



HAL
open science

Champs aléatoires projetés sur l'espace de Hough pour une détection robuste de la pupille de l'oeil

R Kacete, Renaud Segulier, Jérôme Royan, M Collobert

► **To cite this version:**

R Kacete, Renaud Segulier, Jérôme Royan, M Collobert. Champs aléatoires projetés sur l'espace de Hough pour une détection robuste de la pupille de l'oeil. 28èmes journées de l'Association Française en Informatique Graphique, Nov 2015, Lyon, France. hal-01393613

HAL Id: hal-01393613

<https://hal.science/hal-01393613>

Submitted on 25 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Champs aléatoires projetés sur l'espace de Hough pour une détection robuste de la pupille de l'œil

A. Kacete, R. Séguier, J. Royan et M. Collobert

Institut de Recherche Technologique B-com, Rennes

Résumé

Les yeux sont l'une des caractéristiques les plus saillantes du visage humain. La localisation de la pupille de l'œil permet d'obtenir des informations importantes qui peuvent être utilisées dans plusieurs applications de vision par ordinateur. Plusieurs eye-trackers commercialisés aujourd'hui peuvent estimer avec précision sa position. Cependant les spécifications matérielles très complexes et l'environnement de l'utilisateur contrôlé (résolution suffisante de l'image de l'œil, bonnes conditions d'éclairage, légères variations de la pose de la tête) rendent ces solutions impossibles à utiliser dans un environnement arbitraire. Dans cet article, nous présentons une approche basée sur les arbres de régression aléatoires projetés sur l'espace de Hough pour estimer la position de la pupille. Nous démontrons à partir de plusieurs évaluations sur des bases de données publiques références que notre approche est très robuste aux conditions d'éclairage, à l'échelle, aux mouvements des yeux et aux fortes variations de la pose de la tête, et qu'elle apporte une amélioration significative par rapport aux méthodes existantes de l'état de l'art.

Eyes are one of the most salient feature of the human face. The location of the eye-pupil allows to access to important informations and can be used in several computer vision applications. Several commercial eye-trackers can estimate with a good accuracy the eye-pupil location, but need a complex hardware specifications with a controlled user environment (sufficient eye image resolution, good illumination, small head pose variations) making these solutions impossible to use on an arbitrary environment. In this paper, we present an approach based on Hough random forest, we demonstrate by several evaluations on challenging public datasets that our approach is very robust to illumination, scale, eye movements and high head pose variations and yields a significant improvement regarding to a wide range of state of the art methods

Mots clé : Champs aléatoires, arbres de décision, régression, classification, détection de la pupille, détection du regard

1. Introduction

La localisation de la pupille de l'œil joue un rôle primordial dans plusieurs applications dans le domaine de vision par ordinateur spécialement dans l'analyse du visage. Dans la reconnaissance d'expressions faciales, elle peut remonter une information très parlante sur l'état expressif et cognitif de la personne. En biométrie, l'identification et la reconnaissance de visage sont étroitement liées à la pupille de l'œil. D'une manière générale, les interactions homme-machine IHM se basent souvent sur cette caractéristique.

Plusieurs solutions industrielles sont aujourd'hui commercialisées et donnent de très bonnes localisations de la pupille. Certaines ont des spécifications matérielles très complexes (des caméras embarquées sur un casque ou des lu-

nettes qui filment les yeux) et restent non accessibles au grand public et peuvent être très intrusives. D'autres utilisent un champ de caméras infra-rouge et se basent sur l'information du reflet cornéen pour estimer la position de la pupille d'une manière très précise mais restent très sensibles aux conditions d'éclairage.

Hansen et al. donne un état de l'art très détaillé sur les méthodes de détection et de suivi de la pupille de l'œil dans [HJ10]. Dans ce papier, nous présentons une méthode de localisation de la pupille non intrusive et très robuste à partir d'une simple caméra monoculaire. Dans la section 2, nous présentons un survol sur les dernières méthodes de localisation de la pupille en montrant les limitations de ces approches. Dans la section 2, nous décrivons notre méthode en donnant un aperçu global sur la partie théorique de l'algorithme et de son utilisation dans diverses applications en vision par ordinateur. Dans la section 3 nous présentons nos différentes expérimentations sur des bases de données de références et soulignons l'apport de précision et de robustesse

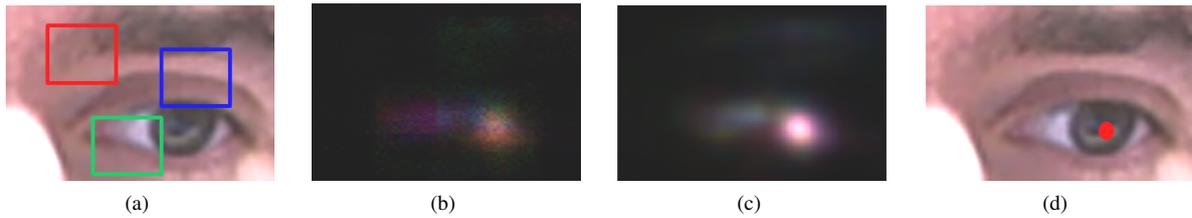


Figure 1: Stratégie de vote par notre forêt. (a) Extraction de patches dans l'image test. (b) Chaque patch vote pour une position probable de la pupille (la couleur indique la contribution de chaque patch). (c) Accumulation de tout les patches de toutes les échelles dans un seul espace de vote. (d) Mapping du maximum global dans l'espace Hough vers l'espace image.

par rapport à l'état de l'art. Nous finissons dans la section ?? par une conclusion générale.

2. État de l'art

Notre travail s'inscrit dans un large champ de méthodes existantes qui essaient de détecter avec une précision suffisante la pupille de l'œil à partir d'une simple caméra monoculaire. [TB11] [VG08] et [MFP*14] décrivent les méthodes les plus pertinentes de l'état de l'art. Le principe général est de détecter le visage en premier en utilisant la méthode Viola Jones [VJ01], puis d'extraire des régions (avec des relations anthropomorphiques) autour des yeux pour estimer finalement la position 2D de la pupille dans le repère image.

Timm et al. [TB11] utilisent l'aspect géométrique de la pupille et définissent une fonction objective basée sur l'image gradient qui atteint son maximum à l'intersection des vecteurs gradients. Cependant la circularité de l'iris est une hypothèse qui n'est plus vérifiée sur des images à fortes variations de la pose de la tête.

Valenti et al. [VG08] utilisent le concept de courbure des isophotes qui représente les courbes reliant les points de même intensité pour construire un espace global de vote. Par la suite, ils étendent leur méthode à l'utilisation des SIFT pour chaque courbe candidate afin obtenir un score de vote final par des comparaisons à des exemples prédéfinis dans une base de donnée. Cette méthode répond aux contraintes de basse résolution et de conditions d'illuminations mais reste sensible aux variations de pose.

Markus et al. [MFP*14] ne prennent pas en compte les aspects géométriques de la pupille. Ils utilisent une approche par apprentissage en entraînant des arbres de régression dégénérés en cascade. Chaque ensemble d'arbres opère à une échelle donnée et constitue l'entrée de l'ensemble suivant jusqu'à la sortie finale définie par le nombre d'échelles fixé au préalable. Leur apprentissage final comporte 100 arbres avec six millions de données d'apprentissage. Cette méthode semble la plus robuste de l'état de l'art, et révèle un très bon comportement aux contraintes standard telle que la basse résolution, les conditions d'éclairages et les orientations de la pose de la tête. Néanmoins elle souffre d'un handicap d'initialisation dû à la conception de leur framework.

Notre solution est relativement proche de la dernière approche. Nous utilisons la même méthode d'apprentissage,

mais en intégrant l'espace de Hough dans l'estimation finale. Notre contribution majeure réside dans l'intégration de l'espace de Hough comme dans [GYR*11] avec la suppression de la dimension échelle. Nous montrons qu'avec un ensemble d'arbres moins grand que [MFP*14] nous obtenons des résultats similaires voire supérieurs à [MFP*14]. Nous montrons aussi que cette méthode peut être étendue à des problématiques de régressions différentes de la nôtre. section Notre approche

2.1. Les champs aléatoires pour la vision par ordinateur

Les champs aléatoires ont été largement utilisés ces dernières années en vision par ordinateurs pour résoudre des problématiques de différentes natures : classification, [GEW06, LF06], régression, [CSRK11, MFP*14] ainsi que l'estimation de densité, [BBC13, SJC08]. Cette large utilisation est la conséquence de leur grande généralisation ainsi que leur rapidité d'exécution. Ils ont été introduit par Breiman et al. [Bre01].

L'idée de base consiste à séparer d'une manière récurrente le problème initial en deux problèmes de complexité réduite. Cette séparation est effectuée à chaque nœud et les nœuds terminaux contiennent les modèles qui approximent le mieux la sortie désirée. Dans notre configuration, comme dans [GYR*11], nous construisons des arbres qui vont simultanément capturer des informations de classification et de régression sur un ensemble de données d'apprentissage, représentées sous la forme de patches $\{\mathcal{P}_i = \{f_i, c_i, y_i\}\}$ (un patch représente une zone d'une taille précise dans l'image) où :

- f_i représente le vecteur caractéristique d'entrée pour un patch de dimension fixe.
- c_i traduit la classe du patches ($c_i = 1$ si $w_i \in \mathcal{I}$, $c_i = 0$ si non)
- y_i représente la variable de sortie à régresser (la distance entre le centre du patch et le celui de la pupille dans l'image ou il a été extrait).

La construction de chaque arbre est un processus supervisé. Tout les nœuds sont déterminés à partir d'un certains nombre de tests binaires définis par :

$$\begin{cases} 1, & \text{si } |f_i(x_1) - f_i(x_2)| \leq \tau \\ 0, & \text{sinon} \end{cases}$$

où $\{f_i(x_1) - f_i(x_2)\}$ représente la différence en intensité entre deux pixels x_1 et x_2 du patch et τ est un seuil aléatoire. [LF06] ont montré que cette caractéristique a un très

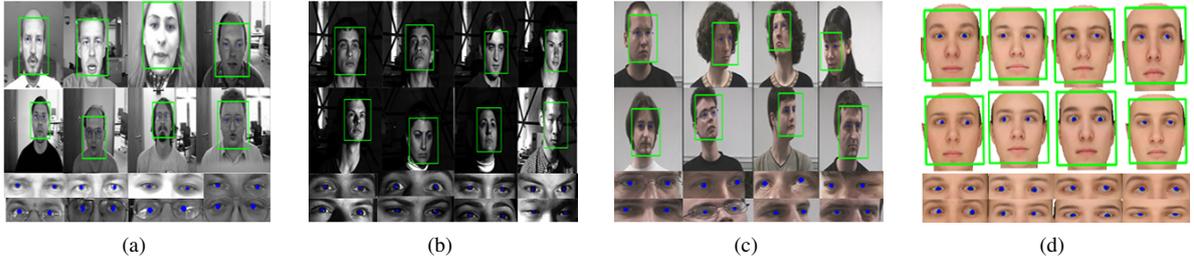


Figure 2: Quelques résultats correctes de l'estimation de la pupille par notre méthode sur certaines bases de données publiques. (a) BioID, (b) Extended-Yale, (c) Pointing'04, (d) Données synthétiques.

grand potentiel de discrimination. La supervision de l'entraînement d'un arbre consiste à fixer à chaque nœud S le test binaire optimal $\theta^* = \{x_1^*, x_2^*, \tau^*\}$ qui maximise la séparation des données et donc la pureté de ses nœuds fils S^L et S^R . Deux fonctions objectives sont utilisées :

– pour la classification :

$$Q_0 = H(S) - \sum_{n \in \{L, R\}} \frac{|S^n|}{|S|} H(S^n) \quad (1)$$

où

$$H(S) = - \sum_{c \in \{0,1\}} p(c/S) \ln(p(c/S)) \quad (2)$$

Q_0 traduit le gain d'information sur un nœud, égal à la différence de l'entropie H du parent S moins la somme pondérée de l'entropie des fils S^L et S^R respectivement.

– pour la régression :

$$Q_1 = \sum_{n \in \{L, R\}} \sum_{i \in S^n} \|y_i - \bar{y}_n\|_2^2 \quad (3)$$

avec y_n ($n \in \{L, R\}$) représentent les moyennes des distances arrivant dans les nœuds fils S^L et S^R respectivement.

2.2. Apprentissage

Nous avons utilisé les deux bases de données [VPSS*13] et [WLSN07]. Nous avons introduit une perturbation en échelle avec $[\pm 30\%, -30\%]$ par rapport à l'original, ainsi qu'une autre perturbation par rapport à la position spatiale de la pupille en x et y $[-25\%, +25\%]$. 10k de données d'apprentissage ont été construites, une extraction de 50 patchs de dimension fixe de (16×16) est effectuée aboutissant à 500k de données d'apprentissage par classe. La profondeur des arbres est fixée à 15. Une stratégie de boosting est appliquée à la construction de certains arbres.

2.3. Test

Pendant le test, nous extrayons deux régions autour des yeux, et construisons une pyramide d'échelle sur chaque région. Pour chaque échelle, nous extrayons un certain nombre de patchs qui constituera l'entrée des arbres entraînés précédemment. Chaque patch est évalué par chaque arbre de la forêt en utilisant les tests binaires optimisés pendant l'apprentissage. Arrivé au nœud terminal, on récupère

les informations sauvegardées dans ce même nœud. Nous projetons toutes les distances de tous les arbres de tous les patchs sur un espace de vote qui est l'espace de Hough généralisé. Afin de sélectionner un maximum global correspondant à la position 2D de la pupille, nous appliquons un fenêtrage de Parzen par échelle. Nous pondérons toutes les échelles avec leurs minima locaux. La figure 2 illustre notre procédé.

3. Résultats expérimentaux

Pour évaluer notre méthode et se comparer à l'état de l'art, la base de donnée BioID contenant 1521 images en niveau de gris est utilisée. BioID est considérée parmi les bases de données les plus défavorables pour les tests de détection de la pupille due à ses variations significatives en terme de conditions d'éclairage, d'échelle, de pose de la tête ainsi que les variations intra-utilisateurs comme le port de lunettes. Comme la majorité des algorithmes d'estimation de la pupille, la métrique proposée par [JKF01] est retenue durant l'évaluation, elle se traduit par :

$$e = \frac{\max\{D_L, D_R\}}{D} \quad (4)$$

Où D_L et D_R représentent les distances euclidiennes entre les centres de la pupille de l'œil gauche et droit estimés et ceux de la vérité terrain. D est la distance euclidienne entre les deux yeux de la vérité terrain. Nous comparons notre méthode à l'état de l'art sur les images où le détecteur de visage réussit. Cela est dû au fait que notre algorithme prend en entrée des régions extraites autour des yeux à partir de relations anthropométriques comme [TB11], [VG08] et [MFP*14]. La figure 1 montre quelques résultats de l'estimation de la pupille sur BioID ainsi que d'autres bases de données. Le tableau ?? compare notre approche aux méthodes de l'état de l'art, il représente le pourcentage des estimations correctes pour une distance e donnée.

Méthodes	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.25$
Jesorsky et al. [JKF01]	38.0	78.8	91.8
Timm et al. [TB11]	82.5	93.4	98.0
Valenti et al. [VG08]	84.1	91.0	96.6
Markus et al. [MFP*14]	89.9	97.1	99.7
Notre méthode	91.3	97.9	99.6

Table 1: Comparaison de notre méthode avec l'état de l'art.



Figure 3: Exemples d'échecs de l'estimation de la position de la pupille de l'oeil par notre méthode.

- $e \leq 0.25$ correspond à des distances comprises entre le centre de la pupille et le coin de l'œil ce qui implique une très faible précision d'estimation. Les résultats sont très comparables.
- $e \leq 0.10$ notre méthode est au dessus de [VG08] et [TB11]. La circularité non validée de la pupille due aux variations de la pose de la tête paralyse ces deux méthodes, tandis que notre approche se montre plus robuste grâce aux perturbations incluses dans la base d'apprentissage.
- $e \leq 0.05$ correspond à des distances très proches de la pupille impliquant une forte précision d'estimation. Notre méthode surpasse [MFP*14] grâce au potentiel étendu de régression de nos arbres lié à la projection dans l'espace de Hough ainsi qu'à l'absence de certains cas typiques dans la base de données d'apprentissage de [MFP*14] comme le port de lunettes.

Sur certaines images, notre algorithme échoue lors de l'estimation dans des contextes avec des contraintes extrêmes dues aux reflets de lunettes, de fermeture des yeux ou d'éclairages trop sombres. La figure 3 illustre ces cas.

4. Conclusion

Nous avons proposé une méthode basée sur les arbres de décision couplant des informations de régression et de classification, chaque arbre renvoie un ensemble de candidats sur la position 2D de la pupille d'une manière pyramidale. Tous les candidats sont projetés après pondération sur l'espace de Hough généralisé. Un fenêtrage de Parzen est appliqué sur cet espace pour sélectionner un maximum d'intensité correspondant à la position spatiale de la pupille. Comparé à l'état de l'art, notre algorithme donne de très bon résultats face à des situations très défavorables grâce aux perturbations introduites sur les données d'apprentissage ainsi qu'à l'extension de la capacité de généralisation des arbres couplés à l'espace de Hough. La robustesse de notre méthode peut répondre aux exigences des applications nécessitant un suivi précis du regard.

Références

[BBC13] BADRINARAYANAN V., BUDVYTIS I., CIPOLLA R. : Semi-supervised video segmentation using decision forests. In *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013, pp. 229–244.

[Bre01] BREIMAN L. : Random forests. *Machine learning*. Vol. 45, Num. 1 (2001), 5–32.

[CSRK11] CRIMINISI A., SHOTTON J., ROBERTSON D., KONUKOGLU E. : Regression forests for efficient anatomy detection and localization in ct studies. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*. Springer, 2011, pp. 106–117.

[GEW06] GEURTS P., ERNST D., WEHENKEL L. : Extremely randomized trees. *Machine learning*. Vol. 63, Num. 1 (2006), 3–42.

[GYR*11] GALL J., YAO A., RAZAVI N., VAN GOOL L., LEMPITSKY V. : Hough forests for object detection, tracking, and action recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. Vol. 33, Num. 11 (2011), 2188–2202.

[HJ10] HANSEN D. W., JI Q. : In the eye of the beholder : A survey of models for eyes and gaze. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. Vol. 32, Num. 3 (2010), 478–500.

[JKF01] JESORSKY O., KIRCHBERG K. J., FRISCHHOLZ R. W. : Robust face detection using the hausdorff distance. In *Audio-and video-based biometric person authentication* (2001), Springer, pp. 90–95.

[LF06] LEPETIT V., FUA P. : Keypoint recognition using randomized trees. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. Vol. 28, Num. 9 (2006), 1465–1479.

[MFP*14] MARKUŠ N., FRLJAK M., PANDŽIĆ I. S., AHLBERG J., FORCHHEIMER R. : Eye pupil localization with an ensemble of randomized trees. *Pattern recognition*. Vol. 47, Num. 2 (2014), 578–587.

[SJC08] SHOTTON J., JOHNSON M., CIPOLLA R. : Semantic texton forests for image categorization and segmentation. In *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE Conference on* (2008), IEEE, pp. 1–8.

[TB11] TIMM F., BARTH E. : Accurate eye centre localisation by means of gradients. In *VISAPP* (2011), pp. 125–130.

[VG08] VALENTI R., GEVERS T. : Accurate eye center location and tracking using isophote curvature. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008), IEEE, pp. 1–8.

[VJ01] VIOLA P., JONES M. : Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (2001), vol. 1, IEEE, pp. 1–511.

[VPSS*13] VILLANUEVA A., PONZ V., SESMA-SANCHEZ L., ARIZ M., PORTA S., CABEZA R. : Hybrid method based on topography for robust detection of iris center and eye corners. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*. Vol. 9, Num. 4 (2013), 25.

[WLSN07] WEIDENBACHER U., LAYHER G., STRAUSS P.-M., NEUMANN H. : A comprehensive head pose and gaze database.