

L'écrit et le document

Analyse de Partitions Musicales

Music Score Analysis

par V. POULAIN D'ANDECY*, J. CAMILLERAPP**, I. LEPLUMEY**

* ITESOFT, Immeuble Alliance, Parc Scientifique Georges Besse, 30000 Nîmes
Tél : 66 29 00 92, Fax : 66 29 02 37

**IRISA/INSA, 20 avenue des buttes de coësmes 35043 Rennes Cedex
Tél : 99 28 64 00, e-mail : Jean.Camillerapp@irisa.fr

Résumé

Les partitions musicales sont des documents qui comportent de nombreux symboles constitués de segments de droite. Dans le but d'extraire ces segments, nous avons mis au point un détecteur basé sur la technique du filtrage de Kalman. Cette méthode modélise les caractéristiques des segments des images binaires à analyser (épaisseurs, biais, légères courbures, bruits, superpositions, discontinuités). En appliquant méthodiquement ce détecteur et en utilisant quelques règles simples de classification sur les segments trouvés, on reconnaît les portées, les queues de note, les liaisons, les barres de croche et les têtes noires. On peut alors reconstruire les croches et les groupes de croches.

Mots clés : segmentation, vectorisation, suivi de ligne, filtrage de Kalman, reconnaissance de partitions musicales

Abstract

Many symbols in music scores are linear segments. In this context, we designed an extractor of segments, which modelizes the characteristics of binary images (scale factor, curvature, bias and noises). This algorithm is based on Kalman filtering technique. We were able to recognize music symbols (staves, stems, slurs, beams, bar lines, black note heads, quavers and note groups) by applying both this extractor and simple rules of classification for the detected segments, to the defined layers.

Key words : Segmentation, Vectorization, Line Tracking, Kalman Filtering, Music Scores Analysis

1. Introduction

Certains documents, tels que les chèques, les dessins industriels, les plans cadastraux, ou plus largement les documents techniques, peuvent être vus comme la superposition de couches de graphiques et de textes [BT92]. Un chèque se décompose en deux couches : une couche de fond (information pré-imprimée, lignes de base, ...) et une couche d'écriture manuscrite [LCQ95]. De même, une partition musicale imprimée, tout comme une partition manuscrite, se décompose naturellement en deux couches [BB92b] : une couche contenant les lignes de portée et une couche contenant tous les autres symboles. La séparation de ces deux couches est considérée par la plupart des auteurs comme un traitement nécessaire à leur interprétation.

Pour ce faire, des techniques de segmentation et de classification très différentes ont été mises en œuvre : projections [KI90, Fuj88], transformée de Hough [BB92a], calcul des cordes discrètes [Mar92], étiquetage des régions blanches [LC91], analyse du Graphe des Lignes Adjacentes [Car89], "pattern matching" ou réseaux neuronaux [Mar92], etc... Souvent combinées entre-elles,

ces techniques ont permis d'obtenir des résultats partiels, très dépendant de l'ensemble des paramètres (facteur d'échelle, angle des portées par rapport à l'horizontale) et des défauts des images binaires (distorsions, courbures, discontinuités, bruits). Ces défauts sont parfois dus aux techniques d'acquisition, mais ils sont souvent liés à la mauvaise qualité du document original lui-même. Ainsi, un document peut-il être mal imprimé et comporter des lignes de portée discontinues. Le support papier peut être trop fin et laisser apparaître à l'acquisition une silhouette des informations imprimées au verso. En conséquence, il semble important de définir une méthode traitant au mieux ces difficultés.

Nous souhaitons parvenir à une méthode d'analyse des partitions musicales imprimées d'édition classique [Rea79] dans l'optique à long terme, d'une transcription automatique en braille. Nous désirons que la méthode ait une certaine souplesse pour pouvoir s'adapter aux différentes caractéristiques des images à analyser (bruit, échelle, biais...).

On constate comme Kato [KI90] qu'une grande partie des symboles musicaux ont une structure linéaire (lignes de portée, queues de note, barres de mesure...) avec des épaisseurs différentes.

L'utilisation d'un extracteur de segments semble être un choix judicieux qui doit permettre l'obtention d'un maximum de résultats. On peut définir une première classification en quatre couches à partir de cet extracteur :

- couche 1 : les segments "longs" à tendance horizontale (lignes de portée, barres de croche et liaisons);
- couche 2 : les segments "longs" à tendance verticale (queues de note et barres de mesure);
- couche 3 : les segments "très courts et très épais" (les têtes de notes noires);
- couche 4 : tous les autres symboles.

Après avoir présenté ce détecteur qui est une application du filtre de Kalman et qui sera l'unique outil de segmentation, on verra successivement comment repérer et classer les objets de type segment (trois premières classes). Le classement sera basé sur des critères structurels simples déduits des propriétés morphologiques des segments.

Les images de test ont été acquises sur un scanner binaire avec une définition de 300 dpi. Elles correspondent à une douzaine de partitions d'édition classique au format A4. Le traitement de partitions d'écriture contemporaine n'entre pas dans le cadre de cet article, l'étape de classification étant très différente. Seule une sous image [figure 1] est présentée en résultat par la suite, mais les tests et le calcul des statistiques ont été réalisés sur l'ensemble des images.



Figure 1. – Extrait d'une partition musicale d'édition classique acquise avec du biais

2. Détection de segments

Avant de pouvoir parler d'extracteurs de segments, il est nécessaire de définir les propriétés des segments recherchés dans les images à analyser, elles influent sur le choix du détecteur de segment :

- (1) Importance du voisinage : un alignement de points isolés aura peu de sens en terme de segment.
- (2) Présence possible de discontinuités (surtout sur les lignes de portée) : il est utile de pouvoir permettre localement une absence

de points due par exemple à des images de mauvaise qualité. Ces discontinuités peuvent être aussi dues à la nature de l'objet à extraire, par exemple, pour la recherche de segments en pointillés (ensemble de segments courts situés sur une même droite).

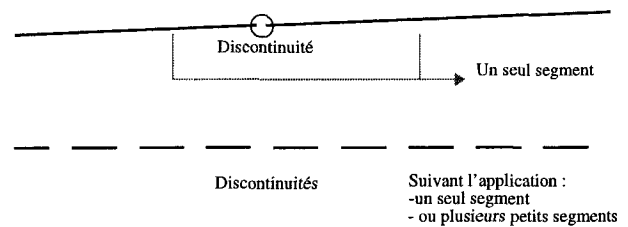


Figure 2. – Prise en compte de discontinuités

- (3) Prise en compte de l'épaisseur.

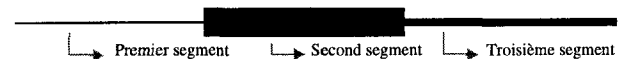


Figure 3. – Prise en compte de l'épaisseur

- (4) Taille très variable des segments, par exemple de quelques dizaines (queues de notes) à plusieurs centaines de points représentatifs (lignes de portée).

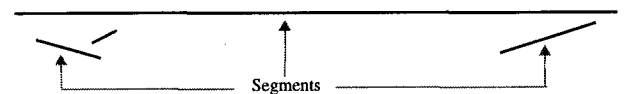


Figure 4. – Prise en compte de variations de taille

- (5) Prise en compte de la superposition de segments : les lignes de portées d'une partition musicale sont recouvertes par les autres symboles musicaux. On veut pouvoir extraire directement ces lignes recouvertes par une autre information. Cette propriété de superposition est issue directement de la notion de couches de documents que nous avons introduite [BB92b].

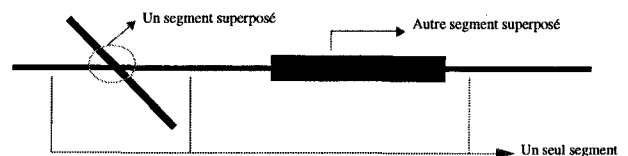


Figure 5. – Prise en compte de superpositions

- (6) Un dernier point qui peut sembler contradictoire avec la notion de segment : nous nous réservons la possibilité d'intégrer une certaine courbure du tracé. Localement, nous avons bien un tracé rectiligne, mais sur l'ensemble de l'image, on accepte une courbure du segment. On peut ainsi utiliser un même modèle pour les lignes de portée et les liaisons.

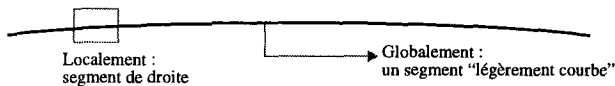


Figure 6. – Prise en compte de courbures

Dans l'image, un empan est une suite connexe de pixels de même couleur et situés dans une même colonne (ou une même ligne) d'une image. Un segment idéal correspond à une suite connexe d'empans noirs de même largeur et dont les milieux appartiennent à une même droite [figure 7]. On appellera par la suite, observation, un empan noir de l'image.

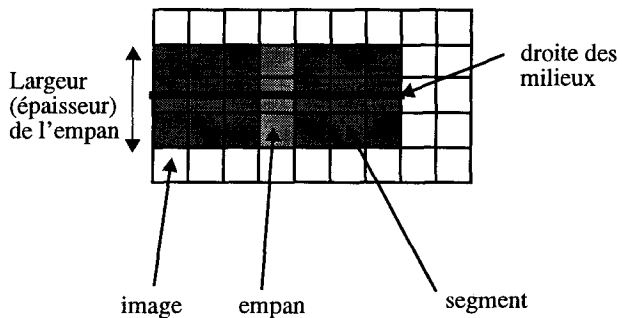


Figure 7. – Exemple de segment et des empans qui le constituent

Le détecteur proposé va extraire directement les segments en évitant si possible leur sur-segmentation (segment coupé en plusieurs morceaux), en se basant sur la cohérence des paramètres des empans (position, épaisseur). Il utilise une technique de prédiction et il réagit en cas d'absence d'empan cohérent, ce qui permet la gestion à la fois des discontinuités et des superpositions. De plus, il intègre une modélisation de l'erreur qui prend en compte à la fois des variations sur les paramètres des empans et une certaine courbure sur la droite porteuse. Ce dernier point donne à la méthode sa souplesse : il suffit de modifier les paramètres d'erreur pour s'adapter aux dégradations que l'on tolère dans l'image.

2.1. PRINCIPE GÉNÉRAL DE LA PRÉDICTION/VÉRIFICATION

Le but de la prédiction/vérification est de suivre, ou d'estimer, l'évolution d'un objet. Pour réaliser cela, il faut caractériser cet objet par un état et concevoir un estimateur qui, connaissant l'évolution de l'état jusqu'à un instant t , permette de prédire l'état en $t+1$.

A partir de cette prédiction pour l'instant $t+1$, on fait la vérification; autrement dit, on regarde l'état réel de l'objet, ou observation, correspondant à la prédiction et on corrige l'estimateur en fonction de l'erreur commise entre l'observation et la prédiction. Ainsi l'estimateur s'améliore avec le temps. La puissance de cette technique réside dans sa tolérance aux erreurs et dans la possibilité

d'effectuer dynamiquement des corrections de l'estimateur. Evidemment, un modèle théorique de l'évolution de l'état de l'objet doit être connu a priori.

Nous avons vu qu'un segment était un regroupement d'empans. En balayant une image ligne par ligne, ou colonne par colonne, un segment est vu comme la suite temporelle des évolutions d'un empan au fil des lignes, ou des colonnes. Cet empan évolue donc dans l'image suivant une droite (la droite support du segment) et doit avoir une épaisseur sensiblement constante (celle du segment). L'état caractérisant un empan est donc :

- sa largeur (i.e. l'épaisseur du segment);
- sa position orthogonale au sens de balayage;
- la pente de la droite support.

Enfin, la prédiction/vérification pose ici le problème du choix de l'observation correspondant à une prédiction. En effet, en balayant une image, on suit simultanément plusieurs segments. On obtient alors, pour un pas de balayage, n prédictions et N empans réellement observés (avec n et N positifs et pas forcément égaux).

2.2. EQUATIONS THÉORIQUES DU FILTRE DE KALMAN

Le filtre de Kalman formalise le principe de la prédiction/vérification. C'est un estimateur linéaire récursif qui permet d'intégrer automatiquement les erreurs constatées par rapport aux estimations. Une des qualités du processus de gestion des erreurs par le filtre de Kalman est de fournir des indications pour choisir l'observation à associer à une prédiction [FY88].

Nous allons revoir rapidement dans cette section les équations théoriques de ce filtre [Sor85].

2.2.1. Définition des modèles

Le filtrage de Kalman demande deux modèles : un vecteur d'état S évoluant avec le temps et un vecteur de mesure X déduit de S . Le vecteur S évolue récursivement avec le temps (passage de l'état k à l'état $k+1$) suivant l'équation :

$$S(k+1) = A(k)S(k) + W(k)$$

A est une matrice caractérisant l'évolution du vecteur S . Le bruit sur l'évolution de l'état $W(k)$ est tel que $E(W(k)) = 0$ et sa variance V_W est connue a priori. Le vecteur X peut être déduit de l'état S par l'équation :

$$X(k) = C(k)S(k) + N(k)$$

C est une matrice permettant de déduire le vecteur de mesure X du vecteur d'état S .

Le bruit de mesure $N(k)$ est tel que $E(N(k)) = 0$ et sa variance V_N est connue a priori.

Par ailleurs, il est important de noter que W et N sont décorrélés.

2.2.2. Calcul de la prédiction

Soit \hat{S} l'état prédit et \hat{X} la mesure prédite, en supposant le modèle parfait, on obtient :

$$\hat{S}(k/k-1) = A(k-1)S(k-1) \quad (1)$$

$$\hat{X}(k/k-1) = C(k-1)\hat{S}(k/k-1) \quad (2)$$

Un intérêt de cette méthode est de pouvoir estimer pour une prédiction, l'erreur possible sur celle-ci.

Soit \tilde{S} l'erreur de prédiction, on peut la définir ainsi :

$$\tilde{S}(k/k-1) = S(k) - \hat{S}(k/k-1)$$

$$\tilde{S}(k/k-1) = A(k-1)S(k-1) + W(k-1) - \hat{S}(k/k-1)$$

Comme nous ne connaissons que la variance de W , on ne calcule que la variance H de \tilde{S} . On montre que :

$$H_{\tilde{S}(k/k-1)} = A(k-1)H_{\tilde{S}(k-1/k-1)}A(k-1)^T + V_w \quad (3)$$

La matrice H de covariance de l'erreur de prédiction va être calculée récursivement. Notons que H est linéairement fonction de V_w . Plus l'erreur de modèle est forte, plus l'erreur de prédiction est forte, ce qui est logique.

2.2.3. Intégration de la mesure

Ayant une mesure réelle $X(k)$, il faut mettre à jour l'état S :

$$S(k) = \hat{S}(k/k-1) + G(k)(X(k) - \hat{X}(k/k-1)) \quad (4)$$

où G , le gain du filtre, est donné par :

$$G(k) = H_{\tilde{S}(k/k-1)}C^T(k)(C(k)H_{\tilde{S}(k/k-1)}C^T(k) + V_n)^{-1} \quad (5)$$

Ce gain pondère le présent par rapport au passé. On remarque que G est quasi inversement proportionnel à V_n . Donc, plus l'erreur de mesure sera forte, plus G aura tendance à diminuer, donnant davantage de poids à la prédiction et diminuant alors la sensibilité aux erreurs. Ceci est logique puisque les mesures n'ont pas été considérées comme fiables.

Enfin, il faut remettre à jour la matrice de covariance de l'erreur de prédiction, pour la prédiction suivante :

$$H_{\tilde{S}(k/k)} = (1 - G(k)C(k))H_{\tilde{S}(k/k-1)} \quad (6)$$

(1), (2), (3), (4), (5) et (6) constituent les équations du filtre de Kalman.

2.3. APPLICATION DU FILTRAGE DE KALMAN À LA DÉTECTION DE SEGMENTS

2.3.1. Choix des matrices A et C pour les équations (1) et (2)

On cherche à suivre des empan dont l'évolution de position est linéaire (modèle "vitesse constante"), et l'évolution d'épaisseur quasi-nulle. Le modèle d'état peut se caractériser par :

$$\begin{bmatrix} Y(k+1) \\ \dot{Y}(k+1) \\ E(k+1) \\ S(k-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ & & & A(k) \end{bmatrix} \begin{bmatrix} Y(k) \\ \dot{Y}(k) \\ E(k) \\ S(k) \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W(k) \end{bmatrix} \quad (7)$$

- Y représente la position (l'ordonnée) du milieu de l'empan suivi, et W_1 le bruit associé.
- \dot{Y} représente la pente de la droite porteuse, et W_2 le bruit associé.
- E représente l'épaisseur du segment, et W_3 le bruit associé.

Le modèle suppose l'indépendance entre l'épaisseur et la position.

Pour diminuer la complexité des calculs, on peut décomposer le filtre en deux parties, l'une ayant pour état le vecteur (Y, \dot{Y}) et l'autre le scalaire E .

Le modèle de mesure du filtre est simple :

$$\begin{bmatrix} X_1(k) \\ X_2(k) \\ X(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ & & & C(k) \end{bmatrix} \begin{bmatrix} Y(k) \\ \dot{Y}(k) \\ E(k) \\ S(k) \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \\ N(k) \end{bmatrix} \quad (8)$$

- X_1 représente la position (l'ordonnée) dans l'espace de mesure, et N_1 le bruit de mesure associé.
- X_2 représente l'épaisseur dans l'espace de mesure, et N_2 le bruit de mesure associé.

Il faut fixer a priori les variances V_{N_1} et V_{N_2} sur les bruits de mesure.

2.3.2. Choix des paramètres W de bruit sur l'état

W_3 a une variance V_E fixée a priori.

W_1 et W_2 peuvent être considérés comme représentant les variations du modèle vitesse, dues aux accélérations (trajectoire curviligne). On utilise un développement du second ordre pour donner des valeurs pertinentes à ces paramètres :

$$\begin{pmatrix} Y(k+1) = Y(k) + \dot{Y}(k) + \frac{1}{2}\ddot{Y}(k) \\ \dot{Y}(k+1) = \dot{Y}(k) + \ddot{Y}(k) \end{pmatrix} \\ \Rightarrow W_1 = \frac{1}{2}\ddot{Y}(k) \text{ et } W_2 = \ddot{Y}(k)$$

dont la variance V_W peut être calculée :

$V_W = V_{\hat{Y}} \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$ avec $V_{\hat{Y}}$, la variance de l'accélération qui devra être donnée a priori.

Ce modèle d'erreur demanderait cependant une étude plus approfondie.

2.3.3. Equations de la prédiction de la mesure

En appliquant A et C aux équations (1) et (2), on obtient les prédictions de mesure suivantes :

$$\hat{Y}(k) = Y(k-1) + \dot{Y}(k-1)$$

$$\hat{E}(k) = E(k-1)$$

On peut aussi calculer les matrices de covariance de l'erreur :

- pour la position,

$$H_{\hat{Y}(k-1/k-1)} = \begin{bmatrix} \sigma_{Y^2} & \sigma_{Y\dot{Y}} \\ \sigma_{Y\dot{Y}} & \sigma_{\dot{Y}^2} \end{bmatrix}$$

$$\Rightarrow$$

$$H_{\hat{Y}(k/k-1)} = \begin{bmatrix} \sigma_{Y^2} + 2\sigma_{Y\dot{Y}} + \sigma_{\dot{Y}^2} & \sigma_{Y\dot{Y}} + \sigma_{\dot{Y}^2} \\ \sigma_{Y\dot{Y}} + \sigma_{\dot{Y}^2} & \sigma_{\dot{Y}^2} \end{bmatrix}$$

$$+ V_{\hat{Y}} \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

- pour l'épaisseur,

$$H_{\hat{E}(k/k-1)} = H_{\hat{E}(k-1/k-1)} + V_E$$

2.3.4. Mise en correspondance et terminaison

Il s'agit de retrouver dans l'image, l'observation réelle correspondant à l'empan dont nous suivons l'évolution. La mise en correspondance se fait sur des critères d'épaisseur (la différence des épaisseurs entre prédiction et observation devra être inférieure à $3\sigma_E$) et de position (la distance entre les milieux des empan devra être inférieure à $3\sigma_Y$). $3\sigma_Y$ et $3\sigma_E$ seront évidemment donnés par les matrices H comme nous l'avons vu dans le paragraphe précédent. Le coefficient 3 permet, en supposant la loi normale, d'être sûr que l'appariement est correct dans 99% des cas.

Si la mise en correspondance échoue, l'étape de filtrage est sautée (il n'y a pas d'observation à intégrer) mais la prédiction sera reprise dans les pas de balayage suivants pour essayer de prolonger au maximum le segment suivi.

Deux cas d'échec de la mise en correspondance sont à considérer :

- l'empan observé est trop épais. Il y a vraisemblablement une superposition d'objets et le segment pourra se retrouver plus tard.
- l'empan observé est trop fin ou même n'existe pas, le segment devrait alors s'arrêter. Mais, pour permettre des discontinuités, on utilise encore la prédiction pendant quelques pas avant de décider de la fin du segment.

2.3.5. Partage d'empan

Il est possible également qu'un empan observé appartienne à plusieurs hypothèses de présence de segment. Deux cas sont à distinguer :

- croisement de deux segments. Il n'y a pas de problème à noter l'appartenance d'un empan à plusieurs segments.
- fusion de deux segments (une fin de liaison fusionnant avec une ligne de portée, ...) impliquant qu'un segment sera suivi par deux hypothèses. Il faut en conséquence arrêter une des hypothèses pour éviter que plusieurs segments ne représentent la même information. On choisit de favoriser la détection des lignes les plus longues.

Pour distinguer ces deux cas, un test sur la pente des hypothèses est réalisable lors de la détection d'un partage d'empan. Lorsque les valeurs sont voisines, vis-à-vis d'une erreur déduite de la matrice de covariance, il y a fusion.

2.3.6. Initialisation

Pour tout nouvel empan qui n'a pas encore été affecté à une hypothèse déjà existante, on crée une nouvelle hypothèse de segment et on initialise l'état et l'erreur du filtre associé (position et épaisseur) :

Pour la position :

- $Y(0)$ = milieu de l'empan générateur de l'hypothèse
- $\dot{Y}(0) = 0$ (on suppose a priori que le segment à suivre est parallèle au sens de balayage)

La valeur initiale de H correspond à une erreur de position fixée à la valeur de l'épaisseur d'une ligne de portée. Cette épaisseur sera calculée lors d'un prétraitement décrit dans la section 3.1 ci-dessous. Il est intéressant de prendre une erreur initiale assez grande pour permettre en début de prédiction, de s'adapter à la pente inconnue du segment.

Pour l'épaisseur :

- $E(0)$ = épaisseur de l'empan générateur de l'hypothèse
- $H_E(0)$ est fixée à deux pixels (justifiés par l'erreur d'échantillonnage de plus ou moins un pixel sur les extrémités d'un empan)

3. Traitement des segments horizontaux

3.1. SEGMENTATION

L'utilisation de filtres de Kalman avec un balayage colonne permet de détecter les segments à tendance horizontale. La figure 8 est le résultat de ce traitement appliqué à l'image de la figure 1.

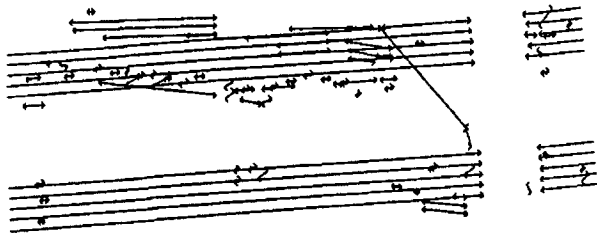


Figure 8. – Segments horizontaux détectés dans l'image de la figure 1

Comme les filtres travaillent sur des empan, un pré-traitement est nécessaire pour convertir les images représentées au départ sous forme de matrices de pixels binaires. Au cours de ce pré-traitement, l'épaisseur des lignes de portées (resp. des interlignes) est calculée comme étant l'épaisseur la plus fréquente des empan noirs (resp. blancs). Ces deux informations, comme Fujinaga [Fuj88] l'a fait remarquer, donnent une idée de l'échelle du document et permettent de normaliser la taille de ses symboles. Elles seront utiles pour classer les segments détectés. Sur la figure 8, deux constatations sont intéressantes : tous les objets linéaires horizontaux (lignes de portées et barres de croche) ont été détectés ainsi que les liaisons horizontales. Il s'agit là d'un effet de bord dû à la tolérance de notre détecteur, car les liaisons ne correspondent pas vraiment à la définition des segments recherchés. Aussi, ne s'étonnera-t-on pas de trouver quelques liaisons interrompues en raison d'une courbure trop importante. En relâchant les paramètres d'erreur des filtres, il est possible de mieux s'accomoder de ces grandes courbures, au détriment évidemment de la qualité du suivi des objets rectilignes.

Dans l'image de la figure 1, nous avons fixé le bruit sur la mesure à 2 pixels (cf justification de l'initialisation de H_E). Pour les bruits sur les états, on fixe expérimentalement une valeur :

- très petite pour la position (écart type de l'erreur de position à 0,017 pixels) assurant que les segments suivis sont bien rectilignes, en accordant une grande confiance à la pente prédite.
- moyennement petite pour l'épaisseur (écart type de l'erreur d'épaisseur à 0,3 pixels) assurant de tolérer de petites variations d'épaisseur, comme pour les liaisons par exemple qui sont légèrement plus épaisses en leur milieu qu'en leurs extrémités.

3.2. CLASSIFICATION

Les objets à classer sont les segments représentant les signes musicaux suivants : les lignes de portée, les liaisons et les barres de croche. Grâce à des critères structurels simples (courbure, épaisseur, parallélisme entre segments, connexité des segments à leurs extrémités), il est possible de les isoler facilement des autres segments trouvés.

Les lignes de portée sont des segments d'épaisseur connue dès le pré-traitement, parallèles entre elles, espacées d'une distance

égale à un interligne, et regroupées par cinq pour constituer une portée. Tous les segments détectés qui répondent à ces caractéristiques sont étiquetés comme des lignes de portée probables.

De même, les barres de croche sont assimilables à des segments plus épais que les lignes de portée et parfois parallèles entre elles. Expérimentalement, on étiquette ainsi les segments au moins deux fois plus épais que les lignes de portée. Même si nous ne cherchons pas ici à quantifier la confiance de l'étiquetage, on note que cette confiance peut croître si les probables barres de croche sont parallèles entre elles.

Enfin, les liaisons sont des segments courbes qui peuvent être découpés en plusieurs tronçons possédant des extrémités communes. La courbure est mesurée grâce à la différence de pente entre l'origine et la fin du segment.

La figure 9 montre les classements obtenus sur l'image de test [figure 1].

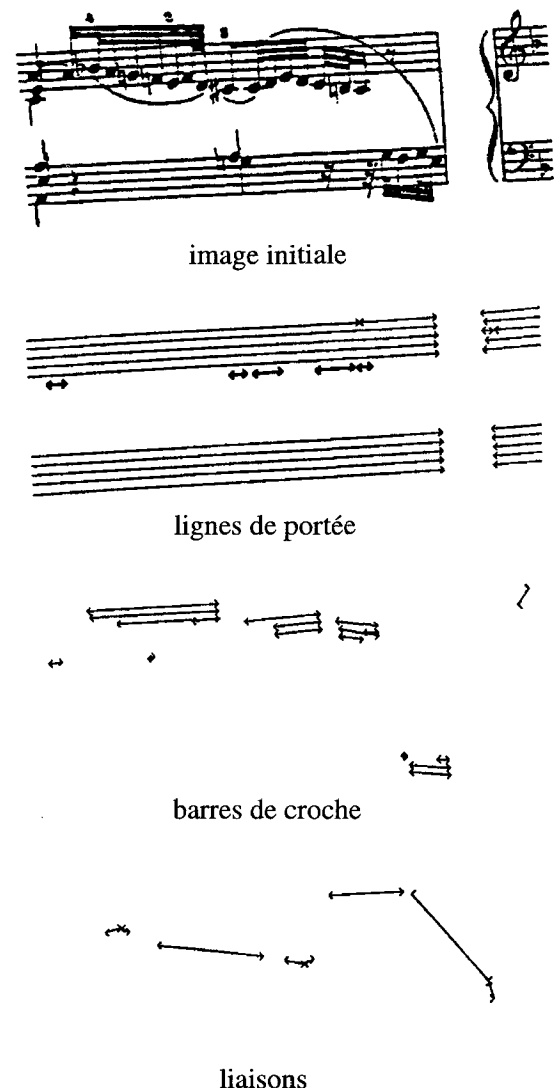


Figure 9. – Classification des segments horizontaux détectés

On constate un taux de reconnaissance de 100% sur les lignes de portée normales. Quelques lignes supplémentaires ont aussi été détectées (81%) même si ce n'était pas l'objectif principal. Pour les barres de croche, nous obtenons quasiment 100% de reconnaissance, mais avec quelques segments de symboles, tels des bouts de clé assez épais, qui ont été confondus. Ces erreurs ne sont pas graves dans la mesure où elles devraient pouvoir être corrigées avec la connaissance d'un contexte, comme la présence de queues de note aux extrémités du segment.

Pour les liaisons, le taux de reconnaissance est bien plus faible. Il est, dans l'image présentée, de 80% et en moyenne sur toutes nos images de 87%. Ce résultat est logique étant donnée l'inadéquation de ces objets avec notre modèle.

3.3. ELIMINATION DES SYMBOLES DE LA PREMIÈRE COUCHE

Avant d'examiner les autres couches, il faut enlever de l'image les pixels noirs déjà étiquetés afin de ne pas les retraiter. Cependant, l'élimination des empan des symboles étiquetés doit se faire "intelligemment" en ôtant seulement les empan appartenant exclusivement à ces symboles. Ainsi, les empan larges traversés par ces symboles, comme les queues de notes recouvertes par des lignes de portées, ne seront pas découpés.

Techniquement, cette opération d'élimination est simple, car la méthode de regroupement des empan noirs assure qu'un empan affecté à un segment a peu de chance d'appartenir à un autre objet. Néanmoins, cette confusion reste possible lorsque les objets sont tangents et se recouvrent parfaitement. C'est le cas des clés de fa tangentes aux lignes de portée. Ceci est un problème classique [Mar92] et on ne cherchera pas à le résoudre à ce niveau. De même, si des segments "bavent", ces bavures peuvent être considérées comme un objet recouvrant le segment, et ne pas être éliminées. Ceci se produit notamment sur la figure 1 où des barres de croches très proches fusionnent malheureusement pour former une barre plus épaisse. Il y a ici une amélioration à apporter au niveau du contrôle des recouvrements de segment. Il faudrait pouvoir interpréter les fusions à partir des segments suivis jusqu'alors afin de redécouper les "bavures".

La figure 10 montre la figure 1 épurée des symboles de la couche 1.

4. Traitement des segments verticaux

4.1. SEGMENTATION

Après l'opération d'élimination, les segments à tendance verticale sont détectés par un balayage ligne. Tous les segments correspondant aux queues de notes et aux barres de mesures sont détectés [figure 11].

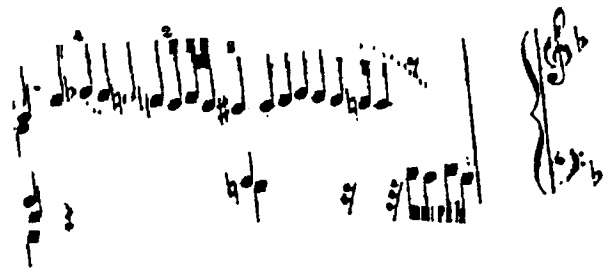


Figure 10. – Partition sans les symboles déjà reconnus

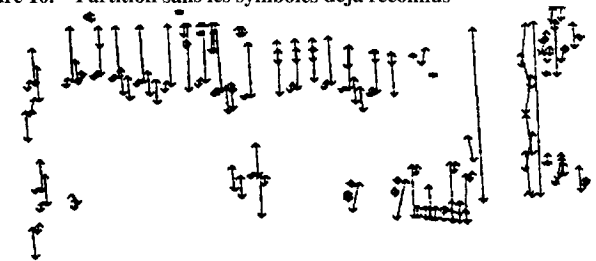


Figure 11. – Segments verticaux détectés

4.2. CLASSIFICATION

Sur des critères structurels simples, tels que l'épaisseur et la longueur, qui se déduisent de la taille des objets de la première couche, il est possible de distinguer les queues et les barres d'une bonne partie des autres objets. Par contre, il est impossible de faire une distinction entre ces deux types d'objets. Une bonne décision exige de tenir compte d'un contexte. Dans un premier temps, ce contexte peut être obtenu en faisant coopérer les objets des couches 1 et 2 : les barres de mesure doivent commencer sur une première ligne de portée et finir sur une dernière; les segments verticaux connexes à des barres de croches sont des queues de notes.

En combinant les barres de croches et les segments verticaux, nous avons ainsi reconstruit à 99.8% les groupes de notes [figure 12]. Les 0.2% oubliés sont dus à la mauvaise qualité de certains documents. Il faut noter deux autres problèmes. Les queues des notes n'appartenant pas à un groupe peuvent toujours être confondues avec des barres de mesures. Et surtout, des clés, qui possèdent une barre épaisse reconnue à tort comme barre

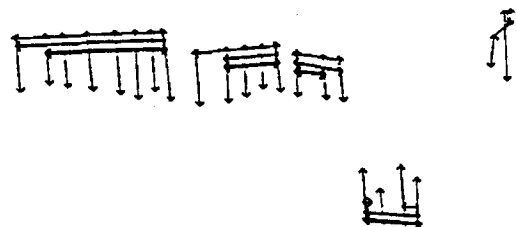


Figure 12. – Groupes de notes reconstruits

de croche, génèrent, du même coup, des groupes de notes puisque dans ces clés sont aussi détectés des segments verticaux.

Ce contexte n'est donc pas suffisant pour étiqueter correctement les symboles de la couche 2. Il faudra attendre le traitement de la couche 3 pour l'enrichir et mieux étiqueter ces symboles.

5. Traitement des têtes de notes noires

5.1. SEGMENTATION

Après avoir éliminé les segments verticaux potentiellement symboles de la couche 2 [figure 13], les têtes noires, qui sont presque des segments aussi longs que larges, se détectent aussi bien par un balayage ligne que par un balayage colonne. Avec un balayage ligne, tous les segments correspondant aux têtes de notes ont été détectés. Pourtant, lorsque deux têtes sont proches et sont l'une au-dessus de l'autre, une seule "longue tête" est trouvée (ceci à cause de la tolérance des filtres de Kalman). Pour découper cette "tête double", on utilise un critère de longueur déduit de la taille des symboles de la couche 1 : toute tête plus grande qu'un interligne doit être découpée suivant la taille de l'interligne. Néanmoins, une solution plus sûre consisterait à effectuer en plus un balayage colonne et à mettre en correspondance les deux lots de segments obtenus.

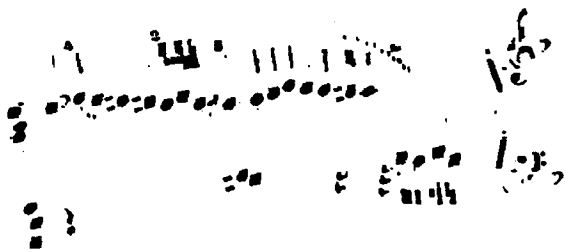


Figure 13. – Elimination des symboles de la couche 2

5.2. CLASSIFICATION

Un simple critère d'épaisseur (la taille de l'interligne) permet de classer toutes les têtes noires avec moins de 2% d'erreur sur les objets. Ces erreurs devraient pouvoir être corrigées en considérant le contexte.

En effet, avec la coopération des symboles des trois couches et en rajoutant la règle "Toute tête noire doit être connexe à une queue", on arrive à reconstruire les groupes de notes [figure 13] et à éliminer les problèmes de confusion et d'erreur liés à ces symboles. Aucun autre objet n'est alors confondu avec une tête et le problème des clés reconnues comme groupes de croches ne se pose plus vu l'absence de tête noire dans ce cas.



Figure 14. – Groupes de notes avec les têtes

Globalement, 96% des groupes de croches et des notes noires isolées sont correctement reconstruits. Les 4% d'omissions viennent de têtes qui n'ont pas été raccrochées à des queues.

6. Conclusion et Perspectives

L'utilisation d'un filtre de Kalman a permis d'obtenir un détecteur de segments intégrant les caractéristiques des images (biais, courbures, bruit, ...). En appliquant méthodiquement ce seul et unique outil aux trois premières couches structurelles d'une partition, on segmente et on interprète un très grand nombre de symboles avec un taux moyen de reconnaissance des objets visés de 94,5%. Maintenant, il nous reste à traiter la dernière couche de la partition dont les symboles se trouvent segmentés mais non interprétés. Des techniques du style "pattern matching" ou des réseaux neuronaux semblent envisageables.

Par ailleurs, la méthode de classification mise en oeuvre est très primaire et utilise un contexte assez pauvre. Aussi, avons-nous développé dans l'équipe une grammaire pour contrôler les processus de reconnaissance [CC95], par rapport à un contexte plus important. Celui-ci permet de guider et de fiabiliser les divers traitements de segmentation et de classification, et par voie de conséquence, d'améliorer la qualité de la reconnaissance. Nous pensons que ce détecteur de segments peut s'étendre à d'autres styles d'édition (édition moderne [Rea79], partition manuscrite, ...) ainsi qu'à d'autres documents techniques (dessins industriels).

De même, on peut extraire des segments dans des images en niveaux de gris en ajoutant au filtre de Kalman la gestion d'un paramètre de luminosité. Cela a été fait dans [LCQ95] pour extraire les lignes de base de chèques, et dans [Ver93] pour suivre les routes sur des images de photo aérienne.

BIBLIOGRAPHIE

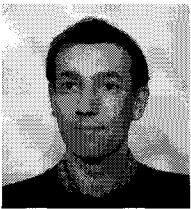
- [BB92a] A. Belaïd and Y. Belaïd. *Reconnaissance des formes*. InterEditions, Paris, 1992.
- [BB92b] D. Blostein and H. Baird. A critical survey of music image analysis. In Springer-Verlag, editor, *Structured Document Image Analysis*, pages 405–434. Eds. H.S. Baird, H. Bunke, K. Yamamoto, 1992.

- [BT92] A. Belaïd and K. Tombre. Analyse de documents : de l'image à la sémantique. In *Colloque National sur l'Écrit et le Document (CNED)*, pages 3–29, Nancy, France, 1992.
- [Car89] N. P. Carter. *Automatic Recognition of Printed Music in the Context of Electronic Publishing*. PhD thesis, University of Surrey, February 1989.
- [CC95] B. Coïasnon and J. Camillerapp. A way to separate knowledge from program in structured document analysis : Application to optical music recognition. In *ICDAR, International Conference on Document Analysis and Recognition*, volume 2, pages 1092–1097, Montreal, Canada, August 1995.
- [Fuj88] I. Fujinaga. Optical music recognition using projections. Master's thesis, McGill University, Faculty of Music, Montreal, Canada, 1988.
- [FY88] T.E. Fortmann and Bar-Shalom Y. Tracking and data association. *Mathematics in Science and Engineering*, 179, 1988.
- [KI90] H. Kato and S. Inokuchi. The recognition system for printed piano music using musical knowledge and constraints. In *Proceedings of the International Association for Pattern Recognition Workshop on Syntactic and Structural Pattern Recognition*, pages 231–248, Murray Hill, NJ, June 1990.
- [LC91] Ivan Leplumey and Jean Camillerapp. Comparison of region labelling for musical scores. In *ICDAR, International Conference on Document Analysis and Recognition*, volume 2, pages 674–682, Saint-Malo, France, September 1991.
- [LCQ95] I. Leplumey, J. Camillerapp, and C. Quéguiner. Kalman filter contributions towards document segmentation. In *ICDAR, International Conference on Document Analysis and Recognition*, volume 2, pages 765–769, Montreal, Canada, August 1995.
- [Mar92] Philippe Martin. *Réseaux de neurones artificiels : Application à la reconnaissance optique de partitions musicales*. PhD thesis, IMAG, Grenoble, France, 1992.
- [Rea79] G. Read. *Music Notation*. Taplinger Publishing Company, New York, 1979.
- [Sor85] H. Sorenson. *Kalman filtering : Theory and Application*. IEEE Press, 1985.
- [Ver93] J.-P. Veran. Suivi de routes dans une image aérienne par filtrage de kalman. Technical report, Ecole Nationale Supérieure des Télécommunications, Département Images, March 1993.

Manuscrit reçu le 2 février 1995.

LES AUTEURS

Vincent POULAIN D'ANDECY



Vincent Poulain d'Andecy a été diplômé en 1993 ingénieur en informatique de l'Institut National des Sciences Appliquées de Rennes. Il a obtenu la même année le D.E.A. d'informatique de l'Université de Rennes I dans la filière Imagerie Numérique. Depuis, il a intégré l'équipe d'Etude et de Recherche de la société ITESOF, à Nîmes, où il s'intéresse plus particulièrement au problème de la lecture optique des formulaires, des documents comptables et des documents administratifs.

Jean CAMILLERAPP



Jean Camillerapp est professeur à l'Institut National des Sciences Appliquées de Rennes depuis 1976. Il est directeur du département d'informatique de l'INSA et responsable du projet IMADOC de l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA). Actuellement, il effectue ses recherches sur l'amélioration de l'extraction du tracé grâce à des documents numérisés en niveaux de gris, ainsi que sur l'introduction des connaissances a priori pour guider le processus de reconnaissance.

Ivan LEPLUMEY



Ivan Leplumey est maître de conférences à l'Institut National des Sciences Appliquées de Rennes depuis 1988. Il fait partie de l'équipe IMADOC de l'IRISA de Rennes où il effectue des travaux sur la reconnaissance de documents (partitions musicales, chèques, braille...).