PANDORA
Recognizing pigments

LAURA FLOREA
University Politehnica of Bucharest, Image Processing and Analysis Laboratory
December 13, 2016

Contents

1 Identification of Chemical Compounds

1.1 Database: content, classes

Database was provided by courtesy of *National History Museum of Romania* and consists of samples of Kremer pigments on arabic gum, and on white/black background photographed with microscopic camera in both visible (sensors sensitive to red, green, blue) and infrared (IR) domains. The Kremer pigments used are described in the dedicated catalog [1]. The database is *property of the National History Museum of Romania and is it not a public database*.

In a first step we analyzed the samples available and we removed those not corresponding to Kremer pigments or those improperly labelled. It has been found that some had infrared images had artifacts and we had to remove them too. Wherever possible, samples that have been removed were replaced with other samples of the same pigment. After discussions with our collaborators at the National History Museum of Romania we decided to divide the database into two distinct parts:

1. pigments samples applied on gum arabic - 70 samples
2. samples containing pigments applied on paper - 157 samples.

Each of the respective samples having 342×683 pixels contains two images, one corresponding to data from the visible spectrum and one for the infrared spectrum. Samples containing pigments applied on paper have an area with white background and an area with black background (areas

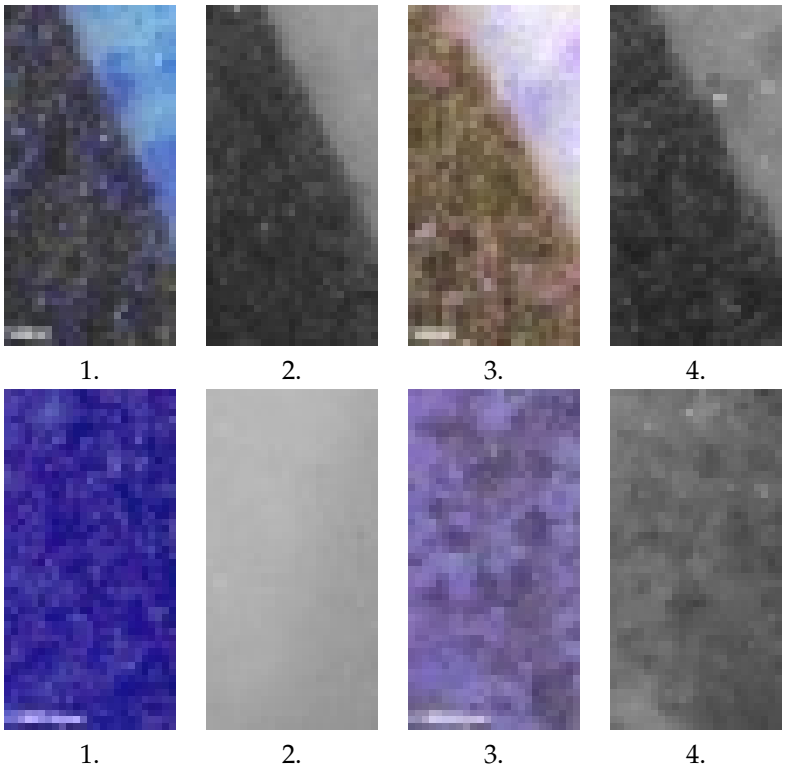

Figure 1: Two samples of pigments applied on paper (above) or the gum arabic (below). Each sample contains the image in the visible spectrum (1 and 3) and infrared (2 and 4 respectively).

Table 1: Dividing the database as implemented in the two phases of the project

Phase	**Phase I - 2015**	**Phase II - 2016**	
Pigment support	Arabic gum or paper	Arabic Gum	paper
No. of samples	270	70	157
No. of classes	15	70	157

that are not always positioned identically). Examples of the two parts of the database are shown in figure 1.

Unlike the first phase of the project, when in the experiments, we have considered in the same class several types of pigments, in this stage experiments (described below), we considered each sample as a separate class. If the first stage there were only 15 relatively homogenous classes, while now it resulted in two databases of 70 and 157 classes respectively (see table 1).

Each sample was divided into several distinct instances to form the final database that has been used for further experiments. Taking into account that are inhomogeneous samples, we have considered non-overlapping instances from each sample. Sample size is chosen so to ensure enough information to preserve the color and the texture of the pigment, but small enough to have as many samples in the database as possible. There have been tests of samples having various dimensions: 64×64 pixels, 96×96 pixel, 128×128 pixels, respectively 340×340 pixels.

We have found that choosing the size smaller than 64×64 the resulting patches no longer

retain enough information to recognize the pigments texture. On the other hand, if we use larger samples, then there is a drastic decrease in the number of samples in the training and testing sets and the results may be inconclusive. Thus the resolution of 64×64 pixels for sampling, will be kept for all experiments following in the paragraphs below. Following this division resulted in a total of 7065 samples per class. So finally we have two databases as follows:

1. 70 samples that contain pigments applied on arabic gum
   - 70 classes * 7065 = 494 550 samples
2. 157 samples containing pigments applied on paper
   - 157 samples class * 7065 = 1109205 samples.

1.2 Algorithm: Implementation and results

In pigment recognition solution built there were implemented and tested different methods based on a classical system of machine learning. Thus, samples extracted from various pigments were considered as input data for features that provide color, texture and combined descriptions. These features were then provided to a machine learning system, namely Support Vector Machine (SVM) [3].

1.3 Features

For each sample were extracted several type of features. These features can be divided into color, texture and mixed characteristics, as follows:

1. **HOG (Histogram of Oriented Gradients)** [4] - considered only the information intensity of the sample in the visible spectrum, rezulting into a texture descriptor.
2. **ColorHOG** - HOG features were calculated on each of the three RGB color planes of the sample in the visible spectrum, to which was added a fourth plane from infrared space (image with gray levels). There was thus introduced and a part which depends on color.
3. **pHOG** - it was considered the intensity information from the visible spectrum, taken on multiple levels of a Gaussian pyramid; thus, it is a texture descriptor.
4. **LBP (Local Binary Pattern)** - we considered the information from the visible, equivalent gray-scale image.
5. **pLBP (pyramid Local Binary Pattern)** - this descriptor too is computed on a pyramid of gray-level images in the visible spectrum.
6. **HoT (Histogram of Topographic Features)** [5]. This is a histogram of the texture descriptor based on the first and second derivative gathered in 6 histograms concatenated in the feature vector. This descriptor is used only for information intensity of the sample in the visible spectrum and is a pure texture descriptor.
7. **colorHoT** We have taken the HoT features computed independetly on each color plane. A forth plane was added with data taken from the infrared domain.
8. **HistLABI** Here, first the RGB color space was turned into Lab. The intensity histogram was taken on each color plane and the forth, infrared histogram was added. All histogram were computed on number of bins empirically founded.

Table 2: Pigment recognition: results

Feature	Paper pigments			Arabic gum pigments			No. elements
	SVM		Accuracy	SVM parameters		Accuracy	
	cost	gamma	[%]	cost	gamma	[%]	
HOG	1	1	7.4	1	-1	15.32	32
pHOG	15	-13	6.58	1	-3	15.56	128
LBP	3	-1	17.46	1	-1	31.67	58
pLBP	1	-3	16.26	15	-3	35.04	232
HoT	3	-1	56.72	9	-7	93.81	60
colorHOG	3	-1	23.04	3	-3	62.94	128
ColorHoT	9	-7	59.24	15	-3	96.47	180
HistLABI	5	-3	79.21	13	-9	98.25	220
HistLABI + HoT	7	-5	81.76	3	-3	98.41	280

1.4 Classification

For the classification task (i.e. pigment recognition) we have used an SVM as implemented in LibSVM [2]. Each query was preceded by an exhaustive search in the cost-gamma space for the optimal classification. Training and testing of the SVM is into a k-fold system : the database is split into $k = 4$ folds and each on of them is taken separately as the testing part. Reported results is the average of individual results fora each fold. These may be follow, together with found cost, C and gamma, γ in table 2.

If one analyzes the results, he will find that color features (e.g. HistLABI) provide better results than plain texture features. Yet, taking into account that query are related to color pigments.

However, there exists texture features, such as colorHoT, which have a positive contribution to overall performance. As one cans see in table 2 the combination HoT with HistLABI gives the best results, reaching 81.76% accuracy on the 157 sample database and respectively to 98.41% on the 70 sample database.

References

1. Kremer - pigmente. product catalog. Catalog de pigmenti disponibil on-line la adresa $http : kremerpigments.comdownloadkrp_katalog_US_140806_web.pdf$.
2. C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3), May 2011.
3. C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273297, 1995.
4. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of CVPR*, pages 886–893, 2005.
5. C. Florea, L. Florea, and C. Vertan. *Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III*, chapter Learning Pain from Emotion: Transferred HoT Data Representation for Pain Intensity Estimation, pages 778–790. 2015.