

# Dating Color Images with Ordinal Classification

Paul Martin, Antoine Doucet and Frédéric Jurie  
GREYC, CNRS UMR 6072  
University of Normandy, Unicaen, France  
firstname.lastname@unicaen.fr

## ABSTRACT

This paper proposes a new approach for automatically dating a photograph, based solely on its content. Building on recent advances in computer vision, the images are first described by a set of features. Then, the age group of every image is predicted by a classifier trained with annotated data. The key strength of our approach – which makes it perform better than existing ones – is the introduction of an ordinal classification framework, particularly adapted to the type of data to be predicted (age groups). The approach is validated on a recent challenging dataset for which it produces state-of-the-art results.

### Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *selection process*

### General Terms

Algorithms, Experimentation, Measurement, Theory

### Keywords

Image Retrieval, Ordinal Classification, Temporal Information, Dating Photography, Image Time-Stamping

## 1. INTRODUCTION

The amount of images available online is seeing a rapid and sustained increase. An important part of them stems from the digitization process of archives, led on a large scale by institutions such as museums and national libraries. The recent surge in the use of social networks has also triggered an enormous amount of user-contributed images. Many of those are digitally-born pictures, but a significant share is digitized, for instance with the help of scanners.

Naturally, as this sum of digitized photographs is provided by numerous sources, they come with inconsistent meta-data, in which the date of the photograph is often absent, and sometimes even unknown to the person in charge of the digitization process. Even in digitally-born pictures, the provided date is often wrong, as this bit of information relies on the assumption of a proper device configuration – an illustration of this is the existence of numerous digital pictures from the beginning of the year 1970, simply due to the fact that 1 January 1970 is the default initial date on

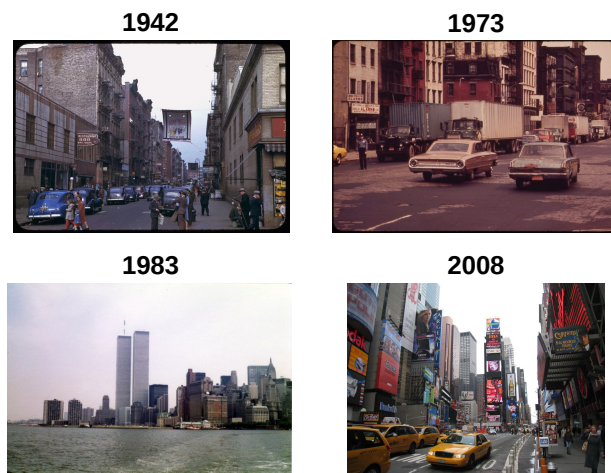


Figure 1: Four different pictures of New York City. The overall objective of this work is to automatically predict the date when such images were taken.

many devices.

In this paper, we address the problem of the automatic forecasting of the date when a digitized photograph was originally taken. The task of automatically dating images has recently been simultaneously introduced in the communities of information retrieval [8] and computer vision [12]. This topic has also aroused interest in the photography industry, as in 2010 Kodak recorded a patent [14] describing a process to determine the date range when a picture was developed, based on watermarks visible on its front and reverse side.

Formally, the problem of automatically dating an image simply consists in associating it with the date when it was taken. While this task can be seen as a classification problem, where the classes are various date ranges, it is important to also take into account the fact that there exists an order relation connecting all the pictures over a timeline. Hence the automatic dating of an image shall not only be assessed in a binary fashion, as correct or incorrect, but should be evaluated by a measure of the distance separating the forecast and the truth, its optimal score being zero. It is indeed preferable to obtain estimations temporally close to real shooting dates.

Getting accurate temporal predictions is important for several applications. The enrichment of existing images with a time-stamp shall for instance allow information retrieval to satisfy users expressing topical and temporal information

needs, such as “*I want pictures of NYC in the 1920s*”.

Enriching existing images with their temporal dimension will have multiple implications on the daily use of search engines. This will allow disambiguation between similar terms with multiple related temporal epochs, or image reorganization, exhibiting the visual evolution of a monument or a celebrity. For instance, the information need “*I want pictures of NYC*” could then be answered visually as shown by Figure 1.

As the technique presented in this paper only relies on the visual content of images, it requires no textual description and its usability therefore spans beyond the Web. It can notably be used to organize collections of images stored in archives (e.g., of national libraries or museums), as an assistant or as a full replacement for manual annotation by domain experts.

In this paper, we present a technique that accounts for the ordinal nature of this image classification problem, in a way that allows significant improvement over the current state of the art. After a review of the related work on dating images and ordinal classification (Section 2), we extensively describe our method and its specificities in Section 3. Our experimental framework and subsequent results are detailed in Section 4, before we draw conclusions and discuss future directions in Section 5.

## 2. RELATED WORK

Recent advances in image classification and image retrieval have provoked a renewed interest for automatic dating technologies, which can provide useful information regarding images. Early methods were relying on manual image analysis of the physical properties of the media (e.g. paper type, size, coated, color scheme) [6]. In contrast, if one is interested in dating an image from its content (as opposed to its support), many references offer tools to help dating based on visual characteristics of objects, places or people in the pictures. To cite just two examples, the Web site of the Victoria and Albert Museum [1] explains how to date an image from clothes, over the period 1840-1930, while the Web site of the English Museums Association [2] offers a complete methodology, including people’s clothes, poses taken by persons, image composition, scenes, to name but a few. However, the methods mentioned above require human intervention by experts of the relevant period, and is therefore not applicable on a large scale.

However, few automatic methods exist to our knowledge. One possible approach is to use time-tags, but this piece of information is only available for a small set of recent digitally-born images and is dependent on the device. The second approach could be the definition of a learning model with specific visual features capable of classifying the temporal dimensions of images with high accuracy. Within this context, the recent Kodak 2010 patent [14] focuses on dating photos based on distinguishing marks that may be on the back of the photo (e.g. brand and type of paper used, written dates) or on the over-print of the image (e.g. date, logo). Detected watermarks would then be matched with watermarks whose time-span is known, hence allowing to retrieve a range of possible dates when the picture had been developed. Unfortunately this method is very specific and requires scans of both sides of a picture, and it is obviously only in rare cases that the reverse side of a photograph is digitized and made available together with its front side. As

a consequence, the scope of this methodology is limited.

More recently, Palermo *et al.* [12] have proposed a new approach allowing to predict the age group of images on the basis of their content. It captures temporally discriminative information based on the evolution of color imaging processes over time. In addition to three popular image features (Gist Descriptor, Tiny Images and  $L^*a^*b^*$  Color Histogram), they built and used four new domain specific image features: Process Similarity Feature, Color Co-occurrence Histogram, Conditional Probability of Saturation Given Hue and Hue Histogram. Their method gives significantly greater accuracy than that of untrained humans on the same data set.

We build on Palermo *et al.* [12] by overcoming one intrinsic limitation of their approach. Indeed, they consider image dating as a multi-class task, ignoring the relative order of the temporal information. In contrast, following the recent works on ordinal classification [3, 4, 5, 7, 10, 11] we do consider dates as relative attributes and propose a specific framework for handling them.

## 3. OUR METHOD

This section presents our proposed method for ordinal image dating, which relates to ordinal classification works of Chang *et al.* [5]. The overall objective is to predict the age group of an image using its content, represented as a set of visual features.

This problem has been addressed in the past as a standard multi-class classification problem *e.g.* using the one versus one strategy [12]. One big limitation of this strategy is that the ordinal nature of the data is not taken into account. For instance, when the classifier does not predict the correct date class, the importance of the error is not taken into account: An error of 100 years in a prediction is not penalized more heavily than a error of 10 years. Another limitation is that each classifier is trained using a small amount of the training data (only data from two classes is used per classifier).

In contrast, we postulate that it is easier to answer questions such as “*Was this image taken before or after this date?*” rather than predicting the date. The rationale is that imaging technologies as well as image contents have evolved over the history of photography, and can be characterized by milestones (*e.g.* date of appearance of color images). Therefore, we suggest to formulate age prediction as a combination of before/after binary classifiers. This strategy as been used in the past in the context of ordinal classification by Frank and Hall [10], who have shown how to cast any  $K$ -class classification problem into a set of  $K - 1$  simpler binary sub-tasks encoding ordinal information.

One of the advantages of this *Before versus After (BvA)* strategy, is that – in addition to encoding ordinal information – each one of the  $K - 1$  binary classifiers will learn from the whole training set. Furthermore, it is computationally more efficient as only  $K - 1$  classifiers have to be trained where the one versus one strategy requires training  $\binom{K}{2}$  classifiers.

Our *BvA* strategy has two steps, one for the before/after prediction, the other for predicting age groups from sets of before/after binary predictions. Let us define more precisely these two steps.

**Binary Before/After classifiers.** We are considering age classes as ordered discrete finite sets of rank labels denoted as  $L = \{1 < \dots < k < \dots < K\}$  where  $K$  is the number of age classes. Thus, a training image will be given

as a pair  $(x_i, y_i)$ , where  $x_i \in \mathbb{R}^d$  represents the  $d$ -dimensional visual features of the image and  $y_i \in L$  its age class. We define a set of  $K - 1$  classifiers, where each classifier  $\Psi_n$  is specialized in predicting whether or not an image belongs to an age class greater than  $n$ .

The classifiers at the beginning (respectively, the end) of the dataset's date interval will have more *before* than *after* examples (respectively, more *after* than *before* examples). To deal with this imbalance, we are weighting the classifiers relative to the number of original classes (decades) present in the before and after parts using cost-sensitive techniques such as those of Elkan [9].

In addition, it is worth noting that examples far from the date separation threshold are the most representative examples of the concerned period of time. In contrast, the ones closer to the boundaries are less informative. To take this specificity into account, we must give importance to examples relative to their distances from the actual classifier split. The solution we propose is to give heavier weights to training examples far from the date threshold, the rationale being that errors made on examples closer to the date thresholds are more acceptable. As an example, if we consider the classifier  $\Psi_n$ , the weights denoted as  $w(n, k)$  applied to the training examples accordingly to their original classes values  $k$  are defined by:

$$w(n, k) = \begin{cases} |n - k| & \text{if } n < k, \\ |n - (k - 1)| & \text{else.} \end{cases} \quad (1)$$

Other weighting functions are possible, for instance cumulative score [5] or squared distance, but in practice they all give similar performance for the application considered, which may be due to the low number of classes.

In practice, the before/after classifiers we use are binary linear classifiers. In the next step, for the prediction of the date class from the binary before/after classifiers, we need to have the probability that the class of the image is above a date threshold. We therefore need an additional step to turn the SVM output into probabilities. For the instance  $x_i$ , the probability given by the classifier  $n$  will be noted  $P(\Psi_n|x_i)$ . We rely on the Platt model [13] to compute the subsequent  $P(\Psi_n|x_i)$ , the probability of the image  $x_i$  to be after the date threshold  $n$ .

**Predicting the age class from before/after binary classifiers.** In the second step, the before/after predictions given by the  $K - 1$  classifiers will be combined to obtain class predictions. For this, we adopt a simple probabilistic model assuming that binary classifiers are independent given an image. The probability that an image  $x_i$  belongs to the class  $k \in L$  is given as:

$$P(k|x_i) = \prod_{n=1}^{k-1} (P(\Psi_n|x_i)) \prod_{m=k}^{K-1} (1 - P(\Psi_m|x_i)) \quad (2)$$

The class of image  $x_i$  is computed by maximizing  $P(k|x_i)$   $\forall k \in L$ , which is to say that:

$$\hat{y}_i = \hat{y}(x_i) = \arg \max_{k \in L} P(k|x_i). \quad (3)$$

While being different, this formulation is related to the one presented by Frank and Hall [10] and Cardoso *et al.* [4]. The key advantage of our approach is that it is based on an original combination of binary classifiers which incorporates more information to compute the prediction than other dating strategies.

The formula we propose in Equation 2 is intuitive, sound and takes advantage of the set of all classifiers producing more robust and consistent predictions. Existing ordinal methods with probability combination such as those of Frank and Hall [10] or Cardoso *et al.* [4] respectively use the difference between probabilities of two classifiers or the product of a portion of the classifiers probabilities. In contrast, our method proposes a more complete combination as it uses more intersecting information from all classifiers.

## 4. EXPERIMENTS

The proposed ordinal classification framework for dating color image is experimentally validated on Palermo *et al.*'s dataset [12], which is the reference dataset for color image dating. The dataset and their code is available online<sup>1</sup>. The dataset consists of 1,325 dated color images downloaded from Flickr and manually inspected to remove non-photographic content, and split into 5 age groups (classes) corresponding to the 5 decades from 1930s to 1970s. The dataset is balanced, with a total of 265 images per class. For fair comparison, we take exactly the same setup (including the same seven visual features) and use the codes provided by the authors.

Performance is evaluated using the Mean Absolute Error (MAE), also called ranking loss, which is defined as the average deviation of the predicted class from the true class

$$MAE = \frac{1}{|T|} \sum_{x_i \in T} |y(x_i) - \hat{y}(x_i)|, \quad (4)$$

where  $T$  is the test set,  $y(x_i)$  the true class and  $\hat{y}(x_i)$  the predicted class of a given image  $x_i$ .

The MAE is one of the metrics of reference for evaluating ordinal classification [3, 4, 5, 7, 11]. In addition, we report the mean accuracy (ACC), to allow for comparison with the works of Palermo *et al.* [12], even though this metric is not very adequate for the evaluation of image dating, as it does not take the distance of error into account.

We use the same evaluation protocol as Palermo *et al.* [12], in which 215 randomly selected images from each decade are used for training<sup>2</sup> while the remaining 50 images are used for testing. Observe that in the original one versus one strategy over this dataset, each classifier uses only 430 examples while in our approach, each classifier can exploit the 1,075 examples of the dataset. We believe that this allows to improve the quality of the subsequent classification.

The comparative evaluation is performed against the reference work of Palermo *et al.* [12] for color image dating. In addition, we computed 3 ordinal classification approaches from the state-of-the-art, namely (i) the reduction framework of Li and Lin [11], (ii) the probabilistic combination model of Frank and Hall [10] and finally (iii) the probabilistic combination model of Cardoso and Pinto da Costa [4].

The three probability combination methods (ours, Frank and Hall's [10] and Cardoso and Pinto da Costa's [4]) are based on the same *BvA* framework. The specificity of our approach is the use of cost-sensitive techniques based on absolute cost, to take into account the distance relation existing between the classes.

<sup>1</sup><http://graphics.cs.cmu.edu/projects/historicalColor/>

<sup>2</sup>In the original paper, the size of the training set is mistakenly indicated to be 225 images per class.

	ACC ( $\pm$ std)	MAE ( $\pm$ std)
Untrained human annotators ([12])	26.0 ( $\pm$ N/A)	N/A
Frank and Hall [10]	41.36 ( $\pm 1.89$ ) <sup>(*)</sup>	0.99 ( $\pm 0.05$ ) <sup>(**)</sup>
Li and Lin [11]	35.92( $\pm 4.69$ ) <sup>(*)</sup>	0.96( $\pm 0.06$ ) <sup>(**)</sup>
Cardoso and Pinto da Costa [4]	41.32 ( $\pm 2.76$ )	0.95 ( $\pm 0.04$ ) <sup>(**)</sup>
Palermo <i>et al.</i> [12]	44.92( $\pm 3.69$ )	0.93( $\pm 0.08$ ) <sup>(*)</sup>
Proposed approach	42.76 ( $\pm 1.33$ )	<b>0.87 (<math>\pm 0.05</math>)</b>

**Table 1: Performance of the proposed image dating framework on Palermo’s dataset [12], and comparisons with related approaches. Comparisons are done using the same visual features and the same experimental protocol as defined in [12]. We also remind their report on the accuracy of untrained human annotators on this dataset. Statistical significance versus our approach has been systematically computed: <sup>(\*)</sup> denotes  $p < 0.05$ , <sup>(\*\*)</sup> denotes  $p < 0.01$ .**

The MAE (and the ACC) of the classifiers on the test sets are reported, averaged over ten random training/testing splits as in Palermo *et al.*’s [12] framework. We also report standard deviation values. The corresponding results are presented in Table 1, where it is shown that our approach performs better than any of its competitors, by a large margin. For each of the results, we computed statistical significant values using the independent two-sample t-test. In terms of MAE, this demonstrated that our approach significantly outperforms all of its competitors, with  $p < 0.05$  versus Palermo *et al.* [12], and  $p < 0.01$  versus the others. Interestingly, in terms of ACC, there is no statistically significant difference between our method and those of Palermo *et al.* [12] and Cardoso and Pinto da Costa [4], while we significantly outperform the other two approaches. Hence, in statistical terms, it is fair to say that our approach performs best for MAE and is tied for first place in terms of ACC.

## 5. CONCLUSION

In this paper, we introduced to the state-of-the-art a new technique that permits an increase in the precision of image dating. This particular classification problem has the specificity that its classes follow an order relation. We designed the *Before versus After* approach to take advantage of this fact, by combining the results of binary classifiers at each class threshold, meant to answer the question “Was a given image taken before or after a given date?”. This way, every classifier makes use of the whole learning set.

Through this combination of binary classifiers, we successfully took advantage of the ordinal nature of time, as we demonstrated in our experiments over the sole publicly available image dating benchmark. Our approach significantly outperformed the state-of-the-art techniques.

Crucially, our technique minimizes the mean average error, while maintaining state-of-the-art performance in terms of accuracy. In other words, our approach makes the same number of mis-classifications, but on average it falls closer to the truth.

In future work, we are planning to learn dates rather than classes, under the assumption that this will trigger more precise predictions. We do not develop this line of work

in the present paper because it would then not be possible to perform a fair comparison with standard classification approaches. Nonetheless, the exploitation of more precise data should in theory allow for further performance gains.

We also believe that the study of image dating needs wider datasets, with more classes and more images, so as to take full advantage of order- and distance-aware approaches to image dating.

## 6. ACKNOWLEDGMENTS

This work is partially supported by the *Conseil Régional de Basse-Normandie* (France).

## 7. REFERENCES

- [1] <http://www.vam.ac.uk/page/d/dating-photographs-from-old-clothes/>; Last accessed 30 November 2013.
- [2] <http://www.museumsassociation.org/>; Last accessed 30 November 2013.
- [3] S. Baccianella, A. Esuli, and F. Sebastiani. Evaluation measures for ordinal regression. In *Ninth International Conference on Intelligent Systems Design and Applications (ISDA'09)*, pages 283–287. IEEE, 2009.
- [4] J. S. Cardoso and J. F. Pinto da Costa. Learning to classify ordinal data: The data replication method. *Journal of Machine Learning Research*, 8:1393–1429, Dec. 2007.
- [5] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 585–592. IEEE, 2011.
- [6] B. Coe. *Guide to Early Photographic Processes*. Victoria & Albert Museum, 1983.
- [7] K. Crammer and Y. Singer. Pranking with ranking. In *Advances in Neural Information Processing Systems 14*, pages 641–647. MIT Press, 2001.
- [8] G. Dias, J. G. Moreno, A. Jatowt, and R. Campos. Temporal web image retrieval. In *19th International Conference on String Processing and Information Retrieval, SPIRE'12*, pages 199–204, 2012.
- [9] C. Elkan. The foundations of cost-sensitive learning. In *17th International Joint Conference on Artificial Intelligence, IJCAI*, pages 973–978, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [10] E. Frank and M. Hall. A simple approach to ordinal classification. In *12th European Conference on Machine Learning (ECML)*, pages 145–156. Springer, 2001.
- [11] L. Li and H. Lin. Ordinal regression by extended binary classification. *Advances in neural information processing systems*, 19:865, 2007.
- [12] F. Palermo, J. Hays, and A. A. Efros. Dating historical color images. In *ECCV (6)*, pages 499–512, 2012.
- [13] J. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In *Advances in Large Margin Classifiers*, 2000.
- [14] A. Scalise and A. C. Blose. Dating images from scanned watermarks, 2010. US Patent Application No: 2010/0329,575. Class: 382218 (USPTO). SERIAL NO 12/491,268. US8290205.