https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-232

The Benefits of Color over Black-and-White Images in Task-Oriented Reconnaissance Applications

Cicely DiPaulo and Larry Scarff, UTC Aerospace Systems, Westford, MA USA

Abstract

Color imaging is such a ubiquitous capability in daily life that a general preference for color over black-and-white images is often simply assumed. However, tactical reconnaissance applications that involve visual detection and identification have historically relied on spatial information alone. In addition, real-time transmission over narrow communication channels often restricts the amount of image data, requiring tradeoffs in spectral vs. spatial content. For these reasons, an assessment of the discrimination differences between color and monochrome systems is of significant interest to optimize the visual detection and identification of objects of interest.

We demonstrate the amount of visual image "utility" difference provided by color systems through a series of subjective experiments that pair spatially degraded color images with a reference monochrome sample. The quality comparisons show a performance improvement in intelligence value equivalent to that achieved from a spatial improvement of about a factor of two (approximately 1.0 NIIRS). Observers were also asked to perform specific detection tasks with both types of systems and their performance and confidence results were measured. On average, a 25 percent accuracy improvement and a 30 percent corresponding confidence improvement were measured for the color presentation vs. the same image presented in black-and-white (monochrome).

1. Introduction

The evolution of color digital imaging began in 1991 with Kodak's 1.3 megapixel CCD that utilized the now-common color filter array invented by Bruce Bayer, coupled to a Nikon F3 camera body. Today, color digital imaging is so commonplace that the very rare experience of panchromatic (black-and-white) photography is primarily encountered as a means for artistic expression. It seems almost unnecessary to question the benefits of color over black-and-white pictures, but in this paper we describe the historical basis for visual image analysis and the interesting challenges to document the benefits offered by color images when performing task-based exploitation and analysis via trained observers.

Background

Two primary application areas that utilize experts to observe and detect anomalies via images are in medical diagnostics (radiology, cardiology, etc.) and military reconnaissance. These two groups have historically reviewed both hardcopy and softcopy representations that are primarily grayscale presentations. While both fields are finding some use for spectral (color) information, the foundation of the tradecraft for each group is based on detecting spatial characteristics and features in images that may not present significant spectral content, and thus the advantage of color is not as obvious. We wish to determine the amount – if any – of added value when both spectral and spatial information is presented in combination.

Motivation

For tactical military applications, there is significant complexity involved in creating a sensing system that can collect a high-resolution true-color (or multispectral) image of a large area when used in fast-jet applications (e.g., F16 aircraft). While color video camera systems are often used in low-altitude remotely piloted aerial vehicles, they do not attain the wide-area or high-resolution capabilities needed for long-range, oblique photography. For these systems, it is important to understand the intelligence tasks that need to be addressed and the cost/benefit tradeoff that color offers over panchromatic imaging.

Image Quality Assessments

In Intelligence, Surveillance, and Reconnaissance (ISR) imaging, a standard metric for subjective image quality is the National Imagery Interpretability Rating Scale (NIIRS) [1]. This metric is strictly based on spatial characteristics (e.g., "identify automobiles as sedans or station wagons") and thus the value of spectral (color) information is not directly captured in the assessment. When designing color and multispectral task-based imaging systems, a metric that demonstrates the advantage of improved color discrimination in observer-based object detection and identification is desired, and ideally could be analogous to a "NIIRS improvement" offered by the inclusion of spectral information – either in a true-color or false-color presentation. Knowledge of the impact of color information on visual discrimination – including confidence, efficiency, and accuracy – can aid both the system designer and the acquisition planner when optimizing reconnaissance systems over the vast range of intelligence problems encountered.

Anticipated Benefits

By characterizing the degree of intelligence benefit derived from color over monochrome images, the image collection process can be better optimized for greater flexibility in target acquisition. In addition, improvements in analyst efficiency and accuracy can be demonstrated that will provide increased intelligence value to the warfighter. While spatial and spectral content each provide a different type of intelligence information to the observer, it is their combination that can yield advantages in recognition and identification beyond that available through spatial feature analysis on grayscale imagery.

2. Experimental Design and Methods

Our initial hypothesis was the following: Given a monochrome and color image with equivalent spatial characteristics, an analyst would choose to exploit the color image, since there is no degradation in spatial characteristics and the additional spectral information could impart relevant target information which might make it easier (i.e., quicker, more efficient) and/or lead to more confidence (accuracy) in their assessment.

Additionally, we hypothesized that if the spatial content of the color image is successively degraded (reducing its resolution), there will be a point at which the analyst will switch and choose to exploit the monochrome image rather than the reduced-resolution color. This point where the two images have equivalent intelligence value provides an indication of the amount of spatial information loss – which can be characterized in terms of NIIRS – that is compensated by the added spectral content of the color image.

Two Evaluation Scenarios

Our study was split into two parts: The first focused on efficiency and confidence comparisons when viewing identical color vs. monochrome images. This study was comprised of ten images in two sets: the first set contained five scenes presented in monochrome, while the second set contained the same fives scenes but presented in color. A specific object of interest was identified for each scene, and participants were given 60 seconds to view each of the images looking for that specific object and counting their number. These objects were: umbrellas, people, flower pots, garbage cans and swimming pools (Fig. 1 provides an image example of these detection tasks). The observers were then asked to record three values: (1) the total number of objects found, (2) their confidence that each item they counted was correctly identified, and (3) their confidence that they had found all of the objects in the scene. Between the monochrome and color image sets, the participants were asked to comment on the difficulty of the task using monochrome imagery. At the end of the evaluation, they were again asked to comment on the difficulty of the tasks, this time when using the color images.

The second part of our evaluation compared the overall intelligence value between a reference monochrome image and corresponding color images of varying spatial quality. The observers were told that these two images were examples from two competing systems, and they were asked to choose which system would provide them with the best *overall* intelligence value, based on their experience and all of the tactical intelligence problems they might encounter. We provided a NATO guide [2] of specific target reporting categories to aid in their assessment.

The image pairs always contained a full-resolution monochrome image paired with a color image of the same scene, sometimes of equivalent spatial resolution, and other times reduced in resolution by factors of 1.4, 2.0, and 4 (equivalent to an approximate NIIRS reduction of 0.5, 1.0, and 2.0 units). In this evaluation, both nadir and oblique images were evaluated (see the next section - “Imagery Sources”). The presentation order of the images was randomized, with slight adjustments made to ensure that identical scenes with different spatial resolutions never appeared immediately following each other.

Figure 1 – (Left) Enlargement of the five objects to be detected in the first evaluation. (Right) Example of scenes used for evaluation [4] (Top = garbage can scene, Bottom = umbrella scene.)

Both evaluations were provided to the observers in the form of PowerPoint presentations, with the images always filling the same area on the screen. Participants in the first study viewed the images on the same monitor (27” flat panel, 1920x1080 resolution). For the second evaluation, remote participants were asked to ensure that they were using monitors that provided at least a 1920x1080 display resolution.

Imagery Sources

Reconnaissance cameras can collect both nadir and oblique images, and therefore we wished to include both vertical (nadir) and oblique image examples in our evaluations. The military nature of the images collected with most reconnaissance systems precludes their open distribution and public disclosure. To implement this evaluation, we obtained permission to utilize open-source imagery from two Internet sites [3] and [4]. This imagery not only provides very similar characteristics to common reconnaissance systems but also demonstrates that the results may be applied to commercial and civil systems as well. We have provided some representative examples of the scenes in Fig. 2. Because hardcopy reproduction of this paper may not provide color output, we encourage the reader to obtain a softcopy version of this paper or visit the websites listed in the references to observe the color characteristics of the images.

Figure 2 - Example imagery: nadir [3] (left) and oblique [4] (right)

Participants

There were twelve participants in the first part of our evaluation (*efficiency and confidence comparison*) divided equally into two groups: six participants were US Air Force-trained imagery analysts, and six others were skilled image scientists and engineers. The second part (*intelligence value*) utilized 16 observers, eleven of which were US Air Force-trained analysts that have had previous experience evaluating multispectral imagery, and five were image scientists/engineers. All participants were current employees of UTC Aerospace Systems, including five analysts who were deployed at remote sites worldwide – the remaining observers were located at the UTC Aerospace Systems facility in Westford, MA.

Procedure

The two evaluations were conducted separately, with about two months’ time elapsed between them. Both evaluations were provided to the observers as a PowerPoint file, along with a score sheet and instructions, allowing both evaluations to be self-administered.

In the first evaluation, the images automatically advanced every 60 seconds to maintain a consistent viewing time among all observers. After the first five (monochrome) images, a pause was inserted into the automatic slide advance to allow the observers time to record their thoughts about task difficulty. When the evaluation concluded, the observer was asked to provide a final comment on task difficulty, this time based on the color images. The observers provided the object count and the requested confidence levels for their object identifications for each image on the scoresheet.

In the second evaluation, the observers could take as much time as desired to make their comparison judgement between the two paired images (reference monochrome vs. spatially-degraded color). The requested response was simply to choose which of the two images they preferred (mono or color) to provide the best overall intelligence value. The specific levels of spatial degradation for the color images were selected to ensure that sufficient degradation was applied so that the monochrome reference scene would be chosen at least once by every observer. A monochrome preference was necessary to provide an estimate of the "equivalence" point in intelligence value for the color and monochrome image pair.

In designing the evaluation, three primary characteristics were of interest: (1) various levels of spatial degradation, (2) the variety of image content and (3) nadir vs. oblique acquisition perspectives. We also wanted to limit the evaluation time to not longer than one hour to ensure that observer fatigue was not an issue. These parameters and time constraints combined to produce our final experimental distribution of eight nadir images, eight oblique images, and four different levels of spatial degradation for the color image (bicubic resampling factors of 0.0, 0.7, 0.5 and 0.25). This produced a total of 64 image pairs to be viewed by each observer.

3. Results and Analysis

The two parts of this evaluation characterize different aspects of the benefit of color imagery over monochrome. The first evaluation addresses the level of improvement in both *efficiency* and *confidence* when performing detection-type tasks with color compared to black-and-white imagery. The second evaluation establishes a method for characterizing one dimension of intelligence value – that of the amount of knowledge gain associated with spectral information – by equating it with the equivalent level in the other dimension (spatial detail), as measured by the NIIRS difference between the color and monochrome images. This difference is approximated by adjusting the spatial sampling of the color images by an amount proportional to a 0.5, 1.0 and 2.0 reduction in NIIRS. The two sections that follow provide a description of the results obtained from these evaluations.

Evaluation #1: Efficiency and Confidence

Three responses were requested from each observer in this portion of the evaluation – the number of objects found, their confidence that every object found was correctly identified, and their confidence that they had successfully found all the objects within the viewing time constraint of 60 seconds. To evaluate efficiency differences, the object count provided by each observer for both monochrome and color samples was compared with the count obtained through careful and detailed study of the images by the authors. The difference between the correct count and the observer's estimate was converted into a percent error, and the difference in error between monochrome and color for each observer was computed to produce an average error *change* between the monochrome and color image. This error change is shown by the solid black squares in Fig. 3 for each of the five objects. While there was significant variation in performance across observers as well as the objects being counted, the median performance improvement was found to provide about a 25 percent accuracy improvement with the color images over the same 60 second viewing period.

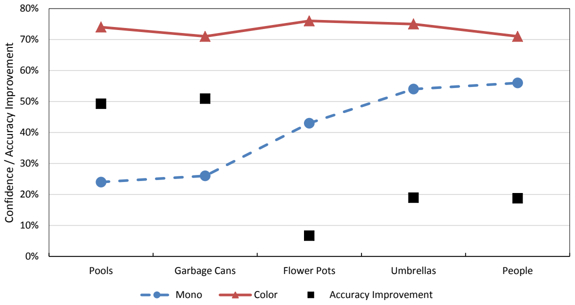

Figure 3 - Average observer confidence in correct object count for monochrome (dashed) and color (solid) viewing. The accuracy improvement in the reported number is shown by the solid squares.

We also evaluated the observer's confidence in their estimate of the total count for both monochrome and color viewing. The confidence provides an indication of the observer's belief that they were able to find all the desired objects. These results are also shown in Fig. 3; the dashed line shows the confidence for monochrome viewing, and the solid line the results for color viewing. As can be observed, the confidence improvement is greatest when the objects have some color component, such as the blue color of a swimming pool. However it is seen that confidence improved universally for all the color image presentations. Here, the median improvement was about 33 percent in confidence when viewing the color image over the monochrome version.

To assist the reader in understanding the range of variation in the raw observer data, Fig. 4 shows the individual observers' confidence *change* between color and monochrome. A value of zero on this plot indicates no change in confidence – the two images provided equivalent confidence as reported by the observer. A negative value on the plot indicates a confidence improvement with the monochrome image, and a positive value indicates a confidence improvement with color. As can be noted, only one observer when reviewing one object (umbrellas) reported a greater confidence using the monochrome image. For every other condition (observers and images) the color image *always* provided equal or improved confidence for all observers.

Finally, the observers were asked to provide comments related to their ability to perform the detection (counting) tasks after they had completed the monochrome phase of the evaluation, as well as after completing the color phase. These comments serve to highlight the individual observer's impressions and experiences during the evaluation:

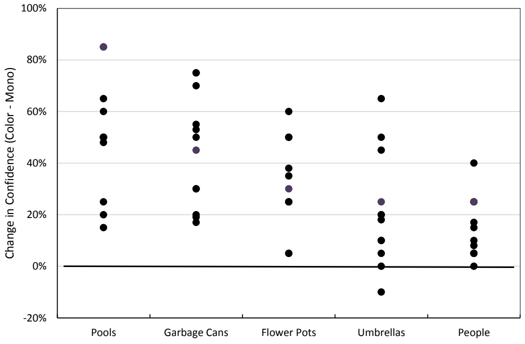

Figure 4 - Individual observer confidence differences between color and monochrome images. Note that only one observer when counting the umbrella objects reported improved confidence with monochrome.

Observer comments after monochrome image set

- *"It was difficult to say with certainty that I was identifying the correct targets."*
- *"I often looked where I 'expected' things to be but the lack of color made ID difficult."*
- *"Most objects were difficult to find. Generally low confidence that I had found all objects."*

Observer comments after color image set

- *"Targets I was unsure about became very obvious. It was much easier to quickly scan the image. The color added much more information."*
- *"Much easier to pick things out because of the color difference"*
- *"Color helps significantly with speed and recognition. More cues to look for."*

Evaluation #2: Intelligence Value of Spatially Degraded Color vs. Monochrome Imagery

The collected data was comprised of the individual responses from 16 observers evaluating four levels of color spatial degradation when viewing 16 image pairs. The response variable was the image that was preferred for each pair, either color or monochrome. The data was first analyzed to assess the correlation of responses between observers. Among the observers, one of the trained analysts selected the color image for *every* pair, which did not allow a determination of the "equivalence" point. Four other observers had selections that produced very low statistical correlation with the average of the observers, and they were also inconsistent in some of their responses – that is, making a choice of monochrome for an intermediate resolution, followed by preference for color at even lower spatial degradation. Our statistical analysis of observer correlation showed correlation values of 0.12, 0.14, 0.24 and 0.31 for the four lowest observers, and was "indeterminate" for the observer that always chose the color image. The eleven other observers were reasonably well correlated with each other, the next lowest having a correlation value of 0.61. Thus for our analysis we removed these five lowest correlating observers from the population.

This analysis attempts to quantify the subjective contribution to intelligence value from two image characteristics that are largely unrelated; (a) spatial detail and (b) features associated with the object's color. The goal is to determine the equivalent amount of spatial information (NIIRS) that is provided by the addition of the color information. All other characteristics being equal, a spatial change by a factor of two approximates a one-unit NIIRS impact. By finding the "crossover point" where the observer chooses the monochrome reference image over the spatially-degraded color image, we will have identified the "spatial-change equivalence" point where the spectral (color) information has become equivalent to the spatial contribution (NIIRS) of the reference monochrome image.

The graph shown in Fig. 5 demonstrates each observer's responses, averaged over the 16 images they evaluated. It can be observed that there is a "cluster" of observers in reasonable agreement, with two that deviate from this average; one has a preference for color (Obs. #6), the other more weighted toward monochrome (Obs. #15). The heavy dashed line shows the average for all observers. From these observer responses, we can calculate their individual "crossover" point – that is, the level of NIIRS change which correlates to a change in their choice from the color back to the monochrome image.

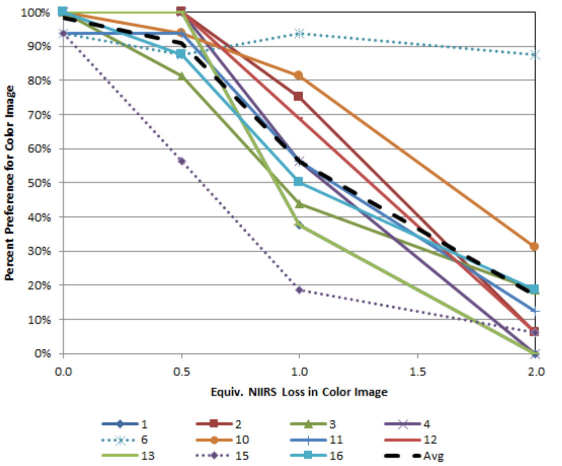

Figure 5 - Individual observer results showing preference for the color image over monochrome as a function of the amount of spatial degradation applied to the color image.

Figure 6 shows the individual observers' level at which the intelligence value is equivalent (in terms of NIIRS/spatial loss) between the lower spatial quality color image and the higher spatial quality monochrome image. We have identified two levels of "equivalence." The first point we have called "50%" which is where each observer chose half of the color images as preferred, and the other half were better in the monochrome presentation. A more conservative estimate of the performance gain by spectral information is the point where each observer preferred the color image 80 percent of the time (i.e., they nearly always preferred the color over the monochrome scene.) You will note that one observer (Obs. #6) never reached the 80% level, always preferring the color images more than 80% of the time, no matter what the spatial degradation. In this analysis, we find that the median level of "NIIRS degradation" that produced an equivalent color and monochrome image is at 1.1 "equivalent NIIRS" or eNIIRS for the 50% point, and about 0.7 eNIIRS for the 80% point.

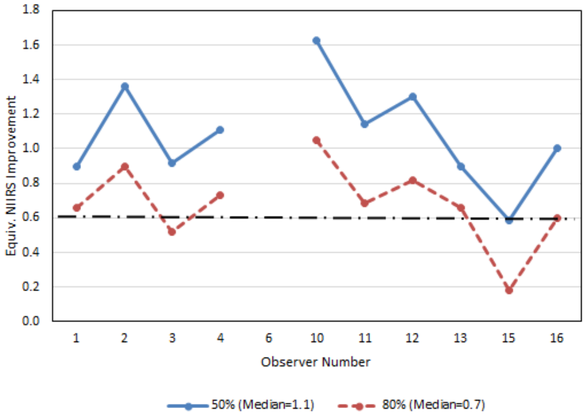

Figure 6 – Individual observer eNIIRS results when color was selected as the preferred image 50% of the time (solid line) and 80% of the time (dashed line). Average improvement: 1.1 eNIIRS (50%) and 0.7 eNIIRS (80%)

We have indicated a horizontal line on the graph in Fig. 6 at the 0.6 eNIIRS level. At this level we observe that all the individual 50% scores are at or above that level, and all but one are nearly at or above it for the more conservative 80% point. Thus, a very conservative estimate of the amount of eNIIRS gain offered by the spectral contribution from color images amounts to at least 0.6 eNIIRS, and a more reasonable average result is slightly above 1.0 eNIIRS.

4. Summary and Discussion

In this study we have introduced our approach toward quantifying the overall benefits of color over black-and-white images for visual interpretation. When we began, we recognized that this characterization would be difficult, primarily because it is impractical to evaluate performance differences on a task-by-task basis, given the variety and diversity of intelligence problems and tactical information gathering requirements that these imaging systems address. An exhaustive literature search for previous analytical work in this area yielded no significant prior research, and thus we offer this work as an introductory exploration into the subjective analysis of the multidimensional aspects of evaluating the gain in *intelligence value* offered by the combination of spectral and spatial information from color imaging systems.

The hypotheses we have explored in this study are that there are three principal areas for improvement offered by color systems over black-and-white: (1) Increased *intelligence value* of the imagery due to the spectral content, (2) increased efficiency in the observer's ability to perform their assessment tasks, and (3) improved confidence in their conclusions drawn from the information. The first part of our evaluation addressed efficiency and confidence and we have demonstrated a unified agreement from both trained image analysts and skilled image scientists that color imagery enhances detection-based tasks. In this evaluation, observers were able to more quickly identify specific targets of interest, and both their confidence and accuracy of target identification was improved about 25 to 30 percent.

The second part of this evaluation has shown that color images provide an increased information benefit that is equivalent (on average) to about a 1.0 "effective-NIIRS" improvement. It is important to recognize that this study compared only a "true-color" presentation with its monochrome countepart. In multispectral systems there are additional spectral channels that can also be used in combination to provide additional spectral information beyond the tasks we have evaluated in this study. An example would be using a near-infrared (NIR) channel to aid in identifying crop/vegetation stress based on the NIR reflectance characteristics. A true-color example would be to ask an observer to count the red cars in an image of a parking lot. These tasks add specific spectral conditions that are not easily achievable in a monochrome image. However, instead of focusing on spectrally specific problems our evaluation asked the observers to consider *all* the diverse intelligence tasks they might encounter and choose the most appropriate system, thereby removing any specific reference to spectrally-weighted information that might bias the results toward a color-based solution.

We anticipate further investigations into this area of research and hope that this paper can serve as a basis for expanding and furthering the exploration of this highly relevant psychophysical topic.

References

[1] L. A. Maver, C. D. Erdman, K. Riehl, "Imagery Interpretability Rating Scales", SID Digest, 1996, pp. 117-120

[2] "NATO Target Reporting Categories," NCAP - National Collection of Aerial Photography. [Online]. Available: http://ncap.org.uk/feature/nato-target-reporting-categories. [Accessed: 14-Nov-2017].

[3] "MassGIS Data - USGS Color Ortho Imagery," *The Official Website of the Executive Office for Administration and Finance*, 04-Nov-2016. [Online]. Available: http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/colororthos2013.html#maincontent. [Accessed: 10-Oct-2017].

[4] K. Adelman and G. Adelman, "Aerial Photographs of the California Coastline," California Coastal Records Project, 2002. [Online]. Available: http://www.californiacoastline.org/. [Accessed: 14-Nov-2017].
Note: All images Copyright (C) 2002-2018 Kenneth & Gabrielle Adelman, California Coastal Records Project, www.californiacoastline.org

Author Biography

Cicely DiPaulo is an Image Science Engineer with UTC Aerospace Systems in Westford, MA, where she has worked since graduating in 2015 from RIT with a BS degree in Imaging Science. She has been primarily involved in image analytics and image quality assessments for the UTC domestic aerial reconnaissance systems. Her most recent project involves subjective visual comparisons of the intelligence-related performance between multispectral systems and monochrome/black-and-white sensors. This area of psychometric scaling has attracted her technical focus and she hopes to continue to develop her expertise in this area in the future.

Larry Scarff is currently an image scientist with UTC Aerospace Systems in Westford, MA specializing in optimizing image information and quantifying knowledge gain via the human visual system augmented by automated analysis. Larry's career in image science has covered a variety of technologies and imaging fields; intelligence and reconnaissance, medical imaging, diagnostic radiology, and consumer-focused mobile imaging applications.