7. AUTENTIFICAREA CONŢINUTULUI MULTIMEDIA

Semnalele multimedia pot fi uşor reproduse, modificate şi falsificate. Chiar dacă nu putem percepe aceste modificări, ceea ce vedem sau ascultăm poate să fi fost modificat sau alterat în mod intenţionat. În anumite cazuri, aceste modificări efectuate în conţinutul multimedia pot avea un efect negativ major prin expunerea lor în mass-media, în sălile de judecată, în campanii politice, etc. Autentificarea conţinutului multimedia, un subdomeniul al securităţii multimedia, îşi propune să determine veridicitatea conţinutului multimedia. Un sistem de autentificare multimedia trebuie să răspundă la următoarele întrebări:

- A fost modificat conţinutul în vreun fel?
- Dacă a fost modificat, unde şi în ce măsură a fost modificat?

Există în principal două abordări care pot da răspunsurile la aceste întrebări: sistemele de autentificare ce utilizează tehnici de watermaring digital şi sisteme de autentificare bazate pe proprietăţile statistice ale materialului multimedia. Metodele din prima categorie se bazează pe inserarea unui watermark sau a unei semnături invizibile în materialul multimedia original, iar cele din a doua categorie nu necesită inserarea de informaţii adiţionale, detectând eventuale modificări din proprietăţile intrinseci ale conţinutului multimedia. În acest capitol vom discuta în special despre autentificarea imaginilor şi secvenţelor video. Tehnicile din a doua categorie pot fi clasificate după cum urmează:

- Tehnici la nivel de pixel, ce detectează anomalii statistice introduse la nivel de pixel
- Tehnici pe baza formatului, ce exploatează corelaţiile statistice introduse de schemele de compresie
- Tehnici bazate pe analiza artefactelor introduse de camerele digitale prin intermediul lentilelor, senzorilor sau procesărilor interne

- Tehnici ce exploatează anomalii între interacţiunea tridimensionala dintre obiecte fizice, lumină si cameră
- Tehnici geometrice prin care se realizează măsurători ale obiectelor din scenă şi poziţia lor relativă faţă de cameră.

Metodele ce utilizează tehnici de watermarking pot detecta modificările de conţinut, independent de modul în care ele au fost realizate aceste modificări, oferind o probabilitate mai bună de detecţie corectă, pe când metodele din a doua categorie pot detecta modificări realizate într-un singur mod (dublă compresie, iluminare inconsecventă a obiectelor din scenă, înlocuire de obiecte din scenă, etc), fiind necesară aplicarea unui număr mare de astfel de metode pentru a realiza o detecţie corectă.

Acest capitol îşi propune prezentarea celor mai importante aspecte şi tehnici legate de autentificarea imaginilor şi secvenţelor video folosind metode de watermarking digital.

Autentificarea conţinutului digital este una dintre aplicaţiile de bază ale watermarking-ului digital. În comparaţie cu alte aplicaţii, ca de exemplu protecţia drepturilor de autor sau amprentarea, obiectivul metodelor de autentificare este verificarea integrităţii conţinutului digital şi detecţia eventualelor manipulări sau modificări ale datelor gazdă. Autentificarea conţinutului multimedia este o problemă actuală în România, în special din cauza faptului că ţara noastră are de suferit de pe urma corupţiei la diferite niveluri. În ultimul timp, o serie de procese ce implicau imagini şi secvenţe video ca probe importante nu au putut fi rezolvate din cauza lipsei de sisteme eficiente de autentificare.

În Figura 7.1 este prezentată schema unui sistem de autentificare a imaginilor bazat pe watermarking digital. Un astfel de sistem de autentificare implică două operaţii: una de inserare şi una de detecţie a watermark-ului. În sistemele de watermarking pentru autentificare, watermark-ul este folosit ca un cod de autentificare sub forma unei secvenţe aleatoare de biţi, unui logo vizual binar sau unor caracteristici legate de conţinut. Acest cod de autentificare este inserat în mod imperceptibil în imaginea sau secvenţa video originală. Chiar dacă este transparent pentru ochiul uman, codul poate fi extras, în anumite circumstanţe, de către detectorul de watermark. Codul extras este comparat cu cel original pentru a verifica integritatea imaginii sau secvenţei video. Eventualele diferenţe dintre cele două indică faptul că datele originale au fost manipulate.

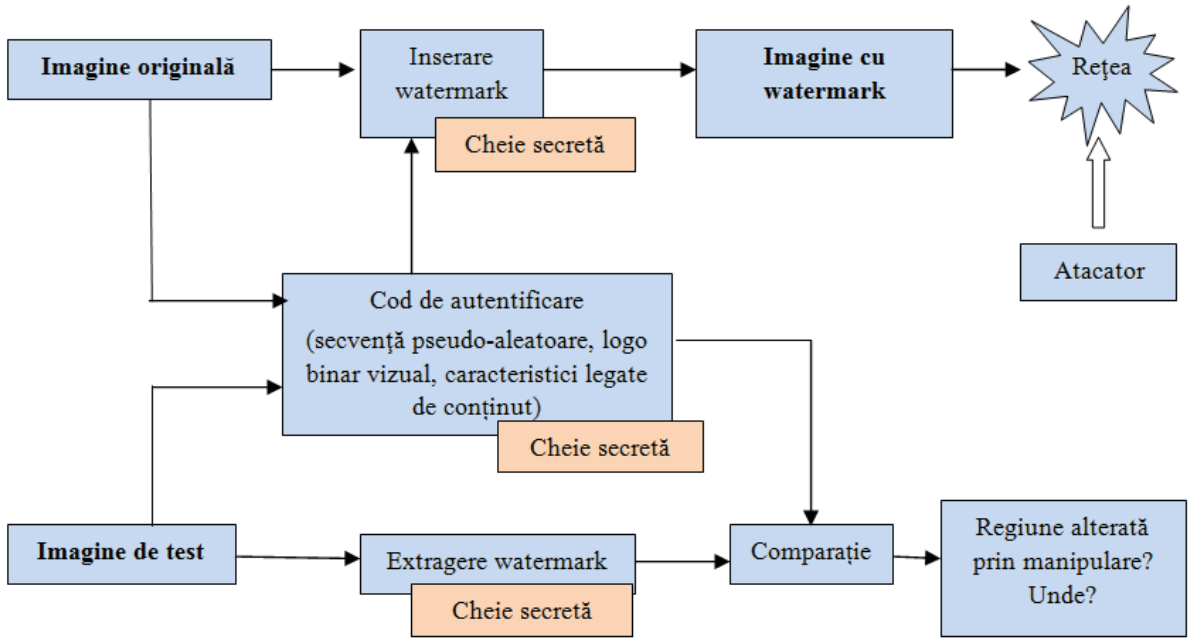

Figura 7.1. Schema unui sistem de autentificare a imaginilor digitale bazat pe watermarking digital

Pentru un sistem eficient de watermarking pentru autentificarea conţinutului multimedia, următoarele caracteristici sunt dorite:

- Transparenţa perceptuală: watermark-ul ar trebui să fie invizibil pentru sistemul vizual uman; altfel spus, trebuie păstrată o calitate înaltă a imaginilor sau secvenţelor video.
- Capacitatea de a detecta dacă o imagine a fost manipulată în mod voit;
- Capacitatea de a localiza modificarea cu o precizie bună: autentificatorul trebuie să poată identifica locaţia regiunilor manipulate cu o anumită rezoluţie şi, pe de altă parte, să poată verifica autenticitatea altor regiuni;
- Compatibilitatea: sistemul de autentificare trebuie să poată supravieţui, într-o anumită măsură, în urma distorsiunilor neintenţionate cauzate de procesări uzuale de imagini, de exemplu să poată deosebi distorsiunile neintenţionate de manipulări intenţionate, maliţioase;
- Portabilitate: informaţia de autentificare ar trebui inserată în imaginea/secvenţa video gazdă şi să nu fie necesare date adiţionale separate;
- Securitatea împotriva falsificării şi a operaţiunilor neautorizate

Pentru a proiecta un sistem eficient de watermarking, trebuie găsit un compromis rezonabil între aceste caracteristici, în conformitate cu cerinţele specifice aplicaţiei.

În literatura de specialitate au fost propuse o serie de tehnici de watermarking pentru autentificarea imaginilor şi secvenţelor video. În funcţie de tipul de autentificare, pe care îl oferă, ele pot fi clasificate în două mari categorii: watermarking pentru autentificare exactă/brută şi watermarking pentru autentificare selectivă/fină. Prima categorie de algoritmi utilizează de regulă watermark-uri fragile iar cea de-a doua categorie watermark-uri semi-fragile.

7.1. Autentificarea exactă (brută)

Watermark-urile fragile permit o autentificare strictă fără toleranţă pentru manipularea conţinutului. O modificare a unui singur bit va face imaginea sau secvenţa video neautentică. Multe tehnici de watermarking fragile utilizează tehnici criptografice pentru a atinge un nivel de securitate înalt.

Cel mai simplu algoritm de marcare fragilă este metoda LSB (Least Significant Bit), în care cei mai puţin semnificativi biţi sunt modificaţi pentru a insera watermark-ul. Schimbarea LSB trece ca neobservabilă din punct de vedere perceptual, de aceea este posibilă inserarea imperceptibilă a watermark-ului.

În [193] este propusă o schemă de autentificare LSB cu watermark fragil inserat cu ajutorul unei chei publice. Imaginea este împărţită în blocuri iar în fiecare bloc planul cel mai puţin semnificativ de biţi este înlocuit cu rezultatul operaţiei de SAU exclusiv (XOR) dintre biţii watermark-ului şi suma hash a imaginii. Suma hash este obţinută din dimensiunile imaginii şi restul valorilor pixelilor din bloc mai puţin bitul cel mai puţin semnificativ, care sunt toţi trecuţi printr-o funcţie hash. Rezultatul funcţiei XOR este criptat înainte de inserare folosind cheia privată a utilizatorului. La detecţie, planul de biţi corespunzător biţilor celor mai puţin semnificativi este decriptat folosind cheia publică corespunzătoare. Apoi, pentru fiecare bloc, fiecare bit al watermark-ului este recuperat calculând XOR între rezultatul decriptării şi suma hash recalculată din acelaşi bloc. Toţi biţii watermark-ului recuperat sunt comparaţi cu cei ai watermark-ul original şi dacă toţi sunt identici atunci imaginea marcată este declarată autentică. Dacă nu, fiecare bit care nu corespunde cu cel original, va indica faptul că blocul respectiv a fost alterat.

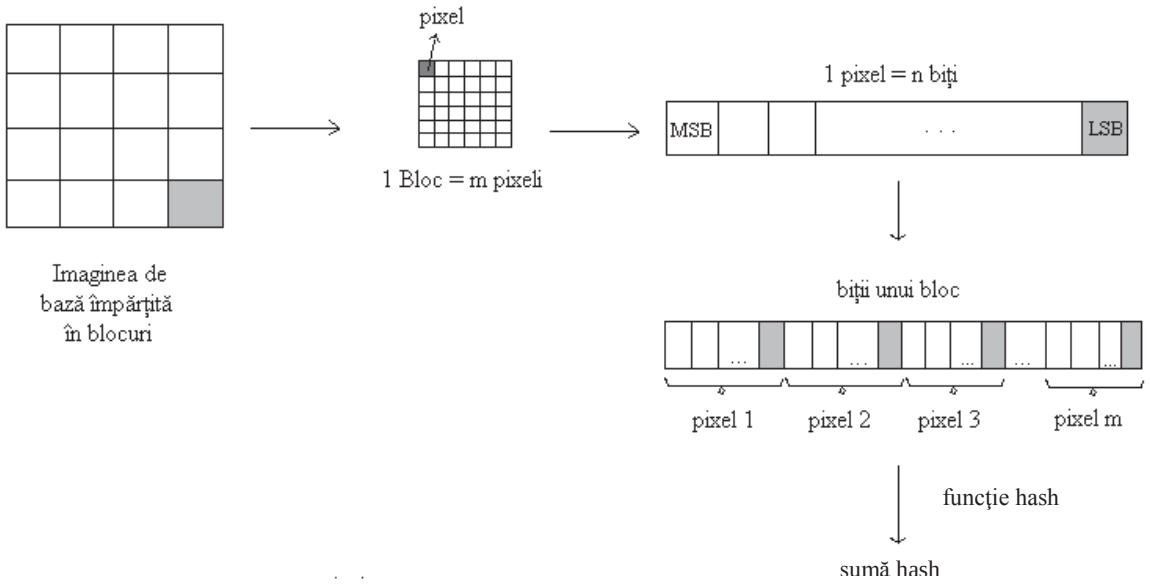

Figura 7.2. Schema de obţinere a sumei hash

Din cauza faptului că algoritmul autentifică fiecare bloc din imagine, acesta este susceptibil atacurilor de cuantizare vectorială, care înlocuieşte blocuri din imagine între ele sau din diferite imagini cu watermark. Pentru a combate acest tip de atac pot fi folosite blocuri suprapuse, în loc de blocuri disjuncte, sau se poate încorpora un identificator de bloc în suma hash.

Un alt algoritm folosit este cel propus în [194], numit şi schema Yeung-Mintzer. Algoritmul foloseşte o funcţie binară (tabel look-up, LUT – Look-Up Table) pentru a forţa fiecare pixel să se asocieze câte unui bit dintr-un set binar aleator de date. La inserare se foloseşte un proces de difuzie a erorii în scopul îmbunătăţirii calităţii imaginii marcate. Având în vedere faptul că inserarea watermark-ului se face la nivel de pixel, alterările pot fi localizate tot la nivel de pixel, însemnând o creştere a rezoluţiei localizării. Un dezavantaj apare din cauza folosirii aceleiaşi funcţii de marcare la mai multe imagini, chiar dacă secvenţa e secretă, favorizând atacurile de cuantizare vectorială.

O îmbunătăţire a schemei Yeung-Mintzer, este dezbătută în [195] care propune evitarea atacurilor de cuantizare vectorială prin luarea în considerare a dependenţei de pixelii vecini la încorporare.

În [196], se studiază securitatea imaginilor marcate ce au capacitatea de localizare a alterării. Autorii au ajuns la concluzia că atacurile de cuantizare vectorială sunt permise

tocmai de marcarea la nivel de pixel. De aceea, este propusă o schema ce presupune marcarea la nivel de bloc şi nu de pixel. Dezavantajul este că scade rezoluţia capacităţii de localizare.

În [197], este propusă o structură ierarhică pentru marcare. Marcarea la nivelele inferioare are avantajul de a oferi o capacitate sporită de localizare, în timp ce marcarea la nivele superioare ale ierarhiei oferă o protecţie sporită în faţa atacurilor de cuantizare vectorială.

În [198], algoritmul propune crearea watermark-ului din primii 5 cei mai semnificativi biţi ai fiecărui pixel în combinaţie cu o secvenţă aleatoare. Watermark-ul astfel format este inserat în ultimii 3 biţi ai fiecărui pixel, adică în 3 cei mai puţin semnificativi biţi. Dezavantajul algoritmului este că nu poate detecta alterările în cei 3 mai puţin semnificativi biţi. În plus, înlocuind tot planul de biţi corespunzător fiecăruia din cei 3 biţi calitatea imaginii cu watermark este redusă.

Pe lângă toate schemele de marcare propuse în domeniul spaţial, câteva amintite mai sus, în literatura de specialitate au fost dezbătute şi scheme în domeniul transformat cum ar fi cel al Transformatei Cosinus Discrete [198] şi cel al Transformatei Wavelet Discrete [199]. Avantajele inserării watermark-ului în domeniul transformat se referă în primul rând la compatibilitatea cu standardele de compresie uzuale cum ar fi compresia JPEG. Marcarea poate fi făcută în timpul compresiei sau poate fi aplicată ulterior imaginii comprimate. Un alt avantaj e oferit de un control mai bun al distorsiunilor perceptuale decât în domeniul spaţial. Astfel, se îmbunătăţeşte calitatea imaginii marcate. În plus, distorsiunile pot fi localizate atât spaţial cât şi temporal. Dezavantajele folosirii marcării în domeniul transformat sunt în primul rând faptul că marcarea făcându-se în domeniul transformat, mici modificări ale pixelilor în domeniul spaţial pot să nu fie detectate de schemele de autentificare în domeniul transformat. Un alt dezavantaj este faptul că precizia localizării alterării depinde de dimensiunile blocurilor din imagine pe cere se face analiza (de exemplu cazul DCT) [199].

7.2. Autentificarea selectivă (fină)

Pentru ca autentificatorul să poată deosebi manipulările întâmplătoare de cele maliţioase, au fost propuse tehnici de watermarking semi-fragile. În comparaţie cu autentificarea exactă oferită de watermark-urile fragile, aceste tehnici oferă o autentificare selectivă/fină. Ele monitorizează mai degrabă conţinutul imaginilor sau cadrelor video decât

reprezentarea digitală a acestora şi permit astfel mici modificări cauzate de operaţii uzuale de procesare de imagini sau video ca de exemplu compresie JPEG sau MPEG cu pierderi mici, filtrare şi îmbunătăţire de contrast, dar pot detecta manipulări ce modifică conţinutul, ca de exemplu inserarea, ştergerea sau înlocuirea de obiecte. Watermark-urile semi-fragile sunt inserate de obicei într-un domeniu transformat pentru a obţine o robusteţe moderată, calitate perceptuală bună şi compatibilitate cu standardele de compresie. Cele mai des utilizate transformate sunt DCT şi DWT, deoarece acestea sunt folosite şi la cele mai populare standarde de compresie ca JPEG, JPEG2000, MPEG-1-4, H.26x. În plus, modele ale sistemului vizual uman (SVU) realizate în aceste domenii pot fi folosite pentru a controla în mod adaptiv puterea de inserarea a watermark-ului pentru a îmbunătăţi calitatea perceptuală a imaginilor/cadrelor cu watermark. Proprietăţile transformatei wavelet, în special, oferă posibilităţi foarte interesante de localizare spaţială a manipulărilor în procesul de autentificare a conţinutului.

Schemele semi-fragile de watermarking utilizează în principal două categorii de tehnici de inserare. Prima dintre ele este metoda cu spectru împrăştiat [200, 201], care a fost propusă pentru prima oară de către Cox în [44]. Aceste metode înserează watermark-ul prin împrăştierea acestuia în spectrul disponibil al semnalului (vezi Figura 7.3).

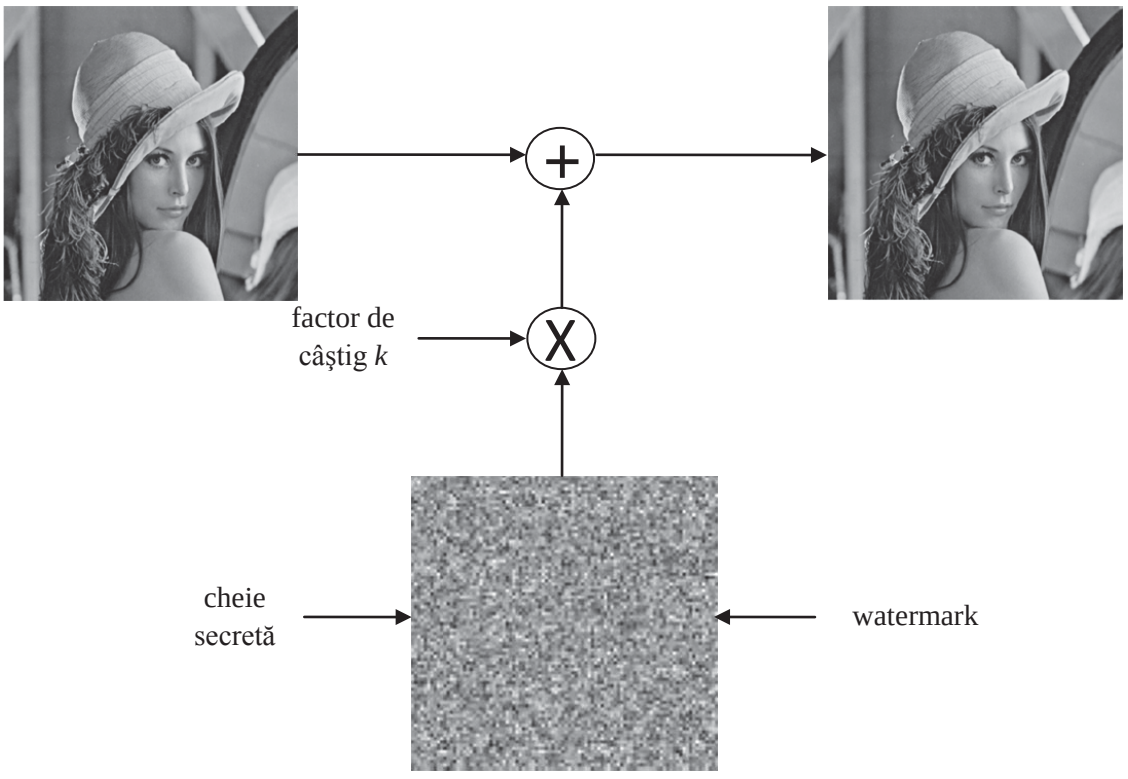

Figura 7.3. Schema de principiu de inserare folosind metoda de watermarking cu spectru împrăştiat

Detecţia este realizată făcând corelaţia între watermark-ul original şi imaginea cu watermark.

Cealaltă metodă populară de inserare este modulaţia indicelui de cuantizare (QIM – Quantization Index Modulation) [193], [202-204]. În cazul acestui tip de metode, watermark-ul este inserat prin cuantizarea unor coeficienţi selectaţi în domeniul spaţial sau transformat. Puterea de inserare şi robusteţea watermark-ului pot fi controlate cantitativ prin intermediul pasului de cuantizare.

Un exemplu de implementare a unei scheme de autentificare selectivă este schema propusă de Fridrich [205, 206]. Imaginea este împărţită în blocuri de mărime medie şi în fiecare bloc este inserat aditiv watermark-ul cu spectru împrăştiat în 30% dintre coeficienţii DCT, cei de frecvenţă medie. Pentru verificarea autenticităţii imaginii se va căuta watermark-ul în fiecare bloc. Dacă watermark-ul este detectat în fiecare bloc cu o rată ridicată atunci putem spune că imaginea nu a fost semnificativ manipulată. Dacă răspunsul detectorului este mai scăzut pentru toate blocurile, putem spune că a fost aplicată o operaţie neagresivă de procesare de semnal. Dacă numai în unele blocuri detecţia e scăzută comparativ cu alte blocuri, se poate estima, în funcţie de răspunsul detectorului, probabilitatea, dacă un bloc a fost alterat. Nu se poate determina cu exactitate în ce bloc este prezentă modificarea, ci numai în ce măsură a fost alterată imaginea. Acest lucru se întâmplă din cauza faptului că blocurile au o dimensiune mare (64x64). Dacă ar fi fost mai mici, atunci ele ar fi generat scăderea performanţei procedeului.

În [207], Lin et al. propune o metodă de autentificare cu watermark semi-fragil folosind metoda modulării indicelui de cuantizare în domeniul DCT. Algoritmul permite compresia cu pierderi JPEG până la un anumit factor de calitate şi e capabil să identifice manipulările maliţioase. Algoritmul foloseşte două proprietăţi ale transformatei DCT: prima – după cuantizarea unui coeficient DCT cu un multiplu întreg al pasului de cuantizare, acesta poate fi recuperat după compresia JPEG, folosind un pas de cuantizare mai mic; a doua – ordinea perechilor de coeficienţi DCT rămâne neschimbată după şi înainte de compresia JPEG. Prima proprietate este folosită pentru marcarea robustă la compresii JPEG acceptabile. A doua proprietate este folosită pentru a genera mesajul de autentificare. În schema propusa de Lin, pentru autentificare şi inserare sunt folosite blocuri de dimensiune 8x8 care nu se suprapun. La autentificare, biţii watermark-ului extras sunt comparaţi cu cei ai watermark-ului regenerat. Localizarea alterării şi recuperarea blocurilor alterate este posibilă numai dacă şi biţii de recuperare sunt inseraţi. Aceasta depinde şi de mărimea blocurilor. Un dezavantaj major, comun tuturor schemelor de autentificare ce folosesc watermark-uri semi-fragile este

acela că este foarte dificil să se insereze watermark-ul în zonele uniforme ale imaginii oferind totodată şi o robusteţe rezonabilă.

Schema propusă în [198] foloseşte inserarea în domeniul DWT. Metoda inserează în coeficienţii DWT o secvenţă pseudo-aleatoare fără legătură cu conţinutul imaginii originale. Imaginea originală este descompusă pe 4 nivele de rezoluţie folosind familia wavelet Haar. Biţii watermark-ului sunt inseraţi în subbenzile celor 4 nivele de descompunere folosind metoda de modulaţie a indexului de cuantizare par-impar. Decizia de asociere a valorilor binare „0" şi „1" coeficienţilor pari sau impari este luată folosind o cheie secretă.

O metodă alternativă la schema anterioară, propusă în [209], presupune inserarea unei secvenţe pseudo-aleatoare de watermark în valoarea medie a coeficienţilor wavelet. Zona alterată este estimată pe baza analizei rezultatelor detecţiei pe mai multe nivele.

În [210], Winne et al. propun o schemă bazată pe modificarea unui algoritm de marcare robust. După aplicarea DWT asupra imaginii, watermark-ul, o secvenţă pseudo-aleatoare, este încorporat în primul nivel de descompunere wavelet. Se construieşte un vector din cei 3 coeficienţi wavelet de detaliu aflaţi la aceeaşi frecvenţă dar de orientări diferite în subbenzile LH, HL, HH. Valoarea coeficientului median este cuantizată în funcţie de bitul watermark-ului folosind un pas adaptiv de cuantizare. Pentru eficienţa sporită se poate folosi şi o etapă de pre-distorsiune. La autentificare se obţine o localizare exactă a alterării cu o mare acurateţe, putându-se determina forma zonei alterate.

În mod ideal un detector de watermark semi-fragil ar trebui să treacă cu vederea modificări neagresive ale conţinutului datorate, de exemplu, editării postproducţie, compresiei uşoare, filtrării, sau operaţiei de îmbunătăţire a contrastului, dar ar trebui să dea alarma atunci când are loc o modificare a conţinutului. Tehnicile de watermarking semi-fragile validează conţinutul multimedia şi nu reprezentarea sa, fiind concepute special pentru a fi robuste la modificări permise şi sensibile la modificări nepermise. Desigur, că, dacă operaţiile de procesare de semnal depăşesc anumite limite, ca de exemplu rate de compresie mari, acestea încep să modifice nu doar modul de reprezentare, ci şi conţinutul documentului în sine, şi în astfel de cazuri ar trebui privite ca atacuri intenţionate. Graniţa dintre un atac neintenţionat şi unul intenţionat nu este bine definită şi depinde de domeniul aplicaţiei şi de tipul de document multimedia. Totuşi, exactitatea detecţiei watermark-ului poate fi folosită ca o măsură a autenticităţii documentului şi se poate stabili autenticitatea acestuia prin stabilirea unui prag.

Tehnicile semi-fragile de detecţie a manipulării şi de autentificare a conţinutului pot fi clasificate in funcţie de modul în care se face autentificarea după cum urmează:

- Tehnici de watermarking semi-fragile cu autentificare vizuală, unde, de regulă, o imagine binară sau un logo vizual este inserat în documentul multimedia, iar detecţia manipulării este bazată pe evaluarea vizuală a diferenţelor percepute de către un operator uman, ca de exemplu la metodele propuse Bhattacharyya în [211] şi Marvel în [212];
- Tehnici de watermarking semi-fragile cu autentificare statistică, la care se obţine un estimat al probabilităţii de manipulare din coeficientul de corelaţie dintre watermark-ul de autentificare inserat şi cel recuperat. La aceste tehnici performanţa poate fi determinată prin probabilitatea de detecţie a secvenţei corecte de watermark ca funcţie de rata de alarmă falsă [213, 214];
- Tehnici cu auto-inserare pentru autentificare şi protecţie a imaginilor, ca cele propuse de Fridrich şi Goljan [215], Lan şi Tewfik [216] pentru imagini Robie şi Mersereau [217] pentru secvenţe video.

Algoritmii de autentificare pot fi cu auto-autentificare sau cu autentificare independentă:

- Algoritmi cu auto-autentificare au la bază validarea unei funcţii hash robuste, ce a fost inserată şi trebuie extrasă din imaginea de test, ca de exemplu în metodele propuse de Piva, Bartolini şi Caldelliy în [218] Lin şi Chang [214], Fridrich [205, 206], Xie, Arce, şi Graverman [219] şi, parţial, Hung, Cheng, şi Chen [220].
- Algoritmi cu autentificare independentă ce stabilesc autenticitatea prin intermediul unei secvenţe de autentificare independentă de imagine, ca în metodele propuse de Eggers [221], Lin, Podilchuk şi Delp [213], Queluz şi Lamy [222], Kundur şi Hatzinakos [223], Lan, Mansour şi Tewfik [224] şi Gwo, Lu şi Liao [225], Ouda şi El-Sakka [226].

Performanţele algoritmilor pot fi măsurate prin probabilitatea de a nu detecta manipularea atunci când ea există şi probabilitatea de alarmă falsă atunci când nu există manipulare intenţionată, dar materialul digital este supus unor operaţii blânde de procesare de semnal. Desigur că se doreşte o probabilitate mică de a nu detecta manipularea atunci când ea există şi, în acelaşi timp, o probabilitate mică de alarmă falsă atunci când avem de a face cu operaţii permise de procesare de semnal. Astfel, algoritmii trebuie comparaţi în ceea ce priveşte semi-fragilitatea lor, adică rezistenţa la operaţii blânde de procesare de semnal şi capacităţii lor de a detecta manipulări.

Cele mai importante atacuri de modificare intenţionată a conţinutului multimedia sunt:

- atacuri de substituţie a unor regiuni din materialul cu watermark cu conţinut diferit

- atacuri de ştergere a unor detalii din materialul gazdă
- atacuri de inserare de noi obiecte în scenă
- manipulări care modifică geometria obiectelor, de exemplu prin rotaţie, oglindire, translaţie şi scalare
- manipulări ce schimbă apariţia obiectelor, ca de exemplu culoarea sau umbra
- schimbări ale fundalului scenei, de exemplu schimbarea orei din zi sau schimbări ale texturii fundalului (pădure, ocean, etc.)
- trunchiere

Dintre aceste atacuri, pentru a compara capacităţile de detecţie a diferiţilor algoritmi, folosim atacul de substituţie, înlocuind un bloc al imaginii cu watermark cu versiunea sa originală, fără watermark. Acest tip de atac este probabil cel mai greu de detectat, constituind astfel cea mai mare provocare.

Pentru a testa rezistenţa algoritmilor propuşi la modificări neintenţionate, am ales următoarea listă de procesări de semnal, având la bază exemple din literatura de specialitate. Ne aşteptăm ca algoritmii semi-fragili să nu dea alarme false la aceste modificări permise:

- compresia slabă, de exemplu JPEG până la 80%
- egalizare de histogramă (distribuţie uniformă)
- Ascuţirea contururilor (sharpening)
- filtrare trece-jos pe blocuri de 3x3 pixeli
- filtrare mediană pe blocuri de 3x3 pixeli
- zgomot aditiv Gaussian până la un PSNR de 35 dB
- zgomot de tip „salt and perper” cu 1% dintre pixeli modificaţi (valoare setată la 0 sau 255)
- erori aleatoare de bit la transmisia şi stocarea în format raw, cu o probabilitate de eroare de 0.001

Metodele semi-fragile de watermarking testate şi modurile de funcţionare ale algoritmilor sunt date pe scurt în Tabelul 7.1. Algoritmii au fost dezvoltaţi pentru a proteja regiuni de diferite mărimi din imagini, unii dintre ei verifică autenticitatea la nivel de pixel, alţii verifică blocuri de diferite dimensiuni (4x4, 8x8, 16x16, 64x64), pe linii sau coloane. Pentru a putea realiza o comparaţie între metode, s-a stabilit o dimensiune fixă a blocului de bază. O regiune de 64x64 de pixeli este declarată neautentică, dacă un număr suficient de blocuri native (de dimensiune propusă de fiecare metodă în parte) sunt neautentice.

Tabelul 7.1. Caracteristici ale metodelor semi-fragile

Algoritm	Domeniu de inserare şi metodă de inserare	Dimensiune a ariei de control	Informaţie/metodă de autentificare
Lin-Chang	Coeficienţi DCT şi cuantizare JPEG-50	Perechi de blocuri 8x8	Relaţie de ordine a trei coeficienţi DCT aleşi aleator din bloc
Lin-Podilchuk-Delp	Mixare spaţială şi aditivă	Blocuri 16x16	DCT inversă a unei secvenţe aleatoare de zgomot poziţionată în banda medie a blocului de coeficienţi
Eggers-Girod	Coeficienţi DCT şi QIM (Modulaţie a Indicelui de Cuantizare) binară	Blocuri de 8x8	Schema Costa scalară: cuantizare par/impar a coeficienţilor DCT cu dither
Fridrich	Mixare spaţială şi aditivă	Blocuri de 64x64	Funcţie hash robustă a blocului, obţinută prin proiecţii cuantizate ale blocului pe 30 de baze aleatoare, folosite pentru generarea unui zgomot aleator şi a unei secvenţe separate de frecvenţă joasă
Kundur-Hatzinakos	QIM a coeficienţilor Wavelet din familia Haar	Blocuri de 4x4	Secvenţă de autentificare inserată prin cuantizare par/impar a coeficienţilor wavelet de nivel 4
Gwo-Lu-Liao	QIM a mediei blocului de coeficienţi wavelet	Blocuri de 4x4	Secvenţă de autentificare inserată prin cuantizare par/impar a valorii medii a grupuri de coeficienţi wavelet
Queluz	Cuantizare spaţială şi adaptivă a proiecţiilor liniilor/coloanelor	triadă de linii/coloane	Secvenţă de autentificare inserată prin cuantizarea proiecţiilor unei triade de linii (coloane) pe baze aleatoare
Lan-Mansour-Tewfik	Coeficienţi DCT, proiecţii Hadamard şi cuantizare	Grupuri de blocuri 8x8	Secvenţă de autentificare inserată prin cuantizare par/impar de proiecţii ale vectorilor de coeficienţi DCT pe baze Hadamard

De asemenea, dimensiunea watermark-ului a fost aleasă astfel încât inserarea acestuia să producă un PSNR de 38 dB. Imaginile de test utilizate au fost alese de rezoluţie 512x512 pixeli. Rezultatele obţinute de fiecare metodă sunt date în Tabelul 7.2.

Tabelul 7.2. Performanţele metodelor semi-fragile de autentificare

Metodă	Prob. detecţie atac	Probabilitate de alarmă falsă la procesări de semnal						
		Fără atac	Filtrare	Egalizare histogramă	„saltandpepper” 1%	AWGN 35 dB	JPEG 80%	Ascuţire contururi
Chang	100%	0.0%	100%	100%	100%	0.0%	0.0%	100%
Delp	99.8%	0.2%	37.1%	0.5%	1.2%	0.5%	0.3%	0.3%
Eggers	100%	0.0%	25.3%	87.3%	0.9%	0.0%	0.0%	75.0%
Fridrich	99.4%	1.1%	43%	3.1%	11.4%	1.1%	5.0%	20.0%
Kundur	99.9%	0.0%	68.2%	98.9%	31.7%	3.5%	0.6%	95.9%
Queluz	99.99%	0.01%	15.5%	87.6%	7.8%	0.01%	0.01%	99.5%
Liao	83.9%	0.2%	12.9%	59.4%	2.0%	0.3%	0.1%	57.7%
Tewfik	83.1%	1.4%	83.5%	82.7%	82.6%	85.5%	81.4%	76.1%

Deoarece succesul oricărui algoritm depinde exclusiv de contextul aplicării sale, următoarele observaţii pot fi făcute în ceea ce priveşte semi-fragilitatea (rezistenţa la diferite operaţii de procesare de semnal), probabilitatea de detecţie şi dimensiunea blocului. Majoritatea algoritmilor, exceptând Gwo-Lu-Liao şi Lan-Mansour- Tewfik detectează bine manipularea prin substituţie şi au şi o probabilitate mică de alarmă falsă în lipsa atacului. Pe de altă parte, este interesant de observat că toate metodele sunt destul de sensibile la operaţii de filtrare, ceva mai puţin algoritmii Queluz şi Liao.

Probabilitatea de detecţie în lipsa atacului de substituţie şi în prezenţa operaţiilor de procesare de semnal ar trebui să tindă către zero. Observăm că din aceasta privinţă, algoritmul Delp are, în medie, cea mai mică probabilitate de alarmă falsă, urmat de algoritmii Fridrich şi Eggers-Girod.

Mărimea blocului pentru detecţia manipulării ar trebui să fie flexibilă, deoarece atacul poate varia de la o linie conţinând câteva zeci de pixeli până la întregul plan al imaginii. Din acest punct de vedere, algoritmii propuşi de Fridrich şi Lan-Mansour-Tewfik sunt cei mai puţin flexibili, deoarece ei nu funcţionează bine pentru blocuri mai mici de 64x64. De asemenea, algoritmul Lin-Podilchuk-Delp necesită blocuri de 16x16 pixeli. Mai există încă un aspect legat de dimensiunea blocului şi anume uşurinţa cu care detecţia manipulării pe blocuri mai mici poate fi integrată în detecţia manipulării unui bloc de dimensiune mai mare, dacă suspectăm, că a fost afectată o regiune mai mare decât dimensiunea nativă a algoritmului. Astfel, dacă dimensiunea nativă a blocului este 8x8, ne-am dori să prezicem probabilitatea de manipulare pe o regiune mai mare, de exemplu 32x32 sau 64x64. La calculul scorurilor alarmele false sporadice ce nu sunt datorate atacului intenţionat dispar prin mediere, în timp ce detecţiile corecte în blocuri mici dau o indicaţie mai sigură. Din acest punct de vedere, algoritmul Eggers-Girod utilizează ce mai directă formulă pentru integrare, pur şi simplu însumând probabilităţile de manipulare. Algoritmul Lin-Podilchuk-Delp poate, de asemenea, să funcţioneze cu uşurinţa cu diferite dimensiuni ale blocului.

De remarcat că pot exista şi alte criterii de comparaţie, ca de exemplu capacitatea de autocorecţie şi uşurinţa cu care o metodă poate fi atacată, etc.

7.3. Autentificarea imaginilor cu regiuni de interes (ROI)

În unele aplicaţii, fidelitatea imaginii sau secvenţei video originale este de importanţă deosebită, ca de exemplu în cazul imaginilor medicale, satelitare şi militare. În astfel de aplicaţii chiar şi cele mai mici modificări nu sunt acceptabile, în special în regiuni importante ale imaginilor/cadrelor. Pe de altă parte, în alte aplicaţii, doar una sau mai multe regiuni particulare sunt potrivite pentru inserarea watermark-ului, pe când celelalte părţi au o capacitate de inserare inexistentă sau foarte mică, iar inserarea watermark-ului în aceste regiuni ar produce pierderi mari de calitate (ex. autentificarea fotografiilor de identitate). Cele două cazuri menţionate mai sus pot fi abordate prin aceeaşi soluţie, utilizând tehnici de autentificare care suportă regiuni de interes (ROI – Regions Of Interest).

În comparaţie cu criptografia, unul dintre avantajele tehnicilor de watermarking pentru aplicaţii de autentificare este posibilitatea localizării regiunilor modificate. Pentru autentificarea conţinutului imaginilor, această proprietate este o caracteristică de dorit pentru majoritatea aplicaţiilor. În afara de examinarea integrităţii generale a conţinutului imaginii, informaţia despre poziţiei, unde a avut loc manipularea, este, de asemenea foarte utilă în practică. Cu ajutorul acestei informaţii, părţile nealterate ale imaginii pot rămâne de încredere şi utile atunci când datele originale nu sunt disponibile. Pot, de asemenea, să divulge motivul atacatorului, ca de exemplu în cazul dovezilor juridice.

7.3.1. Autentificarea imaginilor şi secvenţelor video folosind tehnici de watermarking bazate pe blocuri

Pentru a implementa posibilitatea detecţiei regiunilor manipulate, multe scheme de watermarking existente inserează watermark-ul în blocuri din domeniul spaţial sau transformat [227], [228-232]. Imaginea/cadrul este împărţit în blocuri, iar watermark-urile sunt inserate în fiecare bloc. Autentificarea blocului respectiv se face verificând dacă watermark-ul poate fi extras cu succes din acel bloc.

În continuare, vom identifica câteva dintre problemele metodelor bazate pe blocuri şi vom identifica soluţii pentru aceste probleme.

La metodele bazate pe blocuri, rezoluţia maximă de detecţie a zonelor alterate este dimensiunea blocului utilizat pentru inserarea watermark-ului. De exemplu, algoritmii propuşi în [227] şi [229] utilizează blocuri de 8x8, rezoluţia de detecţie fiind astfel limitată la 8x8

pixeli. Deoarece blocul este unitatea minimă care poate conţine cel puţin un bit de watermark, rezoluţia maximă de detecţie este proporţională cu dimensiunea watermark-ului. Pentru a creşte rezoluţia de detecţie, trebuie utilizate dimensiuni mai mici ale blocurilor, dar, în acelaşi timp, acest lucru va conduce şi la o dimensiune mai mare a watermark-ului. Ca urmare, o dimensiune mai mare a watermark-ului va produce mai multe artefacte şi va degrada mai puternic calitatea imaginii/cadrului. De exemplu, în [230] sensibilitatea detecţie este îmbunătăţită la 2x2 pixeli, dar dimensiunea watermark-ului este, de asemenea, crescută la 1 bit pentru fiecare bloc de 2x2 pixeli. Din aceste motive, provocarea nu este doar compromisul dintre cei doi factori competitivi, rezoluţie de detecţie şi dimensiunea watermark-ului, ci şi necesitatea de a găsi o modalitate de a creşte rezoluţia de detecţie, înserând, în acelaşi timp, aceeaşi sau chiar mai puţină informaţie de watermark.

În plus, pentru a proteja întreaga imagine printr-o schemă bazată pe blocuri, datele de autentificare, adică watermark-ul, trebuie inserate local în întreaga imagine. Totuşi, este foarte dificil să inserezi date în regiuni uniforme fără a cauza artefacte vizuale importante, deoarece capacitatea de inserare în aceste regiuni este mult mai mică decât în regiuni cu textură. Astfel, rata de detecţie a watermark-ului va fi mult mai mică în astfel de regiuni uniforme. Problema se agravează la inserarea watermark-ului în blocuri mai mici. În [233], este utilizată permutarea aleatoare a blocurilor pentru a rezolva distribuţia neuniformă a capacităţii watermark-ului, în vederea folosirii mai eficiente a capacităţii totale de inserare a imaginii. Scopul este utilizarea capacităţii disponibile de inserare a watermark-ului pentru a ascunde cât mai multă informaţie. Deci, o astfel de abordare este interesantă nu doar pentru a creşte rezoluţia de localizare a manipulării cu un watermark de aceeaşi dimensiune sau chiar una mai mică, ci şi pentru a rezolva problema distribuţiei neuniforme a capacităţii watermark-ului în imagine/cadru video.

O altă problemă a metodelor bazate pe blocuri este vulnerabilitatea lor la diferite tipuri de atacuri locale. Deoarece schemele bazate pe blocuri inserează watermark-ul local, ele au puncte slabe la atacuri ca de exemplu copiere şi inserare de blocuri, cuantizarea vectorilor (VQ – Vector Quantization), etc. Atacurile VQ, cunoscute şi ca atacuri prin colaj, care se manifestă prin interschimbarea de blocuri în aceeaşi imagine sau între mai multe imagini cu watermark. Aproape toate metodele bazate pe blocuri sunt, într-o anumită măsură, vulnerabile la astfel de atacuri locale, în special dacă datele pentru autentificare şi procedeul de inserare sunt independente de bloc. Nu doar metodele bazate pe blocuri în domeniul DCT, ci şi cele în domeniul DWT au de suferit de pe urma atacurilor VQ, datorită proprietăţii de localizare

spaţiu-frecvenţă a transformatei wavelet. Pericolul atacurilor locale devine chiar mai mare atunci când algoritmul de watermarking este cunoscut de către atacator, aşa cum este cazul cu algoritmii publici.

Securitatea strategiei de inserare este, de asemenea, o provocare importantă. La tehnicile existente de watermarking pentru autentificarea imaginilor şi secvenţelor video sunt utilizate, de regulă, metode de inserare bazate pe cuantizare. Atunci când algoritmul de inserare îi este cunoscut atacatorului, acesta poate modifica datele în orice fel. Această problemă de securitate poate fi redusă prin trei mijloace diferite: combinând schema de watermarking cu mecanisme criptografice, asigurând securitatea prin extragerea caracteristicilor şi îmbunătăţirea algoritmului de inserare în sine. De exemplu, tehnici criptografice tradiţionale, ca de exemplu funcţiile hash, pot fi utilizate la sistemele de watermarking pentru a creşte securitatea sistemului. Aceste tehnici implică, însă, de cele mai multe ori mai mulţi pixeli sau coeficienţi. Astfel, algoritmii de watermarking bazaţi pe criptografie nu asigura întotdeauna localizarea regiunilor manipulate la o rezoluţie fină. Schemele bazate pe extragerea de caracteristici au o problemă asemănătoare, deoarece o caracteristică este definită ca fiind o anumită proprietate a unui set de eşantioane ale imaginii. Cu privire la această problemă, în [234] este propusă o îmbunătăţire bazată pe inserarea utilizând tabele look-up, unde numerele maxim permise de biţi „0” şi „1” pot fi adaptate. De exemplu, raportul aceste două numere poate fi crescut la 2 în comparaţie cu inserarea simplă utilizând cuantizarea la un număr par/impar de paşi de cuantizare (care este aceea cu inserarea LUT cu număr întotdeauna egal de biţi „0” şi „1”). Pe de altă parte, inserarea LUT va degrada mai puternic calitatea imaginii cu watermark, deoarece sunt introduse mai multe distorsiuni prin utilizarea unui raport mai mare decât 1.

7.3.2. Autentificarea imaginilor şi secvenţelor video folosind tehnici de watermarking ce nu sunt bazate pe blocuri

Schemele de autentificare bazate pe blocuri au o limitare importantă, şi anume faptul că rezoluţia de detecţie a manipulării este limitată la dimensiunea blocului. Pentru a creşte rezoluţia de detecţie a acestor metode este necesară scăderea dimensiunii blocurilor şi astfel creşterea dimensiunii watermark-ului, ceea ce conduce la degradare importantă a calităţii imaginii. Utilizarea Transformatei Wavelet Discrete (DWT) în locul Transformatei Cosinus Discrete (DCT) permite o mai bună localizare spaţială a atacului, putând îmbunătăţi simţitor performanţele algoritmului.

Pentru a elimina dezavantajele metodelor bazate pe blocuri au fost dezvoltate metode ce lucrează în domeniul DWT. Securitatea este asigurată prin permutarea coeficienţilor wavelet în care se inserează watermark-ul. În continuare vom prezenta o astfel de tehnică, propusă în [235]. Schema îşi propune îmbunătăţirea preciziei localizării manipulării, scăderea capacităţii watermark-ului, crescând calitatea perceptuală şi are capacitatea de a deosebi modificările intenţionate ale conţinutului imaginilor de procesările de semnal neagresive. Tehnici de morfologie matematică pot fi, de asemenea, utilizate pentru îmbunătăţirea detecţiei.

Diagramele bloc ale codorului, respectiv decodorului (autentificare imagine) sunt date în Figurile 7.4 şi 7.5.

Metoda de inserare a watermark-ului este prezentată pe scurt în cele ce urmează. În primul rând imaginea originală, cu nuanţe de gri, este transformată în domeniul wavelet folosind o descompunere 2D-DWT pe L nivele de rezoluţie şi sunt selectaţi coeficienţii Wavelet de detaliu din sub-benzile Wavelet LH_n, HL_n şi HH_n de nivel de rezoluţie $n \le L$. Aceştia sunt concatenaţi într-un vector 1D C şi permutaţi aleator folosind o cheie secretă K, obţinând un nou vector C'. Acest proces ne asigură că coeficienţii ce corespund aceleiaşi locaţii spaţiale din imagine sunt separaţi în C'.

Secvenţa C' de coeficienţi permutaţi este divizată în grupuri de câte d coeficienţi. Parametrul d controlează capacitatea watermark-ului. Un bit de watermark este inserat în fiecare grup de d coeficienţi. O dimensiune mai mică a grupului va produce o capacitate mai mare a watermark-ului şi astfel o degradare mai puternică a calităţii imaginii cu watermark şi o dimensiune mai mică a ariei maxim localizabile, dar nu va scădea rezoluţia de detecţie a schemei.

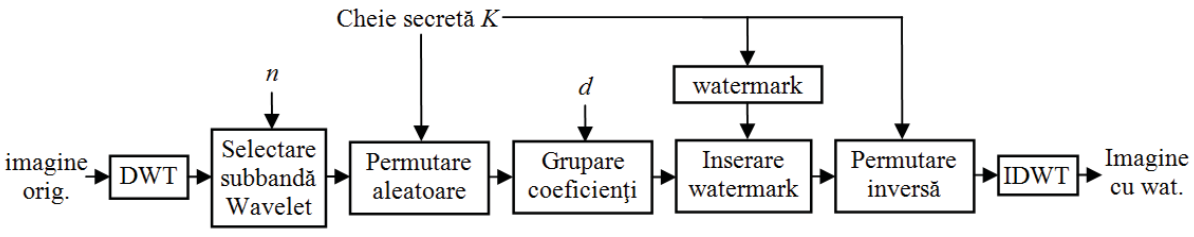

Figura 7.4. Diagrama bloc a codorului

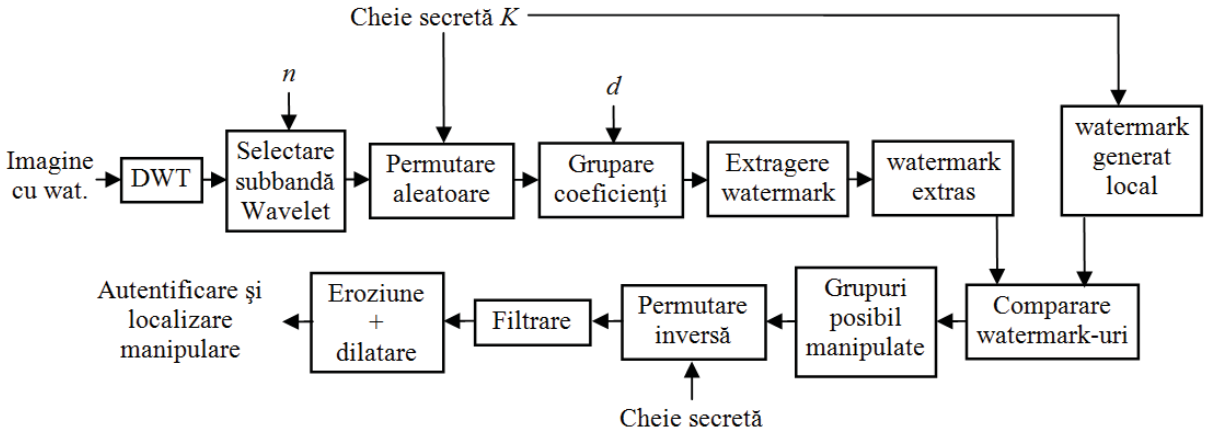

Figura 7.5. Diagrama bloc a detectorului de watermark şi a blocului de autentificare

Watermark-ul, o secvenţă pseudo-aleatoare *w*, este generat cu ajutorul cheii secrete *K* şi serveşte pe post de secvenţă de autentificare. *w* are lungimea egală cu numărul de grupuri de coeficienţi wavelet.

Se calculează media m_i a fiecărui grup i de coeficienţi permutaţi folosind relaţia (7.1) :

$$m_i = \sum_{j=1}^{d} (-1)^j \left| c_i(j) \right| \tag{7.1}$$

unde $c_i(j)$ este coeficientul *j* al grupului *i*. Coeficientul de pondere $(-1)^j$ este folosit pentru a face schema mai robustă la operaţii neagresive de procesare de semnal, care, de regulă, modifică întreaga imagine şi nu afectează media ponderată.

Valoarea medie cu watermark inserat m_i^w se obţine prin cuantizarea lui m_i la cel mai apropiat număr par sau impar de paşi de cuantizare în funcţie de valoarea bitului corespunzător al watermark-ului:

$$m_i^w = \begin{cases} \lfloor m_i/Q \rfloor \cdot Q & if \ \bmod 2(\lfloor m_i/Q \rfloor) = w_i \\ \lfloor m_i/Q \rfloor \cdot Q + Q & if \ \bmod 2(\lfloor m_i/Q \rfloor) \neq w_i \end{cases}, \tag{7.2}$$

unde $\lfloor \ \rfloor$ este operatorul de parte întreagă şi mod2 este restul împărţirii la 2.

Pentru fiecare grup *i* de coeficienţi, valoarea mediei ponderate m_i este schimbată cu noua medie m_i^w prin modificarea coeficientului wavelet $c_{i,\max}(j)$, ce are cea mai mare valoare absolută din grup. Datorită operaţiei de permutare aleatoare fiecare grup ar trebui să aibă cel puţin un coeficient de valoare absolută mare. Actualizarea coeficientului $c_{i,\max}(j)$ se face după cum urmează:

$$c_{i,\max}^{w}(j) = c_{i,\max}(j) + (-1)^{j} \cdot semn(c_{i,\max}(j)) \cdot (m_i^{w} - m_i) \tag{7.3}$$

unde $c_{i,\max}^{w}(j)$ este coeficientul cu watermark şi

$$semn(x) = \begin{cases} -1, & if\ x \le 0 \\ 1, & if\ x > 0 \end{cases} \tag{7.4}$$

După actualizarea coeficienţilor cu cea mai mare valoare absolută din fiecare grup, se aplică permutarea inversă folosind aceeaşi cheie secretă *K*. Ca ultim pas, se calculează Transformata Wavelet Discretă 2D Inversă, obţinând imaginea cu watermark.

BIBLIOGRAFIE

[1] K. M. Martin, *Everyday Cryptography: Fundamental Principles and Applications*, Oxford University Press, 2012.

[2] A. Stanoyevitch, *Introduction to Cryptography with Mathematical Foundations and Computer Implementations*, Chapman and Hall/CRC, Boca Raton, FL, 2010.

[3] D. R. Stinson, *Cryptography: Theory And Practice*, Chapman & Hall/CRC, 2006.

[4] [14] I.J. Cox, M.l. Miller, *Electronic watermarking: the first 50 years*, IEEE Fourth Workshop on Multimedia Signal Processing, pp. 225-230, 2001.

[5] D. Kahn, *The Codebreakers* , MacMillan, New York, 1967.

[6] Mitchell D. Swanson, Mei Kobayashi, Ahmed H. Tewfik, *Multimedia Data-Embedding and Watermarking Technologies*, Proceedings of the IEEE, vol. 86(6), pp. 1064-1087, 1998.

[7] Neil F. Johnson, Sushil Jajodia, *Exploring Steganography: Seeing the Unseen*, IEEE Computer, vol. 31, no. 2, pp. 26-34, 1998.

[8] M. Kobayashi, *Digital Watermarking: Historical Roots*, IBM Research Report, RT0199, Japan, 1997.

[9] J.R. Hernandez Martin, M. Kutter, *Information retrieval in digital watermarking*, Special Issue of IEEE Communication Magazine on Digital Watermarking for Copyright Protection: A Communication Perspective, vol. 39(8), pp. 110-116, 2001.

[10] I.J. Cox, M.L. Miller, J.A. Bloom, *Digital Watermarking*, Morgan Kaufmann Publishers, San Mateo, USA, 2002.

[11] D. Hunter, *Handmade Paper and its Watermarks: A Bibiliography*, B. Franklin, New York, 1962.