Semantic Person Retrieval in Surveillance Using Soft Biometrics

Michael Halstead, Simon Denman, Clinton Fookes, YingLi Tian, Mark Nixon

November 27th, 2018

Table of contents

1. Schedule
2. Introduction
3. Challenges
4. Final remarks

Workshop Schedule

Session 1:

- 9:00am Organisers: Challenge Summary and Overview.
- 9:25am Gabriel R. Gonalves, Antonio C. Nazare, Matheus Diniz, Luiz Eduardo Lima Coelho and William R. Schwartz: Soft Biometric Retrieval Using Deep Multi-Task Network
- 9:50am Emil Barbuta Cipcigan and Mark Nixon: Feature Selection for Subject Ranking using Soft Biometric Queries
- 10:10am Morning Tea

Workshop Schedule

Session 2:

- 11:00am Takuya Yaguchi and Mark Nixon: Transfer Learning Based Approach for Semantic Person Retrieval
- 11:25am Arne Schumann and Andreas Specker: Attribute-based Person Retrieval and Search in Video Sequences
- 11:50am Hiren Galiyawala, Kenil S. Shah, Vandit J. Gajjar and Mehul Raval: Person Retrieval in Surveillance Video using Height, Color and Gender
- 12:15pm Organisers: Discussions, conclusions, and results presentation
- 12:30pm Workshop Close and Lunch

The Organisers

Michael Halstead Queensland University of Technology

Simon Denman Queensland University of Technology

Clinton Fookes Queensland University of Technology

YingLi Tian City College of New York

Mark Nixon University of Southampton

Motivation

- Australian Federal Police recognised a disconnect between research output and their requirements.
- Facial recognition (traditional biometric), gait recognition (soft biometric), and re-identification (soft biometrics) contribute to the state-of-the-art.
  - Facial recognition (primarily higher definition).
  - Gait requires enrollment.
  - Re-Id requires enrollment through a visual representation of the target subject.

What about when enrollment is not possible?

Motivation

Figure: Example of soft biometrics in use where traditional biometrics may fail [7].

Motivation

What did that figure tell us?

- Perpetrators often hide their face.
- Soft biometrics such as clothing are still highly visible.
- You do not need the subjects express consent to search.
- Ideal for general purpose surveillance.

Motivation

- Soft biometrics like clothing are easy to comprehend and pass on to authorities or security operators.
  - Compared to facial recognition, how easily can you describe someones face?
- No requirement for pre-enrollment.
- Finally, this challenge is bringing researchers in this field together.

Semantic Search Overview

Semantic search is based around soft biometrics, which are:

- Easy to extract visually by a person and relay to another person.
- Non-permenant (clothing can be changed).
- Lack distinctiveness (multiple people with green eyes).
- Good as a short term solution to narrow a search field.

Semantic Search Overview

Developing field of research.

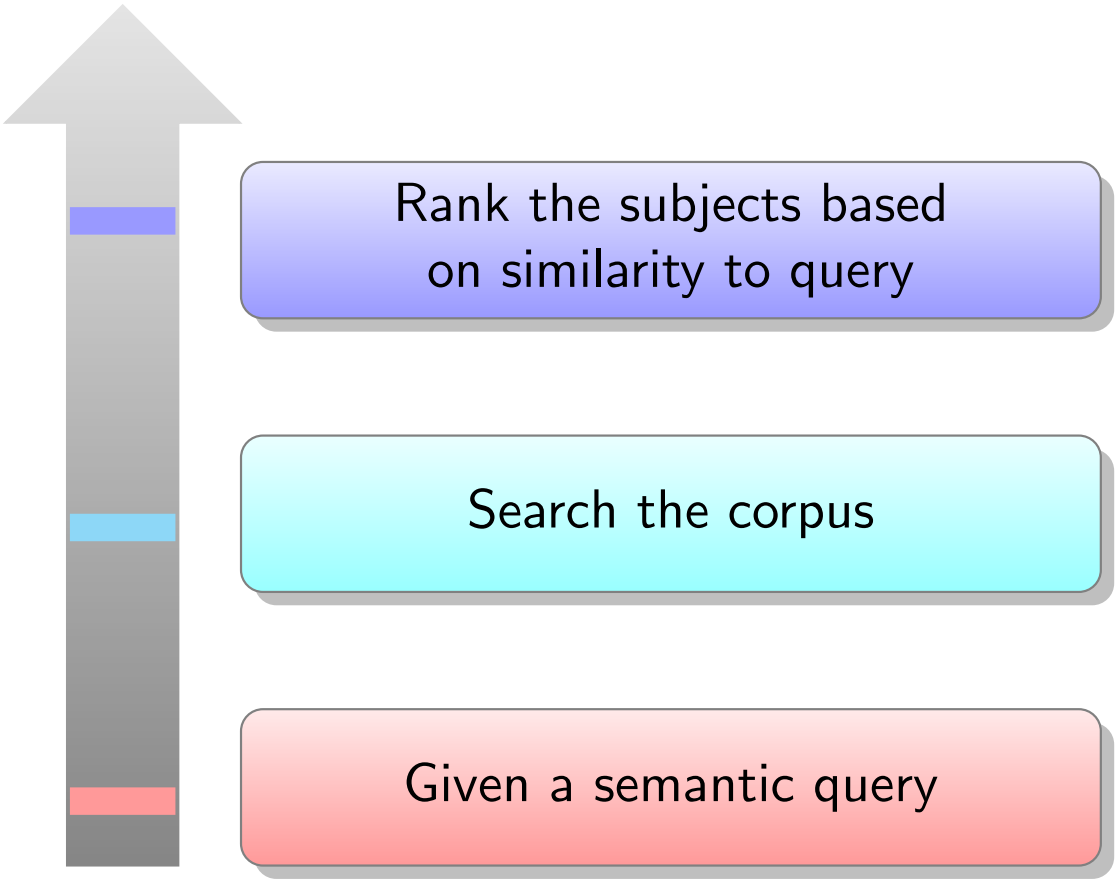
- No common protocols.
- Few real datasets, existing issues:
  - Sterile backgrounds [6];
  - Extensions to re-id (ViPER [4] - colour by [10]);
  - Not specific for this purpose (RAP [9], attribute recognition [2]).

Challenge Overview

- Task 1: Retrieval/Indexing Task.
- Task 2: Video Search Task.
- Why two tasks?
  - Capture different use cases.

Task 1

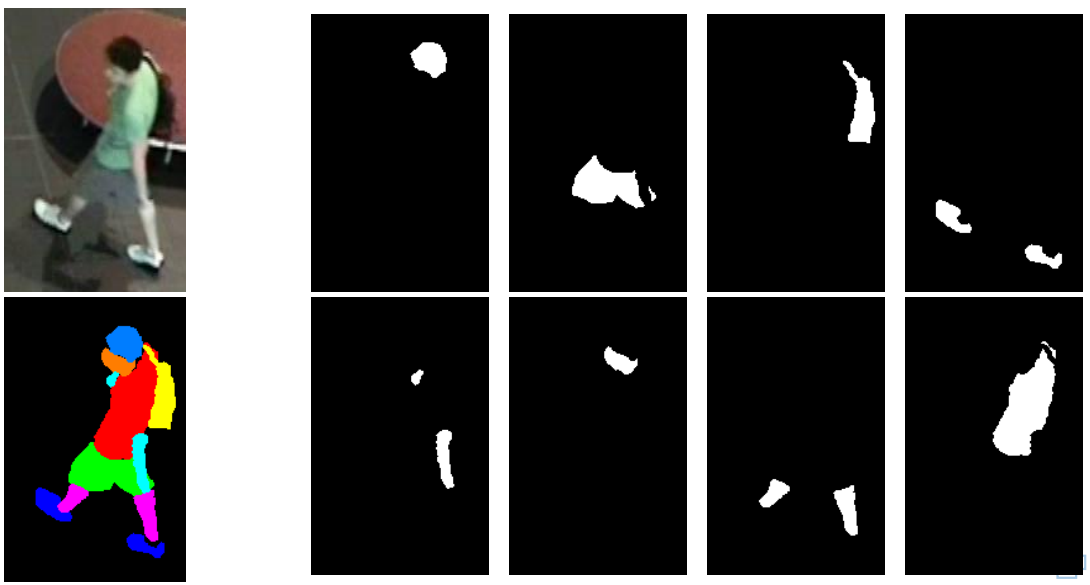
Analogous to re-identification



Task 1

Data:

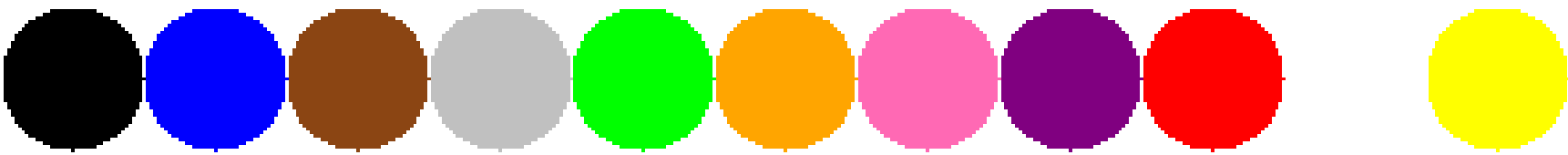
- Separated into training (520 images) and evaluation (196 images) sets.
- Subjects were captured across two locations on a university campus, using ten cameras.
- Each "image" has an RGB image, a soft biometric query, and in the training set a semantic mask is included.



Task 1

What's included in the query?

- Clothing Colour for leg and torso:
  - Three each: primary, secondary, and ancillary;
  - Eleven colours based on "Culture Colours" [1].
- Clothing patterns (textures) leg and torso:
  - Eight leg and torso textures describing a broad range.

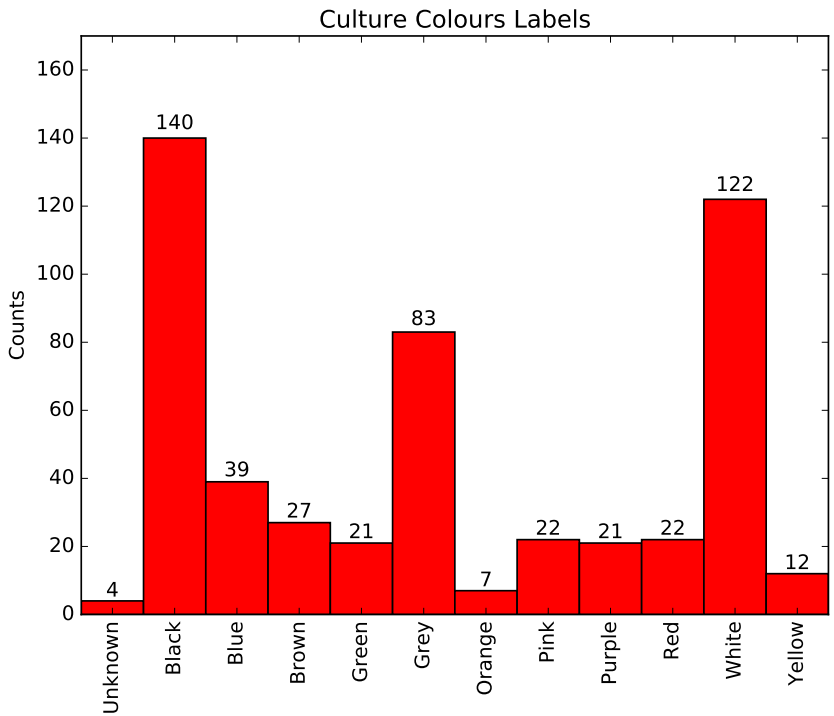


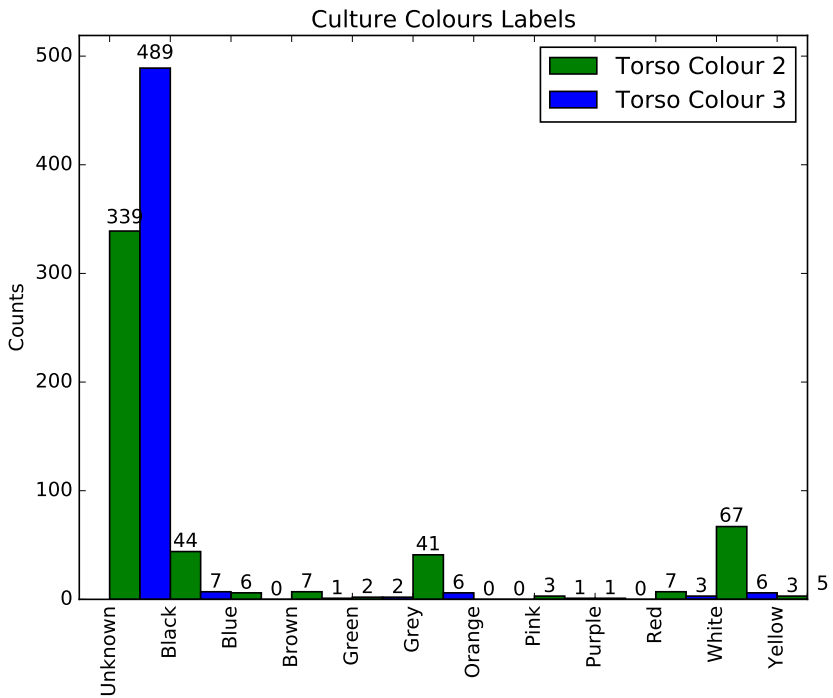
Task 1

- Leg and torso type as either long or short.
- Gender.
- Luggage (carrying or not).
- Pose (front, back, 45°, or 90°).

Task 1

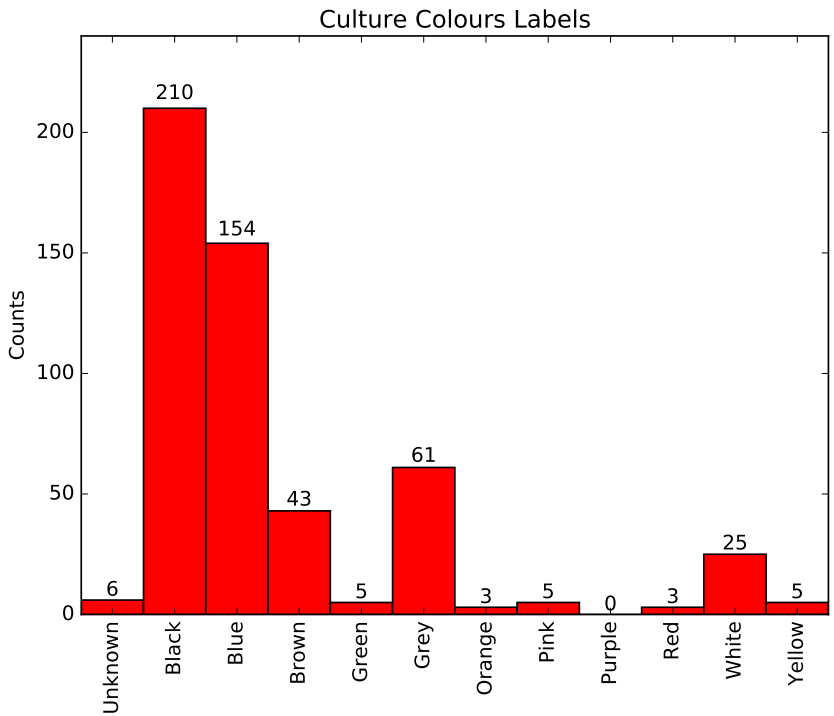
Distribution of traits (training set).

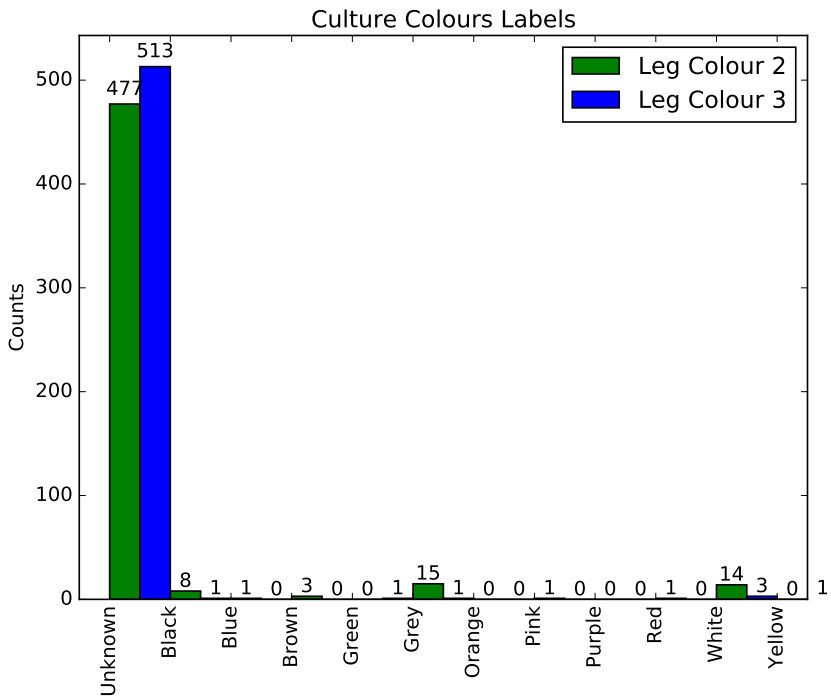




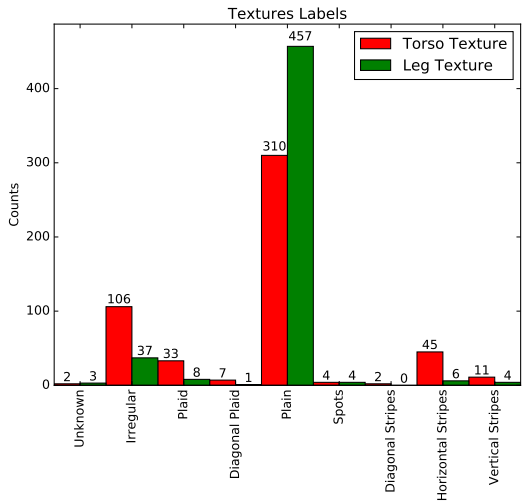
Task 1

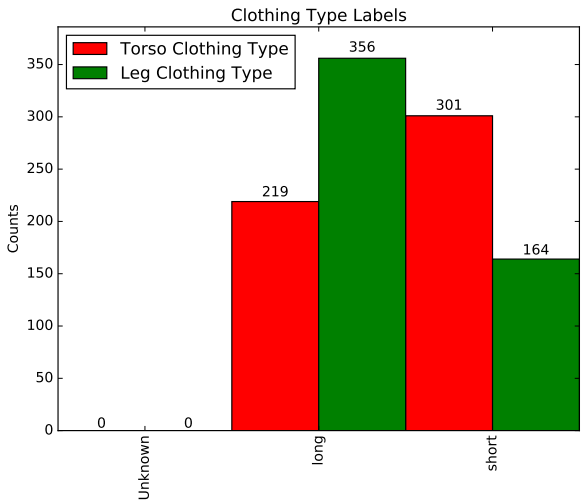
Distribution of traits (training set).

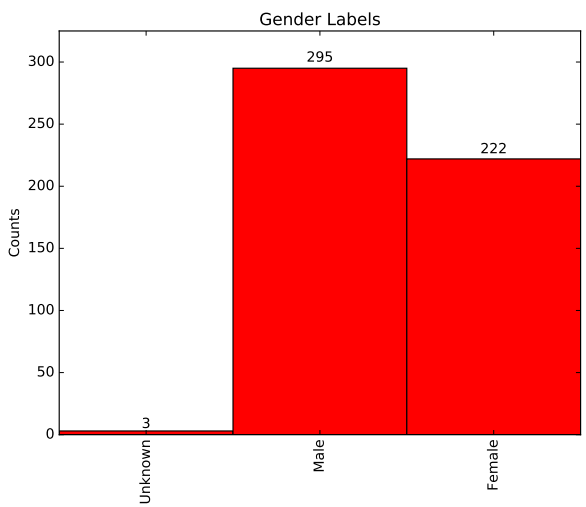


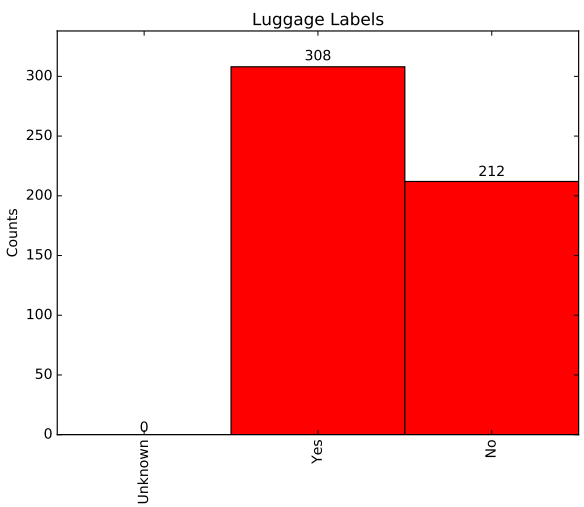


Task 1


Textures Labels
Torso Texture
Leg Texture
Counts
0
100
200
300
400
2
3
106
37
33
8
7
1
310
457
4
4
2
0
45
6
11
4
Unknown
Irregular
Plaid
Diagonal Plaid
Plain
Spots
Diagonal Stripes
Horizontal Stripes
Vertical Stripes

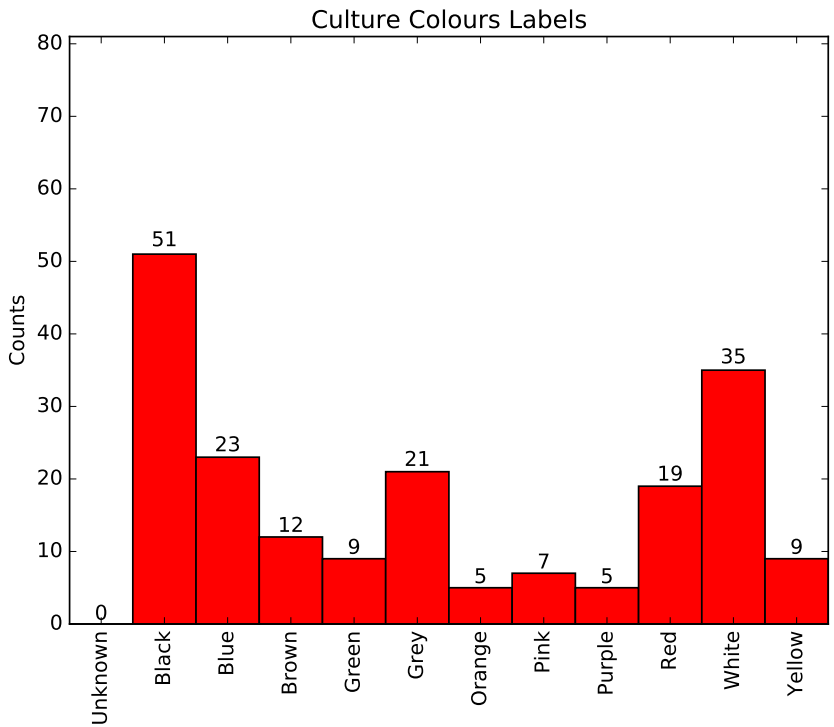

Clothing Type Labels
Torso Clothing Type
Leg Clothing Type
Counts
0
50
100
150
200
250
300
350
0
0
219
356
301
164
Unknown
long
short

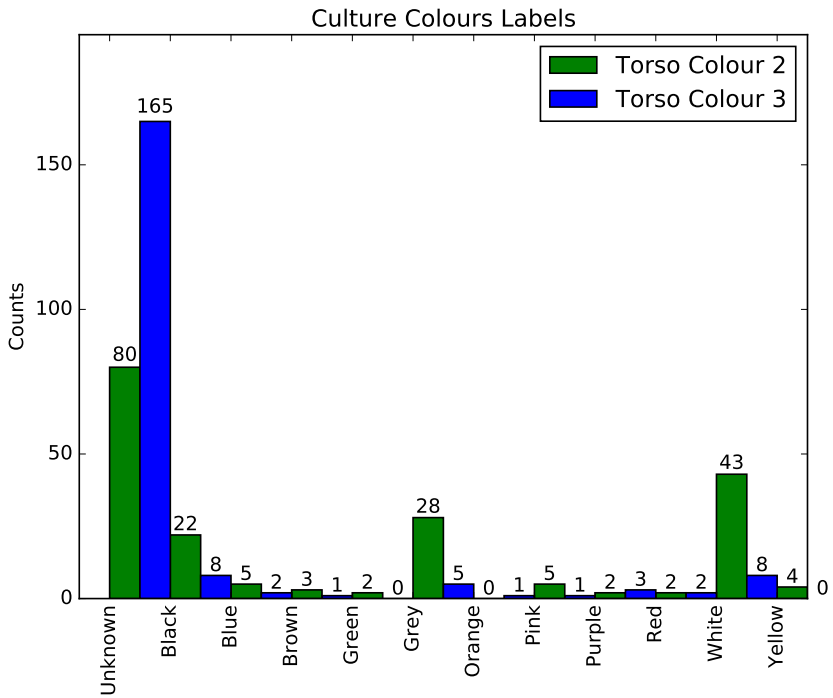

Gender Labels
Counts
0
50
100
150
200
250
300
3
295
222
Unknown
Male
Female


Luggage Labels
Counts
0
50
100
150
200
250
300
0
308
212
Unknown
Yes
No

Task 1

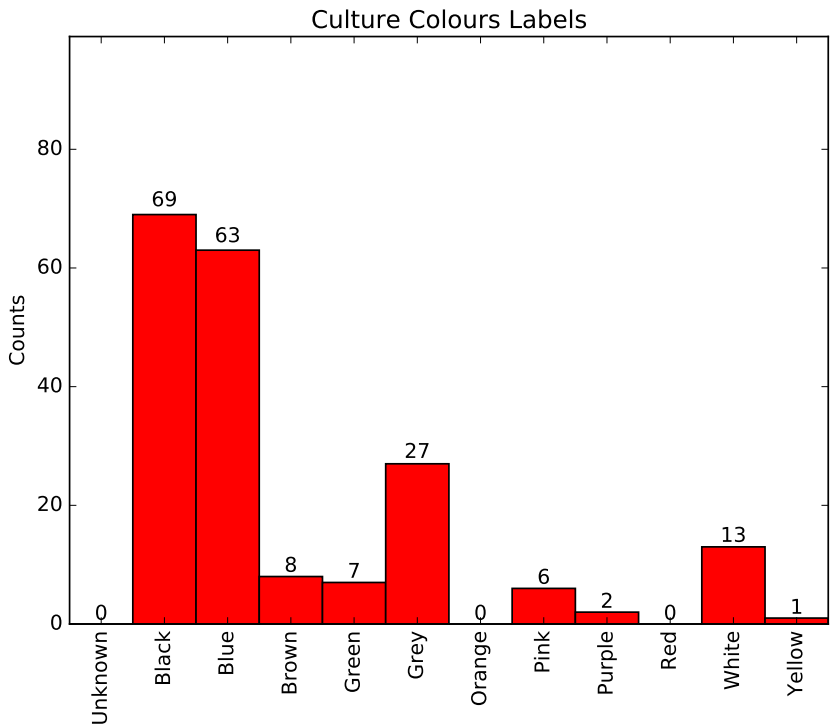
Distribution of traits (testing set).

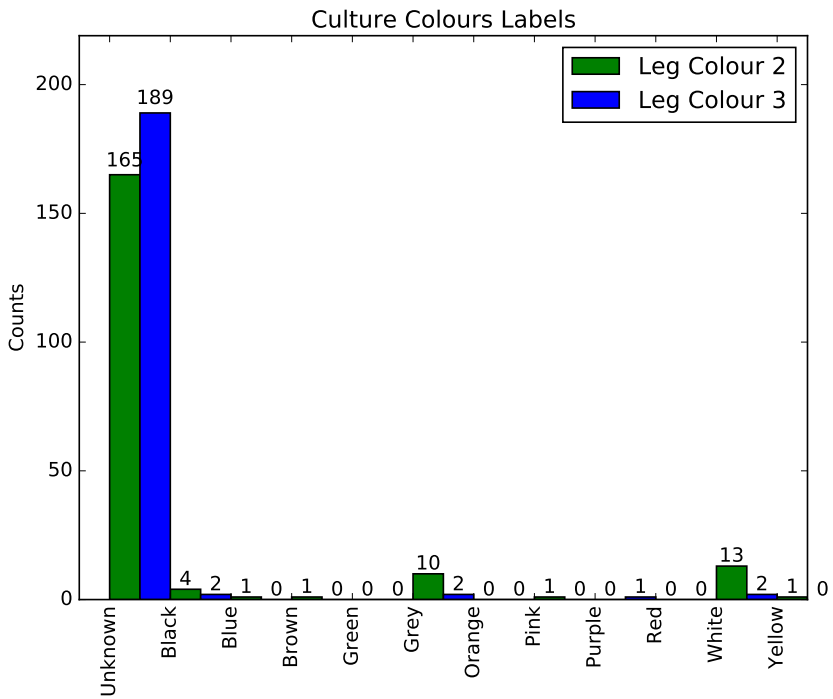




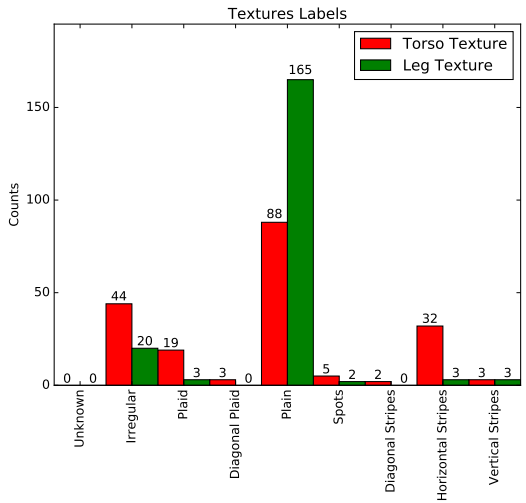
Task 1

Distribution of traits (testing set).

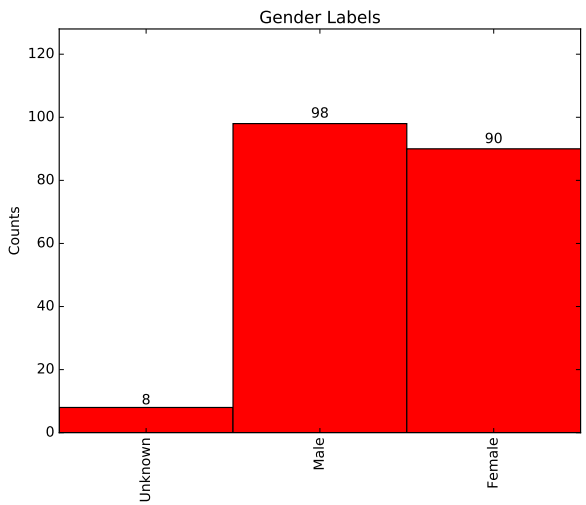


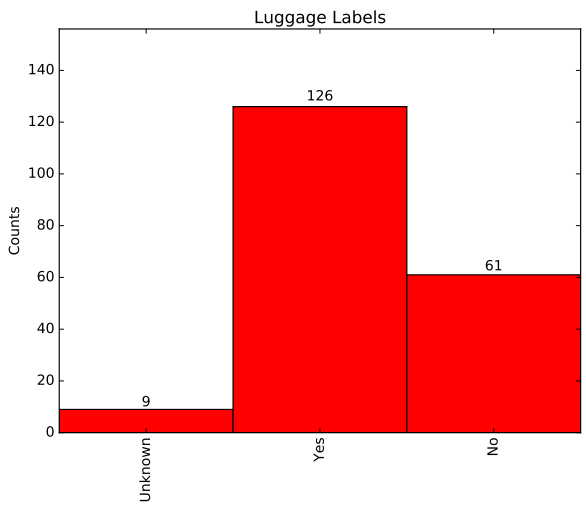


Task 1


Textures Labels
Torso Texture
Leg Texture
Counts
0
50
100
150
0
0
44
20
19
3
3
0
88
165
5
2
2
0
32
3
3
3
Unknown
Irregular
Plaid
Diagonal Plaid
Plain
Spots
Diagonal Stripes
Horizontal Stripes
Vertical Stripes


Clothing Type Labels
Torso Clothing Type
Leg Clothing Type
Counts
0
20
40
60
80
100
120
140
0
0
75
114
121
82
Unknown
long
short


Gender Labels
Counts
0
20
40
60
80
100
120
8
98
90
Unknown
Male
Female


Luggage Labels
Counts
0
20
40
60
80
100
120
140
9
126
61
Unknown
Yes
No

Task 1

Evaluation metrics:

- Evaluated similar to re-identification techniques [8].
- Uses cumulative match characteristic (CMC) curves and performance evaluated at:
  - Rank 1
  - Rank 5
  - Rank 10
  - Rank 25

Task 2

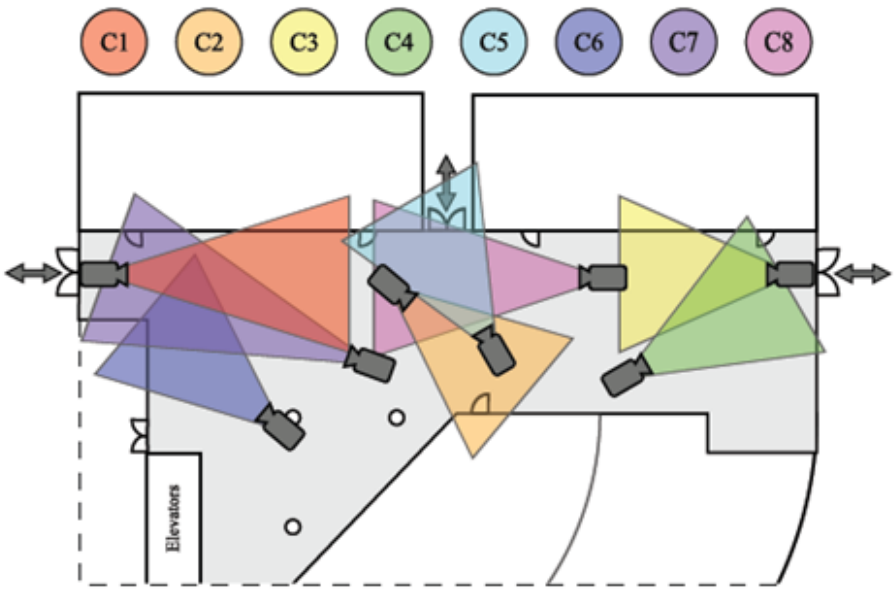
How is this different from Task 1?

- More inline with a surveillance situation.
- The query is used to search a video for the desired subject.
- Multiple potential pedestrians in a scene.
- Varying levels of crowd density, crowd flow, occlusions, and illumination.

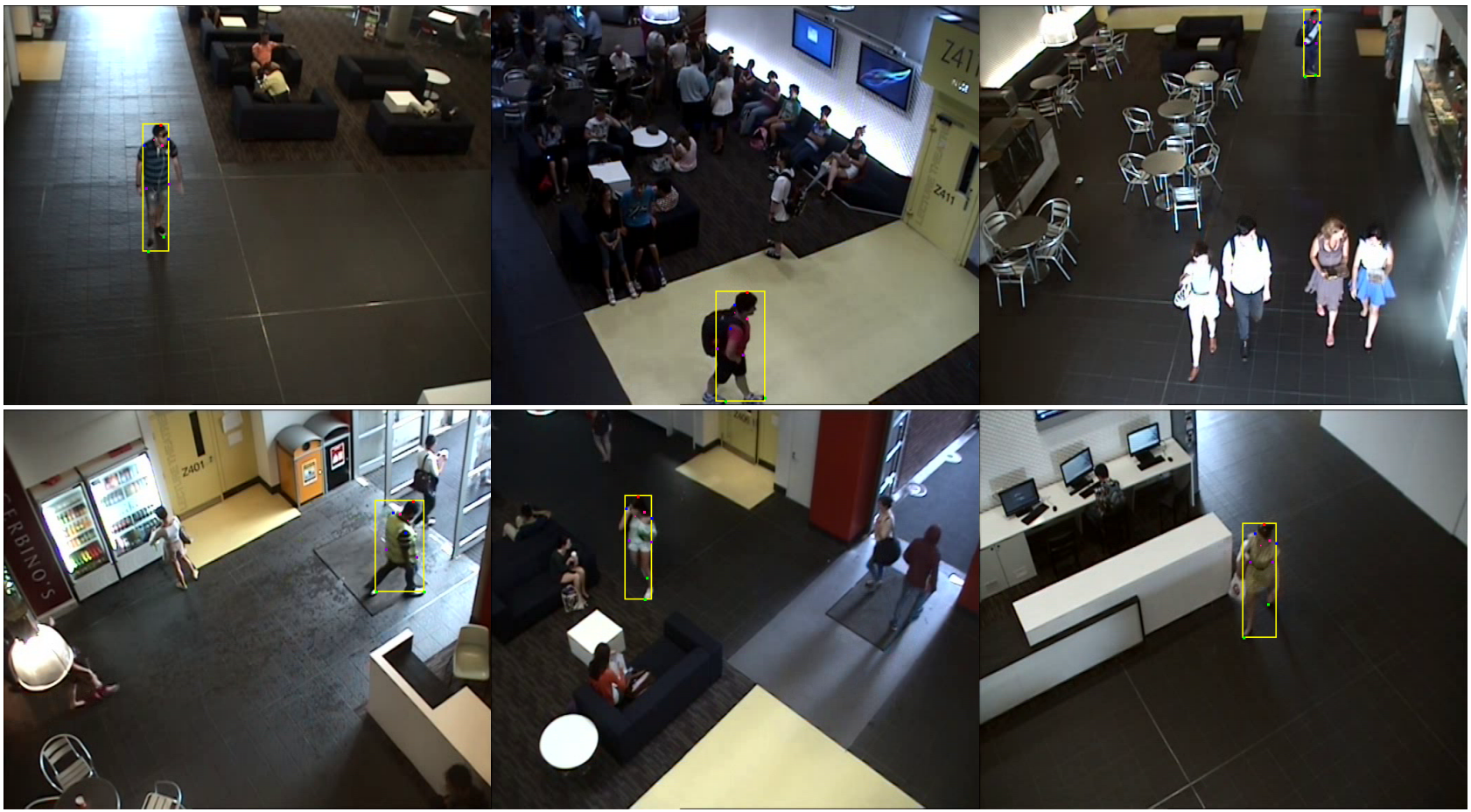
Task 2

Where did the database come from?

- Captured at university setting in one location - 6 cameras in total.
- Training set contains 110 short video segments taken from [5].
- Evaluation set contains 41 segments captured from 4 of the 6 cameras.

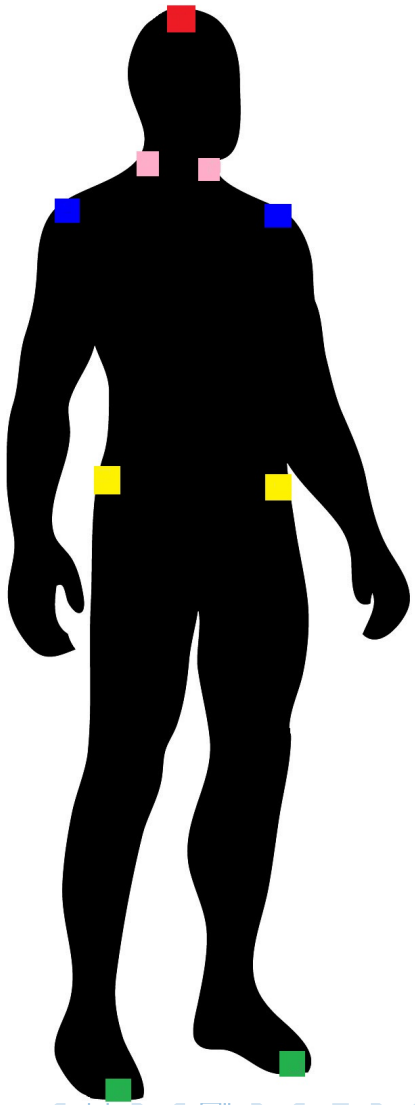


Task 2



Task 2

- Each camera is callibrated with Tsai's callibration scheme [11].
- First 30 frames left un-annotated regardless of subject presence (to initialise any models).
- Annotated by a single person with checks when ambiguous.
- Annotations include soft biometric attributes and key body locations.
- Ambiguity results in a "-1" in annotation.



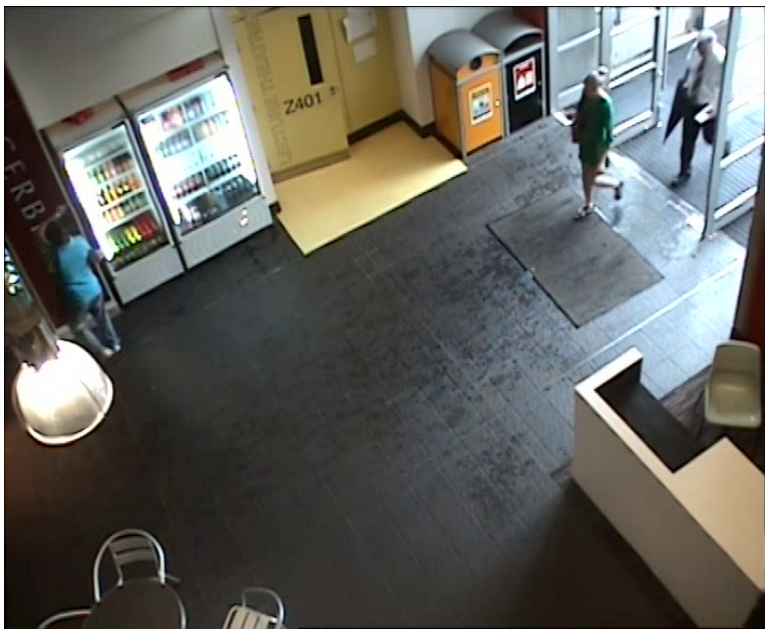
Task 2

Annotations vary slightly from Task 1.

- Colours again based on "Culture Colours", two colours per article (rather than three).
- Texture uses:
  - Plain, check, diagonal stripe, vertical stripe, horizontal stripe, spots, pictures.
- Clothing type:
  - Torso clothing is from long sleeve, short sleeve, no sleeve;
  - Leg clothing is from long pants, dress, skirt, long shorts, short shorts.

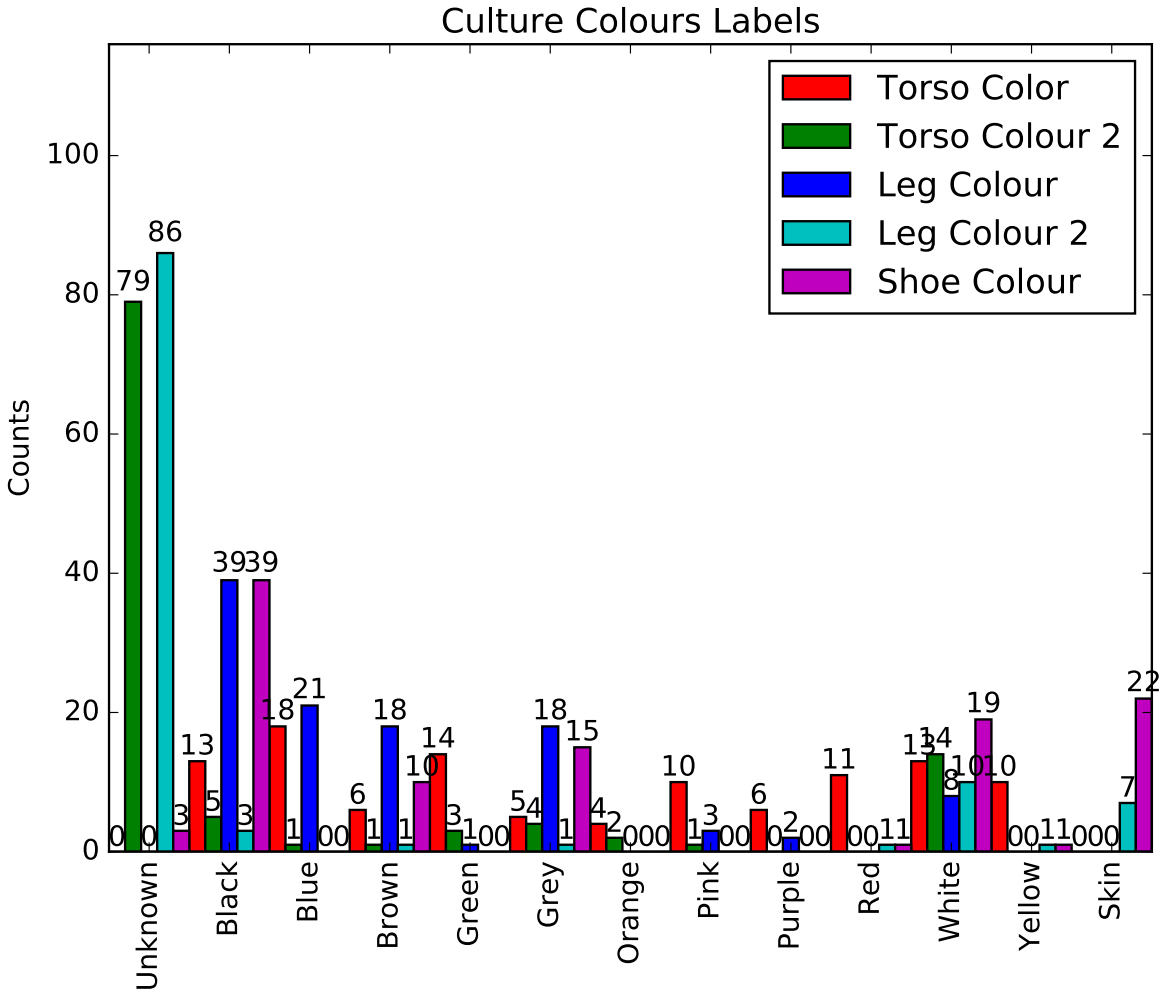
Task 2

- Also includes annotations for: age, gender, height, build, skin, hair.

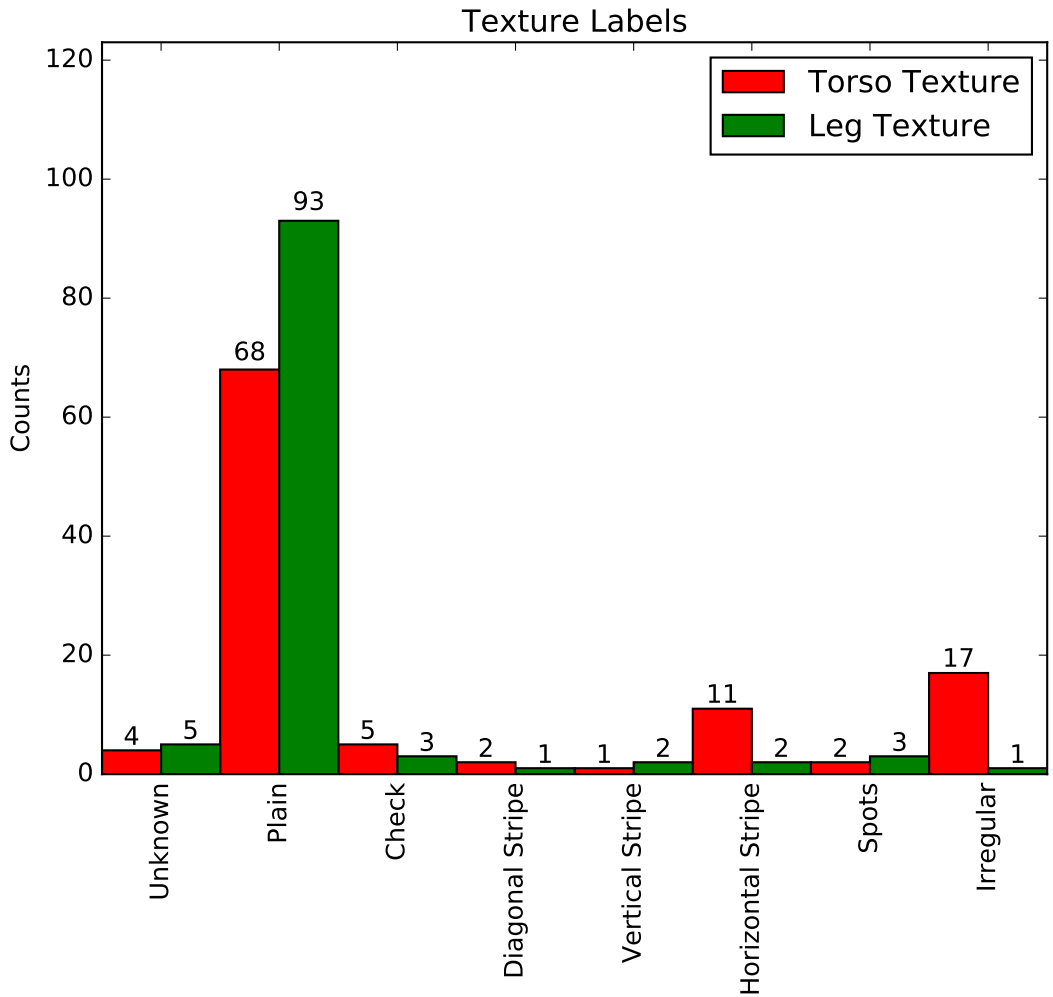


Task 2

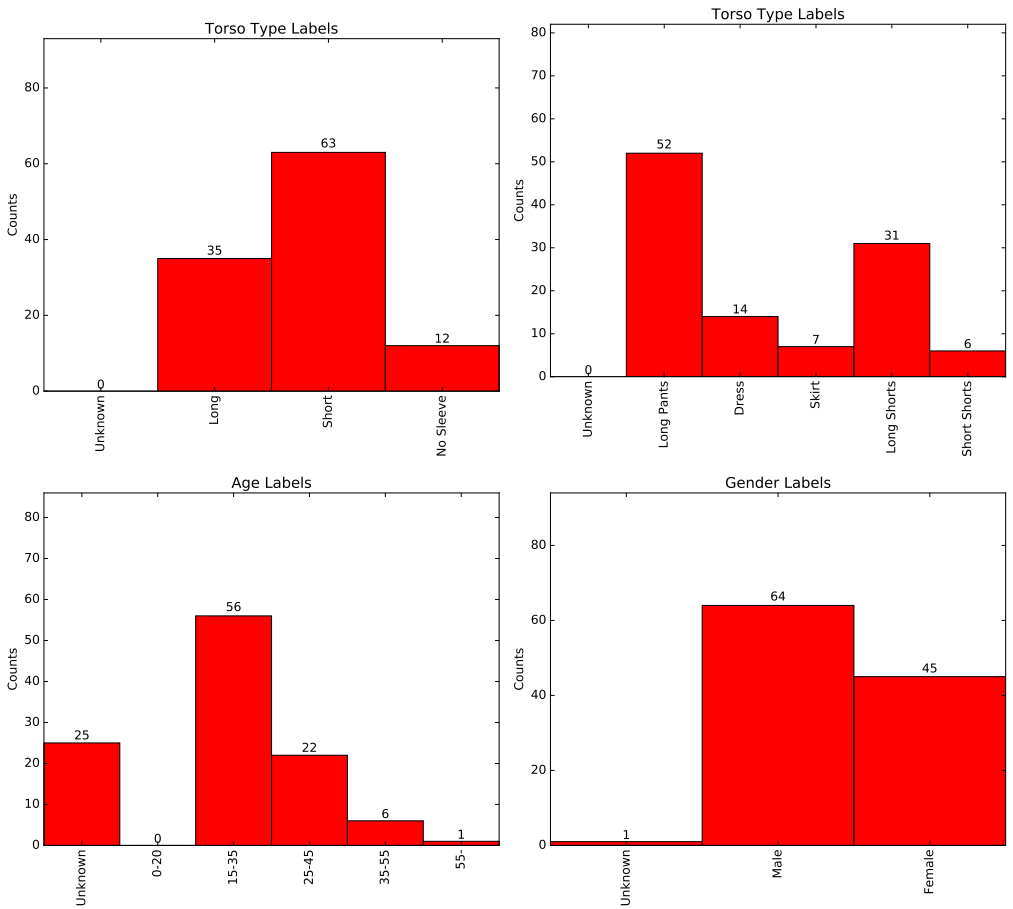
Distribution of traits (training set).



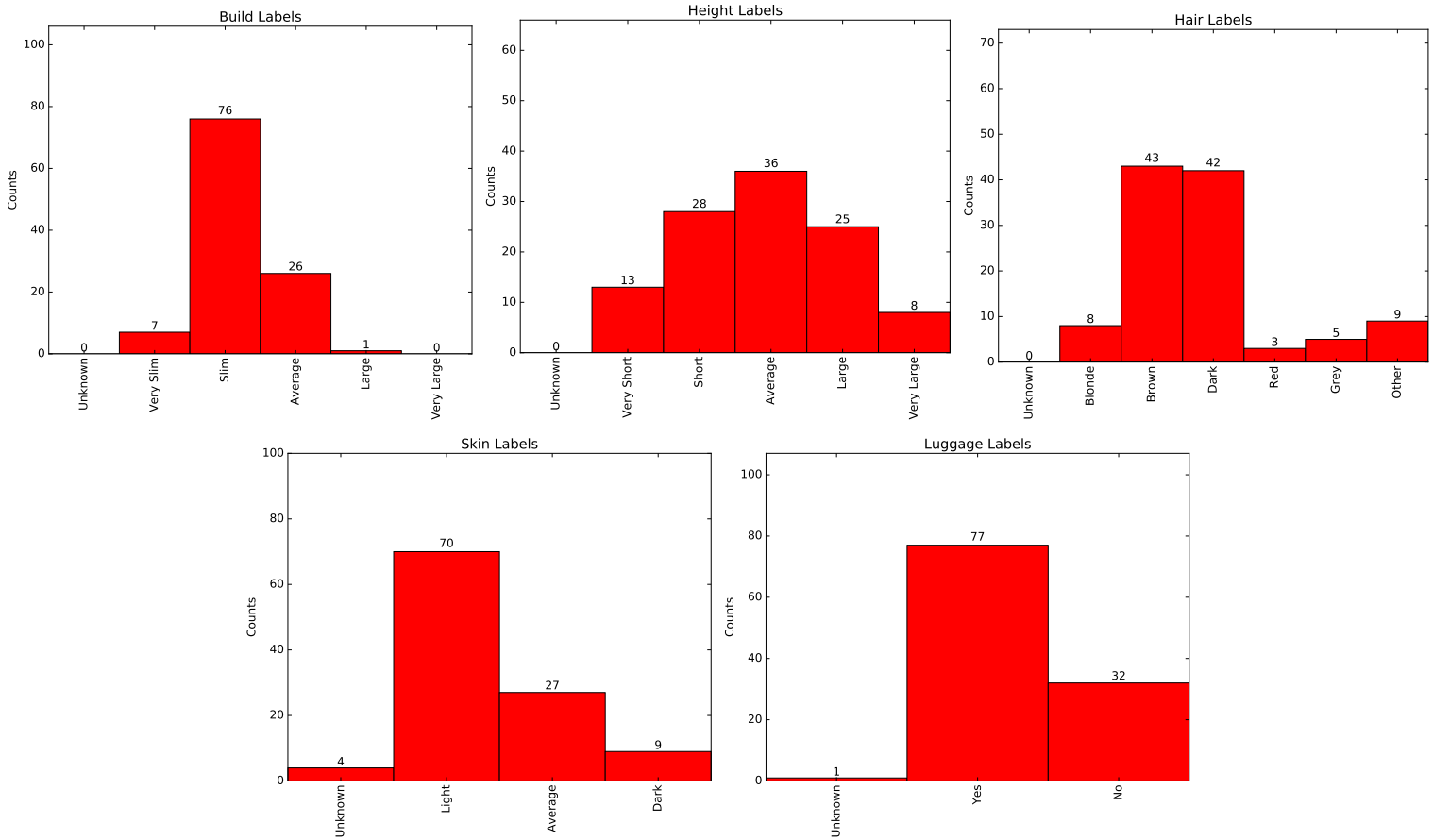
Task 2


Texture Labels
Torso Texture
Leg Texture
Counts
120
100
80
60
40
20
0
4
5
68
93
5
3
2
1
1
2
11
2
2
3
17
1
Unknown
Plain
Check
Diagonal Stripe
Vertical Stripe
Horizontal Stripe
Spots
Irregular

Task 2

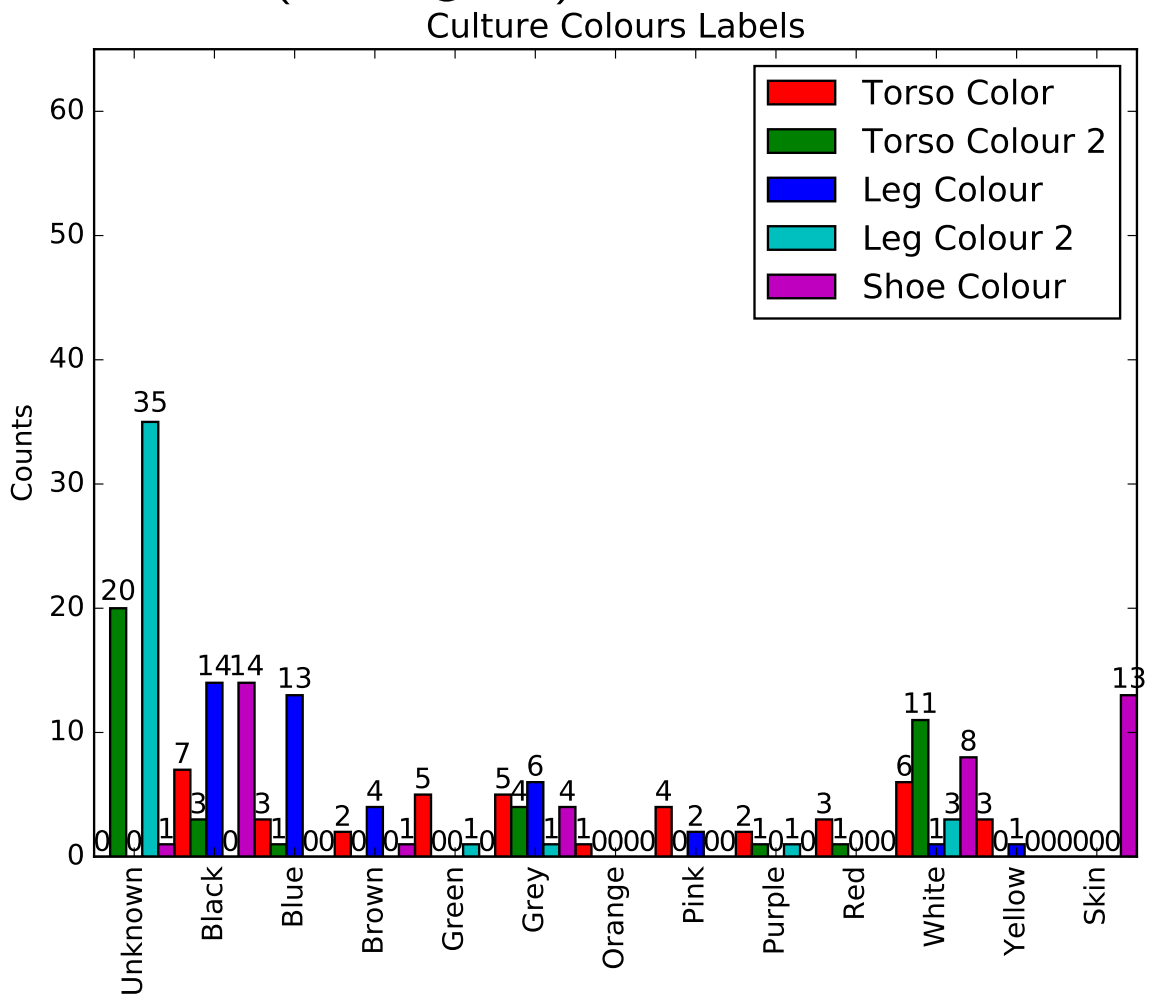

Torso Type Labels
Counts
0
35
63
12
Unknown
Long
Short
No Sleeve
Torso Type Labels
Counts
0
52
14
7
31
6
Unknown
Long Pants
Dress
Skirt
Long Shorts
Short Shorts
Age Labels
Counts
25
0
56
22
6
1
Unknown
0-20
15-35
25-45
35-55
55-
Gender Labels
Counts
1
64
45
Unknown
Male
Female

Task 2

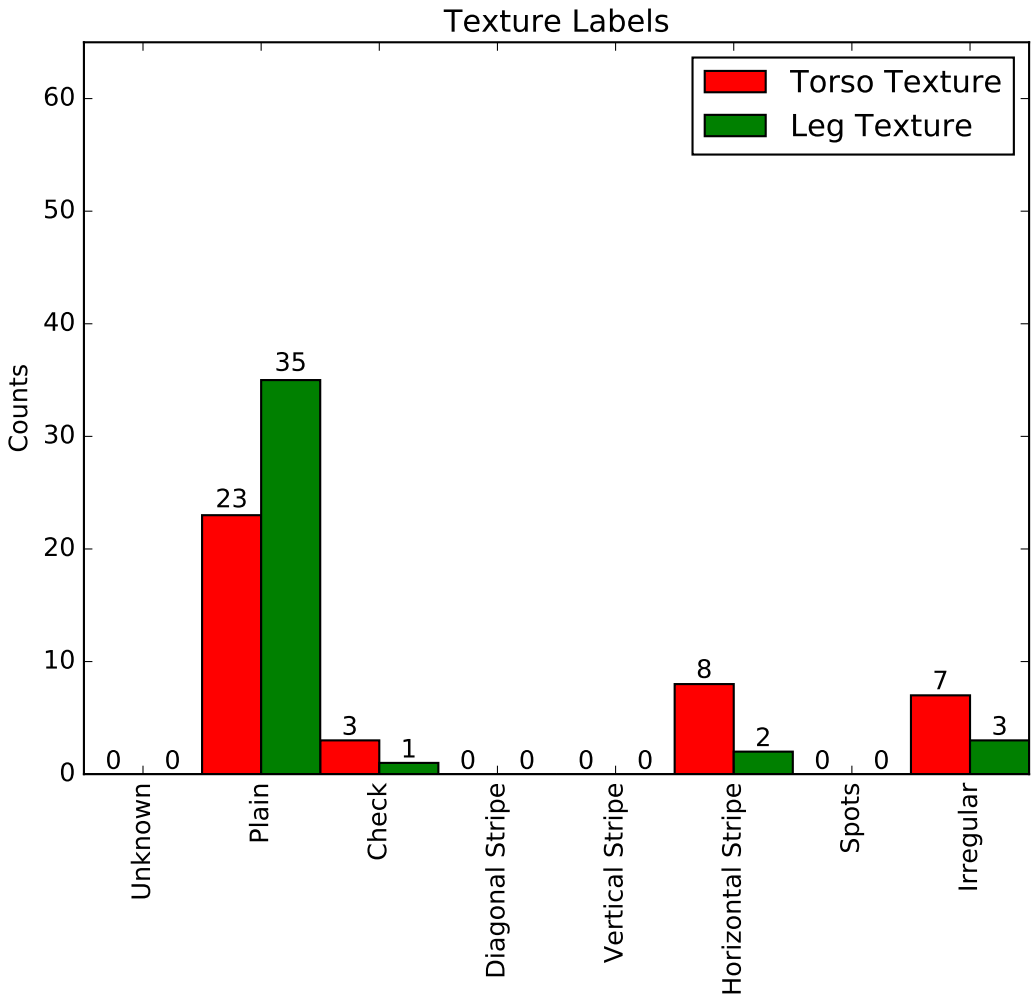

Build Labels
Counts
Unknown
Very Slim
Slim
Average
Large
Very Large
0
7
76
26
1
0
Height Labels
Counts
Unknown
Very Short
Short
Average
Large
Very Large
0
13
28
36
25
8
Hair Labels
Counts
Unknown
Blonde
Brown
Dark
Red
Grey
Other
0
8
43
42
3
5
9
Skin Labels
Counts
Unknown
Light
Average
Dark
4
70
27
9
Luggage Labels
Counts
Unknown
Yes
No
1
77
32

Task 2

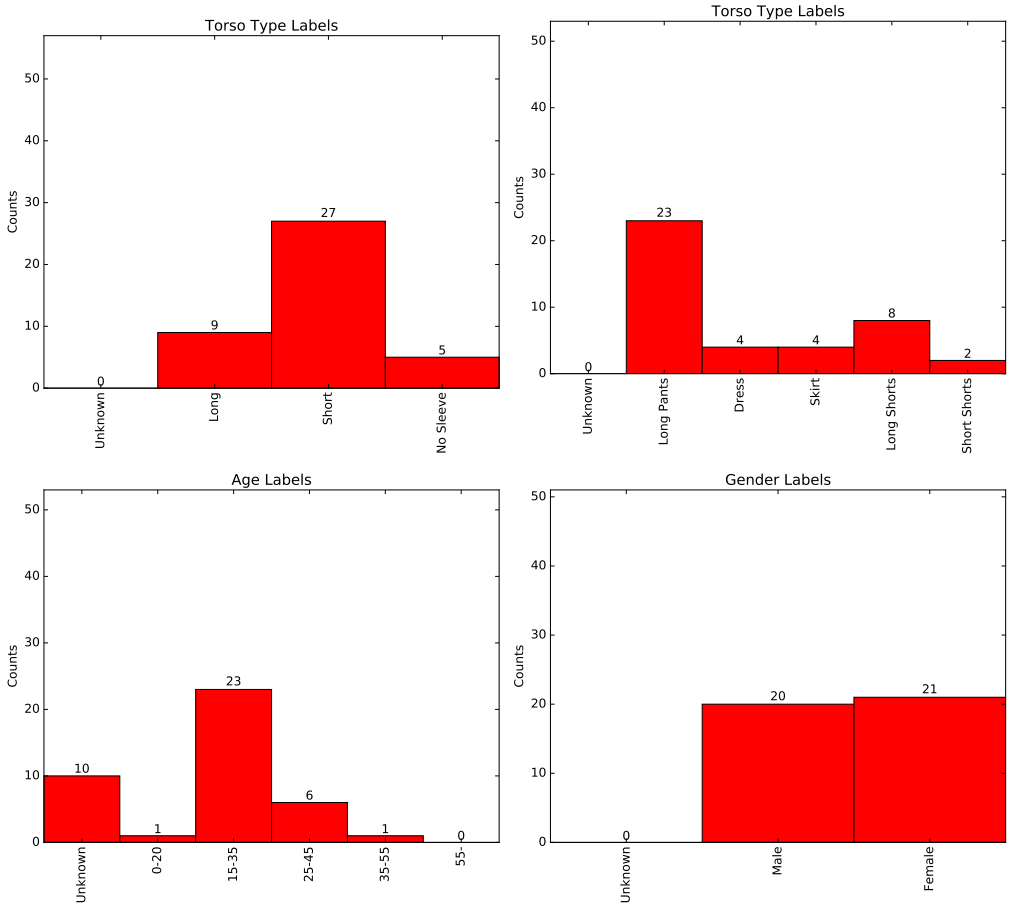
Distribution of traits (testing set)



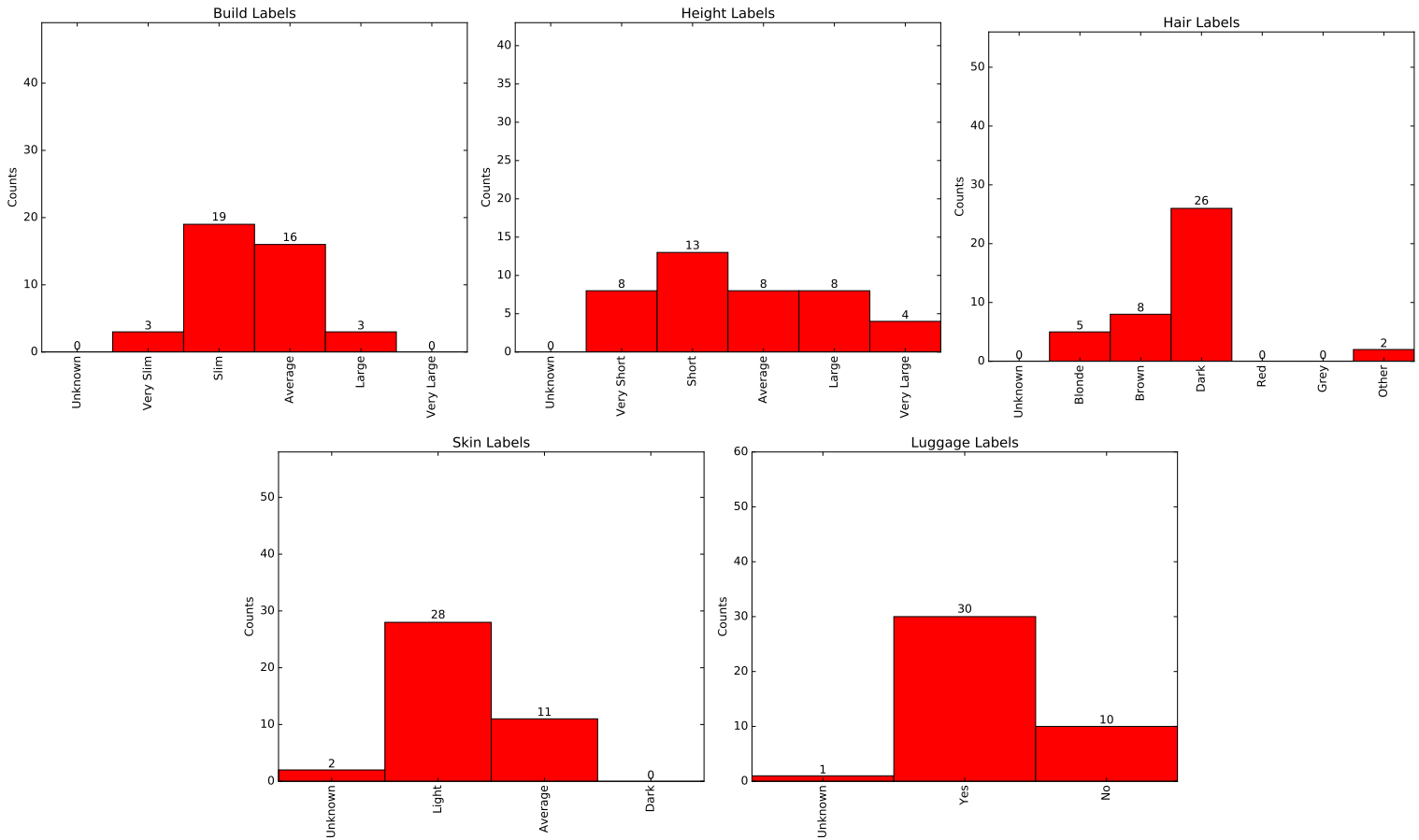
Task 2


Texture Labels
Torso Texture
Leg Texture
Counts
0
10
20
30
40
50
60
0 0
23 35
3 1
0 0
0 0
8 2
0 0
7 3
Unknown
Plain
Check
Diagonal Stripe
Vertical Stripe
Horizontal Stripe
Spots
Irregular

Task 2


Torso Type Labels
Counts
0
9
27
5
Unknown
Long
Short
No Sleeve
Torso Type Labels
Counts
0
23
4
4
8
2
Unknown
Long Pants
Dress
Skirt
Long Shorts
Short Shorts
Age Labels
Counts
10
1
23
6
1
0
Unknown
0-20
15-35
25-45
35-55
55-
Gender Labels
Counts
0
20
21
Unknown
Male
Female

Task 2


Build Labels
Counts
Unknown
Very Slim
Slim
Average
Large
Very Large
0
3
19
16
3
0
Height Labels
Counts
Unknown
Very Short
Short
Average
Large
Very Large
0
8
13
8
8
4
Hair Labels
Counts
Unknown
Blonde
Brown
Dark
Red
Grey
Other
0
5
8
26
0
0
2
Skin Labels
Counts
Unknown
Light
Average
Dark
2
28
11
0
Luggage Labels
Counts
Unknown
Yes
No
1
30
10

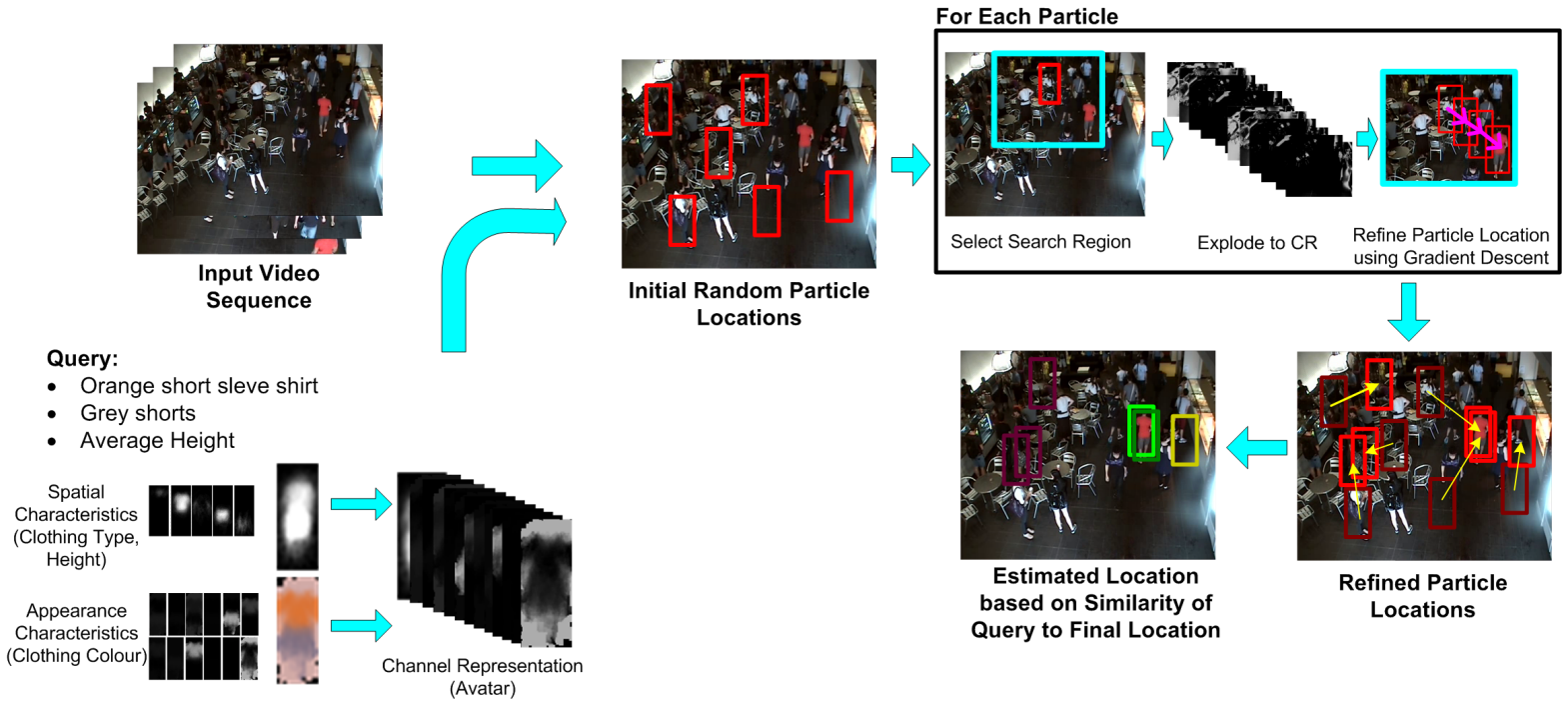
Task 2

Baseline System [3]:

- Creates an avatar as a channel representation, and uses a particle filter driven search to locate the subject
- Prone to errors in high clutter
- Avatar very simplistic: only colour and broad (long/short) clothing types are considered

Task 2

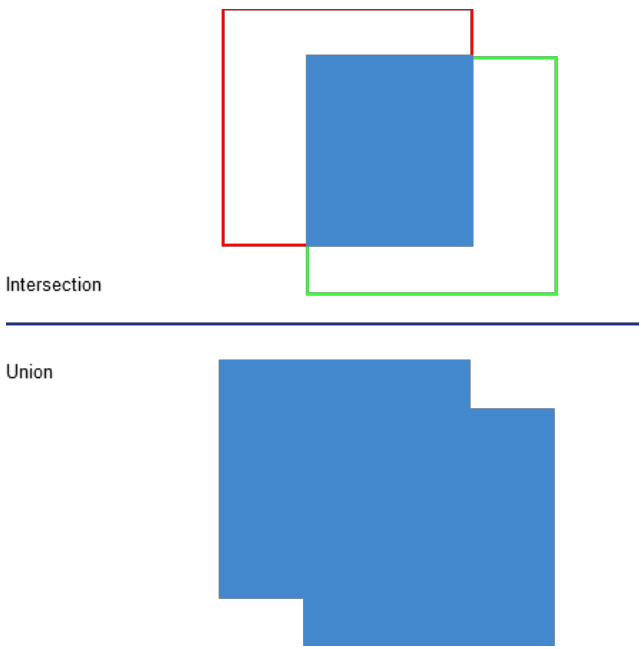
Baseline System [3]:



Task 2

Metric used for evaluation:

- Intersection over Union of detection and ground truth bounding boxes, as per [5].
- The average IoU for a sequence is calculated.
- Final score is the average of all sequences.



Final Remarks

- Two semantic search problems.
- One relating to re-identification
- One more in-line with surveillance type searches.
- Two publicly available datasets each with a training and testing set.
- Semantic queries for both databases and sets available.

Bibliography I

B. Berlin and P. Kay.

Basic color terms: Their universality and evolution.
Univ of California Press, 1991.

L. Bourdev and J. Malik.

Poslets: Body part detectors trained using 3d human pose annotations.
In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1365–1372, 2009.

S. Denman, M. Halstead, C. Fookes, and S. Sridharan.

Searching for people using semantic soft biometric descriptions.
Pattern Recognition Letters, 68:306–315, 2015.

D. Gray, S. Brennan, and H. Tao.

Evaluating appearance models for recognition, reacquisition, and tracking.
In *Proceedings of the IEEE International workshop on Performance Evaluation of Tracking and Surveillance*, 2007.

M. Halstead, S. Denman, S. Sridharan, and C. B. Fookes.

Locating people in video from semantic descriptions: A new database and approach.
In *Proceedings of the 22nd International Conference on Pattern Recognition*, pages 4501–4506. IEEE, 2014.

E. S. Jaha and M. S. Nixon.

Soft biometrics for subject identification using clothing attributes.
In *Proceedings of the IEEE International Joint Conference on Biometrics*, pages 1–6, 2014.

Bibliography II

E. S. Jaha and M. S. Nixon.

From clothing to identity: Manual and automatic soft biometrics.
IEEE Transactions on Information Forensics and Security, 11(10):2377–2390, 2016.

A. Khatun, S. Denman, S. Sridharan, and C. Fookes.

A deep four-stream siamese convolutional neural network with joint verification and identification loss for person re-detection.
In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1292–1301. IEEE, 2018.

D. Li, Z. Zhang, X. Chen, H. Ling, and K. Huang.

A richly annotated dataset for pedestrian attribute recognition.
arXiv preprint arXiv:1603.07054, 2016.

R. Satta, F. Pala, G. Fumera, and F. Roli.

People search with textual queries about clothing appearance attributes.
In *Person Re-Identification*, pages 371–389. Springer, 2014.

R. Y. Tsai.

An efficient and accurate camera calibration technique for 3d machine vision.
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 364–374, 1986.