Questionable Observer Detection Evaluation (QuODE) 2011

Jeremiah R. Barr, Kevin W. Bowyer and Patrick J. Flynn
Department of Computer Science and Engineering
University of Notre Dame
Email: {jbarr1,kwb,flynn}@nd.edu

Abstract

The intent of the competition posed here is to evaluate how well face clustering algorithms detect questionable observers, i.e. people that appear frequently in a video collection that captures crowds watching related events. Participants are called upon to test their algorithms on a challenging crowd video dataset containing wide variations in facial illumination, expression and resolution as well as moderate head pose variations. We ask researchers to measure clustering performance in terms of how strongly the groups within a clustering correspond to the identities of the recorded subjects and the extent to which a clustering facilitates the detection of questionable observers.

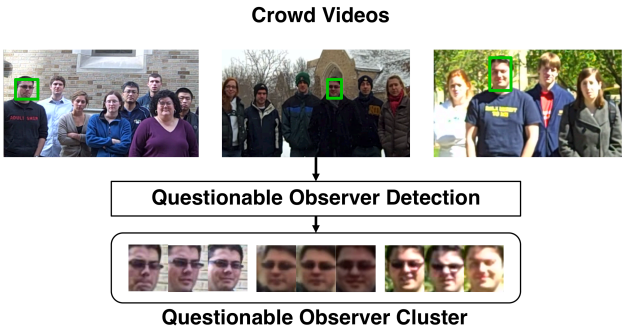

Figure 1. A questionable observer detection algorithm should accept a set of crowd videos as input and return face clusters with patterns from more than a specified number of videos.

1. Introduction

Recently, Barr et al.[3] formalized a challenging face clustering problem, *questionable observer detection*, which entails determining who appears unusually often across a collection of crowd videos. This problem arises in scenarios where we have videos of crowds observing the aftermath of some series of criminal activities, e.g. bombings caused by improvised explosive devices. We may not have prior knowledge about the people that contributed to these crimes, but we may gain insight into where we should start an investigation by determining if some crowd members appear in related crime scenes. These frequently appearing individuals are called *questionable observers*, as they appear at the crime scenes suspiciously often, whereas people that observe too few scenes to arouse interest are called *casual observers*. The objective is to detect all of the questionable observers while not mistaking any casual observer for a questionable one.

The questionable observer detection problem has a number of challenging properties. First, we do not assume that a comprehensive watch list is available, so we have no prior information about who we wish to detect or about how many different persons appear across the set of video clips. Second, most of the practical applications in which this problem arises do not allow for controlled video acquisitions. Changes in pose, illumination, scale, etc. can cause images of the same person that are taken in distinct scenes to appear significantly different and, hence, increase the intra-personal appearance variability. The detection task is further complicated by the fact that crowd members that lie nearer to the video camera can occlude people that lie behind them. For the competition posed here, the *Questionable Observer Detection Evaluation (QuODE) 2011*, we ask participants to mitigate these complications using advanced face recognition technology within a clustering scheme.

2. Problem Statement

More specifically, given a collection of m videos, $\{V_1, V_2, ..., V_m\}$, each of which shows a crowd of people observing a distinct event, we intend to identify those individuals that appear in more than some specified number of scenes, v. Each video V_i contains m_i face image sequences, $S_i = \{s_{i,1}, s_{i,2}, \ldots, s_{i,m_i}\}$, where every sequence $s_{i,j}$ consists of a unique ordered set of face images that represent the same person. We do not assume that we know where and when faces appear across the video set, i.e. the face im-

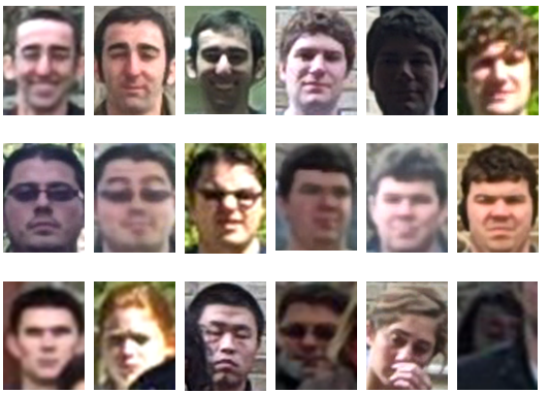

Figure 2. Complicating factors in the ND-QO-Flip dataset. Top row: images of two questionable observers taken under varying illumination conditions. Middle row: images of another two questionable observers making distinct facial expressions in different videos. Bottom row: instances where subjects were occluded by other crowd members or their own body parts.

ages sequences must be extracted automatically using a face tracker or a face detector. The individual elements in a face image sequence need not come from a contiguous sequence of video frames, but no two elements can come from the same video frame. In addition, the face image extraction algorithm may extract multiple sequences from a single video that contain images of the same person if she intermittently leaves the view of the camera.

We frame the questionable observer detection problem as one of unsupervised learning in which we assign a label, $l(s_{i,j})$, to each face image sequence. The set of face image sequences that share a common identity should be assigned the same label. Additionally, the face image sequences with the same label form a *cluster* C_L:

$$C_L = \{s_{i,j} \in \cup_{k=1}^{m} S_k : l(s_{i,j}) = L\}. \quad (1)$$

The *questionable observer detection problem* requires that we mark any cluster C_L as questionable if the number of videos from which its constituent sequences were extracted surpasses the video count threshold v. That is, an individual whose face tracks make up the majority of a cluster for which

$$|\{i \in 1, 2, \ldots, m : \exists j \; such \; that \; s_{i,j} \in C_L\}| > v \quad (2)$$

is considered to be a questionable observer.

3. Evaluation Data Set

Participants must test their algorithms on the ND-QO-Flip Crowd Video Dataset [3], which consists of 14 crowd video clips recorded around the University of Notre Dame Campus over a seven-month period with a Flip camcorder. 12 of these videos were acquired outdoors in overcast, sunny or snowy conditions, whereas the other two were acquired indoors with either artificial or natural illumination. In every scene, the camera panned over the crowd multiple times at various zoom levels. The video clips each

- have a resolution of 640x480 and a frame rate of 30 frames per second;
- contain 25-40 seconds of video per clip;
- were compressed using H.264 compression; and
- capture crowds of four to 12 people.

The subject pool consists of 90 people, five of whom appear in multiple videos and should be detected as questionable observers. The face images tend to have a low spatial resolution in the sense that the mean distance between the eyes is 20 pixels while the most frequent interocular distance is 12 pixels. Further, the Flip camcorder often lost focus of the crowd. Although pose variations were limited insofar as the subjects tended to face toward the camera, the crowd members changed their facial expressions freely and the illumination conditions were not constrained. In other words, this dataset captures a number of challenging nuisance factors that impact facial appearance, including facial expression, illumination, occlusion and spatial resolution.

The ND-QO-Flip Crowd Video Dataset includes metadata describing when and where the subjects initially appear in the videos. The metadata indicates the video frame index and image position of every subject's initial appearance across the video set, using the University of Notre Dame subject number as the subject identifier. Images of the faces described by the metadata and a table of the University of Notre Dame subject numbers for the questionable observers are included as well. The various components of the data set are described in more detail within its accompanying documentation.

4. Performance Metrics

Questionable observer detection performance can be evaluated in terms of how well the face patterns are organized into clusters and to what extent the questionable observers are distinguished from the casual observers. We treat each cluster as though it represents the individual whose face patterns comprise the majority of its constituent patterns. In the ideal clustering, all face patterns associated with a particular individual would be assigned to the same cluster and all individuals would have a distinct cluster. The classification accuracy would be perfect in this case, as would the questionable observer detection rate. The clusterings produced in practice differ from this ideal when some

of the face patterns of one subject are assigned to a cluster that better represents another person. The *self-organization rate* (SOR), which was introduced by Raytchev and Murase [4], accounts for the frequency of cluster assignment errors in a way that is similar to the classification accuracy performance metric, yet the SOR discounts clusters that do not contain a clear majority:

$$SOR = (1 - \frac{\Sigma n_{ab} + n_e}{n}), \tag{3}$$

where n_{ab} denotes the number of patterns representing individual a that were assigned to a cluster dominated by the patterns of individual b, n_e represents the number of patterns that are assigned to a cluster in which no single individual corresponds to more than half of the patterns, and n denotes the number of patterns in the clustering. The SOR varies within $[0, 1]$ and has a positive polarity.

With respect to questionable observer detection performance, false positives and false negatives are the primary error types. A *false positive* occurs when any cluster that represents a casual observer has face patterns from more than v videos, whereas a *false negative* occurs when none of the clusters that correspond to a questionable observer contain patterns from more than v videos. Conversely, a *true positive* arises when at least one of the clusters that represent a questionable observer includes face patterns from more than v videos, and a *true negative* takes place when none of the clusters that correspond to a casual observer contain patterns from more than v videos. For this competition, v is set to 1. In other words, the objective posed here is to distinguish the questionable observers that appear in more than one video from the casual observers that appear in exactly one video without the benefit of an existing database of faces with known labels.

Let tp, fn, tn, and fp be the number of true positives, false negatives, true negatives and false positives that are yielded by a particular clustering. We measure detection performance using the *false positive rate (FPR)* and *false negative rate (FNR)*:

$$FNR = \frac{fn}{tp + fn}, \text{and} \tag{4}$$

$$FPR = \frac{fp}{tn + fp}. \tag{5}$$

The FPR and FNR vary within $[0, 1]$ and have negative polarity.

5. Baseline Algorithms

Thus far, two distinct algorithms have been evaluated on the ND-QO-Flip Crowd Video Dataset, as shown in [3]. The primary algorithm proposed in that work, *QuOD v1*, tracks the faces within the videos, normalizes the associated face images, and then merges the face sequences that correspond to the same person on a video-by-video basis. An outlier detection algorithm determines a representative collection of face images within each face sequence to provide robustness to intermittent occlusions and alignment issues. Hierarchical agglomerative clustering is subsequently performed on the face image sequences from all of the videos to cluster the data by identity. Another algorithm described in [3], *QuOD v1 with VeriLook 4.0*, performs a similar series of operations, but uses the matcher provided with the VeriLook 4.0 Standard SDK from Neurotechnology [1] as the means to compute the similarity scores required by the clustering algorithm. Participants should treat the results that these methods yielded, as shown in Table 1, as performance baselines.

Table 1. Baseline results for a video count threshold of one video.

Method	SOR	FPR	FNR
HAC/VeriLook	0.895	0.061	0.40
Proposed algorithm	0.960	0.056	0.00

6. Participation Guidelines

The objective of QuODE 2011 is to establish the performance of state-of-the-art face clustering algorithms on the ND-QO-Flip Crowd Video Dataset in terms of the SOR, FNR and FPR. Participants should email Jeremiah Barr at jbarr1@nd.edu to register and obtain download instructions for the ND-QO-Flip Crowd Video Dataset, using QuODE 2011 as the subject line.

At the end of the evaluation, participants should submit the following three items:

1. A text file or Word document, called "summary", that includes the names of the individuals involved with their project, their institutional affiliation(s), the name of the algorithm, a brief description about how it operates and a summary of the results it obtained with respect to the SOR, FNR and FPR metrics.

2. A directory named "data", which contains the detected or tracked face images used during clustering.

3. A CSV result file titled "clustering" that describes the cluster assignments made by the algorithm, with each line containing the name of a face image, the unique identifier of the cluster to which it was assigned, and the University of Notre Dame subject number of the person contained in the face image as indicated by the ground truth files included with the ND-QO-Flip Crowd Video Dataset.

These items should be aggregated into a zip or tar.gz file and uploaded to the location provided during the registration process. The results will be presented on the BeFIT website [2]. Participants are encouraged to present their algorithms and results in other publications as they see fit; please remember to cite [3].

7. Acknowledgements

The figures in this paper originally appeared in [3]. Additionally, the ND-QO-Flip Crowd Video Dataset was acquired with the support of the Central Intelligence Agency, the Biometrics Task Force and the Technical Support Working Group through US Army contract W91CRB-08-C-0093. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of our sponsors.

References

[1] VeriLook Standard SDK and Extended SDK. http://www.neurotechnology.com/vl_sdk.html, 2010.

[2] BeFIT 2011 First International Workshop on Benchmarking Facial Image Analysis Technologies: Call for Challenges. http://fipa.cs.kit.edu/befit/workshop2011/call_for_challenges.php, 2011.

[3] J. Barr, K. Bowyer, and P. Flynn. Detecting questionable observers using face track clustering. *Proc. 2011 Workshop on Applications of Computer Vision*, 2011.

[4] B. Raytchev and H. Murase. Unsupervised face recognition from image sequences based on clustering with attraction and repulsion. *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:II–25–30, 2001.