

Spécification du filtrage mean-shift pour la classification non supervisée de séries temporelles multidimensionnelles

Simon MURE^{1*}, Thomas GRENIER¹, Pierre GANÇARSKI², Hugues BENOIT-CATTIN¹

¹Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1, France

²ICUBE-UMR 7357, Pole API, Bd Sébastien Brant, BP 10413, 67412 Illkirch Cedex, France

{simon.mure, thomas.grenier, hugues.benoit-cattin}@creatis.insa-lyon.fr
gancarski@unistra.fr

Résumé – Une nouvelle méthode combinant filtrage et classification non supervisée d’images spatiotemporelles multidimensionnelles, basée sur le formalisme mean-shift, est proposée dans cet article. Le formalisme mean-shift a été adapté afin de non plus faire converger des mesures associées à des échantillons, mais de faire converger l’évolution temporelle de ces mesures dans l’espace amplitude. Pour ce faire, nous avons d’une part modifié le domaine joint spatial-amplitude afin de prendre en compte les évolutions temporelles des mesures et d’autre part, intégré au formalisme une contrainte sur la similarité des trajectoires des échantillons dans l’espace amplitude avec l’utilisation de la norme infini. L’efficacité de la méthode a dans un premier temps été évaluée quantitativement sur des données spatiotemporelles simulées. Nous présentons ensuite de manière qualitative les résultats obtenus dans un contexte applicatif réel, sur une séquence d’images satellitaires multispectrales.

Abstract – A novel approach based on the mean-shift procedure, combining both unsupervised spatiotemporal filtering and clustering of multivariate image sequences, is proposed in this paper. Our contributions are the spatial-range domains modification to integrate the temporal feature into the mean-shift formalism, and the addition of a trajectory constraint in the range domain with the use of the infinity norm. The efficiency of the proposed method was evaluated quantitatively on synthetic data and its pertinence was qualitatively assessed using a multispectral satellite image sequence.

1 Introduction

Pour aller toujours plus loin dans l’analyse, les techniques de traitement d’images tirent profit du maximum d’information disponible. L’augmentation du nombre de caractéristiques étudiées, l’ajout d’a priori, la diversification et la multiplication des acquisitions en sont la conséquence. Depuis longtemps le traitement de l’information temporelle suscite un réel intérêt, à travers l’analyse des séries temporelles, pour comprendre ou identifier des processus en fonction de leurs évolutions [1].

Nous nous intéressons ici au problème du filtrage et de la classification non supervisée de séquences d’images multidimensionnelles, c’est à dire sans utilisation d’a priori et sans phase d’apprentissage. De telles données peuvent être des vidéos couleurs, des suivis longitudinaux en médecine (avec plusieurs modalités/séquences IRM) ou encore des séquences d’images multispectrales (images satellites par exemple). De nombreuses approches de filtrage et de classification non supervisée ont été proposées dans ces contextes [2, 3, 4, 5, 6].

Dans cet article, nous proposons une modification de l’approche de filtrage mean-shift (M-S) [7, 8] pour qu’elle puisse filtrer et

classifier de manière non supervisée des séquences d’images multidimensionnelles. L’approche M-S a déjà été utilisée dans ce cadre (filtrage spatiotemporel [9], analyse d’images satellites [10]), cependant aucune méthode n’exploite conjointement l’information temporelle et les mesures multidimensionnelles. C’est ce que nous revendiquons, ainsi que la possibilité de classifier les échantillons simultanément au filtrage.

La partie 2 est consacrée à la présentation de notre proposition, STM-S, permettant le filtrage et la classification non supervisée de séries temporelles multidimensionnelles. L’approche STM-S sera testée sur des données simulées où seule une technique spatiotemporelle prenant en compte des mesures multidimensionnelles peut regrouper les évolutions similaires. Nous appliquerons ensuite l’approche STM-S au filtrage et à la classification d’une séquence d’images satellite multispectrales. Les expérimentations ainsi que leurs résultats seront présentés et discutés dans la partie 3.

2 Mean-shift spatiotemporel (STM-S)

Dans cette partie, nous adaptons le formalisme M-S au filtrage de séries temporelles (§2.1). Celui-ci sera étendu aux mesures multidimensionnelles (§2.2) et intégrera une étape de classification (§2.3) dans le processus itératif de filtrage.

*Ce travail a été réalisé grâce au soutien financier apporté par la bourse CNRS PEPS INS2I. Il a aussi été réalisé au sein du LABEX PRIMES (ANR-11-LABX-0063) de l’Université de Lyon, dans le cadre du programme "Investissements d’Avenir" (ANR-11-IDEX-0007) de l’Etat Français, géré par l’Agence Nationale de la Recherche (ANR).

2.1 Formalisme

L'approche STM-S est basée sur la procédure M-S "blurring", procédure dont la convergence a été démontrée [11].

Dans un premier temps, les n échantillons $\{\mathbf{x}_i\}_{i=1\dots n}$ à traiter sont définis par :

$$\mathbf{x}_i = [\mathbf{x}_{s,i} \ \mathbf{x}_{t,i}] \in \mathbf{X} \quad \text{avec} \quad \begin{array}{l} \mathbf{x}_{s,i} \in \mathbb{R}^S : \text{support spatial} \\ \mathbf{x}_{t,i} \in \mathbb{R}^T : \text{support amplitude} \end{array} \quad (1)$$

Un échantillon \mathbf{x}_i est donc constitué d'une position $\mathbf{x}_{s,i}$, à valeurs dans \mathbb{R}^S , et d'une série de mesures $\mathbf{x}_{t,i}$ observées à T instants, à valeurs dans \mathbb{R}^T . Ceci étant, l'équation régissant le filtrage STM-S d'un échantillon à l'itération k est notée :

$$\mathbf{x}_i^{[k+1]} = \frac{\sum_{j=1}^n Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) \cdot Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}) \cdot \mathbf{x}_j^{[k]}}{\sum_{j=1}^n Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) \cdot Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]})} \quad (2)$$

où $Sp_{i,j}(\cdot)$ et $Ra_{i,j}(\cdot)$ correspondent respectivement aux distances pondérées, entre deux échantillons \mathbf{x}_i et \mathbf{x}_j , utilisées avec les supports spatial et amplitude :

$$Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) = g_s \left(d_s^2(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}, \mathbf{H}_s) \right) \quad (3)$$

$$Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}) = g_r \left(d_r^2(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}, \mathbf{H}_r) \right) \quad (4)$$

D'une part, la distance séparant deux positions \mathbf{u}_s et \mathbf{v}_s est calculée via la distance de Mahalanobis $d_s(\mathbf{u}_s, \mathbf{v}_s, \mathbf{H}_s)$. \mathbf{H}_s étant une matrice échelle de taille $S \times S$, diagonale et définie positive. D'autre part, la distance utilisée pour calculer l'éloignement entre deux mesures \mathbf{u}_r et \mathbf{v}_r est la norme infini :

$$d_r(\mathbf{u}_r, \mathbf{v}_r, \mathbf{H}_r) = \|\mathbf{H}_r^{-\frac{1}{2}}(\mathbf{u}_r - \mathbf{v}_r)\|_\infty \quad (5)$$

Celle-ci transcrit l'éloignement maximal, temps à temps, entre deux évolutions de mesures. \mathbf{H}_r est une matrice échelle de taille $T \times T$ appliquée aux mesures, diagonale et définie positive. Les distances calculées sur chaque support sont ensuite pondérées par une fonction déterminant la participation ou non d'un échantillon candidat \mathbf{x}_j au filtrage d'un échantillon référence \mathbf{x}_i en fonction de leur éloignement :

$$g_s(d_s^2(\cdot)) = g_r(d_r^2(\cdot)) = \begin{cases} 1 & \text{si } d_s^2(\cdot), d_r^2(\cdot) \leq 1 \\ 0 & \text{sinon} \end{cases} \quad (6)$$

Si l'éloignement sur un des supports dépasse la valeur fixée par la matrice échelle associée, alors l'échantillon candidat sera exclu du filtrage de l'échantillon référence pour l'itération courante. Par conséquent, la combinaison des définitions (3), (4) et (6) dans la formule générale du filtrage STM-S d'un échantillon (2) mène à uniquement considérer les échantillons qui lui seront proches spatialement au regard de l'échelle \mathbf{H}_s , et dont les évolutions de mesures lui resteront suffisamment proches, temps à temps, au regard de l'échelle \mathbf{H}_r .

2.2 Mesures multidimensionnelles

Nous considérons maintenant des mesures spatiotemporelles multidimensionnelles comme les composantes rouge, vert et bleu d'une image couleur.

Les mesures $\mathbf{x}_{t,i}$ s'expriment désormais $[\mathbf{x}_{t,i}^{C_1} \dots \mathbf{x}_{t,i}^{C_f}]$. Chacune des mesures C d'indice $f = 1, \dots, F$ est réalisée à T instants. Le vecteur de mesures $\mathbf{x}_{t,i}$ est donc à valeurs dans $\mathbb{R}^{T \times F}$.

La distance $d_r(\cdot)$ est conservée dans le cas multidimensionnel. La matrice d'échelle en amplitude \mathbf{H}_r de taille $T.F \times T.F$ est diagonale par blocs de taille $T \times T$, ce qui permet de combiner ou non les mesures C_f . La forme diagonale permettra de considérer chacune des mesures de manière indépendante.

2.3 Classification non supervisée

Conjointement au filtrage, une classification des échantillons est intégrée dans la procédure STM-S, ce qui n'est pas le cas avec les approches M-S standard.

La classification consiste à regrouper les échantillons très proches sur tous les supports à la fin de chaque itération, en faisant l'hypothèse qu'ils vont converger aux mêmes modes. Ainsi, seul le représentant des échantillons participera au filtrage et le poids lui étant associé sera proportionnel au nombre d'échantillons représentés. L'équation (2) s'exprime désormais :

$$\mathbf{x}_i^{[k+1]} = \frac{\sum_{j=1}^n Sp_{i,j}(\cdot) \cdot Ra_{i,j}(\cdot) \cdot w_j^{[k]} \cdot \mathbf{x}_j^{[k]}}{\sum_{j=1}^n Sp_{i,j}(\cdot) \cdot Ra_{i,j}(\cdot) \cdot w_j^{[k]}} \quad (7)$$

où $w_j^{[k]}$ correspond au nombre d'échantillons représentés par l'échantillon j et où n correspond au nombre d'échantillons à l'itération k .

Les poids $w_j^{[k+1]}$ sont actualisés après le calcul des $\mathbf{x}_i^{[k+1]}$ en regroupant les échantillons très proches. Cette proximité est déduite des valeurs de $d_s(\cdot)$ et $d_r(\cdot)$ en utilisant des matrices échelles \mathbf{H}_s et \mathbf{H}_r très sélectives. Afin de minimiser l'impact du regroupement sur le processus STM-S c'est le barycentre des échantillons à regrouper, pondérés par $w^{[k]}$, qui les représentera à l'itération suivante.

Considérant \bar{k} le nombre moyen d'itérations pour converger, le coût algorithmique du filtrage STM-S est de l'ordre de $O(\bar{k}n^2)$. Nous avons observé que la classification permet de diminuer, en moyenne, le nombre d'échantillons de moitié à chaque itération ($n^{[k+1]} \lesssim n^{[k]}/2$). Par conséquent, le coût algorithmique du filtrage combiné à la classification est de l'ordre de $O(4n^2)$. Le nombre moyen d'itérations n'ayant plus d'impact sur le coût algorithmique, le gain est potentiellement considérable.

3 Résultats et discussion

Deux expérimentations ont été conduites afin d'évaluer les résultats de la méthode proposée. Dans un premier temps, une évaluation quantitative de la robustesse de l'approche STM-S au flou et au bruit a été faite sur des données simulées. Dans un second temps, les premiers résultats obtenus par l'approche STM-S sur une série d'images multispectrales acquises par satellite sont présentés qualitativement.

3.1 Mesures quantitatives

La qualité du filtrage est évaluée avec un rapport signal sur bruit (*PSNR*) adapté aux données spatiotemporelles et multidimensionnelles. Cette mesure dépend de l'erreur quadratique moyenne à chaque instant entre le résultat du filtrage \mathbf{X} et les évolutions de référence \mathbf{X}^* :

$$PSNR(\mathbf{X}, \mathbf{X}^*) = 10 \log_{10} \left(\frac{\text{MAX}_{\mathbf{X}_{in}}^2}{\frac{1}{nT} \sum_{i=1}^n \|\mathbf{x}_{t,i} - \mathbf{x}_{t,i}^*\|^2} \right) \quad (8)$$

avec \mathbf{X}_{in} la séquence d'images à filtrer et $\text{MAX}_{\mathbf{X}_{in}}$ sa dynamique couleur.

La classification \mathcal{C}_X obtenue par l'approche STM-S est évaluée avec le calcul du coefficient de *DICE* pour chaque classe r par rapport à la référence \mathcal{C}_{X^*} .

$$DICE(\mathcal{C}_X^r, \mathcal{C}_{X^*}^r) = \frac{2|\mathcal{C}_X^r \cap \mathcal{C}_{X^*}^r|}{|\mathcal{C}_X^r| + |\mathcal{C}_{X^*}^r|} \quad (9)$$

avec $|\cdot|$ le cardinal d'un ensemble. On rappelle que l'approche STM-S n'a pas d'a priori sur le nombre de classes.

3.2 Images simulées

Afin de tester l'efficacité de l'approche STM-S, nous avons simulé une évolution des valeurs d'intensités rouge, vert et bleu des pixels d'une image couleur. Ces valeurs sont des nombres réels compris entre 0 et 4 et seront considérées comme références. Une série d'images 64×64 de 8 instants a été générée. Ces images ont ensuite été dégradées à l'aide d'un filtre moyen de taille 3×3 ainsi qu'avec un bruit additif gaussien centré et d'écart-type 0.3, appliqués indépendamment sur chaque dimension couleur et à chaque instant. La figure 1 illustre ces images simulées, la dynamique des intensités ayant été étendue pour l'affichage. Il est important de préciser que les évolutions choisies pour les intensités ne permettent pas la séparation des cinq classes sans prendre en compte simultanément les évolutions d'intensités dans les trois canaux couleur. De plus, les dégradations introduites rendent difficiles la détermination des frontières des classes.

Sur la figure 1, les résultats du filtrage et de la classification STM-S de la série d'images sont présentés pour des valeurs $\mathbf{H}_s = \text{diag}(\infty, \infty)$ et $\mathbf{H}_r = \text{diag}(1.75^2, 1.75^2, 1.75^2)$. L'utilisation de très grandes valeurs pour l'échelle spatiale permet de regrouper les échantillons en tenant uniquement compte de leurs évolutions dans l'espace amplitude sans introduire de contrainte sur le voisinage spatial. La méthode exploite ainsi toutes les évolutions dans l'espace amplitude pour filtrer chacun des échantillons. Nous avons aussi choisi, empiriquement, de regrouper après chaque itération les échantillons dont l'éloignement dans l'espace amplitude est au moins 30 fois inférieur à l'échelle fixée via la matrice \mathbf{H}_r .

On peut remarquer que malgré les dégradations introduites, le résultat du filtrage STM-S est visuellement satisfaisant (Fig.1). Quantitativement, le *PSNR* égal à 19dB pour la séquence d'

Classe	1	2	3	4	5
<i>DICE</i>	1.000	0.999	0.997	0.999	1.000

TABLE 1 – Coefficients *DICE* des 5 classes obtenues par STM-S avec les images simulées.

entrée est de 31dB après filtrage. De même, la classification obtenue par STM-S est très semblable à l'image de référence avec des coefficients de *DICE* supérieurs à 0.990 (Table 1).

3.3 Images satellites

Les données réelles sont constituées de 15 acquisitions réalisées en 2007 par le satellite Formosat-2 dans le sud-ouest de la France. Chaque image est orthorectifiée et composée de quatre bandes spectrales. Par la suite, nous avons choisi de n'utiliser que les bandes proche infrarouge (NIR), rouge et vert. La classification obtenue par l'approche STM-S est comparée à celle proposée dans [10]. Nous avons retenu les paramètres d'échelles $\mathbf{H}_s = \text{diag}(10^2, 10^2)$ et $\mathbf{H}_r = \text{diag}(40^2, 40^2, 40^2)$. La figure 2 présente une image satellite, les classifications de la méthode [10] et de STM-S. Les délimitations des classes sont quasiment identiques entre les deux approches mais on observe une surestimation du nombre de classes par STM-S. De nos expériences, cet effet n'est pas toujours atténué par l'augmentation des valeurs de \mathbf{H}_s . Les évolutions des intensités de deux des classes sont représentées avant et après filtrage sur la figure 2. Ces deux classes sont regroupées par l'approche [10]. L'analyse des évolutions des bandes spectrales permet d'expliquer pourquoi l'approche STM-S considère ces deux classes comme différentes. En effet, malgré des évolutions similaires il existe un déphasage temporel (bandes spectrales rouge et verte) ainsi que de trop fortes dissimilarités entre les évolutions d'intensités dans les bandes spectrales NIR et rouge. Cette sélectivité stricte est introduite par l'utilisation de la norme infini. Un post-traitement permettrait de regrouper les classes d'évolutions jugées dissimilaires par notre approche à cause d'un déphasage ou d'un écart ponctuel trop important.

4 Conclusion

Nous avons proposé une approche originale basée sur l'approche mean-shift permettant de filtrer et de classifier, de manière non supervisée, des données spatiotemporelles multidimensionnelles. Les originalités de la méthode résident dans la représentation des échantillons, l'introduction de la norme infini dans le processus de filtrage pour quantifier la similarité entre deux évolutions, la généralisation au cas multidimensionnel et la réalisation d'une étape de classification conjointe aux itérations de filtrage permettant de réduire la complexité de l'algorithme. L'utilisabilité de l'approche a été démontrée sur une séquence d'images simulées, puis sur une série d'acquisitions satellites multispectrales. Des travaux sont en cours sur la prise en compte de déphasages temporels entre des évolutions de signaux dans l'espace amplitude.

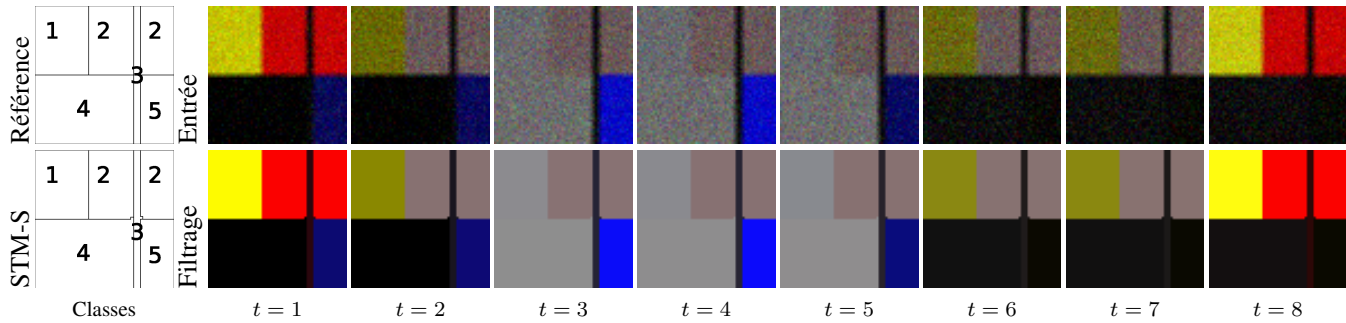


FIGURE 1 – Première ligne : classification de référence et séquence d’images simulées. Deuxième ligne : résultats de la classification non supervisée et du filtrage STM-S.

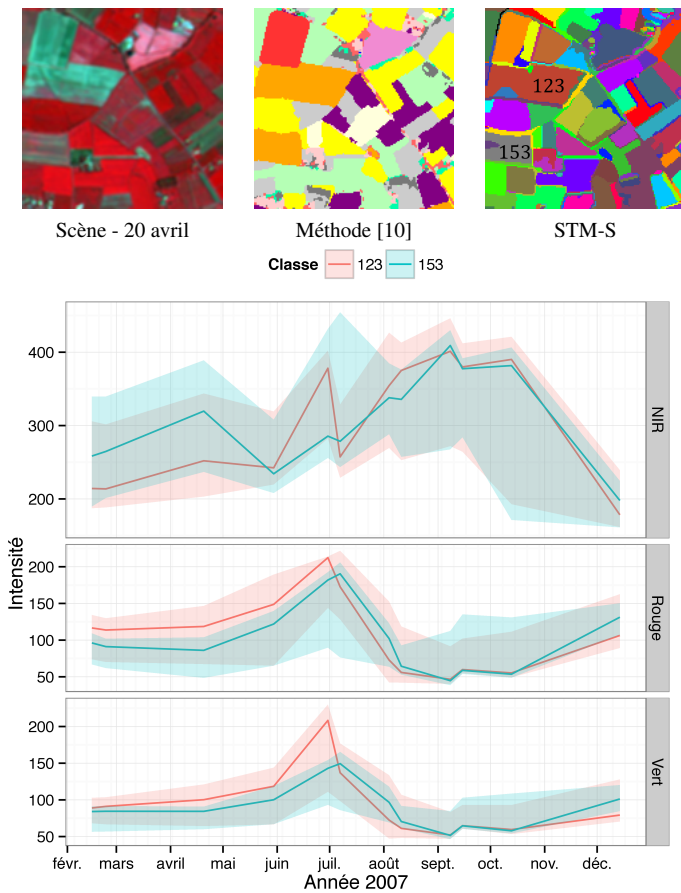


FIGURE 2 – Images : représentation de la scène en fausses couleurs, classifications obtenues par [10] et par STM-S. Courbes : évolutions des intensités filtrées (et non filtrées en clair) des bandes NIR/Rouge/Vert des classes 123 et 153 après classification STM-S, mais regroupées par [10].

Références

[1] C. Aggarwal and C. Reddy, *Data Clustering : Algorithms and Applications*, CRC Press, 2013.

[2] A. Buades, B. Coll, and J. M Morel, “Denoising image sequences does not require motion estimation,” in

Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on, Sept 2005, pp. 70–74.

- [3] S. Bhagavathy and J. Llach, “Adaptive spatio-temporal video noise filtering for high quality applications,” in *Acoustics, Speech and Signal Processing, IEEE International Conference on*, April 2007, vol. 1, pp. I–761–I–764.
- [4] G. Varghese and Z. Wang, “Video denoising based on a spatiotemporal gaussian scale mixture model,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 7, pp. 1032–1040, July 2010.
- [5] F. Petitjean, J. Inglada, and P. Gancarski, “Satellite image time series analysis under time warping,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 8, pp. 3081–3095, Aug 2012.
- [6] M. Maggioni, E. Sanchez-Monge, and A. Foi, “Joint removal of random and fixed-pattern noise through spatiotemporal video filtering,” *Image Processing, IEEE Transactions on*, vol. 23, no. 10, pp. 4282–4296, Oct 2014.
- [7] D. Comaniciu and P. Meer, “Mean shift : A robust approach toward feature space analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [8] T. Grenier, C. Revol-Muller, and G. Gimenez, “Hybrid approach for multiparametric mean shift filtering,” in *Image Processing, IEEE Proceedings of the International Conference on*, Atlanta, USA, 2006, pp. 1541–1544.
- [9] N. Anjum and A. Cavallaro, “Multifeature object trajectory clustering for video analysis,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 11, pp. 1555–1564, Nov 2008.
- [10] F. Petitjean, C. Kurtz, N. Passat, and P. Gancarski, “Spatio-temporal reasoning for the classification of satellite image time series,” *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1805 – 1815, 2012.
- [11] Y. Cheng, “Mean shift, mode seeking, and clustering,” *Pattern Analysis and Machine Intelligence, IEEE Transaction on*, vol. 17, no. 8, pp. 790–799, 1995.