf{s}_j\right)$ is the average of $\tilde{d}(\cdot)$ over the frame sets in $\mathbf{M}_{i,j}$.

$$\begin{aligned}\tilde{D}\left(\mathbf{g}_i, \mathbf{s}_j\right) &= Avg\left[\tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right)\right] \\ &= \frac{1}{|\mathbf{M}_{i,j}|} \sum_{(n_g,n_s)\in\mathbf{M}_{i,j}} \tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right)\end{aligned}$$

Because the (i, j) element of the correspondence table $\mathbf{T}$ is given as $\mathbf{T}_{i,j} = \tilde{D}(\mathbf{g}_i, \mathbf{s}_j)$, this set $\mathbf{O}'$ can alternatively be obtained from the table $\mathbf{T}$ by finding the column index j of the maximal $\mathbf{T}_{i,j}$ for the i-th row in $\mathbf{T}$. (17) can be rewritten as follows,

$$\begin{aligned} m_{\text{UD, PTS}} &= \frac{1}{N} \cdot \sum_{(i,j)\in\mathbf{O}'} N_{i,j} \cdot \\ &\quad \left(\frac{1}{|\mathbf{M}_{i,j}|} \sum_{(n_g,n_s)\in\mathbf{M}_{i,j}} \tilde{d}(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]) \right) \\ &= \frac{1}{N} \cdot \sum_{(i,j)\in\mathbf{O}'} N_{i,j} \cdot \tilde{D}(\mathbf{g}_i, \mathbf{s}_j) \\ &= \frac{1}{N} \cdot \sum_{(i,j)\in\mathbf{O}'} N_{i,j} \cdot \mathbf{T}_{i,j} \end{aligned} \tag{18}$$

where we define the variable $N_{i,j} = |\mathbf{M}_{i,j}|$ and $N = \sum_{(i,j)\in\mathbf{O}'} N_{i,j}$.

As shown in (18), $m_{\text{UD, PTS}}$ is differentiated from $m_{\text{UD, FIMS}}$ in (14) by $\tilde{D}(\cdot)$ - weighted-sum - according to the number of frame occurrence.

In Algorithm 2, we compute the metric $m_{\text{UD, PTS}}$ introduced in (17) for every tracking object ID pair of the entries in the correspondence table. Then, we can choose the tracking object ID of system output that corresponds to that of ground truth by finding the maximum of m_{UD} values in the row of ground truth's tracking object ID. Once the corresponding tracking ID pairs are chosen, a row and a column of its entry in the Correspondence Table are removed in the table for the next iteration. We will repeat this process until finding all matches for tracking object IDs in ground truth (the m_{UD} values for the final corresponding pairs are the maximum values found in the table). We will get the average value of the $m_{\text{UD, PTS}}$ values for the matching tracking object ID pairs).

Algorithm 1 Our benchmarking system with UD metric for PE-FIMS.

for a video $\in$ a set of test videos **do**
  Obtain the set $\mathbf{M}$ in (14)
  L = [ ]
  for $(n_g, n_s) \in \mathbf{M}$ **do**
    Obtain the set $\mathbf{O}_{n_g,n_s}$ in (15)
    K = [ ]
    for $(i,j) \in \mathbf{O}_{n_g,n_s}$ **do**
      $m_{\text{UD}} = \tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right)$
      K.append(m_{UD})
    end for
    L.append(Avg[K])
  end for
  $m_{\text{UD, FIMS}}$ = Avg[L] in (14)
end for
return mean, std of $m_{\text{UD, FIMS}}$ over test videos

Algorithm 2 Our benchmarking system with UD metric for PE-PTS.

for a video $\in$ a set of test videos **do**
  Obtain the set $\mathbf{O}$ in (16)
  $\mathbf{T}_{i,j} = 0$ for $^{\forall}i \in \mathbf{O}^g, ^{\forall}j \in \mathbf{O}^s$
  for $(i,j) \in \mathbf{O}$ **do**
    Obtain the set $\mathbf{M}_{i,j}$ in (17)
    for $(n_g, n_s) \in \mathbf{M}_{i,j}$ **do**
      $m_{\text{UD}} = \tilde{d}\left(\mathbf{g}_i[n_g], \mathbf{s}_j[n_s]\right)$
    end for
    Calculate $\tilde{D}\left(\mathbf{g}_i, \mathbf{s}_j\right)$ in (18)
    $\mathbf{T}_{i,j} = \tilde{D}\left(\mathbf{g}_i, \mathbf{s}_j\right)$
  end for
  Obtain the set $\mathbf{O}'$ in (17) using $\mathbf{T}$
  Calculate $m_{\text{UD, PTS}}$ in (18)
end for
return mean, std of $m_{\text{UD, PTS}}$ over test videos

5 Experiment and Results

Validating of our benchmarking system design for FIMS and PTS, the designed implementation in Sections 4 is tested with actual video data from the cameras of a state-of-the-art security system.

5.1 Overall Framework of Our Benchmarking System for PE-FDA

The overall framwork including generating ground truth for the video data set is given in Figure 8.

First, multiple videos are simulated when evaluating the performance of an individual face detection algorithm in order to be not biased by a particular test video setting. For each video in a list of videos, data files for ground truth and system out are generated. A system output data file is generated by processing a video using an face detection algorithm we want to test. Thus, there is only one data file for the ground truth while a system output data file is generated for each face detection algorithms to be tested.

Second, the data files for ground truth and system output are composed of detected object's information for all frames in a video. For every object in a frame, we generated a detection box indicating the object's face in terms of x and y positions, width and height. While subjects' identity in a database system is offered for testing a face ID matching

algorithm, tracking ID is offered for a pedestrian tracker algorithm.

Third, once metric values are calculated on multiple videos for a tested FDA, we obtain a mean of the evaluation results of a designed algorithm for the final performance evaluation metric.

Note: Mean of Metrics on Multiple Video Streams - For reducing bias caused by a test video, we obtain the mean of the metric values on various video examples. For a video, we can calculate the designed metric with our designed system presented in Section 4. However, in order to minimize bias due to a video source, we want to utilize multiple videos for system performance evaluation as given in Figure 8. Thus, we will provide mean of the calculated values on multiple videos of each system. A well-designed system will have the final outputs of a higher final mean value.

5.2 Ground Truth Generation for Actual Video Data

Because the video data set from actual state-of-the-art security system is used for the experiment of benchmarking system, we generate ground truth of video data.

Generally speaking, generating ground truth is expensive. It is essential that once this ground truth is generated, we do not regenerate the ground truth due to the changed frame rate of FDA. Instead, we will apply the frame mapping algorithm in Section 4.1

We generate ground truth in the blue-colored block of Figure 8. While the previous approaches are using human inputs only, our system framework adapts the pre-processing performed in computer and then verified by human input in order to minimize subjectivity. This minimization is important to maintain ground-truth quality. Generating a ground truth data file, a video is first processed to pre-label objects identity in a video stream. Then, a human input is required to validate the pre-labels and correct only mislabel on detection bounding boxes.

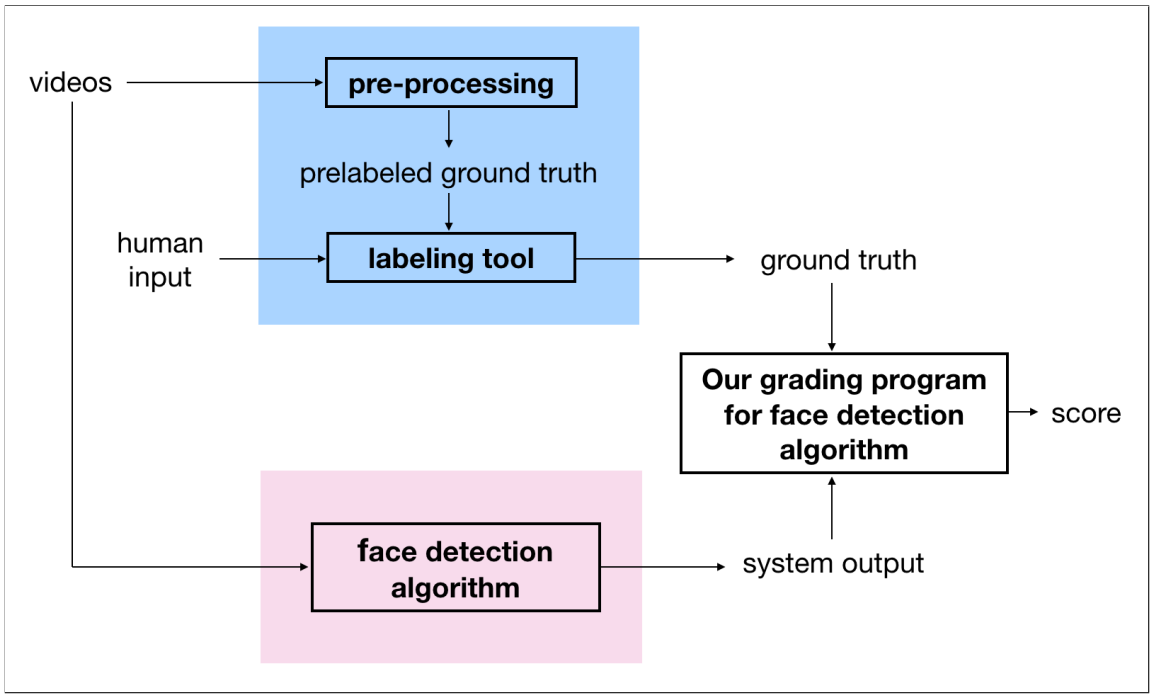


Figure 8: The system framework of our performance evaluation system: (1) this performance evaluation system first performs the pre-processing step of detection in computer, and then processes the verification step by human input. In this way, we can to minimize subjectivity and reduce the cost of ground truth generation. (2) the evaluation score on system $\mathbf{s}_k$ will be provided for each video stream.

5.3 Simulation Setting

Our benchmarking system design of PE-FDA is validated on video data obtained from a state-of-the-art security systems. In the second validation step with actual videos, four videos for FIMS and nine videos for PTS are used. Also, UD metric and Iou metric are applied, separately. It includes a few hundreds frames, and the total frame numbers of the videos are 3348 and 5002 for FIMS and PTS, respectively. (Figure 2 shows a couple of frame images in videos when a pedestrian tracking algorithm [1] is applied.)

For this validation, we chose one of the latest FDAs. An state-of-the-art FIMS is obtained from CMU Cylab Biometrics Laboratory [20], and a PTS is obtained from [1] that is lately implemented with CNN.

5.4 Results

Our simulation results on multiple frame rates and multiple videos from actual security system are given in Tables 1 and 2, respectively.

In Table 1, the evaluation results on multiple frame rates in sample scenarios are given. It is observed in Table 1b that UD metric is less affected by the variation of FDA frame rates than Iou metric is. As mentioned in Section 3.1, this UD metric is even less affected by frame rate variation if p in (6) increases. Also, this result shows that UD metric has a higher values overall as proved in Theorem 2.

In Table 2, the evaluation results on actual video data are given. In this evaluation of a face identity matching algorithm, we compute the mean of calculated values of the final metric in equation (14) on multiple videos. For the purpose of metric quality comparison, the UD metric in (14) and (17) is replaced with the Iou metric for FIMS and PTS, respectively. The simulation results in Table 2 also show that UD metric has a higher values overall as proved in Theorem 2. The average of the evaluation results with our UD metric shows 140% and 300% improvements for the FIMS and the PTS, respectively.

Table 1: Evaluation results on multiple frame rate of sample scenarios

(a) FIMS

	UD metric (14)	Iou metric
r_g=r_s=30.0 fps	0.4848	0.28
r_g=30.0 fps, r_s=15.0	0.6667	0.6667
r_g=30.0 fps, r_s=10.0	0.5	0.5
r_g=30.0 fps, r_s=20.0	0.8889	0.8570

(b) PTS

	UD metric (17)	Iou metric
r_g=r_s=30.0 fps	0.9070	0.7333
r_g=30.0 fps, r_s=15.0	0.8141	0.4667
r_g=30.0 fps, r_s=10.0	0.9303	0.8000
r_g=30.0 fps, r_s=20.0	0.9303	0.8000

Table 2: Evaluation results on multiple videos from actual state-of-the-art security system

(a) FIMS [20]

	UD metric (14)	Iou metric
video # 1	0.0218	0.0150
video # 2	0.1745	0.1256
video # 3	0.1046	0.0600
video # 4	0.4359	0.3271
avg	0.1842	0.1319

(b) PTS [1]

	UD metric (17)	Iou metric
video # 1	0.4230	0.1534
video # 2	0.3721	0.1699
video # 3	0.1975	0.0595
video # 4	0.4147	0.1608
video # 5	0.3249	0.0632
video # 6	0.4165	0.1198
video # 7	0.3998	0.1728
video # 8	0.3708	0.0917
video # 9	0.3522	0.0875
avg	0.3635	0.1198

Therefore, the simulation results demonstrate that our benchmarking system is recommendable to solve the optimization problem of finding PE-FDA for actual security system.

6 Conclusion

In this paper, we designed a benchmarking system for power-efficient face detection algorithm (PE-FDA) that can be implemented in state-of-the-art artificial intelligence (AI) based security system. In a practical security system, power consumption should an important specification when finding an optimal face detection algorithm (FDA). Thus, in order to overcome this practical limitation of power consumption, this paper focuses on a benchmarking system of PE-FDA.

Power consumption in FDA is directly controllable by adjusting its frame rate. However, adjusting a frame rate is a challenge for designing a benchmarking system for FDA due to temporal error and spatial error. In order to overcome these errors in benchmark-

ing system of PE-FDA, we proposed a novel metric, *Unitized-distance* (UD) metric, designed a face mismatch algorithm, and implemented benchmarking systems for face identity matching system (FIMS) and pedestrian tracking system (PTS). Finally, comparing with the most popular conventional metric (Iou metric), our benchmarking system was validated for multiple frame rates and multiple videos obtained from actual state-of-the-art security system. Thus, we conclude that our system is recommendable for the benchmarking evaluation system of PE-FDA.

In the future, we expect that our approaches for developing the benchmarking tool (including the properties of UD metric) will contribute to improve power efficiency in other AI software applications as well as the future benchmarking system for AI-based systems.

Appendices

A Proof of Iou Metric's Property

for any k s.t. $0 \leq k \leq 1$, m_{Iou} satisfies the following inequality,

$$m_{Iou} \leq k \cdot \frac{w_i}{w_u} + (1-k) \cdot \frac{h_i}{h_u}, \tag{19}$$

where $w_u = w_g + w_s - w_i$ and $h_u = h_g + h_s - h_i$.

Proof.

$$m_{Iou} = \frac{w_i \cdot h_i}{w_g h_g + w_s h_s - w_i h_i}, \tag{20}$$

where $w_i \leq \min\{w_g, w_s\}$ and $h_i \leq \min\{h_g, h_s\}$.

(20) can be re-written as follows,

$$\begin{aligned} m_{Iou} &= \frac{w_i \cdot h_i}{(w_g + w_s - w_i) \cdot h_i + (w_g \cdot (h_g - h_i) + w_s \cdot (h_s - h_i))} \\ &= \frac{w_i \cdot h_i}{(w_g + w_s - w_i) \cdot h_i + \epsilon_1} \\ &\leq \frac{w_i \cdot h_i}{(w_g + w_s - w_i) \cdot h_i} = \frac{w_i}{w_g + w_s - w_i}, \end{aligned} \tag{21}$$

where $\epsilon_1 = w_g \cdot (h_g - h_i) + w_s \cdot (h_s - h_i) \geq 0$.

Symmetrically,

$$\begin{aligned} m_{Iou} &= \frac{w_i \cdot h_i}{w_i \cdot (h_g + h_s - h_i) + ((w_g - w_i) \cdot h_i + (w_s - w_i) \cdot h_i)} \\ &= \frac{w_i \cdot h_i}{w_i \cdot (h_g + h_s - h_i) + \epsilon_2} \\ &\leq \frac{w_i \cdot h_i}{w_i \cdot (h_g + h_s - h_i)} = \frac{h_i}{h_g + h_s - h_i}, \end{aligned} \tag{22}$$

where $\epsilon_2 = (w_g - w_i) \cdot h_i + (w_s - w_i) \cdot h_i \geq 0$.

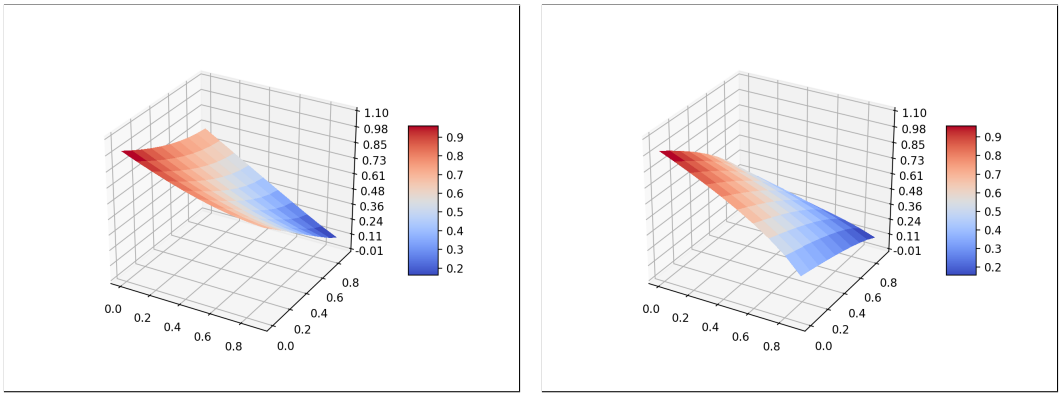
By (21) and (22),

$$m_{Iou} \leq k \cdot \frac{w_i}{w_u} + (1-k) \cdot \frac{h_i}{h_u}. \tag{23}$$

□

B UD Metric's Property with respect to Distance Order p

In Figure 9, UD metric is observed when p is changed. UD metric with (a) p = 2, (b) p=1, (c) p=.1, (d) p=.01 when two unit detection boxes have the relative offsets $\Delta w/w$ and $\Delta h/h$ in the range of 0 to 1 along the x-axis and the y-axis, respectively.



(a) UD metric for p=2

(c) UD metric for p=.1

(b) UD metric for p=1

(d) UD metric for p=.01

Figure 9

In Figure 10, Iou metric and UD metric are plotted when two unit detection boxes have the relative offset $\Delta h/h$ in the range of 0 to 1 and $\Delta w/w = 0$. Iou metric and UD metric for (a) p = 2, (b) p=1, (c) p=.1, (d) p=.01 are plotted when two unit detection boxes have the relative offset $\Delta h/h$ in the range of 0 to 1 along the x-axis while $\Delta w/w = 0$.

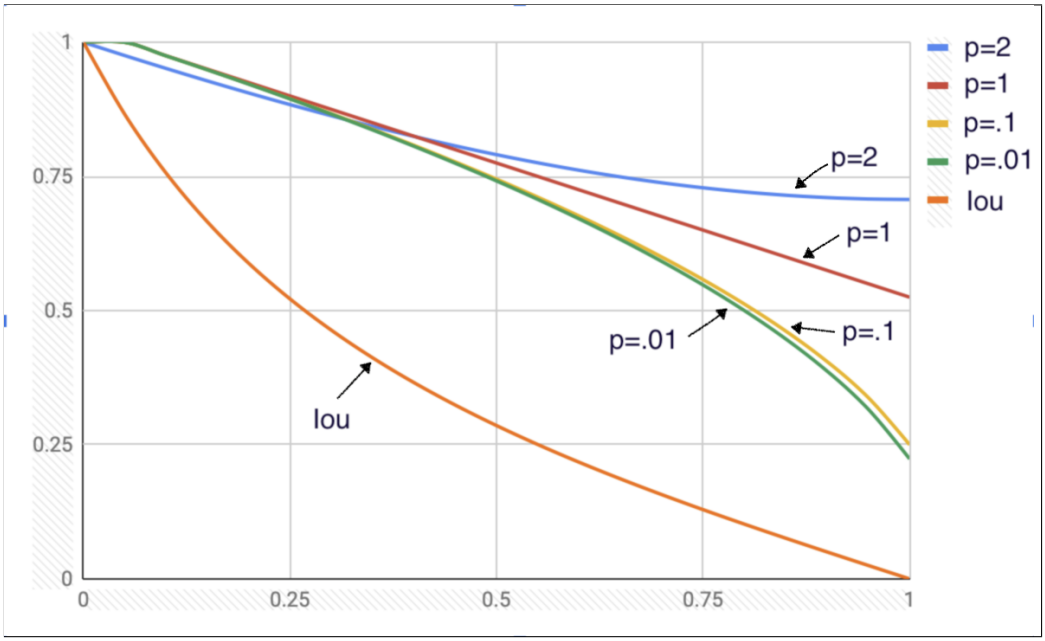


Figure 10

References

[1] Linzaer, "Face-Track-Detect-Extract," https://github.com/Linzaer/Face-Track-Detect-Extract/, 2018, [Online; accessed 18-November-2018].

[2] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear mot metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, no. 1, p. 246309, May 2008. [Online]. Available: https://doi.org/10.1155/2008/246309

[3] Y. Li, A. Dore, and J. Orwell, "Evaluating the performance of systems for tracking football players and ball," in *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*. IEEE, 2005, pp. 632–637.

[4] A. T. Nghiem, F. Bremond, M. Thonnat, and V. Valentin, "Etiseo, performance evaluation for video surveillance systems," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*. IEEE, 2007, pp. 476–481.

[5] K. Smith, D. Gatica-Perez, J.-M. Odobez, and S. Ba, "Evaluating multi-object tracking," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 36–36.

[6] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 319–336, Feb 2009.

[7] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.

[8] J. C. Nascimento and J. S. Marques, "Performance evaluation of object detection algorithms for video surveillance," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 761–774, 2006.

[9] F. Bashir and F. Porikli, "Performance evaluation of object detection and tracking systems," in *Proceedings 9th IEEE International Workshop on PETS*, 2006, pp. 7–14.

[10] T. Ellis, "Performance metrics and methods for tracking in surveillance," in *Proceedings of the 3rd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS'02)*, 2002, pp. 26–31.

[11] N. Lazarevic-McManus, J. Renno, D. Makris, and G. A. Jones, "An object-based comparative methodology for motion detection based on the f-measure," *Computer Vision and Image Understanding*, vol. 111, no. 1, pp. 74–85, 2008.

[12] C. J. Needham and R. D. Boyle, "Performance evaluation metrics and statistics for positional tracker evaluation," in *International Conference on Computer Vision Systems*. Springer, 2003, pp. 278–289.

[13] L. M. Brown, A. W. Senior, Y.-l. Tian, J. Connell, A. Hampapur, C.-F. Shu, H. Merkl, and M. Lu, "Performance evaluation of surveillance systems under varying conditions," in *Proceedings of IEEE Pets Workshop*. Citeseer, 2005, pp. 1–8.

[14] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.

[15] G. G. Chrysos, E. Antonakos, P. Snape, A. Asthana, and S. Zafeiriou, "A comprehensive performance evaluation of deformable face tracking "in-the-wild"," *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 198–232, 2018.

[16] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, "A metric for performance evaluation of multi-target tracking algorithms," *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3452–3457, 2011.

[17] J. R. Hoffman and R. P. Mahler, "Multitarget miss distance via optimal assignment," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 34, no. 3, pp. 327–336, 2004.

[18] L. Rueshendorff, "Wasserstein (vasershtein) metric," 2001.

[19] S. Pingali and J. Segen, "Performance evaluation of people tracking systems," in *Applications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on*. IEEE, 1996, pp. 33–38.

[20] K. Luu, C. Zhu, C. Bhagavatula, T. H. N. Le, and M. Savvides, "A deep learning approach to joint face detection and segmentation," in *Advances in Face Detection and Facial Image Analysis*. Springer, 2016, pp. 1–12.

[21] A.-T. Nghiem, F. Bremond, M. Thonnat, and M. Ruihua, "A new evaluation approach for video processing algorithms," in *IEEE Workshop on Motion and Video Computing*, 2007.

[22] B. E. Fridling and O. E. Drummond, "Performance evaluation methods for multiple-target-tracking algorithms," in *Signal and Data Processing of Small Targets 1991*, vol. 1481. International Society for Optics and Photonics, 1991, pp. 371–384.

[23] F. Yin, D. Makris, and S. A. Velastin, "Performance evaluation of object tracking algorithms," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Rio De Janeiro, Brazil*. Citeseer, 2007, p. 25.