12-2016

# Video annotation by crowd workers with privacy-preserving local disclosure

Apeksha Dipak Kumavat
*Purdue University*

*Graduate School Form*
*30 Updated 12/26/2015*

**PURDUE UNIVERSITY**
**GRADUATE SCHOOL**
**Thesis/Dissertation Acceptance**

This is to certify that the thesis/dissertation prepared

By Apeksha Dipak Kumavat

Entitled
VIDEO ANNOTATION BY CROWD WORKERS WITH PRIVACY-PRESERVING LOCAL DISCLOSURE

For the degree of Master of Science in Electrical and Computer Engineering

Is approved by the final examining committee:

Alexander J. Quinn
Chair

Amy R. Reibman

Edward J. Delp

Vijay Raghunathan

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Alexander J. Quinn

Approved by: Venkataramanan Balakrishnan 12/1/2016
Head of the Departmental Graduate Program Date

# VIDEO ANNOTATION BY CROWD WORKERS WITH PRIVACY-PRESERVING LOCAL DISCLOSURE

A Thesis

Submitted to the Faculty

of

Purdue University

by

Apeksha Dipak Kumavat

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Electrical and Computer Engineering

December 2016

Purdue University

West Lafayette, Indiana

Dedicated to Anjali Kumavat

A.K.A.

*Sheela Vakde*

# ACKNOWLEDGMENTS

First of all, I would like to express my sincere thanks to my major advisor, Professor Alexander J. Quinn, for his invaluable guidance and support. I am extremely grateful to you, Prof. Quinn for giving me this opportunity. I believe I have learned a lot from you, academic and otherwise.

I would like to thank my graduate advisory committee members, Professor Amy R. Reibman, Professor Edward J. Delp and Professor Vijay Raghunathan for their advice, encouragement and supporting my work.

While working on this project, I had an opportunity to interact and learn from the incredibly nice and brilliant colleagues in the Human Powered Systems Lab. I appreciate the support and friendship of: Chaithanya Manam, Gaoping Huang, Jordan Huffaker, Meng-Han Wu and Vipul Bhat.

I would take this oppotunity to thank my parents and my brother, for supporting my career decisions and always believing in me. Thank you for giving me this opportunity to acquire and share knowledge with others.

Finally, to Arjun for all the encouragement, motivation and support. Thank you for all the strength that you have given me throughout this experience.

## TABLE OF CONTENTS

Page

LIST OF TABLES . . . . . . . . . . . . . . . . . . . . . . . . . . . . vi

LIST OF FIGURES . . . . . . . . . . . . . . . . . . . . . . . . . . . vii

ABSTRACT . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ix

1 INTRODUCTION . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1.1 Vision . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

1.2 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

1.3 Contribution . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2 BACKGROUND . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

2.1 Face Recognition in Wild using Machines . . . . . . . . . . . . . . 7

2.2 Content Recognition by Humans . . . . . . . . . . . . . . . . . . . 9

2.3 Obfuscation using Image Processing . . . . . . . . . . . . . . . . . 11

2.4 Crowdsourcing Annotation for Privacy-sensitive Videos . . . . . . . 13

3 SYSTEM DESCRIPTION . . . . . . . . . . . . . . . . . . . . . . . . . 15

3.1 Obfuscation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

3.2 Controlled Local Disclosure . . . . . . . . . . . . . . . . . . . . . 17

3.3 Privacy Protection During Crowd Interaction . . . . . . . . . . . . 20

4 EVALUATION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

4.1 Dataset . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24

4.2 Experimental Setup . . . . . . . . . . . . . . . . . . . . . . . . . 26

4.2.1 System Implementation . . . . . . . . . . . . . . . . . . . . . 26

4.2.2 Crowd Interaction . . . . . . . . . . . . . . . . . . . . . . . . 28

4.2.3 Instructions to Crowd-workers . . . . . . . . . . . . . . . . . . 30

4.2.4 Time, Cost and Incentives . . . . . . . . . . . . . . . . . . . . 30

4.3 Results and Analysis . . . . . . . . . . . . . . . . . . . . . . . . 31

Page

5 SUMMARY . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
5.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
5.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
5.2.1 Obfuscation . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
5.2.2 Forming Hypotheses . . . . . . . . . . . . . . . . . . . . . . . 40
5.2.3 Local Disclosure . . . . . . . . . . . . . . . . . . . . . . . . . 41
5.2.4 Dataset . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
5.3 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
REFERENCES . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
A USER INTERFACE . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
B DATASET . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
C DATA . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

# LIST OF TABLES

Table Page

2.1 Data from [11]. True positive rate for various face detection algorithms against increasing number of allowed false positives . . . . . . . . . . . 7

4.1 Total Correctly Identified Faces and Actions . . . . . . . . . . . . . . . . 32

4.2 Number of clicks by crowd-workers for each video presented . . . . . . 32

4.3 Fitting Coefficients from Logistic Regression For Correct Actions . . . 37

4.4 Fitting Coefficients from Logistic Regression for Correct Face Tags . . 37

4.5 ANOVA Chi-squared test results for *method* as predictor and *correct face tags* as outcome . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38

C.1 Data Collected . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

# LIST OF FIGURES

Figure Page

1.1 Designed System . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.2 Handcrafted illustration of controlled local disclosure (refer Figure 3.7 for actual results). The disclosure is less for face (in red) and more for non-face objects (in green). Original picture from Wikimedia Commons (in public domain) [10]. . . . . . . . . . . . . . . . . . . . . . . . . . 6

2.1 Examples of Human Images Where Skin-color Based Detection Fails. Original images: top-right from wikimedia (public domain) [17], top-left [18], bottom-left [19] and bottom-right [20] from flickr licensed under CC BY 2.0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.2 Significantly degraded celebrity face images that can still be detected as well as recognized by humans. Image reprinted from [21]. . . . . . . . . 10

2.3 Importance of configural features, symmetry and enclosure in human perception. Adding the data from (b) to (a) decreases the time required to find ')' in the image, instead of increasing it due to increased data. Image reprinted from [23]. . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.4 Examples of the filters - gaussian blur, median blur and pixelation for six different radius levels. Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24] . . . . . . . . . 11

2.5 Examples of Human Images as shown in Figure 2.1 Obfuscated Using Median Filter. Original images: top-right from wikimedia (public domain) [17]; top-left [18], bottom-left [19] and bottom-right [20] from flickr licensed under CC BY 2.0. Modification: Blurring using median filter. . . . . . 12

3.1 Process of Median filtering. (a) Original image (b) Median Blur (radius 3) applied to original image in OpenCV . . . . . . . . . . . . . . . . . 15

3.2 Above: Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24]; Below: Median Blur (radius 23) applied to original frame using OpenCV. . . . . . . . . . . . . . . . 16

3.3 Two independent iterations for calculating safe-to-reveal radius for the video frame shown in figure 3.2, each with 100 random simulated clicks. 19

Figure Page

3.4 Controlled Local Disclosure for video frame shown in figure 3.2. (a) Keyhole mode output with marking (b) Keyhole mode output (c) Foveated mode output with marking (d) Foveated mode output. . . . . . . . . . 20

3.5 (a) Output of Fovea system in keyhole mode for 100 random clicks without protection against crowd-collusion (b) Output of Fovea system in keyhole mode for 100 random clicks (same clicks as in (a)) with protection against crowd-collusion. Green circles are not a part of Fovea system's output. Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24] . . . . . . . . . . . . . . . . . . . . . . . 21

3.6 (a) Output of Fovea system in keyhole mode for 1000 random clicks without protection against crowd-collusion (b) Output of Fovea system in keyhole mode for 1000 random clicks (same clicks as in (a)) with protection against crowd-collusion. Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24] . . . . . . . . . 22

4.1 Examples of video clips from the Hollywood2 dataset [24] . . . . . . . . 25

4.2 User Interface for tagging and analyzing videos through Crowd-Sourcing on AMT . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

4.3 (a) Only filter mode; Disclosure of a region surrounding clicks by (b) keyhole mode and (c) foveated mode; Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24] . . . . 29

4.4 Instructions provided to workers for the task . . . . . . . . . . . . . . . 30

4.5 Odds ratios obtained using logistic regression model of the effects of methods—Only filter, Keyhole mode and Foveated mode—on correct actions identified and correct faces identified. Model has been adjusted for the effects of different videos and workers. $**p < 0.01$, $***p < 0.001$. . . . . . . . 36

A.1 Complete user-interface used for the experiment . . . . . . . . . . . . . 45

# ABSTRACT

Kumavat, Apeksha Dipak M.S.E.C.E., Purdue University, December 2016. Video Annotation By Crowd Workers With Privacy-preserving Local Disclosure . Major Professor: Alexander J. Quinn.

Advancements in computer vision are still not reliable enough for detecting video content including humans and their actions. Microtask crowdsourcing on task markets such as Amazon Mechnical Turk and Upwork can bring humans into the loop. However, engaging crowd workers to annotate non-public video footage risks revealing the identities of people in the video who may have a right to anonymity.

This thesis demonstrates how we can engage untrusted crowd workers to detect behaviors and objects, while robustly concealing the identities of all faces. We developed a web-based system that presents obfuscated videos to crowd workers, and provides them with a mechanism to test their hypotheses about what behaviors and/or objects might be present in the videos.

Our system, called *Fovea*, works by initially applying a heavy median blur to the videos. This guarantees privacy but impedes recognition of other content of interest. An algorithm was developed as a part of this thesis to calculate the radius of a safe-to-reveal region around a pixel. It was implemented into an interactive system that allows workers watching the blurred videos to selectively reveal small regions by clicking.

We compared two approaches for local disclosure of information—foveated mode and keyhole mode—together with a non-interactive blur-only mode as a control. The results showed that both modes led to superior recognition of actions while keeping the odds of correct face recognition close to that of the control.

# 1. INTRODUCTION

## 1.1 Vision

This thesis is motivated by our vision of a privacy-preserving crowdsourced video annotation system, called *Fovea*. In this section, we present envisioned interactions with our system.

Officer Dan, after his usual patrolling of the neighborhood while wearing the body-worn camera, returns to the department at the end of the day and handovers his camera to Officer Sarah, who is in-charge of uploading the departments body-worn camera footage on public platforms. After getting the videos from all the cameras handed over to her at the end of the day, each of which contain over 12 hours of video, she first feeds these videos to Fovea.

At the same-time, in a different part of the world, John turns on the surveillance camera system installed around his house and feeds the output stream of videos to Fovea. And so does Sheela, who needs to perform video coding of a large number of videos, for her juvenile interrogation study. She loads all these videos into Fovea.

Fovea takes the input videos provided to it and without altering the original video frames, blurs them [at the server] while presenting them to [the client browser as seen by] the crowd-workers [in form of small video clips or video streams]. Fovea ensures through controlled blurring that the crowd-workers cannot identify any of the faces present in the video, however, are able to make an initial hypothesis about the contents of the video and the location of faces, if any.

After forming the initial hypothesis about a particular region in the video, they then click on that region to reveal a restricted subset of it, however, such that a face is never shown to them. They use the information available to them to provide location of faces, describe contents of the video and to flag a time instance in the video to be containing suspicious behavior.

The location of faces obtained from the crowd, are then used by Fovea to redact all the faces from the original videos. Officer Sarah can now upload these redacted output videos to any public platform without violating privacy issues. Sheela gets the content description for all the videos, obtained from the crowd in a short amount of time, without disclosing the identities of the juvenile subjects of her study. And, John gets an alert from Fovea about an intruder nearing his property in suspicious manner, without revealing the identities of his family on an untrusted crowdsourcing platform.

## 1.2 Motivation

Searching large collections of video for events or behavior is labor intensive. [Traditionally, these tasks have been carried out sequentially by a single human.] This draws us towards utilizing the lucrative solution of splitting and delegating such tasks on the fast, scalable and flexible crowdsourcing platforms. However, most applications of video analysis expect privacy protection of subjects present in these videos, and yet crowdsourcing is normally presumed to be limited to tasks that do not involve private data. This is because crowdsourcing, by definition, involves an "open call" so there are little controls over who participates. Bypassing normal employment relationships allows crowdsourcing to deliver rapid response from many workers on very short notice. However, without an ongoing relationship and the vetting process that normally precedes it, the incentives to keep an employer's data confidential are weaker.

The inability to safely engage crowd workers to perform tasks with sensitive information prevents people from freely delegating information work to an always-available workforce. If not for the issues of privacy and information security, any of the tasks described for the *Fovea* system (section 1.1) could be delegated directly to crowd workers without any need for pre-processing. Besides task markets, such as Mechanical Turk [1] and Upwork [2], this inhibits the potential of any opportunity to engage untrusted human help, such as employees from other divisions, or would-be volunteers from the public.

Ideally, machine automation would make it possible to input video files and find every occurrence of some target event, such as a fight, theft, or a person holding a knife. However, despite significant progress in image understanding algorithms, we are still a long way from general, unconstrained image understanding that rivals the accuracy of human perception. Automated behavior classification is far more challenging yet [3–6].

For searching or coding non-public unconstrained video data, such as surveillance footage or police body camera video (which may sometimes be taken inside private homes), automated detection is not always reliable enough, and yet current models of crowdsourcing would present unacceptable risks to privacy. The increased use of police body cameras in the US—paired with demands for public disclosure and transparency—have brought particular urgency to this issue and led municipalities, such as Seattle to search for technical solutions to the problem [7].

### 1.3 Contribution

This thesis presents a system that allows delegating video analysis tasks to untrusted crowd-workers, while preserving privacy, on task markets. As shown in figure 1.1, videos from surveillance cameras or police body-worn cameras are quantized (sub-system 1), so as to account for the work done by the crowd-workers. The videos are then provided as input to sub-system 2. This system is responsible for presenting obfuscated videos to the workers, such that, the workers are able to “detect” the contents of the videos and location of faces, however, are not able to “identify” any face. This system outputs all the judgements provided by the workers. These judgements

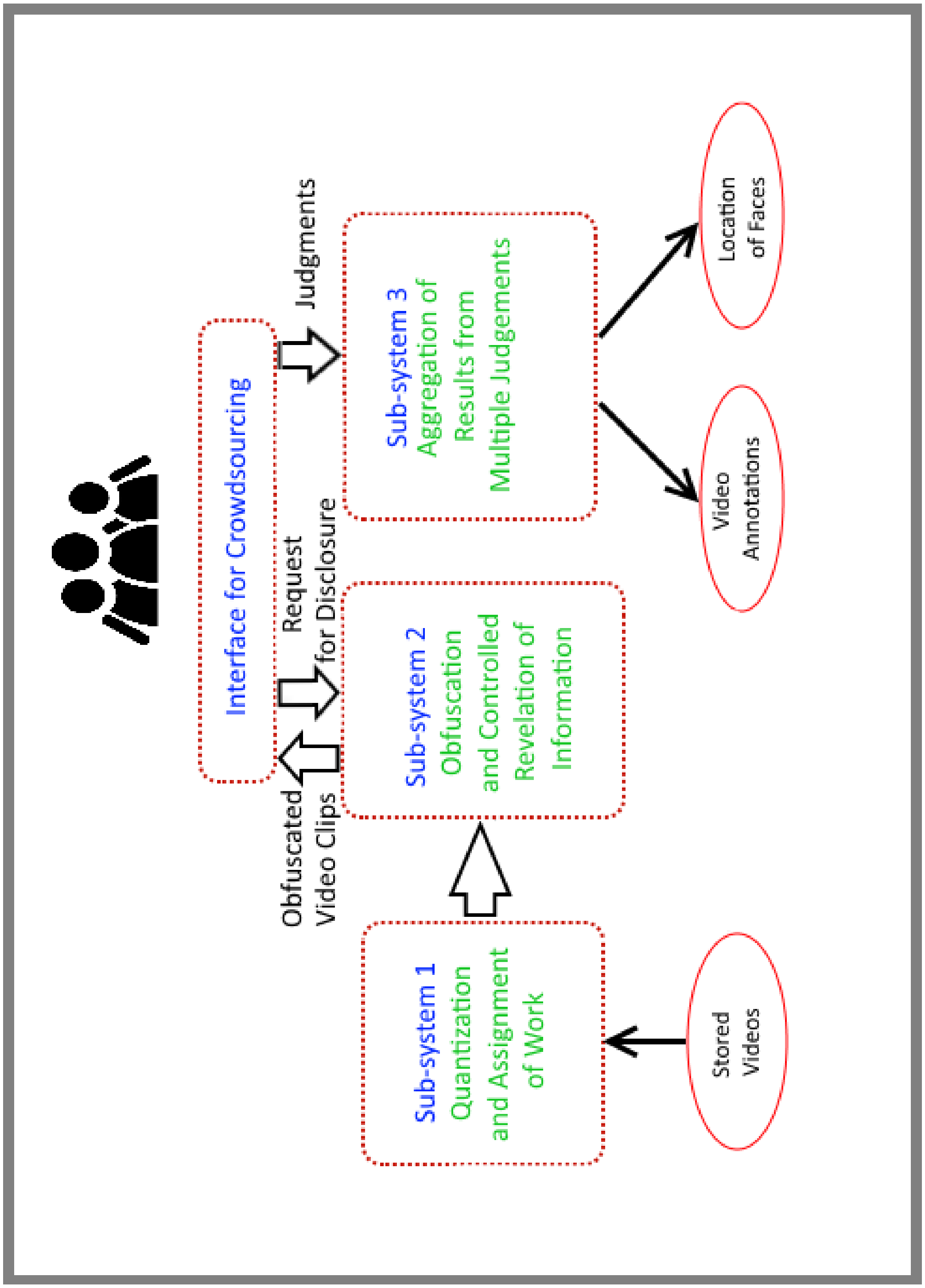

Interface for Crowdsourcing
Obfuscated Video Clips
Request for Disclosure
Judgments
Sub-system 1
Quantization and Assignment of Work
Sub-system 2
Obfuscation and Controlled Revelation of Information
Sub-system 3
Aggregation of Results from Multiple Judgements
Stored Videos
Video Annotations
Location of Faces

Fig. 1.1. Designed System

are then used for extracting information (sub-system 3) about the exact location of faces and video content.

The system, as seen from figure 1.1, receives input as stored videos from body-cameras or surveillance camera footages.

To preserve the privacy of the videos, they are initially obfuscated before showing them to the crowd. The level of obfuscation is set such that it is atleast above the necessary amount to de-identify all the faces in the given video, as well as, just sufficient enough for de-identification of only faces, retaining as much of the other information as possible.

The tradeoff between privacy of identities vs. accuracy of activity recognition have been studied previously for obfuscation techniques like pixelation and blurring [8, 9]. [9] implemented a system that allowed researchers to obtain a trade-off curve in terms of precision and recall over filter level for different obfuscation techniques (such as level of blurring, or mask padding), by providing a set of example videos containing the event to be annotated in a test dataset.

However, these techniques use fixed level of global obfuscation, which results in compromising either the privacy or the accuracy. In our work, at the initial stage, we use an obfuscation level that is biased towards privacy more than accuracy. The prior research and our approach for using obfuscation to preserve privacy is covered in detail in sections 2.3 and 3.1.1.

As a fixed level of global obfuscation, requires compromising either privacy or accuracy, we provide the crowd-workers with a tool to reveal a region on a video frame, which they think would help them to form an accurate judgement about the content of the video. This tool is not intended to form a judgement on its on, however, to confirm a weak hypothesis that is already formed by watching the obfuscated video.

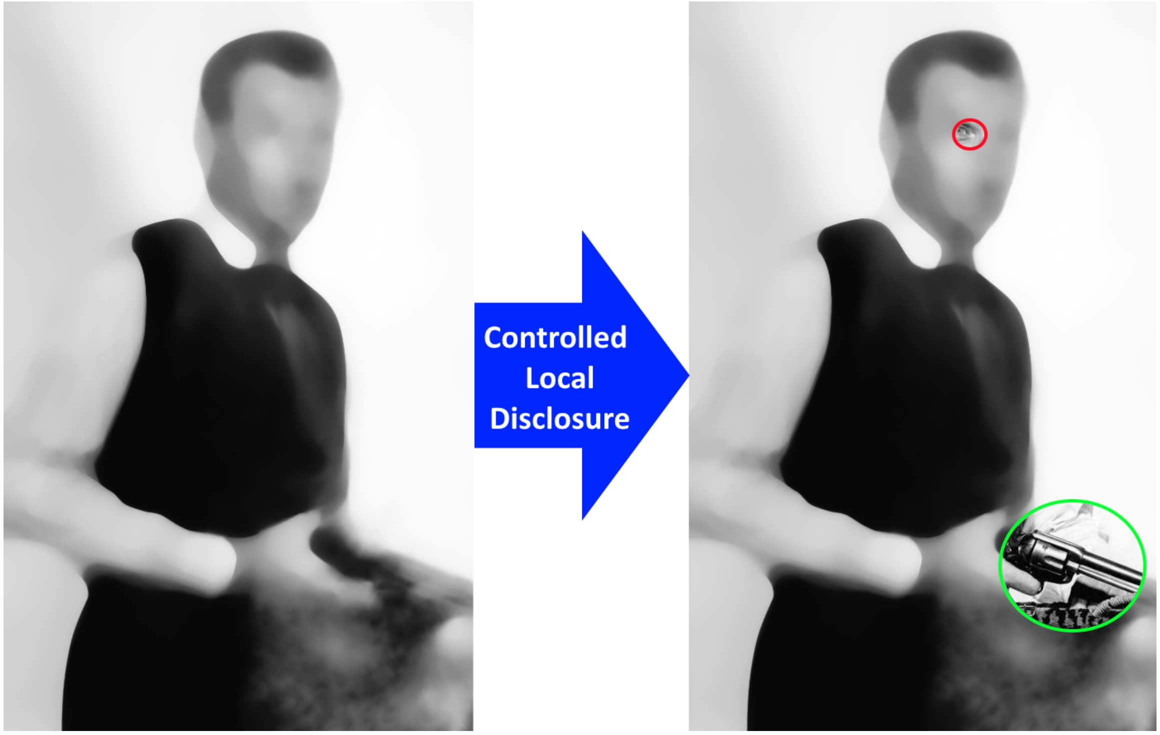


Fig. 1.2. Handcrafted illustration of controlled local disclosure (refer Figure 3.7 for actual results). The disclosure is less for face (in red) and more for non-face objects (in green). Original picture from Wikimedia Commons (in public domain) [10].

This tool is designed to allow the crowd worker to click anywhere on a video frame, and removes the obfuscation in a circular region around this click, such that no face is ever de-identified in this circular region. Our work focused majorly on the development of the algorithm for realizing this tool. The system derives its name—*Fovea*—due to the integration of this tool which tries to create a foveated image for allowing the crowd-workers to focus on regions of interest without ever being exposed to private information.

Finally, in our evaluations, we compare judgements from crowd-workers to confirm the increase in accuracy of activity recognition using the controlled information disclosure tool over using only blurring for privacy protection.

# 2. BACKGROUND

## 2.1 Face Recognition in Wild using Machines

Table 2.1.
Data from [11]. True positive rate for various face detection algorithms against increasing number of allowed false positives

| Method | No. of fasle positives ≈ 10 | No. of fasle positives ≈ 100 | No. of fasle positives ≈ 1000 |
|---|---|---|---|
| Mikolajczyk et al. | 10 | 33 | 54.9 |
| Viola and Jones (OpenCV Version) | 10 | 33 | 59.7 |
| Jain and Learned -Miller | 15.7 | 51 | 67.7 |
| Zhu et al. | 63.8 | 73.3 | 76.6 |
| Shen et al. | 8 | 67.5 | 78.6 |
| Li and Zhang | 69.4 | 80.6 | 83.7 |
| Li et al. | 10 | 73.3 | 80.9 |
| Li et al. | 69.2 | 80.8 | 84.8 |
| Yan et al. | 75.9 | 81.3 | 85.2 |
| Chen et al. | 78.8 | 83.9 | 86.2 |
| Mathias et al. | 72.5 | 83.4 | 87 |
| Yang et al. | 75.4 | 81.6 | 85.2 |
| Yan et al. | - | ∼80 | ∼84.6 |
| Jun et al. | ∼67 | ∼77 | ∼80.6 |

Looking at the extensive development in the field of Pattern Recognition and Computer Vision, a question that arises is “Why should we use the pattern recognition abilities of humans and not machines?” An obvious solution to this problem of selective redaction is to use pattern recognition algorithms that allow detection of sensitive information such as faces, vehicle numbers, etc. and obscure it. Recent works on visual privacy protection such as SmartRedaction by Utility Inc., are indeed directed towards image redaction using computer vision techniques to determine region-of-interest (ROI) in the picture, for instance, tracking moving people, detecting faces in live camera videos as well as skin detection. This ROI may then be removed from the image using the preferred obfuscation technique.

A recent “survey on face detection in the wild: past, present and future” [11], describes, and compares the performance of the state-of-the-art in face detection. The survey report is summarized in table 2.1. One of the conclusions of this survey says “even when allowing a relative large number of false positives (around 1000), there are still around 15 - 20% of faces that are not detected.” This is a huge false negative rate when the application demands guaranteed face detection.

Another highly popular technique for detecting a face involves skin color detection [12–14]. Skin color based face detectors have gained high popularity as they are highly robust to geometric variations such as scale, rotation as well as pose. The survey of skin-color modeling and detection methods [15], shows that the Bayesian network described in [16] shows best performance with 99.4% true positives (and 10% false negatives). However, these techniques are based on skin-color modelling, that looks for “skin” in the modelled color spaces. This color-specific detection of skin fails when the skin color itself is absent in the test image. Figure 2.1 shows examples of human images where the skin-color based detection fails. However, as can be seen, these faces are still recognizable enough for human eye.

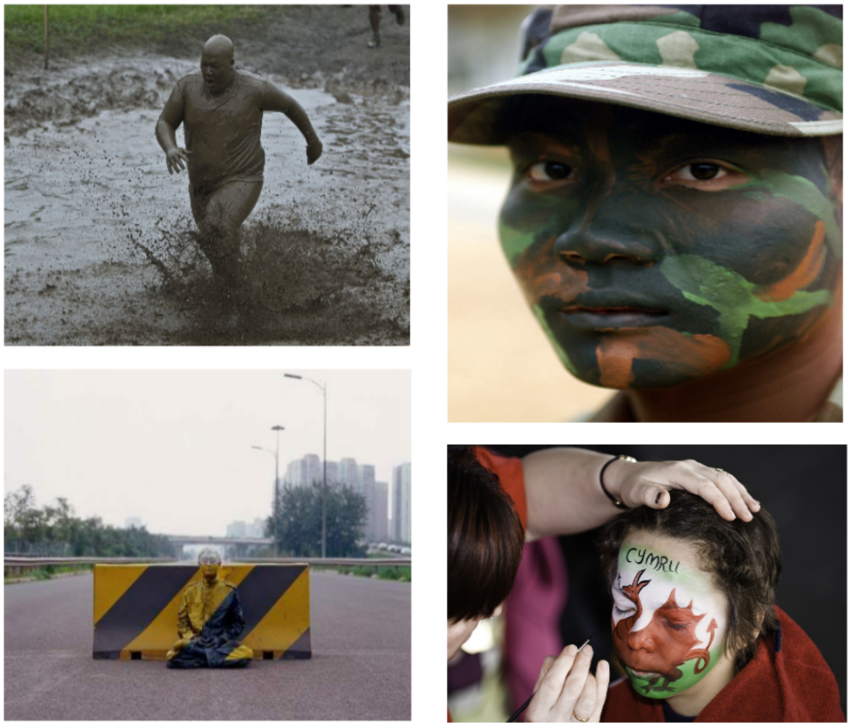


Fig. 2.1. Examples of Human Images Where Skin-color Based Detection Fails. Original images: top-right from wikimedia (public domain) [17], top-left [18], bottom-left [19] and bottom-right [20] from flickr licensed under CC BY 2.0

## 2.2 Content Recognition by Humans

The task of detecting the presence of people and their actions in videos or still images is currently more efficiently accomplished by humans than machines. The study [21] shows that humans are not only able to detect, but also recognize famous celebrities in low-resolution images. Also, humans are able to recognize activities happening even in a blurred video [22].

Humans use not just individual features but configurational information to build their understanding of an image [21]. They are able to locate faces using information such as other body parts, orientation of the body, gait of the person (in the case

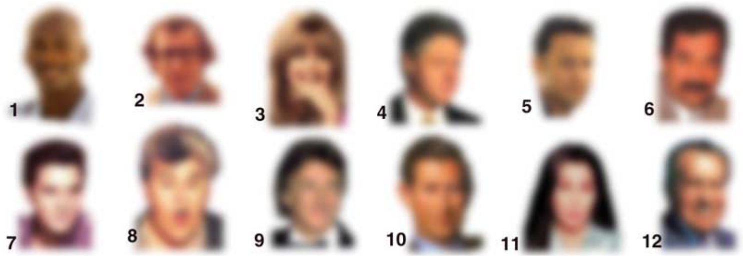


Fig. 2.2. Significantly degraded celebrity face images that can still be detected as well as recognized by humans. Image reprinted from [21].

of videos), etc., to make a hypothesis about the presence of a face. The field of Psychology has extensive work on the way human visual system works. Humans perceive objects, scenes and faces following the Gestalt laws. Utilization of this fact for image segmentation and object detection has been shown to perform better than using just the components of any image [23]. A popular example that shows the bias of human vision towards configural superiority is shown in figure 2.3.

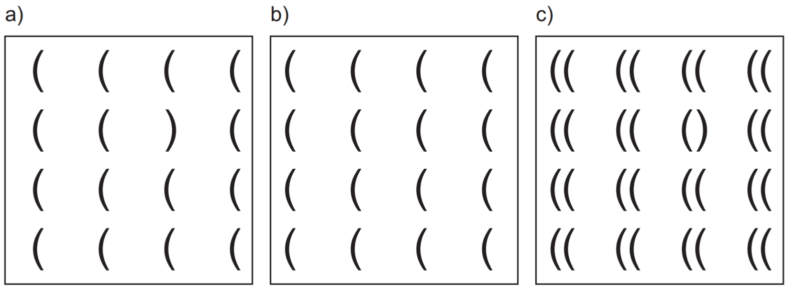


Fig. 2.3. Importance of configural features, symmetry and enclosure in human perception. Adding the data from (b) to (a) decreases the time required to find ')' in the image, instead of increasing it due to increased data. Image reprinted from [23].

The study [22] shows that humans are able to recognize activities in a video, even if some information is distorted. These capabilities of humans can be used for annotating blurred videos, while preserving privacy of the subjects in the video.

## 2.3 Obfuscation using Image Processing

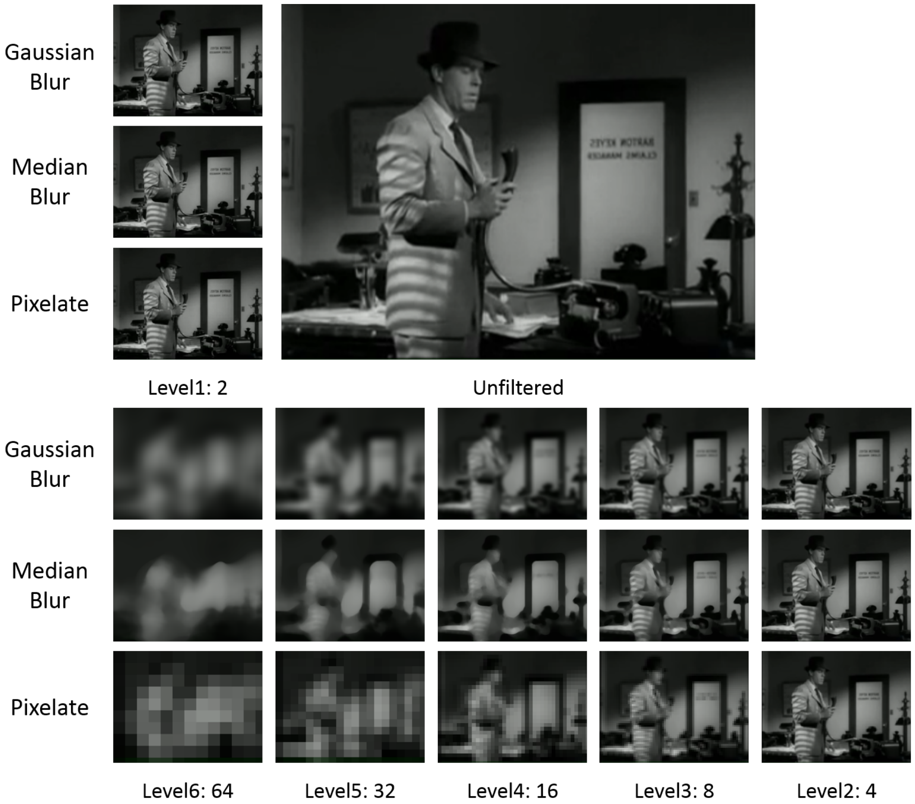


Fig. 2.4. Examples of the filters - gaussian blur, median blur and pixelation for six different radius levels. Original Frame obtained from videoclip “Double Indemnity - 02560.avi” in the Hollywood2 dataset [24]

The strategy we use to avoid exposure of the crowd-workers to private information content is to initially obfuscate the video entirely. The most popular techniques for obfuscation to preserve privacy are blurring and pixelating. The work [8] compares the use of these techniques for preserving privacy.

However, these techniques have shown limitations in securely redacting the images. The extensive work on preserving privacy [25] shows that if the obfuscation—blur or pixelation—used for protecting privacy is mimicked on an available face database, an automatic face recognition works even after obfuscation, and in case of pixelation, even better. The work [26] concludes that there is no general blur level that can be applied to an image that can completely preserve privacy and yet keep the image utilizable. The work [27] provides a promising technique for preserving privacy while maintaining the usability of images by averaging the features of a given face with other faces in the database, creating a new un-identifiable face. However, the technique

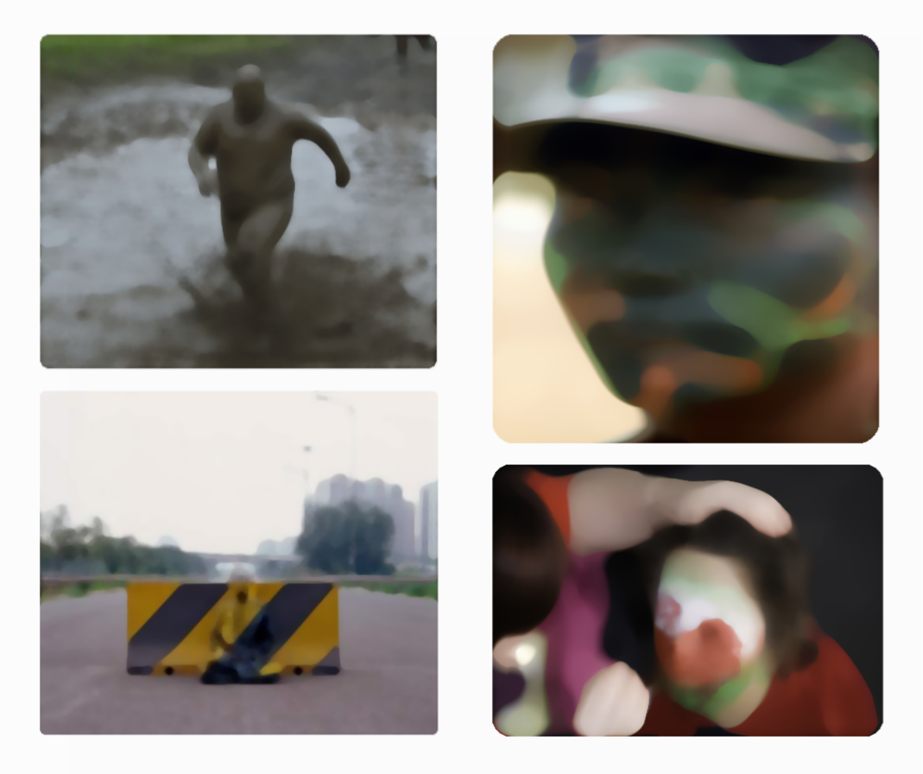


Fig. 2.5. Examples of Human Images as shown in Figure 2.1 Obfuscated Using Median Filter. Original images: top-right from wikimedia (public domain) [17]; top-left [18], bottom-left [19] and bottom-right [20] from flickr licensed under CC BY 2.0. Modification: Blurring using median filter.

works only when the exact location of face is already known. It cannot be applied on a global level.

In our work, we have used median blurring instead of the more popular gaussian blur. Median filtering is widely used for removing noise from an image. However, as the radius of median filter is increased, segmentation of an image starts occurring while the image loses the finer details. This type of filtering allows preserving important configurational information of the image such as spatial location, temporal behaviour, general geometry and object groups.

The human images shown in Figure 2.1 are presented again in Figure 2.2 after applying a median filter to these images. As can be seen from Figure 2.2, the locations of human faces are still discernible, whereas the identities have been redacted to a huge extent.

However, as was also discussed in section 1.3, implementing fixed level of global obfuscation to completely redact sensitive information, impedes the ability of a worker to see the necessary information. Hence, to provide the crowd-workers with the required information to build a hypothesis about the contents of the video, we allow the crowd-workers to click anywhere on the video and reveal a small window of information that can be used to understand the substance of the video. This technique of revealing small portion of the ROI is inspired from the online game called "Bubbles" by Jia Deng, created for their research "Fine-Grained Crowdsourcing for Fine-Grained Recognition" [28].

## 2.4 Crowdsourcing Annotation for Privacy-sensitive Videos

A recent study related to privacy vs. accuracy trade-offs for crowdsourcing annotation of behavioral videos [22], demonstrates the use of blurring for ensuring privacy while compromising accuracy of annotations. Presenting privacy-sensitive videos on task markets, makes them vulnerable to malicious attacks from crowd for identity disclosure. There have been previous studies for understanding behavior of crowd and assigning work based on a worker's reputation [29, 30]. It has been shown that crowdsourced tasks are not only vulnerable to an untrusted worker but can be hacked by an entire malicious crowd [31].

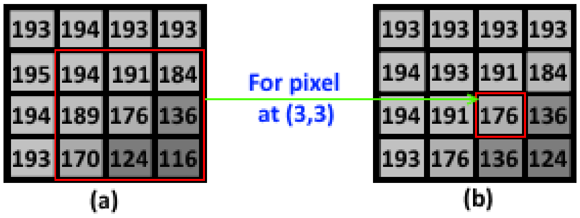
Hence, it becomes extremely important for ensuring that the system is immune to the malicious attacks from the crowd. Section 3.1.3 explains the measures taken to resist attacks from an entire crowd for disclosing an identity.

# 3. SYSTEM DESCRIPTION

The system that we have developed, provides an interface for the crowd-workers to analyze video contents without getting exposed to the private information content of the video. The system encapsulates three major processes:

## 3.1 Obfuscation

Obfuscated video clips are presented to the crowd-workers, as shown in Figure 3.2, which they can pause, play forward or reverse. The purpose of obfuscation is to reveal an initial estimation of the video contents. For this, we use the median filtering technique. This technique is generally used to remove noise in Image Processing, as it re-assigns every pixel the median of its neighbours in a given radius. This decreases the number of outliers in a region, sharpening the image.



| | | | |
|---|---|---|---|
| 193 | 194 | 193 | 193 |
| 195 | 194 | 191 | 184 |
| 194 | 189 | 176 | 136 |
| 193 | 170 | 124 | 116 |

(a)

| | | | |
|---|---|---|---|
| 193 | 193 | 193 | 193 |
| 194 | 193 | 191 | 184 |
| 194 | 191 | 176 | 136 |
| 193 | 176 | 136 | 124 |

(b)

Fig. 3.1. Process of Median filtering. (a) Original image (b) Median Blur (radius 3) applied to original image in OpenCV

However, when this filtering is applied with larger radius, it allows segmentation of the image with less profound boundaries. This converts the image into blob-like structures, that assist in releasing an initial estimation of the video content, however, restricting the finer details of it.

We chose to use the median blur instead of a gaussian blur for obfuscation. As would be describe in the next section (section 3.2), we allow the crowd-workers to request revealing small regions of the original image. If a convolution-type filtering is used and even if a small part of original information is revealed, it would allow a malicious crowd-worker to approximate the original function used for filtering. Using this, the worker may be able to reverse the filtering effect to an extent that might make the faces present in the videos recognizable. In case of median filtering, though it maintains the structural properties of an image, the original information is completely lost. This makes it immune to deconvolution attacks using parts of original image.

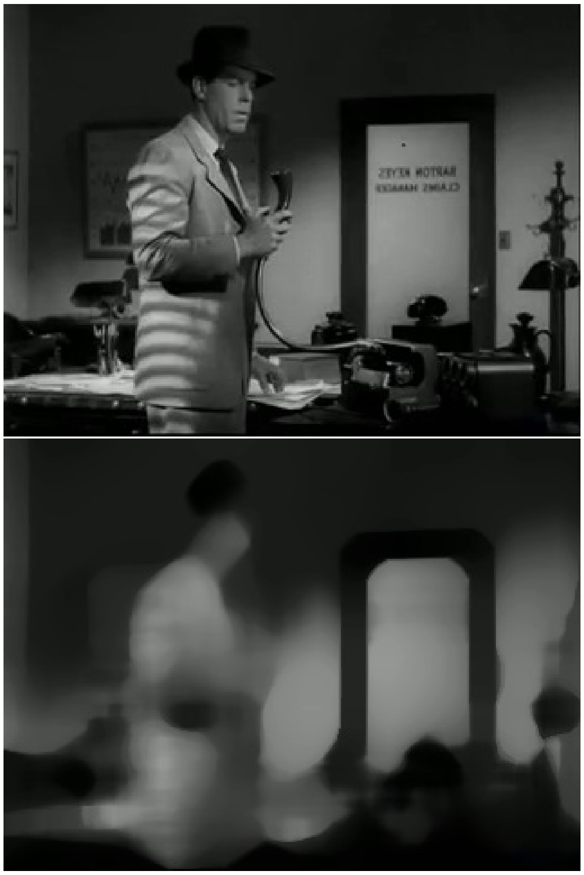


Fig. 3.2. Above: Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24]; Below: Median Blur (radius 23) applied to original frame using OpenCV.

For our system, we apply to all the video frames a median filter of radius 23 using OpenCV function “medianBlur()”. This radius was empirically observed to be enough to deidentify the faces present in the video frames of the dataset.

## 3.2 Controlled Local Disclosure

Once the workers have made their initial judgment about the video contents, the system allows the crowd-workers to pause the video and reveal a particular region by clicking on it. The heart of our system lies in the algorithm that determines the radius of the area that gets exposed after these clicks from the workers.

First of all, we present the method for calculating the radius of the information revealed. We start by removing the noise from the video frame under analysis by performing median filtering of radius 3. We then convert it into a binary image using Canny Edge Detection algorithm [32] with lower threshold set to 50 and upper threshold set to 100.

To calculate the radius of the region that can be safely shown, we find a set of nearest edges for the pixel under consideration by finding the nearest edge in all the directions. We then find a set of four points belonging to the nearest edges set of this pixel, such that these points are separated by 90 degrees angle with reference to this pixel. Using these four points, an estimate of the ratio of the width and length of the overall contour is obtained. This ratio is compared with the general human face width-to-height ratio (FWHR) [33] with an approximation of 20%. If a pixel is found to be a part of such a contour, a restricted region is revealed to the user. The general rule of thumb used by artists to sketch human faces suggests that various features of a face are one-fifth of the face width. Hence, we restrict the radius of the region revealed to one-tenth (i.e. diameter is one-fifth) of the approximated width of the contour, which would prevent showing more than one feature of a face. The pseudocode for the algorithm is shown below (Algorithm 1).

**Algorithm 1** Calculate Safe-To-Reveal Radius

```
procedure CalculateRadius(binaryImage, clickX, clickY)
    for $\theta = 0^\circ$ to $90^\circ$ do
        sumWidths ← 0
        sumHeights ← 0
        templateMatchCount ← 0
        templateMatch ← false
        p1 ← firstEdge(binaryImage,clickX,clickY, $\theta$)
        p2 ← firstEdge(binaryImage,clickX,clickY, $\theta$+90)
        p3 ← firstEdge(binaryImage,clickX,clickY, $\theta$+180)
        p4 ← firstEdge(binaryImage,clickX,clickY, $\theta$+270)
        if p1 $\neq$ NULL and p2 $\neq$ NULL and p3 $\neq$ NULL and p4 $\neq$ NULL then
            templateMatch ← true
            templateMatchCount ← templateMatchCount + 1
            sumWidths ← sumWidths + distance between p1 and p3
            sumHeights ← sumHeights + distance between p2 and p4
    if templateMatch == true then
        width ← sumWidths/templateMatchCount
        height ← sumHeights/templateMatchCount
        if height < width then Swap(width, height)
        if 0.8*(5/6) $\leq$ width/height $\leq$ 1.2*(5/6) then
            radius ← width/10
        else radius ← width/2
    else radius ← distance of (clickX,clickY) to nearest edge
    return radius
```

This radius is extremely conservative and restricts the revealed region even for general objects including lamps, muffins, etc. However, it shows larger areas where the width-to-height ratio cannot be approximated to the face width-to-height ratio. This allows obvious patterns like stripes and elongated ellipses be revealed to the worker. The output of this algorithm for 100 random clicks on the video frame shown in figure 3.2, for 2 independent iterations is shown figure 3.3.

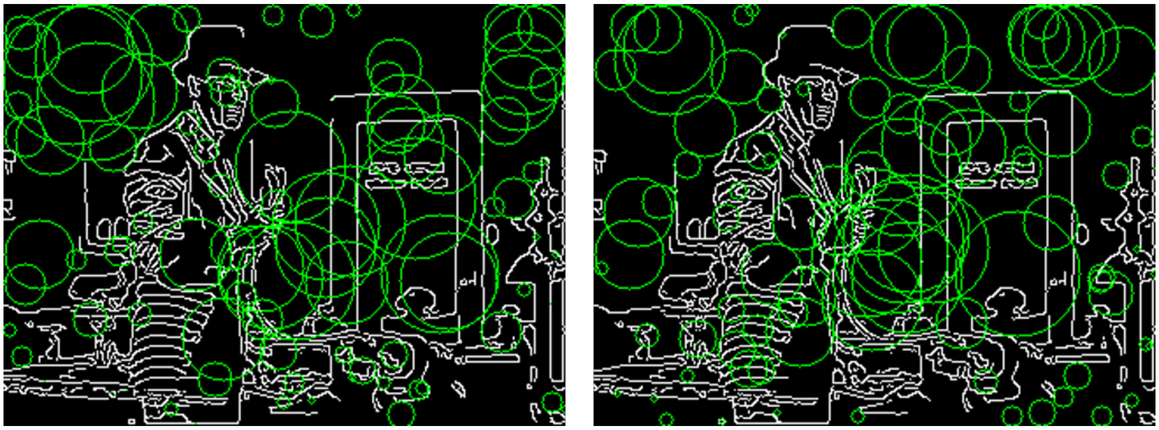


Fig. 3.3. Two independent iterations for calculating safe-to-reveal radius for the video frame shown in figure 3.2, each with 100 random simulated clicks.

Once we know the radius of the region that is safe to be revealed, we consider two different approaches for revealing this area—keyhole mode and foveated mode. In this paper, we compare these two different interaction techniques which differ in the way the information is disclosed using the safe-to-reveal radius as shown in figure 3.4. In the case of *keyhole mode*, the information within the calculated radius is displayed as it is without applying any filter. In the case of *foveated mode*, information within only half of the originally calculated radius is revealed as it is. From there, a median filter is applied with gradually increasing radius up to a distance equal to the initially estimated radius.

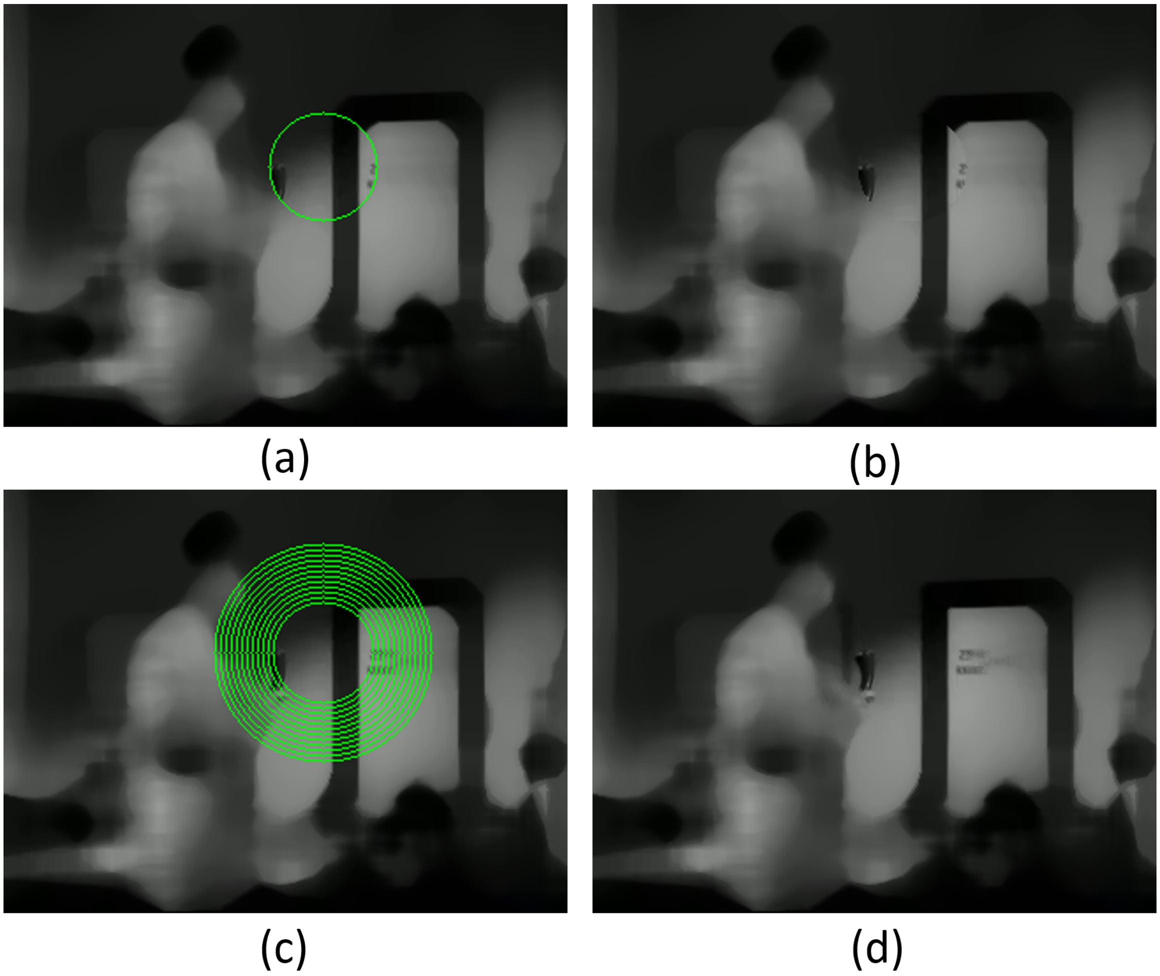


Fig. 3.4. Controlled Local Disclosure for video frame shown in figure 3.2. (a) Keyhole mode output with marking (b) Keyhole mode output (c) Foveated mode output with marking (d) Foveated mode output.

## 3.3 Privacy Protection During Crowd Interaction

However, if we allow revealing the regions based only on the above criteria, the crowd-worker may click at all the points on a face and reveal it entirely. Hence, once a small region is opened for viewing, we constrain the allowed regions for next clicks. The worker is not allowed a click for which twice the radius of the to-be-revealed

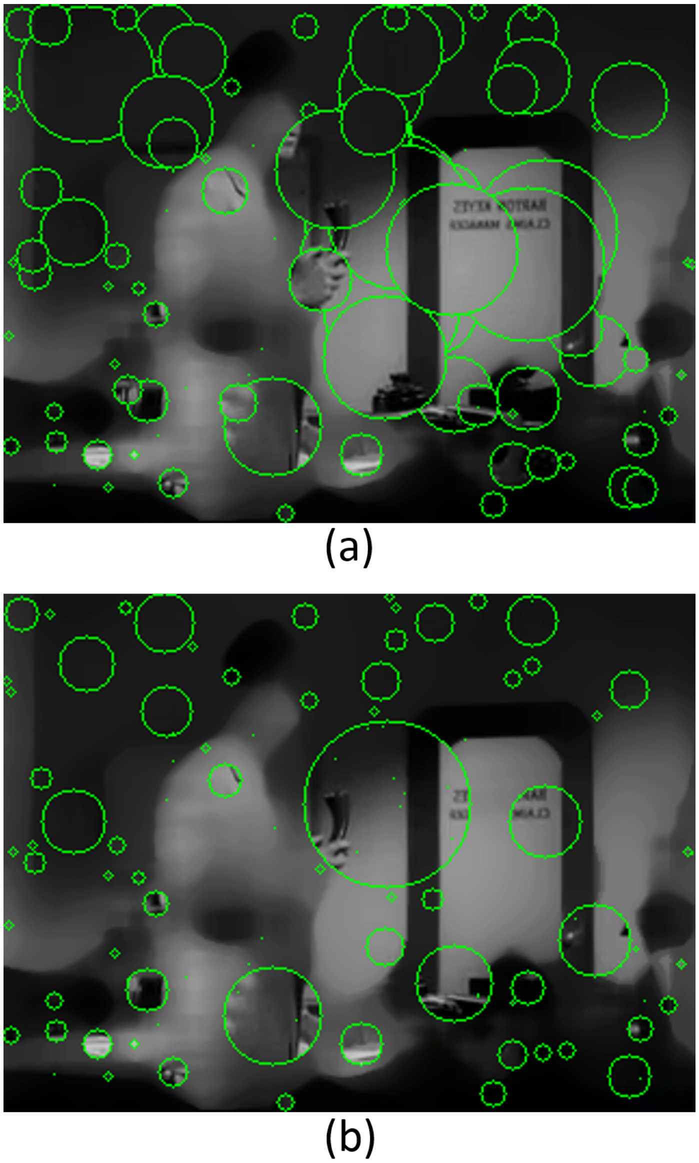


Fig. 3.5. (a) Output of Fovea system in keyhole mode for 100 random clicks without protection against crowd-collusion (b) Output of Fovea system in keyhole mode for 100 random clicks (same clicks as in (a)) with protection against crowd-collusion. Green circles are not a part of Fovea system's output. Original Frame obtained from videoclip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24]

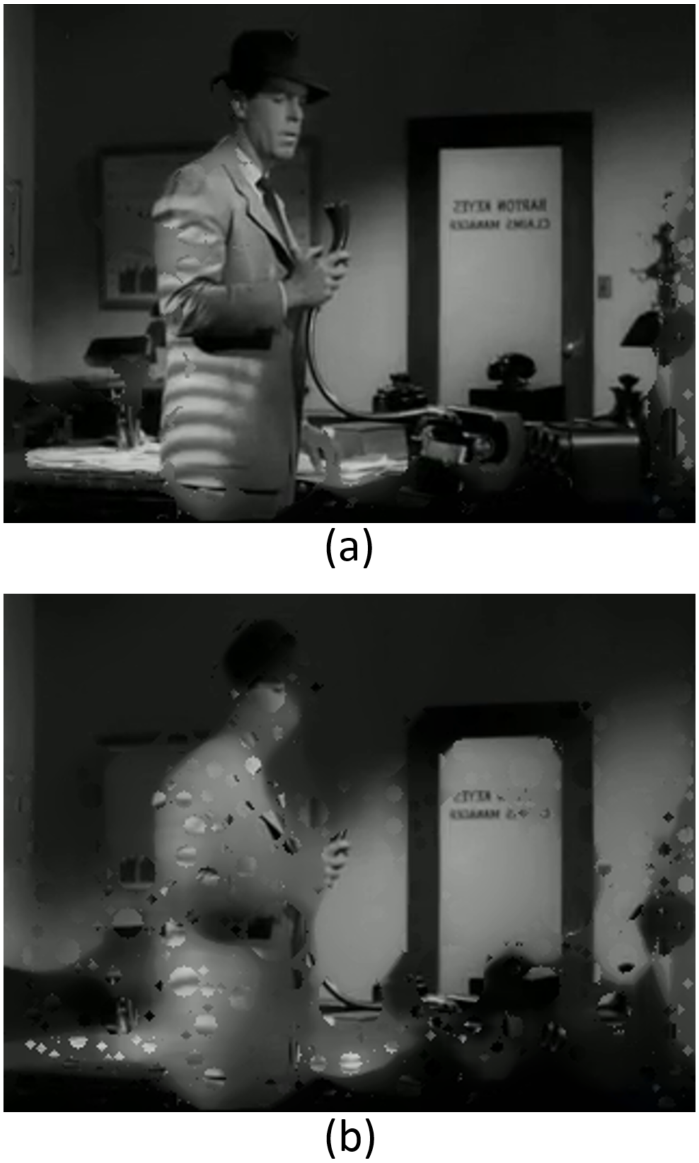


Fig. 3.6. (a) Output of Fovea system in keyhole mode for 1000 random clicks without protection against crowd-collusion (b) Output of Fovea system in keyhole mode for 1000 random clicks (same clicks as in (a)) with protection against crowd-collusion. Original Frame obtained from video-clip "Double Indemnity - 02560.avi" in the Hollywood2 dataset [24]

region overlaps with already revealed area. Also, if the revealed region is kept static throughout the video, the subjects may move in and out of this window, completely revealing their identities. Hence, the revealed region in a particular video frame is propagated to other video frames using optical flow detection technique by [34].

Also, multiple judgments about the contents may be recorded on a single video. To restrict the second worker from clicking close to the regions exposed by the previous worker, the information revealed by the first worker is made available to the next worker for a particular video. By this, we ensure that the workers may not be able to reconstruct the sensitive information, by trading with each other the individual pieces of information that each one of them has. The consecutive workers may choose to build their hypothesis about a video from already revealed information or may choose to reveal further regions as well.

Furthermore, to ensure that the original video frame is not accessible to the crowd-worker, all the image processing is carried out at the server. The video frames are served through a web-application which ensures that nothing except for the safe-to-reveal region around clicks gets shown in the video frame requested by the client.

# 4. EVALUATION

## 4.1 Dataset

We chose 60 video clips randomly from the Hollywood2 video dataset [24]. These clips were restricted to three seconds. These clips were chosen such that there was atleast one human face present in the video clip. An accompanying dataset of faces was created for each video. This dataset of faces contained ten faces for each video. Out of these ten faces one or two faces were of the people present in the video. The rest of the faces were chosen to be of the people visually similar to the subjects present in the video. All the faces present in this dataset of faces were obtained from Wikimedia Commons [35].

For establishing the ground truth, each of the three second video clip was annotated manually for the location of the faces present in the corresponding faces dataset along with the face reference number. The videos were also annotated for the action present in the video. These video consist one or none of the action out of: 'Answering a phone', 'Handshaking', 'Hugging', 'Kissing', 'Sitting down', 'Standing up', 'Climbing up or down the stairs', 'Getting out of/ getting in a car', or 'violence/ weapon'.

Fig. 4.1. Examples of video clips from the Hollywood2 dataset [24]

## 4.2 Experimental Setup

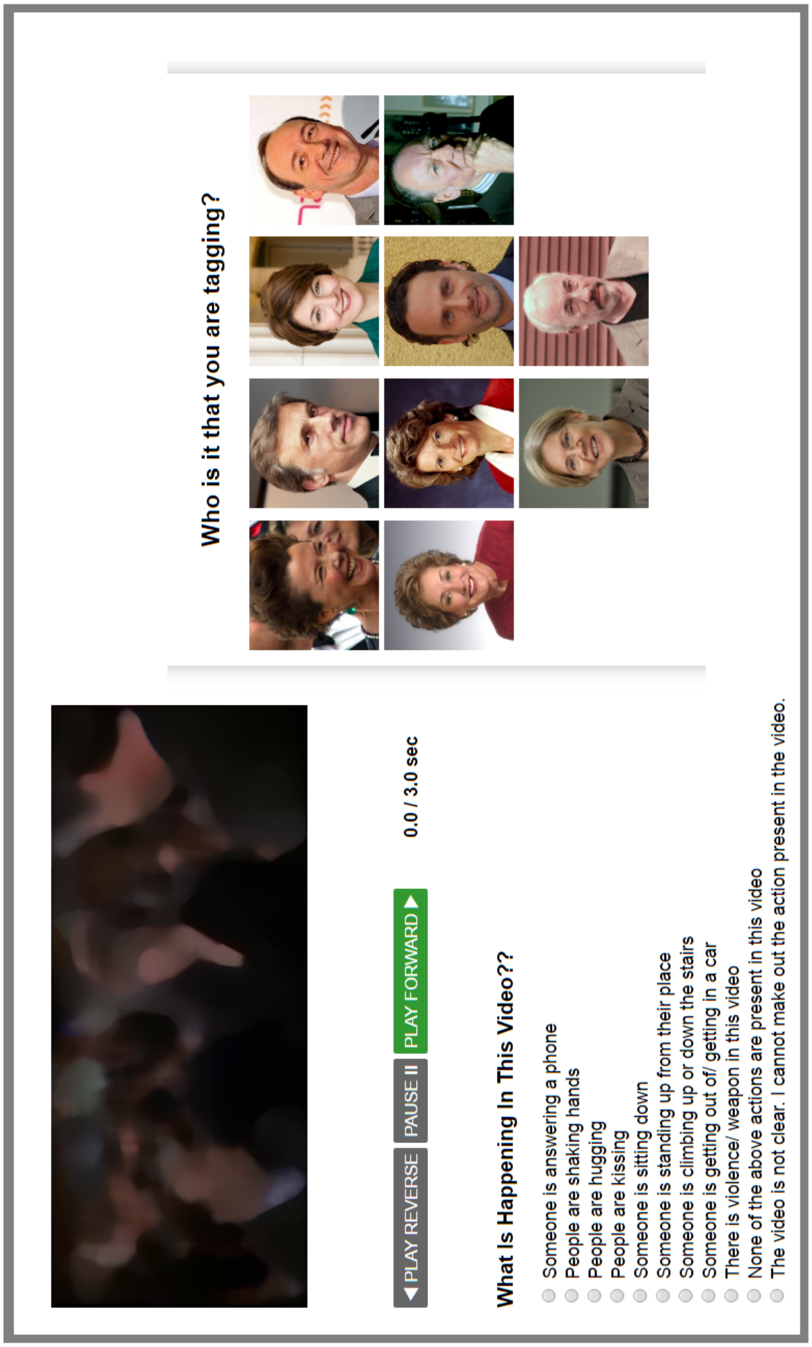
### 4.2.1 System Implementation

For this evaluation of this thesis, We developed an interface that was presented to the crowd-workers on Amazon Mechanical Turk. Figure 4.2 shows one such assignment from the HITS created for the experiment. Each assignment in a HIT included tagging and answering questions for three videos with the three different levels of revelation—*filter only, keyhole mode, foveated mode.*

The experiment was started with 20 HITS, each consisting of one assignment, which had three different movie clips blurred using a median filter of radius 23. The first clip allowed no revelation—*filter only mode*; second clip allowed revealing a small region using discrete boundaries—*keyhole mode*—whereas the third clip allowed revealing regions with gradually decreasing blur—*foveated mode.* Once a particular assignment was completed by a crowd-worker, another assignment was added to that HIT with a maximum number of assignments limited to three per HIT. This was done to ensure that no worker gets to see the same video twice even if a particular assignment was returned several times by different workers. The new assignment added, had the same sequence of videos, however, a new arrangement of the methods, so that all the combinations of videos and methods could be completed. A latin square algorithm was used for assigning jobs (video+method) to a particular assignment.

For each presented video, the crowd-workers were asked to watch the blurred video clip. They had to tag faces then and match them to the person who they thought were present in the tagged location. The interface for this is shown in figure 4.2.

As can be seen from the interface, the crowd workers were asked two questions - 1) What is happening in the video? and 2) Who is it that you are are tagging?



Who is it that you are tagging?

What Is Happening In This Video??

○ Someone is answering a phone
○ People are shaking hands
○ People are hugging
○ People are kissing
○ Someone is sitting down
○ Someone is standing up from their place
○ Someone is climbing up or down the stairs
○ Someone is getting out of/ getting in a car
○ There is violence/ weapon in this video
○ None of the above actions are present in this video
○ The video is not clear. I cannot make out the action present in the video.

Fig. 4.2. User Interface for tagging and analyzing videos through Crowd-Sourcing on AMT

### What is happening in the video?

This question was put forward to analyze how well could the crowd workers understand the content of the video. They were provided with eleven choices including "The video is not clear. I cannot make out the action present in the video." The default option selected was set to "No Response" (not shown on the interface) to filter out the effect of accepted and submitted, however, unattempted assignments. The first nine choices were selected such that any given video would have only one of these options or none; hence, the tenth option of " None of the above actions are present in this video."

### Who is it that you are tagging?

This question was intended for evaluating the extent to which the system gives away information about the identities of the people in the videos. Each video in the dataset contains a random number of subjects. Some of the clips are of a single individual carrying out a task while some have large crowds of people. Each video clip was accompanied by a set of ten faces which had either one or two of the faces present in the video. The rest of the faces in the options provided were chosen to be visually similar to the actual face present. The crowd-workers were not informed of the number of faces from the given options that were present in the video, to decrease the possibility of success by chance.

### 4.2.2 Crowd Interaction

The crowd-workers were asked to pause the video and select a region on the frame where they thought there was a face from the given options. This "tagging" of the face which included the chosen face out of the given options, as well as the coordinates of this face on the frame, was compared to the ground truth annotations for the dataset.

If the face matched the actual face and the locations were within the annotated face regions, the tagging was considered as a success; or failure otherwise.

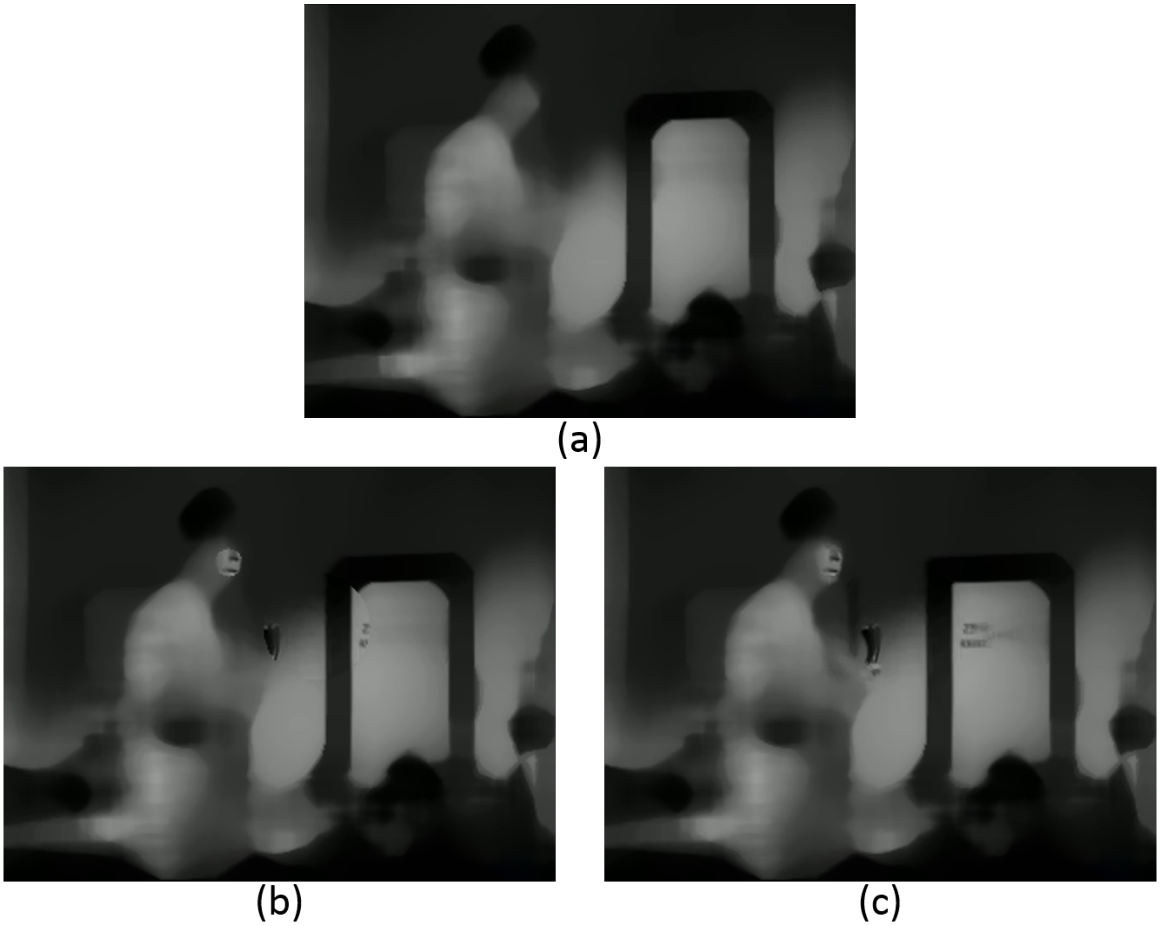


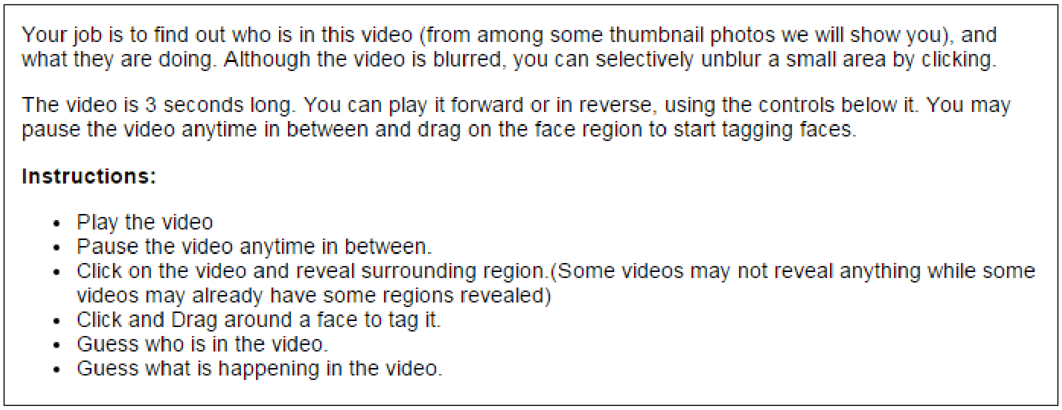
Fig. 4.3. (a) Only filter mode; Disclosure of a region surrounding clicks by (b) keyhole mode and (c) foveated mode; Original Frame obtained from videoclip “Double Indemnity - 02560.avi” in the Hollywood2 dataset [24]

Apart from tagging the regions for a face, the interface also allowed the crowd-workers to pause and click anywhere on the video to reveal a region. Each assignment, as explained earlier, had three videos, one of them allowed revealing a circular window of information using keyhole mode, the other one using foveated mode, while the third one did not allow revealing on clicks at all.

Figure 4.3 shows a frame in an assignment with regions revealed using *Keyhole* and *Fovea* technique. It can be seen how these two approaches provide different levels of exposure to the underlying information present in the actual frame, as well as the added ability it provides to a worker for forming a hypothesis about the contents of the video.

### 4.2.3 Instructions to Crowd-workers

Figure 4.4 shows the instructions provided to the crowd workers for carrying out the task of tagging and analyzing the videos. After several runs and feedback from the crowd-workers about the interface and their experience & understanding of the task, the instructions were improved such that it directed them to do precisely what was desired. With the following instructions, all the submitted tasks were of acceptable quality and none of the submissions were rejected or filtered out.



Your job is to find out who is in this video (from among some thumbnail photos we will show you), and what they are doing. Although the video is blurred, you can selectively unblur a small area by clicking.

The video is 3 seconds long. You can play it forward or in reverse, using the controls below it. You may pause the video anytime in between and drag on the face region to start tagging faces.

**Instructions:**

- Play the video
- Pause the video anytime in between.
- Click on the video and reveal surrounding region.(Some videos may not reveal anything while some videos may already have some regions revealed)
- Click and Drag around a face to tag it.
- Guess who is in the video.
- Guess what is happening in the video.

Fig. 4.4. Instructions provided to workers for the task

### 4.2.4 Time, Cost and Incentives

Each video clip presented was of 3 seconds. On an average, it took about a minute and a half for a crowd-worker to look at a video clip, tag the faces and answer the question about the contents of the video. So for three videos per assignment, which should take about 5 minutes to complete, workers were paid $0.75, which amounts to $9/hour. Apart from this, the workers were provided with an incentive of gaining a bonus of $0.10/correct tag if they correctly tagged a face. However, if they tagged a face incorrectly, a $0.10/incorrect tag was deducted from their bonus. These incentives were provided, to motivate an aggressive approach to breaking the system for revealing identities.

This experimental setup was designed to compare the judgments made by workers for each of the three cases—*Only filter, keyhole mode and foveated mode*—and evaluate how good or poor does a constrained revelation performs as compared to just blurring of the video. This setup provided the basis to assess the algorithms against the hypothesis that they could indeed provide an efficient solution to the problem of conveying non-sensitive information to and concealing sensitive information from the workers for any given video.

## 4.3 Results and Analysis

As described in earlier sections, the expected system should be able to reveal the actions of the subjects in the video; however, the exposed region should not show enough information to enable recognition of identities. To evaluate and validate the system we presented the above-described implementation to the crowd-workers on Amazon Mechanical Turk.

Table 4.1 & 4.2 show the results in a concise form. Actual data obtained during the experiment is presented in Appendix C. Each video had either one or two faces, with total 84 faces in all the 60 videos. As can be seen from Table 4.1, the number of faces correctly identified remains almost constant for all the three modes, whereas the number of correctly identified actions increases after introducing the disclosure modes.

Table 4.1.: Total Correctly Identified Faces and Actions

| | Only Blur Mode | Keyhole Mode | Foveated Mode |
|---|---|---|---|
| Total correctly identified faces (Out of 84) | 20 | 19 | 20 |
| Total correctly identified actions (Out of 60) | 29 | 47 | 54 |

Table 4.2 shows the total number of clicks—requests of disclosure—performed by crowd-workers for each of the 60 videos.

Table 4.2.: Number of clicks by crowd-workers for each video presented

| | Foveated Mode | | Keyhole Mode | |
|---|---|---|---|---|
| Video Number | Number of Clicks | Action Guessed Correctly | Number of Clicks | Action Guessed Correctly |
| 1 | 1 | Yes | 0 | No |
| 2 | 1 | Yes | 0 | No |
| 3 | 4 | Yes | 3 | Yes |
| 4 | 1 | Yes | 0 | No |
| 5 | 1 | Yes | 6 | Yes |
| 6 | 1 | Yes | 0 | No |
| 7 | 1 | Yes | 0 | No |
| 8 | 2 | Yes | 4 | No |
| 9 | 4 | Yes | 4 | No |

*continued on next page*

Table 4.2.: *continued*

| | Foveated Mode | | Keyhole Mode | |
|---|---|---|---|---|
| Video Number | Number of Clicks | Action Guessed Correctly | Number of Clicks | Action Guessed Correctly |
| 10 | 1 | Yes | 1 | Yes |
| 11 | 1 | Yes | 1 | Yes |
| 12 | 1 | Yes | 1 | Yes |
| 13 | 2 | Yes | 1 | Yes |
| 14 | 2 | Yes | 15 | Yes |
| 15 | 2 | Yes | 1 | Yes |
| 16 | 1 | No | 2 | Yes |
| 17 | 2 | Yes | 2 | Yes |
| 18 | 2 | Yes | 2 | Yes |
| 19 | 2 | Yes | 2 | Yes |
| 20 | 1 | Yes | 2 | Yes |
| 21 | 1 | Yes | 1 | Yes |
| 22 | 1 | Yes | 0 | No |
| 23 | 0 | No | 2 | Yes |
| 24 | 4 | Yes | 1 | Yes |
| 25 | 1 | Yes | 1 | Yes |
| 26 | 1 | Yes | 5 | Yes |
| 27 | 5 | Yes | 2 | Yes |
| 28 | 3 | Yes | 1 | Yes |
| 29 | 1 | Yes | 1 | No |
| 30 | 3 | Yes | 1 | Yes |
| 31 | 1 | Yes | 1 | Yes |

*continued on next page*

Table 4.2.: *continued*

| | Foveated Mode | | Keyhole Mode | |
|---|---|---|---|---|
| Video Number | Number of Clicks | Action Guessed Correctly | Number of Clicks | Action Guessed Correctly |
| 32 | 10 | Yes | 1 | Yes |
| 33 | 1 | Yes | 1 | Yes |
| 34 | 0 | Yes | 5 | Yes |
| 35 | 3 | Yes | 4 | Yes |
| 36 | 1 | Yes | 1 | Yes |
| 37 | 1 | No | 0 | Yes |
| 38 | 1 | Yes | 2 | Yes |
| 39 | 1 | No | 1 | Yes |
| 40 | 3 | Yes | 1 | Yes |
| 41 | 1 | Yes | 2 | Yes |
| 42 | 0 | Yes | 1 | Yes |
| 43 | 4 | Yes | 3 | Yes |
| 44 | 7 | Yes | 0 | No |
| 45 | 0 | No | 1 | Yes |
| 46 | 3 | No | 3 | No |
| 47 | 10 | Yes | 2 | Yes |
| 48 | 1 | Yes | 1 | Yes |
| 49 | 0 | Yes | 13 | Yes |
| 50 | 1 | Yes | 2 | Yes |
| 51 | 1 | Yes | 1 | No |
| 52 | 3 | Yes | 1 | Yes |
| 53 | 31 | Yes | 1 | Yes |

*continued on next page*

Table 4.2.: *continued*

| | Foveated Mode | | Keyhole Mode | |
|---|---|---|---|---|
| Video Number | Number of Clicks | Action Guessed Correctly | Number of Clicks | Action Guessed Correctly |
| 54 | 2 | Yes | 4 | Yes |
| 55 | 9 | Yes | 13 | Yes |
| 56 | 3 | Yes | 0 | Yes |
| 57 | 2 | Yes | 2 | Yes |
| 58 | 0 | Yes | 2 | Yes |
| 59 | 5 | No | 1 | Yes |
| 60 | 1 | Yes | 12 | Yes |

Appendix C shows the entire data collected during the experiment. Here, the outcomes for correctly identified faces and action by the crowd workers for a given video and method is either True or False, that is the outcomes are categorical. Hence, we carry out curve fitting for the collected data using logistic regression model.

The dataset used in this experiment includes video clips from Hollywood movies. These video clips have different recording angles, illumination, and color saturation. Hence, we include the effect of the videos for prediction as these videos have different levels of difficulties with respect to recognizing actions or faces. Though it was ensured that the same worker never sees a given video twice even with different methods, a particular worker could work on any number of videos from the available set. These workers could have different levels of abilities to carry out this task of recognizing faces and actions. Some workers may be remarkably good in recognizing faces even in a heavily blurred video while some may find it difficult even when some part of the faces was exposed. So, we include the effect of workers as well for the curve fitting.

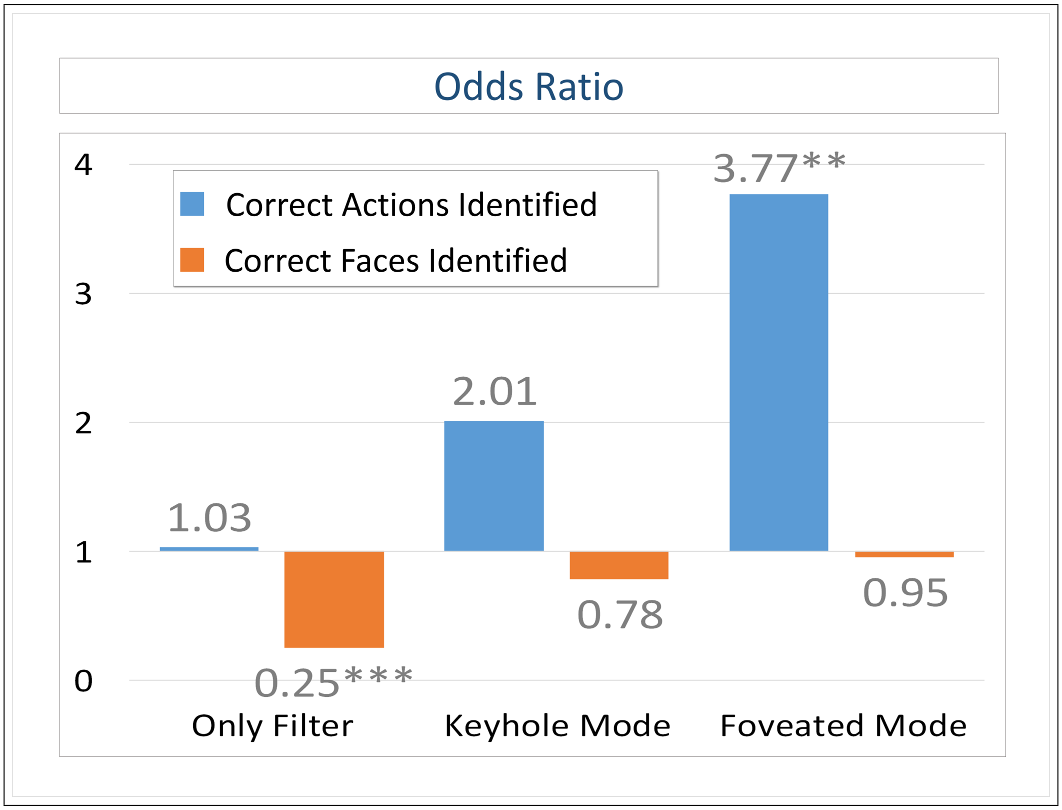


Fig. 4.5. Odds ratios obtained using logistic regression model of the effects of methods—Only filter, Keyhole mode and Foveated mode—on correct actions identified and correct faces identified. Model has been adjusted for the effects of different videos and workers. **$p < 0.01$, ***$p < 0.001$.

Performing the logistic regression for the above model, the fitting parameters obtained are tabulated in Table 4.3 and Table 4.4.

Table 4.3.
Fitting Coefficients from Logistic Regression For Correct Actions

| Methods | Fitting Coefficients |
|---|---|
| Filter Only | 0.02869 |
| Keyhole mode | 0.69938 |
| Foveated mode | 1.32838 |

Table 4.4.
Fitting Coefficients from Logistic Regression for Correct Face Tags

| Methods | Fitting Coefficients |
|---|---|
| Filter Only | -1.379635 |
| Keyhole mode | -0.246928 |
| Foveated mode | -0.049108 |

To compare the effect of the methods on the success rate of recognizing a correct action/ correct tag, we calculated the Odd's Ratio using the fitting parameters obtained above. Figure 4.1 compares the Odds Ratios for these methods. As can be observed, the keyhole and foveated modes perform significantly better than using the only blur approach for revealing the necessary information required to form a hypothesis about the contents of the videos. Moreover, it is observed that Fovea performs much better than Keyhole for conveying this information.

Table 4.5.
ANOVA Chi-squared test results for *method* as predictor and *correct face tags* as outcome

| | Df | Deviance<br>Resid. | Df<br>Resid. | Dev | Pr(<Chi) |
|---|---|---|---|---|---|
| NULL | | | 250 | 273.75 | |
| factor<br>(method) | 2 | 0.057256 | 248 | 273.69 | 0.9718 |

If we look at the results obtained for the success rate of a crowd-worker recognizing a subject present in the video, the Odds Ratios for all the three methods are less than 1. Keyhole and Fovea have higher Odds Ratios, indicating that they allow more information than the only blur case, as is expected. However, performing a chi-square in R for these three methods for the outcome - correct_tags (Table 4.5), gives a $p - value = 0.9718$, indicating that the performance of all the three methods is almost the same for the case of concealing identities.

# 5. SUMMARY

## 5.1 Discussion

In this work we have addressed the following two questions:

1. Is it possible to reveal the nature of behavior and/or presence of an object, without disclosing the identities of any depicted faces?
2. When revealing a small region, is an abrupt transition (keyhole mode) or a gradual transition (foveated mode) from clear to blurred, more useful for helping workers test their hypotheses?

As part of this work, we developed a complete system for answering the above questions as well as to provide a feasible solution for the privacy-preserving crowdsourced video annotation task. The experiments revealed the following:

1. The implemented system allows revealing the nature of behavior and/or presence of an object, disclosing the identities of any depicted persons not significantly more than what is already disclosed by the initial blur. The p-value of 0.9718 obtained from a chi-squared test (refer Table 4.5) shows very little influence of allowing controlled disclosure over identity recognition by workers.
2. The evaluation results show that the workers were able to recognize the actions significantly better using the local disclosure tool. As seen from the Odds ratio (Figure 4.1), probability of a worker recognizing an activity correctly increases more than twice for both the modes. This shows that the human workers are able to test their hypotheses about what behaviors and/or objects may be present using local disclosures.

3. Also, the Odds ratio (Figure 4.1) shows that the foveated mode significantly increases the chances of a worker recognizing an activity correctly, over the keyhole mode. The probability increases by almost four times ($\approx 3.8$), when using foveated mode.

Apart from the evaluation results enlisted above, the implemented system exhibits following capabilities:

1. Interactive video-editing using server-side image processing
2. Protection against crowd collusion
3. Creation of database of video annotations

## 5.2 Future Work

### 5.2.1 Obfuscation

As was observed, the chance of face recognition depends significantly on the level of initial blur. For evaluation of this thesis, we used median filter with empirically decided radius. However,other possible methods of obfuscation as well as different levels of blur in combination to the developed local disclosure tool needs to be studied, to evaluate their effect on the face/action recognition ability of a crowd-worker.

### 5.2.2 Forming Hypotheses

Extending the above discussion about initial obfuscation, our system assumes that the initial blur allows enough information to guide the crowd-worker to the region of interest for clicking and revealing the content necessary for forming hypotheses. The queries used for evaluation of this thesis included recognition of only prominent actions in the video-clips. Effects of other techniques such as *adaptive blur* needs to be studied for allowing a better initial estimation of video content.

### 5.2.3 Local Disclosure

For our system, we disclosed the original image in the calculated safe-to-reveal neighborhood of the click. Instead of showing the original image itself, a modified image may be disclosed such that it increases the chance of accurate activity recognition while decreasing the chance of identity disclosure.

### 5.2.4 Dataset

For the evaluation, the experimental set-up consisted of few, probably known, faces of actors and actresses, present in the video clip. This attempted in modeling the real world situation where a crowd-worker with his/her limited dataset of known faces, gets to see a blurred video for analysis that happens to have a subject from his/her dataset of known faces. A study with surveillance videos or body-camera footage needs to be carried out.

## 5.3 Conclusion

Through our work, we have presented a crowd-powered system for unconstrained video annotation that ensures the privacy of the subjects found in the video. We have designed and evaluated a novel technique of providing a subset of a given visual information, with two different variations, such that no facial identities are revealed. Through our evaluations, we conclude that our system provides more details through revealing restricted regions, which enables the crowd-workers to annotate the contents of the video with increased accuracy. At the same time, the system limits the revelation of facial identities, so that the workers do not have any more information than that available through only blurring.

# REFERENCES

# REFERENCES

[1] Amazon mechanical turk. Accessed: 2016, October 30. [Online]. Available: https://www.mturk.com/mturk/welcome

[2] Upwork. Accessed: 2016, October 30. [Online]. Available: https://www.upwork.com/

[3] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *The Visual Computer*, vol. 29, no. 10, pp. 983–1009, 2013.

[4] O. P. Popoola and K. Wang, "Video-based abnormal human behavior recognitiona review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 865–878, 2012.

[5] M. Javan Roshtkhari and M. D. Levine, "Online dominant and anomalous behavior detection in videos," in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2611–2618.

[6] P. V. K. Borges *et al.*, "Video-based human behavior understanding: a survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1993–2008, 2013.

[7] States grapple with public disclosure of police body-camera footage. Accessed: 2016, October 23. [Online]. Available: http://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2015/09/22/states-grapple-with-public-disclosure-of-police-body-camera-footage

[8] M. Boyle *et al.*, "The effects of filtered video on awareness and privacy," in *2000 ACM Conference on Computer Supported Cooperative Work (CSCW)*, 2000, pp. 1–10.

[9] W. S. Lasecki *et al.*, "Glance: Rapidly coding behavioral video with the crowd," in *2014 ACM Symposium on User Interface Software and Technology (UIST)*, 2014, pp. 551–562.

[10] Wikimedia commons: Robert rockwell man from blackhawk 1959. Accessed: 2016, November 06. [Online]. Available: https://commons.wikimedia.org/wiki/File:Robert_Rockwell_Man_from_Blackhawk_1959.JPG

[11] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: past, present and future," *Computer Vision and Image Understanding*, vol. 138, pp. 1–24, 2015.

[12] M. J. Jones and J. M. Rehg, “Statistical color models with application to skin detection,” *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[13] B. Jedynak *et al.*, “Maximum entropy models for skin detection,” in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, 2003, pp. 180–193.

[14] H. Kruppa *et al.*, “Skin patch detection in real-world images,” in *Joint Pattern Recognition Symposium*, 2002, pp. 109–116.

[15] P. Kakumanu *et al.*, “A survey of skin-color modeling and detection methods,” *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[16] N. Sebe *et al.*, “Skin detection: A bayesian network approach,” in *2004 International Conference on Pattern Recognition*, 2004, pp. 903–906.

[17] Wikimedia commons: Close-up of constructionman apprentice. Accessed: 2016, November 06. [Online]. Available: https://commons.wikimedia.org/wiki/File:US_Marines_DM-SD-02-03639_Cons tructionman_Apprentice_JWTC_camouflage_facepaint.JPEG#filelinks

[18] DVIDSHUB. Creative commons: Tough mudder. Accessed: 2016, November 06. [Online]. Available: https://www.flickr.com/photos/dvids/7250518752

[19] losmundosdemou4. Creative commons: liu_bolin_02. Accessed: 2016, November 06. [Online]. Available: https://www.flickr.com/photos/41428791@N04/4228114805

[20] N. A. for Wales. Creative commons: St david’s day celebration. Accessed: 2016, November 06. [Online]. Available: https://www.flickr.com/photos/nationalassemblyforwales/3677431889

[21] P. Sinha *et al.*, “Face recognition by humans: Nineteen results all computer vision researchers should know about,” *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[22] W. S. Lasecki *et al.*, “Exploring privacy and accuracy trade-offs in crowdsourced behavioral video coding,” in *2015 ACM Conference on Human Factors in Computing Systems (CHI)*, 2015, pp. 1945–1954.

[23] G. Kootstra *et al.*, “Gestalt principles for attention and segmentation in natural and artificial vision systems,” in *2011 ICRA Workshop on Semantic Perception, Mapping and Exploration*, 2011.

[24] I. Laptev *et al.*, “Learning realistic human actions from movies,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[25] E. M. Newton *et al.*, “Preserving privacy by de-identifying face images,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[26] C. G. Neustaedter and S. Greenberg, “Balancing privacy and awareness in home media spaces,” m.S. Thesis, Dept. of Computer Science, University of Calgary, Canada, 2004.

[27] R. Gross *et al.*, “Integrating utility into face de-identification,” in *International Workshop on Privacy Enhancing Technologies*, 2005, pp. 227–242.

[28] J. Deng *et al.*, “Fine-grained crowdsourcing for fine-grained recognition,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 580–587.

[29] S. Jagabathula, L. Subramanian, and A. Venkataraman, “Reputation-based worker filtering in crowdsourcing,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2492–2500.

[30] H. Yu *et al.*, “A reputation-aware decision-making approach for improving the efficiency of crowdsourcing systems,” in *2013 International Conference on Autonomous Agents and Multi-agent Systems*, 2013, pp. 1315–1316.

[31] T. Wang *et al.*, “Characterizing and detecting malicious crowdsourcing,” *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 537–538, 2013.

[32] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

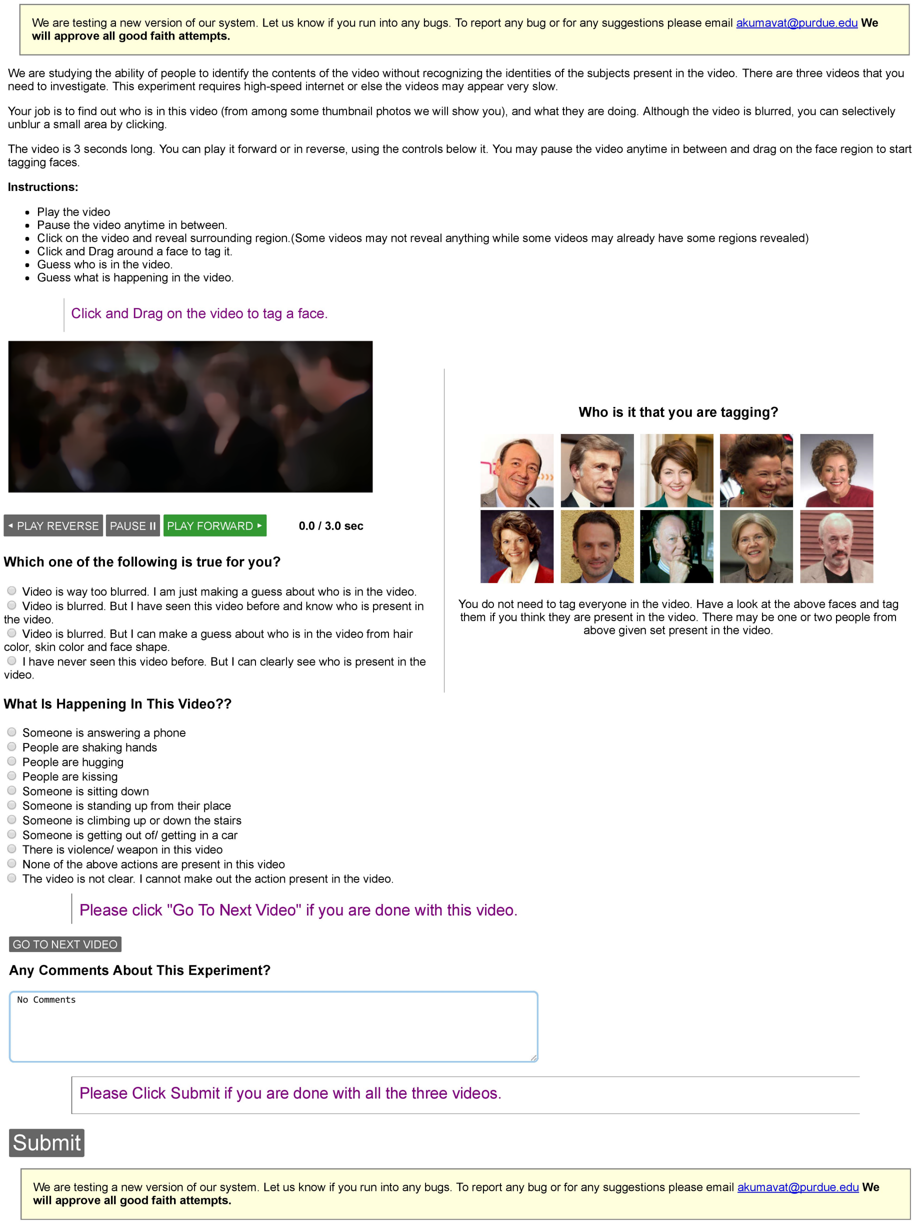
[33] R. S. Kramer, “Facial width-to-height ratio in a large sample of commonwealth games athletes,” *Evolutionary Psychology*, vol. 13, no. 1, pp. 197–209, 2015.

[34] G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Scandinavian Conference on Image Analysis*, 2003, pp. 363–370.

[35] Wikimedia commons. Accessed: 2016, November 06. [Online]. Available: https://commons.wikimedia.org/wiki/Main_Page

# APPENDIX

# A. USER INTERFACE



We are testing a new version of our system. Let us know if you run into any bugs. To report any bug or for any suggestions please email akumavat@purdue.edu **We will approve all good faith attempts.**

We are studying the ability of people to identify the contents of the video without recognizing the identities of the subjects present in the video. There are three videos that you need to investigate. This experiment requires high-speed internet or else the videos may appear very slow.

Your job is to find out who is in this video (from among some thumbnail photos we will show you), and what they are doing. Although the video is blurred, you can selectively unblur a small area by clicking.

The video is 3 seconds long. You can play it forward or in reverse, using the controls below it. You may pause the video anytime in between and drag on the face region to start tagging faces.

**Instructions:**

- Play the video
- Pause the video anytime in between.
- Click on the video and reveal surrounding region.(Some videos may not reveal anything while some videos may already have some regions revealed)
- Click and Drag around a face to tag it.
- Guess who is in the video.
- Guess what is happening in the video.

Click and Drag on the video to tag a face.

**Which one of the following is true for you?**

○ Video is way too blurred. I am just making a guess about who is in the video.
○ Video is blurred. But I have seen this video before and know who is present in the video.
○ Video is blurred. But I can make a guess about who is in the video from hair color, skin color and face shape.
○ I have never seen this video before. But I can clearly see who is present in the video.

**What Is Happening In This Video??**

○ Someone is answering a phone
○ People are shaking hands
○ People are hugging
○ People are kissing
○ Someone is sitting down
○ Someone is standing up from their place
○ Someone is climbing up or down the stairs
○ Someone is getting out of/ getting in a car
○ There is violence/ weapon in this video
○ None of the above actions are present in this video
○ The video is not clear. I cannot make out the action present in the video.

**Who is it that you are tagging?**

You do not need to tag everyone in the video. Have a look at the above faces and tag them if you think they are present in the video. There may be one or two people from above given set present in the video.

Please click "Go To Next Video" if you are done with this video.

**Any Comments About This Experiment?**

No Comments

Please Click Submit if you are done with all the three videos.

We are testing a new version of our system. Let us know if you run into any bugs. To report any bug or for any suggestions please email akumavat@purdue.edu **We will approve all good faith attempts.**

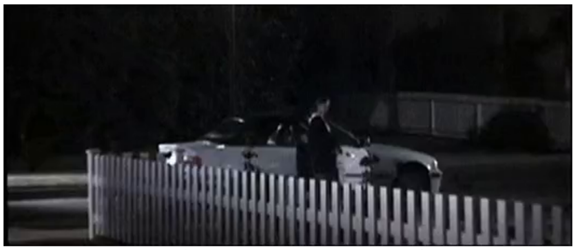
Fig. A.1. Complete user-interface used for the experiment

# B. DATASET

The 60 video clips used in this experiment which were obtained from the Hollywood2—Human Actions—dataset, created by Ivan Laptev [24] are shown below. Name of the file in the dataset, from which the frame was obtained is printed below the frame for reference.

American Beauty - 00170.avi

American Beauty - 00209.avi



American Beauty - 00443.avi

American Beauty - 00706.avi

American Beauty - 01891.avi

American Beauty - 02273.avi

Graduate, The - 03406.avi

Indiana Jones And The Last Crusade - 01483.avi

Big Fish - 02254.avi

Big Lebowski, The - 00818.avi

Being John Malkovich - 01063.avi

Being John Malkovich - 01887.avi

Being John Malkovich - 02001.avi

As Good As It Gets - 01834.avi

Being John Malkovich - 02087.avi

Big Fish - 00077.avi

Godfather, The - 02663.avi

Erin Brockovich - 03073.avi

Butterfly Effect, The - 01376.avi

Butterfly Effect, The - 01495.avi

As Good As It Gets - 01935.avi

Fargo - 01189.avi

Big Fish - 01297.avi

Big Fish - 02027.avi

Casablanca - 01404.avi

Casablanca - 01238.avi

Casablanca - 01168.avi

Double Indemnity - 02568.avi

Casablanca - 00434.avi

Double Indemnity - 02560.avi

Casablanca - 02858.avi

Casablanca - 02722.avi

Casablanca - 01659.avi

Double Indemnity - 00725.avi

Double Indemnity - 01330.avi

Double Indemnity - 01461.avi

Its A Wonderful Life - 02808.avi

Its A Wonderful Life - 03129.avi

Casablanca - 02680.avi

Lost Weekend, The - 01944.avi

Casablanca - 01615.avi

Its A Wonderful Life - 04247.avi

Its A Wonderful Life - 02291.avi

Its A Wonderful Life - 04218.avi

Graduate, The - 03480.avi

Indiana Jones And The Last Crusade - 02359.avi

Lost Highway - 01631.avi

Lost Highway - 01969.avi

Lost Highway - 02167.avi

Lost Highway - 01398.avi

Lost Highway - 01977.avi

Reservoir Dogs - 01353.avi

As Good As It Gets - 02002.avi

Pulp Fiction - 00134.avi

Pulp Fiction - 02081.avi

Crying Game, The - 01989.avi

Lost Weekend, The - 00285.avi

Indiana Jones And The Last Crusade - 00379.avi

Graduate, The - 02463.avi

Its A Wonderful Life - 02175.avi

# C. DATA

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 1 | Only Blur | 1 | - | 1 | 2 | No |
| 1 | Keyhole | 4 | 0 | 1 | 2 | No |
| 1 | Foveated | 2 | 1 | 1 | 2 | Yes |
| 2 | Only Blur | 2 | - | 1 | 2 | Yes |
| 2 | Keyhole | 1 | 0 | 0 | 2 | No |
| 2 | Foveated | 4 | 1 | 1 | 2 | Yes |
| 3 | Only Blur | 4 | - | 0 | 1 | Yes |
| 3 | Keyhole | 2 | 1 | 1 | 1 | Yes |
| 3 | Foveated | 1 | 4 | 1 | 1 | Yes |
| 4 | Only Blur | 8 | - | 0 | 2 | Yes |
| 4 | Keyhole | 17 | 0 | 0 | 2 | No |
| 4 | Foveated | 2 | 1 | 1 | 2 | Yes |
| 5 | Only Blur | 2 | - | 1 | 2 | Yes |
| 5 | Keyhole | 8 | 1 | 1 | 2 | Yes |
| 5 | Foveated | 17 | 1 | 0 | 2 | Yes |
| 6 | Only Blur | 17 | - | 0 | 1 | No |
| 6 | Keyhole | 2 | 0 | 0 | 1 | No |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 6 | Foveated | 8 | 1 | 0 | 1 | Yes |
| 7 | Only Blur | 6 | - | 0 | 1 | Yes |
| 7 | Keyhole | 16 | 0 | 0 | 1 | No |
| 7 | Foveated | 4 | 1 | 0 | 1 | Yes |
| 8 | Only Blur | 4 | - | 0 | 2 | No |
| 8 | Keyhole | 6 | 0 | 1 | 2 | No |
| 8 | Foveated | 16 | 2 | 0 | 2 | Yes |
| 9 | Only Blur | 16 | - | 0 | 1 | Yes |
| 9 | Keyhole | 4 | 0 | 1 | 1 | No |
| 9 | Foveated | 6 | 4 | 0 | 1 | Yes |
| 10 | Only Blur | 2 | - | 0 | 2 | Yes |
| 10 | Keyhole | 4 | 1 | 1 | 2 | Yes |
| 10 | Foveated | 9 | 1 | 0 | 2 | Yes |
| 11 | Only Blur | 9 | - | 0 | 2 | Yes |
| 11 | Keyhole | 2 | 1 | 0 | 2 | Yes |
| 11 | Foveated | 4 | 1 | 0 | 2 | Yes |
| 12 | Only Blur | 4 | - | 1 | 1 | Yes |
| 12 | Keyhole | 9 | 1 | 0 | 1 | Yes |
| 12 | Foveated | 2 | 1 | 0 | 1 | Yes |
| 13 | Only Blur | 7 | - | 0 | 1 | Yes |
| 13 | Keyhole | 13 | 1 | 1 | 1 | Yes |
| 13 | Foveated | 10 | 2 | 0 | 1 | Yes |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 14 | Only Blur | 10 | - | 1 | 1 | No |
| 14 | Keyhole | 7 | 1 | 0 | 1 | Yes |
| 14 | Foveated | 13 | 2 | 1 | 1 | Yes |
| 15 | Only Blur | 13 | - | 0 | 1 | No |
| 15 | Keyhole | 10 | 1 | 0 | 1 | Yes |
| 15 | Foveated | 7 | 2 | 1 | 1 | Yes |
| 16 | Only Blur | 34 | - | 1 | 1 | No |
| 16 | Keyhole | 61 | 1 | 0 | 1 | Yes |
| 16 | Foveated | 55 | 1 | 0 | 1 | Yes |
| 17 | Only Blur | 55 | - | 0 | 1 | No |
| 17 | Keyhole | 34 | 1 | 0 | 1 | Yes |
| 17 | Foveated | 61 | 2 | 1 | 1 | Yes |
| 18 | Only Blur | 61 | - | 0 | 1 | No |
| 18 | Keyhole | 55 | 1 | 0 | 1 | Yes |
| 18 | Foveated | 34 | 2 | 0 | 1 | Yes |
| 19 | Only Blur | 35 | - | 0 | 1 | Yes |
| 19 | Keyhole | 64 | 1 | 0 | 1 | Yes |
| 19 | Foveated | 55 | 2 | 0 | 1 | Yes |
| 20 | Only Blur | 55 | - | 0 | 2 | No |
| 20 | Keyhole | 35 | 1 | 0 | 2 | Yes |
| 20 | Foveated | 64 | 1 | 0 | 2 | Yes |
| 21 | Only Blur | 64 | - | 0 | 1 | No |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 21 | Keyhole | 55 | 1 | 0 | 1 | Yes |
| 21 | Foveated | 35 | 1 | 0 | 1 | Yes |
| 22 | Only Blur | 4 | - | 1 | 2 | No |
| 22 | Keyhole | 1 | 0 | 0 | 2 | No |
| 22 | Foveated | 53 | 1 | 0 | 2 | Yes |
| 23 | Only Blur | 53 | - | 0 | 1 | Yes |
| 23 | Keyhole | 4 | 1 | 0 | 1 | Yes |
| 23 | Foveated | 1 | 0 | 1 | 1 | No |
| 24 | Only Blur | 1 | - | 1 | 2 | No |
| 24 | Keyhole | 53 | 1 | 0 | 2 | Yes |
| 24 | Foveated | 4 | 4 | 1 | 2 | Yes |
| 25 | Only Blur | 38 | - | 1 | 1 | No |
| 25 | Keyhole | 47 | 1 | 0 | 1 | Yes |
| 25 | Foveated | 4 | 1 | 0 | 1 | Yes |
| 26 | Only Blur | 4 | - | 0 | 1 | No |
| 26 | Keyhole | 38 | 1 | 1 | 1 | Yes |
| 26 | Foveated | 47 | 1 | 0 | 1 | Yes |
| 27 | Only Blur | 47 | - | 1 | 2 | No |
| 27 | Keyhole | 4 | 1 | 1 | 2 | Yes |
| 27 | Foveated | 38 | 5 | 1 | 2 | Yes |
| 28 | Only Blur | 26 | - | 0 | 1 | Yes |
| 28 | Keyhole | 60 | 1 | 0 | 1 | Yes |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 28 | Foveated | 55 | 3 | 0 | 1 | Yes |
| 29 | Only Blur | 55 | - | 0 | 1 | No |
| 29 | Keyhole | 26 | 0 | 0 | 1 | No |
| 29 | Foveated | 60 | 1 | 0 | 1 | Yes |
| 30 | Only Blur | 60 | - | 0 | 1 | Yes |
| 30 | Keyhole | 55 | 1 | 0 | 1 | Yes |
| 30 | Foveated | 26 | 3 | 0 | 1 | Yes |
| 31 | Only Blur | 4 | - | 1 | 1 | No |
| 31 | Keyhole | 42 | 1 | 0 | 1 | Yes |
| 31 | Foveated | 38 | 1 | 0 | 1 | Yes |
| 32 | Only Blur | 38 | - | 0 | 2 | No |
| 32 | Keyhole | 4 | 1 | 0 | 2 | Yes |
| 32 | Foveated | 42 | 10 | 0 | 2 | Yes |
| 33 | Only Blur | 42 | - | 0 | 1 | Yes |
| 33 | Keyhole | 38 | 1 | 0 | 1 | Yes |
| 33 | Foveated | 4 | 1 | 0 | 1 | Yes |
| 34 | Only Blur | 4 | - | 0 | 1 | No |
| 34 | Keyhole | 59 | 1 | 0 | 1 | Yes |
| 34 | Foveated | 55 | 0 | 0 | 1 | Yes |
| 35 | Only Blur | 55 | - | 0 | 2 | Yes |
| 35 | Keyhole | 4 | 1 | 2 | 2 | Yes |
| 35 | Foveated | 59 | 3 | 0 | 2 | Yes |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 36 | Only Blur | 59 | - | 0 | 1 | No |
| 36 | Keyhole | 55 | 1 | 0 | 1 | Yes |
| 36 | Foveated | 4 | 1 | 1 | 1 | Yes |
| 37 | Only Blur | 29 | - | 0 | 2 | No |
| 37 | Keyhole | 57 | 1 | 0 | 2 | Yes |
| 37 | Foveated | 44 | 1 | 0 | 2 | No |
| 38 | Only Blur | 44 | - | 1 | 1 | Yes |
| 38 | Keyhole | 29 | 1 | 0 | 1 | Yes |
| 38 | Foveated | 57 | 1 | 0 | 1 | Yes |
| 39 | Only Blur | 57 | - | 0 | 1 | No |
| 39 | Keyhole | 44 | 1 | 0 | 1 | Yes |
| 39 | Foveated | 29 | 1 | 0 | 1 | No |
| 40 | Only Blur | 23 | - | 0 | 1 | No |
| 40 | Keyhole | 55 | 1 | 0 | 1 | Yes |
| 40 | Foveated | 1 | 3 | 0 | 1 | Yes |
| 41 | Only Blur | 1 | - | 1 | 1 | No |
| 41 | Keyhole | 23 | 1 | 0 | 1 | Yes |
| 41 | Foveated | 55 | 1 | 0 | 1 | Yes |
| 42 | Only Blur | 55 | - | 1 | 2 | Yes |
| 42 | Keyhole | 1 | 1 | 1 | 2 | Yes |
| 42 | Foveated | 23 | 0 | 1 | 2 | Yes |
| 43 | Only Blur | 28 | - | 0 | 2 | Yes |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 43 | Keyhole | 62 | 1 | 0 | 2 | Yes |
| 43 | Foveated | 51 | 4 | 0 | 2 | Yes |
| 44 | Only Blur | 51 | - | 0 | 1 | No |
| 44 | Keyhole | 28 | 0 | 0 | 1 | No |
| 44 | Foveated | 62 | 7 | 0 | 1 | Yes |
| 45 | Only Blur | 62 | - | 0 | 2 | No |
| 45 | Keyhole | 51 | 1 | 1 | 2 | Yes |
| 45 | Foveated | 28 | 0 | 0 | 2 | No |
| 46 | Only Blur | 39 | - | 1 | 2 | Yes |
| 46 | Keyhole | 52 | 0 | 0 | 2 | No |
| 46 | Foveated | 38 | 3 | 1 | 2 | No |
| 47 | Only Blur | 38 | - | 1 | 2 | No |
| 47 | Keyhole | 39 | 1 | 1 | 2 | Yes |
| 47 | Foveated | 52 | 10 | 0 | 2 | Yes |
| 48 | Only Blur | 52 | - | 0 | 1 | No |
| 48 | Keyhole | 38 | 1 | 0 | 1 | Yes |
| 48 | Foveated | 39 | 1 | 0 | 1 | Yes |
| 49 | Only Blur | 33 | - | 0 | 1 | Yes |
| 49 | Keyhole | 55 | 1 | 0 | 1 | Yes |
| 49 | Foveated | 46 | 0 | 0 | 1 | Yes |
| 50 | Only Blur | 46 | - | 0 | 2 | Yes |
| 50 | Keyhole | 33 | 1 | 0 | 2 | Yes |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 50 | Foveated | 55 | 1 | 0 | 2 | Yes |
| 51 | Only Blur | 55 | - | 0 | 2 | No |
| 51 | Keyhole | 46 | 0 | 0 | 2 | No |
| 51 | Foveated | 33 | 1 | 1 | 2 | Yes |
| 52 | Only Blur | 27 | - | 1 | 2 | No |
| 52 | Keyhole | 41 | 1 | 1 | 2 | Yes |
| 52 | Foveated | 38 | 3 | 2 | 2 | Yes |
| 53 | Only Blur | 38 | - | 0 | 1 | No |
| 53 | Keyhole | 27 | 0 | 0 | 1 | No |
| 53 | Foveated | 41 | 31 | 0 | 1 | Yes |
| 54 | Only Blur | 41 | - | 1 | 2 | Yes |
| 54 | Keyhole | 38 | 1 | 2 | 2 | Yes |
| 54 | Foveated | 27 | 2 | 1 | 2 | Yes |
| 55 | Only Blur | 4 | - | 0 | 1 | Yes |
| 55 | Keyhole | 40 | 1 | 0 | 1 | Yes |
| 55 | Foveated | 38 | 9 | 0 | 1 | Yes |
| 56 | Only Blur | 38 | - | 0 | 1 | Yes |
| 56 | Keyhole | 4 | 1 | 1 | 1 | Yes |
| 56 | Foveated | 40 | 3 | 1 | 1 | Yes |
| 57 | Only Blur | 40 | - | 0 | 2 | Yes |
| 57 | Keyhole | 38 | 1 | 1 | 2 | Yes |
| 57 | Foveated | 4 | 2 | 2 | 2 | Yes |

*continued on next page*

Table C.1.: Data Collected

| Video | Mode | Worker | Number of clicks | Correctly identified faces | Total faces to be identified | Actions correctly identified |
|---|---|---|---|---|---|---|
| 58 | Only Blur | 66 | - | 1 | 1 | Yes |
| 58 | Keyhole | 68 | 1 | 0 | 1 | Yes |
| 58 | Foveated | 67 | 0 | 0 | 1 | Yes |
| 59 | Only Blur | 67 | - | 0 | 1 | Yes |
| 59 | Keyhole | 66 | 1 | 0 | 1 | Yes |
| 59 | Foveated | 68 | 5 | 0 | 1 | No |
| 60 | Only Blur | 68 | - | 1 | 1 | Yes |
| 60 | Keyhole | 67 | 1 | 0 | 1 | Yes |
| 60 | Foveated | 66 | 1 | 0 | 1 | Yes |