The Role of Parametric Factors on Visual Assessment of Camouflage Substrates

Juan Lin, Renzo Shamey, Color Science and Imaging Laboratory, North Carolina State University, Raleigh, NC 27695-8301, USA*

Abstract

In order to obtain a rapid, accurate, repeatable and reproducible quality control protocol for the visual assessment of Universal Camouflage Pattern (UCP) substrates, various methodologies were examined. The role of parametric factors in repeatability and reproducibility of visual assessments was determined. Each of the colors within the pattern was measured spectrophotometrically and samples were then assessed by a panel of naïve subjects under various conditions. An acceptability tolerance volume for each color was obtained under simulated daylight illumination at 7500K (equivalent to illuminant D75) using eighteen subjects that repeated assessments six times on separate dates. Visual assessment techniques included a set of sixty woven camouflage samples, some of which contained a repeat pattern that we refer to as "key" and some that did not. Identical subjects repeated assessments under various viewing and surround conditions (e.g. placing samples side by side or juxtaposing samples). A total of 31,320 visual assessments were thus obtained. The role of viewing conditions on the level of inter- and intra-subject variability in pass/fail assessments, as well as color difference evaluation of individual colors, based on an AATCC Gray Scale evaluation method, was examined. STRESS was used to compare the degree of variability among subjects. Analysis of results indicates that subjects' agreement in determining pass/fail responses is improved when a visual reference (such as the print repeat pattern) is included as reference. In addition masking the surround improved subject consistency and repeatability in assessments.

Introduction

Camouflage patterns are developed and used for a variety of reasons. Universal Camouflage Patterns (UCP) is a commonly used camouflage pattern in the United States [1] which was characterized in three general patterns: all over brush, track and shadow/line [2]. The UCP color scheme is composed of three colors: gray (officially named Urban Gray 501), tan (Desert Sand 500) and sage green (Foliage Green 502) printed in a specific pattern [3]. Camouflage is used to conceal objects/personnel from detection by incorporating features that blend them with the background. The better the match between the camouflage and its background the less likely it is to be distinguished by a detector [4,5].

Various visual assessment methodologies have been used in the development of diverse color difference models. However, it is well known that variations in viewing conditions and surround affect the accuracy and repeatability of the relationship between the perceived and measured differences of two color patches. In general visual experiments tend to fall into two broad categories:

1. Threshold and matching experiments which are designed to measure the visual sensitivity to small changes in stimuli [6,7], and;
2. Scaling experiments, which are intended to generate a relationship between the physical and perceptual magnitude of a stimulus.

It is critical to determine which class of experiment is appropriate for a desired application. Matching experiments have been widely applied in the development of models that aim to establish a relationship between the perceived and measured differences among stimuli. However, matching comparisons are rarely used to generate 'scales' because the procedure is time-consuming.

The technique of paired comparisons, which is a type of scaling method, is attributed to Gustav Fechner [8], in which stimuli are compared against a standard, or an anchor. A general evaluation procedure recommended for the visual assessment of colored objects is described in the AATCC Evaluation Procedure 9, which is based on a pair-comparison method. In this procedure a 9 step standard gray scale consisting of 9 pairs of contrast gray chips is used, where pairs correspond to progressive differences in contrast between gray samples against the same standard. Grade 5 is represented on the scale by two reference chips mounted side by side, neutral gray in color and having a Y tristimulus value of 12±1. The color difference of the pair is 0.0 (with a tolerance of +0.2 to account for production and measurement variations). A rating of 5 indicates no color difference exists between the stimuli. At the other end of the scale a rating of 1 indicates a relatively large color difference between samples (13.6 ± 1). The magnitude of the color difference between pairs increases perceptually geometrically from pair 5 to pair 1 including intermediate ratings, such as 3-4. During assessments, the gray scale is often placed along the edges of the test sample pair, which consists of a standard sample and its corresponding batch, and the pair is isolated from the surround using a gray mask. The perceived visual difference between samples is then compared against that of the gray sample pairs. Gray scale ratings are typically transformed into CIELAB color differences, based on the colorimetric values of the gray standards, to yield visual differences, which are shown as DV [9,10].

In addition to the role of the visual assessment methodology on perception of color differences a number of additional parameters must also be considered in the design of the experiment. Parametric factors, for instance, have been shown to have a significant impact on perceptual differences, particularly when results are aimed at establishing an ideal acceptability model. Many of the experimental factors, such as context, surround, adaptive scale, etc. comprise the key variables that influence the development of an accurate relationship between average perceived color differences against the predicted/measured data [11,12]. No reported studies concerning the effect of parametric factors on visual assessment of camouflage patterns could be found in the literature.

In this study, three different visual methodologies were used to assess the appearance of a set of camouflage materials against a standard and subjects' performances under each method was also examined.

The purpose of these different methodologies was to investigate the role of parametric factors on perceptual color differences of multicolored substrates. Samples were either placed side by side with the standard or juxtaposed on top of it. In addition, assessments were carried out in a region that contained the repeat print pattern "the key" and in areas that did not. Variability in perceptual results, under otherwise closely controlled conditions, and using the same set of sample pairs and identical subjects, was then determined. Exact details of the procedure are described in the following sections.

Experimental

Sample Preparation

Sixty universal camouflage printed samples were selected such that they represented a sufficient number of Pass and Fail samples as determined by a US Army expert visual assessor. A Standard Universal Camouflage Pattern sample, obtained from the US Army, was used as reference in all visual assessments.

The repeat pattern in the production samples contains a section that resembles the shape of a key, hereafter referred to as the key (identified in Figure 1). Out of the 60 samples selected for visual assessments 40 included the key pattern but 20 sampled did not.

Figure 1. *The Key pattern (in Tan) identified in the loop.*

The reason for the selection of the key pattern was to provide the subjects with a visual reference when determining pass/fail ratings for the individual colors in batch samples as well as the overall substrate. This also minimized other sources of potential variability amongst samples, e.g. due to mechanical or other printing issues. It must be noted that the full width of fabric may contain several repeat patterns with slight variations amongst 'keys' within the same substrate, however, such variability was ignored. Figure 2a-c shows the colorimetric distributions of three solid colors: Urban Gray, Foliage Green and Desert Sand respectively, for the 60 selected samples. The solid dots represent samples that contained the key pattern and * is used to represent samples that did not have the key pattern.

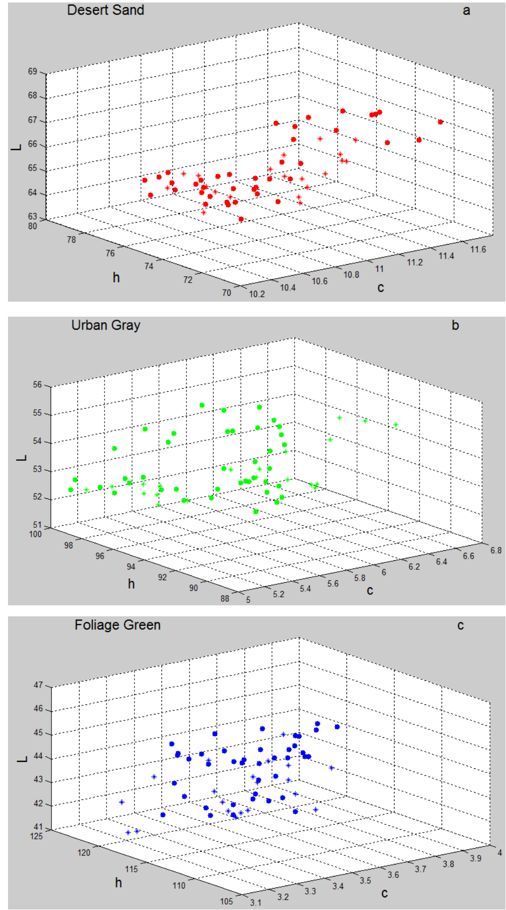

Figure 2a-c. *the Distribution of Tan- Desert Sand (a), Light gray-Urban Gray (b) and Olive- Foliage Green (c) colors in the L*C*h Color Space.*

Visual Assessments

A group of 18 subjects (7M, 11F, Mean age = 22) with normal color vision, as determined by the Nietz test for color vision [13], took part in repeated assessments of samples. Assessments were carried out in six separate trials to avoid visual fatigue during the evaluations. Most subjects did not have any prior experience in assessing color quality of multi-colored samples.

Three types of visual assessment techniques were utilized in this study. These techniques were based on the AATCC Evaluation Procedure 9, option B [14], involving a 0/45 illumination/viewing geometry as shown in Figure 3. The samples were placed in a SpectraLight III (X-Rite) viewing booth, illuminated with filtered tungsten lamps simulating daylight with a correlated color temperature of 7500K approximating illuminant D75. All extraneous lights were eliminated and the illumination conditions were carefully controlled in the course of the experiments in order to minimize variability. Each subject wore a mid-gray laboratory coat and a pair of mid-gray gloves to minimize color variability of the surround during the course of the experiment and to prevent damaging the camouflage samples.

Figure 3. Illumination/Viewing Geometry Employed for the Assessment of Camouflage Samples.

In the beginning of the experiment, each subject viewed the empty illuminated viewing booth for 2 minutes to adapt to the light source, during which time the experiment was explained. The standard was placed adjacent to the production batch being evaluated. Subjects distance from the samples during the evaluation was approximately 60 centimeters. Subjects were not allowed to lift the samples during the course of the visual assessments.

Sample viewing was strived to be limited to 30 minutes whenever possible to avoid fatigue. However, some subjects required up to one hour to evaluate 30 samples in each trial.

Visual Assessment Methods Examined

A Standard sample was supplied by the US Army which was used in all evaluations. The standard was placed on the left hand side of the batch samples being evaluated. Three different visual evaluation methods for samples with and without the key pattern were then carried out:

Method 1:

a. Pass/Fail determination of each color, as well as the overall pattern, in the region containing the key for 20 batch samples placed adjacent to the standard (as shown in Figure 4);

Method 2:

b. Pass/Fail determination of each color, as well as the overall pattern, in a region containing the key pattern for the same 20 batch samples juxtaposed, in part, on top of the standard (as shown in Figures 5);

c. Pass/Fail determination of each color, as well as the overall pattern, for the remaining 10 batch samples juxtaposed, in part, on top of the standard (as shown in Figures 6), in a region outside of the key pattern;

Method 3:

d. Gray scale visual color difference evaluation of individual colors for 20 batch samples juxtaposed on the standard (as shown in Figures 5) in a region containing the key pattern;

e. Gray scale visual color difference evaluation of individual colors for 10 batch samples juxtaposed on the standard (as shown in Figures 6), in a region outside of the key pattern.

In this arrangement each of the colors was assessed for P/F and visual difference in 6 separate trials, on different days, with each trial involving 30 samples (20 with key and 10 without). The overall pattern was also evaluated in terms of P/F when samples were placed side by side and also juxtaposed, separately. In all methods the same batch samples were employed. Each sample was assessed by each subject according to the three assessment techniques indicated above before proceeding to the next sample. Each subject in each trial provided 200 P/F as well as 90 visual difference ratings. Thus a total of 31,320 assessments were obtained and utilized for analysis.

In the first method the subject was asked to visually identify the key pattern on batch samples and compare each individual color around the key pattern with the corresponding colors in the key pattern of the standard placed adjacent to the batch and provide a Pass/Fail rating. Subjects also determined whether the overall appearance of the batch sample was acceptable compared to the standard. While no gap was allowed between the batch and standard samples, the actual distance between the key patterns on the standard and batch varied, slightly, due to variations in the cut size of the production samples received. The potential effect of this variation could not be examined in this study.

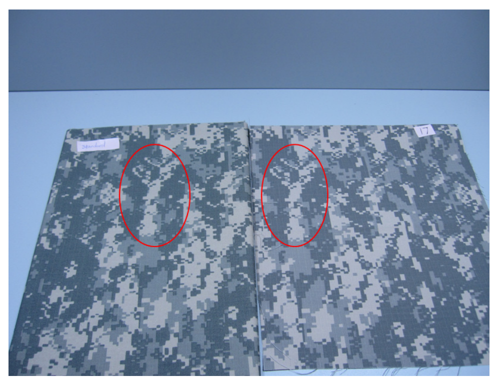

Figure 4. The Key Pattern Identified with Red Ovals on the Batch (right) and Standard (left) Samples Placed Adjacent to Each Other.

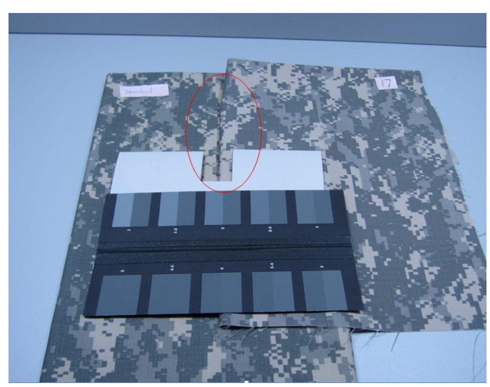

Figure 5. Placement of the Batch Sample Containing the Key Pattern Juxtaposed on top of the Standard for Visual Assessments.

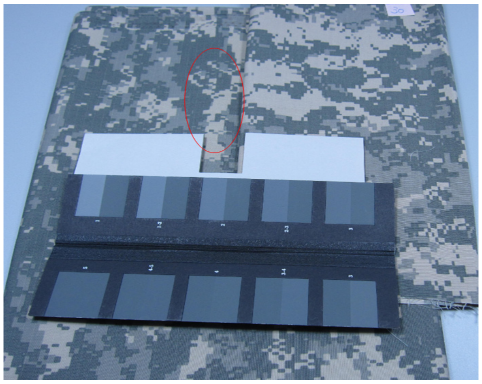

Figure 6. *Placement of Batch Sample with no Key Pattern (right) Juxtaposed on the Standard for Visual Assessments*

In the case of the samples that did not include the key pattern subjects were asked to determine Pass/Fail ratings for each color as well as the overall pattern based on a random region of their choice. The same procedure was repeated for all samples in each trial. For the gray scale ratings each of the individual colors was compared and the difference was evaluated.

The purpose of the first technique was to determine the overall effect of the surrounding colors on the visual perception of each individual color when determining acceptability. The second technique aimed to minimize the effect of surround on the perception of color and color differences and compare the results to those obtained from the first method. The final method aimed to generate a numerical dataset for comparison against spectrophotometric measurements and to assess the magnitude of inter- and intra-subject variability.

Gray Scale Transformations to Visual Difference

For the assessments that involved gray scale visual difference determination (Method 3: d and e), a total of 3,240 evaluations were obtained (30 sample pairs × 6 trials × 18 subjects). The contrast pairs in the standard AATCC gray scale were measured using a Datacolor SF600 spectrophotometer using illuminant D75, 2^o Standard Observer, 9 mm aperture, SPIN, UVEX settings and with an average of 3 readings. The color difference between the standard and each of the gray contrast samples was then calculated using ΔE^*_{ab} equation [9, 14].

The gray scale ratings given by each subject, G_i, were then transformed to visual differences, ΔV, expressed in terms of ΔE^*_{ab} using a third-degree polynomial function shown in Eq. (1).

$$\Delta V = -0.205G_i^3 + 2.6583G_i^2 - 12.469G_i + 21.599 \qquad (1)$$

The ΔV values can be used to determine the degree of variability between trials for a given subject (intra-subject variability) as well as variability in responses obtained from one subject in comparison to the group response (inter-subject variability). Results are discussed in detail in the following sections [15-16].

Subject Variability

The standardized residual sum of squares (STRESS) index, shown in Eq. (2), can be used as a tool for measurement of the strength of the relationship between the visual color difference (ΔV) and the computed color difference (ΔE) for a color pair [15-16].

$$STRESS = 100\ [\ \sum(\Delta E_i - F_1 \Delta V_i)^2 / \sum F_i \Delta V_i^2\ ]^{½} \qquad (2)$$

$$F_1 = \sum \Delta E_i^2 / \sum E_i \Delta V_i^2$$

For a given set of $i = 1, ..., N$ color pairs, the visually perceived color difference is designated by ΔV_i, and the computed color difference by ΔE_i. STRESS can also be employed to determine subject variability in visual assessments. Two types of variability are often calculated, namely intra-subject variability, which indicates the repeatability of the same subject in different trials, and inter-subject variability, which determines the reproducibility of results given by a subject in relation to the general response from a group of subjects. A STRESS value of zero indicates perfect agreement between results from different trials or amongst different subjects.

The inter-subject variability, also named subject accuracy, is deviation between mean results from each subject against the mean results of a panel of subjects, while intra-subject variability is deviation among results of a given subject in replicated trials in an experiment. For the assessment of intra-subject variability ΔV_i and ΔE_i are replaced by the visual responses of a given subject in two different assessment sessions. For the assessment of inter-subject variability ΔV_i and ΔE_i are replaced by the mean responses obtained from a given subject and grand mean response from all subjects respectively.

The subject repeatability in trials for Method 3:d-e was calculated using gray scale ratings and the mean value for the group as well as each subject was obtained. Tables 1-3 show the STRESS values for intra-subject variability in determination of color differences between standard and batch samples for each of the three colors used in the Universal Camouflage Pattern substrates.

Results shown in Tables 1-3 indicate that the mean intra-subject variability in assessment of each color in different trials for all samples ranges from approximately 62 to 70 STRESS units. A closer examination of results indicates that subjects 3 and 9 may be outliers. Removing these subjects from assessment results reduces the mean STRESS for intra-subject variability from approximately 66 to about 63, a reduction of roughly 5%. Despite the modest reduction in mean STRESS, the high values shown in Tables 1-3 would be generally considered too large in the assessment of suprathreshold color differences for solid color samples. However, a comparison of assessments for the three colors shows that the level of variability is similar for each of the three colors examined.

It should also be noted that samples examined in this study contained low chroma (3• C* •12), tan, olive and urban gray colors which are usually considered challenging in terms of visual assessment of color differences. STRESS results obtained in this study also indicate that variability among subjects in assessing these colors is high. Notwithstanding, while small differences among already high STRESS values for each of the colors may not be significant, variability is apparently increased from Olive to Urban Gray to Tan colors (i.e. from lower to higher C* samples), with the difference between highest and lowest mean of 3.5 STRESS units. It has been established that tolerances are generally tighter for lower chroma samples, thus higher variability results for higher chroma samples are in line with the relatively higher tolerances.

Table 1. *Intra-Subject Variability in the Assessment of Tan (Desert Sand) Color expressed in terms of STRESS for All Samples*

Subject	Trials 1&2	Trials 2&3	Trials 1&3	Mean
1	65.04	59.04	60.47	61.52
2	72.31	73.39	63.51	69.74
3	92.87	97.77	97.09	95.91
4	52.43	58.87	66.13	59.15
5	72.74	78.62	62.91	71.42
6	70.12	60.08	58.31	62.84
7	89.13	83.49	81.27	84.63
8	64.56	74.40	71.49	70.15
9	84.87	94.76	94.89	91.51
10	67.55	66.66	68.90	67.70
11	70.81	53.74	60.86	61.81
12	77.20	76.93	66.36	73.50
13	51.79	64.01	62.79	59.53
14	42.82	46.81	45.35	44.99
15	66.59	63.15	64.03	64.59
16	70.13	68.29	54.90	64.44
17	68.28	68.34	63.62	66.75
18	35.99	47.72	45.08	42.93
Mean	**65.71**	**68.67**	**66.00**	**67.40**

Table 2. *Intra-Subject Variability in the Assessment of Light Gray (Urban Gray) Color Expressed in Terms of STRESS*

Subject	Trials 1&2	Trials 2&3	Trials 1&3	Mean
1	62.07	71.97	56.77	63.61
2	76.88	66.99	72.91	72.26
3	99.15	98.64	99.69	99.16
4	52.28	69.50	67.28	63.02
5	61.91	69.00	57.75	62.89
6	65.34	61.35	67.08	64.59
7	65.30	85.32	72.90	74.51
8	69.95	80.60	74.95	75.17
9	83.18	90.34	87.79	87.10
10	51.53	52.69	45.03	49.75
11	73.51	65.49	71.74	70.25
12	75.22	70.92	72.57	72.90
13	67.09	69.18	55.44	63.90
14	48.69	41.53	35.12	41.78
15	77.35	68.84	64.52	70.24
16	71.08	66.32	59.42	65.61
17	57.74	71.31	59.73	62.93
18	37.61	51.51	46.54	45.22
Mean	**66.44**	**69.53**	**64.85**	**66.94**

Table 3. *Intra-Subject Variability in the Assessment of Olive (Foliage Green) Color Expressed in Terms of STRESS*

Subject	Trials 1&2	Trials 2&3	Trials 1&3	Mean
1	59.39	55.43	52.77	55.86
2	65.71	74.45	71.79	70.65
3	96.32	99.50	87.27	94.36
4	71.22	59.66	70.96	67.28
5	65.93	60.67	52.41	59.67
6	67.70	71.46	55.74	64.97
7	80.95	81.28	76.76	79.66
8	62.90	62.99	65.72	63.87
9	97.46	92.05	82.55	90.69
10	72.18	73.87	62.66	69.57
11	63.37	57.63	60.55	60.51
12	63.78	63.33	63.98	63.70
13	45.81	46.12	44.09	45.34
14	33.13	36.90	31.17	33.73
15	60.23	59.64	57.17	59.01
16	62.70	61.03	60.05	61.26
17	48.51	61.17	65.56	58.41
18	36.01	61.48	56.22	51.24
Mean	**64.07**	**65.48**	**62.08**	**63.88**

In addition, the presence of neighboring colors is well known to affect visual judgments, due to the simultaneous contrast effect. Thus it was expected that a change in the color of the surround, or the print pattern, would have an important effect on the perception of the colors considered. This was considered in this study. The AATCC Evaluation Procedure 9 involving the gray scale assessments recommends the use of a viewing cardboard (shown in Figures 5, 6) to mask the surround while viewing a batch against the standard. A pilot study involving a limited number of assessments that involved the mask, and those without indicated that while the use of a mask may be beneficial in reducing the effect of surround, it is considered impractical by users. This aspect of work was not further examined in this study.

A different approach was to examine visual assessment results for samples that contained similar patterns against those that did not. Table 4 shows two sets of intra-subject variability results. The first set involves assessments that included the key pattern and the second set those that did not. Unfortunately sample population in two sample groups was not identical. As described earlier samples comprised 40 that had the key and 20 that did not. Nevertheless, an examination of results shows that the mean intra-subject variability is lower for Light Gray samples without a reference pattern, while it is slightly higher for Tan and unchanged for the Olive colors. Thus, no firm conclusions based on the effect of surround on intra-subject variability could be drawn. In addition, while results in absence of the two outlier subjects are slightly better (shown in the last row of the Table 4 with * and in red), the general pattern is not affected.

The inter-subject variability in all trials and for each of the three colors was also calculated as shown in Tables 5 and 6. STRESS results compared to the intra-subject variability for the assessment of the three colors for all trials are relatively lower. This implies that subjects' assessments within the group are reasonably similar. However, again, the mean values are relatively large, especially in comparison to assessments of suprathreshold color differences of solid colors, as reported in the development of recent color difference models [17].

***Table 4.** Mean Intra-Subject Variability in the Assessment of All Colors in Samples Containing the Key and Those without a Key Pattern Expressed in Terms of STRESS*

	Light Gray		Tan		Olive	
	No Key	Key	No Key	Key	No Key	Key
1	58.59	54.05	59.81	60.96	55.47	66.51
2	57.67	78.46	64.83	77.02	60.47	76.84
3*	92.20	96.86	98.53	87.51	99.29	98.91
4	56.26	70.61	55.40	60.61	64.55	60.49
5	63.45	55.39	73.55	70.16	71.85	56.50
6	44.34	77.34	71.86	54.36	69.22	60.63
7	75.29	84.34	85.32	83.42	77.33	83.57
8	59.70	65.84	75.07	66.01	77.62	73.47
9*	86.78	92.99	94.34	87.99	79.52	91.66
10	68.04	69.19	65.33	68.17	48.14	49.72
11	56.43	58.33	56.61	62.49	76.01	66.26
12	52.97	67.17	65.95	72.02	75.53	71.28
13	37.65	49.43	60.92	57.37	81.10	58.80
14	32.63	34.66	45.05	43.70	38.18	42.82
15	57.57	57.04	73.46	59.50	76.94	62.19
16	48.09	68.62	73.30	55.52	59.18	64.25
17	45.95	58.45	55.32	67.80	51.68	69.12
18	42.10	55.07	35.52	46.86	39.65	48.34
Mean	**57.54**	**66.32**	**67.23**	**65.64**	**66.76**	**66.74**
Mean*	53.55	62.75	63.58	62.87	63.93	63.17

Although a direct comparison of results is not possible, it should be pointed out that some of the subjects that took part in this study were participants in previous suprathreshold color difference assessments of solid color samples with a typical inter- and intra-subject variability range of approximately 20-35. Thus, it may be concluded that assessment of suprathreshold color differences in multi-colored patterns is less repeatable for subjects than that of solid colors.

***Table 5.** Inter-Subject Variability in the Assessment of All Camouflage Substrates Expressed in Terms of STRESS Index*

Subject	Light Gray	Tan	Olive
1	37.66	42.21	31.74
2	46.52	52.43	43.17
3	61.27	63.61	63.25
4	34.04	34.13	37.50
5	27.00	50.14	45.50
6	40.83	38.18	52.74
7	41.04	55.51	58.40
8	38.26	42.01	40.53
9	60.03	57.10	62.31
10	39.43	37.48	26.80
11	36.05	52.93	45.98
12	41.97	52.04	53.15
13	32.58	42.81	54.12
14	24.04	44.02	22.63
15	41.82	39.18	39.65
16	29.91	42.27	36.83
17	57.61	48.27	60.97
18	36.85	36.98	43.98
Mean	**40.38**	**46.18**	**45.51**
Mean*	37.85	44.41	43.36

***Table 6.** Inter-Subject Variability for Samples Containing the Key and Those without a Key Pattern Expressed in Terms of STRESS*

	Light Gray		Tan		Olive	
	No Key	Key	No Key	Key	No Key	Key
1	35.34	37.06	28.02	46.07	23.80	35.66
2	42.21	46.63	52.92	47.87	35.87	45.32
3*	49.06	69.64	61.62	67.93	63.27	64.85
4	26.94	36.57	21.32	37.65	38.22	37.47
5	22.93	27.84	45.08	43.10	48.95	41.24
6	37.72	40.80	32.05	34.07	41.20	59.10
7	30.31	45.44	63.22	50.32	56.73	58.62
8	29.00	44.53	42.85	40.62	39.01	42.30
9*	63.94	53.90	65.25	51.27	49.22	66.91
10	37.16	38.10	38.47	35.33	33.32	23.14
11	21.77	44.11	39.46	52.67	39.02	48.39
12	34.83	43.20	53.49	48.51	58.33	46.41
13	25.41	36.95	30.34	41.20	61.62	44.38
14	14.82	29.91	37.65	39.95	23.73	22.00
15	38.11	40.25	45.03	33.47	30.03	44.54
16	32.75	27.13	31.94	39.23	34.62	36.48
17	45.56	62.29	26.53	54.70	47.60	67.24
18	36.12	35.22	33.18	35.32	34.55	47.91
Mean	34.67	42.20	41.58	44.40	42.17	46.22
Mean*	31.94	39.75	38.85	42.51	40.41	43.76

An examination of individual results for subjects also shows that compared to the group mean, Subject 14 was the most consistent and repeatable for foliage green and urban gray colors. Again here Subjects 3 and 9 show relatively high STRESS values indicating relatively poor reproducibility within the group. Results in absence of the two outlier subjects are slightly better (shown in the last row of Tables 4-6 with * and in red), but the general pattern is not affected. Other statistical functions in conjunction with STRESS may be used to indicate subjects' performance and screen subjects in a pilot study prior to conducting larger visual assessments.

Analysis of Results from Various Visual Color Assessment Methodologies

Due to space limitations the analysis shown in this paper does not cover all aspects of the study and only a selected set of results is discussed. The data obtained from various assessments in this study were, however, analyzed separately as well as combined. A Pass rating was assigned a value of (1) and a Fail rating was given a value of (-1) in each trial for each subject. Subjects were also asked to give an overall Pass/Fail rating for the substrate independent of their separate evaluations for each color in each sample. The role of employing different assessment Methods on ratings was also examined. Since each subject repeated the assessments three times, the statistical function "mode" was used to obtain the overall P/F response for each color in each sample. Individual responses from all subjects were also used to determine the overall group response for each sample as well as for each color in each sample for each method. Pass/Fail ratings were also calculated based on the percentage of Pass or Fail responses from all subjects. Responses based on Mode were compared to % P/F ratings for each sample, an example of which is shown in Figure 7.

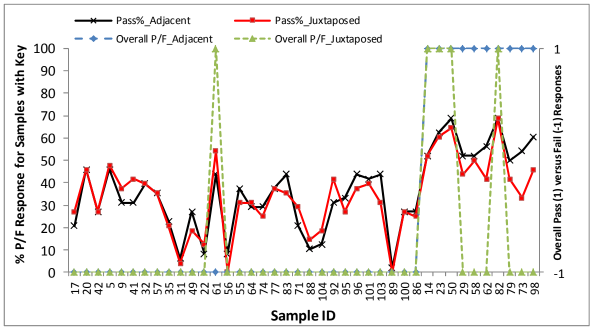

Figure 7. Comparison of Subjects' Overall Mode Ratings for Samples Containing the Key Against % Pass or Fail Ratings.

It can be seen that with the exception of few samples, a unanimous P/F rating was not obtained. In fact in many cases an overall group Pass or Fail rating included a considerable percentage of opposite responses. This is likely due to several factors, including high variability associated with the low chroma colors examined, the complexity of the pattern, the effect of surround and variations in the visual subtense during the assessment. Nonetheless, of particular interest are samples that received an overall Pass rating but included at least one color (from three) that failed as shown in Figure 8.

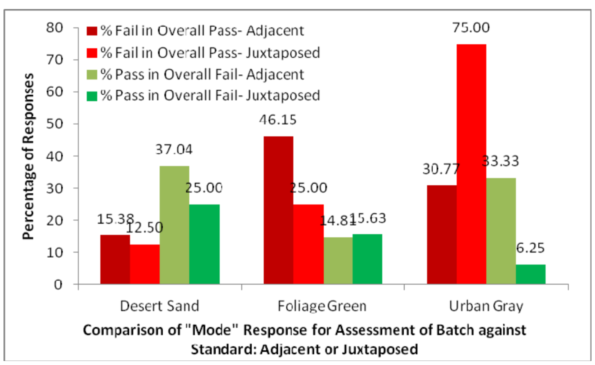

Figure 8. Comparison of the Overall % Pass or Fail for the Substrate Against P/F for Each Color when Samples Were Placed Adjacent to, or on Top of the Standard.

This indicates that the overall *appearance* of the substrate is more influential in subjects' decision to pass or fail a multicolored substrate than variations for individual colors within the pattern.

To examine the role of surround, evaluation results from subjects for Methods 1a and 2b (regions with or without the "key") were compared. The overall P/F responses for each of the substrates from the group of subjects were also compared to the ratings provided by the US Army expert assessor. In the course of the analysis samples were ordered based on the P/F ratings given by the US Army expert assessor from P-1 to P-24 (Pass) and F-1 to F-36 (Fail). This arrangement was used to generate charts and compare results. Figure 9 compares P/F ratings for identical samples that were placed once adjacent to, and another time juxtaposed on top of the standard. This comparison included only the 40 samples that contained the key pattern to avoid inclusion of variability due to other potential issues. It can be seen that the mode of viewing affected the overall P/F ratings for a number of samples. Figure 9 also shows that the overall responses from subjects and those from the US Army expert assessor do not agree in several cases.

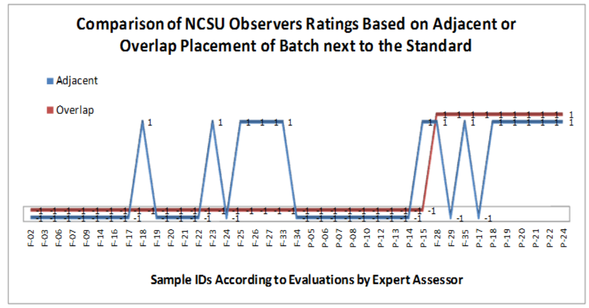

Figure 9. Overall % Pass/Fail for samples containing the Key when Placed Adjacent to or on Top of the Standard Assessed by Subjects Compared to Assessments by the US Army Expert Assessor.

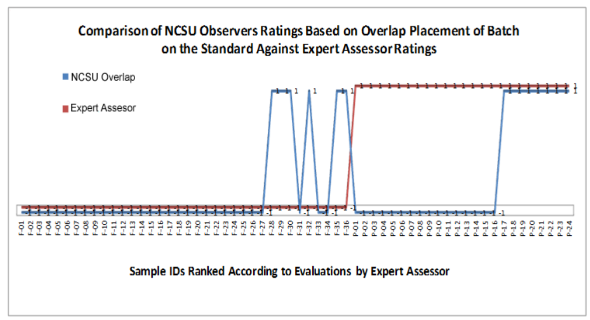

Figure 10. Comparison of the Overall % Pass/Fail Rating for All Samples Placed on Top of the Standard by Subjects and the US Army expert Assessor.

Figure 10 compares the evaluation of samples laid on top of the standard and assessed by the US Army expert assessor and the overall P/F response from subjects. This comparison included all sixty samples. Results show that a significant number of samples evaluated as Pass (P-01 to P-16) were considered to be "Fail" by subjects and vice versa. While this could be due to varying levels of acceptable production tolerances, as determined by different subjects, it could lead to difficulties in determining an agreeable acceptability boundary for the reproduction of samples.

As can be seen only a very small number of samples exhibit a high percentage of agreement among subjects as either pass or fail. In fact in the case of assessments according to Method-1a only one sample (F-03) showed 100% agreement amongst all subjects and the US Army expert assessor to be Fail. In the case of assessments according to method-2 this is increased to 5 samples, which is still a small proportion (8%) of the total number of samples evaluated. Inter- subject variability seems to be particularly high for multicolored samples and this indicates the complexity of reaching consensus to develop acceptability volumes.

Figure 9 shows subjects' P/F responses for samples, with key patterns, placed adjacent to or on top of the standard. Results include assessments from 16 subjects, excluding subject 3 and 9. Several samples were rated as Fail when they were placed on top of the standard. Thus it seems subjects are more critical of variability in samples when the distance between samples is reduced and the field of view is smaller.

Assessment responses for individual colors, desert sand, foliage green and urban grey in adjacent and juxtaposed methods also indicate that variations in individual colors are more likely to affect subjects overall assessment of Pass/Fail ratings when batch samples are laid on top of the standard. An example is shown in Figure 11 for Foliage Green.

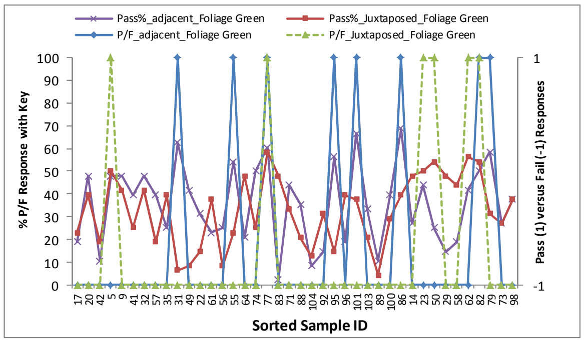

Figure 11. *Comparison of the % Pass Rating for Foliage Green Color in Samples Placed Adjacent to or on Top of the Standard by Subjects.*

As can be seen variations in the assessment method result in disagreements in overall Pass or Fail ratings as well as the percentage of subjects that believe the sample should be rated as Pass. In some cases in fact, results are contradictory, e.g. samples 31, 55, 95, etc.

Conclusions

It is clear that visual assessment of multicolored substrates is challenging, and Pass/Fail responses in comparison to suprathreshold small color differences for solid color assessments are less reproducible. It is also clear that developing standard protocols for the assessment of multicolored objects will be far from trivial. It was attempted to examine the role of some of the parametric factors that may influence subject's decisions in ratings. This study showed that assessing colors in identical regions and placing a mask to negate the effect of the surround on assessments of a given color did not result in *significant* improvements in performance for subjects and may have caused subject fatigue. Visual assessment of individual colors in multicolored samples when batch samples were overlaid on top of the standard seemed to yield more reproducible results compared to assessments involving placing batch samples next to the standard. It was thought that placing a mask in the region of interest may minimize the effect of the surround on the perception of color differences. However, the method was considered impractical by the majority of subjects and was not further examined.

In assessments where a batch sample was placed next to the standard it was not possible to control the distance between stimuli being compared exactly. This was due to varying size and the print pattern and production submits which is very common in practical assessments of production samples. This could result in varying the size and appearance of the stimuli and evaluation of each color within the pattern, which in turn can influence the overall appearance of the object. Conscientious assessors often strive to employ identical fields of view and patterns to minimize variability. Yet while results varied among subjects, this was not found to result in a *significant* improvement in repeatability or reproducibility of responses for majority of subjects tested in this study.

Assessments of color differences according to the ISO (or AATCC) gray scale for color change method for individual colors within the pattern were carried out to obtain numerical color differences for each color. This enabled comparison of visual differences amongst subjects using the STRESS Index. Intra-subject variability results based on STRESS values show that subjects' repeatability in trials involving multicolored patterns is low. However, inter-subject variability results indicate that subject responses within the group were reasonably consistent though somewhat higher than those for suprathreshold color difference evaluation of solid (unrelated) colors. The STRESS function does not exhibit the full complexity of subject behavior during visual assessments, since majority of subjects showed relatively large variability from trial to trial. However, a comparison of subject's STRESS results may be used to screen out potential outlier subjects prior to conducting full scale visual assessments. Further work is required to elucidate the role of various parameters on perception of color differences in multicolored objects.

Acknowledgments

The authors thank Dr. J. Zheng, Ms. V. Halls, & Dr. F. Crivello, from the US Army for the financial support of this project and Ms. M. King for providing camouflage samples with associated expert visual assessor evaluations and all subjects who took part in visual assessments. Dr. D. Hinks provided useful comments during the course of the experiment.

References

[1] Army Combat Uniform (ACU), Presentation, available at http://www.armystudyguide.com/content/powerpoint/Uniforms_Presentations/acu-presentation-2.shtml

[2] Dugas, Anabela; Kramer, F. Matthew. Universal Camouflage For The Future Warrior. U.S. Army Natick Soldier Center. 2004, 27.

[3] New Digital U.S. Army Combat Uniform eliminates Black in pattern, available at http://www.hyperstealth.com/acupat/.

[4] Kenneth S.N., L C.H., An Analysis of Background Color-matching in Amphibians and Reptiles, Ecology, 45: 16, 1964.

[5] Guilford T. Predator Psychology and the Evolution of Prey Coloration. Natural Enemies: Blackwell Scientific Publications. 375-394, 1992.

[6] Engeldrum P.G., Psychometric Scaling: a Toolkit for Imaging Systems Development, 2000.

[7] Sharma G., Bala R, Digital Color Imaging Handbook, CRC Press.

[8] Nunnally J.C., Bernstein IH, Berge JMF, Psychometric Theory, 1967.

[9] AATCC EP1, Gray Scale for Color Change, AATCC RA36 Evaluation Procedure 1; 2004, 378-379.

[10] ASTM D2616-96, Standard Test Method for Evaluation of Visual Color Difference with a Gray Scale, 2003.

[11] Kuo W.G., Luo M.R., Methods for Quantifying Metamerism. Part 1-Visual Assessment, JSDC, 112, 1996, 312-320.

[12] Guan S.S., Luo M.R., Investigation of Parametric Effects using Small Color Differences, Color Res. Appl. 24, 1999, 331-343.
[13] Neitz, J., Manual: Neitz Test for Color Vision, Western Psychological Services, Torrance, CA, USA, 2001.
[14] AATCC EP 9, Visual Assessment of Color Difference of Textiles, AATCC Technical Manual 82, 2007, 394-396.
[15] García P.A., Huertas R, Melgosa M. Cui G. (2007). Measurement of the relationship between perceived and computed color differences, J. Optical Society of America A, 24, (7), 1823-1829.
[16] Melgosa M., Garcia P., Gomez-Robledo L., Shamey R., Hinks D., Cui G., Luo R. (2011). Notes on the application of the standardized residual sum of squares index for the assessment of intra- and inter-observer variability in color difference experiments. J. Optical Society of America A, 28 (5), 949-953.
[17] Luo M. R., Cui G, Rigg B, The development of the CIE 2000 color-difference formula: CIEDE2000. Color Res Appl. 2001; 26:340-350.

Author Biography

Juan Lin obtained her MSc in image processing from Nanjing University, China in 2009 and a PhD in Color Science from North Carolina State University in 2013. She is currently a postdoctoral associate at the Color Science and Imaging Laboratory at NCSU. Her main research interests include color science, image processing, color measurement and color management

Renzo Shamey is the Director of the Polymer and Color Chemistry program and Color Science and Imaging Laboratory at North Carolina State University. He is a Professor of Color Science, a Fellow of the Society of Dyers and Colourists, and author of well over 100 manuscripts. He is active in the CIE Division 1 committees 55, 76, 77 and currently chairs the AATCC RA36 Color Measurement Test Methods Committee.