Ranking Color Correction Algorithms Using Cluster Indices

Jonas Osterloff[1*], Timm Schoening[1], Melanie Bergmann[2], Jennifer M. Durden[3], Henry A. Ruhl[3], and Tim W. Nattkemper[1]

[1] Biodata Mining Group, Faculty of Technology, Bielefeld University, Bielefeld, Germany,
*`josterlo@cebitec.uni-bielefeld.de`
[2] HGF-MPG Group for Deep-Sea Ecology and Technology, Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
[3] National Oceanography Centre, Southampton, UK

Abstract. In the field of underwater imaging it is often necessary to pre-process images to homogenize image quality across a whole image transect and to compensate variations in imaging conditions. A variety of pre-processing methods have been developed in recent years to overcome different problems occurring in underwater imaging, hence performing different on different image sets. Protocols for an objective comparison and scoring of those methods are needed. Here we show how to use cluster indices to rank the differently methods regarding their performances on different image sets. Our results show different rankings for four pre-processing methods for two chosen sets of benthic images from the deep seafloor.

Keywords: underwater image analysis, pre-processing methods, comparison, cluster indices, HAUSGARTEN

1 Introduction

Automated pre-processing of underwater images for color normalization has become important in recent years. The amount of image data has increased exponentially so a manual tuning of color normalization parameters is not feasible. In addition, the growing amount of image data creates a serious bottleneck regarding the visual analysis and interpretation. Color features are important to identify objects of interest (OOI) or geologic patterns [1, 2, 3]. The use of color features either for manual or computational OOI classification demands for color constancy, and thus color normalization is required as a pre-processing step in computational underwater image analysis. Due to backscattering, turbidity and wavelength-dependent absorption, color constancy is difficult to achieve at the time of image acquisition. Thus, a subsequent computational normalization process is needed, such that objects with the same visual properties are represented with almost identical colors in the image. A number of algorithms, targeting different aspects of color restoration and normalization, have been proposed to obtain color constancy [2, 4, 5, 6]. Subjectively the performance of these algorithms varies for different images. Therefore, it is necessary to quantify the performance

of the different algorithms objectively for one complete image set and given context. In this study, cluster indices are used to characterize the results of different color constancy algorithms. These indices were originally proposed to measure the quality of a clustering result [7,8,9]. In contrast to their primary area of application (i.e. a heuristically tuning of cluster algorithm parameters to identify an optimal clustering), here they are used to assess pre-processing algorithms with a given clustering. This clustering is obtained by manually annotating a set of OOIs. The score of a cluster index then indicates the discriminability (or contrast) of different OOIs in the feature space. In this work, a set of three different pre-processing algorithms [4,5,6] that have been proposed for underwater image color normalization are compared with the original images and with a method previously developed by us [2]. The performances of the different pre-processing methods are estimated on two different image sets and their corresponding annotations.

2   Material

For this feature-based evaluation of the different pre-processing methods, two sets of images with corresponding OOI annotations were used. Both image sets were collected using cameras perpendicular orientated to the seabed to achieve evenly illuminated images.

The National Oceanography Centre Southampton UK provided the first set of 1,314 benthic images and 7,504 corresponding annotations of 44 OOI classes. The images were collected using a downward-facing camera mounted on the Autosub6000 AUV (autonomous underwater vehicle) at the Porcupine Abyssal Plain (PAP) in 2011 on research cruise RRS Discovery 377. For some OOI classes only very few annotations occurred in this set. Those very rare classes (number of annotations ≤ 10) were excluded, leaving 7,400 annotations for 27 OOI classes for the PAP image set.

The Alfred Wegener Institute (AWI) provided the second set of 70 images using the online annotation tool BIIGLE [10] to annotate 33 different OOI classes. The images were captured using a towed Ocean Floor Observation System (OFOS) at the HAUSGARTEN IV area located in the eastern Fram Strait in 2004 during expedition ARK XX/1 of the German research icebreaker Polarstern. Again rare classes (number of annotations ≤ 10) were omitted giving 17,950 annotations for 19 OOI classes.

3   Methods

3.1   Pre-Processing Methods

Automated Color Enhancement (ACE). The ACE method was originally not developed for pre-processing underwater images [11]. Nevertheless the method, used for pre-processing of underwater images, could be used to enhanced an automated fish classification system [4]. For the ACE method the CIELAB color space is used. The unsupervised color correction method can be divided into two processing steps. In the first step for each pixel p, differences of

its neighboring pixels are applied to a saturation function. These differences are weighted by their distance to the pixel p and summed. The new pixel values are scaled in the second step using a white patch grey world scaling. The fast ACE implementation by Pascal Getreuer was used here [12].

Unsupervised Color Correction (UCC). The UCC method can be divided into three processing steps [5]. In the first step the RGB color channels are equalized. In the next step, a histogram stretching is applied for each color channel individually using upper and lower limits regarding the dominant color channels. Finally a histogram stretching (clipping using lower and upper limits) using the HSI color space is applied. The stretching is carried out for each color channel individually.

Automated Underwater Image Pre-Processing (AUIP). The AUIP method comprises different processing steps [6]. First, a potential moiré effect is removed. Then, a homomorphic filtering, a wavelet denoising and an anisotropic filtering is carried out using the YCrCb color space. Finally the color mean of the image is equalized in the RGB color space. A C++ adaption of the provided Matlab code was used to compute the AUIP on the images.

Feature Space Based Illumination and Color Enhancement (fSpICE). Our previously developed fSpICE pre-processing method is using a Gaussian filter to correct the illumination variance [2]. The size of the Gaussian kernel is optimized using the manual annotations of the images. For each image set (HAUSGARTEN and PAP) an individually optimized kernel size is calculated. An illumination-corrected image is computed using the pixel-wise difference between the original and the Gaussian filtered image. To equalize the brightness of all images in one image set, for each image and each RGB channel the peak of the color histogram is shifted to the middle of the histogram using a gamma correction.

Examples of original and pre-processed images of the PAP and HAUSGARTEN image sets are shown in Figures 1 and 2.

3.2   Cluster Indices

Cluster indices were originally proposed to measure clustering results. They score the distribution of clusters in the feature space. Here they will be used to score the results of the different pre-processing methods. To this end we consider each OOI class as one cluster. Each cluster shows different characteristics regarding density condition to the applied pre-processing. The cluster indices applied to the class clusters will score the distribution of the different OOI classes in the feature space to indicate how well classes are separated in the feature space. A more distinct separation might result in a better classification of OOIs not only for an automated classification system but also for the human visual system. As it is not clearly defined what a good cluster result looks like, we use three different indices to score the different pre-processing methods.
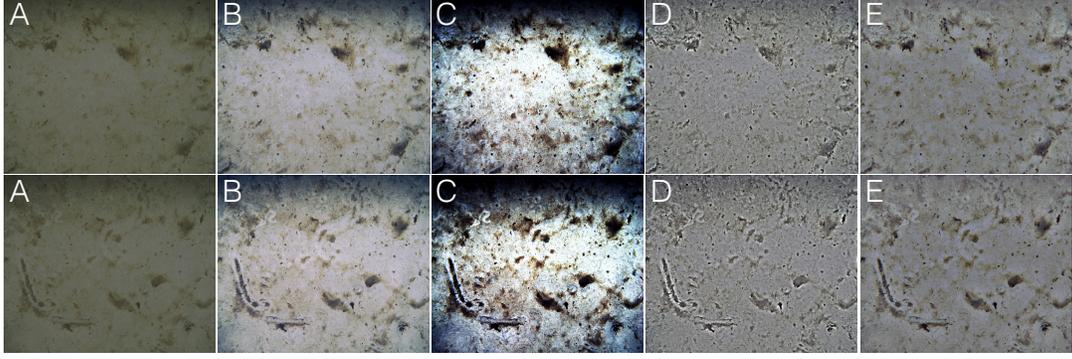
Fig. 1. Exemplary two images of the PAP dataset (rows). (A) Original images and (B-E) results after applying the pre-processing methods: (B) ACE, (C) UCC, (D) AUIP, (E) fSpICE.
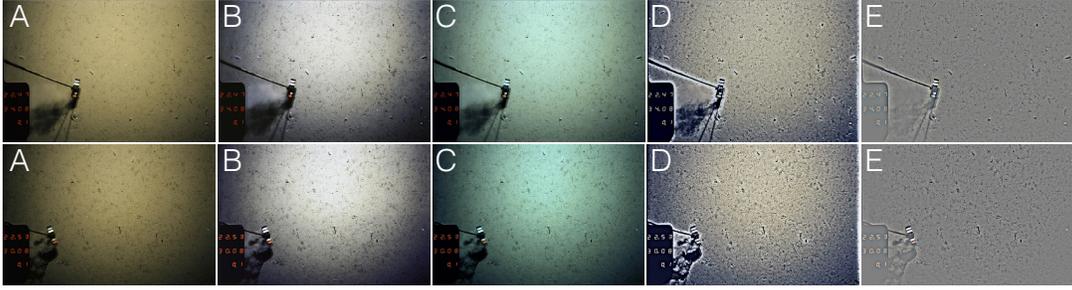
Fig. 2. Exemplary two images of the HAUSGARTEN dataset(rows). (A) Original images and (B-E) results after applying the pre-processing methods: (B) ACE, (C) UCC, (D) AUIP, (E) fSpICE.

Calinski Harabasz (CH). The CH index measures the scatter of the cluster centers in relation to the scatter within the clusters [7].

$$\chi^{\text{CH}}(K) = \frac{\sum_{k=1}^{K} |C_k| d(\boldsymbol{\mu}_k, \boldsymbol{\mu})^2 / (K-1)}{\sum_{k=1}^{K} \sum_{\mathbf{x} \in C_k} d(\mathbf{x}, \boldsymbol{\mu}_k)^2 / (L-K)} \tag{1}$$

where K is the number of clusters, $\boldsymbol{\mu}$ is the mean vector of all data vectors, $\boldsymbol{\mu}_k$ is the mean of the vectors in cluster k, $|C_k|$ is the number of vectors in cluster k and L is the total number vectors.

Index-$\mathcal{I}$ ($\mathcal{I}$). Similar to the CH, the $\mathcal{I}$ index uses the scatter within the clusters, but this time it is weighted reciprocally by the overall scatter E_1 [8].

$$\chi^{\mathcal{I}}(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^v \text{ with } K \text{ as above,} \tag{2}$$

$$E_K = \sum_{k=1}^{K} \sum_{\mathbf{x} \in C_k} d(\mathbf{x}, \boldsymbol{\mu}_k) \ , \ E_1 = \sum_{i=1}^{L} d(\mathbf{x}_i, \boldsymbol{\mu}) \tag{3}$$

and D_k the maximal pairwise cluster distance $D_K = \max\limits_{i,j=1}^{K} d(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j)$.

Davies Bouldin (DB). The DB index uses the within cluster scatter (WCS_k) to score the distribution of the clusters [9]:

$$\text{WCS}_k = \frac{1}{|C_k|} \sum_{\mathbf{x} \in C_k} d(\mathbf{x}, \boldsymbol{\mu}_k) \ . \tag{4}$$

For each cluster k the pairwise maximum of the sum of the variance of cluster k and k' weighed by the distance of the two corresponding cluster centers is computed as $R_k = \max\limits_{1 \leq k' \leq K, k' \neq k} \left\{ \frac{\text{WCS}_k + \text{WCS}_{k'}}{d_{k,k'}} \right\}$. The sum of those maxima gives the score for the index:

$$\chi^{\text{DB}}(K) = \frac{1}{K} \sum_{k=1}^{K} R_k \ . \tag{5}$$

3.3 Feature Descriptors

Features were extracted at the annotated positions to compute the different cluster indices. In this study different color feature sets were used and compared alongside the pre-processing evaluation. In general, histograms with a varying number of bins $(4, 8, 16)$ per color channel were used. To consider the individual size of the OOI, histograms were computed using different patch sizes $(9 \times 9, 17 \times 17, 33 \times 33, 65 \times 65, 129 \times 129)$. The color space used for the extraction was also varied (CIELAB, CIELUV, YCbCr, RGB) as different spaces are optimized e.g. to map distances in the color space to differences perceived by the human visual system [?]. Using a color space like the CIELAB might therefore give higher scores for pre-processing methods optimized for humans. Therefore, five different methods (original, ACE, UCC, AUIP, fSpICE) were evaluated on 60 different feature spaces generating a set of 300 results per image set.

3.4 Normalization

In order to compare different cluster index results χ^*, with $* = \{\text{CH}, \mathcal{I}, \text{DB}\}$, a normalization to the range (0,1) of all $N = 300$ results was required:

$$\hat{\chi}_n^* = \frac{\chi_n^* - \min\limits_{i=1..N} \chi_i^*}{\max\limits_{i=1..N} \chi_i^* - \min\limits_{i=1..N} \chi_i^*} \ . \tag{6}$$

In order to rank the different pre-processing methods directly the harmonic mean (H) of the three cluster indices was calculated:

$$H_n = 3 \times \frac{\hat{\chi}_n^{\text{CH}} \times \hat{\chi}_n^{\mathcal{I}} \times (1 - \hat{\chi}_n^{\text{DB}})}{\hat{\chi}_n^{\text{CH}} \times \hat{\chi}_n^{\mathcal{I}} + \hat{\chi}_n^{\text{CH}} \times (1 - \hat{\chi}_n^{\text{DB}}) + \hat{\chi}_n^{\mathcal{I}} \times (1 - \hat{\chi}_n^{\text{DB}})} \ . \tag{7}$$

The term $1 - \hat{\chi}^{\text{DB}}$ is used here as, despite the other cluster indices, the lowest value identifies the best cluster result for the DB index.

4 Results

4.1 PAP Dataset

The highest χ^{CH} index of $1,266$ for the PAP dataset was reached using the CIELAB color space, an extraction patch size of 129×129, a histogram with 4 bins per color channel and no pre-processing. The same parameter setting also resulted in the best $\chi^{\mathcal{I}}$ index of 0.0556. The lowest and therefore best value for the χ^{DB} index of 8.11 was achieved for a parameter setting using the color space CIELAB, a patch size of 33×33, a histogram with 16 bins per color channel and the fSpICE pre-processing. Looking at the H the best result was achieved using again no pre-processing, a patch size of 129×129 and a histogram with 4 bins per color channel (see Table 1).

Table 1. Extract from the results for the PAP dataset displaying the best scores for each pre-processing method. Overall best scores are highlighted in bold face.

Rank	Method	Color Space	Patch Size	Bins	χ^{CH}	$\chi^{\mathcal{I}}$ [10^{-2}]	χ^{DB}	F Score
1	original	CIELAB	129×129	4	**1266.59**	**5.56**	181.58	**1.00**
2	original	CIELAB	65×65	4	568.08	5.05	94.45	0.687
3	original	CIELAB	33×33	4	401.51	4.72	221.75	0.555
4	original	CIELAB	17×17	4	302.54	4.52	165.68	0.459
5	original	CIELAB	9×9	4	245.76	4.32	86.34	0.394
6	AUIP	RGB	17×17	4	204.16	5.13	50.48	0.354
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	fSpICE	CIELUV	129×129	4	419.53	1.67	26.63	0.324
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
111	ACE	RGB	9×9	4	43.56	4.40	39.52	0.082
112	fSpICE	CIELAB	33×33	16	45.89	0.0222	**8.11**	0.08
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
167	UCC	CIELUV	17×17	8	22.44	3.77	14.63	0.037
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

4.2 HAUSGARTEN Dataset

The best χ^{CH} index of 585 and the best $\chi^{\mathcal{I}}$ index of 0.0853 were achieved with a parameter setting using the RGB color space, a patch size of 9×9, a histogram with 8 bins per color channel and the fSpICE method. The best χ^{DB} index score of 13.12 was achieved using the RGB color space, a patch size of 17×17 and a histogram with 16 bins per color channel and again the fSpICE method (see Table 2). The best score for H was reached by the fSpICE method. The highest H for the remaining color spaces were also achieved by the fSpICE method.

Table 2. Extract from the results for the HAUSGARTEN dataset displaying the best scores for each pre-processing method. Overall best scores are highlighted in bold face.

Rank	Method	Color Space	Patch Size	Bins	χ^{CH}	$\chi^{\mathcal{I}}$ [10^{-2}]	χ^{DB}	H
1	fSpICE	RGB	9×9	8	**585.94**	**8.53**	30.24	**0.990**
2	fSpICE	RGB	9×9	16	584.34	8.37	23.93	0.986
3	fSpICE	RGB	9×9	4	570.20	8.41	68.19	0.953
4	fSpICE	CIELUV	9×9	8	433.80	7.56	35.93	0.849
5	fSpICE	RGB	17×17	16	439.56	6.71	**13.12**	0.824
6	ACE	CIELUV	9×9	8	401.07	7.01	50.83	0.795
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
8	AUIP	RGB	9×9	4	339.49	8.12	44.60	0.780
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
38	UCC	CIELAB	9×9	8	144.56	6.55	41.58	0.460
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
58	original	CIELAB	9×9	16	122.68	6.15	61.96	0.405
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$

5 Discussion

The best results for the $\chi^{\mathrm{CH}}, \chi^{\mathcal{I}}$ and H on the PAP dataset were reached using no pre-processing and the CIELAB color space for the feature extraction. Surprisingly, this suggests that the tested pre-processing methods for the PAP dataset will not improve the classification of the OOIs by machines or the humans. By contrast, the fSpICE method clearly outperforms the other pre-processing methods applied to the HAUSGARTEN dataset. The 50 highest values for H were achieved using a pre-processing method and not the original images. This indicates that a good pre-processing method will boost the classification results for the HAUSGARTEN images. Comparing the two datasets visually, the PAP images do not suffer from a variation of the illumination cone as strongly as the HAUSGARTEN images (see Figures 1 and 2). The PAP images were taken with an AUV which operates at a more stable altitude compared with HAUS-GARTEN images which were obtained by the towed OFOS which is affected by waves and swell (see Material section). The information provided about the imaged area for the two sets proves this impression. The area captured in each PAP image varies between $1.82m^2$ and $2.44m^2$ with a variance of 0.005 and for the HAUSGARTEN set between $3.51m^2$ and $4.49m^2$ with a variance of 0.082. This could explain why a pre-processing for the PAP images seems not as essential as for the HAUSGARTEN images. The image quality in terms of stable conditions and scene illumination during image capture is the crucial factor here determining the need for pre-processing.

The use of cluster indices allows a quantification of the performance of different pre-processing methods in a simple and fast manner. The time-consuming training and tuning of an automated classification system will benefit from this. It now only has to be carried out for the best pre-processing method. The anno-

tations used for the pre-processing method ranking can be reused for the training and tuning of the classification system. In the future, our approach should be tested on different annotated image sets. In addition, we have to test if the cluster index scores reflect the performance of classification systems as expected.

References

1. Schoening T, Steinbrink B, Brün D, Kuhn T, Nattkemper TW (2013) Ultra-fast segmentation and quantification of poly-metallic nodule coverage in high-resolution digital images. Proceedings of the UMI 2013 .
2. Schoening T, Bergmann M, Ontrup J, Taylor J, Dannheim J, et al. (2012) Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN. PloS One 7: e38179.
3. Purser A, Ontrup J, Schoening T, Thomsen L, Tong R, et al. (2013) Microhabitat and shrimp abundance within a Norwegian cold-water coral ecosystem. Biogeosciences Discussions 10.
4. Chambah M, Semani D, Renouf A, Courtellemont P, Rizzi A (2003) Underwater color constancy: enhancement of automatic live fish recognition. In: Electronic Imaging 2004. International Society for Optics and Photonics, pp. 157–168.
5. Iqbal K, Odetayo M, James A, Salam RA, Talib A (2010) Enhancing the low quality images using unsupervised colour correction method. In: 2010 IEEE International Conference on Systems Man and Cybernetics (SMC). IEEE, pp. 1703–1709.
6. Bazeille S, Quidu I, Jaulin L, Malkasse JP, et al. (2006) Automatic underwater image pre-processing. Proceedings of CMM'06 .
7. Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. Communications in Statistics-theory and Methods 3: 1–27.
8. Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence 24: 1650–1654.
9. Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence : 224–227.
10. Ontrup J, Ehnert N, Bergmann M, Nattkemper TW (2009) BIIGLE-Web 2.0 enabled labelling and exploring of images from the Arctic deep-sea observatory HAUSGARTEN. In: OCEANS 2009-EUROPE. IEEE, pp. 1–7.
11. Rizzi A, Marini D, Gatta C (2004) From retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalization. Journal of Electronic Imaging 13: 75–84.
12. Getreuer P (2012) Automatic Color Enhancement (ACE) and its fast implementation. Image Processing On Line 2: 266–277.