
Nouveau modèle pour la datation automatique de photographies à partir de caractéristiques visuelles¹

Paul MARTIN* — Antoine DOUCET** — Frédéric JURIE*

*Laboratoire GREYC [UMR 6072], Université de Caen Normandie, FRANCE 14032
{paul.martin ; frederic.jurie}@unicaen.fr

**Laboratoire L3i, Université de La Rochelle, FRANCE 17042
antoine.doucet@univ-lr.fr

RÉSUMÉ. Nous présentons, dans cet article, une méthode de datation de photographies par l'usage du contenu visuel de celles-ci. Nous nous sommes inspirés de travaux récents de la vision par ordinateur. Nous avons amélioré la méthode de classification utilisée dans ces travaux en dépassant une limite intrinsèque de leur approche. En effet, ils considèrent la datation d'images comme un problème de classification multi-classes, pour lequel une classe représente un ensemble d'années, mais ignorant l'ordre relatif sous-jacent à l'information temporelle. Dans leur approche soit une prédiction est bonne (période valide) soit elle est mauvaise (période invalide) mais aucune différence n'est faite entre se tromper d'une décennie ou de plusieurs. Nos travaux, s'appuient sur des avancées récentes en classification ordinale. Nous considérons les dates comme des attributs à la fois ordonnés et relatifs et nous proposons un cadre spécifique pour les manipuler.

ABSTRACT. We present in this paper, a method for dating photographs by using their visual content. We were inspired by recent works in computer vision. We improved the classification method used in their work by overcoming an intrinsic limit of their approach. Indeed, they consider dating photographs as a multi-class classification problem, for which a class represents a set of years, but ignoring the relative order underlying the temporal information. In their approach a prediction is either good (valid period) or bad (invalid period), but no distinction is made between being wrong of a decade or more. Our work is based on recent advances in ordinal classification. We consider dates as attributes that are both ordered and relative and we propose a specific framework for handling them.

MOTS-CLÉS : Datation de photographies, Classification ordinale, Vision par ordinateur.

KEYWORDS: Photographs dating, Ordinal classification, Computer vision.

1. Traduction étendue d'un article publié en anglais à ICMR 2014 (Martin *et al.*, 2014)

1. Introduction

La quantité d'images disponibles en ligne est en constante augmentation. Une partie importante d'entre elles provient du processus de numérisation des archives, mené à grande échelle par des institutions telles que les musées et les bibliothèques nationales. L'usage intensif des réseaux sociaux draine également une énorme quantité d'images fournies par les utilisateurs. Beaucoup d'entre elles sont des photos « nées numériquement », dont une part importante est issue de la numérisation, notamment par l'usage de scanners.

De ce fait, cette grande quantité de photographies numérisées provient de nombreuses sources. Par conséquent, ces images numériques parviennent avec des métadonnées inconsistantes ou encore incohérentes. La date de la photographie est souvent absente, et parfois même inconnue de la personne en charge du processus de numérisation. Même pour les images « nées numériquement », la date de prise de vue est souvent erronée. En effet, cette information repose sur l'hypothèse d'une configuration appropriée du périphérique d'acquisition. Une illustration de ces cas d'usages est l'existence de nombreuses photos numériques réalisées au début de l'année 1970, tout simplement dû au fait que le 1^{er} janvier 1970 est la date initiale par défaut sur de nombreux appareils.

Dans cet article, nous abordons le problème de la prévision automatique de la date à laquelle une photo numérisée a été prise. La tâche concernant la datation automatique des images a été récemment abordée simultanément dans les communautés de la recherche d'information (Dias *et al.*, 2012) et de la vision par ordinateur (Palermo *et al.*, 2012). Ce sujet a également suscité un vif intérêt dans l'industrie de la photographie. En 2010 la société Kodak a déposé un brevet (Scalise et Blose, 2010) décrivant un processus visant à déterminer la plage de date correspondant à la période durant laquelle une photo a été développée. Cette méthode est basée sur des filigranes visibles sur les faces avant et arrière des photographies. Elle impose donc de disposer des deux faces pour la prédiction.

Formellement, le problème de la datation automatique d'une image consiste simplement à l'associer à la date à laquelle elle a été prise. Bien que cette tâche puisse être considérée comme un problème de classification, où les classes sont différentes plages de dates, il est important de prendre en compte le fait qu'il existe une relation d'ordre reliant toutes les photos sur une frise temporelle. De ce fait, la datation automatique d'une image ne doit pas seulement être évaluée de façon binaire, comme correcte ou incorrecte, mais doit être évaluée selon une mesure de distance entre les prévisions et la véritable date. Il est en effet préférable d'obtenir des estimations temporellement proches des dates réelles de prise de vue.

Obtenir des prévisions temporelles précises des photographies est important pour plusieurs applications. En effet, l'enrichissement d'images existantes avec leurs horodatages permettra d'obtenir des résultats satisfaisant les utilisateurs demandeurs d'informations thématique et temporelle comme « *Je veux des photos de New York dans les années 1920* ».

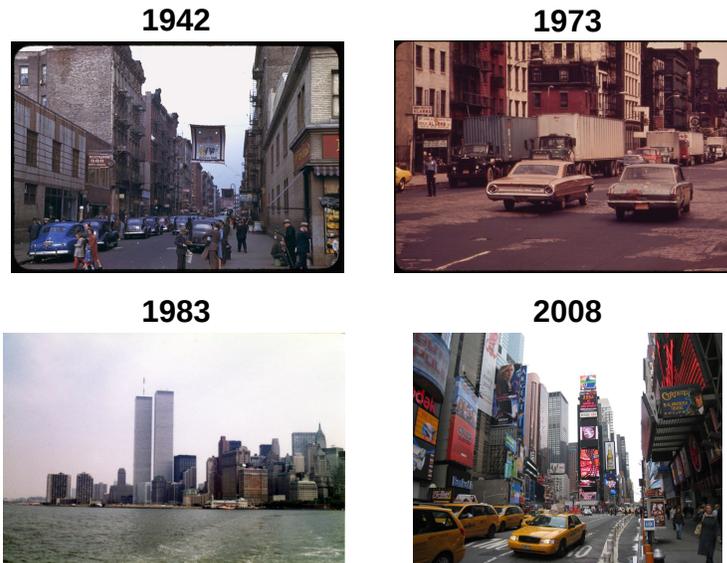


Figure 1 – Quatre vues différentes de la ville de New York. L'objectif général de ce travail est de prévoir automatiquement la date à laquelle ces images ont été prises.

Enrichir les images existantes avec leur dimension temporelle aura de multiples conséquences sur l'utilisation quotidienne des moteurs de recherche. Cela permettra la désambiguïsation entre des termes similaires ayant chacun une époque temporelle associée. Mais aussi la réorganisation d'une collection d'images, présentant ainsi l'évolution visuelle d'un monument ou une célébrité à travers le temps. Par exemple, le besoin d'information « *Je veux des photos de NYC* » pourrait alors être traité visuellement comme le montre la figure 1.

La technique présentée ici ne repose que sur le contenu visuel des images, et ne nécessite aucune description textuelle ; sa facilité d'utilisation dépasse donc l'usage dans le Web. Il peut notamment être utilisé pour organiser des collections d'images stockées dans des archives (par exemple, des bibliothèques ou des musées nationaux), comme un assistant ou un remplacement complet pour l'annotation manuelle par des experts du domaine.

Dans cet article, nous présentons une technique qui tient compte de la nature ordinaire liée au problème de la datation d'images ; de cette manière une amélioration significative sur l'état actuel des techniques de datation est possible. Après un examen des travaux connexes sur la datation des images (section 2) et sur la classification ordinaire (section 3), nous décrivons en détail notre méthode et ses spécificités dans la section 4. Le cadre expérimental et les résultats sont détaillés dans la section 5, puis nous apportons nos conclusions et discutons des orientations futures dans la section 6.

2. Les méthodes existantes pour la datation d'images par le contenu visuel

Dans leurs travaux, Dias *et al.* (2012) proposent de s'attaquer au problème de la datation de photographies grâce à des caractéristiques extraites des images. Étant donné la nouveauté de cette problématique, ils ont alors créé leur jeu de données. Pour ce faire ils ont utilisé une technique d'extension temporelle de requêtes textuelles (de Loupy *et al.*, 1998) basées sur des noms de villes. Ainsi, après filtrage, ils ont obtenu des images des villes aux périodes visées. Ils proposent ensuite d'utiliser une méthode automatique basée sur des caractéristiques visuelles afin de retrouver la période de prise de vue des clichés. À cet effet, ils utilisent 3 caractéristiques visuelles couleurs de l'état de l'art : ScalableColor (Cieplinski, 2001), FCTH (Chatzichristofis et Boutalis, 2008b) et CEDD (Chatzichristofis et Boutalis, 2008a) dans une chaîne de traitement usuelle d'apprentissage supervisé avec un SVM¹ (Vapnik, 1995) linéaire standard avec les paramètres par défaut². Il est important de remarquer qu'ils proposent un découpage du temps déduit d'un *a priori* sur le contexte de cette tâche. En effet, les classes ne sont ni des années ni des décennies mais des périodes temporelles correspondant aux évolutions majeures de la photographie. Leur méthode est testée sur un jeu de données de 1 170 images de villes appartenant aux 5 périodes temporelles suivantes :

- 1826 – 1925 : Premières photographies ;
- 1925 – 1948 : Démocratisation des appareils noir et blanc ;
- 1948 – 1968 : Démocratisation des appareils couleurs ;
- 1968 – 1982 : Apparition des premiers appareils réflexes ;
- 1982 – 2011 : Appareils numériques (jusqu'à la date des expériences réalisées).

Ce problème de classification à 5 classes est évalué avec les métriques standards suivantes : la précision, le rappel, la F-Mesure et l'aire sous la courbe ROC³ et donne des résultats encourageants.

Durant la même année mais dans un domaine différent, Palermo *et al.* (2012) ont proposé une nouvelle approche permettant de prédire la décennie au cours de laquelle une photographie a été réalisée en se basant sur son contenu. Pour ce faire, ils ont constitué un jeu de données comportant 1 325⁴ photographies uniquement en couleur. Les dates de prises de vue de ces photographies s'étalent des années 1930 à la fin des années 1970. Ce jeu de données étalé sur 5 décennies est équilibré, ainsi, chacune des décennies comporte 265 clichés. Leur objectif est la classification de chacune de ces photographies dans l'une de ces 5 décennies. Dans leur approche, la contribution

1. Séparateur à Vaste Marge

2. $C=1$... ils ont utilisé `libsvm` (Chang et Lin, 2011).

3. *Receiver Operating Characteristic*

4. Dans l'article, le nombre d'images de leur jeu de données est annoncé à 1 375 par erreur, nous reportons ici le nombre d'images présentes dans l'archive disponible à l'adresse <http://graphics.cs.cmu.edu/projects/historicalColor/> dernier accès août 2015

principale se concentre sur les caractéristiques visuelles utilisées plutôt que sur le modèle d'apprentissage. Le modèle utilisé est un SVM linéaire utilisant une stratégie *Un contre Un* classique en classification supervisée. Les caractéristiques extraites quant à elles sont principalement des informations de couleurs car leur but est de capturer une information temporellement discriminante basée sur l'évolution au fil du temps des processus photographiques couleurs. Ils utilisent en tout 7 caractéristiques d'images : en plus de trois caractéristiques d'images populaires en classification d'images (Descripteur GIST (Oliva et Torralba, 2001), *Tiny images* (Torralba *et al.*, 2008) et des Histogrammes de couleurs $L^*a^*b^*$ (Xiao *et al.*, 2010)), ils ont mis au point quatre nouvelles caractéristiques d'image spécifiques au problème de la datation de photographies :

- Caractéristique de similarité des processus des pellicules couleurs ;
- Histogrammes de cooccurrence couleur ;
- Probabilité conditionnelle de la saturation sachant la teinte ;
- Histogrammes de teintes.

Ces nouveaux descripteurs sont introduits et présentés en détails dans leur article (Palermo *et al.*, 2012). Ils effectuent une validation croisée à 10 plis et choisissent de manière aléatoire dans chaque pli 80% de la base pour l'apprentissage, soit 1 075 images, et 20% pour le test, soit 250 images. Ils évaluent ensuite leur classification à 5 classes avec la précision⁵ moyenne sur 10 tirages aléatoires. Les résultats de classification montrent une précision de 45,7% dans le meilleur des cas. À titre de comparaison, la précision d'un humain non entraîné sur leur base d'images est de 26% alors que celle d'une classification aléatoire est de 20%. Ceci est donc une approche intéressante car elle offre des résultats supérieurs à l'humain (+19,7 points).

3. Nouvelle approche de datation visuelle grâce à la classification ordinale

Le problème que nous abordons dans cet article peut être formulé de la manière suivante : comment prédire la période de réalisation d'une photographie, en utilisant uniquement ses informations visuelles ? Nous supposons que les dates de prises de vues possibles sont réparties en un nombre fini de classes. Ainsi chaque classe représente une période temporelle. Avec une telle définition, nous nous plaçons dans le cadre d'un problème de classification.

Ce problème a récemment été abordé en temps que problème de classification multi-classes standard *e.g.* au moyen de la stratégie de classification multi-classes *Un contre Un* (Palermo *et al.*, 2012 ; Dias *et al.*, 2012). L'une des principales limitations d'une telle stratégie est que la nature ordinale des données n'est pas prise en compte. En effet, dans cette stratégie, toutes les classes sont opposées entre elles sans prendre en compte l'ordre inhérent au problème de la datation.

⁵. En anglais : *Accuracy*

L'entraînement d'un classifieur est basé sur la minimisation d'une fonction de coût. Dans cette approche, le coût produit par une mauvaise classification est le même, quelque soit l'erreur commise. Il n'y a que deux issues, soit la prédiction est valide soit elle ne l'est pas. Pour que la prédiction soit valide il n'y a qu'une seule configuration possible en revanche toutes les autres configurations entraîneront une prédiction invalide. Ainsi une erreur de prédiction de 40 années n'est pas plus pénalisée qu'une erreur de 10 années. Enfin, une autre limitation vient du fait que chaque classifieur est entraîné en utilisant seulement une petite portion des données d'entraînement. Ceci provient de la nature de la stratégie *Un contre Un* (Knerr *et al.*, 1990) dans laquelle les données utilisées par chacun des sous-classifieurs ne concerne que les 2 classes à opposer.

Notre méthode s'inspire de l'approche de classification ordinale de Frank et Hall (2001) et, de manière plus précise, elle s'apparente aux travaux réalisés par Chang *et al.* (2011). De ce fait, nous supposons, à l'inverse, qu'il est plus facile de répondre à la question : « *Cette photographie a-t-elle été prise avant ou après cette date ?* » que de prédire directement la date, ce qui reviendrait à répondre directement à la question « *Quelle est la date de prise de vue de cette photographie ?* ». Notre raisonnement porte sur le fait que les technologies photographiques ainsi que les contenus des images ont évolué à travers le temps et peuvent être caractérisés par des jalons. Par exemple, l'arrivée de la pellicule couleur *Kodachrome* en 1935. Par conséquent, nous proposons de formuler la classification temporelle comme une combinaison de classifieurs binaires appris selon une succession de séparations temporelles en utilisant la stratégie *Avant contre Après*. Cette stratégie de combinaison a déjà été utilisée par le passé dans le cadre de la classification ordinale par Frank et Hall (2001). Les auteurs ont montré comment transformer tout problème de classification ordinale à K classes en un ensemble simplifié de $K - 1$ sous-problèmes binaires. L'avantage de cette stratégie est que ces sous-problèmes traduisent ainsi l'information ordinale entre les classes.

En plus de prendre en compte l'information temporelle de manière ordinale, la stratégie *Avant contre Après* permet d'utiliser l'intégralité du jeu de données d'apprentissage pour l'entraînement des $K - 1$ classifieurs binaires. Le nombre de sous-classifieurs binaires à apprendre est, en outre, moindre car seulement $K - 1$ classifieurs doivent être formés, là où la stratégie *Un contre Un* exige la formation de $\binom{K}{2}$ classifieurs (Knerr *et al.*, 1990). Nous rappelons également que l'utilisation d'une stratégie *Un contre le Reste* (Vapnik, 1995) va produire elle aussi un petit nombre de classifieurs (K), mais tous les sous-classifieurs appris vont souffrir d'un fort déséquilibre entre les classes positives et négatives.

4. Vers une nouvelle méthode de classification ordinale : *Avant contre Après*

L'usage de la stratégie *Avant contre Après* comporte trois phases. La première est le ré-étiquetage des classes pour l'apprentissage. La seconde concerne la phase d'apprentissage des sous-classifieurs binaires ré-étiquetés. La troisième phase étant

la prédiction de la classe temporelle la plus probable pour une image donnée. Cette prédiction se fait à partir de l'ensemble des sous-classifieurs binaires préalablement appris.

Nous considérons les classes temporelles comme étant un ensemble discret fini d'étiquettes. Les étiquettes des classes temporelles sont désignées par $L = \{1 < \dots < k < \dots < K\}$ où K est le nombre de classes temporelles. Ainsi, lors de l'apprentissage, une image sera donnée comme une paire (x_i, y_i) , où $x_i \in \mathbb{R}^d$ représente un vecteur de caractéristiques visuelles à d dimensions de l'image, appelé aussi signature ou descripteur de l'image, et $y_i \in L$ sa classe temporelle.

En se basant sur cette définition, nous allons présenter ces étapes dans les sous-sections suivantes en illustrant les particularités majeures de cette approche au fil de la discussion. Afin d'illustrer notre méthode de manière cohérente avec la section 5 nous nous plaçons dans le même problème de datation de photographies pour lequel l'objectif est d'attribuer la bonne décennie à un cliché parmi 5 décennies allant des années 1930 aux années 1970. Dans cet exemple, les étiquettes $k \in L$ avec $L = \{1 < 2 < 3 < 4 < 5\}$ représentent respectivement les 5 décennies suivantes : *les années 1930, les années 1940, les années 1950, les années 1960 et les années 1970.*

4.1. Ré-étiquetage des sous-classifieurs binaires

Nous définissons un ensemble de $K - 1$ classifieurs binaires, où chaque classifieur Ψ_n a pour rôle de prédire, si, oui ou non, une image appartient à une classe temporelle supérieure à n . Ici $n \in [1, 2, 3, 4]$ représente les années par lesquelles la séparation temporelle des classes d'apprentissage a lieu. C'est à dire dans notre exemple, les années 1940⁶, 1950, 1960 et 1970. L'illustration de cette première étape de ré-étiquetage est présentée dans le tableau 1.

Classifieurs $\Psi_n \setminus$ Étiquettes $k \in L$	1	2	3	4	5
Ψ_4 : Avant contre Après 1970	-	-	-	-	+
Ψ_3 : Avant contre Après 1960	-	-	-	+	+
Ψ_2 : Avant contre Après 1950	-	-	+	+	+
Ψ_1 : Avant contre Après 1940	-	+	+	+	+

Tableau 1 – Ré-étiquetage des exemples d'apprentissage des sous-classifieurs avec la stratégie Avant contre Après.

Ainsi, à partir des 5 classes de notre exemple, 4 sous-classifieurs issus de la stratégie de ré-étiquetage *Avant contre Après* peuvent être appris aux différents seuils de séparation donnés par n .

6. $n = 1$ représente la séparation de part et d'autre des années 1940, ainsi la classe négative est représentée par la décennie 1930 et la classe positive est représentée par les décennies 1940, 1950, 1960 et 1970.

4.2. Apprentissage des sous-classifieurs binaires

Avec la stratégie *Avant contre Après* chacun des sous-classifieurs est calibré en utilisant l'ensemble des données d'apprentissage. La quantité d'exemples utilisables est donc constante pour tous les sous-classifieurs que le jeu de données soit équilibré ou non. Cependant la proportion d'instances positives et négatives varie selon chaque sous-classifieur comme nous l'illustrons sur notre exemple dans le tableau 2.

Classifieurs $\Psi_n \setminus$ Étiquettes $k \in L$	1	2	3	4	5
Ψ_4 : Avant contre Après 1970	860				215
Ψ_3 : Avant contre Après 1960	645			430	
Ψ_2 : Avant contre Après 1950	430		645		
Ψ_1 : Avant contre Après 1940	215	860			

Tableau 2 – Impact du ré-étiquetage sur la balance des exemples d'apprentissage des sous-classifieurs *Avant contre Après*.

Dans cet exemple, le jeu de données d'apprentissage est équilibré et comporte 215 exemples d'apprentissage par classe. Ce qui fait un total de 1 075 exemples qui sont utilisés par chacun des sous-classifieurs. Nous pouvons observer dans le tableau 2 que le sous-classifieur binaire Ψ_1 appris au début de l'ensemble des classes temporelles utilisera plus d'exemples « après » le seuil de séparation temporelle qu'« avant ». Pour faire face à ce déséquilibre, nous pondérons les classes d'entraînement ré-étiquetées de ces sous-classifieurs par rapport au nombre de classes temporelles originalement présentes dans les parties « avant » et « après ». Nous utilisons des techniques de pondération issues de l'apprentissage sensible aux coûts⁷, telles que celles développées par Elkan (2001). Nous présentons le calcul des poids des classes ré-étiquetées dans le tableau 3.

Classifieurs $\Psi_n \setminus$ Étiquettes $k \in L$	1	2	3	4	5
Ψ_4 : Avant contre Après 1970	1/4				1
Ψ_3 : Avant contre Après 1960	1/3			1/2	
Ψ_2 : Avant contre Après 1950	1/2		1/3		
Ψ_1 : Avant contre Après 1940	1	1/4			

Tableau 3 – Calculs des poids appliqués aux classes ré-étiquetées des sous-classifieurs obtenus avec la stratégie *Avant contre Après*.

De plus, il est à noter que les exemples distants du seuil de séparation temporelle sont des exemples plus représentatifs de la période de temps concernée. À l'inverse, ceux proches de la séparation sont moins informatifs.

7. En anglais : *Cost-sensitive learning*

Pour prendre en compte cette spécificité, nous proposons de donner aux exemples d'apprentissage du modèle une importance qui soit relative à leur distance par rapport à la séparation concernée pour chaque sous-classifieur. La solution, que nous proposons, est de donner des poids plus importants aux exemples lointains de la séparation temporelle, car les erreurs commises sur des exemples plus proches des seuils de date sont plus tolérables.

A titre d'exemple, si nous considérons le classifieur Ψ_n , les poids, que nous noterons, $w(n, k)$ sont appliqués aux exemples d'apprentissage en fonction de la classe originale k associée aux exemples de cette manière :

$$w(n, k) = \begin{cases} |n - k| & \text{si } n < k, \\ |n - (k - 1)| & \text{sinon.} \end{cases} \quad [1]$$

Nous illustrons le calcul des poids des instances originales dans le tableau 4.

Classifieurs $\Psi_n \setminus$ Étiquettes $k \in L$	1	2	3	4	5
Ψ_4 : Avant contre Après 1970	4	3	2	1	1
Ψ_3 : Avant contre Après 1960	3	2	1	1	2
Ψ_2 : Avant contre Après 1950	2	1	1	2	3
Ψ_1 : Avant contre Après 1940	1	1	2	3	4

Tableau 4 – Pondérations des exemples d'apprentissage par le poids $w(n, k)$ de l'équation 1.

D'autres fonctions de pondération sont possibles, par exemple, le score cumulatif (Chang *et al.*, 2011) ou encore le carré des distances, *etc.* En pratique, elles donnent toutes des performances voisines, pour les expériences que nous avons menées, ce qui peut être dû au faible nombre de classes considérées.

Dans nos expériences pratiques, chaque sous-classifieur appris, avec la stratégie *Avant contre Après*, est un classifieur SVM⁸ linéaire (Vapnik, 1995). Dans l'étape suivante, la prédiction de la classe temporelle finale est calculée par un mécanisme probabiliste combinant chacune des probabilités obtenues par tous les sous-classifieurs appris avec la stratégie *Avant contre Après*.

Nous devons donc introduire une étape supplémentaire pour transformer les sorties des SVMs en probabilités. Pour l'instance x_i , la probabilité donnée par le classifieur Ψ appris à la séparation n sera noté $P(\Psi_n|x_i)$. Nous utilisons le modèle de Platt (2000) pour calculer⁹ $P(\Psi_n|x_i)$, la probabilité que l'image x_i appartienne aux classes situées après la séparation temporelle de la date n .

8. Séparateur à Vaste Marge

9. Utilisation de la fonction *sigmoid* pour transformer les distances à l'hyperplan séparateur en probabilités.

4.3. Prédiction à partir des sous-classifieurs binaires

Dans cette troisième étape, les prédictions données par les $K - 1$ sous-classifieurs sont combinées pour obtenir les prédictions de classes temporelles. À ce sujet, il existe plusieurs méthodes de combinaison (Frank et Hall, 2001 ; Cardoso et Pinto da Costa, 2007).

Dans nos travaux, nous nous sommes inspirés des travaux de Cardoso et Pinto da Costa (2007). En effet, comme nous pouvons le voir dans l'équation 2, les probabilités obtenues avec leur modèle ne sont pas « consistantes » pour toutes les classes.

$$\begin{aligned}
 P(1|x_i) &= 1 - P(\Psi_1|x_i) \\
 P(k|x_i) &= \prod_{n=1}^{k-1} (P(\Psi_n|x_i))(1 - P(\Psi_k|x_i)), 1 < k < K \\
 P(K|x_i) &= \prod_{n=1}^{K-1} (P(\Psi_n|x_i))
 \end{aligned} \tag{2}$$

Pour les classes $k \in [1, K]$ il n'y a que k classifieurs qui participent au vote final. Une amélioration serait de considérer tous les classifieurs pour toutes les classes.

Ainsi, pour garantir la cohérence de part et d'autre de la séparation temporelle sur l'ensemble des classes, le calcul des probabilités par classe pourrait être fait de cette manière :

$$\begin{aligned}
 P(1|x_i) &= \prod_{m=1}^{K-1} (1 - P(\Psi_m|x_i)) \\
 P(k|x_i) &= \prod_{n=1}^{k-1} (P(\Psi_n|x_i)) \prod_{m=k}^{K-1} (1 - P(\Psi_m|x_i)), 1 < k < K \\
 P(K|x_i) &= \prod_{n=1}^{K-1} (P(\Psi_n|x_i))
 \end{aligned} \tag{3}$$

Dans la formule 3 nous proposons une version améliorée de la formule 2 de Cardoso et Pinto da Costa (2007). Ainsi, grâce à notre formulation, pour chacune des classes tous les classifieurs participeront au vote. Nous pensons que cette information est alors plus cohérente.

Décomposons ce qui se produit avec ce modèle de combinaison pour la classe k : pour tous les classifieurs avant k , nous prenons les probabilités qu'ils votent après k et inversement pour tous ceux après k , les probabilités qu'ils votent avant k .

Comme pour les autres stratégies, la classe estimée pour l'instance x_i est la classe (k) ayant la probabilité la plus élevée.

Du fait de la consistance de notre modèle, nous pouvons simplifier sa notation (formule 3) en une seule expression ainsi :

$$P(k|x_i) = \prod_{n=1}^{k-1} (P(\Psi_n|x_i)) \prod_{m=k}^{K-1} (1 - P(\Psi_m|x_i)) \quad [4]$$

Ce modèle donne la probabilité qu'une image x_i appartienne aux classes $k \in L$. Ceci représente donc un nouveau et simple modèle probabiliste de combinaison de classifieurs binaires appris avec la stratégie *Avant contre Après*.

Enfin, la classe de l'image x_i est calculée en maximisant $P(k|x_i) \forall k \in L$, i.e. :

$$\hat{y}_i = \hat{y}(x_i) = \arg \max_{k \in L} P(k|x_i). \quad [5]$$

Nous pouvons appliquer la formule 4 sur notre exemple illustratif, et observer le détail de la combinaison des probabilités de chacune des 5 classes à partir des 4 classifieurs issus de la stratégie *Avant contre Après* ci dessous :

$$\begin{aligned} P(5|x_1) &= P(\Psi_1|x_1) * P(\Psi_2|x_1) * P(\Psi_3|x_1) * P(\Psi_4|x_1) \\ P(4|x_1) &= P(\Psi_1|x_1) * P(\Psi_2|x_1) * P(\Psi_3|x_1) * (1 - P(\Psi_4|x_1)) \\ P(3|x_1) &= P(\Psi_1|x_1) * P(\Psi_2|x_1) * (1 - P(\Psi_3|x_1)) * (1 - P(\Psi_4|x_1)) \\ P(2|x_1) &= P(\Psi_1|x_1) * (1 - P(\Psi_2|x_1)) * (1 - P(\Psi_3|x_1)) * (1 - P(\Psi_4|x_1)) \\ P(1|x_1) &= (1 - P(\Psi_1|x_1)) * (1 - P(\Psi_2|x_1)) * (1 - P(\Psi_3|x_1)) * (1 - P(\Psi_4|x_1)) \end{aligned}$$

Comme dit précédemment, la décennie qui sera attribuée à la photographie sera celle recevant la probabilité maximale (5).

Bien qu'étant différente, cette formulation est liée à celle présentée par Frank et Hall (2001) et Cardoso et Pinto da Costa (2007). L'avantage et la différence majeure de notre approche résident dans le fait que pour chacune des classes tous les classifieurs sont utilisés dans la phase de combinaison. Le modèle ainsi appris, incorpore plus d'informations pour calculer la prédiction à l'inverse des autres stratégies.

L'expression que nous proposons dans l'équation 4 est intuitive ; elle tire profit de l'ensemble de tous les classifieurs produisant ainsi des prédictions plus cohérentes et plus complètes. Les méthodes ordinales existantes effectuant des combinaisons de probabilités utilisent respectivement la différence entre les probabilités de deux classifieurs ou le produit d'une partie des probabilités des classifieurs (Frank et Hall, 2001 ; Cardoso et Pinto da Costa, 2007). À l'inverse, notre méthode propose une combinaison plus complète car elle exploite l'information d'intersection entre tous les classifieurs.

En outre, la combinaison proposée est basée sur le produit des probabilités. Ce fonctionnement est préférable aux méthodes qui opèrent avec la somme des valeurs

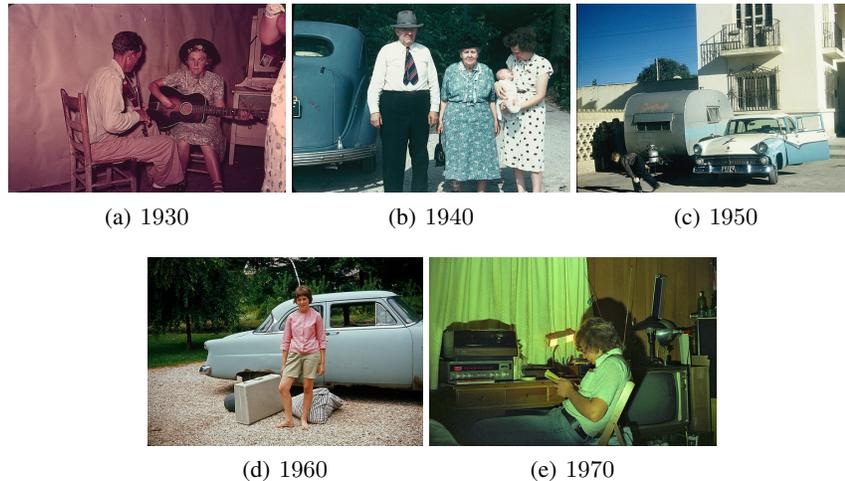


Figure 2 – Exemples d’images de la collection de Palermo et al. (2012).

de confiance seuillées issues des classifieurs (Li et Lin, 2007). En effet, ces dernières soulèvent un problème vis-à-vis de la statistique puisqu’elles combinent des probabilités en utilisant des sommes et non des produits.

5. Expérimentations

Le *framework* de classification ordinaire que nous avons proposé pour la datation des images couleur est validé expérimentalement sur le jeu de données de référence pour la datation d’images couleur (Palermo *et al.*, 2012). Quelques images sont présentées en exemple dans l’illustration 2. Le jeu de données et leurs codes sources sont disponibles en ligne¹⁰. La base de données est constituée de 1 325¹¹ images couleur pour lesquelles les dates de prises de vues sont connues. Ces images proviennent du site *flickr* et ont été inspectées manuellement afin de supprimer le contenu non photographique et parasite. Elles ont été divisées en 5 classes d’âges correspondant aux 5 décennies de 1930 à 1970. L’ensemble est équilibré, avec un total de 265 images par classe. Pour que notre comparaison soit équitable, nous prenons exactement la même configuration (y compris les sept mêmes caractéristiques visuelles) et nous utilisons les codes fournis par les auteurs.

10. <http://graphics.cs.cmu.edu/projects/historicalColor/> dernier accès août 2015

11. Dans l’article, le nombre d’images de leur jeu de données est annoncé à 1 375 par erreur, nous reportons ici le nombre d’images présentes dans l’archive disponible à l’adresse <http://graphics.cs.cmu.edu/projects/historicalColor/> dernier accès août 2015

La performance est évaluée en utilisant l'Erreur Absolue Moyenne (MAE)¹², qui est définie comme étant l'écart moyen, ici exprimé en décennies, entre la classe prédite et la véritable classe,

$$MAE = \frac{1}{|T|} \sum_{x_i \in T} |y(x_i) - \hat{y}(x_i)|, \quad [6]$$

Dans l'équation 6, T désigne l'ensemble de test, $y(x_i)$ la véritable classe et $\hat{y}(x_i)$ la classe prédite pour une image donnée x_i .

Le MAE (6) est l'une des mesures de référence pour évaluer la classification ordinaire (Li et Lin, 2007 ; Cardoso et Pinto da Costa, 2007 ; Baccianella *et al.*, 2009 ; Chang *et al.*, 2011). Bien que la précision¹³ moyenne (ACC) ne soit pas très adéquate pour l'évaluation de la datation d'images, car elle ne tient pas compte de l'importance de l'erreur, nous rapportons les résultats obtenus avec celle-ci afin de permettre la comparaison avec les travaux de Palermo *et al.* (2012).

Nous utilisons le même protocole d'évaluation que Palermo *et al.* (2012). Dans chacune des décennies, 215 images sont sélectionnées au hasard et sont utilisées pour l'apprentissage¹⁴ tandis que les 50 images restantes sont utilisées pour les tests. On peut remarquer que, dans la stratégie standard *Un contre Un* sur ce jeu de données, chaque classifieur opposant donc 2 classes à la fois, utilise seulement 430 exemples. Dans notre approche, chaque classifieur exploite cette fois, les 1 075 exemples¹⁵ du jeu de données d'apprentissage. Selon nous, cela permet d'améliorer la qualité de la classification finale.

L'évaluation comparative est effectuée par rapport aux travaux de référence pour la datation des images couleur (Palermo *et al.*, 2012). Nous avons aussi calculé les résultats avec 3 méthodes de classification ordinaire de l'état de l'art, à savoir : un *framework* de réduction (Li et Lin, 2007) et deux modèles de combinaison probabiliste (Frank et Hall, 2001 ; Cardoso et Pinto da Costa, 2007). Les 3 méthodes de combinaison probabilistes (la nôtre et les deux citées précédemment) sont basées sur la même stratégie *Avant contre Après*.

La spécificité de notre approche est l'utilisation de techniques de sensibilité aux coûts¹⁶ basés sur le calcul d'un coût absolu, afin de tenir compte de la relation de distance existante entre les classes. Nous rappelons ici que ces techniques ne sont pas utilisées dans les travaux originaux de Frank et Hall (2001) ni dans ceux de Cardoso et Pinto da Costa (2007).

12. En anglais : *Mean Absolut Error*, également appelée *Ranking loss*

13. En anglais : *Accuracy*

14. Dans le document original, la taille de l'ensemble d'apprentissage par classe est indiqué par erreur à 225 images par classe.

15. Comme nous l'avons présenté dans le tableau 2 à la page 8.

16. En anglais : *Cost-sensitive*

Approches \ Mesures	ACC (\pm écart-type)	MAE (\pm écart-type)
Classe aléatoire	20	1.6
Valeur centrale : années 1950	20	1.2
Annotateurs humains non formés Palermo <i>et al.</i> (2012)	26.0 (\pm N/A)	N/A
Frank et Hall (2001)	41.36 (\pm 1.89) (*)	0.99 (\pm 0.05) (**)
Li et Lin (2007)	35.92(\pm 4.69) (*)	0.96(\pm 0.06) (**)
Cardoso et Pinto da Costa (2007)	41.32 (\pm 2.76)	0.95 (\pm 0.04)(**)
Palermo <i>et al.</i> (2012)	44.92(\pm 3.69)	0.93(\pm 0.08) (*)
Approche proposée	42.76 (\pm 1.33)	0.87 (\pm0.05)

Tableau 5 – Performance de l’approche proposée sur le jeu de données de (Palermo et al., 2012), ainsi que les comparaisons avec les approches connexes.

La *MAE* (et l’*ACC*) de ces 5 approches sur le jeu de données de test sont calculées et moyennées sur 10 découpages aléatoires entre les ensembles d’apprentissage et de test, comme spécifié dans le protocole expérimental initial (Palermo *et al.*, 2012). Nous rapportons également les valeurs d’écart-type. Les résultats correspondants sont présentés dans le tableau 5, où il est montré que notre approche fonctionne significativement mieux que celles des concurrents.

Les comparaisons sont effectuées en utilisant les mêmes caractéristiques visuelles et le même protocole expérimental. Nous rappelons également les résultats qu’ils ont obtenus sur la précision d’annotateurs humains non formés sur cette problématique et ce jeu de données. Nous avons de plus ajouté les résultats obtenus par un tirage aléatoire des classes parmi les 5 disponibles ainsi que la prédiction constante de la classe centrale. À ce propos, nous pouvons observer que bien qu’ayant des *ACC* similaires, le choix de la classe centrale lui, minimise la distance d’erreur moyenne (*MAE*) dans ce problème équilibré en comparaison du choix aléatoire.

Pour chacun des résultats, nous avons calculé si nos résultats sont statistiquement significatifs en utilisant le test de Student (*t-test*) sur deux échantillons indépendants. La signification statistique par rapport à notre approche a été systématiquement calculée : (*) veut dire que $p < 0.05$, (**) veut dire que $p < 0.01$. En termes de *MAE*, cela démontre que notre approche est nettement meilleure que toutes les autres, avec $p < 0,05$ par rapport à Palermo *et al.* (2012), et $p < 0,01$ par rapport aux autres. Fait intéressant, en termes d’*ACC*, la différence n’est pas statistiquement significative entre notre méthode et celle de Palermo *et al.* (2012) et Cardoso et Pinto da Costa (2007), alors que nous surpassons nettement les deux autres approches. Par conséquent, en termes statistiques, il est juste de dire que notre approche fonctionne mieux concernant la *MAE* et est à égalité pour la première place en termes d’*ACC*.

6. Conclusion

Nous avons présenté une nouvelle technique qui améliore la datation d'images. Ce problème de classification est particulier en ce sens que ses classes suivent une relation d'ordre. Nous avons redéfini la problématique de la datation de photographies à travers cette question : « *Cette photographie a-t-elle été prise avant ou après cette date ?* ». Nous avons mis en place un *framework* de classification ordinaire utilisant une stratégie *Avant contre Après*. Nous avons proposé une nouvelle stratégie de combinaisons des résultats des sous-classifieurs binaires obtenus à chaque seuil de séparation temporelle. Cette stratégie utilise les réponses de chaque sous-classifieur.

Grâce à cette combinaison de classifieurs binaires, nous avons tiré parti de la nature ordinaire du temps, comme nous l'avons démontré dans nos expériences. Notre approche a dépassé les techniques de l'état de l'art. Fondamentalement, notre technique minimise la distance d'erreur moyenne, tout en maintenant une précision à la hauteur de l'état de l'art. En d'autres termes, notre approche donne le même nombre de mauvaises classifications, mais en moyenne, celles-ci tombent plus près de la valeur réelle.

À l'avenir, nous prévoyons d'apprendre des années plutôt que des périodes plus larges, dans l'hypothèse que les prédictions plus précises pourront être obtenues. Une autre hypothèse de travail serait de « deviner » à partir des données, des seuils de séparation temporels qui soient caractéristiques de la photographie.

7. Bibliographie

- Baccianella S., Esuli A., Sebastiani F., « Evaluation measures for ordinal regression », *Ninth International Conference on Intelligent Systems Design and Applications, 2009. ISDA'09.*, IEEE, p. 283-287, 2009.
- Cardoso J. S., Pinto da Costa J. F., « Learning to Classify Ordinal Data: The Data Replication Method », *Journal of Machine Learning Research*, vol. 8, p. 1393-1429, December, 2007.
- Chang C.-C., Lin C.-J., « LIBSVM: A library for support vector machines », *ACM Transactions on Intelligent Systems and Technology*, vol. 2, p. 27:1-27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chang K.-Y., Chen C.-S., Hung Y.-P., « Ordinal hyperplanes ranker with cost sensitivities for age estimation », *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, p. 585-592, 2011.
- Chatzichristofis S. A., Boutalis Y. S., « CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval », *Proceedings of the 6th International Conference on Computer Vision Systems, ICVS'08*, Springer-Verlag, Berlin, Heidelberg, p. 312-322, 2008a.
- Chatzichristofis S. A., Boutalis Y. S., « FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval », *Proceedings of the 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS '08*, IEEE Computer Society, Washington, DC, USA, p. 191-196, 2008b.

- Cieplinski L., « MPEG-7 Color Descriptors and Their Applications. », in W. Skarbek (ed.), *CAIP*, vol. 2124 of *Lecture Notes in Computer Science*, Springer, p. 11-20, 2001.
- de Loupy C., Bellot P., El-Bèze M., Marteau P., « Query Expansion and Classification of Retrieved Documents », *Proceedings of The Seventh Text REtrieval Conference, TREC 1998, Gaithersburg, Maryland, USA, November 9-11, 1998*, p. 382-389, 1998.
- Dias G., Moreno J. G., Jatowt A., Campos R., « Temporal web image retrieval », *String Processing and Information Retrieval*, Springer, p. 199-204, 2012.
- Elkan C., « The Foundations of Cost-sensitive Learning », *17th International Joint Conference on Artificial Intelligence, IJCAI*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 973-978, 2001.
- Frank E., Hall M., « A simple approach to ordinal classification », *12th European Conference on Machine Learning (ECML)*, Springer, p. 145-156, 2001.
- Knerr S., Personnaz L., Dreyfus G., « Single-layer learning revisited: a stepwise procedure for building and training a neural network », in F. Soulié, J. Hérault (eds), *Neurocomputing*, vol. 68 of *NATO ASI Series*, Springer Berlin Heidelberg, p. 41-50, 1990.
- Li L., Lin H., « Ordinal regression by extended binary classification », *Advances in neural information processing systems*, vol. 19, p. 865, 2007.
- Martin P., Doucet A., Jurie F., « Dating Color Images with Ordinal Classification », *Proceedings of International Conference on Multimedia Retrieval*, ACM, p. 447, 2014.
- Oliva A., Torralba A., « Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope », *International Journal of Computer Vision*, vol. 42, n° 3, p. 145-175, May, 2001.
- Palermo F., Hays J., Efros A. A., « Dating Historical Color Images », *Proceedings of the 12th European conference on Computer Vision-Volume Part VI*, p. 499-512, 2012.
- Platt J., « Probabilistic outputs for support vector machines and comparison to regularized likelihood methods », *Advances in Large Margin Classifiers*, 2000.
- Scalise A., Blose A. C., « Dating images from scanned watermarks », December 30, 2010. US Patent App. 12/491,268.
- Torralba A., Fergus R., Freeman W. T., « 80 million tiny images: a large dataset for non-parametric object and scene recognition », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.
- Vapnik V. N., *The nature of statistical learning theory*, Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- Xiao J., Hays J., Ehinger K. A., Oliva A., Torralba A., « Sun database: Largescale scene recognition from abbey to zoo », *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.