Journal of Chromatography A, 1218 (2011) 2820–2825

Contents lists available at ScienceDirect

Journal of Chromatography A

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

The start-to-end chemometric image processing of 2D thin-layer videoscans

Łukasz Komsta[a], Łukasz Cieśla[b], Anna Bogucka-Kocka[c], Aleksandra Józefczyk[d], Jakub Kryszeń[e], Monika Waksmundzka-Hajnos[b,*]

[a] Department of Medicinal Chemistry, Medical University of Lublin, Jaczewskiego 4, 20-090 Lublin, Poland
[b] Department of Inorganic Chemistry, Medical University of Lublin, Chodźki, 6, 20-081 Lublin, Poland
[c] Department of Pharmaceutical Botany, Medical University of Lublin, Chodźki, 1, 20-093 Lublin, Poland
[d] Department of Pharmacognosy, Medical University of Lublin, Chodźki, 1, 20-093 Lublin, Poland
[e] Students' Scientific Association at the Department of Inorganic Chemistry, Medical University of Lublin, Chodźki, 6, 20-081 Lublin, Poland

ARTICLE INFO

Article history:
Available online 9 December 2010

Keywords:
2D-TLC
Image processing
Fingerprinting
Carex spp.

ABSTRACT

The purpose of the research was to recommend a unified procedure of image preprocessing of 2D thin layer videoscans for further supervised or unsupervised chemometric analysis. All work was done with open source software. The videoscans saved as JPG files underwent the following procedures: denoising using a median filter, baseline removal with the rollerball algorithm and nonlinear warping using spline functions. The application of the proposed procedure enabled filtration of random difference between images (background intensity changes and spatial differences of the spots location). After the preprocessing only spot intensities have an influence on the performed PCA or other techniques. The proposed technique was successfully applied to recognize the differences between three *Carex* species from the 2D videoscans of the extracts. The proposed solution may be of value for the any chemometric task – both unsupervised and supervised.

1. Introduction

Thin-layer chromatography is a well-established analytical technique and new inventions and improvements in this field still appear in literature [1–3]. The easiness of applying multidimensionality in this technique is one of its main advantages [4]. Using a medium-polar adsorbent (such as CN, DIOL or NH_2 plates) a chromatographer has an opportunity to develop a plate in normal-phase mode in one direction and reversed-phase mode in the other. This can significantly improve selectivity when compared to one-dimensional TLC, especially during analysis of very complex samples, such as plant extracts or herbal drugs [5].

One-dimensional chromatograms are often treated chemometrically as so-called "fingerprints" – unique signal-like vectors [6]. This approach can be used without identification of any peak, moreover no quantitative analysis is needed. The fingerprints, processed as vectors or matrices, can be used in different supervised or unsupervised approaches. This allows to explore similarity of the samples, to perform quantitative estimation of a complex parameter of the sample and also to classify samples to particular classes. The significant advantage of this approach is that there is no need to know what features (peaks, spots) are responsible for a modelled parameter, because the supervised chemometric algorithm (such as PLS) finds them automatically. However, the noise, baseline drift and time shifts of the peaks are often the main source of variance, totally obscuring the variance caused by differences among the samples themselves. Therefore, correct chemometric processing of such signals is a critical issue – they must be correctly denoised [7], warped [8–10] and also filtered from baseline drifts [6]. These issues are well studied and elaborated in the case of one-dimensional signals.

In the case of two-dimensional signals, like images of TLC plates, the same issues must be addressed during chemometrical preprocessing. However, the topic of 2D-TLC fingerprinting is not present in the literature. The very well starting point for our study was a series of papers related to preprocessing of two-dimensional gel electropherograms [11–17]. The studies performed in the field of 2D electrophoresis were a great challenge for the authors, due to the fact, that electropherograms contained a set of small spots and the main problem was to find corresponding features during warping. For this task, a special fuzzy matching strategy was developed. In this approach the baseline is filtered from these images by two-dimensional tensor products of the splines.

Preprocessing of thin layer chromatographic videoscans is a bit easier task than in the case of electropherograms, and there is no need to develop in-house complicated routines. Our current aim is to build a proposal of unified strategy for denoising, baseline removal and warping of 2D TLC signals for further supervised

* Corresponding author. Tel.: +48 81 532 4532; fax: +48 81 5320413.
E-mail address: monika.hajnos@umlub.pl (M. Waksmundzka-Hajnos).

0021-9673/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2010.12.004

or unsupervised chemometric processing. All these issues were addressed with ready-to-use free open source software. This study is a continuation of previous researches done in our faculty: denoising optimization of densitograms [18], univariate videoscans [19] and videoscan images [20]. The example of application of this technique is the recognition of plant species based on two-dimensional TLC digital photo.

Sedges (*Carex*) belong to one of the largest genus of vascular plants in the world and constitute a group of about 2000 species, widespread mainly in the northern hemisphere [21]. Flavonoids, stilbene derivatives, phenolic acids and fatty acids are the most common compounds identified in the genus [22]. There are a lot of questions related to the taxonomy of this genus caused by the difficulties in the species recognition due to morphological similarity of species, as well as the lack of habitat preference of the individual species. Additional problem is high intraspecific variability and the ability to create hybrids within particular sections [23]. Thus the taxonomic status of some taxa is still ambiguous and causes heterogeneity in their classification. The analysis of chemical compounds present in *Carex* species may bring useful information and contribute to their better distinction, along with other techniques used in studying the taxonomical relationships.

For comparative studies, especially in the case of closely related and morphologically similar plant species, it is important to get as much information from the sample as it is only possible. The analysis of complex samples of natural origin is difficult and not always successful with the application of traditional, one-dimensional separation techniques. If one-dimensional elution gives ambiguous results, then multidimensional techniques can provide more reliable information [24]. Multidimensionality can be easily realized by means of planar chromatography. It has been proved that the application of special modes of chromatogram development may be beneficial in the preliminary comparative studies of different plant species [25,26]. An overview of the use of two-dimensional thin-layer chromatography in phytochemical analysis has been also recently published [27].

2. Materials and methods

2.1. Apparatus and reagents

Methanol and water (HPLC grade), used as the extracting solvents, were obtained from J.T. Baker (WITKO, Poland). Methanol, ethyl acetate and *n*-heptane used as the mobile phase components, were purchased from Merck (Darmstadt, Germany). Acetic acid was obtained from Polish Reagents POCH (Gliwice, Poland). All solvents were of the analytical purity grade. TLC was performed on the 10 cm × 10 cm glass-backed HPTLC 60 CN_{254S} plates, purchased from Merck.

Solutions of plant extracts were applied to the chromatographic plates spot-wise by means of a Camag automatic TLC sampler (Camag, Muttenz, Switzerland) and developed in the horizontal DS chambers (Chromdes, Lublin, Poland). Location of the spots was carried out in UV light, 366 nm. Chromatograms were documented with use of the Camag TLC Reprostar 3 device with Videostore.

2.2. Plant material and extraction procedure

The study was performed for three exemplary species of sedges which are comparatively widespread and relatively common for the Central European Lowland, namely: *Carex gracilis*, *Carex praecox* and *Carex rostrata*. The aboveground parts of the plants were collected from natural sites, located in habitats typical for each species. The plant material was collected in the second half of June 2006, that means in a period, when sedges have simultaneously formed flowers and young fruits, and their vegetative organs are dynamically growing. Thus, the collected plant material represented organs in all possible developmental phases, what allowed to eliminate eventual differences in biochemical features among organs being in particular phases of development. Representatives of particular studied species were collected from 3 localities each.

The plant material was collected and identified by Dr. Magdalena Janyszek from Department of Botany, Poznan University of Life Sciences. Specimens of the species were deposited in the Department of Botany, Poznan University of Life Sciences (POZNAN).

Authenticated plant material (herbs) was dried at room temperature, and then powdered in an electrical mill (IKA A11). Each of the studied species was represented by three samples, coming from different collection sites, and analyzed separately. The powdered material (100 g) was extracted three times (3× 1.5 h) with boiling 80 °C methanol (800 mL). The methanol extracts were evaporated in the vacuum. After evaporation the residues were eluted with hot water (150 mL) and left at room temperature for 24 h. Then the aqueous extracts were filtrated. Phenolic compounds were isolated in a typical way for this group of compounds [28]. After the isolation step, the obtained fractions underwent acid hydrolysis in order to obtain purely aglycon fractions, according to method described elsewhere [29]. The all obtained fractions were individually dissolved, in 5 mL methanol and filtered through 0.45 μm membrane filters SUPELCO (Iso-Disc Filters, USA) and consequently analyzed by 2D-TLC. These stocks solutions were stored under refrigeration (4 °C) before use.

2.3. Two-dimensional thin-layer chromatography

Sample solutions were applied to the 10 cm × 10 cm silica TLC plates spot-wise (1 cm from both plate's edges). 10 μl aliquots of the investigated *Carex* samples were applied to the plates. The plates were dried in a hood for 10 min before development. Developments were performed in the horizontal DS chambers. The plates were conditioned, for 10 min, in the chambers saturated with mobile phase vapors, before each development. First the samples were chromatographed with use of mobile phase I: ethyl acetate–*n*-heptane–acetic acid (50 + 50 + 1, v/v/v). In the perpendicular direction the following mobile phase was applied: methanol–water–acetic acid (50 + 50 + 1, v/v/v). After the first development all the plates were dried, in the stream of cold air, for 30 min. The plates were developed to the distance of 90 mm, in both directions.

3. Results and discussion

The developed chromatographic plates were photographed with the aid of CCD camera and the obtained photos were saved as JPG files with maximum available quality (resolution 708 × 708 pixels) (Fig. 1A). As all the spots of the compounds were blue, there was no need to process three RGB channels and images were converted to grayscale before any further processing. All image processing steps were done using ImageJ free open source software [30]. PCA analysis of processed images was performed inside free open source GNU R environment [31].

The first step of image processing was denoising of the images. It was done using built-in median filter (width 10 pixels) according to our previous study [20]. Next, the background baseline was removed using built-in rollerball algorithm with rollerball radius equal to 100 pixels. The rollerball algorithm simulates the rolling of the sphere of a given diameter on the three-dimensional surface of the image. The diameter of the ball is larger than the spot width, therefore the ball cannot fall into the cavity representing

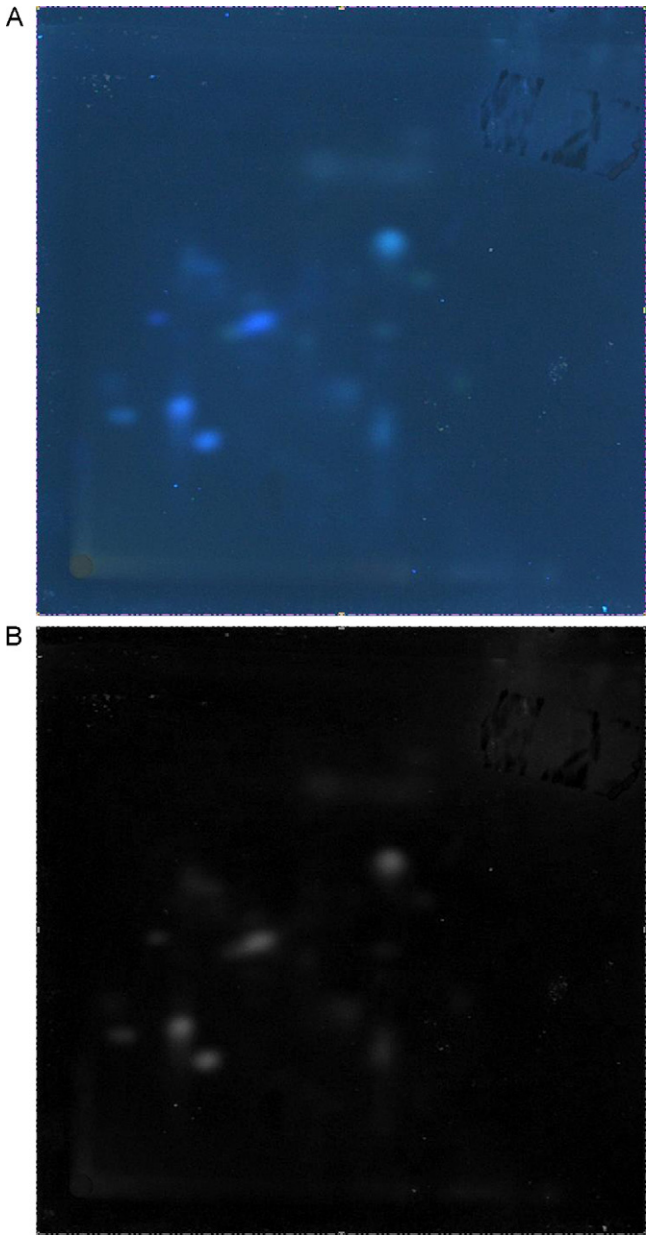

Fig. 1. An example of *Carex gracilis* 2D TLC videoscan: raw image (A) and image after grayscale conversion, denoising and background removal (B).

the spot. On the contrary, the background drift is preserved. The position of the ball center, created after its rolling, is treated as the background estimate. As TLC images are less complex than electrophoretic ones, the rollerball algorithm performs here very well and there is no need to use more complicated techniques such as multidimensional spline tensors [14]. In general, the algorithm is not very sensitive to ball radius parameter, it can be almost any value larger than the diameter of largest spots. The example result of such preprocessing is shown in Fig. 1.

The principal component analysis was made on vectors formed from all corresponding image pixels (each image was expanded from matrix to a long vector). The PCA of unprocessed images (Fig. 2A) did not show any grouping of the samples of the same species. Although two first principal components explain 82% of total variance, the analysis of the loading vectors (Fig. 2B and C) confirms that the difference of the background intensity between images and the spatial differences of the spot location are the main source of overall variance. Eliminating background by rollerball

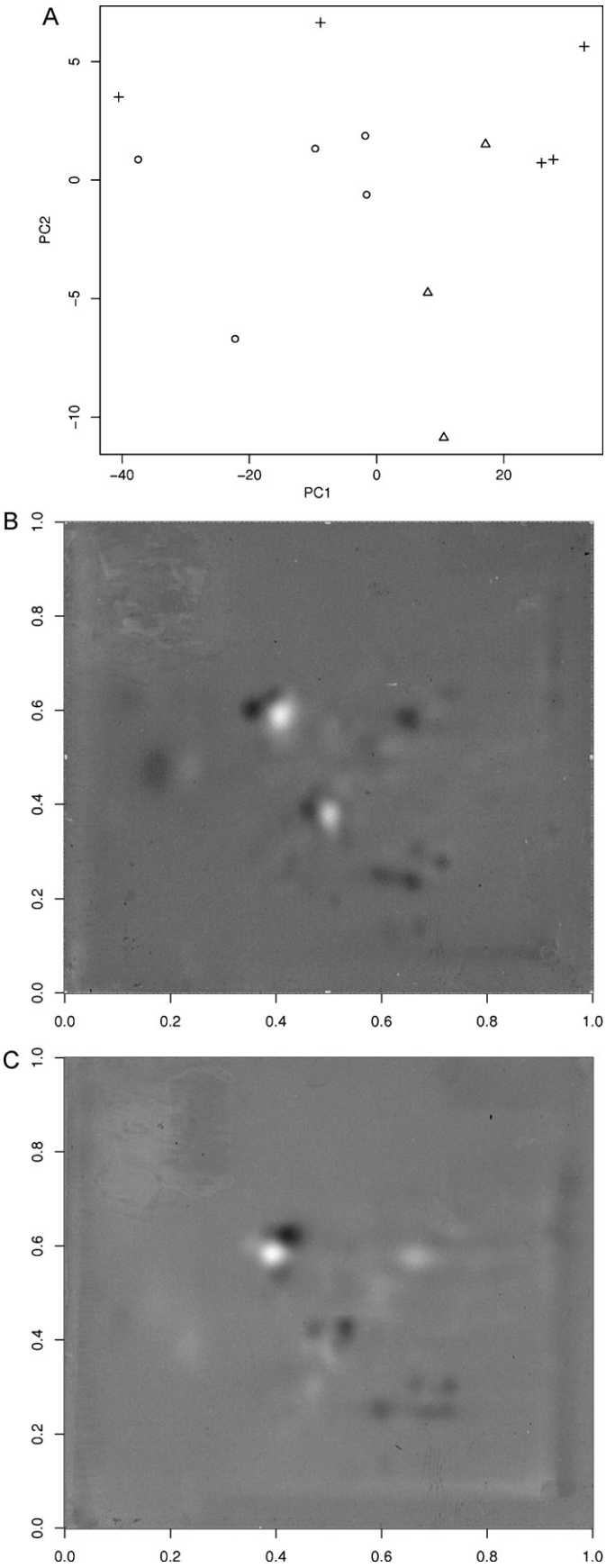

Fig. 2. Principal component analysis done on unprocessed images (A), with corresponding first (B) and second (C) loadings (*C. gracilis* is plotted as circles, *C. praecox* as pluses, *C. rostrata* as triangles).

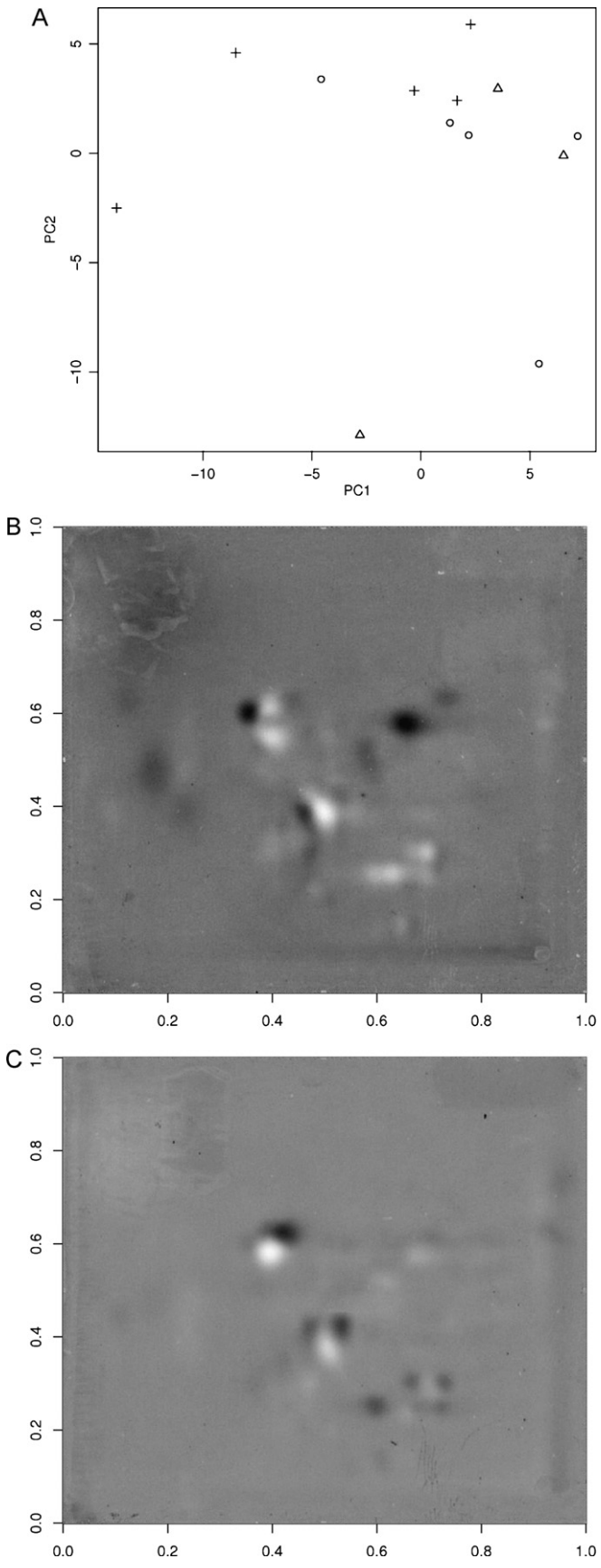

Fig. 3. Principal component analysis done on images after denoising and background removal (A), with corresponding first (B) and second (C) loadings (*C. gracilis* is plotted as circles, *C. praecox* as pluses, *C. rostrata* as triangles).

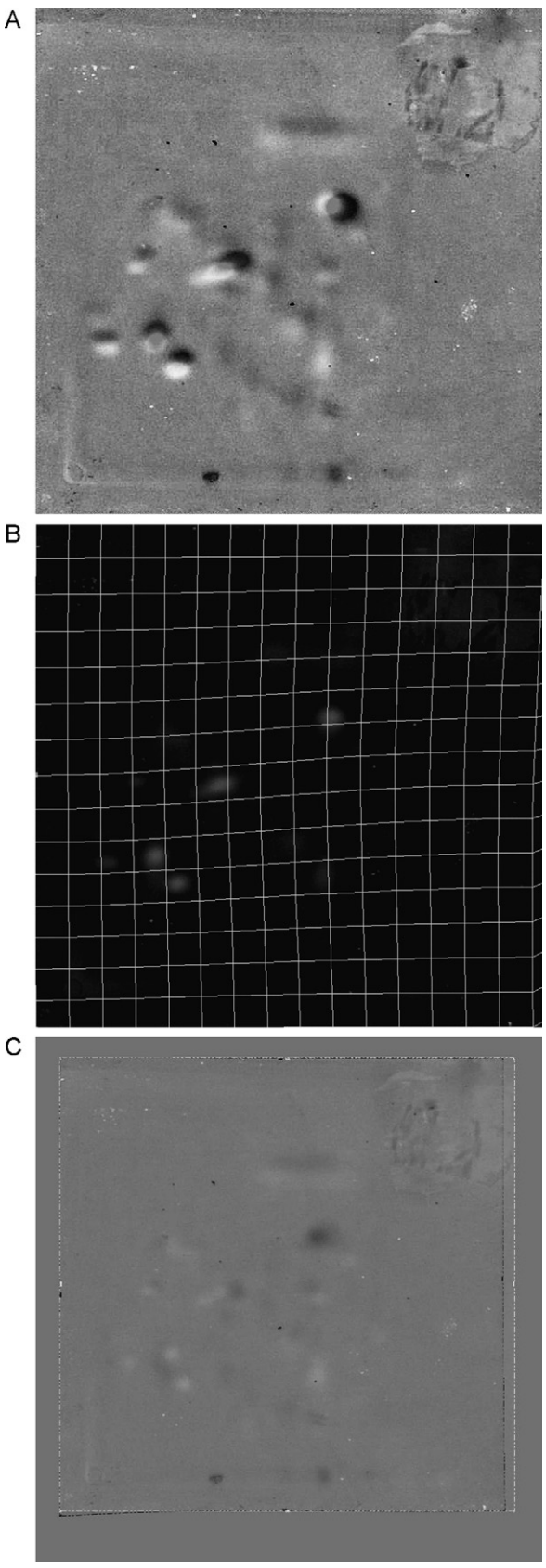

Fig. 4. An example of two *Carex gracilis* videoscans registration: differences between two plates before registration (A), the registration grid (B) and differences after registration (C).

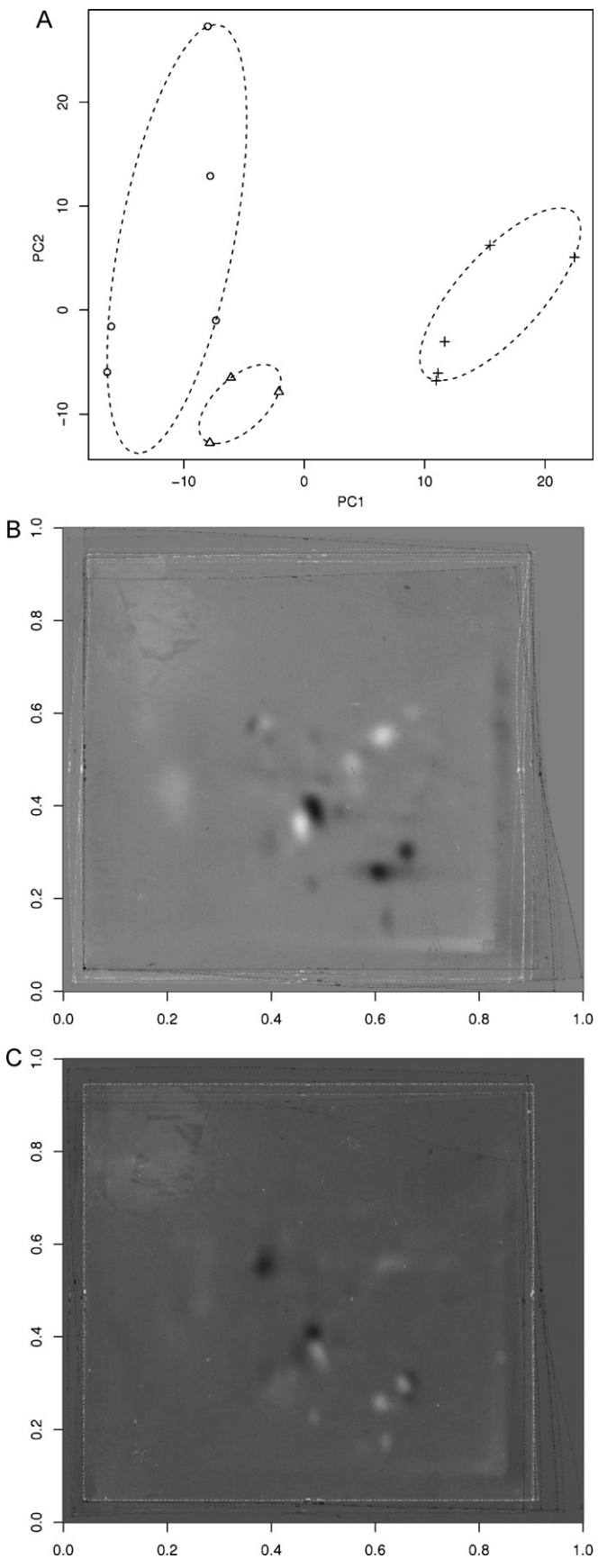

Fig. 5. Principal component analysis done on fully processed images (A), with corresponding first (B) and second (C) loadings (*C. gracilis* is plotted as circles, *C. praecox* as pluses, *C. rostrata* as triangles).

method is not enough, as the spatial differences are still the main source of variance. However PCA done on the images with eliminated background (Fig. 3A) still does not show any genus-based clustering. The loadings (Fig. 3B and C) extract more spots and features and they are not affected by baseline differences, two first PCs explain only 31% of total information.

Therefore, there is an absolute need to perform warping (called also image registering) of the images to remove spatial changes of the spots in the dataset. The warping was done using bUnwarpJ plugin, automated with "register virtual stack slices" script. This software warps a set of images using a nonlinear transforms approximated by a sets of special piecewise polynomial functions called splines and detailed description is given in a reference paper made by its authors [32]. The random spatial changes of spots' position in TLC plates are nonlinear in their nature and simple linear registration algorithms (based on rotation, translation and scaling only) are not sufficient, so the nonlinear approaches are needed to give more flexibility of possible image transformations. This algorithm is very simple and efficient in its standard parameters, the only tuning made was an increase of the parameter called "feature" (detected spot size) to 30 pixels to avoid aggressive deformation and warping of non-corresponding small spots. According to recommendations in [6], the reference image was selected to have the highest mean correlation coefficient with all other ones. The example of warping is shown in Fig. 4.

The PCA done on warped images is shown in Fig. 5. Almost all spatial differences were filtered during the image warping and first principal components depicts mainly differences between species in particular spot intensities. There is a visible clustering trend in analyzed samples (Fig. 5A) and analysis of corresponding loadings (Fig. 5B and C) can depict and detect differences in the TLC spots, responsible for species' differences.

4. Conclusions

The presented methodology of preprocessing images can be recommended in any chemometric task – not only unsupervised (as PCA here), but also supervised, both in classification and calibration context. The random differences between images, caused by background differences and spatial (spots location) differences are filtered and only spots intensities information is preserved. The only one prerequisite is that all analyzed samples should have several main common ingredients, but this requirement is also valid in warping of 1D signals. The use of publicly available free open source software is an additional advantage, because no additional software costs are needed. The specialized CCD camera can be in many cases replaced by photo camera with good CCD matrix, located on tripod.

References

[1] C.F. Poole, J. Chromatogr. A 856 (1999) 399.
[2] C.F. Poole, N.C. Dias, J. Chromatogr. A 892 (2000) 123.
[3] C.F. Poole, J. Chromatogr. A 1000 (2003) 963.
[4] C.F. Poole, J. Chromatogr. A 703 (1995) 573.
[5] M. Waksmundzka-Hajnos, J. Sherma, T. Kowalska, Thin Layer Chromatography in Phytochemistry, CRC Press, Francis and Taylor, Boca Raton, USA, 2008.
[6] M. Daszykowski, B. Walczak, Trends Anal. Chem. 25 (2006) 1081.
[7] L. Pasti, B. Walczak, D.L. Massart, P. Reschiglian, Chemom. Intell. Lab. Syst. 48 (1999) 21.
[8] B. Walczak, W. Wu, Chemom. Intell. Lab. Syst. 77 (2005) 173.
[9] V. Pravdova, B. Walczak, D.L. Massart, Anal. Chim. Acta 456 (2002) 77.
[10] A.M. van Nederkassel, M. Daszykowski, P.H.C. Eilers, Y. Vander Heyden, J. Chromatogr. A 1118 (2006) 199.
[11] M. Daszykowski, A. Bierczynska-Krzysik, J. Silberring, B. Walczak, Chemom. Intell. Lab. Syst. 104 (2010) 2.
[12] M. Daszykowski, E. Mosleth Faegestad, H. Grove, H. Martens, B. Walczak, Chemom. Intell. Lab. Syst. 96 (2009) 188.
[13] M. Daszykowski, I. Stanimirova, A. Bodzon-Kułakowska, J. Silberring, G. Lubec, B. Walczak, J. Chromatogr. A 1158 (2007) 306.

[14] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Acta Chromatogr. 15 (2005) 82.
[15] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Proteomics 4 (2004) 2377.
[16] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, J. Chem. Inf. Comput. Sci. 43 (2003) 978.
[17] K. Kaczmarek, B. Walczak, S. de Jong, B.G.M. Vandeginste, Acta Chromatogr. 13 (2003) 7.
[18] Ł. Komsta, Anal. Chim. Acta 641 (2009) 52.
[19] Ł. Komsta, J. Chromatogr. A 1216 (2009) 2548.
[20] Ł. Komsta, Acta Chromatogr. 21 (2009) 355.
[21] M. Hendricks, F. Oberwinkler, D. Begerow, R. Bauer, Plant Syst. Evol. 246 (2004) 89.
[22] A. Fiorentiono, B. D'Abrosca, S. Pacifico, A. Izzo, M. Letizia, A. Esposito, P. Monaco, Biochem. Syst. Ecol. 36 (2008) 691.
[23] M. Janyszek, A.M. Jagodziński, Plant Syst. Evol. 278 (2009) 87.
[24] M.A. Hawrył, A. Hawrył, E. Soczewiński, J. Planar Chromatogr. 15 (2002) 4.
[25] Ł. Cieśla, A. Bogucka-Kocka, M. Hajnos, A. Petruczynik, M. Waksmundzka-Hajnos, J. Chromatogr. A 1207 (2008) 160.
[26] Ł. Cieśla, K. Skalicka-Woźniak, M. Hajnos, M. Hawrył, M. Waksmundzka-Hajnos, Acta Chromatogr. 21 (2009) 641.
[27] Ł. Cieśla, M. Waksmundzka-Hajnos, J. Chromatogr. A 1216 (2009) 1035.
[28] L. Świątek, J. Góra, Herba Polonica 24 (1987) 125.
[29] H. Schmidtlain, K. Herman, J. Chromatogr. 115 (1975) 123.
[30] http://rsbweb.nih.gov/ij/.
[31] R Development Core Team, R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0, http://www.R-project.org.
[32] I. Arganda-Carreras, C.O.S. Sorzano, R. Marabini, J.M. Carazo, C.O. de Solorzano, J. Kybic, Lect. Notes Comput. Sci. 4241 (2006).