RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2026, **16**, 12511

Rediscovering lost colors: film color restoration by hyperspectral imaging and cluster-based spectral correction algorithm (CBSCA)

Emilio Catelli,[a] Lingxi Liu,[b] Jacopo Fadanni,[c] Jošt Stergar,[de] Matija Milanič,[de] Giorgia Sciutto,[a] Francesco Zerbetto*[f] and Silvia Prati *[a]

Received 5th November 2025
Accepted 25th February 2026

DOI: 10.1039/d5ra08504g

rsc.li/rsc-advances

The present study proposes a new method for the digital restoration of color films affected by irregular color degradation. Color fading is a major problem affecting analog color films and while most existing digital methods address homogeneous dye fading, the correction of irregular color degradation remains a challenging and time-consuming process. The main objective of this work is to develop a scalable and minimally user-guided strategy capable of restoring such complex degradation patterns. The proposed approach combines visible-range hyperspectral imaging of degraded film frames to obtain detailed spectral information with a novel computational algorithm referred as the Cluster-Based Spectral Correction Algorithm (CBSCA). This algorithm is specifically designed to handle the large volume of data recorded by hyperspectral acquisition, enabling robust color restoration while ensuring reduced computational demand and limited subjective intervention. In comparison with conventional RGB scanners and color restoration software, this approach offers a methodology to perform accurate color acquisition and correction while effectively managing large hyperspectral datasets. The ultimate goal is to recover the film's original visual content, ensuring its accessibility to the public.

Introduction

At the time of their creation, movies were considered as a form of entertainment.[1] Today, after more than a century of movie history, they are legitimately recognized as a part of our cultural heritage[2] and considered as both a form of art and a valuable historical resource.

Before the digital age, movies were shot on a thin plastic strip coated with a light-sensitive emulsion, hence the name film.[3] The repetitive use of the hard copy and inappropriate storage in humid/high-temperature conditions have severely compromised film's appearance and legibility. Particularly, either the plastic support or the silver salts/colorants in the emulsion turn out to be physico-chemically unstable, producing degradation effects.[4] Film degradation is progressive and manifests though phenomena including delamination, embrittlement, distortion, scratches, dye and silver fading, biological attack and many others.[5] Conservation efforts aim to slow deterioration, while restoration is the ensemble of technical procedures finalized to return artifacts and especially their image content to their initial quality.[3,6] Both conservation and restoration are part of the broad preservation actions, which are defined as "all the practices and procedures necessary to ensure permanent accessibility (with minimum loss of quality) to the visual or sonic content of the materials."[3,7]

In the last years, digital restoration has emerged as a key step in the preservation practice as it is able to recover the film's original content, ensuring its accessibility to the public.

A digital restoration process begins with tracing and reconvening all the possible (multiple) versions of a film, individuating the best-preserved scenes.[3]

Physical restoration is initially used to repair the analog copies of the movie in order to provide the film with the properties required for scanning. In particular, this step consists of operations aimed at fixing the support, cleaning the surface of the film, stabilizing, and correctly mounting the scene.[3,8] Then, the film is digitally scanned at high resolution. The digital scanner acquires the RGB scene one by one at 2 K (2048 × 1080 pixels) or 4 K (4096 × 2160 pixels) resolutions, which can be visualized by computers.[3]

Color vision is governed by three retinal cone cells in eyes, sensitive to red, green and blue colors.[9] To reproduce human color perception, conventional scanners capture the light

[a]*Department of Chemistry "G. Ciamician", University of Bologna, Ravenna Campus, Via G. Guaccimanni 42, Ravenna, 48121, Italy. E-mail: s.prati@unibo.it*
[b]*Department of Art History, Musicology and Theatre Studies, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium*
[c]*Department of Physics and Astronomy, University of Padova, Via F. Marzolo 8, Padova, 35131, Italy*
[d]*Jozef Stefan Institute, Jamova cesta 39, Ljubljana, SI-1000, Slovenia*
[e]*Faculty of Mathematics and Physics, University of Ljubliana, Jadranska ulica 19, Ljubljana, SI-1000, Slovenia*
[f]*Department of Chemistry "G. Ciamician", University of Bologna, Via P. Gobetti 85, Bologna, 40129, Italy. E-mail: francesco.zerbetto@unibo.it*

reflected from or transmitted through a material in these three primary colors, and combine them to obtain all the possible hues.[10] Film scanners are sensitive to the dyes used in the film industry, therefore are settled to acquire images at the characteristic wavelengths of the dye maximum absorption around 450 nm for the yellow dye, 550 nm for the magenta, and 650 nm for the cyan one.[10]

With the help of software, the restorer adjusts the scratches, the cuts, and reconstructs the missing parts.[8] More deeply, the restorer undertakes meticulous work, each frame individually, on the colors and brightness, reconstructing the presumed original colors, by keeping as a reference the best-preserved frame of each scene when available. At the end, the audio track is corrected, and a digital and analog copy of the restored film is created for archive.[4,11]

Among the key steps for digital restoration, color restoration is the most critical one. First, it necessitates the best acquisition of the colors, which sometimes is not fully achieved with conventional scanners due to the uncorrected illumination and/or the inappropriate spectral sensitivity of the instrument (*i.e.* a RGB scanner equipped with blue, green and red LEDs with a relatively small bandwidth might not fully acquire the absorption peaks of the dyes in the specimen).[12] Second, it faces the difficulty of tracing and recognizing the original colors, especially when the fading effect has affected the full image. This process becomes more complex when there are multiple types of fading (inhomogeneous fading) on the same image. Third, the color reconstruction often relies on the personal experience and subjective taste of the restorer, especially when reference frames carrying the original (or best-preserved) colors are missing. Even when well-preserved reference frames are available, the restoration is not free from subjective influences.

In the past decades, such problems have raised much attention among the experts working in Cinematheques or Universities. Focused mainly on diminishing the subjective intervention of the restorers, the current leading computational algorithms are based on the idea of restoring faded films in a way that looks natural to the human eye. In presence of homogeneous color cast, these algorithms mainly adjust the contrast and color within predefined ranges[13–16] without the need for an original color reference. Other methods combine the use of new scanners for color acquisition, with algorithms able to reconstruct the original color appearance, leveraging spectroscopic knowledge of the film's dye characteristics.[17–20]

Moving beyond the conventional three-band RGB scanners, Trumphy *et al.* proposed a new type of scanner based on multiband (multispectral) acquisitions in transmission mode, versatile and accurate in capturing the multitude of colors of film stocks.[12] Further, coupling the three RGB bands of the multispectral scanner with a spectral-based computational method called "dye purification", Trumphy *et al.* were able to reconstruct the concentration of the pure original dyes in the emulsion, an information that was successively employed to operate a coherent and objective digital color restoration of the images and its faded parts.[17–19]

An advanced imaging scanner combined with a clustering-based "vector quantization" algorithm was used by Liu *et al.* to restore complex faded frames with inhomogeneous discoloration.[20] In such research, the frames were scanned by a custom-made visible hyperspectral imaging camera, with spectral resolution of 2.5 nm and spatial resolution about 100 μm pixel at 30 cm distance.[21] The method first creates a library of faded spectra and associates each faded spectrum to a well-preserved one by strictly registering the reference damaged image and the best preserved one. Then, the digital restoration is achieved by searching inside the library, pixel by pixel, for a reference spectrum that contains the most similar degradation characteristics and substituting the damaged image with the corresponding well-preserved ones. The library was established by overlapping the references, which leads to significant redundancy, and the spectrum-wise substitution introduces significant noise in the reconstructed images. Moreover, the pixel-wise comparison between the restoration target and the library imposes a substantial computational load and thus limits the applicability to large images.[20] To address this limitation, Liu *et al.* further investigated clustering algorithms on degraded color films to efficiently extract distinct degradation patterns. They initially applied a grid-based simplification approach[22] and later adopted superpixel segmentation as a data reduction strategy.[23] This method was combined with a soft clustering algorithm that enabled accurate segmentation while generating a probability matrix, reflecting the likelihood of each pixel belonging to multiple clusters, thereby reducing the risk of losing small structures by forcing pixels to belong solely to a single cluster.[23] In particular, different methods were employed, and the most promising results were obtained by combining Simple Linear Iterative Clustering (SLIC) superpixel algorithm[24] with the soft clustering algorithm, the Gaussian Mixture Model (GMM).[25] The developed approach was applied to single frames separately.

Given the ability of SLIC superpixel and Gaussian mixture model (GMM) to identify color degradation patterns in high-dimensional spectroscopic data,[23] the present work advances digital film restoration by proposing a computational algorithm that integrates these methods with a spectroscopy-based color correction technique to address inhomogeneous color fading. SLIC and GMM provide a solid foundation for this task, as they effectively capture fine-scale degradation details while maintaining a low computational load and preserving the spatial resolution of the restored images. The proposed computational algorithm-referred to as the Cluster-Based Spectral Correction Algorithm (CBSCA)- is specifically designed for hyperspectral imaging data, enabling the full exploitation of the rich information encoded in spectral profiles for restoration.

Inhomogeneous(irregular) color fading poses a significant challenge in digital film restoration due to the non-uniform spatial distribution of degradation and the dependence of many existing methods on manual, region-wise correction. The proposed color correction approach robustly identifies degraded regions by operating in the spectral domain, relying on spectral shape rather than RGB intensities. Furthermore, CBSCA supports the simultaneous processing of multiple frames, enabling consistent correction of recurring degradation patterns across multiple frames within the same scene, thereby

improving chromatic coherence while reducing manual intervention and computational overhead. As a first proof of concept, the method has been developed employing four frames, three degraded and one identified as the best preserved, previously studied by Liu *et al.*[20,22,23] The approach depends on this best-preserved reference frame, which provides the source of undegraded spectra used for restoring the degraded regions.

Experimental methods

Samples

The samples used in the study were provided by "L'immagine Ritrovata Film Restoration Laboratory" of Bologna (Cinematheque of Bologna) and are a series of four color frames from a 35 mm Fuji N4 1986 film. The support is cellulose acetate while the emulsion has a trichromatic structure with yellow, cyan and magenta dyes allocated in each pertinent layer.[20,22,23] The frame considered as the best-preserved (S1) presents a relatively uniform blue coloration, with magenta details. In contrast, the three faded frames (S2–S4) show inhomogeneous color fading ranging from magenta to yellow.

Hyperspectral imaging

The instrumentation used in this study is a custom-made VNIR push-broom hyperspectral camera, developed at the University of Ljubljana, Faculty of Mathematics and Physics.

The camera combines an imaging spectrograph V10E (Specim, Finland), with a 5 MP monochrome CMOS camera (Flir System USA) and a 50 mm Xenoplan lens (Schneider-Kreutznach,Bad Kreuznach, Germany).[21] The number of across track pixels is 2448 (y-direction). The spectral range spans from 338 to 1025 nm, comprising 2048 wavelengths acquired at 0.3 nm interval. The spectral resolution is 2.5 nm and spatial resolution about 100 μm per pixel at 30 cm distance. Illumination is provided by two blocks of water-cooled broadband LED light sources with integrated polarizer-diffuser (Bolder Vision Optik, Boulder, CO, USA), positioned on either side of the camera. Each LED block mounts ten white LEDs (LCW H9GP), ten 780 nm LEDs (SMB1N780D), ten 850 nm LEDs (SFH 4715S) and ten 940 nm LEDs (SFH 4725S) LEDS. The LEDs are interlaced, resulting in a total of 40 LEDS per each light source. Detailed information on the illumination system and its spectral distribution is provided in ref. 21. Custom-made software developed in the Matlab enviornment (The MatWorks, USA) controls image acquisition and allows definition of scanning parameters such as speed, integration time and number of pixels along the x-direction.

The four film samples were scanned sequentially in reflection geometry, with the dyed layers oriented toward the camera. The samples were illuminated using the above-mentioned LED-light sources, and a wire-grid polarizer (Bolder Vision Optik, Boulder, CO, USA) was employed between the camera and the sample to mitigate the specular reflection contribution. The distance between the camera and the samples was set to 30 cm, with an integration time of 300 ms and speed of 1000 rpm. Commercial white Spectralon standard (Labsphere, USA) with reflectance near 100% was used both to convert the images from digital number to absolute reflectance values and as a support to hold the fames. The frame samples were kept flat against the white support using a custom-made frame.

The film frames were fully scanned including the lateral perforations and film margins, resulting in hyperspectral datasets of approximately 1600 × 2448 pixels and 2048 wavelengths. The acquisition time per each frame required few minutes. The images were first normalized using the white reference and subsequently binned 5× along the spectral dimension to reduced noise. Additional corrections were applied in both spatial and spectral dimensions. Spatially, the hyperspectral cube was cropped to retain only the image content. Spectrally, the wavelength range was restricted to 380–780 nm. The resulting hyperspectral cube consists of 1000 × 1200 pixels and 240 wavelengths.

RGB images were generated from hyperspectral data using the CIE XYZ color matching functions and the spectrum of a D65 illuminant, following the method proposed by Magnusson *et al.*[26] The procedure here adopted consists of two steps: (1) conversion of the spectra to the XYZ colorspace, implemented using an in-house MATLAB algorithm and (2) conversion from the XYZ to the RGB colorspace using a in-built MATLAB function. In first step, after defining the spectral boundaries of both the XYZ color match functions and the hyperspectral image, the reference XYZ color vectors were interpolated to match the spectral dimension of the hyperspectral data The resulting curves were used as scaling factors to compute the X, Y, and Z image channels separately. These three images were then combined to form the CIE XYZ color space image. In the second step, the MATALB function *xyz2rgb* was used to convert XYZ images to RGB images.

Segmentation

Image segmentation was performed in two steps: (1) simple linear iterative clustering (SLIC) superpixel[24] and (2) Gaussian Mixture Model (GMM).[25]

The procedure was implemented in Matlab environment (The MathWorks, Inc., version 2023b).

To reduce computational complexity while preserving spectral details, we first applied the SLIC superpixel algorithm to segment the images into macro areas, where each area presents relatively homogeneous characteristics. For performing this step, the color images of the frames are in the RGB color space, obtained from the hyperspectral data.

To evaluate whether a three-channel representation is sufficient, we performed Principal Component Analysis (PCA) on the hyperspectral datacubes. PCA orders the principal components by decreasing variance, and cumulative explained variance quantifies the intrinsic dimensionality of the data. We found that the first three principal components account for more than 88% of the total variance, indicating that the dominant variability of the hyperspectral data lies in a three-dimensional subspace. This justifies the use of RGB representation for subsequent processing (SI Fig. SM1). Calculation of the PCA was obtained using hypertools[27] in MATLAB environment.

Giving a number of superpixels K of an image with N total pixels, the algorithm begins by initializing K cluster centers $C_k = [R_k\, G_k\, B_k\, x_k\, y_k]$ evenly across the image, where $k = [1, K]$. R_k, G_k, B_k are the color values per each cluster k, while x_k and y_k are the spatial coordinates. For each center, the algorithm searches within a local grid interval S, defined as (eqn (1)):

$$S = \sqrt{(N/k)} \tag{1}$$

to assign nearby pixels to the center using a distance metric D. The metric D express how well a pixel matches a cluster center simultaneously considering how similar their colors are and how close they are in spatial location within the image (eqn (2)):

$$D = \sqrt{{d_c}^2 + \left(\frac{d_s}{S}\right)^2 m^2} \tag{2}$$

where d_c is the measured color distance, d_s is the spatial distance to the center, and m is compactness parameter that controls the regularity of area shapes, *i.e.*, the lower the compactness, the better adherence to object boundaries, and the more irregularly shaped. According to eqn (2), the superpixels are calculated both from spectral (projected in the RGB space) and spatial similarity. The update step recomputes cluster centers as the mean of the feature vectors of pixels assigned to that cluster (eqn (3))

$$\boldsymbol{C}_k^{\text{new}} = \frac{1}{M_k} \sum_{i=1}^{M} \boldsymbol{f}_i \tag{3}$$

where $\boldsymbol{C}_k^{\text{new}}$ is the new cluster center, M_k is the number of pixels assigned to cluster k and $\boldsymbol{f}_i$ is the feature vector of pixel i.

These assignment and update steps are iteratively repeated until the change in position of the cluster centers is negligible (convergence). A label matrix L_{SLIC} is at the end generated that contains the membership of each pixel to the superpixels, with the final number of superpixels less than the input K.

Aligning with the best results reported in ref. 23, the optimal values for the images were obtained using a compactness value $m = 10$ and $K = 2000$. The number of superpixels K was optimized considering the total sum of the internal cluster variance as a function of the number of clusters. The resulting analysis, presented in the SI (Fig. SM2), shows that the internal cluster variance stabilizes at approximately 2000 superpixels, indicating that the clusters become nearly homogeneous in their composition. This quantitative assessment was further complemented by visual inspection of the resulting image of the superpixel grids, which verifies that the clustering was able to capture all independent structural features associated with image degradation (*e.g.*, human contours and thin lamp supports).

The four film frames were segmented into 1822, 1864, 1898, and 1875 superpixels, reducing the overall data size from 4.8 million spectra to only 7.5k spectra. The centroids of each superpixel were extracted and stored in a bidimensional matrix (cluster centers x spectral wavelengths), to form the input matrix for the following spectral clustering.

Then, to further group the macro areas by their spectral signatures, the Gaussian mixture model (GMM) is applied to the centroid matrix obtained from the previous step. GMM models the distribution of each spectrum x_i as a weighted sum of k Gaussian components, defined by the cluster center μ, covariance matrix Σ, and the weight π (eqn (4)):

$$p(x_i) = \sum_{j=1}^{k} \pi_j \mathcal{N}(x_i \mu_j, \Sigma_j) \tag{4}$$

For a specified number of components k, the model is initialized using k-means algorithm, then iteratively refined by the Expectation–Maximization (EM) algorithm.[28] The model obtained, with well-defined distribution approximation of cluster centers and covariance matrices, can be applied to classify unseen data, which in our case is the full resolution original data without superpixel reduction. Finally, each original spectrum is assigned to the closest cluster for visualizing the segmentation, with the probability matrix of it belonging simultaneously to other clusters saved for further color correction step. The training is performed in Matlab (The MathWorks, Inc., version 2023b) using the *fitgmdist* function, with a regularization value set to 0.01 and other algorithmic parameters as default. The optimal number of clusters was initially evaluated using the Bayesian Information Criterion (BIC)[29] and the Akaike Information Criterion (AIC)[30] across varying numbers of clusters. However, both criteria suggested a number of clusters that was too low to capture the complexity of the degradation patterns adequately. Previous studies have shown that, in high-dimensional data, AIC and BIC can become unreliable, as their estimates may be adversely affected by the high dimensionality.[31,32] This limitation is particularly relevant in the present case, where the superpixels' spectral matrix comprises a large number of spectral variables.

As a result, an alternative model section strategy based on the Within-Cluster Sum of Squares (WCSS) as a function of the number of clusters was considered, using the "elbow" method. Inspection of the corresponding plot reveals a flattening in the WCSS curve between 10 and 13 clusters (Fig. SM3, SI). Although the "elbow" method generally favors smaller numbers of clusters, 13 clusters were selected here to provide a more faithful representation of the color variations. This choice is further supported by visual inspection of the resulting cluster maps, which indicates that the selected number of clusters meaningfully describes the principal color variation in the faded frames.

Spectra–spatial curve correction

The color correction is achieved by reducing the color distance between the damaged frame and the reference one.

Two frames differ in color not only for the different degree of damage but also for the differences in the scene itself. To avoid the unwanted effect of correction, good clustering is crucial. In that step, similar regions are grouped with a double positive effect: on one side, it allows a proper correction of the different damages, on the other, it assures that the fine details that define a scene are preserved.

The spectra correction proceeds through the following steps:

Given a damaged frame f, for each cluster i, a representative spectrum S_i^{f} is computed as the average of the pixels spectra of that cluster (eqn (5)).

$$S_i^{\mathrm{f}} = \frac{1}{N_i}\sum_{j}^{N_i} v_j^{\mathrm{f}} \tag{5}$$

N_i is the number of pixels in the cluster i, v_j^{f} is the spectrum of pixel j.

On the reference frame, a representative spectrum is computed as the average of the pixels in the same spatial region that correspond to a cluster in the damaged frame (eqn (6))

$$S_i^{\mathrm{r}} = \frac{1}{N_i}\sum_{j}^{N_i} v_j^{\mathrm{r}} \tag{6}$$

With S_i^{r} being the representative spectrum of the reference frame r, N_i is the number of pixels in the cluster i, v_j^{r} is the spectrum of pixel j.

While working with soft clusters, in the computation of the representative spectra, a pixel is associated with the cluster with the highest probability.

For each cluster and frame, a correction factor C_i^{f} is then computed (eqn (7)):

$$C_i^{\mathrm{f}} = S_i^{\mathrm{r}} - S_i^{\mathrm{f}} \tag{7}$$

Given the correction factors for all the clusters in a frame, a properly weighted correction is applied to each individual pixel producing a (partially) corrected color $\tilde{v}_{xy}^{\mathrm{f}}$ (eqn (8))

$$v_{xy}^{\prime\mathrm{f}} = \sum_{i}^{N_c} C_i^{\mathrm{f}} \times p_{xy}^{i} \tag{8}$$

With N_c the number of clusters, p_{xy}^{i} the probability of a pixel of belonging to cluster i.

This procedure produces a frame with strong artifacts due to the border effects of the clusters. To remove the artifacts, a second correction is applied.

This second correction starts by considering the partially corrected spectra to compute the correction factor (eqn (9)).

$$C_i^{\prime\mathrm{f}} = S_i^{\mathrm{r}} - \frac{1}{N_i}\sum_{j}^{N_i} v_j^{\prime\mathrm{f}} \tag{9}$$

The new correction factor is applied considering as weights the fraction of pixels in the neighborhood that belong to the given cluster, where $C_i^{\prime\mathrm{f}}$ is the new correction factor and $v_j^{\prime\mathrm{f}}$ is the updated spectrum of the damaged frame. Consider x_j, y_j the coordinates in pixels of the j-th pixel, then the weights are computed as (eqn (10)):

$$w_i(j) = \frac{\mathrm{count}i\in\left[x_j - n : x_j + n; y_j - n : y_j + n\right]}{(2n)^2} \tag{10}$$

with n half the size of the neighborhood. A detailed representation of the calculation of the weights is reported in Fig. SM4 in SI.

With the computed weights, the corrected spectrum $\tilde{v}^{\mathrm{f}}(j)$ becomes (eqn (11)):

$$\tilde{v}^{\mathrm{f}}(j) = \sum_{i}^{N_c} C_i^{\prime\mathrm{f}} \times w_i(j) \tag{11}$$

The evaluation of the optimal number of neighborhood pixels was determined by computing the color distance between pixel-to-pixel aligned corrected frame and the reference one for different neighborhood sizes. The optimal neighborhood size corresponds to the minimum color distance and was found to be 35 for frame S2, 20 for frame S3 and 10 for frame S4. The results of this evaluation are presented in SI, Fig. SM5.

Visualization and evaluation

The visualization of the films in the RGB domain was obtained by transformation of the hyperspectral data cubes through the CIE XYZ color space.[26] The evaluation of the quality of restoration was perform by calculating, in turn, the color distance ΔE between the reference best-preserved and the restored images. The procedure required first a thorough alignment of the images in Matlab (The MathWorks, Inc., version 2023b) using the *imregister* function and secondly the calculation of the ΔE pixel by pixel through Euclidean distance (eqn (12)):

$$\Delta E = \sqrt{\left(R_{\mathrm{ref}} - R_{\mathrm{corr}}\right)^2 + \left(G_{\mathrm{ref}} - G_{\mathrm{corr}}\right)^2 + \left(B_{\mathrm{ref}} - B_{\mathrm{corr}}\right)^2} \tag{12}$$

where, R_{ref}, G_{ref} and B_{ref} represent the three color channels of the reference image, while R_{corr}, G_{corr} and B_{corr} are those of the corrected ones. The color difference is displayed as a heatmap where the highest ΔE values are in red and the lowest in blue. For providing a quantitative estimation of the goodness of the restoration, the dE values obtained from each pair of images were averaged across all the pixels (eqn (13)):

$$\mathrm{d}E = \frac{1}{M}\left(\sum_{i=1}^{M}\Delta E_{\ i}\right) \tag{13}$$

where M is the total number of pixels in the images.

The alternative restoration approach was conducted on faded RGB images using the DaVinci Resolve software, version 18.6 (Blackmagic Design, Australia). The images are imported into the software and processed with the same 1000 × 1200 pixels resolution as the hyperspectral images. The restoration was performed in ten to thirteen steps (nodes). Two to three nodes were employed to perform initial color correction of hue, tint, temperature, contrast, and saturation. The successive nodes were used to correct selectively specific areas affected by severe fading by using masks to select the pertinent areas. The last node was used to homogenize the overall color by acting again on hue and saturation. The images were saved in the RGB color space.

Results and discussion

The different restoration steps are summarized in Fig. 1: (1) hyperspectral imaging scanning, (2) image segmentation, (3)

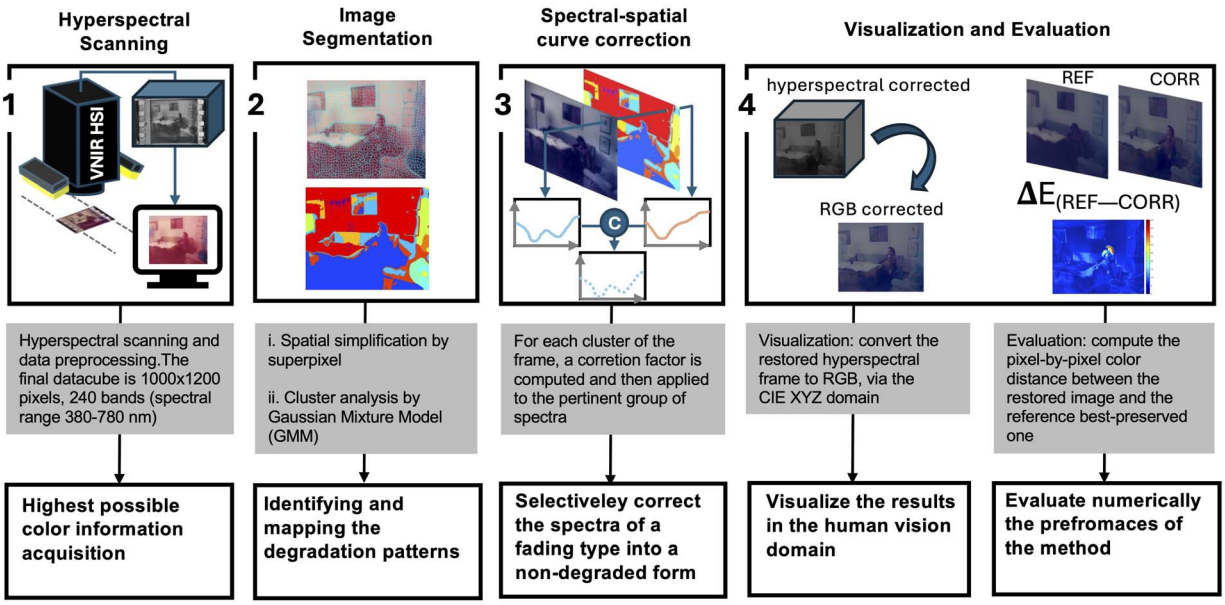

Fig. 1 Schematic representation of the cinematic film restoration approach. (1) Hyperspectral imaging scanning; (2) segmentation using clustering-based methods; (3) spectra correction (spectral–spatial curve correction) where the best-preserved frame acts as a source of not degraded spectra and it is used to correct the cluster-spectra of the faded image; (4) visualization of the results as RGB images and evaluation of the RGB color differences between the best-preserved frame and the restored ones.

spectra–spatial curve correction, and (4) visualization and evaluation.

Non-invasive hyperspectral imaging scanning (Fig. 1(1)) provides the initial spectral dataset of best-preserved and faded films. Image segmentation and spectral–spatial curve correction stages (Fig. 1(2) and (3)) allow the identification of similar faded areas across all the samples and successive perform color restoration. The two stages together are indicated as Cluster-Based Spectral Correction Algorithm (CBSCA). The last step of visualization and evaluation (Fig. 1(4)) implied first the transformation of the hyperspectral restored images into the RGB images, and afterward the comparison between the RGB corrected with the RGB best-preserved to assess the pixel-by-pixel color diversity. A conventional approach to frame restoration based on the commercial DaVinci Resolve software[33] was also considered as a further comparison method.

Hyperspectral imaging scanning

The hyperspectral image collection is the first step of our restoration approach. The data are acquired across two spatial dimensions (pixels) and one spectral dimension (wavelength), forming what is commonly referred to as a hyperspectral datacube.[34] The datacube can be initially inspected as a series of images collected at different narrow wavelengths or as a series of spectra extracted for each pixel of the image.

The four frames employed in this research are presented in Fig. 2a. They are visualized as RGB images, derived from the hyperspectral data through conversion *via* the CIE XYZ color space.[26] The frames were fully scanned including the lateral perforation, but only the central image of dimension 1000 × 1200 pixels (~1 K resolution) was submitted to digital color restoration.

The analyzed samples are positive chromogenic film frames, characterized by a three-layer structure applied to a cellulose base, with each layer containing yellow, cyan and magenta dyes, respectively.[20] From previous studies, it is known that the cyan fades first, followed by the yellow dye.[17,35] In the presence of a homogenous cyan dye fading, the result is a magenta color cast that is spread overall on the image.[35] Temperature and humidity are the main factors that trigger the dye deterioration.[36] Bearing in mind the degradation pathways, the degradation effects can be recognized by visually inspecting the RGB

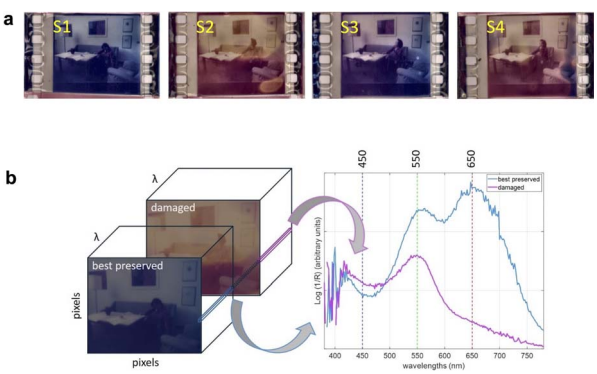

Fig. 2 RGB Images and spectra visualization from the scanning. (a) RGB frames obtained from the hyperspectral data; (b) spectra profiles from the same region of two aligned images, best preserved and faded. The difference in absorption bands of the two spectra clearly describe a specific fading effect.

images: the first frame from the left in Fig. 2a(S1), having the dominant blue color, was considered the least affected by the dye degradation and thus retained as the best-preserved. The other three show complex degradation patterns that involve either the discoloration of the cyan dye (Fig. 2a(S3)) or the discoloration of both cyan and magenta, resulting in a yellow strips or circles (Fig. 2a(S2) and (S4)). Some of the degraded frames show the original blue color exceptionally in limited areas (Fig. 2a(S4)).

By extracting a representative full spectrum from the same pixel position in the faded and not faded hyperspectral data cubes (Fig. 2b), it is clear that the technique can acquire superior information. What in RGB images is displayed as an intensity difference, in the spectral domain shows additional information such as peak shifts and curve rise, all of which can further contribute to better differentiate the degradation patterns and create unique correspondences between a specific degradation pattern and the color to be restored to. At the pixel level, the comparison of the two spectra (Fig. 2b) further indicates that the cyan dye (~650 nm) is almost completely degraded while the magenta dye (~550 nm) is only partially faded. The yellow dye does not show relevant differences due to degradatation.

Image segmentation

The second step (Fig. 1(2)) consists of the identification and grouping of the degraded parts by image segmentation and corresponds to part 1 of 2 of the Cluster-Based Spectral Correction Algorithm (CBSCA). The process divides the image into discrete regions, where pixels inside each region share similar characteristics.

In the present work, image segmentation is obtained by applying the method developed by Liu *et al.*,[23] namely, combining the Simple Linear Iterative Clustering (SLIC) algorithm with the Gaussian Mixture Model (GMM) clustering method. Particularly, initial spatial simplification of the images is achieved through the superpixel method, using the SLIC algorithm.[24]

The method groups adjacent pixels with high similarity in color properties and returns a new image fragmented into irregular but highly homogeneous macro pixels. For the sake of calculation efficiency, the superpixels method was applied to the RGB representations of the frames, obtained by converting the hyperspectral data first to CIE XYZ space and then to the RGB space. Afterwards, the calculated superpixel grid was applied to the hyperspectral data, and the average spectra of each super pixel were computed. Successively, we grouped the superpixels previously created for all four images independently[23] employing the GMM method.

Compared to previous studies,[23] the GMM has been built on the full set of hyperspectral images, thus, each superpixel is associated simultaneously with multiple clusters, with the final assignment determined by the highest likelihood. This probabilistic framework is especially beneficial for pixels lying near the boundaries between different groups, as it allows for more accurate and flexible cluster membership assignment, thereby improving the modeling of transitional regions. This multi-image strategy enhances the ability to identify and group similar colors across varying states of degradation, offering a richer understanding of both original colors and degradation patterns through the characteristic spectral profiles.

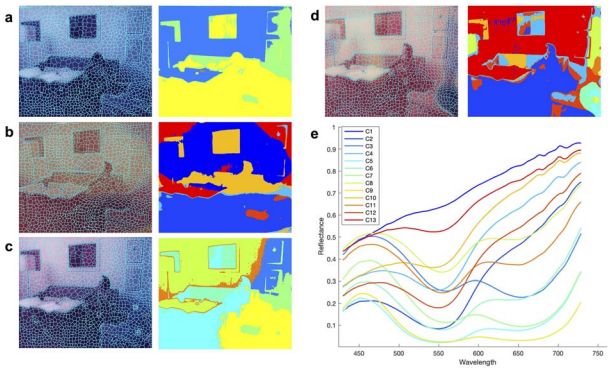

Fig. 3 Clustering superpixels. (a–d) The four frames partitioned into superpixels, alongside the result of the cluster analysis by Gaussian Mixture Models (GMM); (e) cluster centers for the 13 identified clusters.

The resulting clusters, visualized across all frames in Fig. 3, provide as a robust foundation for consistent and reproducible full-resolution color restoration.

The application of such a segmentation method to multiple hyperspectral images simultaneously represents an important step forward in film color restoration. It offers the potential to comprehensively describe all the degradation patterns within an entire film scene and shifts the perspective on the practical feasibility of employing such technique in full film restoration.

Spectral-spatial curve correction

The core of the method, indicated as spectral-spatial curve correction, is a spectral correction factor that is calculated independently for each cluster spectrum. In the restoration framework (Fig. 1(3)), the color restoration is undertaken in the spectral domain using the results of the GMM clustering to correct the faded spectral profiles to their original un-degraded shape. A cluster representing fading effect and its cluster spectrum are initially chosen and visualized (Fig. 4, step 1). The same spatial regions are selected inside the best-preserved image by overlapping the superpixel grid of each faded image to the best-preserved image (Fig. 4, step 2). A new cluster spectrum, corresponding to the selected areas in the best-preserved image is then computed. Having now two cluster spectra obtained from the same cluster extension, one from the faded image (S^f) and one from the best preserved (S^r), it is possible to calculate the correction factor (C^f) by simply subtracting the faded spectrum from the reference spectrum (Fig. 4, step 3). The correction factor is applied at each wavelength of the faded spectrum and later extended to all the original pixels belonging to the cluster to obtain the necessary image correction (Fig. 4, step 4). This approach provides a more accurate and objective color correction in the spectral domain, than

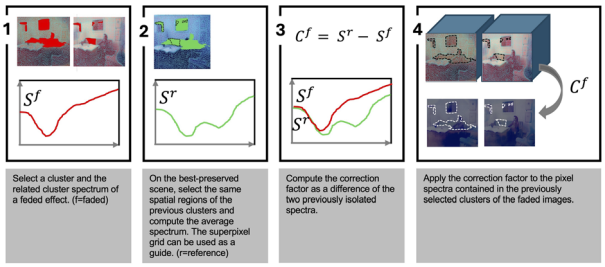

Fig. 4 Schematic representation of the spectral-spatial curve correction method. (1) Clusters describing a fading effect and representative averaged spectrum. (2) Application of the cluster grid of the faded images to the best-preserved scene and extraction of the averaged spectrum. (3) Estimation of the correction factor (C^f) as a difference between the spectrum extracted from the faded areas (S^f) and the spectrum extracted from the exact same area but on the best-preserved scene (S^r). (4) Application of the correction factor to the original hyperspectral data of the faded frames, paying attention at enabling only the pixels belonging to the clusters.

adjusting manually the hue, saturation, and other parameters in RGB-domain software.

An important concept that must be considered when developing our color correction methods is that each previously defined cluster describes a nearly homogeneous type of fading regardless of whether it is due to a single or a combination of multiple (linear or non-linear) effects. In this approximation, it is possible to consider only a single correction factor for all the pixels of a cluster and successively extend it to all the pixels belonging to the same cluster.

It is worth mentioning that the pixel spectra inside a cluster retain minimal but important spectral variation from the centroid. These variations maintain the color difference among the items in the image. Being summed to the single-pixel spectra, the correction factor retains and does not flatten the preexisting spectral color difference within an area. In brief, the spectral–spatial curve correction preserves the spectral heterogeneity while assuming similar, spatially more homogeneous deterioration patterns of the dyes.

To appreciate and further evaluate the results of our approach, the restored hyperspectral images were converted to RGB images. Adhering to the conventional method, the pixel spectra were initially converted in the XYZ domain and successively to RGB values. In this way, the additive process visualizes the color images with the superimposition of red, green, and blue images.

In the RGB domain, it was noted that the method is effective when restoring the internal portion of the faded area, but fails in the contour regions of the clusters, producing a too marked spectral gap between adjacent segments. This effect has an impact on the final color RGB images, which present an undesirable color contour. A smoother transition was thus obtained considering also the probabilities of the clustering as a weight for the correction. In the previous GMM soft clustering method, we used weights parameters, among others, to define the probability of membership of a superpixel to a cluster (assigning the pixel to the cluster with the highest probability). In this step, the matrix of the weights was used to modulate the correction factor, thus obtaining a final spectrum that retains the characteristics of the adjacent clusters, in accordance with the probability assignation. The final correction is obtained further refining the edges of the cluster areas. In particular, a new correction factor is computed comparing the partially corrected image with the reference one. The new correction for each pixel is obtained weighting the different correction factor by the number of pixels belonging to a given cluster in a selected neighborhood, a square region around the selected pixel of nearly 20 × 20 pixels over the contours. As an example, a pixel highly surrounded by the pixels of one cluster will be mostly corrected with the correction factor of that cluster. In this way, the strong differences between clusters edges are reduced assuring a smooth transition between regions with different corrections. This expands the previous assumption of homogeneity with variable degrees of the same degradation type within a superpixel.

Visualization and evaluation

The restored frames, obtained as hyperspectral datacubes were converted into RGB images *via* the CIE XYZ color space. In this domain, a better appreciation of the digital restoration and immediate visual evaluation of the performance of the method can be performed. In step four of the approach (Fig. 1(4)), we evaluate the goodness of our restoration approach by comparing the restored images with the reference best-preserved frame.

The results obtained by applying our CBSCA computational methods are shown in Fig. 5. By simple visual observation, it is possible to notice high similarity between the best-preserved image (Fig. 5a) and the corrected ones (Fig. 5c). Also, the color difference between the degraded areas is now barely perceptible thanks to the double-step approach for smoothing the contour of the clusters. The image R3, which originally presented only a magenta color cast on the left side of the image, is here completely restored with minimal color differences from the best-preserved. Regarding images R2 and R4, which originally presented yellow stains over a magenta color cast, there are still some minimal ghost effects due to the non-perfect color correction. However, the overall images appear clearer, and the pieces of the interior decoration are well diversified in color. Not always successful was the preservation of the magenta color of the lady's shirt and the painting hanging on the wall. This was because they were not considered as individual clusters, but the relative superpixels were grouped into an extended cluster with a more dominant bluish color correction factor. As proof of this, the magenta item on the desk, clustered together with more purplish areas, retained a more magenta color than it is present in the original image.

A more objective evaluation of the color differences is presented in Fig. 5d where a color difference ΔE defined as the difference between the RGB values of two colors, was calculated between the restored and the best-preserved images at the pixel level. This is a common method previously used by several authors[14,17] to evaluate the quality of a restoration. This

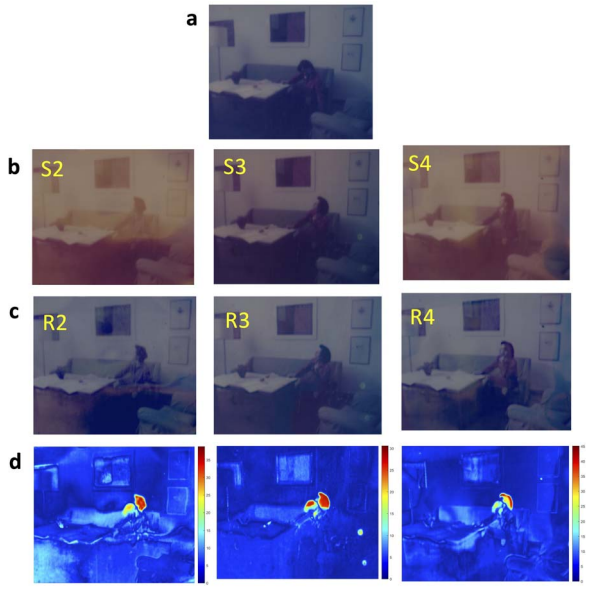

Fig. 5 RGB images of best-preserved, degraded and restored frames using CBSCA. (a) Image of the best-preserved frame; (b) images of the faded frames; (c) restored images by CBSCA (R); (d) ΔE differential maps of the best-preserved against restored.

evaluation is made possible since the images of the faded and best-preserved frames represent nearly the same scene and were submitted to a meticulous alignment work prior to analysis.

In general, low values of ΔE indicate high accuracy among the colors while high values of ΔE point to a significant mismatch. Performed at the pixel level, ΔE images were displayed in a false color scale where low ΔE is indicated in blue and high ΔE is in red. An average value of color difference dE per each restored *versus* best-preserved image is also provided, to allow a comparison of restoration through numerical quantity.

By observing the ΔE false color images, it is immediately clear that a marked red color, which corresponds to a high mismatch of the images, can be noted in correspondence with the head of the lady. The mismatch is not a mistake in color restoration, but as a natural difference between consequential scenes where the lady's head is moving. Conversely, considering the frame background, the major color differences due to restoration are enhanced in green, while the highest similarity is expressed in blue. In general, all four restored images (Fig. 5d) show a good level of similarity with the best-preserved ones, especially concerning the color of the wall/background which is in all cases dark blue. The decorative items, such as the sofa and the paintings, still present some green colors and stripes due to a minimal, non-perfect color correction. In the color maps, the contours of the original yellow stains also emerged, but with a minimal color variation. The average dE, reported in Table 1 shows the best color correction of R3 and a slightly worse value for the frames R2 and R4. Given the complexity of the above three frames and considering the low values of ΔE the restoration results can be considered highly satisfactory.

To provide a comparison between our CBSCA method and a common digital method used in the film restoration community, the faded images were corrected using the DaVinci Resolve software.[33] The best-preserved frame was still used as a reference, and the difference color maps were computed again between the restored and best-preserved (Fig. 6). It is worth mentioning that the restoration was performed directly on the RGB images, where the quality of the monitor/device used, and the skills of the restorer have a relevant impact on the final result. In this case, the approach used followed the steps reported by Liu *et al.*,[20] with the only difference that here we processed the projected RGB images obtained by the hyperspectral datacubes. A primary color balance is followed by a correction of hue, tint, temperature, contrast, and saturation. With a procedure called masking, the different degraded areas were selected and isolated from the rest of the images and further adjusted separately in color. Since different degraded areas are present, multiple masks were used. The masking approach can be viewed as a similar step to our segmentation, unavoidable when treating with inhomogeneous fading. However, in this case, the masking is less precise than the clustering method since the extension of the areas is decided by the user by visual observation, and not objectively calculated by the spectra similarity as our algorithm does. Furthermore, homogeneous corrections are applied to a much smaller number of masked areas than in the case of CBSCA, where corrections are performed on multiple thousands of superpixels.

The results obtained with DaVinci are presented in Fig. 6. It can be noted that the stains are still visible, and a blue color cast is more evident, thus losing the magenta hue that is a key color of some items in the original best preserved. By observing the ΔE differential images (Fig. 6d) it is possible to observe that DV3 is well corrected. Uniform color cast is, in fact, an easy task with DaVinci which is the reason the software performed well with this frame. In the other images, affected by a complex color degradation, the contour of the stains is more evident, as no edge smoothing step was applied. In addition, the internal parts of the stains and decorative items show a uniform greenish false

Table 1 Average color difference dE calculated between the best preserved S1 and the frames S2, S3 and S4, before and after color correction

	Color difference (dE)		
Method	Frame S2	Frame S3	Frame S4
Original color difference	28.3258	6.0306	24.4813
Cluster-based spectral correction algorithm (CBSCA)	5.1943	2.6355	4.1199
Da Vinci Resolve	6.6016	2.9266	4.5950

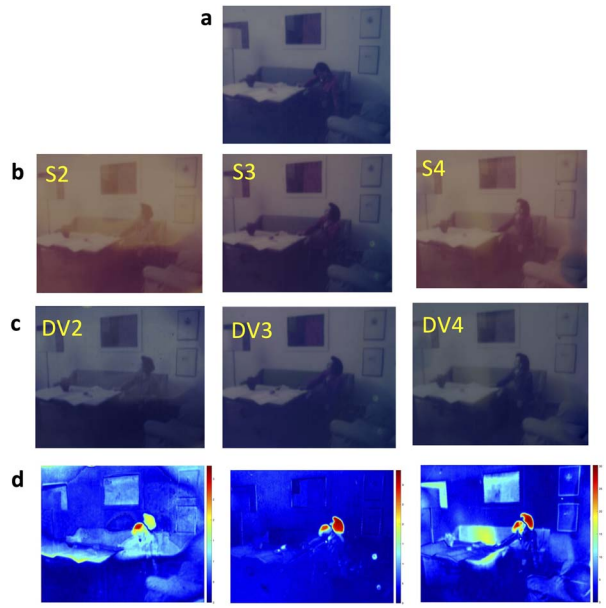

Fig. 6 RGB images of best-preserved, degraded and restored frames with DaVinci Resolve software. (a) Image of the best-preserved frame; (b) images of the faded frames; (c) restored images by Da Vinci Resolve software (DV); (d) ΔE differential maps of the best-preserved against restored.

color, meaning a relevant difference in color with the best preserved. Indeed, the dE average values (Table 1) of our method are lower than those obtained with DaVinci, demonstrating the superior quality of using hyperspectral imaging and the CBSCA for digital color restoration. Even for the image DV3, where the software application produce results with a very low dE, the color correction achieved with CBSCA is slightly better than the Da Vinci correction.

The dE color differences presented in Table 1 highlight an overall better performance of the CBSCA method compared to the work performed on DaVinci Resolve, although the closeness of the final values indicates that good results in restoration can be achieved with both methods. In the paper proposed by Liu *et al.*,[20] dE was used to evaluate the restoration of the S2 and S3 frames using Simple and Multicodebook approaches as well as DaVinci Resolve. Although the dE results are not directly comparable with those obtained by our method due to the difference in image resolution (4× lower in the previous work), the reduction from the original color difference to the post-restoration dE is larger in our work, suggesting better performances in color correction.

The comparison of the results of the CBSCA and conventional software suggests that CBSCA not only allows users to easily obtain better results, but also overcomes the major drawback of the conventional software, namely the highly subjective intervention in the inhomogenoeus color correction. Besides, the operation of color correction undertaken by the DaVinci software was revealed to be highly time-consuming, particularly due to the numerous masks (segments) created to correct the inhomogeneous fading. Conversely, by combining hyperspectral data with CBSCA, we demonstrated the possibility of having an objective method that efficiently addresses step-by-step the inhomogeneous color degradation.

Methodological considerations

The method proposed in this study represents a step forward in the automatic restoration procedure of inhomogeneous color-faded films. However, the method requires a non-degraded reference frame. This may appear to be an intrinsic limitation, as the method cannot be applied when such a reference does not exist. Nevertheless, a well-preserved reference frame – which is often difficult to find – could be obtained by singly restoring the best-preserved frame among a series of degraded frames. Therefore, a manual, software-based restoration of the least degraded frame could be used to generate a suitable reference for applying CBSCA to a given scene. While this initial step would be time-consuming and inherently subjective, the restoration of the remaining frames affected by inhomogeneous color degradation could be significantly accelerated using CBSCA. It is important to note that the identification of degradation patterns using SLIC and GMM is not scene-dependent, as GMM may assign the same clusters to similar degradation effects across multiple frames. In contrast, the spectral curve correction is scene-specific, and the source of well-preserved spectral information must be provided for each scene to be restored. Moreover, the spectral correction must be applied to temporally adjacent frames, as it is a linear subtractive method. In fact, for temporally adjacent frames, the non-linear effects (*e.g.*, log-linear or multiplicative in absorption) are small and linear subtraction is approximately correct.

The CBSCA algorithm comprises three distinct steps: SLIC, GMM, and spectral–spatial curve correction. Although cluster selection in SLIC and GMM can be guided by quantitative criteria, the optimal number of clusters is also determined by qualitative assessment. Specifically, the selection of the clusters is conducted by integrating quantitative measurements with visual inspection of the RGB images and the discoloration effects. This introduces a limitation, as the approach is neither fully systematic nor entirely objective and may require additional time to determine the final number of clusters. Strictly considering the resources required by the algorithms, the computational complexity of the cluster-based SLIC and GMM algorithm is estimated with Big-O notation, with the variables defined as n = number of data points (pixels or superpixels) in an image, k = number of components (clusters),d = spectral dimensionality, f = number of frames, and i= number of iterations. The computational complexity of SLIC is relatively low, since $O(n \times d)$, where n corresponds to the number of pixels in the initial image and d includes only three RGB bands and the two spatial coordinates. Since $d = 5$, the complexity for RGB images is effectively linear in the number of data points.[37] Increasing the data to the full hyperspectral data is unnecessary, as we demonstrated in the Methods section that the three-band images are sufficient to model the degradation variability. Conversely, the computational complexity of GMM is higher, $O(n \times k \times i \times d^2)$,[38] and it becomes even more significant when

GMM is applied simultaneously across multiple frames, $O(n \times k \times i \times d^2 \times f)$. Therefore, to maintain manageable computational complexity while increasing the number of frames, the number of pixels must be reduced. This is achieved by introducing a superpixel step before GMM, reducing the GMM input size while preserving essential information.

These considerations can be quantitatively measured in terms of algorithm runtime, *i.e.* computational efficiency. When GMM is applied directly to the full hyperspectral dataset (1200 pixels × 1000 pixel × 240 wavelengths), it requires approximately 214 seconds per frame. By first applying SILC which reduces the data to approximately 7.5 k centroids × 240 wavelengths, the runtime of the full process (SLIC and GMM) decreases dramatically to approximately 30 seconds for processing all the four images together, with SLIC accounting for approximately 6 seconds per image. In contrast, evaluating the computational complexity and efficiency of Da Vinci Resolve is not straightforward. While individual operations such as color, hue and saturation adjustments are relatively fast, applying manual masks – necessary to independently correct differently degraded areas within each frame – can take tens of minutes to several hours, depending on the user's experience and the desired level of refinement. In cases of inhomogeneous fading, DaVinci Resolve requires correcting each frame individually, making the process considerably more time-consuming than the CBSCA approach.

Some considerations must also be made regarding the hyperspectral scanning. Film frames may suffer from distortion of the support, which represents a major source of registration error in image acquisition and can compromise both the quality of digital restoration and the reproducibility of the algorithm application across different frames. To avoid this issue, a custom-made film holder was designed to keep the different frames flat over the Spectralon white support. Specular reflection from the films' translucent surfaces can also introduce registration errors. We addressed this problem by using a set of cross-polarizers in the acquisition pathway. Under these conditions, external sources of errors are minimized while hyperspectral acquisition parameters remain constant, enabling high reproducibility in image acquisition.

The acquisition in reflection geometry was adopted as it represents the commonly used configuration in hyperspectral imaging. The decision to use reflection scanning was also motivated by the high-quality spectra in the visible range. In part, this is due to light travelling through the sample twice, thereby effectively doubling the interaction path and improving signal quality. Although the reflectance values collected from a faded dye are not a rigorous proxy for the dye concentration in color film (unlike transmission scanning), we optimized the scanning conditions to obtain reliable spectra for restoration. Since we do not attempt to estimate absolute dye concentration, maintaining constant scanning parameters and illumination conditions, comparing frames that are close in time, and using cluster aggregations that reduce noise and average local variability allow for effective capture of dye fading through reflectance scanning. However, the proposed method's versatility also makes it suitable for transmission-mode hyperspectral images.

Conclusion

The approach we developed for digital restoration has resulted in low dE values, indicating high-quality outputs that closely match the best-preserved references even when starting from severely degraded frames. This success is largely due to the synergistic integration of hyperspectral imaging (HSI) with a novel computational pipeline, the Cluster-Based Spectral Correction Algorithm (CBSCA), capable of tackling the challenge of inhomogeneous dye fading in an objective and reproducible way (Fig. 1).

HSI offers a significant advantage in color acquisition by capturing the full reflectance spectrum for pixel, rather than a simple RGB triplet as with conventional scanners. This spectral richness not only allows for precise color reconstruction, but also provides more information about degradation phenomena such as dye fading. However, this benefit comes at the cost of data overload, since each pixel in each frame generates hundreds of spectral bands.

To address this issue, we adopted a superpixel-based dimensionality reduction method. While the use of SLIC superpixels and GMM clustering was previously introduced by Liu *et al.*[23] for individual frames, the present study introduces a key innovation: applying this segmentation framework simultaneously across multiple hyperspectral images (Fig. 3). This multi-frame approach enables the identification of common degradation patterns that recur across the sequence, significantly enhancing the interpretability and consistency of the restoration.

The CBSCA introduces a spectral–spatial curve correction strategy (Fig. 4), which operates entirely in the spectral domain. Unlike conventional RGB-based methods that rely on manual masking and subjective tuning, our framework computes spectral correction factors based on matched regions between degraded and best-preserved frames. This correction is then applied to all pixels belonging to a given cluster, achieving an objective and scalable method that maintains the original spectral variations within each region.

The results demonstrate not only superior restoration quality compared to standard tools, but also the scalability of the method to higher-resolution hyperspectral systems and extended film sequences.

In conclusion, our work advances the field of digital film restoration by introducing a multi-image clustering strategy for hyperspectral segmentation, and by applying a spectral-domain restoration correction that preserves fine details and spatial coherence.

This framework is well-suited for large-scale applications and offers a promising path toward automated and high-fidelity restoration workflows for cinematographic films.

Author contributions

SP and FZ were responsible for the conceptualization of the study and provided overall supervision throughout the project. JS, EC and MM performed the hyperspectral scanning acquisition. LL and JF designed the research methodology and

developed the Cluster-based spectral correction algorithm (CBSCA). EC and LL developed the evaluation process of the restored frames. EC, LL and JF wrote the original draft. SP, FZ, LL, JF, MM, and JS reviewed and edited the final manuscript. SP submitted the manuscript with the consent and approval of all authors.

Conflicts of interest

There are no conflicts to declare.

Data availability

The hyperspectral data of the frames, along with the algorithms developed in MATLAB and Python for digital color restoration, are retained by the authors. Requests for additional information or data availability should be directed to the corresponding authors.

Supplementary information (SI): Fig. SM1: principal component analysis for the S1–S4 hyperspectral data cubes. Fig. SM2: average internal variance of the superpixels as a function of the number of superpixels. Fig. SM3: within-cluster sum of square (WCSS) as a function of the number of clusters computed using the GMM. Fig. SM4: graphical representation for the calculation of the weights, used to remove the border effects of the clusters. Fig. SM5: color distance between the corrected frames and the reference one at different neighborhood size values. See DOI: https://doi.org/10.1039/d5ra08504g.

Acknowledgements

The authors would like to express their gratitude to the Fondazione Cineteca di Bologna (Bologna, Italy) and to the laboratory "L'Immagine Ritrovata" for providing the valuable positive films that were scanned and restored in this study. The film scanning was made possible through financial support from the Italia-Slovenia bilateral project entitled "Advanced imaging technologies for the study of works of art".

References

1 T. Gunning, The cinema of attractions Early film, its spectator and the avant-garde, *Theater and Film: A Comparative Anthology*, R. Knopf, Yale University Press, New Haven and London, 2005, pp. 37–45.
2 Recommendation for the Safeguarding and Preservation of Moving Images, UNESCO Archives, https://www.unesco.org/en/legal-affairs/recommendation-safeguarding-and-preservation-moving-images, accessed 09 September 2024.
3 S. Bellotti and A. Plutino, Color Design & Technology A Multidisciplinary Approach to Colour Part 1, *Research Culture and Science Book Series*, A. Plutino, G. Simone and A. Rizzi, Gruppo del Colore – Associazione Italiana del Colore, 2022, vol 5, pp. 73–100.
4 M. Chambah, Digital film restoration and image quality, *Paper Presented at the ICA-Belgium Colour Symposium 2019 - Connecting Colour in Design, Art & Science*, Belgium, 2019.
5 https://filmcare.org/visual_decay, accessed 10 January 2026.
6 V. Martínez Marín, *Intervención*, 2020, **11**, 34–80.
7 https://www.nfsa.gov.au/preservation/preservation-glossary/preservation, accessed 10 January 2026.
8 https://www.immagineritrovata.it/en/film-restoration/servizi/, accessed 04 July 2025.
9 S. Solomon and P. Lennie, *Nat. Rev. Neurosci.*, 2007, **8**, 276–286.
10 B. Flueckiger, D. Pfluger, G. T. S. Croci, T. Aydn and A. Smolic, *Investigation of Film Material–Scanner Interaction*, Report Ver, Zurich, 2018, vol 1.
11 https://www.ilpost.it/2024/01/18/restauro-film/, Accessed 16 September 2024.
12 G. Trumpy, J. Y. Hardeberg, S. George and B. Flueckiger, *Proc. SPIE*, 2021, **11784**, 138–146.
13 O.-M. Machidon and M. Ivanovici, *J. Cult. Herit.*, 2018, **33**, 181–190.
14 A. Rizzi, C. Gatta, C. Slanzi, G. Ciocca and R. Schettini, in *8th International Conference, VISUAL 2005*, Springer Berlin Heidelberg, Amsterdam, The Netherlands, 2006, pp. 1–12.
15 A. Rizzi, A. Berolo, C. Bonanomi and D. Gadia, *Multimed. Tool. Appl.*, 2016, **75**, 3747–3765.
16 A. Plutino and M. Tarini, *IEEE Trans. Image Process.*, 2023, **32**, 2786–2799.
17 G. Trumpy and B. Flueckiger, in *2018 Colour and Visual Computing Symposium (CVCS) IEEE*, 2018, pp. 1–5.
18 G. TrumpyS. Chatterjee, *Proceedings of the Archiving 2022*, Society for Imaging Science and Technology, 2022, pp. 101–105.
19 S. Chatterjee, G. Trumpy and U. Ruedel, *Heritage*, 2023, **6**, 3418–3428.
20 L. Liu, E. Catelli, A. Katsaggelos, G. Sciutto, R. Mazzeo, M. Milanic, J. Stergar, S. Prati and M. Walton, *Sci. Rep.*, 2022, **12**, 21982.
21 J. Stergar, R. Hren and M. Milanič, *Sensors*, 2022, **22**, 6274.
22 L. Liu, G. Delnevo and S. Mirri, *J. Big Data*, 2023, **10**, 1–16.
23 L. Liu, A. Malmert, E. Pouyet, S. Mirri and G. Delnevo, *IEEE Access*, 2025, **13**, 108266–108283.
24 R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, 2274–2282.
25 H. Deng and J. Han, in *Data Clustering. Algorithms and Applications*, ed. C. C. Aggarwal and C. K. Reddy, CRC Press, 2018, pp. 61–86.
26 M. Magnusson, J. Sigurdsson, S. E. Armansson, M. O. Ulfarsson, H. Deborah and J. R. Sveinsson, in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2020, pp. 2045–2048.
27 N. Mobaraki and J. M. Amigo, *Chemom. Intell. Lab. Syst.*, 2018, **172**, 174–187.
28 G. J. McLachlan, S. X. Lee and S. I. Rathnayake, *Annu. Rev. Stat. Appl.*, 2019, **6**, 355–378.
29 G. Schwarz, *Ann. Math. Stat.*, 1978, 461–464.
30 H. Bozdogan, *Psychometrika*, 1987, **52**, 345–370.
31 N. Amruthanath and T. Gupta, *Ind. Eng. Manag.*, 2019, **8**, 1.
32 S. Bhattacharya and P. D. McNicholas, *Adv. Data Anal. Classif.*, 2014, **8**, 45–61.

33 DaVinci Resolve, *ver. 18.6*, (Blackmagic design, Australia), https://www.blackmagicdesign.com/it/products/davinciresolve/color, accessed 06 July 2025.
34 M. Picollo, C. Cucci, A. Casini and L. Stefani, *Sensors*, 2020, **20**, 2843.
35 G. Weaver, Z. Long, Chromogenic Characterization: A Study of Kodak Color Prints, 1942-2008, in *Topics in Photographic Preservation*, Photographic Materials Group, AIC, 2009, vol. 13, pp. 67–82.
36 https://www.filmcare.org/vd_dyefade.php, accessed 06 June 2025.
37 R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, Slic superpixels, *Tech. Rep.*, 2010, **149300**, 155–162.
38 S. Salwig, T. Kahlke, F. Hirschberger, D. Forster and J. Lücke, *arXiv*, 2025, preprint, arXiv:2501.12299, DOI: 10.48550/arXiv.2501.12299.