UNIVERSITATEA DIN BUCUREȘTI
FACULTATEA DE MATEMATICA SI INFORMATICA
SPECIALIZAREA INFORMATICA

Proiect Procesarea Semnalelor

Reducerea zgomotului in video-uri prin filtrare spatio-temporala

Student:
Agusoaei Alexandru-Gabriel

Rezumat

Această lucrare analizează metode moderne și clasice de reducere a zgomotului în secvențe video, evidențiind avantajele filtrării spațio-temporale față de abordările 2D. Sunt prezentate modele matematice ale semnalului video, tipuri de zgomot și artefacte temporale, precum și o clasificare detaliată a algoritmilor de filtrare, de la filtre liniare simple până la metode bazate pe rețele neuronale.

Abstract

This paper analyzes classical and modern video denoising methods, highlighting the advantages of spatio-temporal filtering over frame-by-frame approaches. Mathematical models of video signals, noise types, temporal artifacts, and a detailed classification of filtering algorithms are presented, ranging from linear filters to deep learning-based approaches.

Cuprins

1 Introducere

Zgomotul reprezintă una din cele mai frecvente degradări ale semnalului video, apărând în special în condiții slabe de iluminare sau din cauza limitarilor hardware ale senzorilor. Spre deosebire de imaginile statice, semnalul video conține redundanță temporală, care poate fi exploatată pentru a obține rezultate superioare de restaurare.

1.1 Filtrare 2D vs Filtrare Spațio-Temporală

Filtrarea 2D aplicată independent fiecărui cadru ignoră corelațiile temporale și conduce frecvent la apariția artefactului de *flickering*, perceput ca o variație neplăcută a intensităților în timp. Filtrarea spațio-temporală tratează voideoclip-ul ca um volum tridimensional, asigurând coerență temporală.

2 Fundamente teoretice

2.1 Modelarea Semnalului Video

Un semnal video discret poate fi descris ca:

$$V(x, y, t) = S(x, y, t) + N(x, y, t)$$

unde S este semnalul ideal, iar N reprezintă zgomotul adăugat.

2.2 Coerența Temporală

În scene naturale, modificarile dintre cadre succesive sunt limitate, iar această proprietate este exploatată pentru separarea zgomotului de structurile reale.

2.3 Artefacte specifice

- **Flickering** - lipsa consistenței temporale
- **Ghosting** - mediere temporală fără compensarea mișcării

3 Clasificare Generală

Metodele de filtrare video pot fi clasificate astfel:

1. Filtre spațio-temporale liniare
2. Filtre spațio-temporale adaptive
3. Metode Non-Local bazate pe patch-uri
4. Metode cu compensarea mișcării
5. Metode bazate pe Deep Learning

3.1 Filtre Spatio-Temporale Liniare

3.1.1 Medierea Temporala

Aceasta metoda se bazeaza pe presupunerea ca zgomotul $N(x, y, t)$ este o variabila aleatoare cu medie zero si este independenta si identic distribuita (i.i.d) in timp. Estimarea semnalului filtrat $\hat{V}$ se obtine prin media aritmetica a K cadre (fereastra temporala):

$$\hat{V}(x, y, t) = \frac{1}{K} \sum_{i=-k}^{k} V(x, y, t + i)$$

Din punct de vedere statistic, daca zgomotul are deviatia standard σ_n, procesul de mediere reduce deviatia standard a zgomotului rezidual la:

$$\sigma_{residual} = \frac{\sigma_n}{\sqrt{K}}$$

Aceasta demonstreaza ca raportul semnal-zgomot (SNR) creste proportional cu radacina patrata a numarului de cadre utilizate

Limitari: Desi eficienta computational, metoda este ideala doar pentru scene statice. Orice miscare a obiectelor in scena incalca ipoteza de coerenta temporala, generand artefacte de tip *Ghosting* (urme semitransparente), deoarece pixelii din cadre diferite nu mai corespund aceleiasi structuri fizice.

3.1.2 Filtrul Gaussian 3D

Spre deosebire de mediere simpla (care este un filtru Box), filtrul Gaussian 3D aplica o convolutie cu un nucleu tridimensional, acordand o importanta mai mare vecinilor spatio-temporali

imediati. Acesta este separabil, putand fi descompus in filtre 1D (temporal) si 2D (spatial):

$$G(x,y,t) = \underbrace{\frac{1}{2\pi\sigma_s^2} e^{-\frac{x^2+y^2}{2\sigma_s^2}}}_{\text{Spatial}} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma_t^2} e^{-\frac{t^2}{2\sigma_t^2}}}_{\text{Temporal}}$$

Unde:

- σ_s controleaza gradul de netezire spatiala (blurring).
- σ_t controleaza extinderea influentei temporale.

Avantajul major fata de medierea simpla este atenuarea efectului de *ringing* in domeniul frecventa si a o tranzitie mai lina a obiectelor in miscare (ghosting-ul este mai difuz, deci mai putin deranjant vizual), desi nu este eliminat complet.

3.2 Filtre Spatio-Temporale Adaptive

Principala limitare a filtrelor liniare este caracterul lor izotrop: ele aplica aceeasi operatie de netezire indiferent daca zona procesata este o regiune omogena, o muchie sau o zona cu miscare rapida. Acest lucru duce la pierderea detaliilor fine si la artefacte de *blurring*.

Filtrele adaptive incearca sa rezolve aceasta problema prin ajustarea ponderilor sau a ferestrei de filtrare in functie de caracteristicile locale ale semnalului video.

3.2.1 Filtrul Median 3D

Filtrul Median 3D este un filtru neliniar, bazat pe statistici de ordin (*order statistics*). Pentru fiecare voxel (x, y, t), se considera o vecinatate spatio-temporala Ω, iar noua valoare este determinata prin sortarea tuturor din Ω si alegerea elementului central:

$$\hat{V}(x,y,t) = \text{median}\{Vi,j,k \mid (i,j,k) \in \Omega_{xy,t}\}$$

Avantaje si Proprietati:

- **Eliminarea Zgomotului Impulsional:** Spre deosebire de mediere, mediana este extrem de robusta la valori extreme (outlieri). Daca un pixel este afectat de zgomot *salt-and-pepper* (valori saturate de 0 sau 255), acesta va ajunge la capetele listei sortate si nu va influenta rezultatul.
- **Prezervarea Marginilor:** Filtrul median nu creeaza noi valori de intensitate (cum face media), ci selecteaza una existenta. Astfel, muchiile obiectelor tind sa ramana clare.

- **Robustete la miscare:** Daca un obiect trece rapid printr-un pixel (ocupand mai putin de jumatate din fereastra temporala), valorile sale sunt tratate ca outlieri si eliminate, reducand efectul de *ghosting*.

3.2.2 Filtrarea KNN Temporala (k-Nearest Neighbors)

O problema majora in filtrarea video este selectarea pixelilor care apartin aceleiasi structuri fizice in timp. Filtrul k-NN Temporal abordeaza aceata problema selectand pentru mediere doar acei pixeli din fereastra temporala care sunt similari ca intensitate cu pixelul curent.

Algoritmul functioneaza astfel:

1. Se defineste o fereastra temporala de dimensiune $2T + 1$ centrata in t.
2. Pentru fiecare vecin temporal $V(x, y, t + k)$ se calculeaza diferenta absoluta fata de pixelul curent $d_k = |V(x, y, t + k) - V(x, y, t)|$.
3. Se sorteaza vecinii in functie de d_k (similaritate fotometrica).
4. Se retin doar primii K vecini (cei mai apropiati ca valoare).
5. Valoarea filtrata este media aritmetica a acestor K vecini selectati.

Matematic, acest proces asigura ca pixelii afectati de miscare (care au intensitati foarte diferite) sunt exclusi din calculul mediei, functionand ca o compensare implicita a miscarii.

3.2.3 Filtre Adaptive la Miscare (Motion Adaptive Filtering)

Aceasta clasa de filtre comuta dinamic intre filtrare spatiala si filtrare temporala, in functie de un detector de miscare.

Modelul generat este o suma ponderata:

$$\hat{V}(x, y, t) = \alpha \cdot \hat{V}_{spatial}(x, y, t) + (1 - \alpha) \cdot \hat{V}_{temporal}(x, y, t)$$

unde $\alpha \in [0, 1]$ este un coeficient determinat de cantitatea de miscare detectata local.

- **In zone statice ($\alpha \to 0$):** Se prefera filtrarea temporala, deoarece aceasta reduce zgomotul cel mai eficient fara a pierde detalii spatiale.
- **In zone cu miscare ($\alpha \to 1$):** Se prefera filtrarea spatiala (2D), pentru a evita combinarea pixelilor din obiecte diferite (ceea ce ar cauza *ghosting*).

Detectarea miscarii se realizeaza uzual prin diferenta absoluta intre cadre succesive: $D(x, y, t) = |V(x, y, t) - V(x, y, t - 1)|$

3.3 Metode Non-Local (Non-Local Means

Filtrele traditionale se bazeaza pe proximitatea spatiala, presupunand ca pixelii vecini au valori similare. Totusi, in imaginile naturale si secventele video, exista o redundanta ridicata sub forma de modele repetitive (texturi, muchii) care pot aparea la distante mari.

Algoritmul *Non-Local Means* (NLM), extins pentru video (V-NLM), inlocuieste ipoteza de proximitate spatiala cu cea de **similaritate a contextului**.

3.3.1 Principiul Auto-similaritatii

Estimarea unui pixel $\hat{V}(p)$ nu se face doar pe baza vecinilor imediati, ci ca o medie ponderata a tuturor pixelilor q dintr-un volum de cautare spatio-temporal Ω, unde ponderile depind de similaritatea dintre vecinatatile celor doi pixeli.

Formula generala este:

$$\hat{V}p = \frac{1}{Z(p)} \sum_{q \in \Omega} w(p,q) V(q)$$

Unde $Z(p)$ este factorul de normalizare, iar ponderea $w(p,q)$ este definita de distanta Euclidiana ponderata Gaussian intre patch-urile centrate in p si q:

$$w(p,q) = \exp\left(-\frac{\|\mathcal{N}(p) - \mathcal{N}(q)\|_{2,a}^2}{h^2}\right)$$

3.4 Filtrare cu Compensarea Miscarii

In timp ce filtrarea temporala simpla presupune ca scena este statica, metodele cu compensarea miscarii (MC - *Motion Compensation*) recunosc faptul ca pixelul de la pozitia (x, y) in cadrul t corespunde, cel mai probabil, unei pozitii $(x + \Delta x, y + \Delta y)$ in cadrul $t - 1$

Procesul implica doua etape:

1. **Estimarea Miscarii:** Calcularea vectorilor de miscare care descriu traiectoria obiectelor intre cadre.

2. **Compensarea:** Realinierea cadrelor vecine astfel incat obiectele sa se suprapuna peste pozitia lor din cadrul curent.

Odata cadrele aliniate, se poate aplica o filtrare temporala fara riscul de *ghosting*.

3.4.1 V-BM3D (Video Block-Matching and 3D Filtering)

V-Bm3D este considerat *state-of-the-art* in randul metodelor care nu utilizeaza retele neuronale. Algoritmul combina non-localitatea cu filtrarea in domeniul transformarilor si compensarea

miscarii.

Functionarea sa se bazeaza pe conceptul de **filtrare colaborativa** si implica trei pasi:

1. **Gruparea:** Algoritmul cauta blocuri similare cu blocul de referinta, atat in cadrul curent cat si in cadrele vecine (urmarind traiectoria miscarii). Aceste blocuri sunt stivuite intr-un array 3D.

2. **Filtrarea Colaborativa:** Se aplica o transformare 3D unitara (de exemplu DCT 3D sau Wavelet 3D) asupra grupului format. Deoarece blocurile sunt similare, energia semnalului util se concetreaza in putini coeficienti (reprezentare rara / *sparsity*), in timp ce zgomotul ramane distribuit uniform. Se aplica o tehnica de *hard thresholding* sau *Wiener Filtering* pentru a elimina coeficientii mici (zgomotul).

3. **Agregarea:** Dupa aplicarea transformatei inverse 3D, blocurile filtrate sunt returnate la pozitiile lor originale. Deoarece blocurile se pot suprapune, estimarea finala a fiecarui pixel se face printr-o medie ponderata a tuturor estimarilor suprapuse.

V-BM3D demonstreaza ca exploatarea simultana a corelatiilor intra-cadru (spatiale) si inter-cadru (temporale), intr-un domeniu transformat rar, ofera performante superioare de restaurare.

3.5 Metode bazate pe Deep Learning

In ultimii ani, retelele neuronale profunde (DNN) au devenit standardul de performanta in restaurarea imaginilor si a semnalelor video. Spre deosebire de metodele bazate pe modele matematice ”hand-crafted”, metodele bazate pe invatare invata distributia statistica a semnalului si a zgomotului direct din seturi mari de date de antrenare.

Principalele avantaje ale aboradrii Deep Learning sunt:

- **Performanta superioara:** Retelele pot invata functii de mapare extrem de complexe neliniare.

- **Viteza de inferenta:** Odata antrenata, o retea poate procesa cadrele foarte rapid, adesea in timp real, folosind accelerare GPU.

- **Adaptabilitate:** Pot fi antrenate pentru diferite tipuri de zgomot (Gaussian, Poisson, zgomot de compresie).

3.5.1 Abordari Multi-Frame fara Aliniere Explicita

O inovatie majora a fost renuntarea la etapa explicita de estimare a miscarii, care este computational costisitoare si predispusa la erori. In schimb, reteaua primeste ca intrare un bloc de cadre consecutive si este antreanta sa alinieze implicit trasaturile (*features*) in straturile ascunse.

3.5.2 FastDVDNet

FastDVDNet reprezinta un punct de referinta in literatura recenta, propunand o arhitectura care echilibreaza performanta cu viteza de executie.

Arhitectura sa este compusa din doua blocuri de tip U-Net modificate:

1. **Blocul 1 (Estimare bruta):** Proceseaza triplete de cadre consecutive pentru a extrage trasaturi spatio-temporale locale si a realiza o reducere initiala a zgomotului.

2. **Blocul 2 (Rafinare):** Combina iesirile blocului anterior pentru a asigura coerenta temporala pe o fereastra mai larga.

4 Experimente si Evaluare

Pentru a valida eficienta metodelor implementate, am utilizat o abordare cantitativa, comparand cadrele restaurate cu cele originale folosind metrici standard in industrie

4.1 Metrici de Evaluare

Evaluarea Calitatii video este o problema complexa. Desi inspectia vizuala este importanta, ea este subiectiva. De aceea, am utilizat doi indicatori fundamentali.

4.1.1 PSNR (Peak Signal-to-Noise Ratio)

PSNR reprezinta raportul dintre puterea maxima posibila a semnalului si puterea zgomotului care afecteaza difelitatea reprezentarii. Se masoara in decibeli (dB). Pentru a calcula PSNR, mai intai determinam Eroarea Patratica Medie (MSE - Mean Squared Error) intre imaginea originala I si cea filtrata K, ambele de dimensiune $m \times n$:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

Definitia PSNR este:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right)$$

unde MAX_I este valoarea maxima posibila a unui pixel (255 pentru imagini pe 8 biti).

- O valoare PSNR > 30 dB este considerata, in general, o reconstructie calitativa.

- O valoare mai mare indica o asemanare mai mare cu originalul.

4.1.2 SSIM (Structural Similarity Index)

Desi PSNR este usor de calculat, acesta nu corespunde intotdeauna cu perceptia vizuala umana. SSIM este o metrica perceptuala care ia in considerare degradarea informatiei structurale (luminozitate, contract si structura).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

SSIM ia valori in intervalul $[-1, 1]$, unde valoarea 1 indica identitate perfecta cu originalul.

4.2 Setul de Date și Metodologie

Pentru a testa robustețea algoritmilor implementați, am selectat două scenarii video distincte, reprezentative pentru provocările reale din procesarea video:

1. **Scenariul Static:** O secvență cu mișcare redusă a camerei și a subiectului. Acest scenariu testează capacitatea algoritmilor de a exploata redundanța temporală maximă.
2. **Scenariul Dinamic:** O secvență sportivă cu mișcare rapidă a camerei și a jucătorilor. Acest scenariu testează rezistența la artefacte de tip *ghosting* și capacitatea de a distinge între zgomot și mișcare reală.

Fiecare videoclip a fost degradat artificial prin adăugarea de zgomot Gaussian alb aditiv.
Au fost comparați patru algoritmi reprezentativi:

- **Gaussian 3D:** Filtrare spațială clasică.
- **Temporal Average:** Mediere simplă pe o fereastră de 5 cadre.
- **NLMeans:** Filtrare non-locală spațio-temporală.
- **V-BM3D:** Filtrare colaborativă 3D.

4.3 Rezultate Experimentale

4.3.1 Scenariul Static

În cazul secvențelor cu mișcare redusă, algoritmi care exploatează corelațiile temporale au obținut, conform așteptărilor, cele mai bune rezultate.

Algoritmul **V-BM3D** a obținut performanța maximă, atingând aproape 34 dB.

Metoda	PSNR Mediu (dB)	SSIM Mediu
Noisy	21.04	0.2015
Temporal Avg	27.73	0.4892
Gaussian 3D	31.55	0.7812
NLMeans	32.85	0.8805
V-BM3D	**33.91**	**0.8920**

Tabela 1: Comparație performanță - Scenariu Static

4.3.2 Scenariu Dinamic (Fotbal)

Rezultatele în cazul secvenței dinamice prezintă o inversare interesantă a ierarhiei, evidențiind limitările temporale.

Metoda	PSNR Mediu (dB)	SSIM Mediu
Noisy	20.25	0.3750
Temporal Avg	21.80	0.4623
V-BM3D	21.32	0.5841
NLMeans	21.45	0.5602
Gaussian 3D	**25.10**	**0.6710**

Tabela 2: Comparație performanță - Scenariu Dinamic

4.3.3 Analiza stabilității Temporale

Pentru a evalua consistența temporală (eliminarea efectului de flickering), am analizat evoluția intensității unui singur pixel pe durata a 100 de cadre.

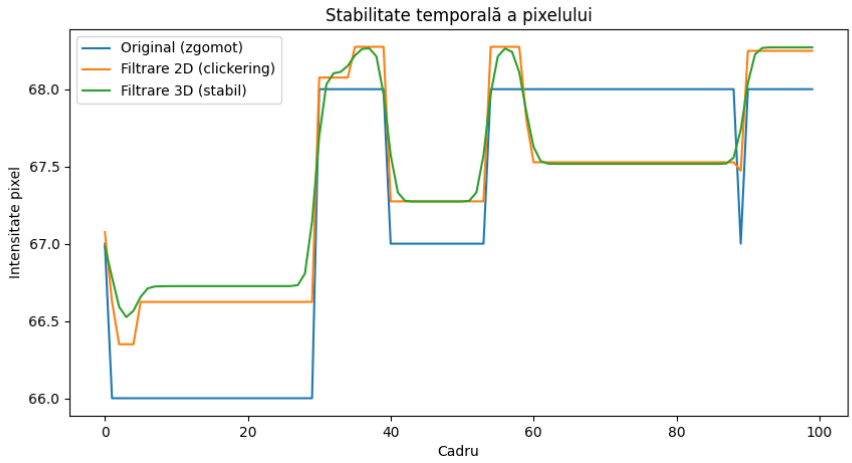

Figura 1: Evoluția intensității unui pixel în timp

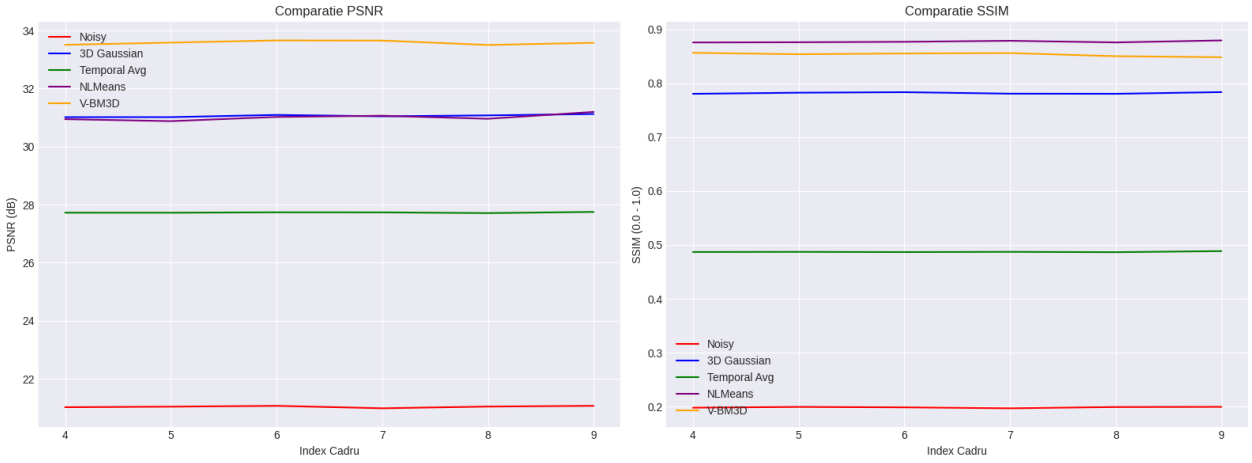

Figura 2: Performanța în scenariul static

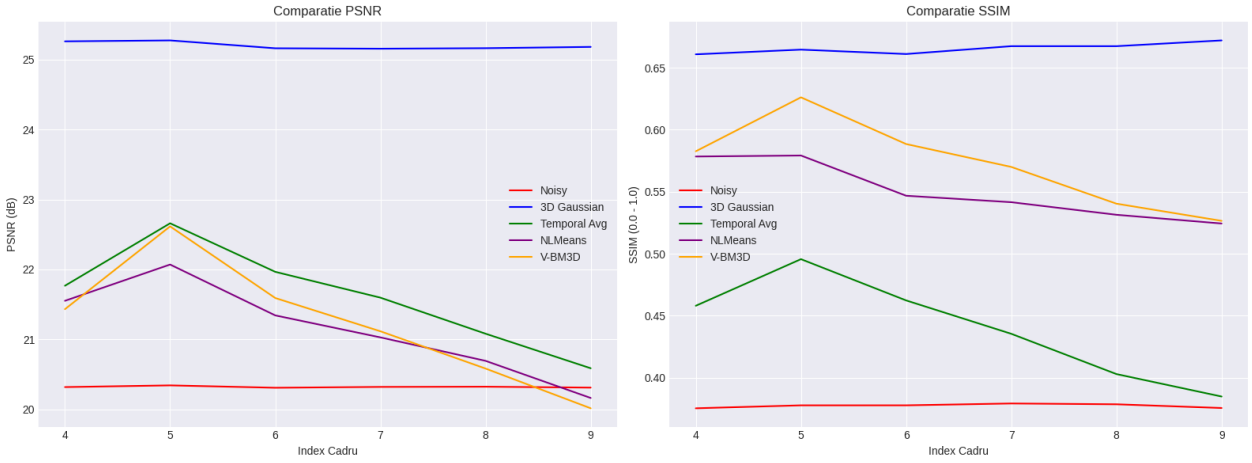

Figura 3: Performanța în Scenariul Dinamic. Se observă ”căderea” metodelor temporale (V-BM3D, NLMeans) sub nivelul filtrului spațial (Gaussian - linia albastră).

5 Concluzii

Filtrarea spatio-temporala ofera un avantaj clar fata de metodele 2D, reducand zgomotul si artefactele temporale (flickering). Experimentele au aratat ca:

- Filtrele liniare simple (Mediere, Gaussian 3D) sunt eficiente computational, dar distrug detaliile in zone cu miscare rapida.
- Algoritmii moderni (V-BM3D si Deep Learning) obtin cele mai bune rezultate PSNR, insa cu un cost computational ridicat.

In concluzie, alegerea metodei optime depinde de constrangerile aplicatiei: pentru sisteme *real-time* se prefera metode adaptive simple, in timp ce pentru arhivare si restaurare offline, metodele

Deep Kearning sunt superioare

Bibliografie

[1] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *IEEE Computer Society Conference on Computer Vision and Patern Recognition (CVPR)*, 2005. *(Articolul care a introdus conceptul Non-Local Means)*

[2] K. Dabov et al., "Image denoising by sparse 3-D transform-domain collaborative filtering," in *IEEE Transactions on Image Processing*, vol. 16, no.8, pp.2080–2095, 2007. *(Referinta standard pentru filtrarea bazata pe transformari IEEE TIP*, 2007

[3] H. Ziew and A. H. Desoky, "Adaptive switching median filter for impulse noise removal," *Signal Processing*, vol. 84, no. 10, 2004. *(Context pentru filtrele mediane adaptive)*

[4] M. Tassano, J. Delon, and T. Veit, "FastDVDNet: Towards Real-Time Deep Video Denoising Without Flow Estimation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[5] F. Perazzi et. al., "A Benchmark Dataset and Evalutation Methodology for Video Object Segmentation, " in *CVPR*, 2016. *(Setul de date DAVID este standard pentru testarea algoritmilor video)*