



**HAL**  
open science

# Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIELab Color Space

P.-A Herrault, David Sheeren, Mathieu Fauvel, Martin Paegelow

► **To cite this version:**

P.-A Herrault, David Sheeren, Mathieu Fauvel, Martin Paegelow. Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIELab Color Space. The 16th AGILE International conference on geographic information science, May 2013, Louvain, Belgium. pp.95 - 112, 10.1007/978-3-319-00615-4\_6 . hal-01452677

**HAL Id: hal-01452677**

**<https://univ-tlse2.hal.science/hal-01452677>**

Submitted on 2 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIELab Color Space

P.-A. Herrault, D. Sheeren, M. Fauvel and M. Paegelow

**Abstract** In this chapter, we describe an automatic procedure to capture features on old maps. Early maps contain specific informations which allow us to reconstruct trajectories over time and space for land use/cover studies or urban area development. The most commonly used approach to extract these elements requires a user intervention for digitizing which widely limits its utilization. Therefore, it is essential to propose automatic methods in order to establish reproducible procedures. Capturing features automatically on scanned paper maps is a major challenge in GIS for many reasons: (1) many planimetric elements can be overlapped, (2) scanning procedure may conduct to a poor image quality, (3) lack of colors complicates the distinction of the elements. Based on a state of art, we propose a method based on color image segmentation and unsupervised classification (K-means algorithm) to extract forest features on the historical 'Map of France'. The first part of the procedure conducts to clean maps and eliminate elevation contour lines with filtering techniques. Then, we perform a color space conversion from RGB to L\*a\*b color space to improve uniformity of the image. To finish, a post processing step based on morphological operators and contextual rules is applied to clean-up features. Results show a high global accuracy of the proposed scheme for different excerpt of this historical map.

---

P.-A. Herrault (✉) · D. Sheeren · M. Fauvel  
University of Toulouse, INP-ENSAT, UMR 1201 DYNAFOR, Av. de l'Agrobiopôle,  
BP 32607, Auzeville Tolosane, Castanet Tolosancedex, Toulouse 31326, France  
e-mail: pierrealexis.herrault@ensat.fr

D. Sheeren  
e-mail: david.sheeren@ensat.fr

M. Fauvel  
e-mail: mathieu.fauvel@ensat.fr

P.-A. Herrault · M. Paegelow  
University of Toulouse, UTM, UMR 5602 GEODE, 5, allée A. Machado,  
Toulouse 31058, France  
e-mail: paegelow@univ-tlse2.fr

## 1 Introduction

Olds maps contain specific spatial information as location of historical places or historical land cover, elevation contour lines, building footprints and hydrography. Capturing this spatial information interest for various studies as those one about long term changes of landscapes, urban development or coastlines evolution (Cousins 2001; Bender et al. 2005; Gimmi et al. 2011; Smith and Cromley 2012). For few years, a lot of these old maps are available for research thanks to the work of scanning by the national archives in different countries.

The traditional approach to capture the cartographic objects in the old maps is based on a user intervention (for digitizing). This approach is obviously very time-consuming, and difficult to reproduce on large areas. Several works attempted to develop automated data captures techniques in order to establish reproducible procedures (Leyk 2006). However, most of them are specific to only one particular map and generally, cannot be applied on other historical maps.

In this chapter, we propose a new method to extract automatically forest features from the historical ‘Map of France’ dating from the 19th century. The method is based on image recognition techniques including a color space transformation and an unsupervised classification of the digital map. The chapter is structured as follows. In Sect. 2 we describe existing works developed to the automatic processing of old maps. Next, the method we propose is presented in Sect. 3. Experiments and results obtained on several excerpts of the considered historical map are presented in Sect. 4. Finally, we draw some conclusions in Sect. 5.

## 2 Previous Works

Capturing features automatically from raster maps is a major challenge in GIS for many reasons. First, there are many planimetric elements that overlap each other such as road lines, elevation contour lines, marks or soil features, and so on. Second, scanning procedure or image compression may conduct to a poor quality of the data that make the recognition more difficult. Last, some old maps may be in black and white (without any other colors) that make the development of an automatic procedure almost impossible (Fig. 1).

Several authors already proposed automatic methods to capture geographical objects in scanned thematic maps (Ansout et al. 1990). Many examples of maps have been investigated for feature recognition such as the topographic maps of the United States Geological Survey, the military maps of the Polish Geographic Institute (Iwaniowski and Kozak 2012) or the Swiss National historical map (Leyk 2006).

In a general way, automatic extraction procedure can be divided in three steps: (1) a cleaning-up step, (2) a feature recognition/extraction step and (3) a post-processing step. The first step is defined to make the feature extraction easier while the last step is carried out to improve the extraction results. Steps 2 and 3 can be



**Fig. 1** **a** Overlapping of planimetric elements on historical ‘map of France’ (~1850); **b** review of the historical ‘map of France’ in black and white (~1900)

viewed as a single integrated process according to some authors (Wise 1999). In the next sections, we described the existing approaches for old maps processing by combining the pre and post-processing steps (1 and 3) in the same part because of the similarity of the techniques.

## 2.1 Existing Pre and Post-processing Methods

Generally, scanned thematic or old maps are complex and contain different kinds of noise. Indeed, many elements that overlap each other (such as road lines, elevation contour lines or text) can be seen as noise when the users want extract specific information. Scanning procedures may also conduct to a shading effect on the maps and alter the quality of the original document. This noise makes difficult the use of the digital maps and in particular the automatic extraction of some features of interest (e.g., the buildings or the forests).

To overcome this problem, digital historical maps are often filtered using image-processing techniques. The filters consist in assigning a new value in each pixel using the pixel values in its neighborhood. They are well adapted to reduce noise in the images while preserving some structural elements like edges or contours. Various filters exist in the image-processing field ranging from convolution filters to morphological filters.

Ansault and Soille (1990) proposed a morphological-based filtering step in their thematic map processing method. They showed how dilatation operator and image reconstruction process can be used successively to remove text in the map which complicates regions extraction. The proposed approach provides a general

framework to deal with this classical issue. Chiang et al. (2012) also adopted morphological filters in a pre-processing step. He applied successive erosions to remove elevation contour lines before to extract the roads which are very similar. A close approach is followed by Samet et al. (2010) in a post-processing stage. A dilatation operator is applied to reconnect broken contour lines and reduce noise in the map after a color image segmentation.

Iwanowski and Kozak (2012) used an image cleaning run two times in their extraction procedure. One time before segmentation, in order to filter the image and remove small pixel scale variations of colors and another time, after image segmentation, using successively a closing by reconstruction and an opening by reconstruction to improve forest detection results. Arrighi and Soille (1999) in a similar perspective, used some morphological filters to produce a clean mask of the elevation contour lines before a real line detection.

## ***2.2 Existing Features Extraction Methods***

Data capture on historical maps may concern various types of features such as (a) text (marks, names of cities or rivers...), (b) regions (land cover, buildings,...), (c) symbol (semantic information or punctual objects like churches, mills...) or (d) lines (elevation contours, edges, roads...). Each of these feature categories may conduct to define specific techniques to extract them automatically.

Text extraction is one of the most complicate tasks to automate in old maps. Indeed, text and graphics are often overlapped that make difficult the separation of the elements and therefore text recognition. Cao and Tan (2002) proposed a method for text recognition based on the observation that the strokes of characters are generally more short segments than those of cartographic elements. The authors used a combination of line continuation with the line width to separate elements and improve text extraction. Text can be also detected using color attributes when the color characters differ from the others objects. Centeno (1998) used for example a Karhunen-Loeve color-space conversion to isolate text and capture it. Last, Myers et al. (1996) performed a character recognition method with a verification-based approach to detect text without requiring pre-segmentation graphical entities.

Concerning regions, existing methods to extract them automatically are mainly based on image segmentation and/or classification techniques. Shaw and Bajcsy (2011) used a combination of a region-growing segmentation and morphological operators to extract the Ontario Lake in several maps. The results shows that this technique is limited when there are numerous overlapped objects. Mahmouda et al. (2011) proposed a similar approach including a multi-resolution segmentation and a (rule-based) classification procedure using color and spatial attributes to extract features areas. Other approaches consist in classifying the image without segmentation step in advance. If the maps contain thematic colors layers, color signature can provide much information to capture objects of interest but means some limitations when processing poor image quality (Leyk 2006).

Chiang et al. (2012) used K-means algorithm to capture road vectors in a raster map. Results show that algorithm is efficient but that the users need to determine a large number of K clusters to separate different features. Henderson and Linton (2009) carried out a comparison between k-means algorithm, expectation-maximization and graph theoretic to separate semantic classes in a raster maps. Threshold-based classification methods are also used by some authors to separate color or grayscale components related to thematic classes. This way is also used as a first step before to set up a more accurate segmentation method (Ansoult 1990; Chi and Yan 1993).

The extraction of map symbols is complicated by the fact they can form themselves discontinuous chains (Gamba and Mecocci 1999). A majority of works employed knowledge-based approaches (Arias et al. 1993) to perform a symbol extraction based on set of rules. Other approaches identify the symbols in the legend, constructing a training set library to then use it in the classification of the geographic symbols (Samet and Soffer 1998). It is also possible to consider symbols as short lines and then, applying line extraction techniques to symbol recognition (Boesch 1996).

To finish, line extraction in maps may concern contour-objects lines, elevation contour lines, traffic connections lines, drainage networks etc. Kaneko (1992) used directional distances propagations for extracting lines structure in line drawings whereas Mariani et al. (1997) proposed several reconstruction algorithms for the identification of drainage networks. When the maps are printed in colors, this information can provide very reliable results for line extraction. For example, Ansoult et al. (1990) used the mean and variance of the hue channel to extract regions contour in a soil map. Arrighi and Soille (1999) adopted a different approach in extracting extremities of the lines before thinning them with a skeletonization algorithm. In last, Chen et al. (1999) proposed a new method based on local window segmentation approach to overcome thick lines.

All these methods appear as efficient in particular contexts but are not suitable for all supports. Thus, each raster-map contains specific problems of overlapping or quality so an identification step is required to identify the characteristics of the map before choosing the procedure to apply.

### 3 Proposed Method

In this section, we expose our extraction method of the forest features for the historical ‘Map of France’ (Fig. 2). Our procedure includes the following steps:

- (1) Pre-processing for map filtering
- (2) Color space transformation from RGB to CIELab space
- (3) Color image classification based on K-means algorithm
- (4) Post-processing of forest features.

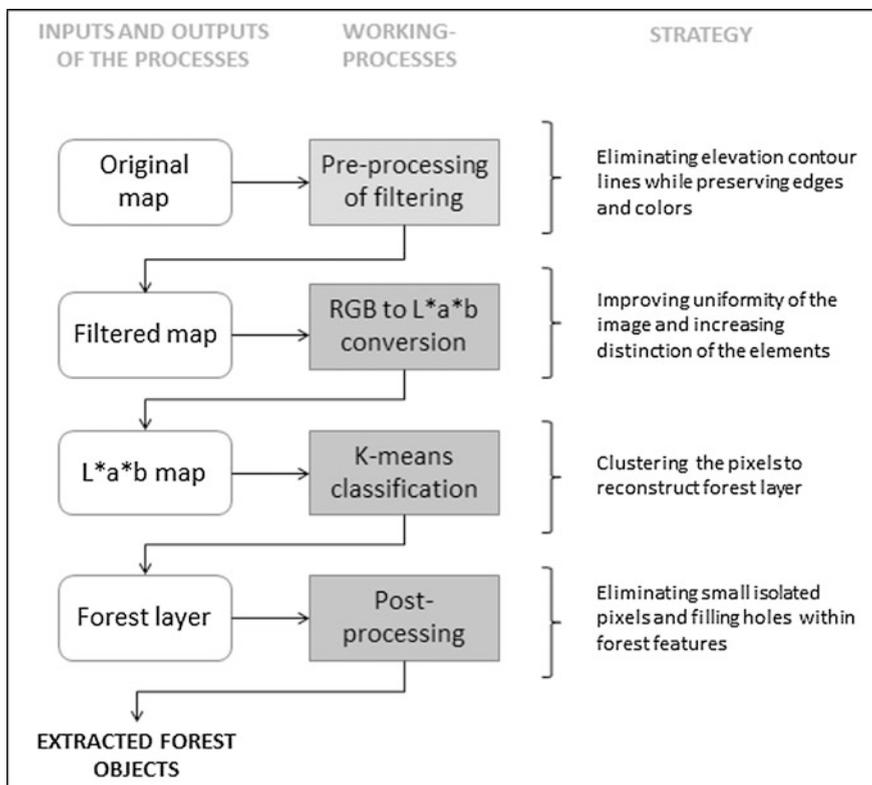


Fig. 2 General methodology (flowchart inspired by Leyk et al. 2006)

### 3.1 Pre-processing for Map Filtering

Before performing Color Image Classification, a clean-up process is applied to eliminate elevation contour lines which overlap forests. The goal is to remove these contour lines while preserving colors and edges of objects. Three steps are performed: (1) dilatation, (2) median filtering, (3) low-pass filtering (Fig. 3).

A first step was dedicated to fill all holes within forest (created by text, symbols and elevation contour lines) except those one created by roads. We choose to use a morphological dilatation operation which allows filling holes thanks to a structuring element (SE). The best SE we found is a square of  $5 \times 5$  pixels. This operator is applied on each band of the image (R, G, B).

Secondly, we want to simultaneously reduce remaining elevations contour lines pixels while preserving edges and colors. As well-known, two-dimensional median filtering appears as efficient for this objective (Samet et al. 2010). Nevertheless, a neighborhood size ill-suited may conduct to important cost of



**Fig. 3** Flowchart of the pre-processing of filtering

blurring and makes edges invisible in the image (Motwani et al. 2004). We choose to perform a median filter in a window of  $5 \times 5$  on each dilated band R, G and B. We applied it five times until obtaining a result that we judge satisfactory.

To finish, we have observed a light remaining background noise. Another application of a median filter would conduct to make edges invisible. To overcome this problem and to proceed to a sweet objects extraction, we propose to use a low pass filter in a window  $5 \times 5$  which allows eliminating background noise while preserving low frequencies.

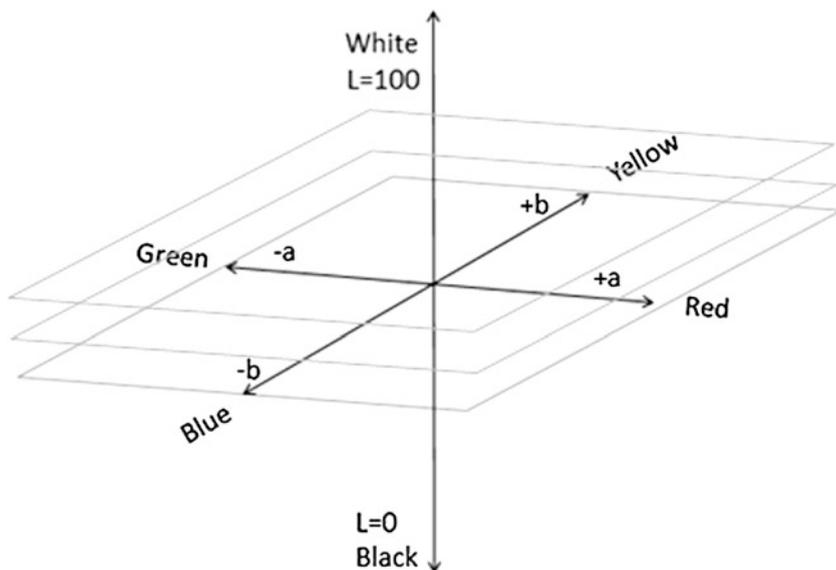
### 3.2 Color Space Transformation from RGB to CIELab Space

Previous steps allowed us to filter objects but colors are disturbed during this process: hue is heterogeneous for a same object class and intensity fluctuates according to low slope areas or high slope areas. Globally, color is rarely homogenous in historical maps so there is a need to choose a color space which assures certain uniformity in luminosity.

The choice of a suitable color space is often a crucial step for old color map processing. The traditional color space is the RGB space. However, this color space presents some limitations for the automatic extraction of forests: non-uniformity of the luminosity, lack of human perception (Angulo and Serra 2003). Other color spaces are well-known for computer graphic applications like Hue, Saturation, Value (HSV) or Hue, Luminosity, Value (HLS) but are less suitable for image processing (Hanbury and Serra 2003).

In this method, we propose to transform the image into the  $L^*a^*b$  (or CIELab) color space. This choice is inspired by previous works showing that CIELab is appropriate for color images with various types of noises but not only (Ganesan et al. 2010). The performance of different unsupervised classification (particularly K-means clustering) has been analyzed in few color spaces like RGB (Brunner et al. 1992), HSV or  $L^*a^*b$  (Wiszeki and Stiles 1982). The general consensus implies that  $L^*a^*b$  is more efficient since removing effects of illuminations yields optimal segmentation results.

In this space, the characteristic known as luminosity is reported on an axis L that is perpendicular on a pile of 'ab' planes where each plane contains all the possible colors for a given luminosity (Fig. 4). This axis represents the sensitivity of the human eye to the luminance. Thus, it is possible to consider each variation



**Fig. 4** The axis  $L$  perpendicular on a pile of ' $ab$ ' planes

of the green color (e.g., forest features) like a succession of pure colors on axis ' $a$ ' since this axis informs us on the degree of green which qualifies the objects. We attempt to search different green colors of the forests for each possible value of  $L$ .

This allows us to take into account the green variations within forest areas. For example, in RGB color space, two green pixels with a different luminosity could be separated into two different classes by a clustering algorithm because this space does not take into account luminosity but only the proportion of green within pixel.

The colors are uniformly distributed in an  $a*b$  plane, from red to green along the  $a$  axis and from blue to yellow along the  $b$  axis. The three coordinates of the  $L*a*b$  color space is given as following:

$L$  the luminosity layer

$a$  indicates where color falls on the axis from red to green

$b$  indicates where color falls on the axis from yellow to blue (Fig. 5).

Our goal is to find the most representative partition of the forests in the  $ab$  chromaticity diagram for each possible value of  $L$ . This partition can be found automatically in the 2-dimensional space by clustering.

**Fig. 5** An 'ab' plane for only one value of L (from Barbu et al. 2012)



### 3.3 Color Image Classification Based on K-Means Algorithm

K-means is one of the simplest and most popular unsupervised learning algorithm. This technique is used in various application domains including computer science, ecology, geostatistic or remote sensing (e.g., Lucchese and Mitra 1999; Jigar et al. 2012).

As a reminder, the main idea of the algorithm is to define a partition of N observations into K clusters in which each observation belongs to the nearest cluster (Barbu et al. 2012). K-means algorithm is given as following:

$$J_{K\text{-means}} = \sum_{k=1}^K \sum_{j \in S_k} d^2(X_j, C_k)$$

- where  $K$  is the number of the cluster (class) evaluated in a space defined by  $S_K$
- $(X_j, C_k)$  is the distance between the observation  $X_j$  and the class  $C_k$ .

The K-means algorithm is performed in four steps:

- (1) Place  $K$  points into the space represented by the observations that are being clustered
- (2) Each observation is assigned to the closest point of  $K$  defined as the centroid of one cluster  $C_k$
- (3) When all the observations have been assigned to the  $K$  clusters, the centroids of the obtained clusters must be recomputed
- (4) This procedure (steps 2 and 3) is iterated until the algorithm converges (i.e. until there are no cluster changes for the observations after the re-computation of the centroid positions) or that it respects a fixed stopping criterion.

The difficulty in K-means algorithm resides in the definition of the optimal number of clusters  $K$ . Because of a real complexity of old maps, number of colors can widely vary between one scan and another. The difficulty in K-means algorithm resides in the definition of the optimal number of clusters  $K$ . Because of a real complexity of old maps, number of colors can widely vary between one scan and another. Here, we decided to compute ten clusters ( $K = 10$ ) in order to extract the forest features. This number enables us to better separate different thematic classes while avoiding mixed clusters. Some clusters are merged after this step in order to obtain a single layer including only forest and non-forest features.

### ***3.4 Post-processing for Features***

A post-processing step was carried out to correct some artifacts in the forest layer after the classification step. This post-processing includes a morphological opening in order to remove the small isolated pixels. The opening operator was applied with a structuring element of  $3 \times 3$ . Some holes within forests were also removed using contextual rules: all the non-forest pixels enclosed by forest pixels were classified as forest. This operation created homogeneous forest objects. Finally, the original image including only the forests was reconstructed.

## **4 Results and Discussion**

The method was applied on the historical “Map of France”. This map was produced from 1825 to 1866 at 1:40,000 scale and is known for its relatively high planimetric accuracy. The map includes several thematic categories (like forests, buildings, grasslands...) which are represented in color. Experiments were conducted on three different excerpts of the map which differ in terms of slope and relief, quality and colors for the forest features (various green levels) (Fig. 6).

The performance of the method was assessed quantitatively by computing a confusion matrix and some related accuracy index. Here, we give also the intermediate results in order to illustrate the interest of each step. The effect of the filtering step (which enables to eliminate elevation-contour lines) is illustrated in Fig. 7.

In Fig. 7a, we can observe on the histogram of the original excerpt 2 (in gray levels) an important number of pixels in low values (gray level  $< 100$ ). These pixels are related to the elevation contour lines which overlap to regions. Thanks to the dilatation operator, a great part of these lines were removed while preserving edges (Fig. 7b). However, this filtering was not sufficient because too many pixels with low values remain.

The second step of filtering conducts to further reduce these dark pixels in order to smooth the image. It is interesting to observe on the histogram of the image (c) that

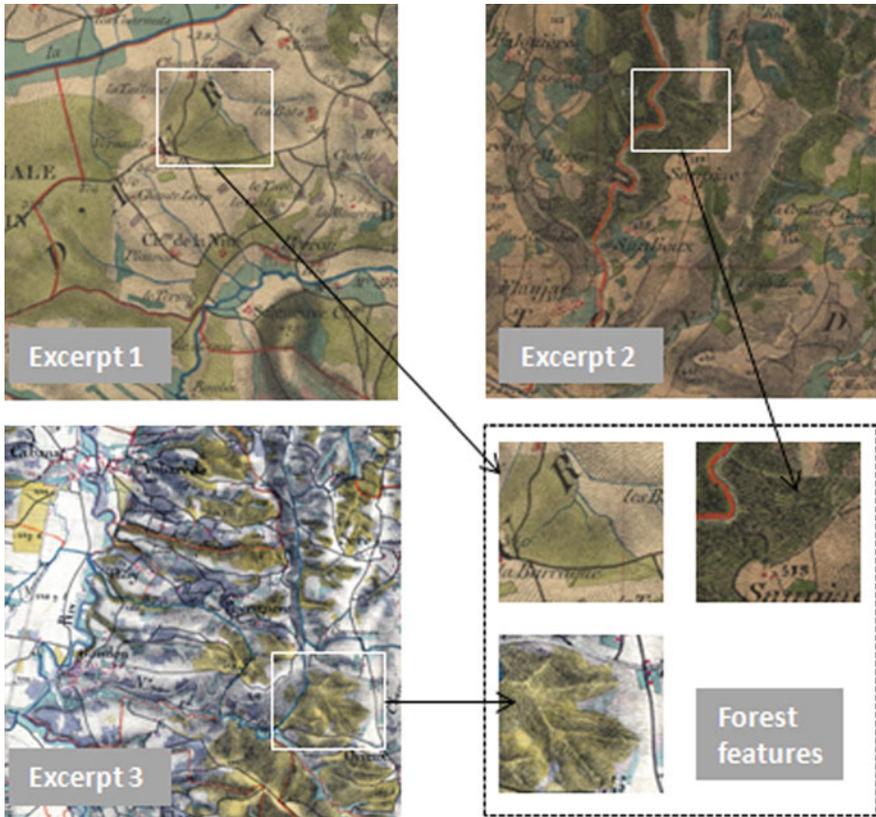


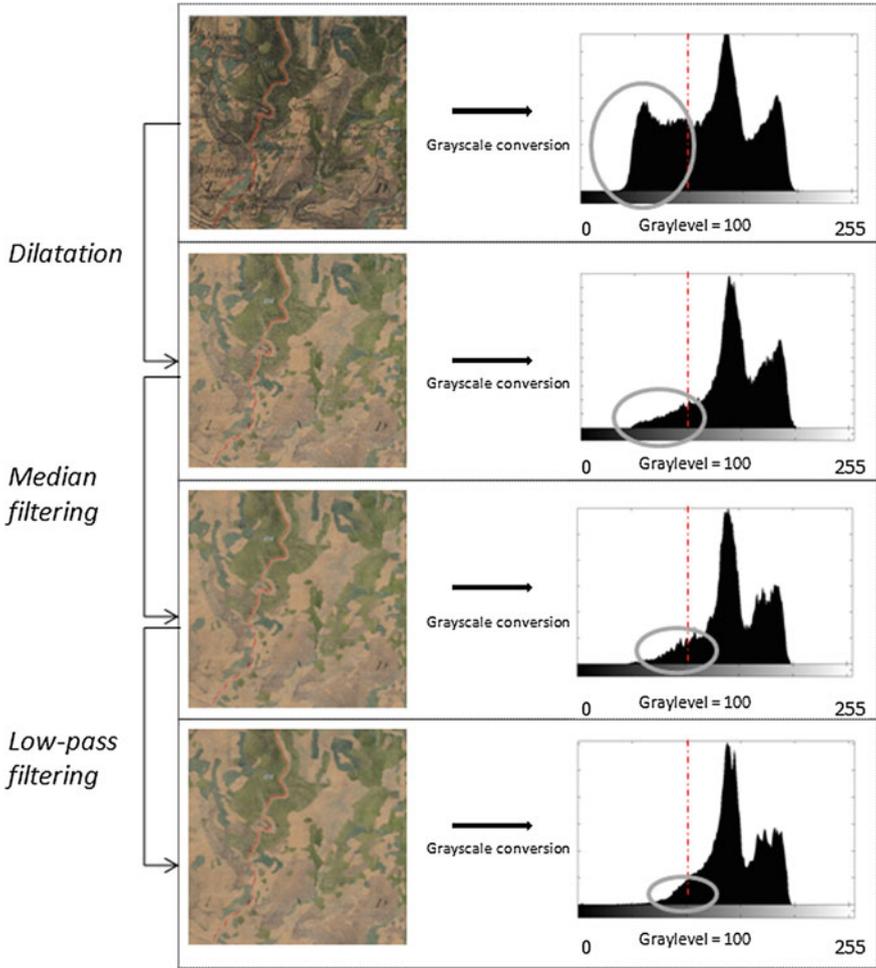
Fig. 6 Three excerpts of historical ‘map of France’ tested

the median filtering is able to remove dark pixels while transferring to higher frequencies. We can also see that a background noise remains: it is represented by small variations in the minimum values of the histogram (c).

A last step of filtering provides us the way to smooth these small variations: the low pass filtering. Thanks to this convolution filter, we go to reduce noisy pixels while correctly preserving dark pixels which correspond to road lines (Fig. 7d). The final result of the filtering procedure allows us to set up color-image classification. Colors and edges are still visible. This conducts to more uniform regions while facilitating pixel classification.

To show the performance of the  $L^*a^*b$  color space, we conducted a K-means classification on the excerpt 2 in RGB color space in order to compare it with  $L^*a^*b$  (Fig. 8).

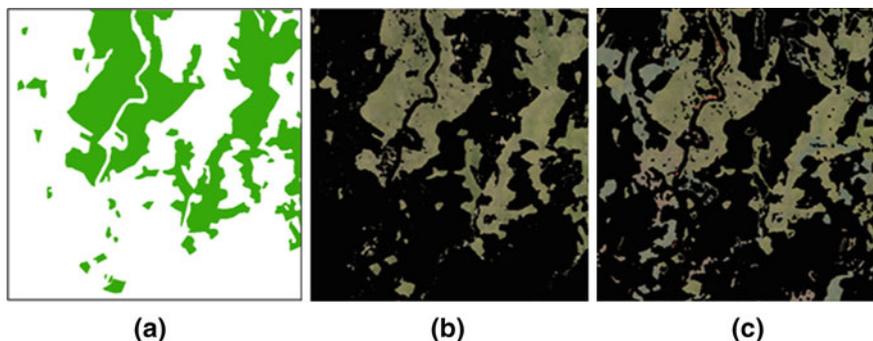
The quality of results differs. In the image (b), the extracted features do not represent an identical land use cover. Indeed, after the fusion of the different clusters which contained forest features, the final result is not satisfactory. Forest features are mixed with built areas (pale red) or grasslands (blue). Probably the



**Fig. 7** Successive steps of pre-processing of filtering

lack of uniformity of luminance in RGB color space does not allow separating the color variations of forest features from other land covers. By contrast, image (a) shows that color-image classification in  $L^*a^*b$  color space is the most efficient. Only forest features were extracted and any confusion with other land cover classes persists.

To finish, post processing step based on mathematical morphology and contextual rule was applied to fill holes within forest features and removing small isolated pixels. Figure 9a shows that extracted forest layer is not satisfactory because of too many holes inside. There are also small isolated pixels which do not represent forest elements and contribute to over detect the feature layer. After a binarization step, Fig. 9c means that morphological opening is efficient to remove



**Fig. 8** **a** Manual extraction layer. **b** K-means classification in  $L^*a^*b$  color space and **c** In RGB color space

small isolated pixels which are non-forest pixels. The contextual rule used proves to be also suited to fill holes within forest features. Indeed, we observe that holes in forest in Fig. 9b have been filled and allocated to ‘white’ class (Fig. 9c). To finish, forest layer of the original image is reconstructed by adding corrected binary image and original image in order to evaluate the performance of the global procedure.

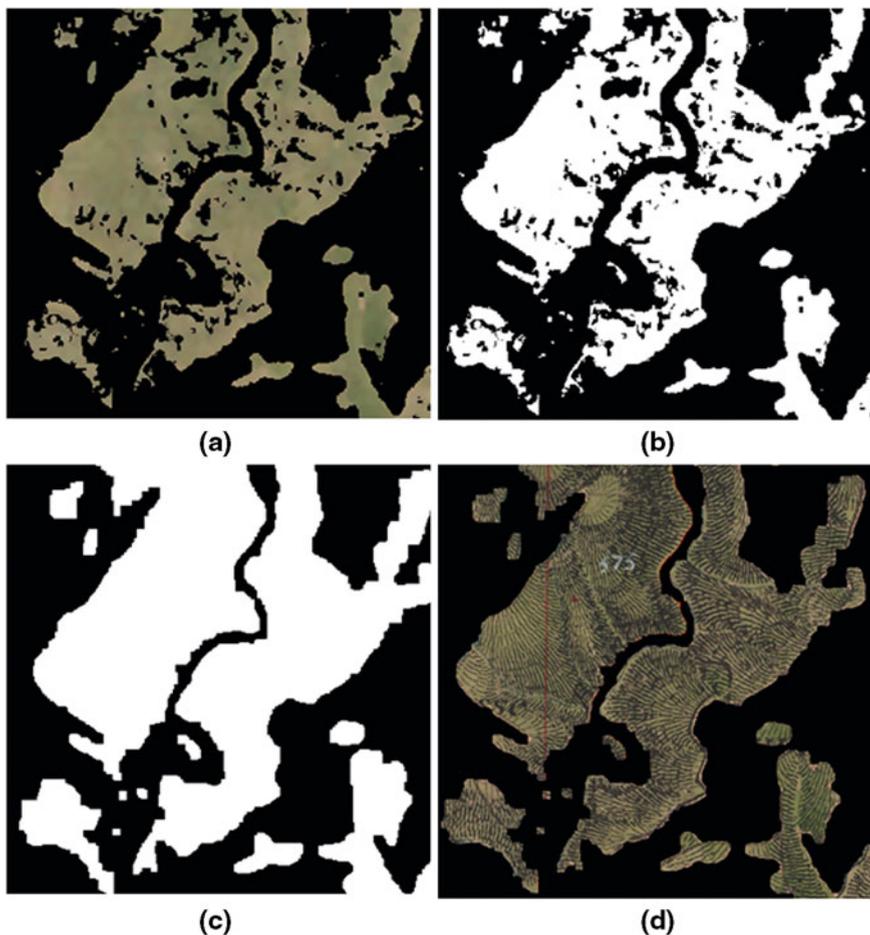
The general procedure has been applied on the three excerpts previously presented (Visual results in Fig. 10). The presented results show a high global accuracy ( $Kappa \approx 0.90$ ). The processes and the different parameters defined proved to be valid for each excerpt tested (Table 1).

Sensitivity Index and Specificity Index measure conditional probabilities that forest and non-forest are correctly classified. For each excerpt, these two indexes exceed 0.92 which conduct us to think the developed method is efficient to separate forest features and other elements on the image. PCC index is also very high for the three image tested ( $>0.90$ ). Kappa index is lower for the three examples but is globally satisfactory for the three examples.

Globally, the extraction procedure shows a real robustness because of homogeneous results on the three excerpts tested from the historical ‘Map of France’.

Successive steps of the filtering process proved to be efficient to remove contour lines elevation. Each filter used in a specific purpose contribute to eliminate noisy pixels (e.g., contour elevations) while preserving edges and contours. At the end of this procedure, the color is not disturbed and it is always possible to proceed to an unsupervised classification.

Concerning  $L^*a^*b$  color space, we can say that it appears well suited to low quality maps. Its structure offers the possibility to consider the green color variations as a unique color which increases uniformity of the forest regions in one scan and between them. This is a crucial advantage because of the heterogeneous quality of historical maps in general and the different color variations which can represent a same class. Accuracy results highlighted some minor problems.



**Fig. 9** **a** Extracted forest feature; **b** binary extracted feature; **c** binary corrected feature; **d** reconstructed forest feature

Firstly, there can be a trend to ‘under-detect’ forest during the unsupervised classification step. Indeed, the pre-processing step is efficient but some high slope areas are still shaded and are not recognized as forest features. This is even truer as the luminance of the forest feature is low: it is the case for the excerpt 2.

The post-processing step may also conduct to some confusion. The definition of the structuring element for morphological opening may change the morphology of several features or connect features to others. Thus, the historical structure of the forest matrix can be slightly modified in a local environment. We can do a similar observation during the step which fills holes within features. Indeed, some areas have been filled while they are not forest features but rather lakes or small

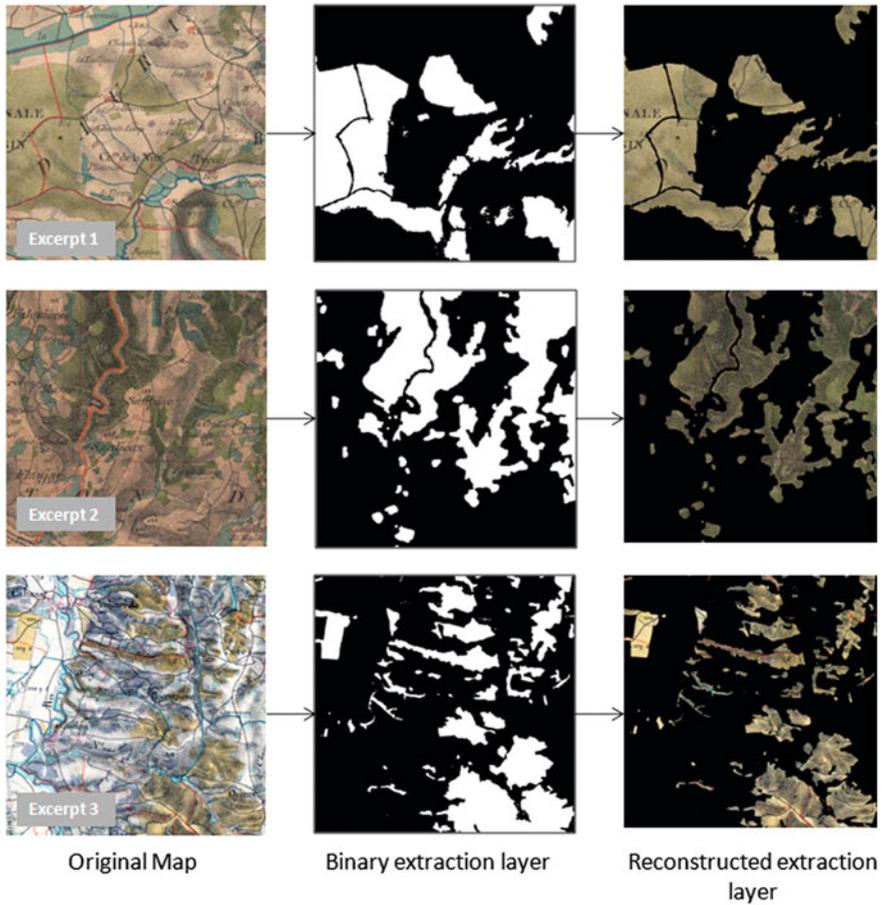


Fig. 10 Visual extraction results on the three excerpts tested

Table 1 Extraction results for three excerpts of  $1,500 \times 1,500$  pixels each

Excerpt tested	Excerpt 1	Excerpt 2	Excerpt 3
Accuracy index			
Sensitivity	0.92	0.92	0.94
Sensitivity	0.97	0.96	0.97
Pcc	0.96	0.95	0.96
Kappa	0.92	0.89	0.9

croplands. This can be explained by the definition of parameters which are not always suited to map specificities.

Last, the quality of the map is crucial in extraction procedure. Some specific areas within one scan are more or less preserved in terms of colors. So, several forest features and croplands (beige color) after pre-processing become confused while creating difficulties in extraction process. Pixels values for these two classes are very close for all the possible values of L. For these specific cases, a user intervention is required.

## 5 Conclusion

In this chapter, a method for extracting forest features from historical ‘Map of France’ was presented.

It is based on a pre-processing step followed by an unsupervised classification in a CIELab color space. Last, a post processing step was performed in order to improve extracted forest features.

Results on the three excerpts of the historical map are shown in Table 1 and present a high global accuracy. For three excerpts whose color of forest features, slope level and quality are very different, Kappa index is in average equal to 0.90. These results show a real robustness of the proposed scheme.

The main advantage of this approach is that it takes into account the color variations of forest features. The pre-processing of filtering allowed us to globally eliminate elevation contour lines which widely improve the distinction between features. It should be also noted that L\*a\*b color space proved to be efficient for increasing uniformity in a low quality map thanks to the taking into account of all possible value of Luminance in the image. Using this color space for other extraction tasks in historical maps could be an interesting alternative.

The accuracy assessment identified some minor problems. Pre-processing and post processing steps include the definition of some parameters which are not always suited to specificities of all sheets of the map. This difficulty may conduct to ‘over-detect’ or ‘under-detect’ forest features in these cases. Moreover, this may conduct to modify features shape or connecting between them which may modify global structure of the forest matrix in a local environment. Last, the poor color quality of the map widely varies from one sheet to another. Several classes can become confused while making color image classification difficult.

Future research will be dedicated to develop more advanced contextual rules for the post processing step. Taking into account spatial relationships between features could allow us to improve extraction results without any risks of modifying global structure of the features. This global procedure will be also tested on other raster color maps which suffer of identical quality problems. The main phases of the proposed scheme could be an interesting framework for these extraction problems.

**Acknowledgments** This research was supported by the French National Research Agency (ANR JCJC MODE-RESPYR 2010 1804 01–01). P.-A. Herrault is also funded through a PRES Toulouse University and Region Midi-Pyrenees grant.

## References

- Angulo J, Serra J (2003) Mathematical morphology in color spaces applied to the analysis of cartographic images. In: Levachkine S, Serra J, Egenhofer M (eds) Semantic processing of spatial data, in proceedings of the GEOPRO 2003-international workshop semantic processing of spatial data, Mexico City, pp 59–66
- Ansoult M, Soille P, Loodts J (1990) Mathematical morphology: a tool for automated GIS data acquisition from scanned thematic maps. *Photogrammetric Eng Remote Sens* 56(9): 1263–1271
- Arias JF, Lai CP, Chandran S, Kasturi R, Chhabra A (1993) Interpretation of telephone system manhole drawings. In: Second international conference on document analysis and recognition, Tsukuba Science City, pp 365–368
- Arrighi P, Soille P (1999) From scanned topographic maps to digital elevation models, in the international symposium on imaging applications in geology geovision '99. University of Liege, Belgium
- Barbu T, Ciobanu A, Mihaela C (2012) Automatic color-based image recognition technique using LAB features and a robust unsupervised clustering algorithm. In: Proceedings of the 13th WSEAS international conference on automation and information (ICAI '12), Iasi, Romania, 13–15 June 2012
- Bender O, Boehmer HJ, Jens D, Schumacher KP (2005) Using GIS to analyse 200 years of cultural landscape change in Southern Germany. *Landscape Urban Planning* 70:111–125
- Boesch R (1996) Detection and extraction of complex map symbols. *Int Arch Photogrammetry Remote Sens* 31 Part B3, Vienna
- Brunner CC, Maristany AG, Butler DA, Vanleuween D, Funck JW (1992) An evaluation of colorspace for detecting defects in Douglas-fir veneer. *Ind Metrol* 2(3 and 4):169–184
- Cao R, Tan C (Eds) (2002) *Text/graphics separation in maps*, vol 2390, Springer, Berlin
- Centeno JS (1998) Segmentation of thematic maps using color and spatial attributes. *Graphics Recognit Algorithms Syst* 1389:221–230. Lecture notes in computer science, Springer
- Chen L, Liao H, Wang J, Fan K (1999) Automatic data capture for geographic information systems. *IEEE Trans Syst* 5(2):205–215
- Chi Z, Yan H (1993) Map image segmentation based on thresholding and fuzzy rules. *Electron lett* 29(27):1841–1843
- Chiang YY, Leyk S, Knoblock CA (2012) Efficient and robust graphics recognition from historical maps. *Graphics Recognit Achievements Challenges Evol Selected Papers of the ninth international workshop on graphics recognition (GREC)*, Lecture notes in computer science
- Cousins S (2001) Analysis of land-cover transitions based on 17th and 18th century cadastral maps and aerial photographs. *Landscape Ecol* 16:41–54
- Gamba P, Mecocci A (1999) Perceptual grouping for symbol chain tracking in digitized topographic maps. *Pattern Recogn Lett* 20:355–365
- Ganesan P, Rajini V, Rajkumar IR (2010) Segmentation and edge detection of color images using CIELAB color space and edge detectors, in emerging trends in robotics and communication technologies (INTERACT), 2010 International conference
- Gimmi U, Lachat T, Burgi M (2011) Reconstructing the collapse of wetland networks in the Swiss lowlands 1850–2000. *Landscape Ecol* 26:1071–1083
- Hanbury A, Serra J (2003) Colour image analysis in 3-D polar coordinates, DAGM congress, p 8
- Henderson TC, Linton T (2009) Automatic segmentation of semantic classes in raster map images. In: 8th IAPR International workshop on graphics recognition, La Rochelle, France, July 2009, pp 253–262
- Iwanowski M, Kozak J (2012) Automatic detection of forest regions on scanned old maps. *Electr Rev* 88:249–252
- Jigar MS, Brijesh S, Satish SK (2012) A new K-mean color image segmentation with cosine distance for satellite images. *Int J Eng Adv Technol (IJEAT)*, 1(5), ISSN: 2249–8958

- Kaneko T (1992) Line structure extraction from line-drawing images. *Pattern Recogn* 25(9):963–973
- Leyk S, Boesch R, Weibel R (2006) Saliency and semantic processing: extracting forest cover from historical topographic maps. *Pattern Recogn* 39(5):953–968. doi:[10.1016/j.patcog.2005.10.018](https://doi.org/10.1016/j.patcog.2005.10.018)
- Lucchese L, Mitra SK (1999) Unsupervised segmentation of color images based on k-means clustering in the chromaticity plane content-based access of image and video libraries. (CBAIVL '99), In: *Proceedings IEEE workshop on digital object identifier*: doi:[10.1109/IVL.1999.781127](https://doi.org/10.1109/IVL.1999.781127), pp 74–78
- Mahmouda A, Elbialya S, Pradhana B, Buchroithner M (2011) Field-based landcover classification using TerraSAR-X texture analysis. *Adv Space Res* 48(5):799–805
- Mariani R, Lecourt F, Deseilligny M, Labiche J, Lecouturier Y (1997) *Interprétation de cartes géographiques: algorithmes de reconstruction de réseaux hydrographiques et routiers. Traitement du Signal* 14(3):317–334
- Motwani MC, Gadiya MC, Motwani RC, Harris FC Jr. (2004) Survey of image denoising techniques. In: *Proceedings of global signal processing*, Santa Clara, CA
- Myers G, Mulgaonkar P, Chen C, De Curtins J, Chen E (1996) Verification-based approach for automated text and feature extraction. In: Kasturi R, Tombre K (eds) *First IAPR workshop on graphics recognition*, *Lecture notes in computer science*, vol 1072, Springer, Berlin, pp 190–203
- Samet H, Soffer A (1998) Magellan: map acquisition of geographic labels by legend analysis. *Int J Doc Anal Recog* 1(2):89–101. doi:[10.1007/s10032005](https://doi.org/10.1007/s10032005)
- Samet R, Askerbeyli INA, Varol C (2010) An implementation of automatic contour line extraction from scanned digital topographic maps. *Appl Comput Math* 9(1):116
- Shaw T, Bajcsy P (2011) Automation of digital historical map analyses. In: *IS&T/SPIE Electronic Imaging*, 7869-09, Session 3, Conference 7869: computer vision and image analysis of art II, 23–27 Jan (oral presentation)
- Smith MJ, Cromley RG (2012) Measuring historical coastal change using GIS and the change polygon approach. *Trans GIS* 16(1):3–15
- Wise S (1999) Extracting raster GIS data from scanned thematic maps. *Trans GIS* 3(3):221–237
- Wyszecki G, Stiles WS (1982) *Color science*, 2nd edn. Wiley, London