Comparison of Five Color Models in Skin Pixel Classification

Benjamin D. Zarit, Boaz J. Super, Francis K. H. Quek
Electrical Engineering and Computer Science
University of Illinois at Chicago
bzarit@inso.com, super@eecs.uic.edu, quek@cs.wright.edu

Abstract

Detection of skin in video is an important component of systems for detecting, recognizing, and tracking faces and hands. Different skin detection methods have used different color spaces. This paper presents a comparative evaluation of pixel classification performance of two skin detection methods in five color spaces. The skin detection methods used in this paper are color-histogram based approaches that are intended to work with a wide variety of individuals, lighting conditions, and skin tones. One is the widely-used lookup table method, the other makes use of Bayesian decision theory. Two types of enhancements, based on spatial and texture analyses, are also evaluated.

1. Introduction

Skin detection in video images is an important first step in a wide variety of recognition and tracking systems. It can be used to begin the process of face recognition or facial expression extraction, and can provide an initial estimate or follow-up verification for face and hand tracking algorithms. Being able to do these types of detection based merely on skin color would eliminate the need for cumbersome tracking devices or artificially placed color keys. Imagine a video conference in which the speakers can move freely, while cameras automatically track their positions, always keeping them centered in the frame [16].

One of the primary problems in skin detection is color constancy. Ambient light, bright lights, and shadows change the apparent color of an image. Different cameras affect the color values as well. Movement of an object can cause blurring of colors. Finally, skin tones vary dramatically within and across individuals.

A primary objective of the current work is to study the effect of color space choice on skin detection performance. Two color histogram based methods are evaluated on five color spaces; one is the widely-used lookup table method, the other makes use of Bayesian decision theory. The effect of using different color histogram resolutions and of enhancements based on region growing and texture analysis are also evaluated. The skin detection methods use off-line training rather than incremental run-time training, and are tested on images representing a wide variety of people, environments, cameras, and lighting conditions.

2. Background

A number of existing systems employ a skin detection algorithm. Face detectors and face trackers make up the majority of these (see [2 - 4, 7, 13, 15 - 17]). In real-time trackers, a skin detector does not do the actual tracking, but instead, does the initial location of the face and acts as a reality check. Face recognition and facial expression analysis systems often use skin detection as an initial step. Hand trackers (such as [1, 9, 11]), while not as common as face trackers, also can make use of skin detectors to aid in tracking.

A majority of the skin detection algorithms use color histograms for segmentation, either directly or for ML estimation ([1, 2, 5, 7 - 12, 14, 16, 17]); others perform pixel classification based on predefined ranges in color space ([6, 13]). Differences among existing skin detection systems occur primarily in the following areas: color space used, incremental run-time training versus off-line training, and the techniques used for sorting and identifying colors corresponding to skin. The skin detection methods used in this paper are based on [11] and [10].

Individual color spaces used in prior skin detection methods include HSV ([9, 12, 13]), a variant of Hue and Saturation ([6]), Normalized RGB ([1, 7, 16, 17]), simple RGB ([10, 11]), YUV ([4]), and transformations from CIE XYZ, including Farnsworth ([3, 15]) and CIE L*a*b* ([2]). Five of these color spaces are compared in the current paper. Further discussion of color spaces and skin detection can be found in [18].

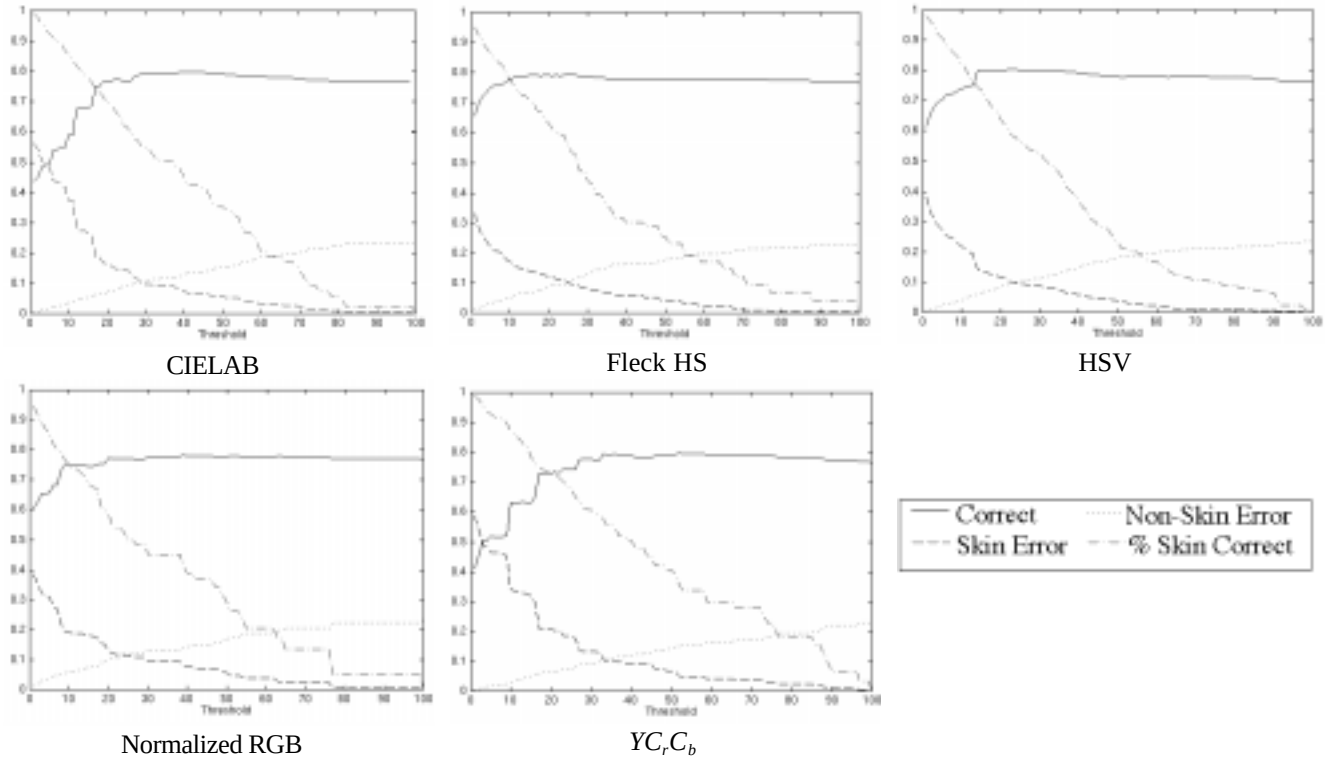

Figure 1: Results for lookup table method across five color spaces for *64x64* resolution.

3. Images Used and Performance Metrics

Both algorithms presented here were trained using 48 images. The testing was done on 64 other images. The images were downloaded from a variety of sources, including frames from movies and television, professional publicity photos and amateur photographs. The images were selected so as to include a wide range of skin tones, environments, cameras, and lighting conditions. Some of the images depicted multiple individuals and the quality of the images varied. The choice of the images is described in greater detail in [18]. To obtain ground truth for training and evaluation of pixel classification performance, the skin regions in all 112 images were marked by hand.

Four different metrics are used to evaluate the results of the skin detection algorithms. **C** (percent correct) is the proportion of all image pixels (both skin and non-skin) identified correctly. **SE** (skin error) is the number of skin pixels identified as non-skin, divided by the number of image pixels. **NSE** (non-skin error) is the number of non-skin pixels identified as skin, divided by the number of image pixels. **S** (% of skin correct) is the proportion of all skin pixels identified correctly.

4. Algorithm 1: Lookup Table Method

The first algorithm presented here uses a color histogram-based approach for segmenting the skin pixels from the remainder of the image. This approach relies on the assumption that skin colors form a cluster in some color measurement space ([10, 11]). A two-dimensional histogram is used to represent the skin tones. By using the two parameters of a color system which do not correspond to intensity or illumination, the histogram should be more stable with respect to differences in illumination and local variations caused by shadows [12].

The two-dimensional histogram used here is referred to as the lookup table (**LT**). Each cell in the **LT** represents the number of pixels with a particular range of color value pairs. A set of training images is used to construct the **LT** as follows: Each image, having been previously segmented by hand, undergoes a color space transformation. Then, for each pixel marked as skin, the appropriate cell in the **LT** is incremented. After all the images have been processed, the values in the **LT** are divided by the largest value present. The normalized values ([0,1]) in the **LT** cells reflect the likelihood that the corresponding colors will correspond to skin.

To perform skin detection, an image is first transformed into the color space. For each pixel in the image, the color values index the normalized value in the **LT**. If this value is greater than a threshold, the pixel is identified as skin. Otherwise, the pixel is considered to be non-skin.

4.1 Lookup Table Results

Initially, ten different **LT**s were constructed, at two different resolutions (*64 x 64* and *128 x 128*) and in five different color spaces (CIEL*a*b*, HSV, an Alternate Hue-Saturation system, referred to here as Fleck HS ([6]), Normalized RGB and YC_rC_b). Each of the ten **LT**s was constructed from the 48 images in the training set and tested with the 64 images from the test set. For each image, the algorithm was run with thresholds from 0 to 1. Figure 1 shows the results for a resolution of *64 x 64*.

For all ten **LT**s, the value of the threshold determined a tradeoff between **NSE** and **SE**. In all cases **C** (% correct) started below 50% and then increased to around 80% where it leveled off. The **S** (% of skin correct) tended to be close to 100% at very low thresholds, and then fell to near zero at high thresholds.

Figures 2 and 3 show these results grouped by **S**, rather than by color space. In each plot, **C**, **SNE** and **SE** are shown for all five color spaces.

At a 60% **S** (Figure 2), HSV and Fleck HS have the best results, with **C** around 80%, **NSE** just over 10% and **SE** just below 10%. The results from the other color spaces are comparable, but with a higher **NSE** and a lower **C**.

At an 80% **S** (Figure 3) the differences between the color spaces become more apparent. Here, Fleck HS has the best results, followed closely by HSV, while YC_rC_b and CIELAB have the worst, with **C** below 70% and **NSE** above 30%. In every case, **SE** is negligible, as it is near or below 5%.

It is interesting that the two Hue-Saturation based color spaces perform better than the two systems designed to accurately reproduce color information (CIELAB and YC_rC_b).

4.2 Adding Double Thresholding

While the **LT** method works reasonably well, it has a tendency to omit pixels (a high **SE** at low thresholds). To overcome this problem, a simple double-thresholding region-growing method was added to the algorithm. A similar technique was used in [6].

Values above a threshold are always considered to be skin, while values below a lower threshold are always considered to be non-skin. All pixels with values above the first threshold are identified as skin. Next, for each pixel, the *5x5* neighborhood around the pixel is examined. If a majority of the pixels in this neighborhood are skin, it is also identified as skin (if not already). Otherwise, the current pixel is identified as non-skin. This has the effect of removing small groups of spatially outlying skin pixels, as well as filling in small areas that were missed.

After applying this simple smoothing technique, the region growing begins. Pixels adjacent to pixels that have been previously identified as skin are examined. If the **LT** indicates a value between the two thresholds, the pixel is identified as skin. This process repeats until the total number of new pixels added from a single pass through the image is less than 1% of the total number of

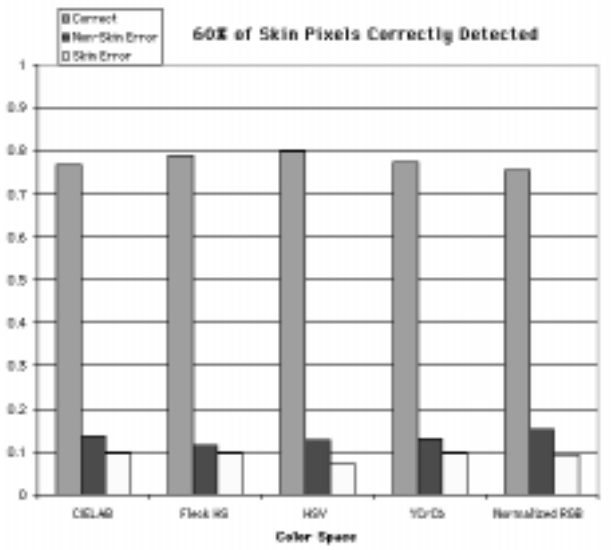

Figure 2: Result at an 60% Skin Correct

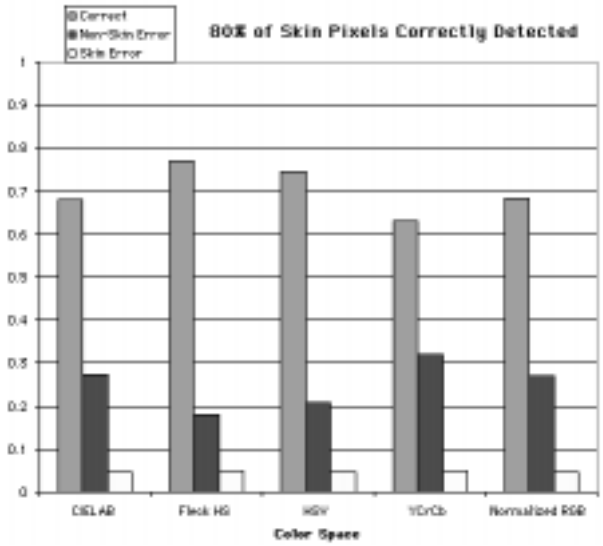

Figure 3: Result at an 80% Skin Correct

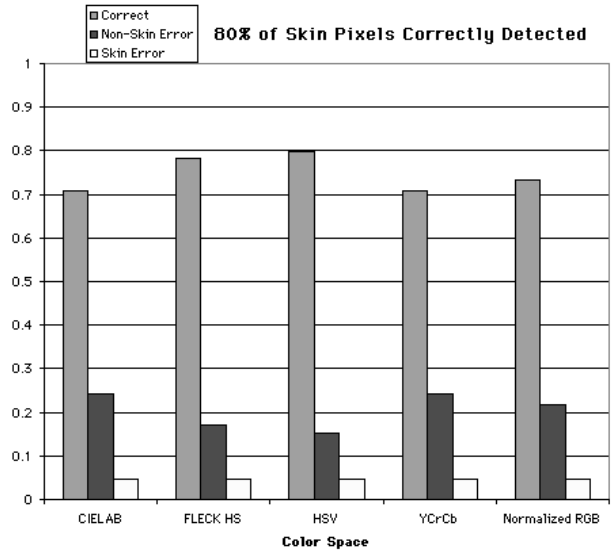

Figure 4: Lookup table method with double thresholding of half of original threshold.

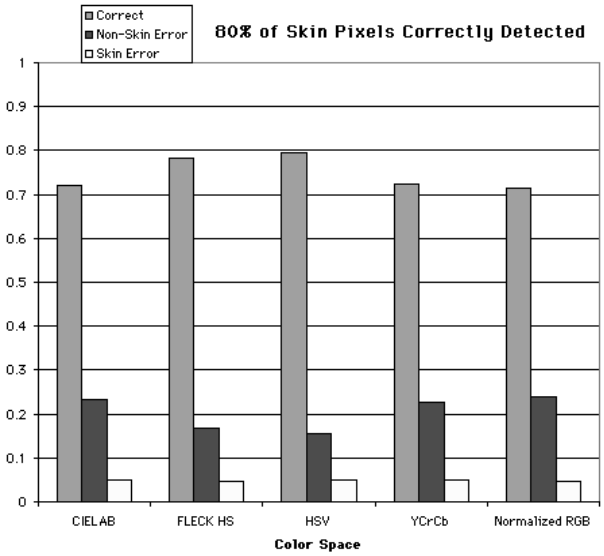

Figure 5: Lookup table method with double thresholding of quarter of original threshold.

pixels in the image.

4.3 Results of Adding Double Thresholding

Instead of testing every combination of thresholds, the second threshold selected was either half or a quarter of the first. Again, testing was done with all five color spaces. Looking at the results for an 80% **S** for both half (Figure 4) and a quarter (Figure 5) of the original threshold, once again HSV and Fleck HS have the best results, while CIELAB and YC_rC_b have the worst results. While the results with double thresholding are improved for the latter two color spaces, overall, using double thresholding with the **LT** does not result in major improvements.

5. Algorithm 2: Bayesian Method

The second method presented here uses Bayes' Theorem to choose the most likely hypothesis, given the value of a feature. Here, the mutually exclusive classes are skin (s) and non-skin ($\neg s$). The feature is the two-dimensional color value $\bar{x}$ of a pixel. In contrast with the lookup table method, the Bayesian method uses two color histograms, one for skin and one for non-skin pixels.

When constructing the probabilities for Bayesian decision making, there are two possible assumptions. In the first case, the probability that a pixel is skin is assumed to be the same as the probability that a pixel is non-skin ($P(s) = P(\neg s)$). This corresponds to maximum likelihood (**ML**) estimation. For any pixel, if the ratio found from (1) is greater than one, then the pixel can be classified as skin. Otherwise, the pixel can be classified as non-skin:

$$\frac{P(s \mid \bar{x})}{P(\neg s \mid \bar{x})} = \frac{P(\bar{x} \mid s)}{P(\bar{x} \mid \neg s)} \quad (1)$$

In the second case, the values of probabilities $P(s)$ and $P(\neg s)$ are estimated from the training data. This corresponds to maximum a posteriori (**MAP**) estimation. If the ratio in (2) is greater than 1, then the pixel can be classified as skin. Otherwise, the pixel can be classified as non-skin:

$$\frac{P(s \mid \bar{x})}{P(\neg s \mid \bar{x})} = \frac{P(\bar{x} \mid s)P(s)}{P(\bar{x} \mid \neg s)P(\neg s)} \quad (2)$$

5.1 Results of Bayesian Method

This method was tested with all five color spaces and with the **MAP** and **ML** techniques. The results are shown in Table 1. The results for each of the color spaces are very close for **ML**. While the results for the different color spaces are very similar, the results for **MAP** and **ML** are quite different. **MAP** has a slightly higher **C**, and a higher **SE**. **ML** has a high **S** (around 90% in all five cases), with a very low **SE**, but a high **NSE** (from 15-20%).

5.2 Bayesian Method with Texture Detection

To reduce **NSE**, a simple texture detection method was used, as skin tends to be smooth and to have little texture. The texture detection method used here measures the variance in a neighborhood. By eliminating areas with a high variance, neighborhoods that are definitely not skin should be eliminated. However, this may have the

undesired effect of reducing the number of correct results, because pixels along the boundary between skin and non-skin regions will also have a high degree of variance. Texture detection was also used in [6].

The algorithm begins by applying the **ML** or **MAP** methods and then applies smoothing as described for the **LT** method. The variance is measured in *3x3* neighborhoods of pixels already identified as skin. Areas with a high texture are reclassified as non-skin.

5.3 Results for Bayesian Method with Texture Detection

In this case, the results were as expected. **NSE** decreased, while **SE** increased and **S** decreased. **C** also increased, which implies that overall, this was a good technique. However, when using **MAP**, the decrease in **S** was very extreme, making the results significantly worse than for other techniques. Overall, the results from **ML** are superior to the results from **MAP.**

6. Conclusions

The **LT** method performed best when used with Fleck HS or HSV, but CIELAB and YC_rC_b yielded poorer results. The results proved to be similar to other skin detection systems ([18]). Double thresholding did not provide a significant improvement over a single threshold for the two color spaces that performed the best. Additionally, the results for a resolution of *128x128* were very close to the results for a resolution of *64x64*.

For the Bayesian method, the biggest difference was between **ML** and **MAP**. **ML** performed significantly better than **MAP** across all five color spaces. Using different color spaces provided very little variation in the results. The **ML** method also worked well when simple texture detection was added, lowering the **NSE**, though also reducing **C** and **S**. The **ML** method produced better results than the **LT** method; in addition, it compares favorably to other skin detection systems ([18]), is computationally efficient, and is easy to implement.

References

[1] Subutai Ahmad, "A usable real-time 3d hand tracker", *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, pp. 1257-1261, 1994.

[2] J. Cai, A. Goshtasby, and C. Yu, "Detecting human faces in color images", *Int. Workshop on Multi-Media Database Management Systems*, pp. 124-131, August 1998.

[3] Qian Chen, Haiyuan Wu, and Masahiko Yachida, "Face detection by fuzzy pattern matching", *IEEE 5th Int.Conf. Computer Vision*, pp. 591-596, 1995.

[4] M. Collobert, R. Feraud, G. Le Tourneur, D. Bernier, J. E. Vaiallet, Y. Mahieux, and D. Collobert, "Listen: A system for locating and tracking individual speakers", *Int. Conf. Automatic Face and Gesture Recognition*, pp. 283-288, October 1996.

[5] T. Darrell, G. Gordon, M. Harville, J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," *Proc. Conf. on Computer Vision and Pattern Recognition*, Santa Barbara, California, June 1998.

[6] Margaret M. Fleck, David A. Forsyth, and Chris Bregler, "Finding naked people", *European Conf. Computer Vision*, vol. 2, pp. 592-602, 1996.

[7] Martin Hunke and Alex Waibel, "Face locating and tracking for human-computer interaction", *IEEE Computer*, pp. 1277-1281, November 1994.

[8] K. Imagawa, S. Lu, S. Igi, "Color-Based Hands Tracking System for Sign Language Recognition," *Proc. 3rd Int. Conf. on Face and Gesture Recognition*, Nara, Japan, April 1998.

[9] Rich Kjeldsen and John Kender, "Finding skin color images", *Int.Conf. Automatic Face and Gesture Recognition*, pp. 312-317, October 1996.

[10] Thomas A. Mysliwiec, "Fingermouse: A Freehand computer pointing interface", Technical report VISLab-94-01, University of Illinois at Chicago, 1994.

[11] Francis K. H. Quek, Thomas Mysliwiec, and Meide Zhao, "Fingermouse: A freehand pointing interface", *Proc. Int. Workshop on Automatic Face- and Gesture-Recognition*, Zurich, Switzerland, pp. 372-377, June 1995.

[12] David Saxe and Richard Foulds, "Toward robust skin identification in video images", *2nd Int. Face and Gesture Recognition Conf.*, September 1996.

[13] Karin Sobottka and Ioannis Pitas, "Segmentation and tracking of faces in color images", *Second Int. Conf. Automatic Face and Gesture Recognition*, pp. 236-241, October 1996.

[14] J-C. Terrillon, M. David, and S. Akamatsu, "Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments," *Proc. 3rd Int. Conf. on Face and Gesture Recognition*, Nara, Japan, April 1998.

[15] Haiyuan Wu, Taro Yokoyama, Dadet Pramadihanto, and Masahiko Yachida, "Face and facial feature extraction from color images",

Int.Conf. Automatic Face and Gesture Recognition, pp. 345-350, October 1996.

[16] Jie Yang and Alex Waibel, "Tracking human faces in real-time", Technical report, Carnegie Mellon University, 1995, CMU-CS-95-210

[17] Jie Yang, Weier Lu, and Alex Waibel, "Skin-color modeling and adaptation", *Proc. of ACCU '98*, vol. 2, pp. 687-694, 1998.

[18] Benjamin D. Zarit, "Skin detection in video images", Technical report VISLAB-99-01, University of Illinois at Chicago, 1999.

Color Space	Technique	C	NSE	SE	S
CIELAB	**ML**	82.72	14.92	2.36	89.99
CIELAB	**MAP**	85.30	8.87	5.83	75.31
Fleck HS	**ML**	77.82	19.73	2.45	89.64
Fleck HS	**MAP**	82.34	8.24	9.42	60.10
HSV	**ML**	79.32	18.05	2.63	88.85
HSV	**MAP**	82.91	8.58	8.51	63.93
Norm. RGB	**ML**	78.42	19.32	2.27	90.40
Norm. RGB	**MAP**	82.48	7.48	10.04	57.45
YC_rC_b	**ML**	83.43	14.14	2.43	89.70
YC_rC_b	**MAP**	85.93	8.33	5.74	75.68

Table 1: Results for Bayesian decision making

Color Space	Technique	C	NSE	SE	S
CIELAB	**ML**	85.52	10.74	3.74	84.17
CIELAB	**MAP**	86.21	6.17	7.62	67.73
Fleck HS	**ML**	84.23	11.89	3.87	83.59
Fleck HS	**MAP**	84.60	3.14	12.26	48.06
HSV	**ML**	84.45	11.38	4.17	82.32
HSV	**MAP**	84.63	3.55	11.81	49.95
Norm. RGB	**ML**	84.25	12.08	3.66	84.49
Norm. RGB	**MAP**	84.04	3.26	12.70	46.19
YC_rC_b	**ML**	85.99	10.25	3.76	84.07
YC_rC_b	**MAP**	86.75	5.53	7.71	67.32

Table 2: Results for Bayesian decision making with texture detection