Journal of Computational Mathematics and Data Science 13 (2024) 100104

Contents lists available at ScienceDirect

Journal of Computational Mathematics and Data Science

journal homepage: www.elsevier.com/locate/jcmds

Novel color space representation extracted by NMF to segment a color image

Ciro Castiello [a], Nicoletta Del Buono [b,*], Flavia Esposito [b]

[a] Department of Computer Science, University of Bari Aldo Moro, Bari, Italy
[b] Department of Mathematics, University of Bari Aldo Moro, Bari, Italy

ARTICLE INFO

Keywords:
Color spaces
Low rank data representation
Feature extraction
Machine learning algorithm
Nonnegative matrix factorization
Image segmentation

ABSTRACT

This paper considers the task of separating pixels in color image into background and foreground classes. Using the machine learning technique known as Nonnegative Matrix Factorization, data pertaining to different color channels – selected by color spaces – are combined, and a novel space representation is extracted.

The novel representation of the image includes additional information, namely "metacolor", which could be related to foreground and background and adopted to improve binary segmentation of the investigated image. In both qualitative and quantitative experiments, the use of novel color space representation produces some improvements in the binary segmentation results when it compared to those obtained applying common simpler thresholding algorithms directly to the original image.

1. Introduction

Several techniques have been developed in literature to individuate image distinct areas or objects according to some image characteristics (as pixel intensity values). Image thresholding is the most basic technique for segmenting images: a binary segmentation can be obtained by simply assigning each pixel to a class by comparing its color intensity with a threshold value [1]. The basic idea in thresholding is that homogeneous areas inside an image correspond to a class of pixels possessing comparable color characteristics. Image thresholding techniques are based on the construction of histograms illustrating the distribution of colors in an image and the detecting a threshold used to segment the given image. Gray-level image segmentation is often used after reducing the color image to a single gray channel. This facilitates the segmentation process assigning all pixels having grey levels lower than the threshold to one set or class and pixels having gray levels higher than the threshold to another set or class. Color images can benefit from techniques developed for grayscale image representation: a histogram is used to segment each component of the color space, and the results are merged to create a final segmentation. [2–4]. However, selecting a threshold within these histograms is challenging [5], and such techniques cannot be always successfully applied, because designed to process only specific color properties [6]. Segmentation results are determined by the three-dimensional nature of the color information used to describe the image. Different color spaces (of which RGB is arguably the most common) can be used in this context to create different nuances of image representations. But referring to a particular color space that can be applied to each segmentation task is not practicable.

In this paper, we propose to use a machine learning technique, i.e. an unsupervised learning method known as Nonnegative Matrix Factorization (NMF), to extract new features from different color image representations to provide a new color space. These new features, called "metacolors", are automatically extracted from a selection of color models of the given image and are then used to perform a binary segmentation of the image.

* Corresponding author.
E-mail addresses: ciro.castiello@uniba.it (C. Castiello), nicoletta.delbuono@uniba.it (N. Del Buono), flavia.esposito@uniba.it (F. Esposito).

https://doi.org/10.1016/j.jcmds.2024.100104
Received 1 July 2024; Received in revised form 15 September 2024; Accepted 6 October 2024

In general, NMF provides a low-rank approximation of a given nonnegative data matrix $A \in \mathbb{R}_+^{m\times n}$, by extracting the columns of the factor $W \in \mathbb{R}_+^{m\times r}$, and the encoding coefficients, that are rows of the factor $E \in \mathbb{R}_+^{r\times n}$ (with $r \leq mn/(m+n)$, where r is generally user-defined), in such a way that $A \approx WE$. One key feature of NMF is its ability to automatically extract parts with some level of sparsity and easy interpretable meaning. Since early works [7,8], where several images representing human faces were decomposed into parts-based representations roughly corresponding to items intelligible by human intuition (noses, mouths, eyes, etc.), NMF has been broadly used in many successful applications, including blind source separation, document clustering, object recognition, hyperspectral image analysis [9,10]. In the context of image processing, NMF demonstrated its usefulness in several tasks: identifying regions or objects into images [11]; improving the representation of positive local data (such as color histograms) [12]; encoding color channels for face recognition task [13]; processing three color channels simultaneously for generating new basis of images [14]. In the work [15], a numerical method based on NMF was already proposed to integrate histograms of different color spaces characterizing a given image; the additional knowledge represented by the so called metahistograms was used as input of the Otsu's thresholding methods demonstrating some improvement in the final segmentation results. Moving from this idea [15], the aim of this paper is illustrating a novel framework which starts from the integration of different color channels into a structured representation, and then extract a new color space representation of the investigated image by means of NMF. This new representation encodes image properties, namely metacolors, which aggregate the expression of multiple color characteristics of an image into a more compact form while allowing an interpretation in terms of basis vectors for the latent color space representation of the image. In the proposed framework the novel space representation in then used to feed a thresholding segmentation algorithm to obtain a segmented image into its background and foreground pixels.

2. Adopted color spaces and thresholding algorithms

In the following sections, we briefly review some color spaces commonly used to parameterize a color image I, as well as the some common thresholding segmentation algorithms. Next, NMF is shown as a method for integrating different representations of the image and for extracting a novel color space that can be utilized for segmenting I.

When color information needs to be formalized, a common practice is to define a color space, i.e., a representation based on a three-dimensional numerical vector where each value designates the amount of the specific attribute of color pixels in I [16].

The RGB space is based on the trichromatic theory [17] that relies on the idea that all colors can be represented using different shades of the three color channels R (red), G (green), and B (blue). It commonly represents a color image I, and from some linearly or nonlinearly transformations it generates other color spaces. In this work, together with the RGB color space, we also considered CIE XYZ, $\mathbf{YC_bC_r}$ and HSV color spaces to represent an image I.

Note that any color space can be (variously) reduced to a one-dimensional gray-scale representation, namely a vector $Gray$ of gray intensity values. Naturally, such a reduction also generates a loss in the quantity of information included in the original color image, even though it simplifies the image processing overall.

Binary thresholding techniques set a single threshold to separate foreground from background pixels according to how closely they relate to the threshold value. In the following, we overview some non-parametric histogram-based global binary thresholding algorithms which will be used in the numerical experiments.

To provide a formal illustration of the thresholding methods, a suitable scenario is formalized as follows.

Let us consider an image composed by $N = pq$ pixels. Let h be the B-bin histogram providing the pixel intensity distribution in a considered channel, that is

$$h(I) = [n_1, n_2, \dots, n_i, \dots, n_B],$$

where n_i is the number of the image pixels at color level i, and B is the number of distinct levels (e.g., if 8-bit gray-scale image representation is considered, then $B = 256$). A histogram threshold k should be set to define two classes C_0 and C_1 including the pixels whose intensity is less than k and those whose intensity ranges in $[k+1, B]$, respectively (i.e., background and foreground regions). By adopting some simplifying assumptions [18], the histograms can be normalized and regarded as probability distribution, so that $P_i = \frac{n_i}{N}$ is the probability of occurrence of the color intensity level i (observe that $N = \sum_{i=1}^{B} n_i$ is the total number of pixel of the image I). In this way, the probability distributions for C_0 and C_1 are defined as $w_0 = \sum_{i=1}^{k} P_i \quad \text{and} \quad w_1 = \sum_{i=k+1}^{B} P_i$

Well-known binary segmentation algorithms are Otsu, Kapur, Tsai and Ridler methods. Otsu algorithm implements a nonparametric and unsupervised method to automatically derive the optimal thresholding from a gray level histogram, based on discriminant analysis [18]. Kapur thresholding algorithm finds the optimal threshold k that maximizes the sum of the entropy of the two classes C_0 and C_1 while computing the probability distribution of the pixel intensities. The method considers the foreground and background of an image as two different signal sources and computes an optimal threshold maximizing the sum of the two class entropy [19]. Tsai method selects the best threshold is selected in such a way that the first three moments of the image ($m_k = \frac{1}{N}\sum_{i=1}^{B} i^k n_i, k = 1, 2, 3$) are preserved in the resultant binary image [20]. Ridler method starts with an initial guess threshold value k_0 to split the image histogram into two portions, and then iterative updates the threshold as $k_i = \frac{\mu_A + \mu_B}{2}$, being μ_A and μ_B are, respectively, the mean intensity value of the pixels below and above the current threshold [21,22]. For a more detailed description of the steps the considered algorithms perform, we refer the reader to the Appendix section.

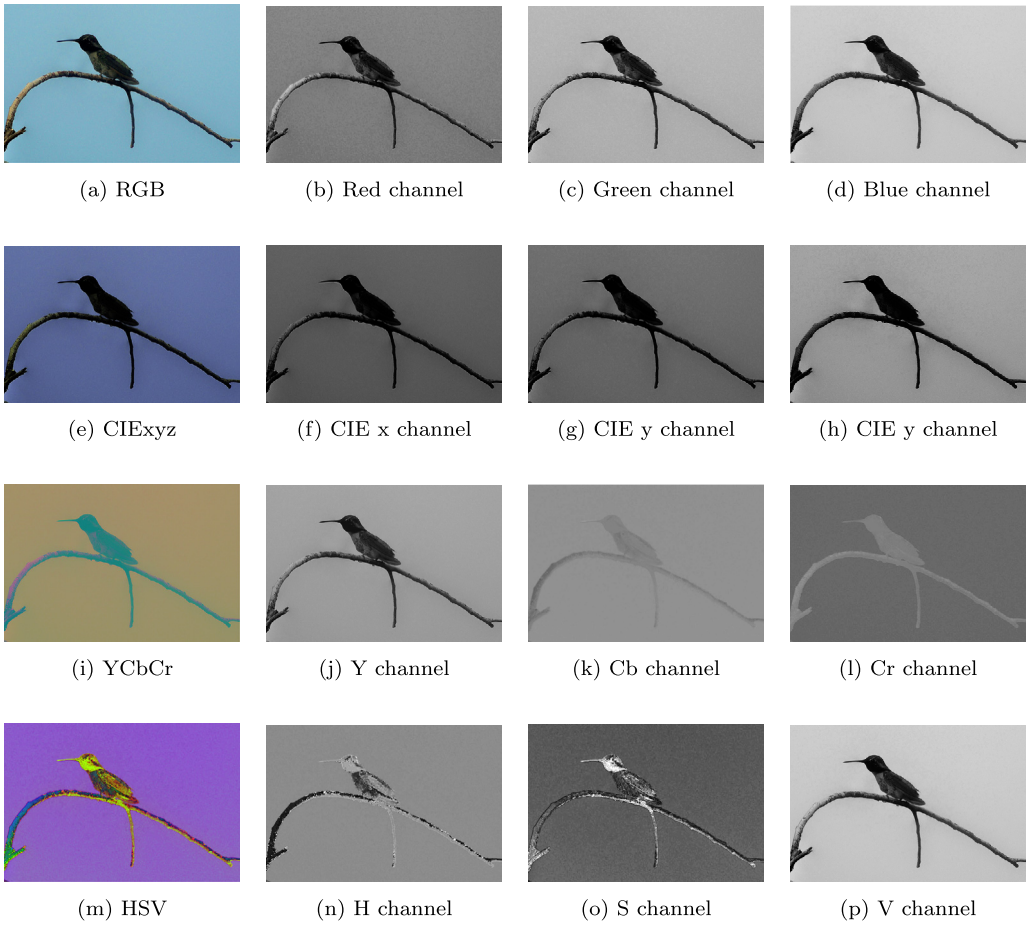


Fig. 1. Bird example image represented in the selected color spaces (from dataset PASCAL VOC2012 ID-image 2008_000123) (a–d) RBG image and the corresponding color channels R, G, and B; (e–h) CIExyz image and the corresponding color channels x, y, and z; (i–l) YCbCr image and the corresponding color channels Y, Cb, and Cr; (m–p) HVS image and the corresponding color channels H, V, and S.

3. Low-rank factorization and color channel integration

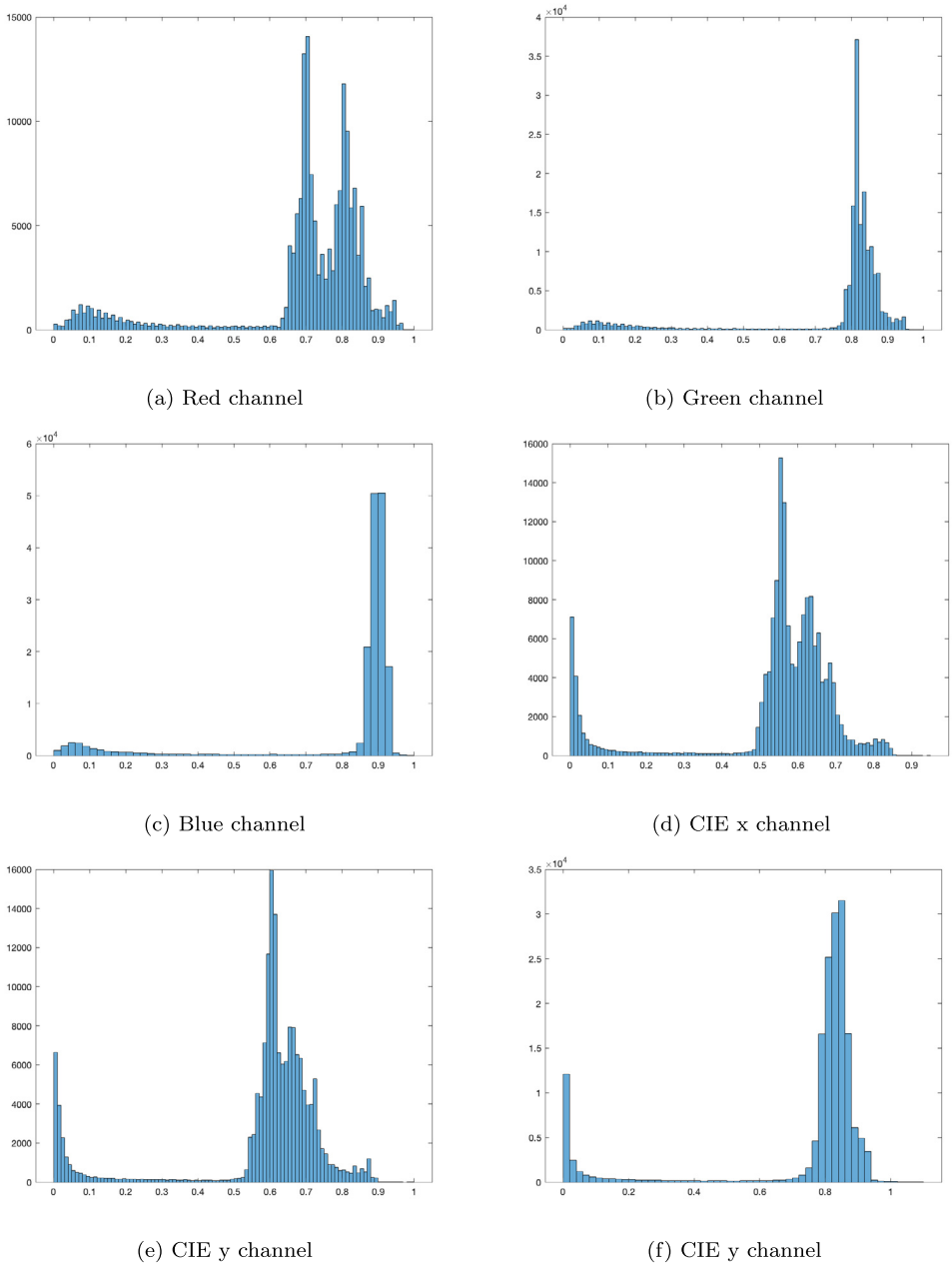
An image can be regarded as a matrix of pixels. The rows and the columns of the matrix indicate the position of each pixel in the image. The values of the pixels indicate the pixel intensity with respect to some specific color space. More precisely, each pixel incorporates three distinct values related to the channels of the color space: the single pixel intensity is a combination of the intensities in three channels.

The adopted color-space representation of an image can influence the result of image segmentation; appropriate color spaces for representing I are usually selected by testing them one by one to determine which is the most convenient for the specific task. However, no color space has proven to be more effective than the others. Thus, selecting the appropriate color space remains a challenge in color image segmentation processes. Rather than looking for the best representation of I, in this work, we propose to derive a new global representation of I by integrating different color image properties. Particularly, we integrate color channels representing I using the NMF method to extract latent features from different color space representations.

Starting with a color image $I \in \mathbb{R}^{p\times q}$ composed by $N = pq$ pixels, we can build a derived matrix A_I including the vectorization of the color intensity values of each pixel evaluated along the different color channels involved in the color spaces representation previously reviewed. In this way, we can write

$$A_I = [R, G, B, X_{CIE}, Y_{CIE}, Z_{CIE}, Y, C_b, C_r, H, S, V, Gray] \in \mathbb{R}_+^{N\times 13}. \tag{1}$$

In other words, A_I constitutes an enlarged representation of I, including the basic RGB color information, together with the device-independent imaginary primaries components (deriving from the CIE XYZ color space), the luminance and the blue and red chromaticity components (deriving from the YCbCr color space), the intensity, hue, and luminosity information (deriving from the HSV color space), and the gray-scale channel component. Fig. 1 illustrates an example image in its different color space representations, while Figs. 2 and 3 provides the histograms of pixel distributions for different color channels included in the extended image representation.



(a) Red channel (b) Green channel

(c) Blue channel (d) CIE x channel

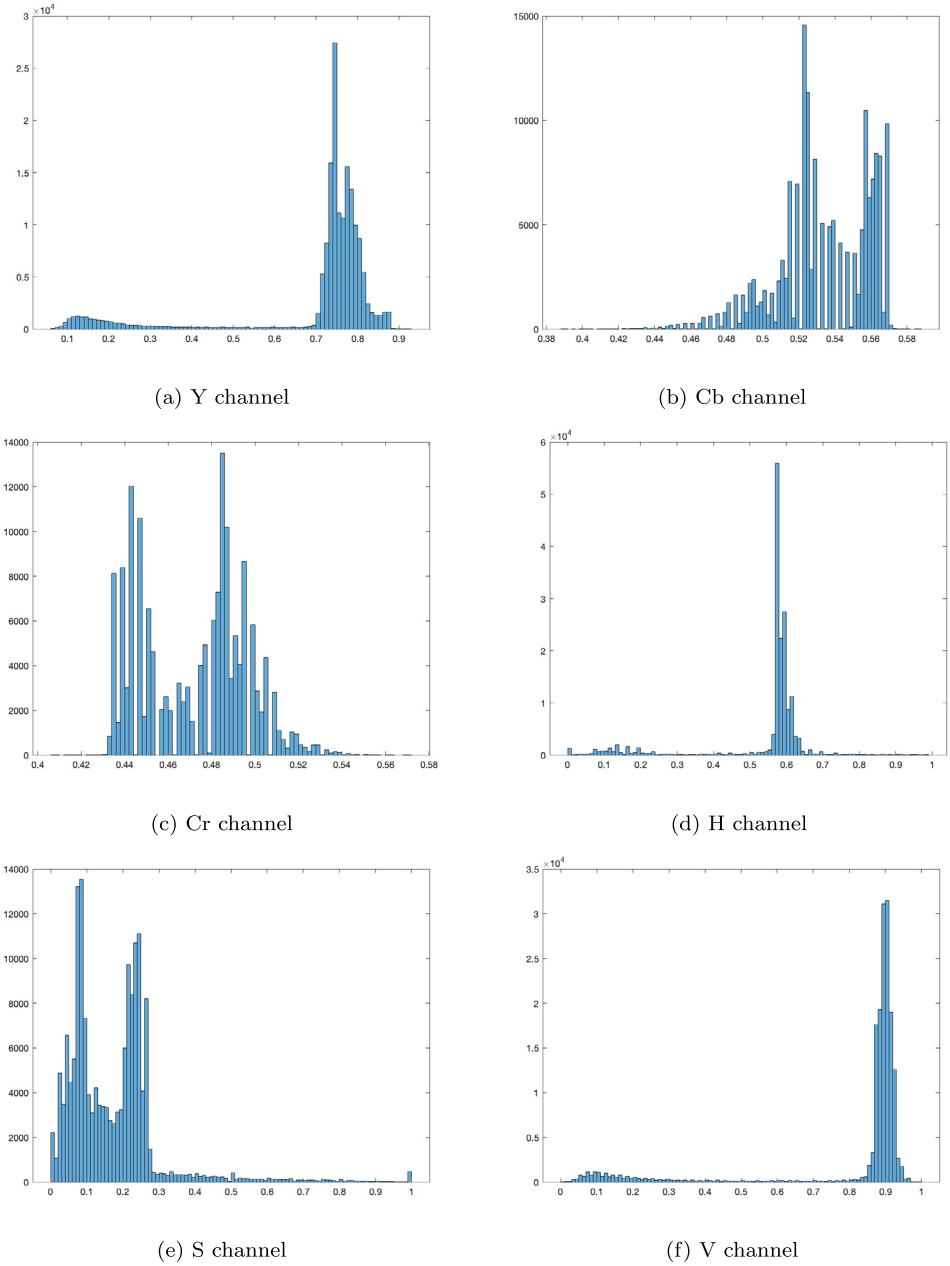
(e) CIE y channel (f) CIE y channel

Fig. 2. Histograms of the channels representing the Bird example image: (a–c) the color channels R, G; (d–f) color channels x, y, and z in CIExyz.

Due to the nonnegative property of elements in A_I, we can use the NMF algorithm [8,9] to extract new meta-features hidden into this enlarged representation of I. We indicate these meta-features as "metacolors".

3.1. Novel color space extraction from A_I

NMF is able to decompose $A_I \in \mathbb{R}_+^{N\times 13}$ into an additive combination of $W \in \mathbb{R}_+^{N\times r}$ (the basis matrix) and $E \in \mathbb{R}_+^{r\times 13}$ (the encoding factor), both with nonnegative elements, and such that $A_I \approx WE$. The rank r is the tuning parameter in NMF and its proper definition is one of the primary problems in NMF. This value is typically unknown beforehand, although it can be imposed a priori in order to move forward with the factorization process provided certain domain information is available. In this work, the rank value has been set to $r = 3$. This choice is motivated by the fact that it represents the usual dimension of the color spaces



(a) Y channel (b) Cb channel

(c) Cr channel (d) H channel

(e) S channel (f) V channel

Fig. 3. Histograms of the channels representing the Bird example image: (a–c) channels Y, Cb, and Cr; (d–f) color channels H, V, and S.

describing a color image I, so that the extracted metacolors can be regarded as the three channels of a new color space offering an alternative representation of the image I.

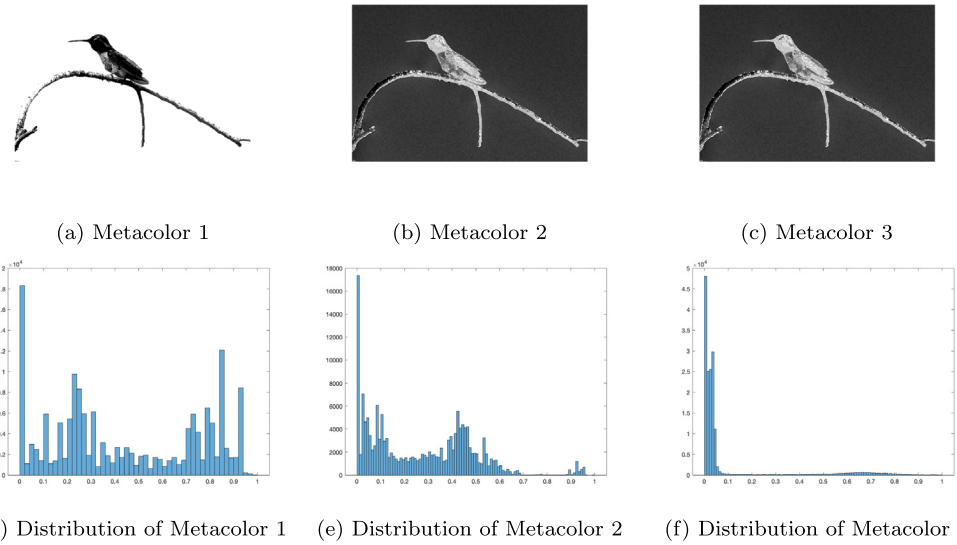
The three columns in W are an additive combination of color channels describing I so that they contain some key latent pixel features and can be considered as a novel color representation of the image. These three columns, briefly named metacolors, store all the information about the pixels of the image in a compact form with respect to the all selected original color channels.

As concerning the computation of W and E from A_I, the ANLS algorithm is used [9]. Starting from random nonnegative factors W_0 [23], ANLS approximates the solution of two convex sub-problems related to the NMF factorization of A_I solving:

$$\min_{E\geq 0} \|A_I - WE\|_F^2, \quad \text{being } W \text{ fixed}$$

and

$$\min_{W\geq 0} \|A_I^\top - E^\top W^\top\|_F^2, \quad \text{with } E \text{ fixed.}$$



(a) Metacolor 1 (b) Metacolor 2 (c) Metacolor 3

(d) Distribution of Metacolor 1 (e) Distribution of Metacolor 2 (f) Distribution of Metacolor 3

Fig. 4. Novel color space representation obtained factorizing the enlarged representation of the bird example image (rank value $r = 3$) (a-c) and the corresponding distribution histograms.

The algorithm used the following update rules [24]:

$$W_{ia}^{it+1} = W_{ia}^{it} \frac{(A_I(E^{it})^\top)_{ia}}{(W^{it}E^{it}(E^{it})^\top)_{ia}}, \quad \forall i, a \tag{2}$$

$$E_{bj}^{it+1} = E_{bj}^{it} \frac{((W^{it+1})^\top A_I)_{bj}}{((W^{it+1})^\top W^{it+1}E^{it})_{bj}}, \quad \forall b, j \tag{3}$$

where it denotes the iteration step. Updates are performed until a stopping criterion is satisfied to guarantee the final solutions are stationary points. Theoretical convergence properties of the ANLS under certain assumptions is stated in [25], while detailed description of the stopping condition is reported in [9,26].

Fig. 4 (a–c) illustrates the novel channels in the color space extracted after the factorization process and their properties (panels (d–f) depicts the distributions and ranges of the metacolors). The image example is related to the Bird image. Metacolors 2 and 3 appear quite similar, but it can be noted that they are quite different in term of value scales (panels (e–f). These differences help to better select the appropriate threshold distinguishing background from foreground pixels when thresholding segmentation is adopted.

Fig. 5 (a-c) illustrates the three novel color channels, their distributions and ranges (panels (e-f), for a more complex image example (the Baby image). In this case all the three Metacolors seem to represent different image pixel property.

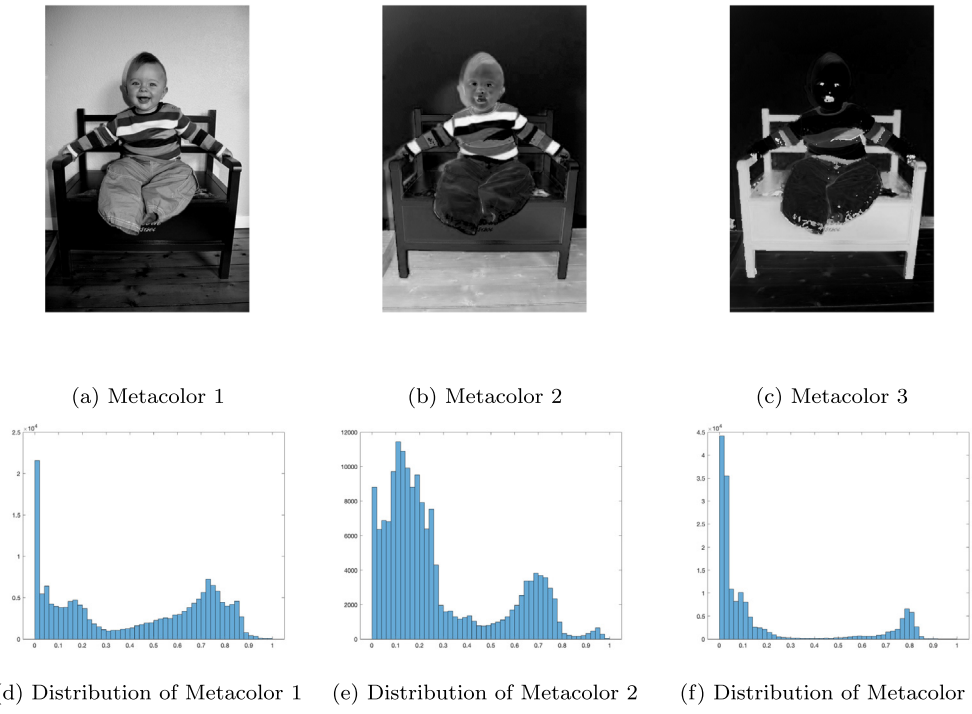
W represent the basis of latent information embedded in different color representations of I. This novel knowledge can be used either to reconstruct I or, as we propose in this work, to undergo a binary segmentation step (thresholding segmentation can be obtained using one of the previously cited algorithms). In particular, we used the three channels coded extracted by NMF as a set of more informative features able to improve the performance of simple thresholding segmentation techniques.

Fig. 6 sketches the phased of the NMF-based segmentation approach from the construction of the enlarged representation of an image I to the extraction of novel color space representation codified into the so called metacolors, and their adoption into a binary segmentation scheme.

4. Numerical results and discussions

The proposed NMF-based method for thresholding segmentation has been evaluated both qualitatively and quantitatively. The qualitative evaluation was conducted on four benchmark color images (represented in Fig. 7 by referring to an empirical goodness inspection simply based on human intuition [27]. The quantitative evaluation was conducted on a subset of 125 images randomly chosen from the PASCAL VOC2012 dataset (http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html). This dataset includes color images of objects (airplanes, birds, cars, etc.) depicted into realistic scenarios (Fig. 8 reports two images of this dataset with their corresponding ground truth binary segmentation).

The dataset includes for many images the corresponding ground truth (GT) binary segmentation, therefore a numerical evaluation has been performed using the quantitative metrics reported in Table 1. All the numerical experiments are conducted on a PC with



(a) Metacolor 1 (b) Metacolor 2 (c) Metacolor 3

(d) Distribution of Metacolor 1 (e) Distribution of Metacolor 2 (f) Distribution of Metacolor 3

Fig. 5. Novel color space representation obtained factorizing the enlarged representation of the bird example image (rank value $r = 3$) (a-c) and the corresponding distribution histograms.

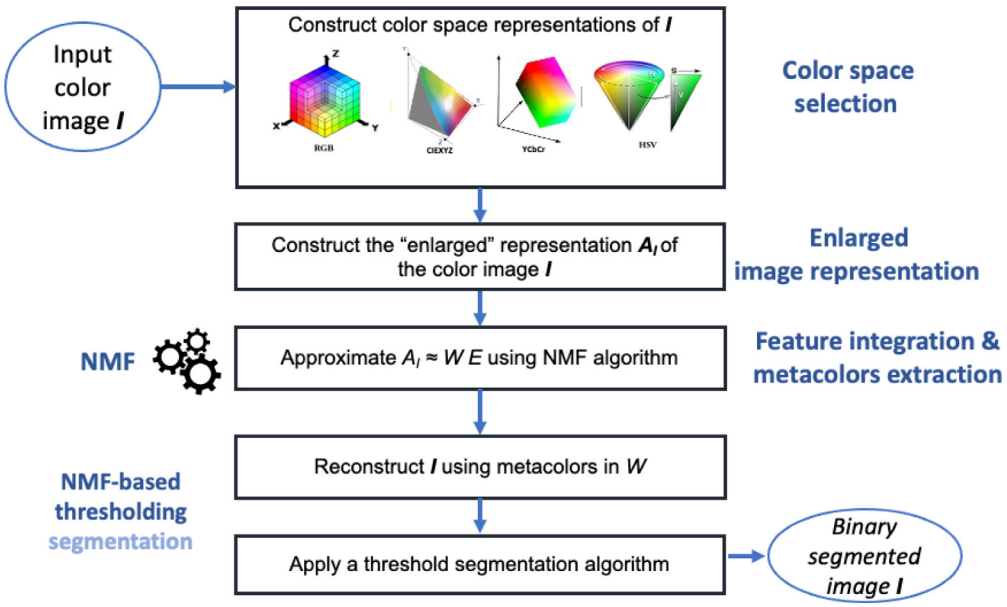


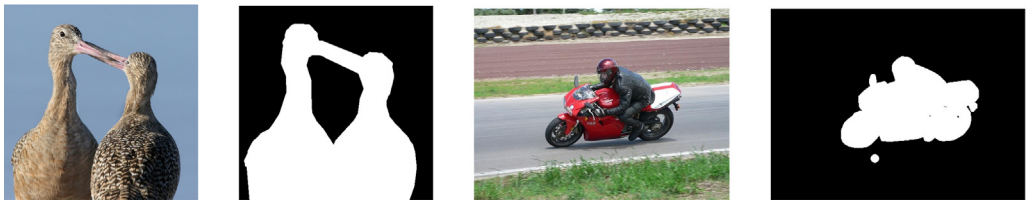
Fig. 6. The NMF-based segmentation algorithm flowchart. The first step builds an enlarged representation A_I of the image I given as input; then NMF performs the feature integration process on A_I extracting the meta-colors. The final step applies the binary threshold segmentation algorithm to the new image representation.

1.6 GHz Intel Core i5 dual-core, 16-GB memory, without external GPU; the segmentation algorithms and the NMF-based approach were run in MATLAB R2023b.

For the sake of comparison, in Fig. 9(b)-Fig. 12(a), we report the results of the segmentation obtained by applying the thresholding algorithms used as reference methods (i.e., Otsu, Kapur, Tsai, Ridler) to the benchmark images involved in the qualitative evaluation. The illustrated results relate to the segmentation obtained by referring to the gray-scale image representation.

(a) Martini Glasses (b) Starfish (c) Airplane (d) Table

Fig. 7. Images used for the qualitative evaluation of the segmentation results obtained.



(a) Goose (b) Ground truth (c) Motorbike (d) Ground truth

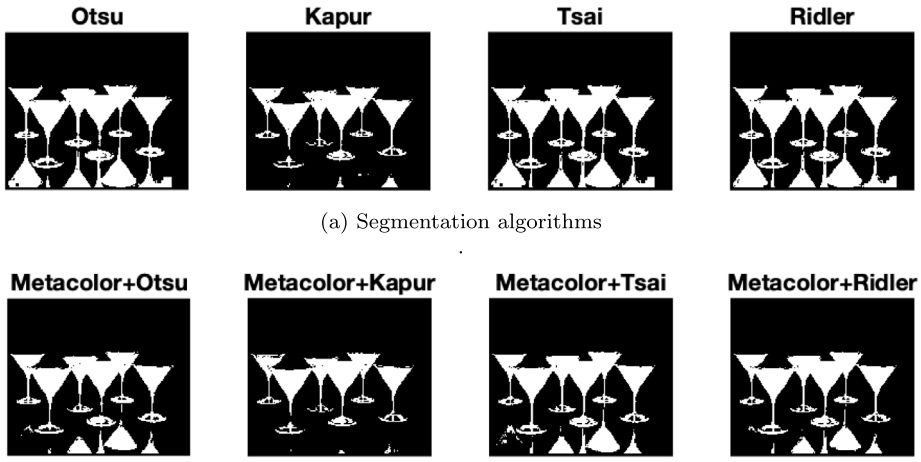
Fig. 8. Two example images included in the dataset used for the quantitative evaluation.

Table 1
List of the quantitative metrics used to compare segmented image with its ground truth (GT). As usual, the number of foreground pixels correctly classified as foreground are denoted by TP, while FP denotes the number of background pixels incorrectly classified as foreground. Analogously, the number of background pixels correctly classified as background are denoted by TN, while FN denotes the number of foreground pixels incorrectly classified as background.

Metric	Equation	Meaning
ACC	$\frac{TP+TN}{TP+TN+FP+FN}$	% pixels correctly classified in class c
$Sens$	$\frac{TP}{TP+FP}$	% pixels corresponding to boundary in GT
F	$\frac{2TP}{2TP+FP+FN}$	predictive performance
$Prec$	$\frac{TP}{TP+FP}$	segment performance
MCC	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	ineffectiveness in classifying FP
$Dice$	$\frac{2TP}{2TP+FP+FN}$	overlapping of predicted segmentation and GT
Jac	$\frac{Dice}{2-Dice}$	similarity of predicted segmentation and GT
$Spec$	$\frac{TN}{TN+FP}$	% pixels correctly identified in background

Analogously, in Fig. 9(b)-Fig. 12 (b), we report the segmentation results obtained by applying the same thresholding algorithms to the benchmark images represented in terms of the novel metacolor space defined through the proposed NMF-based method (as described in 6). As can be observed, the reference thresholding algorithms prove to be effective (producing just a few errors) in identifying the main objects which constitute the foreground of the images. Such inaccuracies are somewhat corrected by the application of the NMF-based segmentation. In fact, a better distinction between glasses and shadows can be observed for the Martini glasses image (Fig. 9(b)), and the foreground objects appear to be well-separated from the background. When we turn to consider the starfish image (Fig. 10) the improvement of the results is more appreciable in a better background identification. The Airplane object (Fig. 11) is better identified, providing a more clear appearance of the silhouette of the object in flight. In the table image (Fig. 12) the proposed method allows a better identification of the actual form of the object (especially when Otsu and Ridler algorithms are considered).

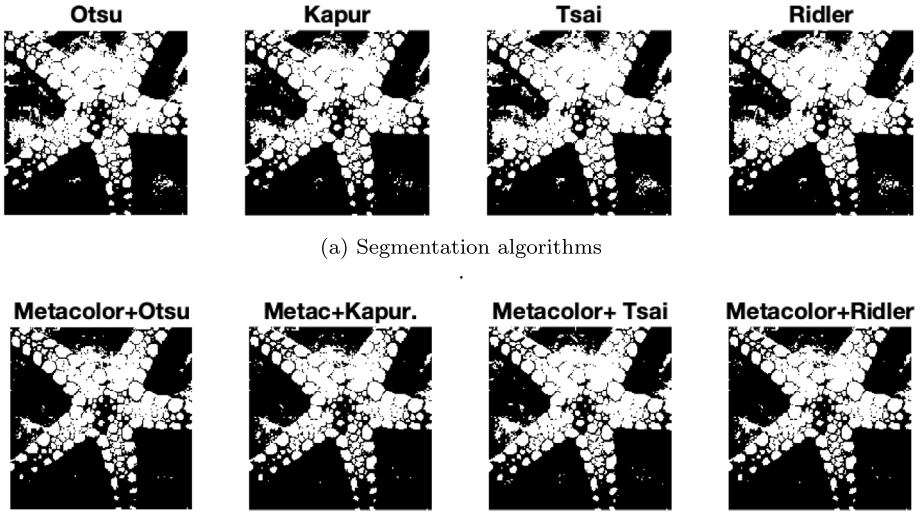
Concerning the quantitative analysis, Tables 2 and 3 report the average values of the previously introduced metrics resulting from the segmentation processes performed over the 125 images collected from the PASCAL VOC12 dataset. The metacolors extracted from the expanded representation of each color image are used as the basis vectors for representing the gray channel used as input to the considered thresholding algorithms. The quantitative results of the classical thresholding algorithms when the novel space representation is used, are generally improved with respect to the performance the same algorithms exhibited on a gray-scale image representation. This is specially true when Otsu and Ridler algorithms are considered: in this case all the metrics are improved when metacolors are used thus confirming an enhanced discriminating ability of the novel color representation extracted by NMF. However, the Kapur method does not receive any improvement in the segmentation quality deriving from the extraction and use



(a) Segmentation algorithms

(b) Proposed NMF-Based segmentation

Fig. 9. Binary segmentation obtained using: (a) the threshold algorithms on the original image and (b) the NMF-based segmentation approach on the Martini Glasses image.



(a) Segmentation algorithms

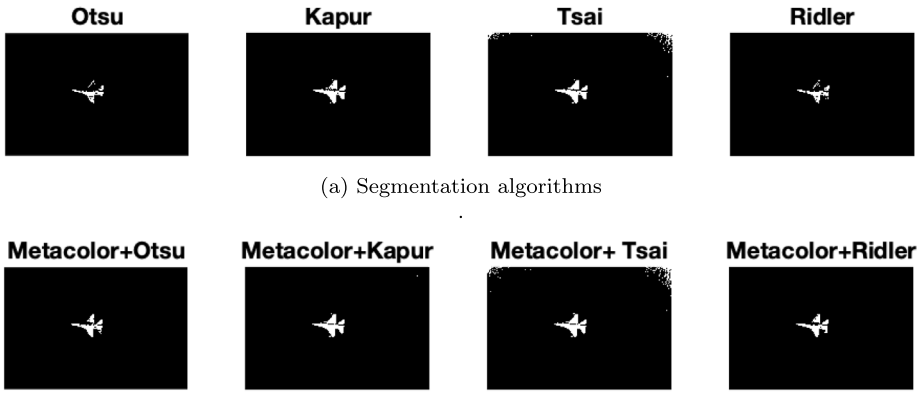
(b) Proposed NMF-Based segmentation

Fig. 10. Binary segmentation obtained using: (a) the threshold algorithms on the original image and (b) the NMF-based segmentation approach on the Starfish image.

of metacolors. Fig. 13 reports the binary segmentation results for an example image with ground truth. The segmentation obtained using the metacolors extracted by NMF helps to identify more pixels of the stone with the bird is backed as background pixels. However, the entropy-based thresholding algorithm produces poor binary segmentation results. A possible explanation could be based on the fact that the original color channels are not able to encode the information useful for discriminating pixels by a method based on entropy.

5. Conclusion

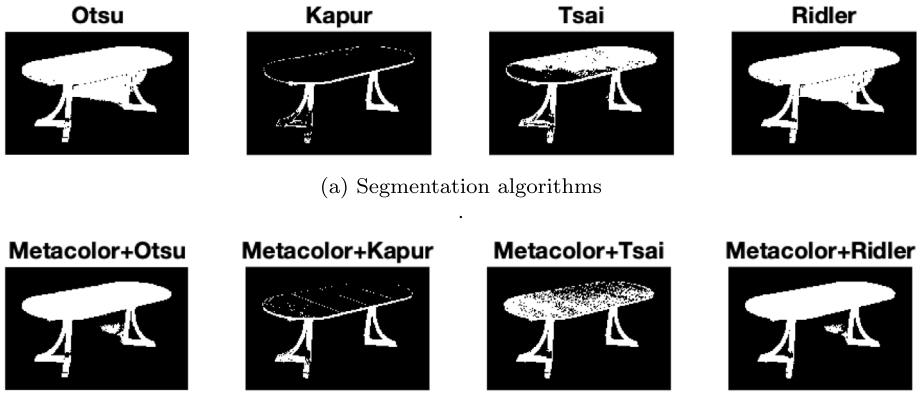
In this paper, information related to chosen color channels are integrated using an NMF algorithm, and a novel space representation is extracted some hidden knowledge of a color image. The extracted novel color representation can be been used for binary segmentation. The first step involves representing a color image I to be segmented using several color channels (four color spaces are employed, each with three color characteristics). Subsequently, each original color feature is retrieved and merged



(a) Segmentation algorithms

(b) Proposed NMF-Based segmentation

Fig. 11. Binary segmentation obtained using: (a) the thresholding algorithms on the original image and (b) the proposed NMF-based approach on the image of the aircraft.



(a) Segmentation algorithms

(b) Proposed NMF-Based segmentation

Fig. 12. Binary segmentation obtained using: (a) the threshold algorithms on the original image and (b) the NMF-based segmentation approach on the table image.

Table 2
Mean and standard deviation of the quantitative measures used to evaluate segmentation results with respect to the known ground-truth on 125 images for Otsu and Kapur thresholding algorithms (without and with the use of metacolors).

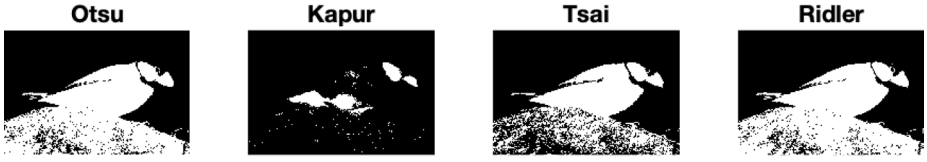
	Otsu	Metac+Otsu	Kapur	Metac+Kapur
ACC	0.76 ± 0.32	**0.77 ± 0.15**	**0.68 ± 0.28**	0.63 ± 0.3
$Sens$	0.75 ± 0.28	**0.79 ± 0.26**	**0.73 ± 0.29**	0.70 ± 0.31
F	0.69 ± 0.24	**0.71 ± 0.24**	**0.62 ± 0.32**	0.50 ± 0.36
$Prec$	0.70 ± 0.22	**0.71 ± 0.24**	**0.65 ± 0.33**	0.49 ± 0.38
MCC	0.38 ± 0.32	**0.42 ± 0.27**	**0.32 ± 0.36**	0.24 ± 0.35
$Dice$	0.69 ± 0.24	**0.71 ± 0.24**	**0.62 ± 0.32**	0.50 ± 0.36
Jac	0.58 ± 0.26	**0.60 ± 0.26**	**0.52 ± 0.32**	0.42 ± 0.34
$Spec$	0.61 ± 0.34	**0.65 ± 0.33**	**0.61 ± 0.33**	0.55 ± 0.33

into a new data matrix A_I encapsulating features so to incorporate several features of a color image in addition to the perceptual characteristics of RGB space. This expanded matrix representation of the image is mined using the Nonnegative Matrix Factorization algorithm to automatically extract metacolors, that constitute the basis vectors of a novel "color space" to be used for image analysis. These metacolors are used to divide the original image into binary segments by applying a thresholding technique.

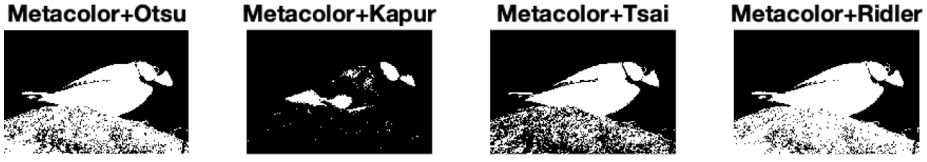
Table 3
Mean and standard deviation of the quantitative measures used to evaluate segmentation results with respect to the known ground-truth on 125 images for considered Tsai and Ridler thresholding algorithms (without and with the use of metacolors).

	Tsai	Metac+Tsai	Ridler	Metac+Ridler
ACC	0.73 ± 0.17	0.73 ± 0.17	0.71 ± 0.19	**0.75 ± 0.17**
$Sens$	0.74 ± 0.29	**0.77 ± 0.26**	0.75 ± 0.29	**0.76 ± 0.28**
F	0.69 ± 0.24	0.69 ± 0.24	0.69 ± 0.24	**0.71 ± 0.24**
$Prec$	0.70 ± 0.20	**0.74 ± 0.22**	0.70 ± 0.20	**0.75 ± 0.21**
MCC	0.36 ± 0.29	0.36 ± 0.26	0.36 ± 0.33	**0.41 ± 0.28**
$Dice$	0.69 ± 0.24	0.69 ± 0.23	0.69 ± 0.24	**0.72 ± 0.25**
Jac	0.58 ± 0.26	**0.59 ± 0.24**	0.57 ± 0.26	**0.61 ± 0.26**
$Spec$	0.59 ± 0.32	0.59 ± 0.32	0.58 ± 0.34	**0.65 ± 0.31**

(a) Original Image (b) Ground Truth



(c) Segmentation of the original image



(d) Segmentation using NMF-Based representation

Fig. 13. (a) Example of an image with (b) its ground truth. (c) Binary segmentation obtained using the threshold algorithms on the original image. (d) Binary segmentation obtained the NMF-based segmentation approach.

When compared with some standard thresholding algorithms that act directly on the gray-scale image representation derived from the RGB channels, the proposed NMF-based approach provides improved binary segmentation results.

As a plan for future work, we are going to consider the object identification inside an image, the interpretation of metacolors with reference to image attributes and the use of some more specific NMF algorithms.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are also affiliated to GNCS-INDAM (Gruppo Nazionale per il Calcolo Scientifico of Istituto Nazionale di Alta Matematica) Francesco Severi, P.le Aldo Moro, Roma, Italy.

Appendix

Otsu. This method aims at deriving an optimal threshold value that maximizes the separation between the two classes C_0, C_1. The basic idea is to minimize the intra-class variance defined as the weighted sum of the variances of the foreground and background regions. It is computed as

$$w_0\sigma_0^2 + w_1\sigma_1^2 \frac{(m_G P_1 - m(k))^2}{P_1(1-P_1)},$$

where σ_0 and σ_1 are variances of the two classes and the sum is weighted by the previously defined class probability distributions w_0 and w_1. m_G is the mean of the image intensity, $m(k)$ is the cumulative mean at a level k, and P_1 is the probability that a pixel is assigned to one class. The optimal threshold value is the one minimizing this intra-class variance. When 2 classes are involved minimizing the intra-class variance is equivalent to maximizing inter-class variance defined as

$$\sigma(k) = w_0(\mu_0 - \mu_k)^2 + w_1(\mu_1 + \mu_k)^2$$

being $\mu_k = \sum_{i=1}^{L} iP_i$ and $\mu_0 = \frac{\mu_k}{w_0}, \mu_1 = \frac{\mu_k}{w_1}$, then it can be computed as $\tilde{k} = argmax_{k=0,\dots,L-1}\sigma(k)$, being B the number of the distinct intensity levels appearing in the image. The main steps performed by the Otsu algorithm are:

1. Calculate the pixel frequency distribution histogram of I and the probability distribution P_i of each intensity level $i = 1, \dots, L$;
2. Normalize the histogram so that it sums to one, and calculate the w_0 and w_1;
3. Calculate the cumulative mean and cumulative variance of pixel values;
4. For each threshold value k calculate the inter-class variance;
5. Find $\tilde{k}$ that maximizes the inter-class variance.
6. Use $\tilde{k}$ to binarize I.

In the end, a final threshold $\tilde{k}$ is provided and it is used to produce a binary segmentation of the image I. **Kapur**. The Kapur, Sahoo, and Wong (KSW) thresholding algorithm (also known as the Kapur entropy-based thresholding algorithm) finds the optimal threshold that provides the greatest amount of information or reduces uncertainty when separating the image into foreground and background [19]. The method uses information theory to search an optimal threshold k that maximizes the sum of the entropy of the two classes C_0 and C_1 while computing the probability distribution of the image pixel intensities. The method considers the foreground and background of an image as two different signal sources and computes an optimal threshold maximizing the sum of the two class entropy. Let us assume that the image I is segmented by a threshold k, and let us consider the entropies H_0 and H_1 of the two classes C_0 and C_1:

$$H_0 = \sum_{i=1}^{k} \frac{P_i}{w_0} \ln \frac{P_i}{w_0} \quad \text{and} \quad H_1 = \sum_{i=k+1}^{L} \frac{P_i}{w_1} \ln \frac{P_i}{w_1}.$$

The main steps performed by Kapur algorithm are:

1. Calculate the pixel frequency distribution histogram of I and the probability distribution of each intensity level P_i, $i = 1, \dots, L$;
2. Normalize the histogram so that it sums to one, and calculate the cumulative sum w_0 and w_1;
3. Calculate the overall entropy $H(k)$ of the two classes, that is $H(k) = H_0 + H_1$;
4. Find $\tilde{k}$ that maximizes $H(k)$.
5. Use $\tilde{k}$ to binarize I.

In the end, a final threshold $\tilde{k}$ is provided and it is used to produce a binary segmentation of the image I.

Tsai. Tsai's moment-preserving method [20] is based on the idea to consider an image as the blurred vision of an ideal binary image. In this method, the best threshold is selected in such a way that the first three moments of the image are preserved in the resultant binary image. Defined the kth moment m_k as

$$m_k = \frac{1}{N} \sum_{i=1}^{L} i^k n_i,$$

then the optimal threshold $\tilde{k}$ is chosen as

$$\tilde{k} = \frac{\frac{1}{2}(c_1 + \sqrt{\Delta}) - m_1}{\sqrt{\Delta}}$$

being $\Delta = c_1^2 - 4c_0$, $c_0 = \frac{m_1 m_3 - m_2^2}{m_2 - m_1^2}$, and $c_1 = \frac{m_1 m_2 - m_3^2}{m_2 - m_1^2}$.

Ridler and Calvard. This method consists in an iterative algorithm that starts with an initial guess threshold value k_0 that is used to initially split the image histogram into two portions. Generally, k_0 is set to the average color level of the overall image or the average color level of a subset of the pixels of the image (e.g., the four corners) that is most likely to contain only background pixels [21,22]. Starting from k_0, the algorithm proceeds at each iteration as follows:

1. Compute the set V_{Below} of pixels in I with intensity level below the threshold k_0;
2. Compute the mean intensity value μ_B of the pixels included into V_{Below};
3. Compute the set V_{Above} of pixels in I with intensity level above the threshold k_0;
4. Compute the mean intensity value μ_A of the pixels included into V_{Above};
5. Evaluate the new threshold value as $k_i = \frac{\mu_A + \mu_B}{2}$;

6. Iterate steps 1–5 while $k_{i+1} - k_i$ is greater than a fixed tolerance value.

In the end, a final threshold $\tilde{k}$ is provided and it is used to produce a binary segmentation of the image I.

References

[1] Shapiro (Autore) LG, Stockman GC. Computer vision. Pearson College Div; 2001.
[2] Sankur B, Sezgin M. Image thresholding techniques: A survey over categories. Pattern Recognit 2001;34(2):1573–607.
[3] Yan F, Zhang H, Kube CR. A multistage adaptive thresholding method. Pattern Recognit Lett 2005;26(8):1183–91.
[4] Gonzalez R, Woods R. Digital image processing. Prentice Prentice Hall; 2002.
[5] Cheng HD, Jiang XH, Sun Y, Wang JL. Color image segmentation: Advances and prospects. Pattern Recognit. 2001;34:2259–81.
[6] Garcia-Lamont F, Cervantes J, López A, Rodriguez L. Segmentation of images by color features: A survey. Neurocomputing 2018;292:1–27. http://dx.doi.org/10.1016/j.neucom.2018.01.091, URL https://www.sciencedirect.com/science/article/pii/S0925231218302364.
[7] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401(6755):788–91.
[8] Lee DD, Seung HS. Algorithms for non-negative matrix factorization. In: Proc adv neur inf proc sys, vol. 13, MIT Press; 2000, p. 556–62.
[9] Gillis N. Nonnegative matrix factorization. SIAM; 2020.
[10] Yuan Z, Oja E. Projective nonnegative matrix factorization for image compression and feature extraction. In: Kalviainen H, Parkkinen J, Kaarna A, editors. Image analysis. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005, p. 333–42.
[11] Guillamet D, Schiele B, Vitria J. Analyzing non-negative matrix factorization for image classification. In: 2002 international conference on pattern recognition, vol. 2. 2002, p. 116–9 vol.2.
[12] Guillamet D, Vitria J, Schiele B. Introducing a weighted non-negative matrix factorization for image classification. Pattern Recognit Lett 2003;24(14):2447–54. http://dx.doi.org/10.1016/S0167-8655(03)00089-8, URL https://www.sciencedirect.com/science/article/pii/S0167865503000898.
[13] Rajapakse M, Tan J, Rajapakse J. Color channel encoding with NMF for face recognition. In: 2004 international conference on image processing, 2004, vol. 3. 2004, p. 2007–10.
[14] Luong TX, Kim B-K, Lee S-Y. Color image processing based on nonnegative matrix factorization with convolutional neural network. In: 2014 international joint conference on neural networks. 2014, p. 2130–5. http://dx.doi.org/10.1109/IJCNN.2014.6889948.
[15] Castiello C, Del Buono N, Esposito F. Improving color image binary segmentation using nonnegative matrix factorization. In: Gervasi O, Murgante B, Rocha AMAC, Garau C, Scorza F, Karaca Y, et al., editors. Computational science and its applications – ICCSA 2023 workshops. Cham: Springer Nature Switzerland; 2023, p. 623–40.
[16] Kahu SY, Raut RB, Bhurchandi KM. Review and evaluation of color spaces for image/video compression. Color Res Appl 2018. http://dx.doi.org/10.1002/col.22291, URL https://onlinelibrary.wiley.com/doi/10.1002/col.22291.
[17] Distante A, Distante C. Color. In: Handbook of image processing and computer vision: volume 1: from energy to image. Cham: Springer International Publishing; 2020, p. 79–176.
[18] Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 1979;9(1):62–6. http://dx.doi.org/10.1109/TSMC.1979.4310076.
[19] Kapur J, Sahoo P, Wong A. A new method for gray-level picture thresholding using the entropy of the histogram. Comput Vis Graph Image Process 1985;29(3):273–85. http://dx.doi.org/10.1016/0734-189X(85)90125-2, URL https://www.sciencedirect.com/science/article/pii/0734189X85901252.
[20] Tsai W-H. Moment-preserving thresolding: A new approach. Comput Vis Graph Image Process 1985;29(3):377–93. http://dx.doi.org/10.1016/0734-189X(85)90133-1, URL https://www.sciencedirect.com/science/article/pii/0734189X85901331.
[21] Ridler TW, Calvard S. Picture thresholding using an iterative selection method. IEEE Trans Syst Man Cybern 1978;8(8):630–2. http://dx.doi.org/10.1109/TSMC.1978.4310039.
[22] Xue J-H, Zhang Y-J. Ridler and Calvard's, Kittler and Illingworth's and Otsu's methods for image thresholding. Pattern Recognit Lett 2012;33(6):793–7. http://dx.doi.org/10.1016/j.patrec.2012.01.002.
[23] Esposito F. A review on initialization methods for nonnegative matrix factorization: Towards omics data experiments. Mathematics 2021;9:1006.
[24] Berry M, Browne M, Langville A, Pauca P, Plemmons R. Algorithms and applications for approximate nonnegative matrix factorization. Comput Statist Data Anal 2007;52(1):155–73.
[25] Kim J, He Y, Park H. Lgorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. J Global Optim 2014;58:285–319. http://dx.doi.org/10.1007/s10898-013-0035-4.
[26] Kim H, Park H. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. SIAM J Matrix Anal Appl 2008;30(2):713–30. http://dx.doi.org/10.1137/07069239X.
[27] Zhang Y. A survey on evaluation methods for image segmentation. Pattern Recognit. 1996;29(8):1335–46. http://dx.doi.org/10.1016/0031-3203(95)00169-7.