Characterizing the Utility of Surveillance Video for Person Re-Identification

Haoyu Chen
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN
chen1562@purdue.edu

Edward J. Delp
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN
ace@ecn.purdue.edu

Amy R. Reibman
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN
reibman@purdue.edu

***Abstract*—Surveillance videos have many applications which can now be accomplished with automated systems. However, the performance of an automated system under realistic circumstances usually does not match the performance produced with a benchmark dataset, due to quality degradations and content variations. In this paper, we discuss several quality and non-quality factors that impact algorithm performance, and design a metric that measures the utility of a video for performing a specific task. Having such measurement allows a user to compress and/or select videos prior to implementing video analytics, reducing both computing power and storage. Here, we choose to study one of the most common applications for video analytics—person re-identification—and we call the corresponding quality measurement "identifiability".**

***Index Terms*—surveillance video, identifiability, task-specific quality evaluation**

I. Introduction

As the number of surveillance systems deployed increases drastically over the years, many automated systems have been designed to assist the human viewers in analyzing these videos, and many of these applications can be used for homeland security purposes as well [1,2]. However, although these automated systems may have acceptable accuracy in laboratory experiments, things become more complicated in realistic scenarios. Intuitively, real-life surveillance videos usually have worse quality than benchmark datasets, due to factors such as compression, resolution, and noise. In addition, for a specific task, many content-related non-quality factors such as lighting or background clutter also have noticeable impact on the system's performance. Hence, having a measurement that characterizes a video's utility for performing a specific video analytic task, by quantifying these factors, will be helpful. Such measurement could assist the user in finding the most suitable videos as the input of a system or adjusting the system according to the utility level of input videos.

In this paper, we are focusing on person re-identification (re-id) task as an example. The problem of re-id was first defined as "to identify object B encountered at time t_2 as the same object as object A encountered at time t_1" in [3]. In our case, we consider a person re-id system with multiple cameras, where we would like to find individuals captured by one camera in the video that has been, or will be, captured by one or more other cameras. The first camera generates the "query" or "probe" images, while the additional cameras constitute the "gallery". Both query and gallery images may be compressed, transmitted, and stored from one location to another. In addition, a person-detector processes each video (before or after compression) to generate candidate locations for the people present in each video. These locations are typically stored as bounding-boxes, or rectangles inside which the person is believed to be located. As is common practice, in our system the person re-id algorithm takes these automatically generated bounded boxes as input.

The performance of re-id algorithms are typically evaluated using top-1, top-5, and top-10 accuracy values, based on the premise that multiple candidate matches will be returned to a human agent who will use these results to take appropriate action. However, these average performance metrics do not tell the full story about how well the system might perform in specific cases. In particular, some individuals may have characteristics that make them significantly easier to identify than others, due to unique personal attributes. Moreover, re-id performance is also affected by other factors, including environmental conditions, image resolution, compression, and the accuracy of the bounding boxes.

Consider the impact of quality. While the very popular benchmark dataset Market-1501 [4] and its full-frame version PRW [5] have a resolution of 1920x1080, most videos in the "Cleveland" dataset [6], collected from real surveillance cameras installed in facilities of the Greater Cleveland Regional Transport Authority (GCRTA), have resolution of 352x240. Some previous work has been done in the past, to investigate the ability for the low-quality Cleveland dataset to perform facial re-id [6] and person re-id [7].

Next, consider the impact of non-quality issues. While most benchmark re-id datasets, including CUHK03 [8], Market-1501/PRW [4,5], and DukeMTMC-ReID [9,10], were taken within certain area (many of them use university campuses), real-life surveillance videos have a much wider range of

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 17STCIN00001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

variation in environment settings such as background, viewpoint, and lighting, which could also greatly affect a system's performance.

Sample images of Cleveland and Market-1501 images can be found in Figure 6 and 7. We can observe that while the Market-1501/PRW dataset consists of scenes with very similar background and lighting condition, scenes from the Cleveland dataset have more diversified background and settings.

Hence, our goal in this paper is to establish a utility measurement that models the relationship between both quality and non-quality factors of an input video and the corresponding system performance. Having such measurement allows the user to reduce storage and computing power waste by selecting suitable videos as the input or by compressing videos with higher utility than system requirement.

For this purpose, we propose a guideline that identifies key factors, quantifies these factors' impact, and finally determines the utility of a surveillance video for performing a specific video analytic task. Since the specific task we consider in this paper is person re-id, we use the word "identifiability" to denote this measurement, which, intuitively, means the ability for performing identification task. Under this scheme, videos with higher identifiability would also have both quality and non-quality factors that are suitable for performing person re-id, and vise versa. We then demonstrate experimentally that our identifiability metric can provide useful information for two sample purposes: to create specifications on compression and bounding-box accuracy, and to indicate when a gallery video does not have high-enough quality to produce the desired re-id accuracy.

II. Related Work

A. Person Re-Identification

Currently, state-of-the-art re-id algorithms have achieved high accuracies on existing datasets. For video-based re-id, algorithms such as patch-based appearance [7] and two-stream Siamese network [11] have reached decent accuracies on both iLIDS-VID dataset [12] and PRID2011 dataset [13]. The best model evaluated with the iLIDS-VID dataset [12], the part appearance mixture model (PAM) [14], has obtained 95% for rank-20 identification rate and 75% for rank-1 identification rate. For single-shot re-id datasets, several deep-learning based [15,16] have reached 90% for rank-1 identification rate and 80% for mean average precision score (mAP) with the popular Market-1501 dataset [4]. In [17], the authors proposed to use GAN to model the relationship between cameras and it performed well on both DukeMTMC-ReID [9,10] and Market-1501 dataset [4]. For smaller datasets that have insufficient images to train a neural network, hand-crafted features such as GOG [18] produces a reasonable result, with rank-10 identification rate of 88% on the VIPeR [19] dataset.

In this paper, we implemented several existing algorithms for performing re-id task with surveillance videos. We used Faster R-CNN [20] as person detector to extract bounding boxes from raw surveillance videos. And for performing re-id task, we chose the GOG [18] feature extractor, one of the best handcrafted feature that does not need any training data. Because the Market-1501 dataset provides a training set, we apply the KISSME metric learning [21] to get better results, but for Cleveland dataset, we used Euclidean distance for matching.

B. Task-Specific Quality Assessment

Traditional video/image quality assessment methods focus on human perception [22,23]. However, as more automated image/video analytic tools have been developed, attention has shifted to also consider how quality affects system performance. Some earlier papers discussing this topic include [24], where the authors provide a guideline for objective quality assessment, and [25], where the authors discuss the quality of biometric systems, using fingerprints as an example.

More recently, people have explored the impact of quality on a specific task. Face identification is one of the popular tasks people have been focusing on. The authors of [26] compare human perception of image quality and the quality score computed from face matchers as quality assessment criteria, and the authors of [27] use two tasks—distorted face identification and verification—to evaluate the quality of compressed surveillance videos. In [28], the authors discuss image degradations including noise, blur, contrast, occlusion, etc., but they do not consider compression. Taking a different path, researchers in [29,30] have proposed a method motivated by visual psychophysics to examine how facial recognition and object recognition algorithms react to different kinds of degradations.

As for other tasks, a quality prediction model for pedestrian detection is proposed in [31], and [32] utilized the same idea to create a quality-adaptive system that improves detection performance. For the object detection and recognition task, the authors of [33]–[35] considers factors causing degradation, such as blur, noise, or resolution, and construct corresponding distorted video dataset for analysis.

For the task of re-id, however, only limited amount of research has been done on the impact of quality on a re-id system. One benchmark paper [36] has briefly discussed the types of variability present within popular re-id datasets, such as viewpoint, illumination, resolution, etc., but it did not quantify the impact of these variations on re-id performance.

In this paper, we propose a method that considers both quality and non-quality factors to determine a utility score, denoted "identifiability".

III. Identifiability Metric

As described briefly in the introduction, our goal is to introduce a video utility metric that quantifies quality/non-quality factors impacting the performance of a specific task. In achieving that goal, there are three essential steps:

Step 1: Identifying key factors that affect task performance.
Step 2: Applying evaluation algorithms that quantify the utility of each factor.

Step 3: Applying score fusion with proper weight distribution, based on how much impact each factor has on the desired task.

The output score from the steps above represents a video's utility for a specific task. In this paper, we use person re-id as our example task for evaluation, where we use the word "identifiability" to denote the output utility measurement. Some applications for its use are discussed in Section III-D.

A. Identifying key factors

To implement step 1, we consider three major aspects that affect a video's identifiability: environment, capture system, and personal attributes of an individual, along with one optional aspect—similarity between query and gallery context.

Environment factors include lighting, background, and occlusion. Inconsistent lighting may cause the captured identities in the scene to appear brighter or darker than they normally look. For a pixel-based feature extraction algorithm, the background may introduce extra texture around a person and therefore cause inaccuracy in the feature descriptor. And occlusion damages the completeness of an identity either by adding extra pattern/texture that does not belong to the identity or by blocking key features.

System factors include initial camera resolution, camera angle, lens specification, and compression applied for transmission or storage purpose. These factors are hard to acquire directly from a video, without knowing the original setup and specifications. However, we are still able to estimate the impact of these factors based on visual cues such as blockiness, blurry edges, and/or the size of a person.

Attributes of an individual, including unusual color, distinct pattern and/or noticeable accessories, reflect whether the identity is distinct to human perception. In evaluating a surveillance video, however, it is impractical to extract all identities and see if they have distinct features. Therefore, we use color vividness and texture sharpness of the observed individuals to represent the video's overall ability of preserving ones' distinct features.

Similarity between query and gallery context is an additional aspect that affects identifiability. For example, if the target identity (query) is extracted from an airport surveillance video, it would be easier to re-identify this identity from another airport's surveillance videos (gallery), than from videos captured at a restaurant or grocery store. However, this information may not be available in all instances, so this would be an optional choice.

B. Quantifying the impact of each factor

In the current paper, we are not considering an automated evaluation system, which requires incorporating components including key-frame extraction, person detection, background subtraction, and texture/color analysis, as it consumes too much time and computing power.

However, a human viewer only needs a short amount of time to identify and assess the color vividness, texture sharpness of one identity appeared in the video as well as the background settings including where people would appear more frequently and if the background/lighting around that area is changing drastically. Therefore, while an automated evaluation system is being developed, we propose to use human viewers' observation as a demonstration of our idea in this paper.

Under this scheme, we need to emphasize that the human scoring must be limited to the specific factors we deem important for the re-id task. This is important because there exist some instance where videos with high subjective quality score may not be suitable for performing identification task. For example, a high-resolution video taking very far away from the ground may have a high subjective score, but each identity observed would be too small for either human or re-id algorithms to extract useful information.

In Table I, we provide a sample questionnaire for evaluating a video's identifiability. For each criterion, a "good" results in +1 point, an "average" is 0, and a "bad" is -1.

C. Weight distribution

With the questionnaire provided in previous section, we then perform a weighted sum to generate the output identifiability score. Weights for each criterion are set by their relative importance. In this case, since the GOG feature extractor uses both color and gradient-based pixel features of the whole bounding box, we choose to distribute more weight to color, texture, and background factors, since they all directly affect the feature descriptor. Also, as mentioned previously, if the query is provided, the similarity between query and gallery video context can have significant impact on a video's identifiability.

Therefore, two sample weight distributions, based on person re-id with GOG feature descriptor, can be found in Table II. Weight Allocation 1 is for instances where query is unknown and the user only wants to evaluate the utility of the gallery video; otherwise, Weight Allocation 2 should be used, with an additional factor being the similarity between query and gallery context.

In this paper, Weight Allocation 2 is adopted as we have the information on both query and gallery.

D. Applying the identifiability metric

Our identifiability metric, computed using the steps above, creates a score indicating a video's utility for performing the person re-id task. Ideally, a higher score indicates better utility, while a lower or even negative score implies that the video might not be suitable for performing the task. Based on that measurement, there are two possible applications: adjust the system design based on input videos' identifiability scores, and prioritize videos with higher identifiability.

In the first application, we can adjust the system design to reduce computation and storage cost. For example, we can envision applying more compression or a weaker person detector to those videos with high identifiability, without impacting the system's ability to do effective re-id, since those videos with higher identifiability tend to be more robust against system-related degradations.

TABLE I: Evaluation Criterion for Identifiability

Criterion	Good	Average	Bad
Background/ Lighting	clear and consistent background with uniform lighting	acceptable background and/or lighting variations	drastic lighting and/or background changes, messy background
Occlusion	no occlusion that affect completeness of a pedestrian	minor or occasional occlusion	heavy occlusion that damages pedestrian completeness
Color	clear color with proper saturation	some color distortion and/or low saturation	severe color distortion and/or obvious unnaturalness
Texture	details and edges that can be observed clearly	edges and details that are blurred but still visible	edges that are hard to identify, insufficient resolution
Distortion	no distortion can be observed	some distortion in corners, not affecting pedestrian completeness	severe distortion that causes visible stretch or compress on people
(optional) Similarity	settings that closely simulate the target scenario	setting that has some similarity, in terms of viewpoint or distance to people	settings that are drastically different from the target scenario

TABLE II: Weight Allocation for Evaluation Criteria

Criterion	Weight Allocation 1	Weight Allocation 2
Background/ Lighting	20%	15%
Occlusion	10%	10%
Color	30%	20%
Texture	30%	20%
Distortion	10%	5%
Similarity	N/A	30%

For the second application, we can prioritize our computational power to focus on videos with a higher identifiability score. According to our design, low or negative identifiability scores would indicates that it may be difficult to do effective re-id with these videos, while videos with higher identifiability are more likely to produce reliable results. In the extreme, it could be imagined that videos with a low enough identifiability score could be discarded completely.

The requirement for our identifiability score for the first application is that we are able to compress more heavily, or have a less accurate bounding box for, those individuals who are intrinsically more identifiable. The requirement for the second application is that our score for a video is well correlated with the accuracy of a re-id algorithm performed on it. We will explore whether our identifiability score satisfies these two requirements in the next section.

IV. Experiment and Results

In this section, we present two experiments to demonstrate the potential applications of our identifiability metric. The first investigates the impact of system-related degradations on identifiability. Our goal is to demonstrate that videos containing highly-identifiable individuals can endure more compression and tolerate less accurate bounding-box detectors. The second experiment explores the degree of correlation between our identifiability score and the level of re-id performance that can be obtained. A high correlation would motivate the ability to prioritize computational effort on videos with high identifiability.

A. Impact of system-related degradations

We begin by investigating the impact of system-related degradations. Under our assumptions, the identifiability of one image or one video is jointly affected by both quality and non-quality factors. Therefore, higher scores in non-quality factors should offset degradations in quality aspects.

For the experiment, we choose personal attributes to represent the non-quality factor, and compression and inaccurate bounding boxes to represent system-related degradations.

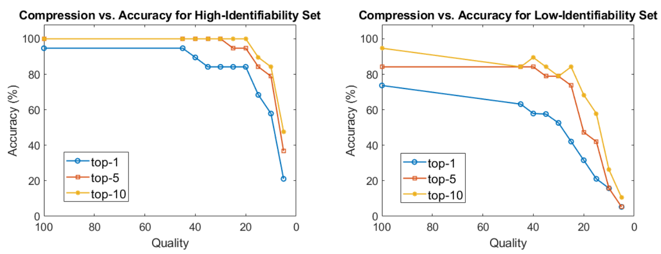
(a) High identifiability examples (b) Low identifiability examples

Fig. 1: Sample identities representing high and low identifiability in terms of personal attributes

We selected 2 small sets of identities from Market-1501 dataset, representing high and low identifiability in terms of attributes, respectively, to serve as queries (some sample identities are shown in Figure 1). Then we used the gallery provided by Market-1501 dataset for performing re-id. For the compression test, we apply JPEG compression with decreasing quality onto each of the two sets (see Figure 2 for two sample queries).

Fig. 2: original image (leftmost) vs. compressed images with quality = 40, 20, and 5

The results are presented in Figure 3. As expected, the high-identifiability collection shows much better robustness against compression than the low-identifiability collection. While both collections have very good baseline accuracies with no extra compression (denoted with quality=100), the accuracy of the low-identifiability collection degrades much faster as the compression increases. As can be seen from the results, the high-identifiability collection experiences almost no degradation performance-wise with quality larger than 20, and still maintains reasonable accuracies with lower quality. The results only experience drastic degradation for quality=5, which is very high compression and is rarely seen even in real-life systems. On the contrary, accuracies of the low-identifiability collection keep degrading and almost drop to zero after the quality decreases beyond 20.



(a) High-identifiability set (b) Low-identifiability set

Fig. 3: Accuracy vs. Compression

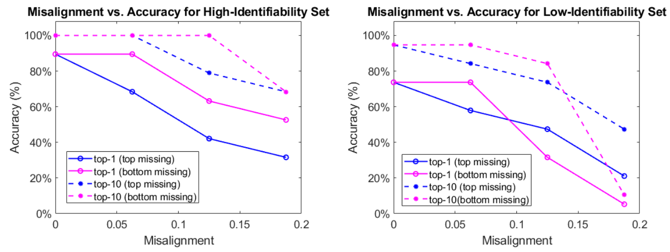
For the case of inaccurate bounding boxes, we simulated bounding box misalignment to both high-identifiability and low-identifiability collections, by only keeping a portion of the original bounding box. The misalignment was not caused by a real person extractor, but by manually removing parts of the original image. Samples can be seen in Figure 4, with 6.25%, 12.5% and 18.75% reduction on both width and height, towards upper-left or lower-right, respectively.

The results of re-id performed with these queries are shown in Figure 5. Similarly, although both sets degrade as the misalignment increases, the high-identifiability collection shows better robustness than the low-identifiability collection.


(a) High-identifiability examples (b) Low-identifiability examples

Fig. 4: original image (leftmost) vs. 6.25%, 12.5%, 18.75% misalignment (from left to right)

From the results above, we conclude that individuals with high identifiability in non-quality factors could resist more



(a) High-identifiability set (b) Low-identifiability set

Fig. 5: Accuracy vs. Misalignment

system-related degradation than low-identifiability ones. Similarly, as videos can be viewed as a collection of these individuals, a video with higher identifiability would show better robustness against system-related degradations, and we can compress high-identifiability videos without significantly compromising accuracy.

B. *Effectiveness of identifiability score*

In the second experiment, we evaluate the effectiveness of identifiability score computed following our procedures. Recall that our goal is to create a metric that quantifies the utility of a video for performing the re-id task. Therefore, we want to see a video's identifiability score correlated with the re-id performance using corresponding the video as the gallery (i.e., videos with high identifiability should generate relatively higher accuracies for re-id).

First, we selected 4 videos from the Cleveland dataset [6] to perform re-id tasks. We use the video taken in Brookpark tunnel 1 (denoted "tunnel 1") as our query video and extract 100 query images from it. Then we search for these queries in the other 3 gallery videos, which are taken in Brookpark tunnel 2 (denoted "tunnel 2"), Brookpark bus stop (denoted "bus stop"), and West Park tunnel (denoted "tunnel 3"), respectively. Sample frames from these videos can be seen in Figure 6. We then apply the procedures mentioned in Section III-B. As one can observe, tunnel 2 and bus stop have similar ground color as tunnel 1 (query), as tunnel 3 has drastically different background patterns; however, the lighting conditions in tunnel 2 and bus stop are not as consistent as tunnel 1. Therefore, tunnel 2 and bus stop is assigned a "0" for "background/lighting" factor, and tunnel 3 is assigned a "-1". We discussed the difference and features observed in these three videos and were able to determine the factor scores for them, shown in Table III. By combining the factors with preset weight allocation, we have 0.45 (tunnel 2), 0.3 (bus stop), and -0.4 (tunnel 3) as their identifiability scores.

Similarly, we applied the same evaluation procedure on the Market-1501 dataset, which is more popular and has better quality. Since we do not have access to the raw surveillance video, we could only estimate the identifiability score based on full-frame images of the dataset. For performing re-id, all queries are from Camera 5, and we search for these queries in galleries extracted from Camera 1, Camera 3, and Camera 6, respectively. Sample images can be seen in Figure 7. Following

TABLE III: Factor Scores for Cleveland videos

criterion	tunnel 2	bus stop	tunnel 3
background/ lighting	0	0	-1
occlusion	1	0	1
color	1	1	1
texture	-1	-1	-1
distortion	1	0	-1
similarity	1	1	-1

(a) Brookpark tunnel 1 (tunnel 1)

(b) Brookpark tunnel 2 (tunnel 2)

(c) Brookpark bus stop (bus stop)

(d) West Park tunnel (tunnel 3)

Fig. 6: Sample images from the 4 Cleveland videos

the same procedure, the identifiability scores are computed to be 0.85 for Camera 3, 0.45 for Camera 1, and 0.4 for Camera 6. Detailed scores can be found in Table IV.

TABLE IV: Factor Scores for Market-1501 dataset

criterion	cam 1	cam 3	cam 6
background/ lighting	0	0	-1
occlusion	0	1	1
color	1	1	1
texture	1	1	1
distortion	1	1	1
similarity	0	1	0

As can be seen in the re-id results (Table V), for both datasets, videos with higher identifiability generate better accuracy. The Pearson correlation coefficients between top-1, 5, and 10 accuracies and the identifiability scores are 0.944, 0.932, and 0.964 for Cleveland dataset, and 0.999, 0.994, and 0.994 for Market-1501 dataset. This result confirms our assumptions that videos with higher identifiability do perform better.

TABLE V: Re-identification Results

	Cleveland dataset			Market-1501 dataset		
Accuracy	tunnel 2	bus stop	tunnel 3	cam 1	cam 3	cam 6
Top-1	44%	51%	12%	23.8%	54.5%	20.3%
Top-5	68%	77%	35%	44.2%	73.2%	35.9%
Top-10	76%	80%	45%	52.9%	79.4%	45.3%
Score	0.45	0.3	-0.4	0.45	0.85	0.4

(a) Camera #5

(b) Camera #1

(c) Camera #3

(d) Camera #6

Fig. 7: Sample images from Market-1501 dataset

Also, note that Market-1501 is very large and contains identities extracted across several days, while each of the Cleveland videos we used only is a few hours in length; also, considering quality, Market-1501 dataset has significantly higher resolution than the Cleveland dataset. Therefore, comparisons of accuracy between the two datasets are not meaningful.

All in all, we still conclude that the identifiability score properly reflects re-id performance using corresponding video. Thus, it is promising to consider performing video selection prior to running re-id task to avoid wasting computing power on videos that are not as useful.

V. Conclusions

In this paper, we presented a new way to evaluate surveillance video utility, by specifying key attributes that affect task-specific performance. By applying this evaluation protocol, one can determine the utility of a video for performing a specific task, which was denoted as identifiability in our case of person re-id. Utilizing this identifiability score could help reduce computing power and storage, for example, by compressing videos with significant higher score and only using those with acceptable scores as the system input.

There are still much future work that can be done in improving the evaluation metric. Our current weight allocation for the evaluation criteria are based on our prior knowledge acquired by observing re-id algorithms; in the future, we plan to design a mathematical strategy to model the relationship between factors and performance, and then allocate weights accordingly. In addition, an automated system that analyzes identity features as well as general background information around the identity would be useful, and we are currently developing an implementation of such a system. Finally, this idea of task-specific evaluation metric should not be limited to person re-identification. It should be expand to more areas such as detection (where we evaluate "detectability"), classification (where we evaluate "classifiability"), tracking (where we evaluate "trackability"), etc., in the future.

References

[1] Q. Wan, K. Panetta, and S. Agaian, "A video forensic technique for detecting frame integrity using human visual system-inspired measure," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, April 2017.

[2] A. Barman and S. K. Shah, "Distance aggregation based score fusion for improving person re-identification," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, April 2017.

[3] Alvin Plantinga, "Things and persons," *Review of Metaphysics*, vol. 14, no. 3, pp. 493–519, 1961.

[4] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian, "Scalable person re-identification: A benchmark," in *IEEE International Conference on Computer Vision (ICCV)*, 2015.

[5] Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, and Qi Tian, "Person re-identification in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[6] P. Li, J. Brogan, and P. J. Flynn, "Toward facial re-identification: Experiments with data from an operational surveillance camera plant," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Sep. 2016, pp. 1–8.

[7] K. Tahboub, B. Delgado, and E. J. Delp, "Person re-identification using a patch-based appearance model," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 764–768.

[8] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[9] Zhedong Zheng, Liang Zheng, and Yi Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in *IEEE International Conference on Computer Vision*, 2017.

[10] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

[11] D. Chung, K. Tahboub, and E. J. Delp, "A two stream Siamese convolutional neural network for person re-identification," in *IEEE International Conference on Computer Vision*, Oct 2017, pp. 1992–2000.

[12] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang, "Person re-identification by video ranking," in *European Conference on Computer Vision*, 2014, pp. 688–703.

[13] Martin Hirzer, Csaba Beleznai, Peter M. Roth, and Horst Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian Conference on Image Analysis (SCIA)*, 2011.

[14] F. M. Khan and F. Brmond, "Multi-shot person re-identification using part appearance mixture," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2017, pp. 605–614.

[15] Alexander Hermans, Lucas Beyer, and Bastian Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.

[16] Wei Li, Xiatian Zhu, and Shaogang Gong, "Harmonious attention network for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[17] D. Chung and E. J. Delp, "Camera-aware image-to-image translation using similarity preserving StarGAN for person re-identification," in *CVPR Workshop on Target Re-identification and Multi-Target Multi-Camera Tracking*, 2019.

[18] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptor for person re-identification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 1363–1372.

[19] Doug Gray, Shane Brennan, and Hai Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, Rio de Janeiro*, 2007.

[20] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2015.

[21] Martin Köstinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof, "Large scale metric learning from equivalence constraints," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[22] D. J. Heeger and P. C. Teo, "A model of perceptual image fidelity," in *International Conference on Image Processing (ICIP)*, Oct 1995, vol. 2, pp. 343–345 vol.2.

[23] Xin Li, "Blind image quality assessment," in *International Conference on Image Processing (ICIP)*, Sep. 2002.

[24] Zhou Wang, Hamid R. Sheikh, and Alan C. Bovik, *Objective Video Quality Assessment*, pp. 1041–1078, CRC Press, 2004.

[25] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 531–543, April 2007.

[26] L. Best-Rowden and A. K. Jain, "Learning face image quality from human assessments," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 3064–3077, Dec 2018.

[27] W. Heng, T. Jiang, and W. Gao, "How to assess the quality of compressed surveillance videos using face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 8, pp. 2229–2243, Aug 2019.

[28] Samil Karahan, Merve Kilinc Yildirum, Kadir Kirtac, Ferhat Sukru Rende, Gultekin Butun, and Hazim Kemal Ekenel, "How image degradations affect deep CNN-based face recognition," in *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*. IEEE, 2016, pp. 1–5.

[29] B. RichardWebster, S. Anthony, and W. Scheirer, "PsyPhy: A psychophysics driven evaluation framework for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2280–2286, Sep. 2019.

[30] B. RichardWebster, So Yon Kwon, C. Clarizio, S. E. Anthony, and W. J. Scheirer, "Visual psychophysics for making face recognition algorithms more explainable," in *European Conference on Computer Vision*, 2018.

[31] K. Tahboub, A. R. Reibman, and E. J. Delp, "Accuracy prediction for pedestrian detection," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 4192–4196.

[32] K. Tahboub, D. Güera, A. R. Reibman, and E. J. Delp, "Quality-adaptive deep learning for pedestrian detection," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 4187–4191.

[33] L. Kong, R. Dai, and Y. Zhang, "A new quality model for object detection using compressed videos," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 3797–3801.

[34] Lingchao Kong, Ademola Ikusan, Rui Dai, Jingyi Zhu, and Dara Ros, "A no-reference image quality model for object detection on embedded cameras," *International Journal of Multimedia Data Engineering and Management*, vol. 10, pp. 22–39, 01 2019.

[35] L. Kong, A. Ikusan, R. Dai, and J. Zhu, "Blind image quality prediction for object detection," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, March 2019, pp. 216–221.

[36] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps, and R. J. Radke, "A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 3, pp. 523–536, March 2019.