

Multispectral Image Segmentation

Roarke Horstmeyer, MIT Media Lab
6.869 Final Project, Spring 2010

Abstract

Cameras using spectral filters that are different than the typical RGB Bayer pattern can extend segmentation algorithm performance. Conventional color-based segmentation, often achieved by clustering perceptually similar values in a two or three dimensional space, suffers from metamers, luminance variation and a low resolution for bright colors. This paper demonstrates how increasing the number of spectral samples in the visible range with a multispectral camera can improve K-means clustering performance. It furthermore demonstrates how extending spectral samples outside of the visible range easily separates otherwise invisible clusters. A novel algorithm is developed to combine the perceptual-based color segmentation used on RGB images with the statistical clustering methods of hyperspectral data. Finally, initial work in determining an optimal set of color filters for segmentation is presented with the goal of designing a learning algorithm for this task in the future.

1. Introduction

A conventional digital color camera contains an interlaced set of red, green, and blue filters over its pixels known as the Bayer pattern. These three filters, close in function to the three color-sensitive cones in human eyes, allow for an image to be reconstructed faithfully on many devices in a number of standard color spaces [1]. Each of the three filters non-uniformly samples a range of overlapping wavelength values ranging roughly from 400nm-750nm. While only these three values can produce a large percentage of colors recognizable to the human eye, such a small projection basis also has its limitations, including metamers, limited spectrum coverage and a low distinction of bright colors.

Segmenting objects in a scene through the clustering of properties like color, texture, and brightness has been studied since the early days of the computer vision field, and is of current interest for applications like matting, object identification and image search, among others. Common color segmentation algorithms like K-means [2] and Mean-Shift [3] often transfer RGB values to a

perceptually uniform space like LAB and LUV. The non-linear transfer process removes a luminance term (L) from a two dimensional color space based on opponent colors, in the case of LAB, of red-green (A) and blue-yellow (B). In this perceptually uniform space, a Euclidean metric can be used to define cluster distances in the two color dimensions for simple segmentation.

Since multispectral cameras are often *not* designed to capture images for human consumption, they typically use a unique application-specific spectral filter basis. Examples range from blood oxygenation detection [4] to coral reef deterioration measurements [5]. Segmentation is an equally important goal in these applications, especially with hyperspectral images where data reduction is paramount. Most segmentation examples from this field work in a high-dimensional space with statistical measures of similarity like correlation and spectral angle, looking for features and paying little attention to human perceptual cues.

This paper attempts to borrow concepts from both of these extremes, while exploring what sort of spectral filtering scheme yields the most salient color segmentation results for a given number of filters. Examples of improving results from the LAB color space will first be presented for added filters outside the visible region, and second for a denser sampling scheme within the visible region. Then, the advantages gained in segmenting in the LAB color space will be applied to improve results from a uniformly sampled multispectral camera image. Finally, the question of optimizing a filter set for the task of segmentation will be discussed with example scenarios presented for non-linear spectral mixing schemes.

2. Background

Image segmentation through feature clustering or spatial division is of primary importance to the computer vision field, and is continually active research area since the field's inception. Likewise, clustering of similar ground cover is a high priority with hyperspectral data. Following is a summary of advances in segmentation in these two

fields, and a discussion of work with multispectral data at their intersection.

Conventional Color Segmentation

Clustering and segmentation functions fall into two basic categories: agglomerative and divisive [2]. Typically, color is used along with other cues like texture, brightness and energy to create a feature space, where like clusters form [6, 3]. K-means and Mean-shift [3] are two widely used agglomeration methods to group these clusters. While performance of these clustering methods has been analyzed in different color spaces like RGB [7], HSV [8,9], and LAB [9], the general consensus implies that removing the effects of illumination variations yield optimal segmentation results [1]. This paper will take a close look at how the LAB color space is used to improve segmentation and clustering results with K-Means and Mean Shift, and will then apply concepts common to the hyperspectral community to explore an optimal segmentation algorithm and filtering scheme for multispectral camera setups.

Hyperspectral Segmentation

Hyperspectral cameras can provide hundreds of non-overlapping spectral channels of a given scene, typically from overhead. Clustering is one method of reducing the size of these large data sets. Specifically, clustering can be defined as supervised, in which the various possible spectra in the scene are known, and unsupervised, which we will be considering in this paper. Standard unsupervised segmentation methods include ISODATA and the chain method [10]. More recent trends consider combining measures of texture along with spectral clustering [11], multiscale analysis [12], or attempt to find illumination-invariant classifiers for a given type of ground cover [13]. This paper will also explore schemes in which scenes can be treated in a brightness invariant manner, but will borrow from conventional segmentation schemes to do so.

Multispectral Segmentation

Multispectral images, which lie somewhere between the tri-filtered scheme of the Bayer pattern and the hundreds of spectral channels acquired with hyperspectral cameras, borrow segmentation methods from both of these distinct fields. Examples of hybrid segmenting approaches are in [11, 14]. Tricks can be played with illumination [15] or new pixel filter arrays [16] to assist in capture. Yasuma et al. previously implemented an optimization procedure in search of a specific multispectral filter set, but used metrics concerned with pixel spacing and size [17]. This paper aims to examine a similar optimization, but doing so from the perspective of improving cluster segmentation.

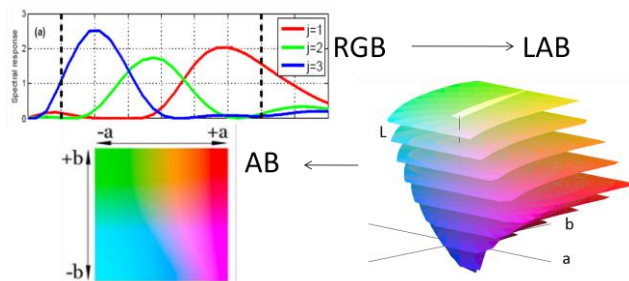


Figure 1: The transfer process from RGB values to AB space. (a) Typical R,G, and B curves of sensitivity vs. wavelength from a digital sensor, where the dotted lines show the 400nm and 700nm cutoff. A 3D representation of LAB values [18], transferred by a non-linear set of equations which can be found in [1]. With the brightness axis L removed, this becomes a perceptually uniform 2D space where clustering can be performed.

3. Perceptual and Shape-based Clustering

The optimal clustering space and distance metric for color segmentation depends upon the spectral filtering scheme used to capture the image. As mentioned in Section 2, the CIE-based LAB (or LUV) color spaces have been shown to be optimal for RGB images, and higher-dimensional spaces are used when more color channels are available. In this section, we will describe how segmentation is achieved with these two types of images, and more importantly, why these different segmentation schemes are believed optimal.

3.1. The Benefit of Removing Brightness

RGB images are transferred to what is known as a perceptually uniform color space like LAB or LUV through a non-linear transformation (Figure 1). This transformation maps points in the 3D non-orthogonal RGB space to a different 3D space, where one axis contains values of “lightness”, on ranges from red values to green values, and the third axis ranges from blue to yellow. The exact nature of this non-linear transformation was derived through human experiment, as the two color axes were chosen to mimic the function of the opponent color process in the human eye [1].

As a result of this transformation, perceptually salient color values can be *approximately* treated using the Minkowski distance metric d_k in LAB space:

$$d_k = \sum_{i=1}^N [(A_i - B_i)^k]^{1/k}, \quad (1)$$

where A and B are the two spectral vectors to be compared and k typically equals 2 (for Euclidean distances). This is the basic similarity measure used in K-means clustering,

where points are assigned to a certain cluster whose center is updated iteratively [2]. Besides providing a convenient distance metric, the separation of a brightness measure onto the L-axis allows for an increasingly brightness-invariant clustering to be performed in AB-space. While it may seem counter-intuitive, working in this two instead of three-dimensional space yields more compact segmentation results, and may share similarities with how humans classify individual colors into groups [19].

3.2. Added Dimensions for Spectral Shape

If one is provided data with more than three (RGB) color channels to work with, then it would be intuitive to assume that performing segmentation in a higher-dimensional space may yield more accurate results. However, the number of optimal dimensions is not obvious. This point will be explored in more detail in Section 4.2. Many standard multispectral segmentation schemes of N channels simply work in an N -D space. One common measure of *magnitude* similarity between two spectrums A and B in this space is correlation,

$$C = \frac{\sum_{i=1}^N (A_i - \mu_A)(B_i - \mu_B)}{(N - 1)\sigma_A\sigma_B}, \quad (2)$$

Where μ and σ are mean and variance. Likewise, a useful measure of similarity in *shape* is given with the N -D measurement of spectral angle,

$$\theta = \cos^{-1} \frac{A^T B}{\|A\| \|B\|}, \quad (3)$$

where A and B are treated here as vectors. As a vector is completely defined by its magnitude and shape, the above two measurements of similarity in an N -D space should provide complimentary and complete similarity measurements for clustering. Generally, illumination variations within an object cause this assumption to be false. We will show this by applying correlation and spectral angle to measure cluster similarity separately with the K-means algorithm.

4. Segmentation with Multiple Spectral Bands

Given that (1) is used as a distance metric to segment RGB images with K-means, while (2) and (3) are used to do the same with hyperspectral images, we will show how multispectral image segmentation can borrow from both approaches to enhance clustering performance. Clustering results in this section are generated using MATLAB’s “kmeans” iterative algorithm as a base. Color transformations from RGB to LAB are likewise performed under MATLAB’s built in “makecform” and “applycform” color functions. An extension to a MATLAB implementation of mean shift from [20] also demonstrated similar results.

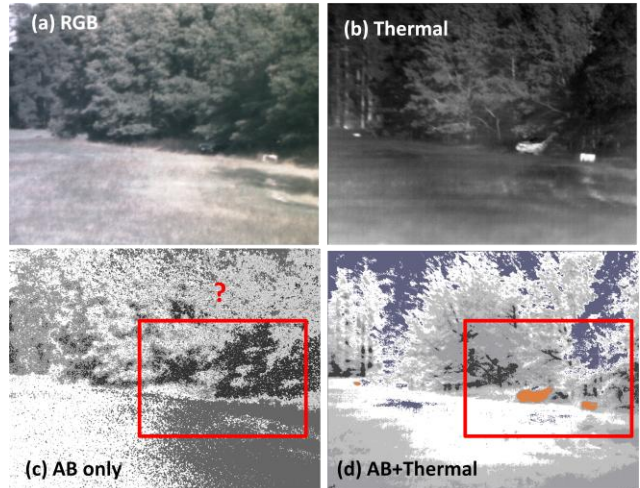


Figure 2: Color segmentation results for an RGB image with a thermal channel. The RGB image (a) is moved to AB space where a Euclidean distance metric is used for K-means segmentation into 5 clusters in (c). The car hidden in the forest is not segmented with just visible information. The addition of a thermal image (b) allows clustering into 5 groups in a 3D space to produce (d), where the car and metallic box clearly become visible. Their cluster is colored orange for clarity.

4.1. Bands Outside of the Visible

As an initial demonstration, we can extend a conventional RGB camera to include a fourth channel outside of the visible range. An example of this is in [21], where a thermal channel ranging from 8-12 μ is coaxially combined. Although no perceptual model exists to describe how this thermal channel should combine with three visible channels, Figure 2 makes it quite clear that it offers an additional layer of information (an otherwise hidden car displays strong thermal contrast). Simply appending the thermal channel as a 3rd orthogonal dimension to the opponent color-based A-B space allows K-means to successfully segment the objects of interest using (1) as a metric. This is a clear demonstration of a multispectral camera providing more information, and thus better results, to a conventional segmentation scheme.

4.2. Additional Bands within the Visible

Since the above demonstration relies on additional information *outside* the RGB spectrum, can multispectral content offer superior segmentation results strictly *within* the band of visible light? To answer this question, we use a dataset of 30 narrowband images provided by [22], where each image is taken over a different 10nm slice of the visible spectrum ranging from 400-700nm. This dataset allows us to simulate the response of a theoretical filter set simply by adding different narrowband images together. First, to determine if more visible bands yield

better clusters, we bin the 30 images into even groups, simulating the response of 2-15 non-overlapping, equally wide filters. Then, the shape-based spectral angle similarity metric is used with K-means to generate the different clustered images in each of these different dimensional spaces. From Figure 3, we can see an increase in segment uniformity when an increasing number of filters are used.

To characterize this uniformity, two previously suggested metrics [23] are applied to the clustered images from different filter sets. The first metric is a measurement of cluster compactness V :

$$V = \sum_{i=1}^N (A_i - C_{k(i)})^2 / N, \quad (4)$$

where A is the spectral vector and C_k is the k^{th} cluster center. The second metric is a measure of continuity,

$$D_i \propto \frac{\# \text{ adjacent pixels } j \text{ where } A_i \neq A_j}{\# \text{ adjacent pixels}}. \quad (5)$$

These two metrics are plotted against the number of filters used for K-means segmentation in Figure 4. Here we can see a clear increase in performance, but a non-linear tapering off after approximately 8 filters used.

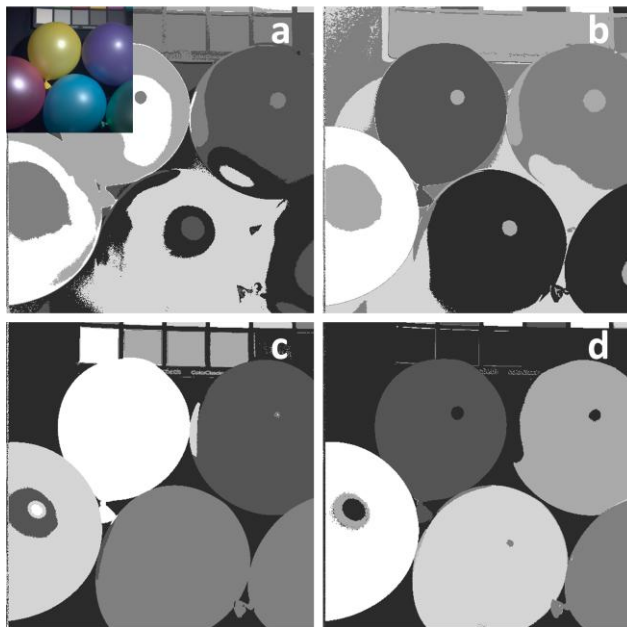


Figure 3: Example segmentation results for different numbers of binned spectral filters. (a) The visible spectrum is split up into *two* filtered segments: 400-550nm, and 550-700nm. K-means clustering with spectral angle is performed in this 2D space. The same is performed for (b) *four* filtered segments in 4D space and (c) *six* segments in 6D space. (d) 30 filtered channels are combined to simulate an RGB filtered image, then transferred and segmented (Euclidean) in AB space, for comparison.

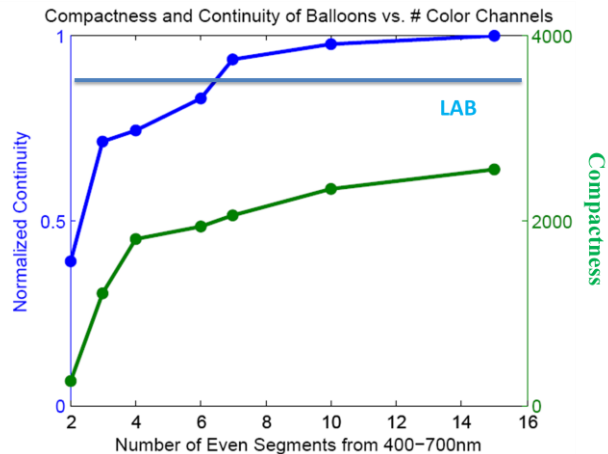


Figure 4: Continuity (left axis) and compactness (right axis) vs. # of simulated color filters for multispectral image clustering. K-means is used to create the 6-clustered images in Figure 3 from an original set of 30 narrowband images of the balloons. The 30 images are binned into a different number of equally wide, non-overlapping segments for each data point along the x-axis.

4.3. A Joint Perceptual-Shape Algorithm

Comparing the segmentation performance of 6 separate spectral channels in Fig. 3(c) and 2 non-linear A-B channels in Fig. 3(d) shows the utility of working in a space with perceptual lightness removed. Both results are nearly equivalent in dividing the specular balloons, but the A-B segmentation result requires 66% less data. However, the 6-channel result is able to segment the grayscale chart in the background of (c), which is not possible in A-B space in (d) because it is brightness-invariant. Based on these mutual benefits, a joint segmentation approach that takes the best of both worlds will be pursued.

In general, dimension reduction should not immediately be connected with a loss of segment quality. Many previous detection schemes not based on perceptual cues (starting with [24]) suggest accurate reductions from hundreds of spectral reflectance channels to as few as three dimensions (e.g. [12,14]). Therefore, it should be possible to accurately combine a higher-dimensional shape or correlation-based segmentation from a multispectral setup with these reduced-space guesses.

An algorithm demonstrating this concept is presented in Figure 5. A simulated multispectral image with 7 separate, equally wide (~40nm) channels is used. First, the algorithm clusters in 7D space with a spectral angle (or correlation) measure. Then, the 7 channels are combined to simulate an RGB-filtered image estimate, attempting to roughly follow the three curves in Fig. 1(a). This estimate is then transferred to 2D A-B-space where familiar perceptual-based clustering is applied. Note from the

above paragraph that the exact A-B-space works well but is not required – other direct low-dimensional representations yield similar results. Finally, these two shape-based and perceptual-based estimates are combined in-phase for a final result. In-phase combination can be achieved through a variety of different methods, including re-clustering in a new space or a weighted shift of the multiple cluster centers C_i ,

$$T_c \propto \sum_{i=1}^n w_i C_i, \quad (6)$$

where T_c is the new center of the c^{th} cluster. Weights can be found with a contrast maximization function or metrics like in (4) and (5). Both approaches were tested, and the former approach was adopted for Fig. 5.

Note that this combination scheme is different than the joint domain combination of [3] or additive log graph weights of [25], which consider other features besides color. Here, instead of increasing the dimensionality of a single space, we perform segmentation in two spaces of completely different dimensions and combine clusters after. To the best of our knowledge, this is a unique approach.

5. In Search of an Optimal Color Filter Basis

The previous section demonstrates how additional color channels can assist in clustering, and that combining orthogonal channels into a perceptually uniform space like LAB further improves results. This indirectly supports the hypothesis that a convex optimization scheme can find an optimal set of n color filters for color-based clustering, because i) a non-orthogonal low dimension basis set performs better than higher dimensions, and ii) this non-orthogonal set attempts to maximize a perceptual contrast function, albeit a non-linear one [1].

This optimization can be viewed from two basic vantage points. First, one can attempt to find optimal filters k_i

given the flexibility to perform any type of transformation F_i on each channel in the set before segmentation. Using contrast as a simple metric, this yields

$$S = \operatorname{argmax} \left(\sum_{i=0}^n \left(\sum_{j \neq i} \frac{F_j(k_j) - F_i(k_i)}{F_j(k_j) + F_i(k_i)} \right) \right), \quad (7)$$

where S is the optimal set of n filters defined by functions k_i . Without clear guidelines or restrictions on the generation of F , however, it is difficult to proceed far with this maximization scheme. One simple example of the almost certain non-linear behavior of F is in Figure 6. Another is the non-linear transformations from near-uniform RGB to LAB or LUV space, defined piecewise without explicit rule entirely on human experiment.

To simplify this problem, one can instead assume that clustering and segmentation will be performed in the basis of the filters to be optimized. This is similar to the operation of many hyperspectral segmentation schemes. In this case, (7) simplifies to

$$S = \operatorname{argmax} \left(\sum_{i=0}^n \left(\sum_{j \neq i} M(k_i, k_j) \right) \right), \quad (8)$$

where M can be any difference metric (contrast is a simple one) and the transformation F_i disappears. This optimization is better posed, especially since non-negativity, continuity and other restrictions can be directly applied to the spectral curves k .

In either case, an optimal spectral sampling scheme will be completely dependent upon a scene’s spectral content. Thus, the data set of natural image scenes in [22] can be used to train a system to identify optimal segmentation classifiers in the spectral domain. For this approach, a ground-truth segmentation set is needed as in [6], which is currently being generated with the multispectral data set.

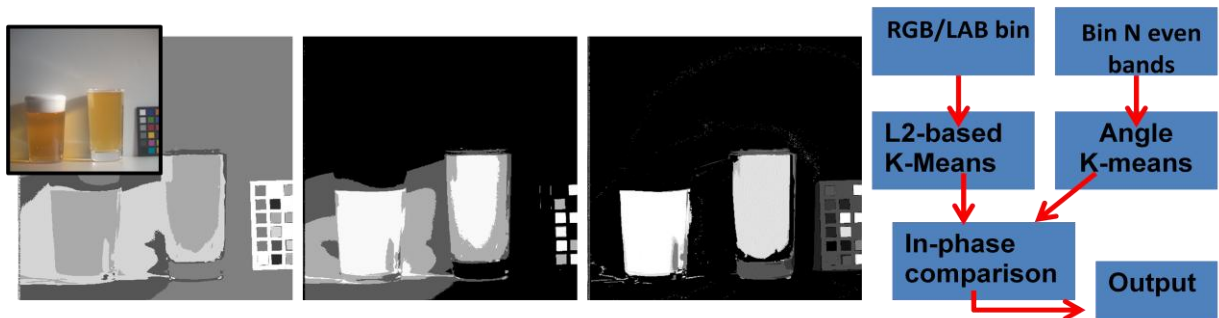


Figure 5: Modifying conventional segmentation methods to work alongside multispectral segmentation can improve clustering performance. A dataset of 30 multispectral channels of two beer cups was combined to simulate 7 equally wide non-overlapping spectral images, which was k-means segmented in 7D space into 5 clusters with spectral angle similarity (left). The 30 channels were used to simulate RGB filters, which was moved to (L)AB space for similar clustering in 2D with Euclidean similarity (middle). Combining the two approaches in-phase yields a better-segmented output (right). A schematic of the algorithm used is also shown.

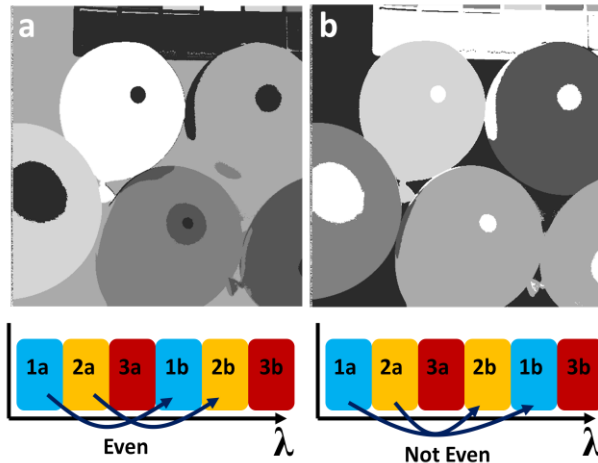


Figure 6: A demonstration of the difficulty in designing an optimal filter set when the transformation basis is included in the optimization procedure. The visible spectrum can be divided into 6 segments and transformed to a new 3D space defined with contrast functions between distinct regions 1a and 1b, 2a and 2b, and 3a and 3b, much like in LAB space. Treating the spectrum linearly by assuming contrast varies evenly (a) produces a worse K-means segmentation result than selecting contrast pairs in a non-even manner (b) (compare middle balloon segmentation).

6. Future Work

Initial results presented in this paper indicate room for color segmentation improvement both in conventional imaging, by capturing a few additional spectral channels, and in hyperspectral imaging, by incorporating the concept of perceptually uniform space. It is clear, though, that optimization procedures presented in Section 5 include a certain degree of complexity that is not addressed with simple evaluation metrics like contrast, compactness or continuity. Human subject-based segmentation results must be used instead for data set optimization or training. Given such a data set, it appears that a search for n contrast-maximizing segmentation filter pairs would be a manageable problem. Extending this search to incorporate all possible segmentation spaces, however, appears not. Therefore, a filter-only based maximization will be the focus of future work.

References

- [1] G. Wyszecki and W.S. Stiles. Color Science, 2nd Edition. John Wiley and Sons, 1982
- [2] Forsyth and Ponce, Computer vision: a modern approach, Prentice Hall, 2002
- [3] D. Comaniciu and P. Meer, Mean Shift: A Robust Approach Toward Feature Space Analysis, IEEE Trans. PAMI 24(5), 2002
- [4] R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, J. Tanida, "Multispectral imaging using compact compound optics," Opt. Express **12**, 1643-1655 (2004).
- [5] ACR Gleason, RP Reid, Automated classification of underwater multispectral imagery for coral reef monitoring, 2005
- [6] D. Martin, C. Fowlkes and J. Malik, Learning to Detect Natural Imaging Boundaries Using Local Brightness, Color and Texture Cues, IEEE PAMI(25) 5, 2004
- [7] C.C. Brunner, A.G. Maristany, D.A. Butler, D. Vanleuween, J.W. Funck, An evaluation of colorspace for detecting defects in Douglas-fir veneer, Ind. Metrol. 2 (3 and 4) (1992) 169-184.
- [8] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in Proc. Eur. Conf. Computer Vision, Copenhagen Denmark, Jun. 2002, pp. 661-675.
- [9] L. Shamir, Human Perception-based Color Segmentation Using Fuzzy Logic, http://www.phy.mtu.edu/~lshamir/publications/fl_color.pdf
- [10] Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern Classification. New York: John Wiley & Sons.
- [11] M. Hauta-Kasari, J. Parkkinen, T. Jaakelainen and R. Lenz, Multispectral Texture Segmentation Based on the Spectral Cooccurrence Matrix, PAA (1999)2:275-284
- [12] S. Miaohong and G. Healey, Hyperspectral Texture Recognition Using a Multiscale Opponent Representation, IEEE GeoSci. and Remote Sensing, 2003
- [13] G. Healey and A Jain, Retrieving Multispectral Satellite Images Using Physics-Based Invariant Representations, IEEE PAMI 22, 1996
- [14] D. Nuzillard and C. Lazar, Partitional Clustering Techniques for Multi-Spectral Image Segmentation, J. Computers (10) 2, 2007
- [15] J. Park, M. Lee, M. Grossberg and S. Nayar, Multispectral Imaging using Multiplexed Illumination, IEEE ICCV 2006
- [16] J. Brauers and T. Aach, A color filter array based multispectral camera, Workshop Farbbildverarbeitung, 2006
- [17] F. Yasuma, T. Mitsunaga, D. Iso and S. Nayar, Generalized Assorted Pixel Camera: Post Capture Control of Resolution and Spectrum, Technical Report CUCS-061-08, 2008
- [18] G. Hoffmann, CIE Color Space, <http://www.fho-emden.de/~hoffmann/cielab03022003.pdf>
- [19] C. Healey and J. Enns, A Perceptual Colour Segmentation Algorithm, Technical Report TR-96-09, 1996
- [20] B. Shai, Mean shift in MATLAB, wisdom.weizmann.ac.il/~bagon/matlab_code/mean_shift.m
- [21] R.L. Morrison, R. Stack, R.A. Athale, and G. Euliss, B.F. Necioglu, R. Horstmeyer, Colin Reese. Dual-Band Imaging System Based on a Compact Coaxial Folded Optic Architecture. OSA COSI, 2009
- [22] S. Nayar, Multispectral Image Database, <http://www.cs.columbia.edu/CAVE/databases/multispectral/>
- [23] J. Theiler and G. Gisler, A contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation, Proc. SPIE 3159, 1997
- [24] D. Judd, D. MacAdam and G. Wyszecki, Spectral Distribution of typical daylight as a function of correlated color temperature, JOSA 54 (8) 1964
- [25] J. Shi and J. Malik, Normalized Cuts and Image Segmentation, IEEE PAMI (22) 7, 2000