

Traitement de vidéos de polysomnographie pour l'estimation de l'état des yeux chez le nouveau-né prématuré

Fabienne PORÉE^{1,2}, Antoine SIMON^{1,2}, Sandie CABON^{1,2}, Anaïs COROLLEUR^{1,3}, Nicolas NARDI^{1,2,4}, Patrick PLADYS^{1,2,4},
Guy CARRAULT^{1,2,3}

¹Université de Rennes 1, LTSI, Rennes, F-35000, France

²INSERM, U1099, Rennes, F-35000, France

³INSERM, CIC-1414, Rennes, F-35000, France

⁴CHU Rennes, Pôle de pédiatrie médico-chirurgicale et génétique clinique, Rennes, F-35000, France

Fabienne.Poree@univ-rennes1.fr, Antoine.Simon@univ-rennes1.fr, cabon.sandie@gmail.com,
a0coroll@enib.fr, Nicolas.Nardi@chu-rennes.fr, Patrick.Pladys@chu-rennes.fr,
Guy.Carrault@univ-rennes1.fr

Résumé – Cette étude a pour objectif une meilleure surveillance des nouveau-nés prématurés placés en Unités de Soins Intensifs par l'utilisation de la vidéo. Un algorithme d'estimation de l'état des yeux du bébé, combinant une étape de suivi avec une étape de segmentation et de caractérisation, est proposé. Les résultats obtenus sur une base de données de 10 vidéos fournissent des valeurs de concordance entre l'algorithme et un scoring manuel supérieures à 95%. Une mise en correspondance de l'état des yeux avec les stades du sommeil permet de valider l'approche.

Abstract – The study presented here aims to better monitor premature infants placed in intensive care units through the use of video. An algorithm estimating the eye state of the baby is proposed, combining tracking with segmentation and characterization steps. Results obtained on 10 videos lead to concordance values between the algorithm and a manual scoring above 95%. A representation of the eye state as a function of sleep stages validates the approach.

1 Introduction

La polysomnographie est une méthode classiquement utilisée en hôpital pour la surveillance des patients présentant des troubles neurologiques, respiratoires et/ou du sommeil. Elle est basée sur l'acquisition de signaux électrophysiologiques (EEG, EMG, EOG, respiration, ECG, actimétrie etc.) sur de longues durées dont l'analyse (ou scoring) aboutit à un hypnogramme, représentant l'enchaînement des stades de sommeil [1]. Chez les nouveau-nés et les nourrissons, les stades du sommeil sont difficiles à identifier à partir des signaux électrophysiologiques seuls. En revanche, ils sont liés au comportement de l'enfant [2].

L'étude que nous présentons ici a pour objectif une meilleure surveillance des nouveau-nés prématurés placés en Unités de Soins Intensifs (USI) selon le principe NIDCAP (*Newborn Individualized Developmental Care and Assessment Program*). Dans cet esprit, la vidéo a l'avantage d'être sans contact et de pouvoir fonctionner pendant des durées illimitées, contrairement aux capteurs classiques. Le principe est d'appliquer aux vidéos des traitements automatiques dans le but d'extraire différentes informations, pouvant traduire l'état du bébé concernant son sommeil mais aussi son état d'agitation, une certaine souffrance... Le degré d'ouverture de l'oeil étant la première étape de cette caractérisation, ce travail est focalisé sur l'esti-

mation automatique de cet indicateur à partir de vidéos.

S'il existe une littérature abondante sur le traitement automatique de la vidéo pour la détection de mouvements pour l'analyse du sommeil [3], il n'y en a pas, à notre connaissance, sur la caractérisation de l'état des yeux en polysomnographie. En effet, de très nombreux travaux ont porté sur la détection des yeux dans des images ou des vidéos, mais la plupart considèrent des environnements d'acquisition contraints, soit par des dispositifs d'acquisition particuliers (éclairage infra-rouge [4]), soit des positionnements spécifiques avec notamment des visages observés à une faible distance et de face [5] ou avec un angle limité [6, 7], ou encore une détection limitée à des yeux ouverts [8, 9]. Par ailleurs, de nombreuses approches reposent sur des méthodes de classification avec apprentissage sur de grandes bases de données [9, 10]. Cependant, dans l'application considérée ici, ces différentes contraintes ne sont généralement pas respectées : les vidéos sont acquises dans la pénombre, le visage du nouveau-né est généralement tourné et partiellement caché, le nouveau-né effectue des mouvements avec des occlusions potentielles... Enfin nous ne disposons pas d'une base de données suffisante pour un apprentissage. Pour répondre à ces difficultés, la méthode que nous proposons se focalise sur un seul oeil et combine une étape de suivi d'une région de référence avec une étape de segmentation et de caractérisation du

degré d'ouverture de l'oeil. Cette information pourra être ensuite ramenée à deux états ('Ouverts' ou 'Fermés') ou conservée telle quelle et ainsi fournir une information sur le degré d'éveil du bébé.

2 Méthode

2.1 Description de l'algorithme

La méthode proposée est décrite par la figure 1 et est composée de trois étapes principales : le suivi d'une région d'intérêt, dite "repère", la segmentation de l'oeil et enfin sa caractérisation par son aire. En effet, comme le nouveau-né est susceptible d'effectuer des mouvements, effectuer un suivi de l'oeil est nécessaire. Cependant, l'oeil changeant d'apparence au cours de l'enregistrement en fonction de son état, le suivi est réalisé non pas sur la région de l'oeil mais sur une région repère du visage, notée ROI_{rep} , laissée au choix de l'utilisateur et supposée conserver son apparence tout au long de la vidéo (ex : zone du nez ou de l'oreille). Le suivi de cette région repère permet de localiser la région de l'oeil (ROI_{oeil}), de le segmenter puis de le caractériser par son aire. Ces étapes sont décrites ci-dessous.

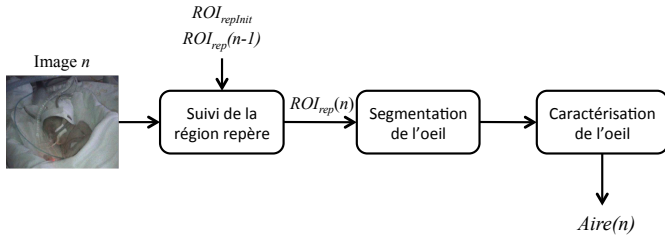


FIG. 1 – Description de l'algorithme.

2.1.1 Suivi de la région repère

Après conversion des images en niveaux de gris (les vidéos sont acquises dans la pénombre et donc l'information de couleur peut être ignorée), le suivi de la région repère ROI_{rep} est effectué par recherche de correspondances (*template matching*) entre la région repère et la nouvelle image. La métrique utilisée est la somme des différences au carré (SSD) entre les intensités des pixels de la région repère et celles de la région testée dans l'image. À partir d'une initialisation manuelle de ROI_{rep} sur la première image de la vidéo ($ROI_{repInit}$), le suivi est effectué autour de la position de la région sur l'image précédente (avec une région de recherche de largeur 64 pixels de chaque côté). Pour chaque image de la vidéo, la région repère mise en correspondance est tout d'abord celle sélectionnée initialement par l'utilisateur ($ROI_{repInit}$). La valeur minimale de la métrique (SSD_{min}) obtenue est comparée à deux seuils ($S_1 = 0.02$ et $S_2 = 0.04$) :

- $SSD_{min} < S_1$: ROI_{rep} est bien détectée dans l'image ($ROI_{rep}(n)$);
- $S_1 \leq SSD_{min} < S_2$: la position de ROI_{rep} doit être affinée ; une seconde mise en correspondance est effectuée avec $ROI_{rep}(n-1)$ comme région repère ;

- $SSD_{min} > S_2$: ROI_{rep} n'est pas détectée, le suivi est interrompu pour une interaction avec l'utilisateur (cf. section 2.1.4).

2.1.2 Segmentation de l'oeil

Une fois que $ROI_{rep}(n)$ est identifiée, la région de l'oeil, notée $ROI_{oeil}(n)$, est positionnée grâce à la position relative δ_{ROI} des deux régions. De façon à initialiser δ_{ROI} , $ROI_{oeil}(1)$ est aussi sélectionnée par l'utilisateur sur la première image.

La segmentation de l'oeil est alors réalisée en appliquant plusieurs traitements successifs sur $ROI_{oeil}(n)$:

- Filtrage morphologique de type *Black Hat* avec un élément structurant de forme ellipsoïdale (dimension 15x15) pour identifier les régions plus sombres que leur voisinage (dans les vidéos considérées, un oeil ouvert correspond à des intensités faibles) ;
- Seuillage avec un seuil T (de valeur choisie interactivement entre 15 et 25) ;
- Filtrage morphologique (ouverture avec un élément structurant ellipsoïdal de dimension 5x5) pour supprimer les petites régions ;
- Détection de contours de façon à extraire la ou les structure(s) restantes.

2.1.3 Caractérisation de l'état de l'oeil

L'oeil est caractérisé par son aire $Aire(n)$ en fonction du nombre de contours k détectés à l'étape de segmentation :

- $k = 0$: l'oeil est considéré comme fermé et $Aire(n) = 0$;
- $k = 1$ ou $k = 2$: l'oeil est considéré comme ouvert, $Aire(n)$ est la somme des aires des k contours et δ_{ROI} est mis à jour en fonction du centre de l'oeil ;
- $k > 2$: l'oeil est considéré comme n'étant pas détecté ; le suivi est interrompu pour une interaction avec l'utilisateur.

Une fois l'ensemble de la vidéo traitée, un lissage est appliqué sur les valeurs d'aires afin de limiter les changements d'états incohérents. Puisqu'il a été observé qu'un clignement d'oeil s'observe au minimum sur 5 images successives, un filtrage médian de largeur 5 est appliqué.

2.1.4 Interruption pour interaction avec l'utilisateur

En cas d'incertitude sur le suivi de la région de référence ou sur la segmentation de l'oeil, l'utilisateur doit soit resélectionner les régions d'intérêt soit, si les régions sont cachées, avancer dans la vidéo jusqu'à ce qu'elles soient à nouveau visibles.

2.2 Evaluation des performances

Pour évaluer les performances de la méthode proposée, ses résultats ont été comparés à un scorage manuel, c'est-à-dire après une analyse visuelle des vidéos.

Le scorage manuel consiste à noter les instants où les yeux sont 'Ouverts' (1) ou 'Fermés' (0) (des états intermédiaires

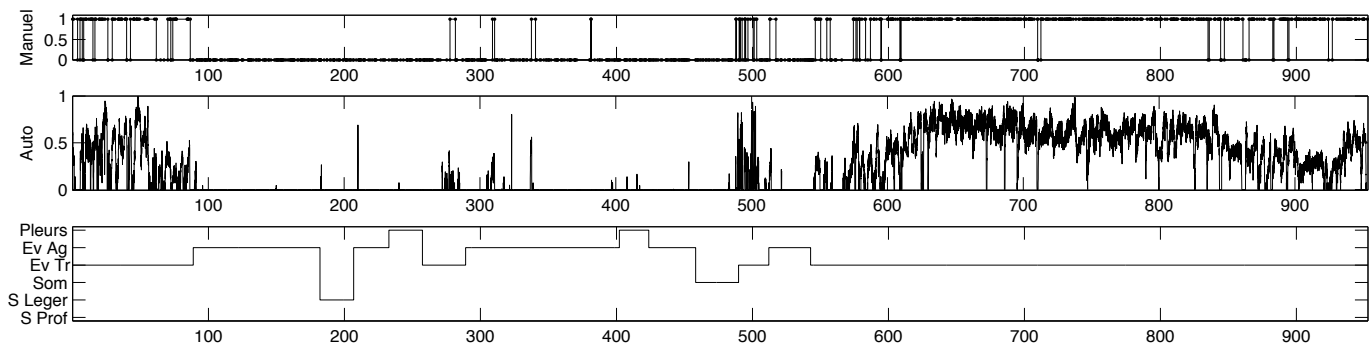


FIG. 2 – Résultats obtenus sur la vidéo 9. Résultat du scoring manuel, résultat du scoring automatique et hypnogramme associé.

étant difficiles à définir). D'autre part, les durées des vidéos et leur fréquence d'acquisition étant respectivement longues et élevées, un scoring image par image n'est pas possible. Une stratégie pourrait consister à réaliser ce scoring en continu, à raison d'une valeur par seconde par exemple, ce qui présenterait l'avantage de fournir une référence sur l'ensemble de l'extrait étudié. En revanche :

- le fait de scorer une valeur par seconde ne fournit pas une résolution temporelle élevée et des événements courts comme des clignements peuvent être manqués ;
- pour pouvoir comparer le scoring automatique (une valeur par image) avec le scoring manuel (une valeur par seconde), un rééchantillonnage d'un des deux scorages devrait être réalisé.

Pour ces différentes raisons, nous avons préféré réaliser un scoring aléatoire d'un pourcentage d'images de chaque vidéo. Nous avons fixé cette valeur à 5%.

L'analyse visuelle ne pouvant aboutir qu'à deux valeurs (0 ou 1), les valeurs d'aires issues de l'algorithme ont été binarisées par une mise à 1 de l'ensemble des valeurs non nulles.

La comparaison des deux scorages a ensuite été réalisée grâce à une table de concordance. En considérant qu'un état ouvert de l'oeil du bébé est l'état positif, et qu'un état fermé correspond à l'état négatif, les valeurs de sensibilité (Se) et de spécificité (Sp) de l'algorithme ont été calculées, ainsi que la concordance globale ($Conc$), c'est-à-dire le pourcentage des valeurs où les deux scorages sont en accord.

3 Résultats

3.1 Protocole d'acquisition

Les vidéos analysées sont issues d'une étude précédente menée au CHU de Rennes, approuvée par le comité de protection des personnes (CPP Ouest 6-598) et conforme aux normes établies par la Déclaration d'Helsinki. La signature d'un consentement éclairé par les parents était obtenue pour chaque enfant.

Les enregistrements ont été réalisés sur des bébés prématurés, d'âge gestationnel inférieur à 32 semaines d'aménorrhées, et ayant entre 4 et 6 jours d'âge postnatal. Une caméra était installée près du lit du bébé, afin d'observer la majorité du corps

du bébé. Les enregistrements ont été réalisés dans la pénombre. La caméra utilisée avait une résolution de 720x576 pixels et une fréquence d'acquisition de 25 images par seconde.

Pour chaque enregistrement, un scoring des stades de sommeil a été réalisé par un expert, en six stades : Sommeil profond (S Prof), Sommeil léger (S Leger), Somnolence (Som), Eveil agité (Ev Ag), Pleurs.

3.2 Traitement des vidéos

Un total de 10 vidéos a été traité. L'algorithme intégrant une étape de détection des états cachés qui génère une pause dans le traitement et éventuellement une réinitialisation, pour chaque vidéo, nous avons sélectionné des intervalles temporels où l'état caché apparaît rarement.

La figure 2 illustre le traitement d'une vidéo. Elle montre que le scoring manuel de 5% des images fournit une représentation satisfaisante de l'enregistrement total et que les scorages automatique et manuel sont cohérents.

Les résultats obtenus sur les 10 vidéos sont présentés dans le tableau 1. On observe que les valeurs de sensibilité et de spécificité sont toujours supérieures à $> 95\%$, sauf pour une vidéo où $Se = 78.57\%$. Dans ce dernier cas, les mouvements du nouveau-né sont rapides et la mise à jour de δ_{ROI} ne s'effectue pas toujours correctement.

4 Discussion et Conclusion

Un objectif étant, à terme, d'utiliser l'information de l'état des yeux du nouveau-né pour caractériser son sommeil, la répartition (moyenne \pm écart-type) des valeurs d'aires fournies par l'algorithme sur l'ensemble des vidéos traitées est représentée, en fonction des stades de sommeil, sur la figure 3.

Les résultats obtenus sont cohérents puisque les yeux sont toujours fermés en sommeil profond et lors des pleurs, et presque toujours en sommeil léger et en éveil agité. En somnolence, les valeurs sont encore de moyenne très basse avec un écart-type qui augmente, correspondant à des ouvertures brèves. C'est lors de l'éveil tranquille qu'il y a le plus de dispersion des valeurs puisque les yeux sont le plus souvent ouverts, mais peuvent aussi être fermés.

TAB. 1 – Caractéristiques des vidéos (Durée, Nombre d’images évaluées, Nombre d’images scorées manuellement ‘Ouvert’ et ‘Fermé’). Performances de l’algorithme (Sensibilité, Spécificité, Concordance globale).

Vidéos	Durée (min’sec)	Nb images évaluées	Nb images ‘Ouvert’	Nb images ‘Fermé’	Se (%)	Sp (%)	Conc (%)
Vidéo 1	17’27	1384	40	1344	98.21	99.68	99.64
Vidéo 2	31’50	2367	343	2024	99.71	99.75	99.75
Vidéo 3	30’58	2357	132	2225	97.73	99.87	99.75
Vidéo 4	27’10	2105	56	2049	98.21	100.00	99.11
Vidéo 5	10’11	831	0	831	-	100.00	100.00
Vidéo 6	24’04	1634	146	1478	99.32	100.00	99.66
Vidéo 7	15’07	1198	28	1170	78.57	99.74	99.25
Vidéo 8	41’02	3064	4	3060	100.00	100.00	100.00
Vidéo 9	16’03	1191	584	607	95.38	97.69	96.56
Vidéo 10	28’22	2156	4	2152	100.00	100.00	100.00

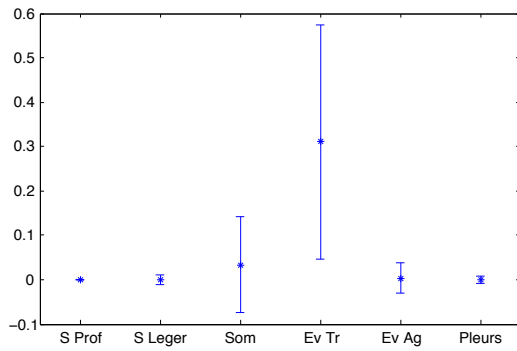


FIG. 3 – Répartition (moyenne \pm écart-type) des valeurs de surfaces de l’oeil en fonction des stades de sommeil.

Dans cet article, nous avons proposé un algorithme pour estimer l’état des yeux d’un bébé prématuré à partir de vidéos de polysomnographie. Les résultats, comparés à un scorage manuel, montrent que les performances de l’algorithme sont élevées. Cependant la méthode doit être améliorée pour permettre une utilisation en temps réel et en continu car certaines difficultés (oeil caché ou bébé très agité) nécessitent une intervention manuelle. Ces améliorations pourront notamment porter sur l’étape de suivi avec intégration d’approches reposant sur des modèles d’apparence.

Enfin, la représentation des valeurs en fonction des stades de sommeil valide l’approche mais montre aussi que la seule information sur les yeux ne peut suffire à estimer les stades de sommeil. Ainsi les perspectives de ce travail portent sur la combinaison de cette information aux mouvements et/ou aux sons du bébés.

Références

[1] L. Curzi-Dascalova and M. Mirmiran, “Manual of methods for recording and analyzing sleep-wakefulness states in preterm and full-term infant,” *Inserm, Paris*, pp. 1–162, 1996.

[2] H. F. Prechtl, “The behavioural states of the newborn infant (a review),” *Brain research*, vol. 76, no. 2, pp. 185–212, 1974.

[3] M. Scatena, S. Dittoni, R. Maviglia, et al., “An integrated video-analysis software system designed for movement detection and sleep analysis. Validation of a tool for the behavioural study of sleep,” *Clinical Neurophysiology*, vol. 123, no. 2, pp. 318–323, 2012.

[4] Z. Zhu and Q. Ji, “Robust real-time eye detection and tracking under variable lighting conditions and various face orientations,” *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 124–154, Apr. 2005.

[5] J. Song, Z. Chi, and J. Liu, “A robust eye detection method using combined binary edge and intensity information,” *Pattern Recognition*, vol. 39, no. 6, pp. 1110–1125, 2006.

[6] Y. Tian, T. Kanade, and J. F. Cohn, “Dual-state parametric eye tracking,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 110–115.

[7] A.M. Bagci, R. Ansari, A. Khokhar, and E. Cetin, “Eye tracking using markov models,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. IEEE, 2004, vol. 3, pp. 818–821.

[8] Y. Li, Xi. Qi, and Y. Wang, “Eye detection by using fuzzy template matching and feature-parameter-based judgement,” *Pattern Recognition Letters*, vol. 22, no. 10, pp. 1111–1124, 2001.

[9] Z. Qian and D. Xu, “Automatic eye detection using intensity filtering and k-means clustering,” *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1633–1640, 2010.

[10] P. Viola and M. J. Jones, “Robust real-time face detection,” *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.