University: Sharif University of Technology

Department: Electrical Engineering

Course Name: Advanced Neuroscience

Homework 8 Report

Visual Attention

Student Name: Ali Shahbazi

Student ID: 98101866

Instructor: Prof. Ali Ghazizadeh

Academic Semester: 2023 Spring

1 Eye Tracking Database

- ***showEyeData():*** I have selected 36 images out of 1003, as my stimuli. Selected images are located in `'./SELSTIMULI'` folder. Eye data of these images are located in `'./SELDATA/tu'` folder for subject *tu*. Figure 1 shows 9 samples of selected images with eye data plotted on it. Most probable saccades are into face and body components.

(a) (b) (c)

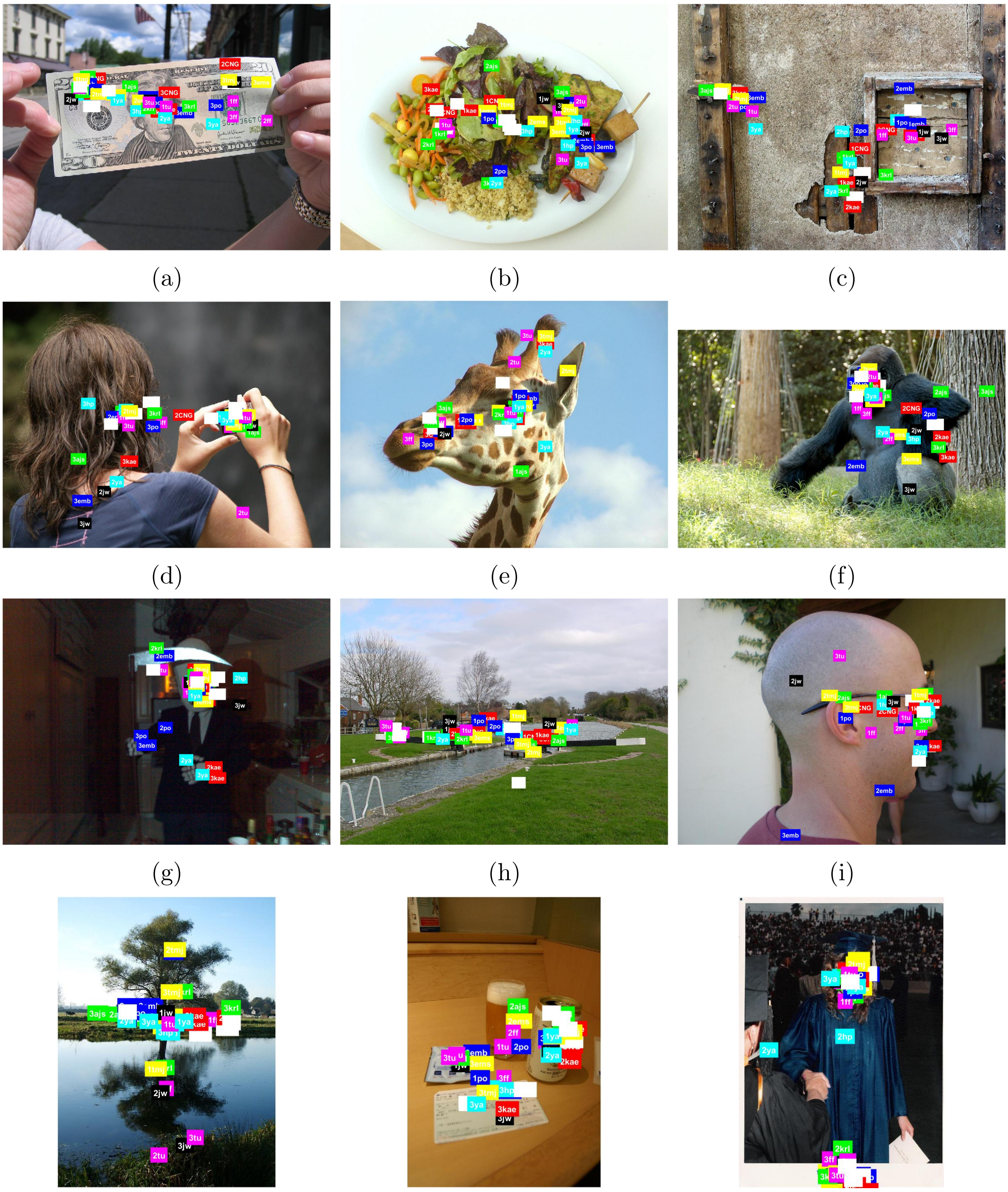
(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 1: Eye Data of Some Selected Images

- ***showEyeDataAcrossUsers():*** Figure 2 shows previously samples of selected images with saccade positions plotted on it. Concentration of saccades shows points of interest.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure 2: Saccade Positions of Some Selected Images

2 Saliency Model

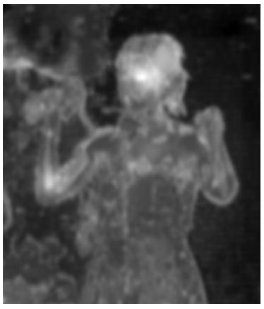
Using `saliency.m` function and commenting unwanted features, we extract saliency model for desired combination of features.

According to the Judd et al. 2009, features that are used in this simulation, can be categorized as:

- **Low-level features:**
  - Subband: Are features related to a band of frequency in the image. Dividing image into several frequency bands and process each band individually.
  - Itti: As proposed in Itti et al. 2000, intensity, orientation, and color contrast are important bottom-up features calculated here.
  - Color: Values of red, green, and blue channels and probabilities of each channel as features.
  - Torralba: This feature is calculated based on subband pyramids.
- **Mid-level features:**
  - Horizon: If objects are placed on the surface of the earth, this feature acts as an horizon line detector.
- **High-level features:**
  - Object: Humans tend to focus on objects and faces and people more. Here these categories are recognized.

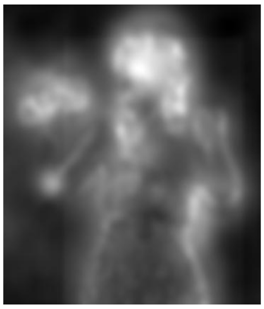
    Note that because of problems in `faceDetec.m`, face recognition is commented.
- **Center prior:**
  - Distance from Center: Important points of interest are more likely to be found in center of the image.

(a) Sample Image 1

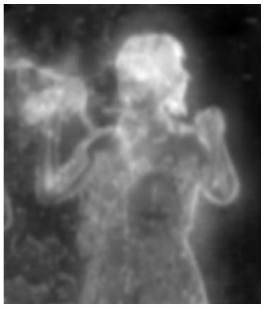


(b) All - Subband

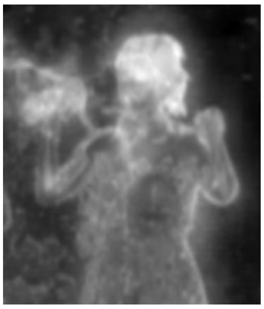
(c) All - Itii



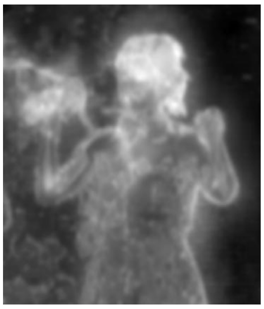
(d) All - Color



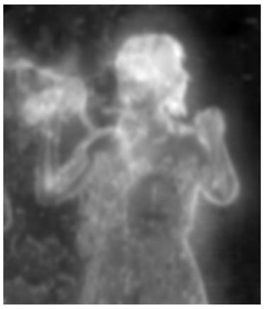
(e) All - Torralba



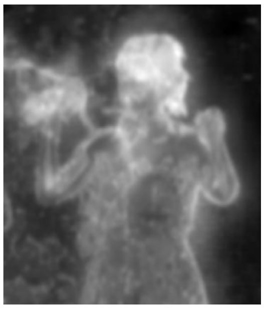
(f) All - Horizon



(g) All - Object

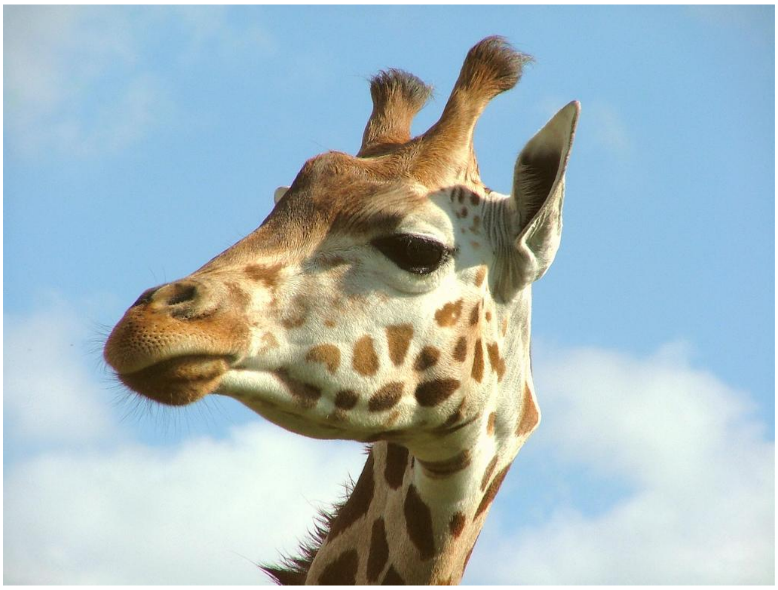


(h) All - Center

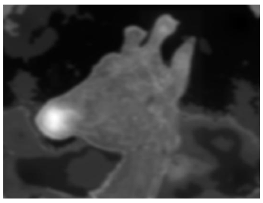


(i) All Features

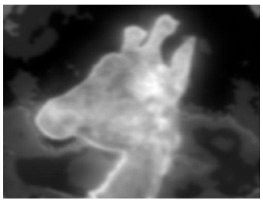
Figure 3: Sample Image and 8 Saliency Models; (b)-(h) lack one the features while (i) has all features



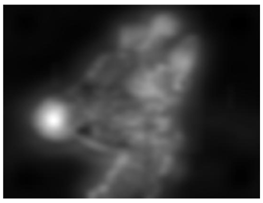
(a) Sample Image 2



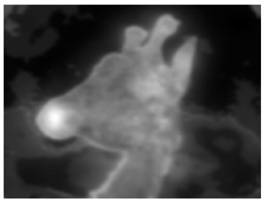
(b) All - Subband



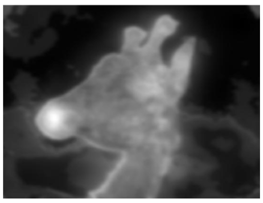
(c) All - Itii



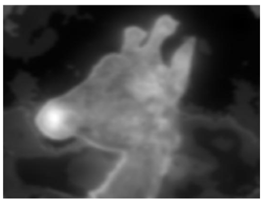
(d) All - Color



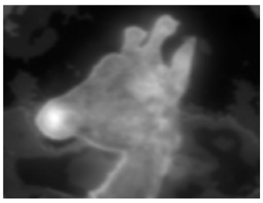
(e) All - Torralba



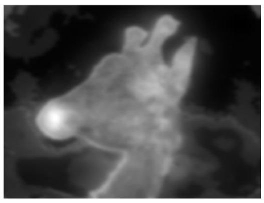
(f) All - Horizon



(g) All - Object



(h) All - Center



(i) All Features

Figure 4: Sample Image and 8 Saliency Models; (b)-(h) lack one the features while (i) has all features

3 Compare Saliency Maps to Fixations

Computing saliency maps for a large number of stimuli and subjects requires extremely huge amount of time. Therefore, I limited the number of stimuli to 120 images and tended to run for 5 subjects without considering *object feature* which was time consuming.

We can plot mean of AUCs for each feature as a measure of performance. According to Figure 5, performance of all features are around 0.6 to 0.7 without a significant difference except the *Distance from Center* feature which has a very low performance. Overall, the performance of all features together are quite better although larger amount of data is needed to argue.

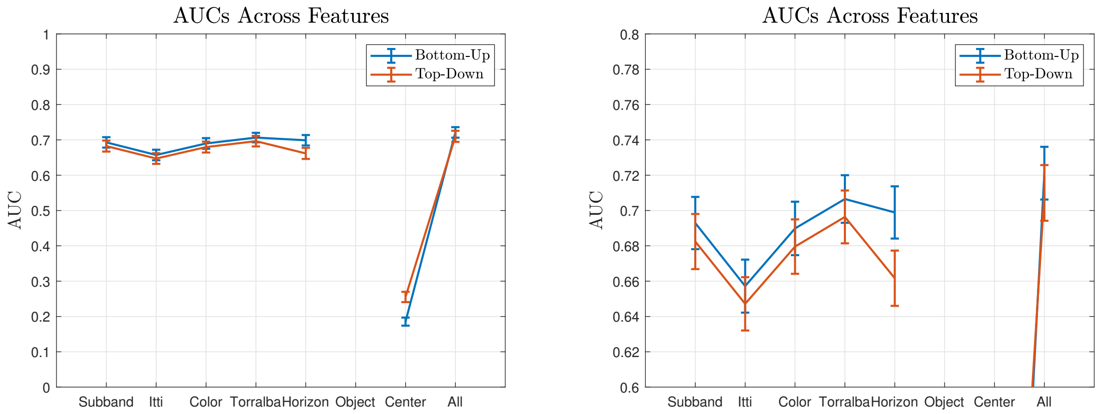
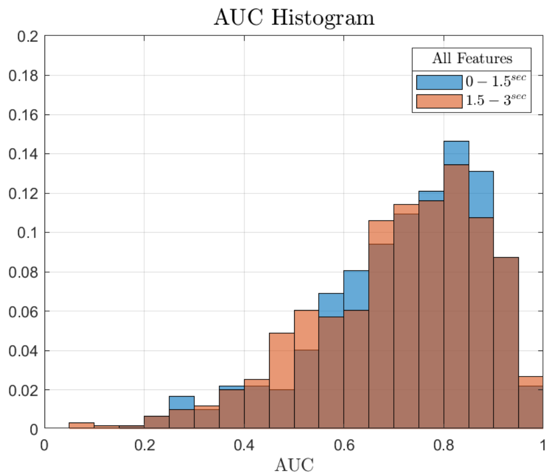


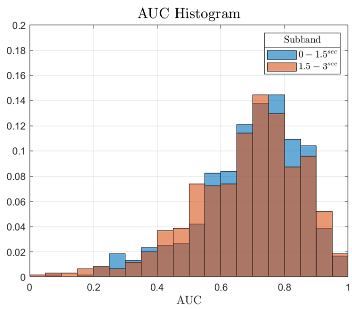
Figure 5: Performance of each Feature; calculated for 5 subjects and 120 stimuli (*object* feature was not included in *all features* mean AUC)

Also, With the AUCs calculated, we can investigate the idea of categorizing features as *bottom-up* and *top-down* by splitting the 3^{sec} trial into two periods. Figure 6 shows distribution of AUCs for each feature in the first period and second. If a feature's distribution for the first period $(0 - 1.5^{sec})$, had a high mean while it's distribution for the second period $(1.5 - 3^{sec})$ was not that good, we can conclude that this feature is related to bottom-up attention. This conclusion is valid for the vice versa condition where a feature can be related to top-down attention.

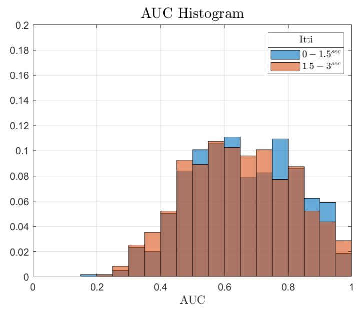
Apparently, non of the features calculated in Figure 6 are sensitive to periods. Although a small difference can be observed in Figure 6g.



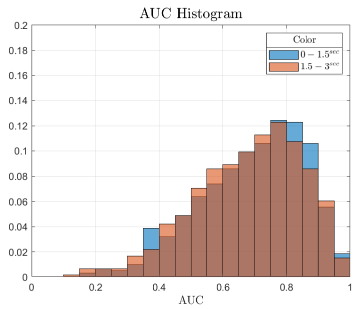
(a) All Features



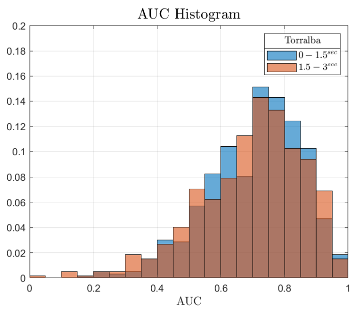
(b) Subband



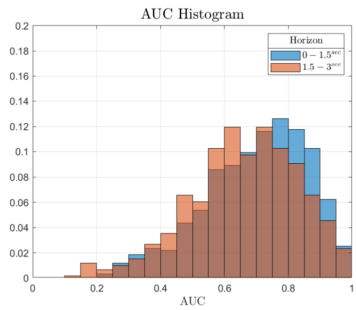
(c) Itti



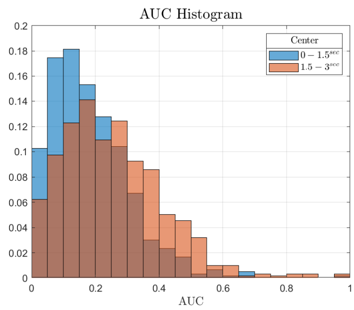
(d) Color



(e) Torralba



(f) Horizon



(g) Distance from Center

Figure 6: Histogram of AUCs for Different Periods; calculated for 5 subjects and 120 stimuli (*object* feature was not included in *all features* histogram)

Once again, I simulated with just one subject (*'ajs'*) and same 120 stimuli. This time the *object* feature was calculated both individually and in *all features*. Figure 7 shows the mean AUC for each feature.

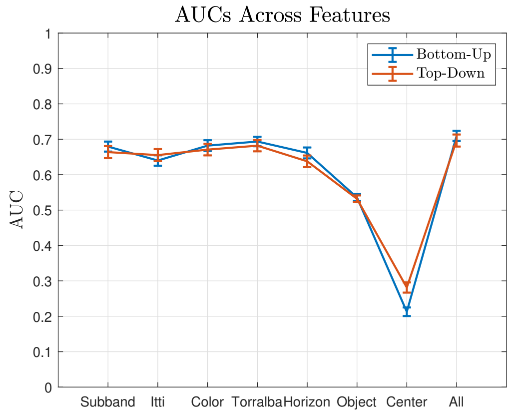
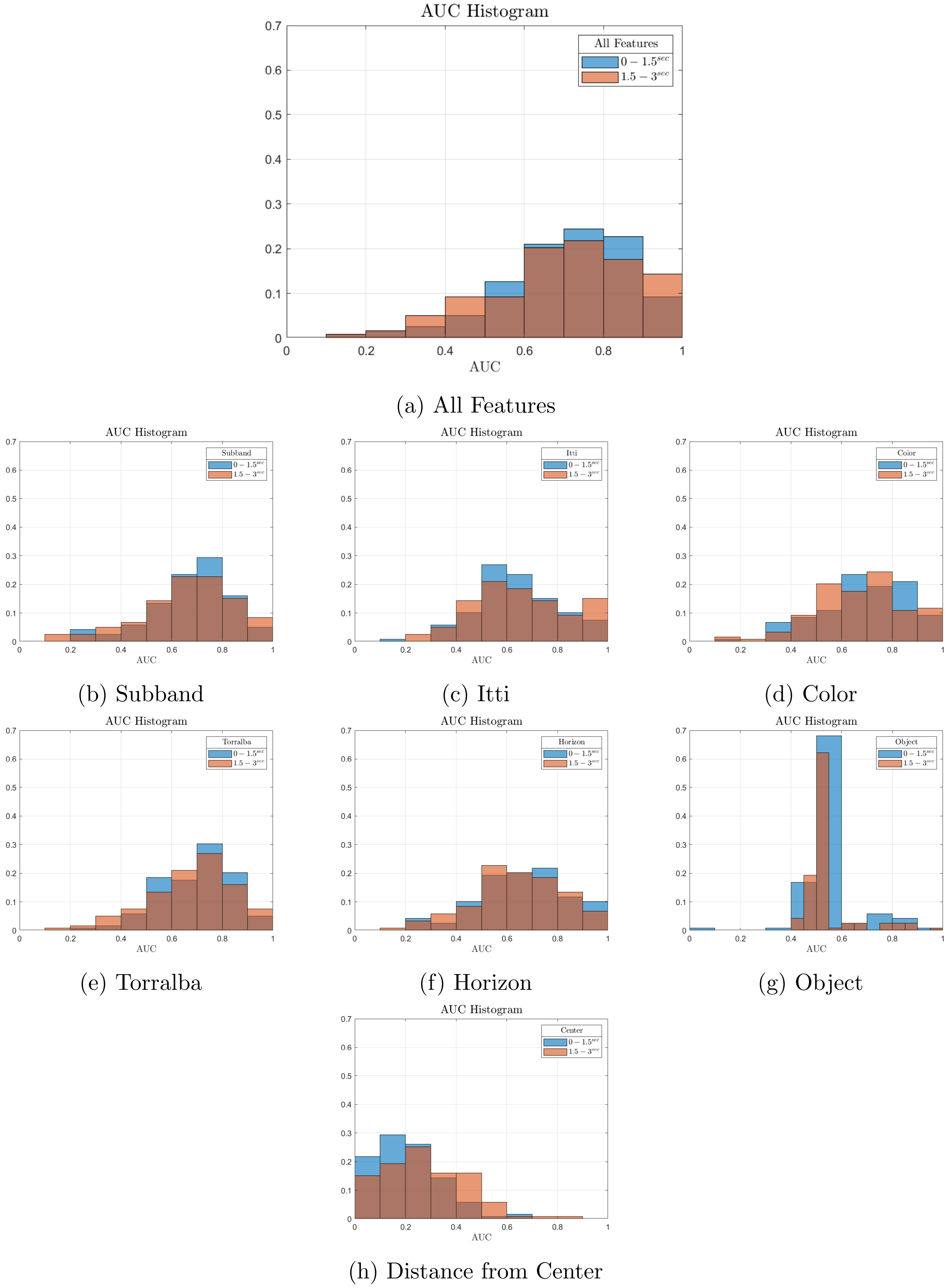


Figure 7: Performance of each Feature; calculated for subject *'ajs'* and 120 stimuli

Again in Figure 8, all previous features remain indifferent for the eye data period. The new feature *Object*, on the other hand, has a very distinct histogram. This abnormal histogram is due to binary function of recognizing object or not.



(a) All Features

(b) Subband

(c) Itti

(d) Color

(e) Torralba

(f) Horizon

(g) Object

(h) Distance from Center

Figure 8: Histogram of AUCs for Different Periods; calculated for the subject *'ajs'* and 120 stimuli

References

Itti, Laurent and Christof Koch (2000). “A saliency-based search mechanism for overt and covert shifts of visual attention”. In: *Vision research* 40.10-12, pp. 1489–1506.

Judd, Tilke, Krista Ehinger, Frédo Durand, and Antonio Torralba (2009). “Learning to predict where humans look”. In: *2009 IEEE 12th international conference on computer vision*. IEEE, pp. 2106–2113.