

Théorie semi-quadratique, régularisation et estimation robuste

Pierre CHARBONNIER

Équipe de recherche ENDSUM, Cerema, 11 rue Jean Mentelin, B.P. 9, 67035 Strasbourg

Résumé – L’hypothèse classique de distribution gaussienne, correspondant à un critère des moindres carrés, s’avère inadaptée dans de nombreux problèmes où les données traitées se composent en réalité de deux catégories. Formalisée il y a environ trente ans dans le cadre de la régularisation des problèmes inverses, la théorie semi-quadratique permet d’établir des formes analytiques appropriées et conduit le plus souvent à des algorithmes déterministes, rapides et faciles à implanter. Cette théorie se situe à la confluence des statistiques robustes (M-estimateurs), de la régularisation markovienne ou variationnelle des problèmes inverses, des équations aux dérivées partielles et de l’analyse convexe. Ces inspirations diverses ont favorisé sa diffusion même si, dans ses multiples résurgences, elle n’est pas toujours identifiée en tant que telle. Nous proposons ici une synthèse de cette approche qui permet de résoudre une grande variété de problèmes pratiques en traitement du signal et des images.

Abstract – The conventional assumption of a Gaussian distribution, corresponding to a least-squares criterion, proves unsuitable in many problems where the data being processed actually comprise two categories. Formalized some thirty years ago in the context of inverse problem regularization, half-quadratic theory makes it possible to establish suitable analytical forms and leads to deterministic algorithms that are fast and easy to implement. This theory lies at the confluence of robust statistics (M-estimators), Markovian or variational regularization of inverse problems, partial differential equations and convex analysis. These diverse inspirations have favored its dissemination, even if, in its many resurgences, it is not always identified as such. We propose here a synthesis of this approach which can be used to solve a wide variety of practical problems in signal and image processing.

1 Introduction

Les critères des moindres carrés (MC) sont universellement employés pour résoudre des problèmes d’estimation de paramètres ou des problèmes inverses. En estimation, ils permettent de contraindre la norme de l’écart entre la prédiction d’une donnée par un modèle et son observation, ou *résidu*. En reconstruction régularisée (Tikhonov), ils interviennent pour pénaliser la norme de l’inconnue x ou d’une de ses dérivées. En représentant l’une ou l’autre de ces quantités, liées à x , par une variable u , l’équivalent probabiliste des critères des MC prend la forme d’une fonction gaussienne \mathcal{P} qui, selon le cas, représente une vraisemblance ou un *a priori*, et s’écrit (en considérant une variance unité pour simplifier) :

$$\mathcal{P} \propto \exp\left(-\frac{1}{2} \sum_{i=1}^N u_i^2\right) = \exp(-\mathcal{J}_{MC}(x)). \quad (1)$$

Les modèles des MC sont notoirement sensibles aux fortes valeurs de u_i . En régression, une seule donnée éloignée du modèle, i.e. un grand résidu, suffit à biaiser fortement la solution. En régularisation, la fonction quadratique s’oppose à la création de forts gradients dans la solution, qui prend alors un aspect trop lisse. Le problème est dû à la forme de la distribution gaussienne \mathcal{P} sous-jacente aux critères des MC : les grandes valeurs de u ont une probabilité d’apparition quasi-nulle. Cela revient à ne considérer qu’un seul type de données alors qu’en réalité, il y en a deux : en régression, des données conformes au modèle (les *inliers*, générant de faibles résidus) et des données atypiques (les *outliers*, forts résidus); en régularisation, du bruit (faibles gradients) et des discontinuités (forts gradients). Dans les deux cas il faudrait éliminer de l’optimisation les fortes valeurs de u , alors que leur influence sur la solution se trouve exacerbée par la quadratique.

Pour améliorer les choses, il s’agit donc, en premier lieu, de modifier la forme de \mathcal{P} pour gérer correctement les deux

sous-populations représentées par la variable u . De nombreux modèles ont été proposés dans la littérature, que ce soit dans le cadre de l’estimation robuste ou dans celui de la régularisation avec préservation des discontinuités. Nous en proposons une synthèse (Sec.2), en montrant comment la *fonction de pondération*, élément central de la théorie semi-quadratique (SQ) apporte une vue unifiée des modèles proposés. La section suivante (3) est consacrée aux deux théorèmes SQ et à l’algorithmique associée. Enfin, la section 4 conclut ce papier.

2 Modèles non quadratiques

On peut envisager deux façons d’améliorer le modèle. La première consiste à utiliser dans (1) une fonctionnelle comme :

$$\mathcal{J}(x) = \frac{1}{2} \sum_{i=1}^N \varphi(u_i), \quad (2)$$

où φ n’est plus forcément quadratique. En estimation robuste, cette formulation généralisée du Maximum de Vraisemblance est connue sous le nom de M-estimateur [22, 23] (φ est notée ρ dans ce cadre). En régularisation, on parle de modèle à *phi-fonction*. Toute la question réside dans le choix de φ . Comment assurer l’insensibilité du critère aux fortes valeurs et comment « mesurer » cette robustesse ?

Une seconde approche [17] introduit une variable *auxiliaire* $b_i \in \{0, 1\}$ pour « casser » la contrainte quadratique :

$$\mathcal{J}^*(x, b) = \frac{1}{2} \sum_{i=1}^N b_i (u_i)^2 + (1 - b_i). \quad (3)$$

La variable b est appelée « processus de ligne » en reconstruction régularisée, où elle modélise explicitement les contours. Ce critère *augmenté*, quadratique à b fixé, est celui des moindres carrés pondérés, très utilisé aussi en estimation robuste. On montre [6] que son optimisation sur b conduit à

Fonction	Abrév.	Réf.	Fonction de potentiel $\varphi(u)$	$\psi(u) = \varphi'(u)$	$\varphi'(u)/2u$	Cond. 4-6?	φ convexe?
Quadratique (Gauss)	Q	-	u^2	$2u$	$\frac{1}{u}$	non	oui
Hyper-surfaces (Quasi-Laplace)	HS	[11, 9]	$2\sqrt{1+u^2} - 2$	$\frac{2u}{\sqrt{1+u^2}}$	$\frac{1}{\sqrt{1+u^2}}$	oui	oui
Hebert-&Leahy (Cauchy)	HL	[21]	$\log(1+u^2)$	$\frac{2u}{(1+u^2)}$	$\frac{1}{(1+u^2)}$	oui	non
Geman & McClure	GM	[18]	$\frac{u^2}{1+u^2}$	$\frac{2u}{(1+u^2)^2}$	$\frac{1}{(1+u^2)^2}$	oui	non
Famille exponentielle lissée	SEF	[32][3]	$\frac{1}{\alpha} [(1+\epsilon^2)^\alpha - 1], \forall \alpha \neq 0$ $\rho(\epsilon) = \ln(1+\epsilon^2), \alpha = 0$	$2u(1+u^2)^{\alpha-1}$	$(1+u^2)^{\alpha-1}$	si $\alpha < 1$	si $\alpha < \frac{1}{2}$

TABLEAU 1 : Exemples de fonctions φ proposées dans la littérature et leurs dénominations les plus usuelles dans les approches déterministe et probabiliste. La fonction SEF permet de retrouver les fonctions Q, HS, HL et GM, selon la valeur de α

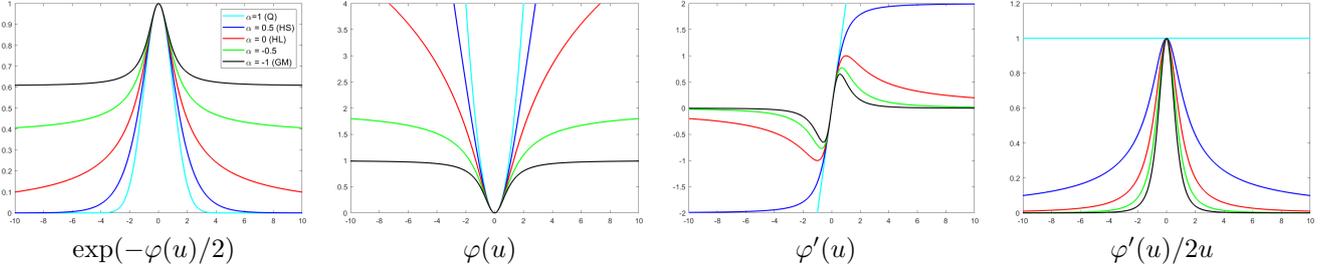


FIGURE 1 : (de gauche à droite) distribution statistique, fonction de potentiel, fonction d'influence et fonction de pondération de la famille exponentielle lissée (SEF) pour différentes valeurs de α . La légende est la même pour tous les graphiques

(2) pour la quadratique tronquée $\varphi(u) = \min(u^2, 1)$. Peut-on généraliser cette technique à une variable auxiliaire à valeurs réelles, permettant un comportement moins « binaire » ?

La théorie SQ apporte une réponse à ces questions.

2.1 Propriétés des fonctions φ

De très nombreuses fonctions de potentiel $\varphi(u)$ ont été proposées dans la littérature, que ce soit en régularisation [9, 35] ou en estimation robuste [15]. Quelques fonctions populaires sont présentées tab. 1. Leurs tracés (fig. 1) ont des allures très différentes. La théorie des M-estimateurs fournit des outils intéressants pour caractériser les fonctions. Cela peut concerner l'optimalité statistique des estimateurs en termes de variance, évaluée par l'*efficacité relative*, ou le pourcentage de valeurs atypiques qu'ils peuvent supporter (*point de rupture*). La *fonction d'influence* $\varphi'(u)$ [20] mesure, elle, l'effet de l'amplitude de u sur le résultat de l'estimation. On peut constater, fig. 1, que l'influence des forts u n'est pas bornée pour la quadratique. Par contre, elle l'est pour toutes les fonctions robustes. Parmi celles-ci, on distingue les fonctions *monotones*, comme HS, et les fonctions *redescendantes*, comme HL et GM. Pour ces dernières, l'influence tend vers 0, ce qui les rend très robustes. En contrepartie, on perd la convexité de $\varphi(u)$, ce qui pourrait causer des problèmes de convergence. Notons que nous n'avons pas représenté de fonctions définies par morceaux, comme la fonction de Huber $\min(u^2, 2u - 1)$ [22], car leur fonction d'influence n'est pas continue, ce qui les rend sensibles à de petites variations de u . Une fonction très intéressante est la famille exponentielle lissée (SEF) qui permet de retrouver les quatre autres fonctions du tableau 1 à l'aide d'une forme analytique paramétrée par un seul coefficient, α . Proposée dans [32], elle a été « redécouverte » récemment [3].

Afin d'établir sous quelles conditions on peut rendre φ insensible aux forts u , on a étudié les équations d'optimalité associées aux critères fondés sur (2). En régularisation [9, 12] on les interprète en termes d'équations aux dérivées partielles

(EDP) de diffusion par un laplacien pondéré. En estimation robuste [10], on se ramène à l'estimation par MC pondérés. Selon le cas, les coefficients de diffusion ou les poids sont donnés par la *fonction de pondération*, paire, $\varphi'(u)/2u$. On montre que, pour obtenir un comportement différencié selon les valeurs de u , il faut :

$$\lim_{u \rightarrow 0} \varphi'(u)/2u = M, \quad 0 < M < +\infty \quad (4)$$

$$\lim_{u \rightarrow +\infty} \varphi'(u)/2u = 0 \quad (5)$$

$$\varphi'(u)/2u \quad \text{continue, strictement décroissante.} \quad (6)$$

En général, M est choisi égal à 1. Ces conditions assurent un comportement quadratique (de type MC) près de l'origine (4), qui se relâche pour les fortes valeurs de u (5), comme souhaité. La condition (6) assure la continuité du modèle. On constate sur les exemples de la fig. 1 que les diverses fonctions φ ont des fonctions de pondération au comportement similaire.

3 Théorèmes et algorithmes

On peut montrer [9] que les fonctions réelles φ paires, croissantes, vérifiant les conditions (4-6), sont telles que :

$$\varphi(u) = \min_{b \in [0,1]} (bu^2 + \Psi(b)), \quad \forall u, \quad (7)$$

où Ψ est une fonction strictement convexe et b est la variable auxiliaire, dont la valeur optimale à u fixée est donnée par la fonction de pondération : $b_{inf} = \varphi'(u)/2u$. On parle d'*extension multiplicative*. Ce résultat généralise un théorème initialement limité aux fonctions à asymptote horizontale, introduit dans [19] en régularisation. On a également, sous les mêmes conditions, une extension *additive* :

$$\varphi(u) = \min_{b \in \mathbb{R}^+} ((b-u)^2 + \xi(b)), \quad \forall u, \quad (8)$$

où ξ est une fonction strictement convexe. La valeur optimale de b à u fixée est donnée par $b_{inf} = (1 - \varphi'(u)/2u)u$. En régularisation, ce résultat a été formalisé à l'origine dans [16].

3.1 Dérivation des théorèmes

Il existe plusieurs façon de prouver les théorèmes SQ. En estimation robuste, Huber [23] utilise une approximation locale de φ par une succession de paraboles. Le lien entre cette approche et la théorie SQ a été mis en évidence dans [5, 14]. Cela constitue un exemple de *transfert d'optimisation* [28] : on parle d'algorithmes à approximation quadratique majorante (ou minorante, si l'on raisonne en probabilité) [1], un cas particulier d'algorithme de majoration-minimisation [31]. Ce principe est à la base de la preuve de convergence de l'algorithme *Expectation-Maximization* (EM), et une interprétation des algorithmes semi-quadratiques dans le cadre de l'EM a été proposée dans [8]. La deuxième famille de démonstration utilise directement la définition de la propriété de convexité, voir [9] pour la forme multiplicative et [10] pour la forme additive. Notons que la justification de l'algorithme Mean-Shift [13] (dont on montre qu'il s'agit d'un algorithme SQ [33]) reprend également ce canevas. Deux autres familles de démonstrations exploitent la dualité. D'abord exploitée pour démontrer le second théorème SQ [16, 9] la dualité de Legendre a ensuite été appliquée pour le premier [24, 2]. Enfin, une dérivation de la théorie SQ a été proposée dans un cadre lagrangien [32].

L'étude de ces démonstrations (voir [10] pour plus de détails) met en lumière le rôle central que jouent les propriétés de convexité et de monotonie dans la théorie SQ : la forme multiplicative repose sur la convexité et la décroissance de la fonction $-\varphi(\sqrt{u})$; la forme additive repose sur la convexité et la croissance de la fonction $\frac{1}{2}[u^2 - \varphi(u)]$. Il est intéressant de noter que ces propriétés sont assurées par les conditions (4-6) sur la fonction de pondération.

3.2 Algorithmique

En reportant les équations (7) et (8) dans (2), on obtient des critères plus simples à minimiser que le critère d'origine. En effet, à b fixée, ils sont quadratiques et leur minimisation conduit à la résolution d'équations normales linéaires. De plus, lorsque la variable x est fixée, le critère augmenté est convexe en b et la valeur du minimum est donnée par une expression analytique qui dépend de la fonction de pondération. D'après (4) et (5), il est clair que la variable auxiliaire (non binaire) b joue le rôle de « marqueur » de données : « inlier / outlier » en estimation, « bruit / discontinuité » en régularisation.

Bien que des algorithmes stochastiques existent [19, 25], les propriétés SQ sont, la plupart du temps, exploitées sous forme d'algorithmes déterministes opérant par minimisations alternées par rapport à chacune des variables. En général, on positionne initialement tous les b_i à 1, ce qui revient à commencer par un algorithme des MC classique.

En estimation robuste, la forme multiplicative conduit au célèbre algorithme *Iteratively Reweighted Least-Squares* (IRLS) proposé dès 1974 dans [4]. En reconstruction sur des bases propres, une version repondérée de MeanShift a été proposée dans [33]. Une interprétation SQ du filtrage bilatéral a été exploitée dans [7]. Pour la forme additive, on obtient un algorithme des MC avec remise à jour des résidus [23], parfois appelé *Residual Steepest Descent* ou RSD [1].

En reconstruction régularisée, les algorithmes sont nommés ARTUR [12] pour la forme multiplicative et LEGEND pour la forme additive [9]. Une version adaptée au bruit de Poisson, baptisée MOISE, a été proposée dans [26].

Notons que la théorie SQ est un moyen justifié de dériver l'algorithme du point fixe que l'on obtiendrait en figeant les pondérations des équations normales à leur valeur calculée à l'itération précédente (technique de linéarisation du gradient [29] ou *lagged diffusivity* [34]).

Ces algorithmes convergent vers l'unique minimum global dans le cas des fonctions φ convexes. Dans le cas non-convexe, ils convergent vers un minimum local, à condition que les points critiques de la fonctionnelle soient isolés [12, 2]. La convergence est généralement plus lente dans le cas de la seconde forme SQ, mais celle-ci peut être largement moins demandeuse de calculs dans certains contextes, voir par exemple [14, 25]. Pour une analyse plus fine et des résultats permettant d'améliorer la vitesse de convergence des algorithmes SQ, on pourra se référer à [30, 1, 27, 29].

Pour améliorer la convergence pour des potentiels non convexes, une stratégie efficace, inspirée du GNC (*Graduated Non Convexity*) de [6], consiste à employer *en continuation* une famille de potentiels d'abord convexes, puis de plus en plus non-convexes. La SEF [32] est très bien adaptée pour cela, car il suffit d'ajouter une boucle autour de l'algorithme SQ, en faisant diminuer progressivement la valeur du paramètre α .

Dans cette brève synthèse nous avons, pour simplifier, édulcoré le paramètre d'échelle dans (1) et (2). En pratique la variable u_i doit être divisée par un facteur de proportionnalité, qui ajuste le comportement de la fonction à la dynamique des données. D'autre part, lorsque le critère est régularisé, un coefficient permet de pondérer les termes de vraisemblance et d'*a priori*. Ces *hyper-paramètres* sont souvent laissés au choix de l'utilisateur. Il est cependant possible de les fixer de manière plus automatique, en utilisant le MAD (*median absolute deviation*) en estimation robuste, ou des méthodes bayésiennes, comme dans [25] en reconstruction régularisée.

4 Conclusion

La théorie semi-quadratique est un outil simple et efficace pour dépasser les MC en s'affranchissant de l'influence des fortes valeurs. Son développement a bénéficié de recherches dans des domaines variés, tels que l'estimation statistique, la régularisation, la théorie des EDP, ou encore l'analyse convexe, qui ont permis d'améliorer modèles et algorithmes. Une étude bibliométrique sur *Web Of Science* confirme que les principales publications sont citées régulièrement, même si un pic est constaté depuis 2020 pour [11], et que la théorie SQ est largement utilisée en reconstruction et traitement du signal et des images, vision par ordinateur, et même en intelligence artificielle, dans des domaines applicatifs très variés.

Références

- [1] M. ALLAIN, J. IDIER et Y. GOUSSARD : On global and local convergence of half-quadratic algorithms. *IEEE Trans. Image Processing*, 15(5):1130–1142, 2006.
- [2] G. AUBERT et P. KORNPBST : *Mathematical Problems in Image Processing : Partial Differential Equations and the Calculus of Variations*. Springer, 2006.
- [3] J. T. BARRON : A general and adaptive robust loss function. *In 2019 IEEE/CVF CVPR*, pages 4326–4334, 2019.

- [4] A. BEATON et J. TUKEY : The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16(2):174–185, 1974.
- [5] M. J. BLACK et A. RANGARAJAN : On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International Journal of Computer Vision*, 19(1):57–92, 1996.
- [6] A. BLAKE et A. ZISSERMAN : *Visual reconstruction*. MIT Press, Cambridge, USA, 1987.
- [7] L. CARAFFA, J.-P. TAREL et P. CHARBONNIER : The guided bilateral filter : When the joint/cross bilateral filter becomes robust. *IEEE Trans. on Image Processing*, 24(4):1199–1208, 2015.
- [8] F. CHAMPAGNAT et J. IDIER : A connection between half-quadratic criteria and EM algorithms. *IEEE Signal Processing Letters*, 11(9):709–712, 2004.
- [9] P. CHARBONNIER : *Reconstruction d'image : régularisation avec prise en compte des discontinuités*. Thèse de doctorat, Univ. Nice-Sophia Antipolis, France, 1994.
- [10] P. CHARBONNIER : *Modèles de forme et d'apparence en traitement d'images*. Habilitation à diriger des recherches, Université de Strasbourg, 2009.
- [11] P. CHARBONNIER, G. AUBERT, L. BLANC-FÉRAUD et M. BARLAUD : Two deterministic half-quadratic regularization algorithms for computed imaging. *In Proc. IEEE ICIP*, volume II, pages 168–172, 1994.
- [12] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT et M. BARLAUD : Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.
- [13] D. COMANICIU et P. MEER : Mean shift : A robust approach toward feature space analysis. *IEEE Trans. Pat. Anal. Mach. Intel.*, 24(5):603–619, 2002.
- [14] R. DAHYOT : *Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier*. Thèse de doct., Univ. Strasbourg I, 2001.
- [15] D.Q.F. DE MENEZES, D.M. PRATA, A.R. SECCHI et J.C. PINTO : A review on robust M-estimators for regression analysis. *Computers & Chemical Engineering*, 147:107254, 2021.
- [16] D. GEMAN et C. YANG : Nonlinear image recovery with half-quadratic regularization and FFT's. *IEEE Trans. Image Processing*, 4(7):932–946, 1995.
- [17] S. GEMAN et D. GEMAN : Stochastic relaxation, gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pat. Anal. Mach. Intel.*, 6(6):721–741, 1984.
- [18] S. GEMAN et D.E. MCCLURE : Bayesian image analysis : an application to single photon emission tomography. *Proc. Statist. Comput. Sect.*, pages 12–18, 1985.
- [19] S. GEMAN et G. REYNOLDS : Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pat. Anal. Mach. Intel.*, 14(3):367–383, 1992.
- [20] F R. HAMPEL : The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.
- [21] T. HEBERT et R. LEAHY : A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.*, (2).
- [22] P.J. HUBER : Robust estimation of a location parameter. *The Annals of Math. Stat.*, 35(1):73–101, 1964.
- [23] P.J. HUBER : *Robust statistics*. John Wiley and Sons, New York, 1981.
- [24] J. IDIER : Convex half-quadratic criteria and interacting auxiliary variables for image restoration. *IEEE Trans. Image Processing*, 10(7):1001–1009, 2001.
- [25] A. JALOBEANU : *Modèles, estimation bayésienne et algorithmes pour la déconvolution d'images aériennes et satellitaires*. Thèse de doctorat, Université de Nice-Sophia Antipolis, France, 2001.
- [26] P.M. KOULIBALY, P. CHARBONNIER, L. BLANC-FÉRAUD, I. LAURETTE, J. DARCOURT et M. BARLAUD : Poisson statistic and half-quadratic regularization for emission tomography reconstruction algorithm. *In Proc. IEEE ICIP*, volume II, pages 729–732, 1996.
- [27] C. LABAT : *Algorithmes d'optimisation de critères pénalisés pour la restauration d'images. Application à la déconvolution de trains d'impulsions en imagerie ultrasonore*. Thèse de doctorat, Univ. Nantes, 2006.
- [28] K. LANGE, D. R. HUNTER et I. YANG : Optimization transfer using surrogate objective functions. *Journal of Comput. and Graph. Stat.*, 9(1):1–20, 2000.
- [29] M. NIKOLOVA et R. CHAN : The equivalence of half-quadratic minimization and the gradient linearization iteration. *IEEE Trans. image Processing*, 16(6):1623–1627, 2007.
- [30] M. NIKOLOVA et M. NG : Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM J. Sci. Comput.*, 27(3):937–966, 2005.
- [31] Y. SUN, P. BABU et D. P. PALOMAR : Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Processing*, 65(3):794–816, 2017.
- [32] J.P. TAREL, S.S. IENG et P. CHARBONNIER : Using robust estimation algorithms for tracking explicit curves. *In 6th ECCV*, volume LNCS 2350, pages 492–507, 2002.
- [33] T. VIK, F. HEITZ et P. CHARBONNIER : Robust pose estimation and recognition using non-gaussian modeling of appearance subspaces. *IEEE Trans. Pat. Anal. Mach. Intel.*, 29(5):901–905, 2007.
- [34] C. VOGEL et M. OMAN : Iterative methods for Total Variation denoising. *SIAM J. Sci. Comput.*, 17(1), 1996.
- [35] Z.ZHANG : Parameter estimation techniques : a tutorial with application to conic fitting. *Image and Vision Computing*, 15(1):59–76, 1997.