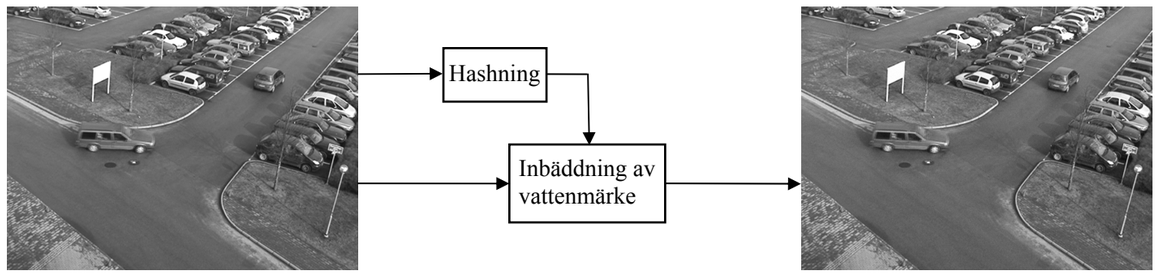
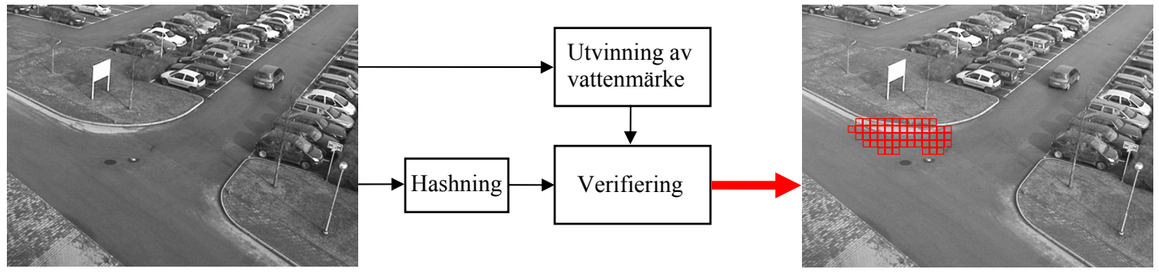
Identifiering och förändringsindikering av sensordata via vattenmärken

En litteraturstudie

K-G STENBORG, MAGNUS HERBERTHSON, GUSTAV HAAPALAHTI





FOI är en huvudsakligen uppdragsfinansierad myndighet under Försvarsdepartementet. Kärnverksamheten är forskning, metod- och teknikutveckling till nytta för försvar och säkerhet. Organisationen har cirka 1000 anställda varav ungefär 800 är forskare. Detta gör organisationen till Sveriges största forskningsinstitut. FOI ger kunderna tillgång till ledande expertis inom ett stort antal tillämpningsområden såsom säkerhetspolitiska studier och analyser inom försvar och säkerhet, bedömning av olika typer av hot, system för ledning och hantering av kriser, skydd mot och hantering av farliga ämnen, IT-säkerhet och nya sensorers möjligheter.

FOI
Totalförsvarets forskningsinstitut
Informationssystem
Box 1165
581 11 Linköping

Tel: 013-37 80 00
Fax: 013-37 81 00

www.foi.se

FOI-R--2699--SE Underlagsrapport **Informationssystem**
ISSN 1650-1942 December 2008

K-G Stenborg, Magnus Herberthson, Gustav Haapalahti

Identifiering och förändringsindikering av sensordata via vattenmärken

En litteraturstudie

Titel	Identifiering och förändringsindikering av sensordata via vattenmärken
Title	Watermarking for Identification and Change Detection of Sensor Data
Rapportnr/Report no	FOI-R--2699--SE
Rapporttyp Report Type	Underlagsrapport
Sidor/Pages	39 p
Månad/Month	December
Utgivningsår/Year	2008
ISSN	ISSN 1650-1942
Kund/Customer	FM
Forskningsområde Programme area	4. Sensorer och signaturanpassning 4. Sensors and Low Observables
Delområde Subcategory	42 Sensorer 42 Above surface Surveillance, Target acquisition and Reconnaissance
Projektnr/Project no	E3093
Godkänd av/Approved by	

FOI, Totalförsvarets Forskningsinstitut
Avdelningen för Informationssystem
Box 1165
581 11 Linköping

FOI, Swedish Defence Research Agency
Information Systems
Box 1165
SE-581 11 Linköping

Sammanfattning

Sensordata som innehåller ljud, bild och video inom olika frekvensområden är viktiga, och kommer i framtiden vara ännu mer betydelsefulla, inom försvaret. Vattenmärkning (Watermarking) är en metod som kan användas för att bädda in information i sensordata. Efter inbäddningen skall det inte finnas någon perceptuell skillnad mellan originaldata och den sensordata som fått ett vattenmärke. Detta vattenmärke kan vid ett senare tillfälle utvinnas ur sensordatan och användas för olika säkerhetstillämpningar. I denna rapport går vi djupare in på två av dessa tillämpningar:

1. Identifiering – Ett vattenmärke bäddas in robust i sensordatan och innehåller viktig information som till exempel sensor-id, tidpunkt, sensorinställningar, GPS-positionering och copyrightuppgifter. Anledningen till att sådan information sparas i ett vattenmärke och inte i en header är att det blir svårare att i efterhand manipulera informationen. Dessutom följer informationen alltid med även efter formatbyten och andra omvandlingar så att viktiga uppgifter inte riskeras att förloras.

2. Förändringsindikation – För digitala sensordata finns det inget negativ som kan användas för att påvisa hur originaldata såg ut. Därför finns det risk för att medvetna manipuleringar eller omedvetna förändringar av sensordata medför att felaktigt grundade beslut fattas. Genom att använda så kallade bräckliga vattenmärken går det avgöra om sensordata har blivit förändrat eller ej. Vissa sådana metoder kan dessutom upptäcka inom vilka områden som till exempel en bild har blivit förändrad och några metoder kan även delvis återskapa originalinformationen i dessa områden.

Denna rapport går först igenom grunderna för vattenmärkning och studerar sedan vad som tidigare gjorts inom de båda tillämpningsområderna. Ett scenario med övervakningskameror runt ett skyddsområde används för att illustrera hur vattenmärkning på olika sätt kan användas för att öka säkerheten.

Nyckelord:

Vattenmärke, vattenmärkning, vattenstämpel, digital signatur, watermark, watermarking

Summary

Sensor data containing sound, images and video with various frequency contents are important, and will in the future be even more so, within the Swedish Defence. Watermarking is a method which can be used to embed information in such data. After the embedding, there should be no conceptual difference between the original and watermarked sensor data. The watermark can at a later stage be extracted from the sensor data and used for security applications. In this report we will we look into two such applications:

1. Identification - A watermark is embedded in the sensor data in a robust manner. It can contain important information like, for example sensor id., timestamps, sensor settings, GPS data and copyright information. The reason for this information to be stored as a watermark rather than in a header is that it is harder to manipulate the information afterwards. In addition, the information will remain even after a format change or another transform, which means that important data is not lost.

2. Change detection - To digital sensor data there is no 'negative' which can be used to prove what the original data looked like. Therefore, there is a risk that deliberate manipulations or accidental changes of sensor data results in decisions made on false premises. By using so called fragile watermarks, it is possible to decide whether sensor data have been changed or not. With some fragile watermarking methods, it is in addition possible to discover in which parts for example an image have been manipulated, and some methods can even partly recover the original information in these areas.

This report starts by addressing the fundamentals of watermarking, and is then studying achievements within the above mentioned application areas. A scenario with surveillance cameras around a restricted area is used to illustrate how watermarking can be used in different ways to improve the security.

Keywords:

Watermark, watermarking, digital signature

Innehållsförteckning

1 Inledning

Detta kapitel inleds med en kort bakgrund, fortsätter med att presentera ett scenario och avslutas med att beskriva innehållet i kommande kapitel.

1.1 Bakgrund till vattenmärkning

Vattenmärkning (Watermarking) är ett sätt att bädda in information i olika typer av sensordata som bilder, ljud och video. Namnet vattenmärke[1] kommer från den typ av märke som sedan flera hundra år används för att märka viktiga papper såsom sedlar och arkiveringspapper [Hartung99].

Det finns många typer av digitala vattenmärkningsmetoder och många olika användningsområden. I denna rapport kommer vi titta närmare på två från varandra väldigt skilda områden av vattenmärken:

- **Identifikation** där viktig information såsom sensoridentitet, ägare av data och tidkod bäddas in sensordata. Denna tillämpning är viktig för att kunna påvisa vem som skapat sensordata och när det skett.

- **Förändringsindikation** där en signatur som bäddas in i sensordata senare kan användas för att upptäcka om informationen har skadats eller om den är opåverkad. Genom att använda denna typ av vattenmärkning kan vi säkerställa om informationen från sensordata är pålitlig eller ej.

Dessa båda användningsområden kräver olika typer av vattenmärkningsmetoder. Vi har genomfört en litteraturstudie för att undersöka vilka metoder som är bäst anpassade till de båda tillämpningarna och den studien presenteras i denna rapport. Det finns andra tillämpningar för vattenmärkning som spårandet av piratkopierare [Stenborg09] och möjligheten att avläsa innehåll i sändningar (broadcast monitoring) [Oostveen02] samt andra närliggande områden som steganografi [Simmons83]. Dessa tillämpningar och områden ligger dock utanför det som denna rapport tar upp.

Kryptering och vattenmärkning bör oftast användas tillsammans eftersom de har olika svagheter och styrkor vilket gör att de kompletterar varandra. Kryptering hindrar obehöriga från att till exempel kunna se en bild men det fungerar bara så länge som bilden inte har blivit dekrypterad. Vattenmärkning däremot kan användas för att knyta en otillåtet spridd kopia av en bild till dess källa, eller användas för att se var och när bilden är tagen och vem som äger den, eller för att verifiera att en bild inte har blivit manipulerad. Vilken typ av kryptering och vattenmärkning som skall användas är beroende på vilken typ av säkerhet som vill uppnås.

[1] Vi har valt att använda namnet ”vattenmärke” även om ”vattenstämpel” är ett namn som på svenska är vanligare för den typ av märkning som görs i papper.

Figur 1: a) Synligt vattenmärke innehållande FOIs logotyp. Denna typ av synlig vattenmärkning är enkel att ta bort exempelvis genom att beskära bilden. b) Osynligt vattenmärke som yttrar sig som ett svagt brus i bilden som bara kan märkas när skillnaden mellan originalbild och vattenmärkt bild beräknas.

Vi har i första hand koncentrerat oss på visuella sensordata[2] som bilder och video. Vattenmärken kan delas in i två klasser, synliga och osynliga.

- **Synliga** vattenmärken är sådana som tydligt kan ses av alla användare. Vanliga typer av synliga vattenmärken är de copyrighttexter som kan finnas i nederkanten på fotografier eller de tv-kanalslogotyper som vanligtvis ligger i övre högra hörnet av tv-bilden. Ett synlig vattenmärke kan ses i figur 1a.

- **Osynliga** vattenmärken är sådana som inte är synliga för blotta ögat utan ligger inbäddat som till exempel ett brus i bilden. Det är denna typ av perceptuellt osynliga vattenmärken som vi är intresserade av. Ett osynligt vattenmärke illustreras i figur 1b.

Resten av denna rapport kommer behandla osynliga märken men även synliga märken kan ha många användningsområden.

1.2 Scenario

Vi har valt att i denna rapport beskriva ett scenario som på olika sätt får belysa hur de olika vattenmärkningsmetoderna kan användas. Många andra intressanta områden för scenarier skulle också kunnat användas, såsom:

- Video/IR-information från UAV.
- Kameraförsedda soldater som lagrar eller distribuerar video.
- Bild och ljudkommunikation mellan två centraler.
- GIS, Geografiska informationssystem.

Vi tycker dock att övervakningskamerascenariet på nästa sida enkelt kan illustrera många av vattenmärkningstillämpningarna.

[2] Med visuella sensordata tänker vi oss här data som representeras visuellt, dvs även tex IR-data som ligger på för ögat inte uppfattbara våglängder är visuella eftersom det går att titta på dess sensordatarepresentation.

Övervakningsscenario:

Ett högt prioriterat skyddsområde övervakas genom ett antal kameror. Dessa övervakningskameror har en låg bildfrekvens med fyra bilder per sekund för att den datamängd som sparas varje dygn inte skall bli för stor. Dessa övervakningsbilder lagras i en månad på hårddiskar. Övervakningsbilderna granskas dessutom i realtid i en övervakningscentral av en vakt. Eftersom det är många kameror kan vakten inte titta på alla samtidigt utan han/hon måste växla mellan kamerorna.

En dag visar det sig att hemligt material förvunnit från området. En första titt på de senaste veckornas övervakningsbilder visar ingenting ovanligt. En misstanke uppstår om att bilderna blivit manipulerade endera under överföringen mellan kamera och övervakningscentral eller efter lagring på hårddiskarna.

I rapporten kommer vi återkomma till detta scenario för att exemplifiera hur de olika vattenmärkningsmetoderna kan användas. Dessa scenariodelar är alla inrutade på liknande sätt som detta så att de skall vara enkla att hitta.

1.3 Översikt

Denna rapport är uppdelad så att **kapitel 2** går igenom grunderna för vattenmärkning och sedan även beskriver de båda metoderna identifiering och förändringsindikation. I **kapitel 3** beskriver vi sedan några metoder för identifiering och i **kapitel 4** metoder för förändringsindikation. Slutsatser redovisas i **kapitel 5**.

2 Vattenmärkning

Detta kapitel går igenom grunderna för digital vattenmärkning, först generellt och sedan mer i specifikt för de båda tillämpningarna identifiering och förändringsindikation. Dessutom görs korta beskrivningar av hur de olika vattenmärkningsdelarna kan yttra sig på övervakningsscenariot.

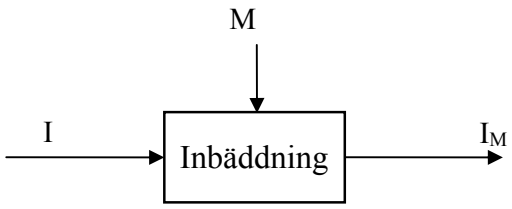
2.1 Generellt om vattenmärkning

Antag att sensordata *I* skall vattenmärkas med märket *M*. För enkelhets skull kommer vi oftast att betrakta sensordata som en stillbild[3]. I så fall skall vi på något sätt ändra lite på pixlarna i *I* så att vårt märke *M* blir inbäddat utan att man med det mänskliga ögat ser någon skillnad. Vi kallar den inbäddade bilden för I_M. Utgående från I_M skall vi sedan kunna utvinna märket *M* eller kontrollera om bilden är märkt med märket *M*.

2.1.1 Inbäddning

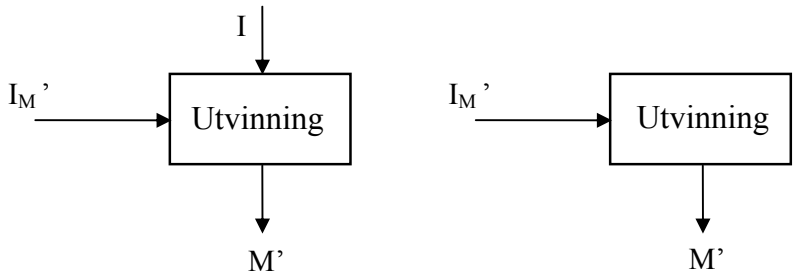
Inbäddningen av märket i bilden kan ske på flera olika sätt. Det går till exempel att bädda in vattenmärket direkt i bilden (spatialdomän) eller efter någon transform, dvs i krusningsdomän (även känd som Wavelett eller vågelement) eller i någon frekvensdomän. Även andra angreppssätt för inbäddning är möjlig som att bygga vattenmärkningen på fraktalkodning [Puate96]. Mer exakta beskrivningar av hur inbäddning kan göras i olika domäner finns i kapitel 3 och 4.

Figur 2 ger en generell beskrivning av hur bilden *I* får vattenmärket *M* inbäddat så I_M skapas. Denna och övriga figurer i rapporten är förenklade. Egentligen använder de flesta praktiska metoder nämligen någon form av nycklar vid inbäddning och utvinning av vattenmärkena.



Figur 2: Inbäddning av vattenmärket *M* i bilden *I* vilket resulterar i den märkta bilden I_M.

[3] Att vi väljer stillbild som sensordata hindrar inte att det mesta som beskrivs i rapporten med vissa modifieringar kan appliceras på ljud, video och andra typer av data.



Figur 3: a) Icke-blind vattenmärkningsutvinning. För att utvinna vattenmärket *M'* ur bilden I_M' krävs originalbilden *I*. b) Blind vattenmärkningsutvinning. Originalbilden behövs inte för att utvinna vattenmärket utan det räcker med den vattenmärkta bilden I_M'.

2.1.2 Utvinning

Om vi får en vattenmärkt bild I_M' måste vi på något vis få fram vilket märke *M'* som finns inbäddat. För att kunna utvinna märket kan olika typ av information behövas. Vi kan därför dela in vattenmärkningsmetoder och deras utvinning i följande klasser:

- **Icke-blind vattenmärkning** (Nonblind Watermarking) – För att kunna utvinna vattenmärket ur en bild I_M' krävs originalbilden *I*. Se figur 3a.

- **Blind vattenmärkning** (Blind Watermarking) – Originalbilden behövs inte för att kunna utvinna vattenmärket ur bilden I_M'. Se figur 3b.

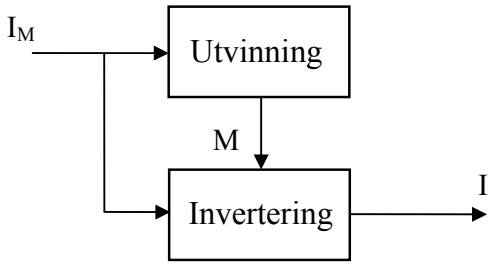
Anledningen till att vi kallar den bild som skall undersökas för I_M' är att den kanske inte längre är exakt likadan som den inbäddade bilden I_M. Efter inbäddningen kan bilden ha exponerats för skador eller medvetna attacker (sektion 2.1.5) vilket också orsakar att det utvunna märket *M'* kanske inte längre är exakt som den inbäddade märket *M*.

För de flesta tillämpningar är blind vattenmärkning att föredra eftersom originalbilden då inte behövs vid kontroll av vattenmärket. Förutom blind och icke-blind vattenmärkning finns följande klasser för utvinningen:

- **Privat vattenmärkning** (Private Watermarking) – Endast auktoriserade parter kan utvinna vattenmärket ur I_M genom att vattenmärkningsmetoden är hemlig eller genom att en hemlig nyckel krävs för att genomföra utvinningen.

- **Öppen vattenmärkning** (Public Watermarking) – Alla kan utvinna vattenmärket då utvinningsmetoden är känd och ingen hemlig nyckel krävs.

Privat vattenmärkning är ofta att föredra eftersom den oftast anses vara säkrare. Fast vid vissa tillämpningar är öppen vattenmärkning önskvärd, tex om vattenmärket innehåller uppgifter om upphovsrättsinnehavare och man vill att alla skall kunna se vem som är bildens ägare.



Figur 4: Inverterbar vattenmärkning. Efter att vattenmärket *M* har utvunnits ur bilden I_M används informationen om *M* för att ta bort *M* ur den märkat bilden I_M och återskapa originalbilden *I*.

2.1.3 Inverterbar vattenmärkning

För vissa situationer kan det vara omöjligt att använda en metod som bäddar in vattenmärken på ett sådant sätt så att originalbilden påverkas. Exemplevis kan det vara väldigt känsligt att ändra på ”medicinska bilder” eftersom det anses att även en liten distortion skulle kunna påverka hur en läkare läser av bildinformationen. För att kunna använda vattenmärkning inom sådana användningsområden måste det då vara möjligt att efter utvinning av vattenmärket kunna återskapa originalbilden. Sådana typer av vattenmärkningsmetoder kallar vi inverterbar vattenmärkning (Erasable watermarks, Invertable watermarks).

Förutsatt att en bild inte har påverkats efter inbäddningen av märket kan det oskadade märket utvinnas ur den vattenmärkta bilden I_M. Om vattenmärkningsmetoden är inverterbar går det därefter använda märken *M* för att återställa I_M till originalbilden *I*, se figur 4.

Eftersom bilden inte får ha blivit påverkad mellan inbäddning och utvinning måste det gå upptäcka att så inte skett. Det kan göras genom digitala signaturer (sektion 2.3.1) eller genom att vattenmärkningsmetoden i sig är förändringsindikerande så att *M* kan användas för att se att bilden inte har påverkats. Ett tredje alternativ är att bilden har två vattenmärken, ytterst ett som inverterbart som verifierar att bilden inte påverkats och sedan ett till inverterbart som innehåller annan information.

I praktiken har det visat sig svårt att skapa fungerande inverterbara vattenmärknings-metoder utan att först begränsa originalbildens omfång. Detta beror på att risken alltid finns för att en åtta bitars pixel hamnar utanför intervallet [*0:255*] vilket i så fall inte går återskapa. Även andra avrundningsproblem kan uppstå. För att en originalbild *I* skall kunna inverteras från den märkta bilden I_M brukar därför krav ställas på att pixlarnas värden inte får understiga eller överstiga vissa värden. Pixelvärdena får alltså inte ligga nära gränserna *0* och *255* för att undvika att märkningen orsakar att pixlarna kommer utanför intervallet. Mer om inverterande vattenmärkning kan läsas i [Goljan01][Ni06].

2.1.4 Robusthet

En väldigt viktig egenskap för alla vattenmärkningsmetoder är vilken nivå av robusthet de har. Om en bild blir lite skadad under distribution så att till exempel en liten del (8x8 pixlar) blir till brus, kommer då vattenmärket ha blivit skadat eller kommer det vara helt efter utvinning?

De skador som en vattenmärkt bild kan utsättas för kan dels vara sådana som sker av misstag eller de som sker genom välplanerade attacker. Mer om sådana attacker i nästa sektion.

Det finns tre nivåer av robusthet som en vattenmärke kan ha:

- **Bräckligt vattenmärke** (Fragile Watermark) är inbäddat på ett sådant vis att även väldigt små förändringar i den inbäddade bilden I_M kommer att förstöra vattenmärket M. Denna typ av vattenmärkning är användbar vid förändringsindikation, sektion 2.3.

- **Halvbräckligt vattenmärke** (Semi-fragile Watermark) klarar av små förändingar som måttlig kompression, men om bilden utsätts för större påverkan skadas vattenmärket M. Även denna typ av vattenmärkning kan användas vid förändringsindikation, sektion 2.3.

- **Robust vattenmärke** (Robust Watermark) skall klara av även större påverkan på bilden I_M utan att märket M blir skadat. Denna typ av vattenmärkning används bland annat till indentifikation, sektion 2.2.

Av dessa tre är robust vattenmärke det absolut vanligaste och det används också inom andra tillämpningar som att hitta läckan till piratkopior. I sektion 2.3.3 finns en längre beskrivning av robusthetsgrader för förändringsidentifiering.

Figur 5: a) Attackerad bild genom tillförande av brus. b) Attackerad bild genom förstörande kompression. Bilden är hårt komprimerad med JPEG. För att vattenmärkningen skall överleva sådana attacker måste den vara robust.

2.1.5 Attacker

Det finns många anledningar till att sensordata kan utsättas för medvetna attacker. Här delar vi in attackerna i två övergripande klasser:

- **Förstörande attack** som utförs för att försöka skada det inbäddade märket så att det inte går identifiera. Denna typ av attacker kan till exempel utföras för att ta bort vattenmärken som innehåller:

    - Information om vem som äger bilden.
    - Tidkod som avslöjar när bilden skapades.
    - Uppgift om vem som sprider vidare information (piratkopieringsfallet).

    Med förstörande menar vi i detta fallet lika mycket att det skadade märket inte innehåller någon information som att märket verkar innehålla en annan slumpmässig korrekt information än den ursprungliga. Exempelvis kan en bild attackeras genom förstörande kompression, beskärning, skalning, translation, filtrering, kvantisering och adderande av brus, figur 5.

- **Missledande attack** utförs för att det skall verka som att bilden innehåller ett annat specifikt märke än det ursprungliga. Attacker av denna typen kan till exempel användas för att:

    - Få det att verka som att en påverkad eller utbytt bild är den äkta genom att lura förändringsindikationen.
    - Påvisa att en annan person äger bilden än det ursprungliga ägaren.
    - Förändra tidkoden så det verkar som att bilden är tagen vid ett annat tillfälle.

    Skillnaden mot den förstörande attacken är alltså att här skall inte bara märket tas bort eller förändras slumpmässigt utan bytas ut mot en av attackeraren vald information.

För att klara av en attack måste vattenmärkesalgoritmen vara robust mot just den typen av attack. Den första punkten av de missledande attackerna handlar om förändringindikation där bräckliga eller halvbräckliga vattenmärken används. Sålunda måste även förändringsindikaernade vattenmärkningsmetoder vara robust mot en missledande attack även om de i övrigt är bräckliga eller halvbräckliga.

Övervakningsscenario:

För övervakningsbilderna kan vi tänka oss två ställen där attacker kan utföras. Under överföring från övervakningskamerorna till centralen eller när övervakningbilderna ligger lagrade. Exempel på attacker kan vara att lägga på brus eller komprimera bilderna hårt. Andra attacker kan vara att byta ut bilder från en tidpunkt med bilder från en annan tidpunkt alternativt byta ut delar av bilderna där människor eller fordon syns.

2.2 Identifikationsmärkning

Om man redan i sensorn har möjlighet att bädda in ett robust vattenmärke kan det användas för att till exempel lagra information som sensorns id, sensorinställningar, vem som är ägare av materielet, tidkod eller GPS-information (om en GPS är kopplad till sensorn). Denna typ av information skulle även kunna lagras som metadata som ligger i en header till bilden eller skickas i en egen parallell ström. Några fördelar med att bädda in informationen som ett vattenmärke är:

- Informationen följer alltid med bilden till skillnad från headerinformation som kan försvinna om till exempel bildens lagringsformat förändras (tiff till jpeg eller liknande).
- Det blir svårare för någon att medvetet ta bort eller manipulera informationen till skillnad mot om den ligger lätt åtkomlig i header eller medföljande informations-fil.
- Tidkod inbäddad som vattenmärke gör att risken för synkroniseringsfel kan minskas.

Om det till en bild behövs mycket sidoinformation finns möjlighet att vattenmärket bara innehåller den viktigaste informationen samt en pekare som länkar till en fil där all nödvändig sidoinformation finns. Fast då måste informationen i den filen skyddas på annat sätt.

Identitetsinbäddningen kan göras på många olika sätt. Några av dem redovisas i kapitel 3. Det går att bädda in direkt i spatialdomänen eller transformera bilden till någon frekvensdomän eller krusningsdomän för att göra inbäddningen.

Övervakningsscenario:

Om det i övervakningskamerorna bäddas in en tidkod kan detta vattenmärke utvinnas vid övervakningscentralen och där se att det är rätt tidpunkt som bilderna visar. Detta kan göras automatisk så att vakten blir larmad om tiden inte stämmer. Om någon då försöker att manipulera övervakningsbilderna genom att koppla in sig på distributionslinan och återanvända gamla bilder kommer detta upptäckas.

På liknande vis går det granska tidkoden på äldre lagrat övervakningsmaterialet och se att tiden stämmer. Skulle det visa sig att material som skall komma från en fredagkväll kommer från en torsdagkväll är risken att materialet blivit manipulerat eller att någonting blivit fel vid lagringsprocessen.

En annan situation som vattenmärkena kan användas för är om en övervakningsbild på något sätt har blivit spridd. Den kanske finns upplagd på en internetsida som hävdar att den är tagen en speciell dag och föreställer någonting speciellt. Genom vattenmärket går det för oss utläsa vilken myndighet eller företag som spelat in bilden, vilken inspelningstidpunkten var och vilken sensor som har använts. På så vis går det kontrollera och påvisa om de uppgifter som hävdas på hemsidan stämmer eller ej.

2.3 Förändringsindikation

För digitala stillbilder finns det inget negativ som visar hur en garanterat oförändrad originalbild ser ut. För att kunna avgöra om en digital bild verkligen är originalbilden kan vattenmärkning användas. Vattenmärkning är inte enda lösningen utan även en annan metod, digitala signaturer, finns.

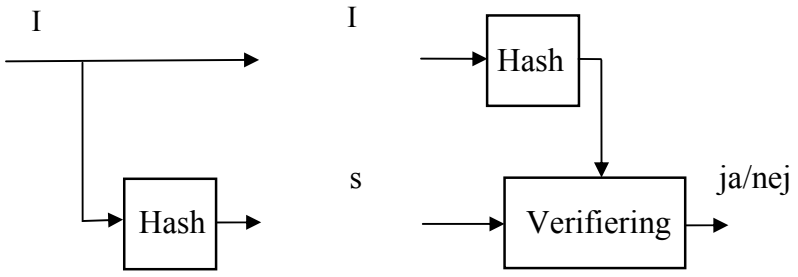
2.3.1 Digitala signaturer

Digitala signaturer kan användas för att autentisera digitala dokument. En bra genomgång av området med inriktning mot multimedia-filer finns i [Sun05].

Här tar vi upp autentisering av dokument som skickas från en sändare A till en mottagare B. Förenklat går det till så att ett hash-värde [Rivest92] beräknas på delar av eller hela det digitala dokumentet hos A. Hash-värdet är den digitala signaturen och den skickas tillsammans med det digitala dokumentet till B. När B har mottagit det digitala dokumentet använder han samma hash-funktion och jämför svaret med den signatur som A skickade, se figur 6. Om de stämmer kan B med stor sannolikhet förlita sig på att det digitala dokumentet inte har blivit förändrat under distributionen. Själva hashningen måste göras med privata respektive öppna nycklar för att ingen tredje part som vill förändra dokumentet skall kunna anpassa signaturen till förändringen.

Digitala signaturer kan även användas i andra sammanhang än distribution, exempelvis för att försäkra sig om att ett lagrat dokument inte har blivit förändrat. Det görs genom att det förutom det lagrade dokumentet finns en autentiseringssignatur lagrad. Den används återigen för att jämföras med resultatet från hashningen av dokumentet.

Ett problem med denna typ av förändringindikation är att signaturen och det digitala dokumentet måste hållas samman. Om det digitala dokumentet exempelvis är en bild så måste signaturen alltid finnas med. Om signaturen då lagras i en header till bilden finns risk för att den försvinner om bilden genomgår någon formatförändring. Skulle en bild spridas och hamna på en internetsida är risken stor att den har bytt format och kanske dessutom komprimerats. Hur skall vi då kunna avgöra om det som visas på bilden stämmer eller om den har blivit manipulerad?



Figur 6: Digital signatur. Förenklad beskrivning av hur en signatur *s* skapas genom en hash av en bild *I*. Vid ett senare tillfälle används signaturen för att verifiera bilden genom jämförelse med en ny hash av bilden.

Figur 7: a) Vattenmärkt bild. b) Manipulerad bild där en bil har tagits bort. c) Global förändringsindikation kan bara notera att bilden är skadad. d) Lokal förändringsindikation noterar var bilden är skadad, i detta fall i bildblock om *16 x 16* pixlar.

2.3.2 Autentiseringsnivå

Det finns olika nivåer på hur mycket information som det går få ut från olika vattenmärkningsmetoder gällande möjliga bildförändringar. Vi listar dem här.

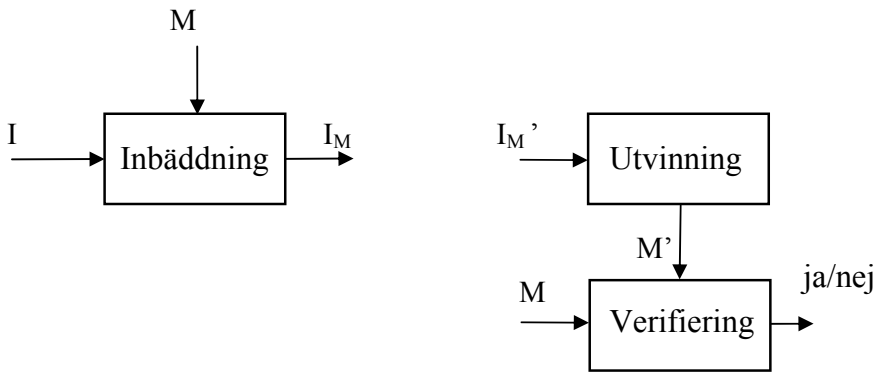
- **Global autentisering** ger bara svar på frågan om bilden har blivit manipulerad eller ej, figur 7c.

- **Lokal autentisering** kan däremot avgöra inom vilka områden som en bild har blivit manipulerad, figur 7d. Vanligtvis delas bilden in i block och varje sådant block verifieras för sig själv. Det finns dessutom metoder som arbetar på pixelnivå och alltså kan påvisa exakt vilka pixlar som har ändrats. Metoder som ger pixelprecision brukar dock ha problem med att vara helt säkra mot attacker och vanligtvis kommer förändrade pixlar bara upptäckas med en viss sannolikhet.

- **Skadetypsupptäckande autentisering** är en typ av metoder där vattenmärkes-metoden ger en uppfattning om inte bara var utan dessutom vilken typ av attack som har utförts och kanske hur mycket som pixlar har förändrats. Denna typ av märkning kallad ”Telltale watermarking” [Kundur99] har det ännu inte genomförts så mycket forskning på. Det är dock av stort intresse att få fram praktiska vattenmärkningsmetoder som ger denna typ av information.

- **Återskapande** (Recovery) betyder att det från en skadad bild går att återskapa originalet eller i alla fall få en uppfattning om hur de skadade partierna såg ut ursprungligen. Mer om återskapande finnns i sektion 2.3.4.

Mer om vilken typ av autentiseringsnivå de olika metoder har beskrivs i kapitel 4.



Figur 8: Förändringsindikation genom att ett för mottagaren känt vattenmärke *M* bäddas in bräckligt eller halvbräckligt hos sändaren. Mottagaren kan sedan utvinna märket *M´*ur bilden och jämföra det mot märket *M* för att se om bilden har blivit förändrad.

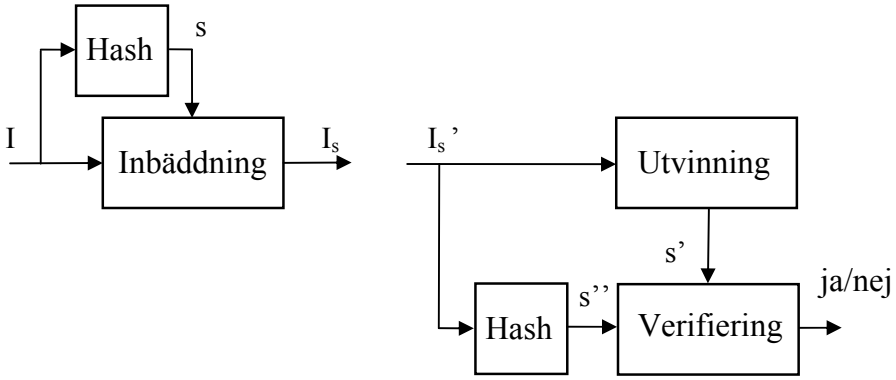
2.3.3 Bildberoende och bildoberoende vattenmärken

För tillämpningen förändringsindikation kan själva märket som bäddas in vara av olika typer, endera är det ett fast märke eller också en signatur av något slag som är beroende av originalbilden. Nedan beskriver vi dessa typer av märken. Notera att vi nu bara diskuterar själva märket innan inbäddningen och inte inbäddningsmetoden.

- **Känt vattenmärke** betyder att vattenmärket som bäddas in redan är känt av mottagaren och att det märket inte är beroende av innehållet i bilden, figur 8.

- **Bräcklig signatur** är en typ av märke som beror på bilden på sådant vis att om bilden förändras lite grann kommer den beroende signaturen att förändras mycket. Att vi kallar denna typ av märken för signaturer kommer från digitala signaturer i sektion 2.3.1. Beroendet kan skapas genom till exempel hashning, figur 9.

- **Halvbräckliga signaturer** är också ett märke som är beroende av bilden fast till skillnad från den helbräckliga signaturen i föregående punkt förblir denna typ av signatur den samma även om bilden genomgår mindre förändringar som exempelvis måttligt förstörande kompression. Om däremot bilden genomgår större förändringar, som att delar raderas eller byts ut, kommer den halvbräckliga signaturen som beräknas utgående från den påverkas. Även halvbräckliga signaturer kan beskrivas av figur 9.



Figur 9: Förändringsindikation genom bräcklig eller halvbräcklig signatur. En signatur *s* beräknas på bilden genom exempelvis en hashfunktion. Denna signatur bäddas in i bilden hos sändaren så I_s bildas. Mottagaren av en bild I_s' kan utvinna det inbäddade vattenmärket *s´*och även använda hashfunktionen för att beräkna bildens signatur *s´´*. Om *s'* och *s''* inte är lika har bilden blivit påverkad. Denna beskrivning är förenklad eftersom inga nycklar används för att beräkna signaturen eller utvinna vattenmärket.

Som vi ser i figur 9 beräknar vi först en bräcklig signatur på bilden och sedan bäddar vi in den i bilden. Då har bilden blivit påverkad och om en ny signatur beräknas på den inbäddade bilden bör den inte vara den samma som den första signaturen. För att lösa detta problem kan inte signaturen beräknas på alla bildens bitar utan vissa bitar (eller delar) av bilden används för att beräkna signaturen och andra bitar (eller delar) av bilden används till inbäddningen av vattenmärket. På så vis kommer signaturen som beräknas på den inbäddade bilden vara den samma som signaturen som beräknades först, förutsatt att ingen annan förändring av bilden har skett.

Beroende på vilka bitar eller delar av bilden som vi väljer för att beräkna signaturen kan vi styra om den kommer vara bräcklig eller halvbräcklig. Om vi till exempel beräknar signaturen på frekvenskomponenter som är invarianta mot JPEG-kompression kommer den vara halvbräcklig och robust mot just den typen av kompression.

Vid utvinningen av det inbäddade märket får vi enligt figur 9 ut märket/signaturen *s'*. Om själva inbäddningsmetoden är robust kommer då *s'=s* även om bilden har utsatts för förändringar. Eftersom signaturberäkningen är bräcklig kommer dock *s''≠ s*. Detta leder till att i verifieringssteget kommer vi upptäcka att *s'≠ s''* och vi vet därmed att bilden har blivit förändrad.

Om inbäddningsmetoden av märket är bräcklig eller halvbräcklig och bilden har blivit förändrad uppstår istället situationen att det utvunna märket är *s'≠ s*. Vi har dessutom att den beräknade signaturen är *s''≠ s*. Med väldigt hög sannolikhet kommer vi då få åter igen få att *s'≠ s''* i verifieringssteget. Med andra ord kan vi välja om vi vill använda en robust, halvbräcklig eller bräcklig inbäddningsmetod.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig	B	B	B
Halvbräcklig	B	HB	HB
Känd=Robust	B	HB	R

Tabell 1: Tabell över de olika bräcklighetsfallen. Det är nio olika fall som kan uppstå beroende på val av signatur och inbäddningsmetod. Den av signaturberäkningsmetoden och märkningen som är mest bräcklig styr hur bräcklig hela vattenmärkningsmetoden blir. B = Bräcklig, HB = Halvbräcklig och R = Robust.

Eftersom det alltså både finns bräckliga, halvbräckliga samt robusta vattenmärkningsalgoritmer (presenterade i sektion 2.1.4) och bräckliga, halvbräckliga samt kända signaturer/märken som bäddas in, uppstår nio olika möjliga fall för en vattenmärkningsmetod vilket illustreras i tabell 1. Den totala robustheten för en metod blir den lägsta robusthetsnivån av vattenmärkningen och signaturen. Därmed uppstår fem fall med bräcklig vattenmärkning, tre halvbräckliga fall samt ett fall som är helt robust (och därmed inte intressant ur förändringsindikeringssynpunkt). Innan val av metod bestäms bör informationen i denna tabell tas i beaktning så att metoden uppfyller de krav som ställs gällande typ av bräcklighet.

I kapitel 4 kommer vi illustrera vilken av de olika fallen som varje förändringsindikationsmetod tillhör genom att rita ut tabellen för varje metod.

Övervakningsscenario:

Om man misstänker att övervakningbilder har blivit manipulerade kan vi använda förändringsindikeringen för att autentisera dem. Om vi använder en lokaliserande förändringsindikering kan vi upptäcka i vilka delar som övervakningsbilden har blivit förändrad. Exempelvis kan en person blivit borttagen eller en nummerplåt blivit utbytt.

Det bräckliga eller halvbräckliga vattenmärket bäddas in i bilden i sensorn och efter överföring till övervakningscentralen kan alla mottagna bilder autentiseras och om förändringar upptäcks skall vakten bli larmad.

För fallet med en övervakningsbild som spridits på nätet eller till press går det genom den bräckliga eller halvbräckliga vattenmärkningen avgöra om övervakningsbilden stämmer eller om den har blivit manipulerad.

2.3.4 Återskapande

Vissa förändringsindikeringsmetoder har dessutom med ett återskapande steg där de delar av bilden som blivit förändrade delvis kan återställas. Det finns flera metoder för återskapande vattenmärkning men de vanligaste metoderna använder sig av två vattenmärken:

- Först skapas en nedsamplad, hårt kvanitiserad och komprimerad version av originalbilden. Denna lågupplösta version av ursprungsbilden bäddas in som ett robust vattenmärke i ursprungsbilden.
- Bilden som skapades i förra punkten blir sedan inbäddad med ett förändrings-indikerande vattenmärke genom någon bräcklig eller halvbräcklig metod.

Vid kontroll av om bilden har förändrats sker då båda steg:

- Först utvinns det bräckliga eller halvbräckliga vattenmärket och en kontroll sker om bilden har förändrats och i så fall i vilket område av bilden.
- Förutsatt att metoden i föregående punkt kommit fram till att bilden blivit förändrad utvinns därefter den robust inbäddade lågupplösta versionen av originalbilden. Denna används får att ersätta de delar av bilden som blivit markerade som förändrade.

De områden i bilden som återskapas med hjälp av den lågupplösta versionen av originalbilden kommer självklart inte innehålla alla detaljer som fanns med i originalbilden men kan förhoppningvis ge en uppfattning om vad som försvunnit. En sak att tänka på är att områden i den inbäddade lågupplösta versionen av bilden skall bäddas in i andra områden av bilden så att till exempel informationen från det övre högra hörnet bäddas in i det nedre vänstra hörnet. Idealt bör dessutom den robusta inbäddningen innehålla redundans så att all vattenmärkesinformation finns lagrad på två eller fler ställen i bilden.

Övervakningsscenario:

Om en medveten manipulation genomförts för att ta bort en bil från en övervakningsbild går det genom återskapandet se att det är en bil som blivit borttagen. Den låga upplösningen och kvalitén på de återskapade delarna av bilden gör det kanske svårt att exakt identifiera bilen (nummerplåten kan kanske inte utläsas) men bara informationen om att det är en bil som tagits bort kan vara värdefull.

3 Metoder för identifikationsmärkning

I detta kapitel presenterar vi kortfattat några olika exempel på metoder som publicerats inom området robust vattenmärkning. Alla dessa metoder kan användas för att bädda in information i bilder. Eftersom antalet publicerade artiklar inom denna del av vattenmärkningsområdet är väldigt stort har vi valt att inte göra någon djupdykning utan letat ut vissa typexempel på metoder.

3.1 Spatialdomän

I denna sektion presenteras vattenmärkningsmetoder som bäddar in i spatialdomän.

A Watermarking Technique for Digital Imagery: Further Studies

Raymond B. Wolfgang and Edward J. Delp

Metoden beskriven i [Wolfgang97] är en gammal så kallade substitutsvattenmärkning där märket bäddas in i spatialdomän. Inbäddningen sker genom att vissa förutbestämda bitar i bildens pixlar byts ut mot vattenmärkets bitar. Vid utvinning av vattenmärket är det känt vilka bitar för pixlarna som innehåller vattenmärket och dessa utläses. För att göra metoden mer robust går det om vattenmärket inte är för stort bädda in märket i flera icke överlappande block i bilden och sedan vid utvinningen beräkna det mest sannolika värdet på varje vattenmärksbit.

Om vattenmärket till exempel är en binär bild och den skall bäddas in robust (dvs i pixelbitar med rätt hög signifikans) finns det risk för att vattenmärket kommer bli synligt. Allmänt brukar denna metod av inbäddning i spatialdomän inte anses vara den bästa.

Spatial Domain Digital Watermarking of Multimedia Objects for Buyer Authentication

Dipti Prasad Mukherjee, Subhamoy Maitra och Scott T. Acton

Metoden i [Mukherjee04] är en så kallad additiv vattenmärkning som märker i spatialdomän. Kortfattat går inbäddningen till så att vattenmärket adderas till bilden på vissa eller alla pixlar genom att värdet på pixlarna förändras om en etta bäddas in eller behåller originalvärde om nolla bäddas in. I just den här metoden har det hela utökats lite så att speciella block med mönster av additioner, subtraktioner eller orörda pixlar används för att utöka robustheten för de inbäddade bitarna.

Till utvinningen av vattenmärket behövs originalbilden. Detta är med andra ord en icke-blind vattenmärkning. Detta är en stor svaghet med additiva vattenmärken.

Additativa vattenmärken kan också bäddas in i frekvensdomän och förutsatt att det finns en linjär transformation mellan frekvens- och spatialdomänen blir det ingen skillnad i vilken domän som inbäddningen görs.

Techniques for data hiding

W. Bender, D. Gruhl, N. Morimoto och A. Lu

I [Bender96] presenterades en ofta använd vattenmärkningsmetod kallad patchwork. Denna metod bygger på statistik och antagandet att om vi slumpmässigt adderar *x* stycken pixelvärden från bilden och subtraherar dem med *x* stycken andra slumpmässigt valda pixlar kommer vi att få en summa som ligger nära noll.

För varje vattenmärkningsbit finns ett antal pixlar i bilden kopplade vars positioner bestäms genom algoritmen eller en hemlig nyckel. Hälften av dessa pixlar ingår i grupp A och hälften i grupp B. Om vi till exempel skall bädda in en etta för den vattenmärknings-biten så adderar vi alla pixlar i grupp A med ett och subtraherar alla pixlar i B med ett. För inbäddning av en nolla gör vi tvärtom, alla pixlar i grupp A subtraheras med ett och alla pixlar i grupp B adderas med ett.

Vid utvinningen så kontrolleras summan av pixlelvärdena i grupp A subtraherad med summan av pixelvärdena i grupp B. Om vi då får ett positivt tal utvinns en etta för den vattenmärkningspositionen och om talet är negativt utvinns en nolla.

Denna metod blir självklart robustare desto större antal pixlar som finns i de båda grupperna för varje vattenmärkningsbit. Fast antalet bitar som kan bäddas in minskar då robustheten ökar så en avvägning måste göras. Om antalet pixlar i de båda grupperna är små ökar dessutom risken för att det blir fel i utvinningen även om bilden inte har utsatts för någon attack eftersom sannolikhetsantagandet i början då inte är lika giltigt. Om vi vill minska sannolikheten att få fel går det sätta ett tröskelvärde för ett intervall kring noll som summan i utvinningen måste ligga utanför, om inte anses den positionen i vattenmärket ha blivit förstörd.

3.2 Frekvensdomän

I denna sektion presenteras ett par vattenmärkningsmetoder som bäddar in meddelanden i någon frekvensdomän.

A Robust DCT-based Watermarking for Copyright Protection

Shinfeng D. Lin och Chin-Feng Chen

Denna metod [LinChen00] använder discreta cosinustransformen (DCT) i bildblock av storlek *8 x 8* pixlar. Metoden är en typ av substitutionsvattenmärkning där vattenmärket bäddas in i den minst signifikanta biten (Least Significant Bit - LSB) för vissa låga frekvenskomponenter. Textur-komplexiteten för varje block beräknas innan inbäddningen och används för att bestämma hur vattenmärket skall bäddas in för att inte påverka bildens visuella kvalité.

A Secure, Robust Watermark for Multimedia

Ingemar J. Cox, Joe Kiliant, Tom Leighton och Talal Shamoon

Denna artikel [Cox96] presenterar den så kallade ”spread spectrum”-vattenmärkningen som bäddar in i frekvensdomän. Dels finns en additativ variant men även två så kallade multiplikativa vattenmärkningsmetoder. Dessa går ut på att ett vattenmärket multipliceras med frekvenskomponenterna i bilden. En viktning används som kontrollerar hur starkt det inbäddade vattenmärket påverkar bilden. Denna viktning kan göras beroende av bilden så att vattenmärket bäddas in kraftfullare på ställen där det inte kommer synas lika tydligt för blotta ögat. Detta gör att inbäddningen kan göras robustare än om en konstant vikt används över alla frekvenskomponenter.

För att utvinnan vattenmärket behövs originalbilden så denna typ av vattenmärkning är icke-blind. Det utvunna vattenmärket kan korreleras mot alla möjliga vattenmärken för att få fram det ursprungliga. Denna typ av vattenmärkning visar sig vara robust mot många attacker. Om två märkta bilder medelvärdesbildas eller om till exempel fyra olika märken alla bäddas in med denna metod i en och samma bild kommer korrelationen mellan vattenmärkena ändå identifiera dem alla.

För att denna metod skall kunna användas för att utvinna en binär kod behövs det troligtvis införas redundant inbäddning av informationen och en bar kod-separation. Den additativ version av spread spectrum har många likheter med additativ inbäddning i spektraldomän. Spread spectrum-metoder för vattenmärkning är väldigt vanliga.

4 Metoder för förändringsindikation

I detta kapitel presenterar vi en litteraturstudie över publicerade artiklar och konferens-bidrag som handlar om hur vattenmärkning använts för förändringsindikation. Vi har delat in metoderna i tre kategorier beroende på om vattenmärket bäddas in i spatial-domän, frekvensdomän eller krusningsdomän. För varje metod besvaras två frågor och markeras i en bräcklighetstabell enligt tabell 1.

4.1 Spatialdomän

I denna sektion beskriver vi metoder som bygger på bräckliga/halvbräckliga-vattenmärkning i spatialdomän.

Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification

P.W. Wong och N. Memon

Denna metod [Wong01] har sitt ursprung i [Wong98] men den visade sig vara sårbar för den så kallade vektorkvantiseringsattacken [Holliman97]. Denna nya version är dock robustare mot den typen av attack. Metoden finns i två versioner, en med hemlig nyckel och en med öppen nyckel. En hemlig nyckel behövs för att bädda in vattenmärket i båda fallen, men för det andra fallet är det en annan öppen nyckel som används för att utvinna vattenmärket och kunna upptäcka om någon förändring av bilden har skett.

Bilden delas in i block och den minst signifikanta biten (Least Significant Bit - LSB) för varje pixel nollställs. Med hjälp av pixlelvärdena (förutom LSB) i ett sådant block samt den hemliga nyckel och några andra parametrar beräknas en hash. Den föreslagna hashfunktionen är MD5 [Rivest92] med en resulterande bitlängd på 128. Denna signatur bäddas sedan in i pixlarnas LSB i blocket.

Mottagaren kan sedan själv beräkna vad signaturen bör vara för varje block givet en hemlig eller publik nyckel (beroende på metod). Sedan är det enkelt att undersöka om den beräknade signaturen stämmer med informationen i LSB.

Metoden saknar möjlighet för återskapande men i [Tao06] beskrivs hur återskapande kan tillföras metoden. Lösningen involverar dock att återskapandeinformationen (index-karta och DCT-koefficienter) sänds utanför själva bilden som exempelvis headerdata. Detta bryter mot tanken om att all information skall finnas inbäddad i själva bilden vilket gör att denna återskapningmetod inte är så intressant.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej (ja för metoden i [Tao06]).

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig	X		
Halvbräcklig			
Känd=Robust			

Toward Secure Public-key Blockwise Fragile Authentication Watermarking

P.S.L.M. Barreto, H.Y. Kim och V. Rijmen

Artikeln [Barreto02] beskriver hur hashning och öppen nyckel-kryptografi kan användas för att autentisera bilder. Författarna klagar på att en metod av Wong [Wong98], som bygger på hashning och kryptering av hash-numret blockvis, inte är säker. Det är lätt att byta ut block genom att ta dem från andra bilder som är kodade med samma algoritm. Det går också att brute-force ändra ett bildblock tills hash-numret stämmer. Wongs metod använder också 64 bitars RSA kryptering som är väldigt osäker.

De beskriver en ny metod som de kallar HBC1 (Hash Block Chaining) som löser delar av problemen. Hash-numret för varje block skapas genom att använda information från närliggande block. Utbyte av block kan då inte göras utan att upptäckas. Det är också svårt att brute-force ändra en bild för att få samma hash-nummer just på grund av det cirkulära beroendet mellan blocken. De visar en ny attack som kan användas mot HBC1 som de kallar transplantationsattack. Det som krävs för att attacken ska kunna utföras verkar dock väldigt osannolikt. De beskriver även en ny algoritm som de kallar HBC2 som står emot transplantationsattacken.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig	X		
Halvbräcklig			
Känd=Robust			

A Hierarchical Digital Watermarking Method for Image Tamper Detection and Recovery

Phen Lan Lin, Chung-Kai Hsieh och Po-Whei Huang

Metoden i [Lin05] går, i en förenklad beskrivning, till på följande sätt:

- En bild delas upp i block om 4x4 pixlar, varefter blocken numreras [1:N]
- Med hjälp av en hemlig nyckel k skapas en 1-1-avbildning från mängden [1:N] till sig själv, det vill säga en permutation. På så sätt pekar varje block på ett annat "hemligt" block, som i och för sig kan vara det aktuella blocket själv.
- I varje block om 4x4 pixlar skapas fyra subblock om 2x2 pixlar, och ur dessa subblock extraheras 8 bitars information. 1 bit avdelas till ett autentiserings-vattenmärke, 1 bit anslås till en paritetskontroll, och 6 bitar används för att beskriva en medelintensitet hos subblocket.
- Om block A länkas till block B nollställs i B de två LSB:erna i varje pixel, och för varje subblock sparas där autentiseringsvattenmärkesbiten och paritetskontroll-biten hos B självt, samt intentsitetinformationen hos motsvarande subblock i block A.

Man kan sedan kontrollera äktheten hos den erhållna biten på olika nivåer, såsom subblocknivå, blocknivå eller blockomgivningsnivå.

I [Chang08] presenteras en attack på denna vattenmärkningsmetod. Angreppet bygger på att vattenmärkningsmetoden är känd sånär som på den hemliga nyckeln k. I korthet går angreppet till så att man dels för varje subblock beräknar intensiteten (med de två LSB nollställda) och skapar en lista med den informationen samt dels ur varje subblock ser efter vilken intensitet (från ett annat block) den är märkt med och skapar en annan lista med denna information.

Genom att utnyttja att subblocken grupperas i block kan man sedan matcha de två listorna mer eller mindre framgångsrikt. Genom ett iterativt förfarande kan tvetydigheter (typiskt) klaras av och man kan till slut lista ut vilken permutation den hemliga nyckeln *k* get upphov till. Med denna kännedom kan sedan bilden modifieras och märkas om.

Författarna nämner också möjligheten att göra ett blint angrepp. Det bygger på att man kan ändra bildinformationen utan att känna den hemliga nyckeln så länge som subblockmedelintensiteten bevaras.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Ja.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig	X		
Halvbräcklig			
Känd=Robust			

A Wavelet-Based Fragile Watermarking Scheme for Secure Image Authentication,

Hong-Jie He, Jia-Shu Zhang och Heng-Ming Tai

Detta papper [He06] beskriver ett bräckligt vattenmärke som i viss mån utntyttjar krusningar.

Genom att lågpassfiltrera bilden fås en bild av mindre format (och med lägre upplösning). Pixlarna i denna lågupplösta bild 4-bitskvantiseras med hjälp av en hemlig kvantiserings-algoritm. Dessa 4 bitar placeras sedan i de minst signifikanta bitarna (LSB) hos den ursprungliga bilden. Det görs dessutom på ett sådant sätt att 4 bitar erhålla från en del av bilden placeras in i en annan del av bilden. Genom att använda denna metod blir vattenmärkningen svårare att angripa, samtidigt som man kan avgöra om eventuella angrepp skett mot det egentliga bildinnehållet eller mot själva vattenmärket.

Metoden kräver att par hemliga nycklar samt en användarsatt tröskelnivå som diskriminerar mellan olika typer av attacker.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig	X		
Känd=Robust			

4.2 Frekvensdomän

I denna sektion beskriver vi metoder som bygger på bräckliga/halvbräckliga-vattenmärkning i frekvensdomän.

Digital Fragile Watermarking Scheme for Authentication of JPEG Images

C.-T. Li

I [Li04] blir bilden DCT-transformerad och en bit-sekvens A genereras med hjälp av en hemlig nyckel. En annan bit-sekvens B skapas som pekar ut vilka DCT koefficienter som är 0 och vilka som inte är 0. Vattenmärket skapas genom att ta XOR mellan A och B. För varje DCT block väljs 4 nollskilda högfrekvens-koefficienter ut och kallas för "watermarkable". Dessa moduleras med de korresponderande bitarna från vattenmärket

och med en hemlig summa som beräknas från nollskilda komponenter från ett grannsystem (block).

Autentiseringen utförs genom att repetera alla stegen baklänges för att få ut vattenmärket, generera ett märke igen från bilden och sedan jämföra dom.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig		X	
Känd=Robust			

Semi-Fragile Watermarking for Authenticating JPEG Visual Content

Ching-Yung Lin och Shih-Fu Chang

Artikeln [Lin00] beskriver en metod som lägger till två typer av information i en bild, autentiseringsinformation och återskapningsinformation. Metoden är blockbaserad och med hjälp av autentiseringsinformationen kan man se om ett block har förstörts. Återskapningsinformationen kan då användas för att återskapa blocket med lägre upplösning.

Authentiseringsbitar genereras genom att titta på skillnaden mellan DCT komponenter från par av block. Om skillnaden är större än 0 så får man en bit som är 1 och om skillnaden är mindre än 0 så får man en bit som är 0. Det finns två satser i artikeln som beskriver varför skillnaden används. Den ska vara invariant när man gör JPEG komprimering.

Återskapningsinformation genereras genom att ta orginalbilden och skala ner den med en faktor 2 i varje axel. Sedan delas bilden upp i 8x8 block och samma teknik som i JPEG används för att kvantisera DCT komponenterna i varje block. Man kan säga att de gör en låg kvalitet JPEG komprimering av en nedskalad bild.

Informationsbitarna läggs in i ursprungsbilden genom att ändra på enskilda DCT komponenter med ett kvantiseringssteg som är större än det som används i JPEG komprimeringen. Då kan man enligt dem rekonstruera den ändring även efter komprimering utförts, vilket ger den halvbräckliga egenskapen. Autentiserings-informationen läggs in i de blocken som de hör till och återskapningsinformation sprids ut i hela bilden för att kunna användas även när en del av bilden har ändrats.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Ja, inbäddar en lågupplöst bild av sig själv.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig		X	
Känd=Robust			

4.3 Krusningsdomän

I denna sektion beskriver vi metoder som bygger på bräckliga/halvbräckliga-vattenmärkning i krusningsdomän.

Multiscale Fragile Watermarking Based on the Gaussian Mixture Model

Hua Yuan och Xiao-Ping Zhang

Detta arbete [Yuan06] bygger på en bilds statistiska egenskaper i krusningsdomänen, och behandlar bräckliga vattenmärken. En tidigare variant av metoden finns beskriven i [Yuan03]. Genom att arbeta i krusningsdomänen kommer, om lämplig krusningsbas är vald, förhållandevis få pixlar i bilddomänen att modifieras. Tanken är att bilden i krusningsdomänen delas i två delar. Vanligtvis en mindre vattenmärkningsdel och en större referensdel. Via en statistisk ansats modelleras sedan krusningskoefficienterna i var och en av de bägge delarna som en summa av två gaussiska fördelningar som skattas fram. Man kan säga att de bägge gaussiska fördelningarna representerar en vanlig gaussisk grundfördelning, samt en svansfördelning som representerar bidrag från förekomster av kanter i orginalbilden. Man skattar sålunda två varianser för vattenmärkningsdelen samt två varianser för referensdelen.

Vattenmärkningen går till så att krusningskoefficienterna i vattenärkningsdelen ändras på så sätt att variansen för svansfördelningen blir precis lika med variansen för svansfördelningen i referensdelen. Finessen är att man ändrar på koefficienter som representerar kanter, varför relativt få bildpixlar kommer att ändras. Fortfarande kommer varje ändring av bilddata att ändra skattade parametrar på ett sådant sätt att ändringen uppmärksammas.

- *Global eller lokal upptäckt av manipulering?* Global.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig	X		
Känd=Robust			

Improve Security of Fragile Watermarking via Parameterized Wavelet

Ji-Wu Huang, Jun-Quan Hu, D. Huang och Y.Q. Shi

Detta arbete [Huang04] behandlar vattenmärkning i krusningsdomänen. Den föreslagna vattenmärkningen är halvbräcklig och bygger på parametriserad heltalig krusnings-transform. Fördelen ligger i parametriseringen, där valet av parameter, vilken styr valet av krusningstransform, gör att det bara är personer med kännedom om parametervärdet som vet exakt vilken krusningstransform som har använts. På så sätt är parametern hemlig. En svårighet med metoden är att man bland alla teoretiskt tänkbara parametrar måste välja sådana som också i praktiken fungerar numeriskt. Olämplig valda parametrar kan vara sådana att numeriska avrundnings- och fortplantningsfel i beräkningarna tas för medvetna modifieringar.

Författarna kan , enligt artikeln, välja parametervärden med tillhörande transfomer, som är sådana att dels numeriken blir den önskade, metoden dels klarar kompresion a la JPEG, men att ändå det extraherade vattenmärket blir oandvändbart om inte den valda parametern är känd mer eller mindre exakt. Slutsatserna bygger på experiment och "noggrant valda tröskelvärden".

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig		X	
Känd=Robust			

Multiresolution Fragile Watermarking using Complex Chirp Signals for Content Authentication

Dan Yu, Farook Sattar, Braham Barkat

Detta arbete [Yu06] beskriver en bräcklig vattenmärkningsprocedur, som nyttjar vattenmärken i form av complexa chirpsignaler. En tidigare version av metoden finns

beskriven i [Yu04]. Vattenmärkningen, som bygger på kända (hemliga) nycklar, erbjuder autenticering av bilden utan att orginalbilden eller de exakt detaljerna kring chirpsignalen är kända.

Vattenmärkningen kan förenklat beskrivas som följer. Givet en komplex chirpsignal, så erhålls en följd av komplexa tal med belopp *1*. Dessa komplexa tal kvantiseras, och deras respektive real- och imaginärdel bäddas in i krusningstransformen enligt förbestämda regler. Bland annat görs en permutation en samplen, och det sker via en hemlig nyckel. Dessutom sker själva inbäddningen, det vill säga förändringar i krusningskoefficienterna, via något som kallas förbättrad kvantiseringsindexmodulation. Detta förbättrar prestanda vid extrakion.

Genom att känna den hemliga nyckeln samt kvantiseringsförfarandet nämnt ovan så blir inbäddningen, som alltså är bräcklig, möjlig att extrahera utan att känna de exakta parametrarna hos den inbäddade chirpsignalen.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig			
Känd=Robust	X		

Semi-Fragile Image Authentication using Generic Wavelet Domain Features and ECC

Qibin Sun och Shib-Fu Chang

Detta arbete [Sun02] beskriver hur man kombinerar många verktyg för att erhålla ett halvbräckligt vattenmärke. Metoden kombinerar krusningar, felrättande koder samt en slutgiltig kryptering av en signatur.

För att göra vattenmärkningen halvbräcklig gör man först en speciell filtrering med sikte på att tillåta vissa acceptabla manipulationer. Processen kallas avbrusning och bygger på ansatser/heuristik. I nästa steg fokuserar man på krusningskoefficienter av viss minsta styrka, det vill säga man gör ett tröskelförfarande i krusningsdomän. Tröskelvärdet styr nivån på bräckligheten. Enligt artikeln visar erfarnheten att metoden blir stabilare om man dessutom gör så kallade morfologiska operationer på data. Efter detta bestäms själva vattenmärket, också ur rådande data. Vattenmärket ges sedan ett visst mått av redundans, vilket innebär att felrättande koder (ECC) kan användas. Vattenmärket bäddas sedan in på ett eller annat sätt.

Utöver detta hashsas även vattenmärket varefter det krypteras med en hemlig nyckel och sparas separat. Det kallas då global signatur. Autenticeraren kan sedan dels dekryptera den globala signaturen med vattenmärkarens öppna nyckel, dels extrahera vattenmärket ur bilden och jämföra de två.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig		X	
Känd=Robust			

A DWT-Based Fragile Watermarking Tolerant of JPEG Compression

Junquan Hu, Ji-Wu Huang, Daren Huang och Yun Q. Shi

Detta arbete [Hu03] beskriver, förhållandevis översiktligt, ett halvbräckligt vattenmärke som bäddas in i krusningsdomänen. Metoden bygger också på människans synuppfattning genom att använda en kvantiseringsmetod som tar hänsyn till hur ögat och synapparaten fungerar. Vidare presenteras en metod för att skilja mellan medveten och omedveten/slumpmässig manipulering. Metoden som helhet kan också avgöra var eventuella manipulationer skett, samtidigt som bilden ändå tål JPEG-komprimering utan att signalera för manipulation.

Själva vattenmärket är en binär mask (logo), som har representationer med olika upplösning. För att dels vara osynliga för ögat, och dels klara JPEG-komprimering, bäddas vattenmärket in i "vissa delband" hos den krusningstransformerade bilden. Inbäddningen görs efter ett parameterval som styr detaljerna i metoden. Samma parameterval styr också metodens bräcklighet samt förmåga att bädda in data.

En metod föreslås för att avgöra om bildförändringar är oavsiktliga eller medvetna, och artikeln avslutas med experimentella illustrationer av den presenterade tekniken.

- *Global eller lokal upptäckt av manipulering?* Lokal.
- *Kan den återskapa förstörda områden?* Nej.

Signatur\Märkning	Bräcklig	Halvbräcklig	Robust
Bräcklig			
Halvbräcklig			
Känd=Robust		X	

5 Slutsatser och framtida arbete

Vattenmärkning är en metod att bädda in information i olika typer av sensordata såsom bilder, video och ljud. Det inbäddade vattenmärket kan utvinnas ur sensordatan för att sedan användas i olika säkerhetsapplikationer. Vattenmärkningen skall inte användas istället för kryptering utan skall ses som ett komplement eftersom de har skilda svagheter och styrkor och dessutom olika tillämpningsområden.

Vattenmärkena kan vara synliga eller osynliga, där framför allt de senare dessutom delas in i rubusta, halvbräckliga eller bräckliga vattenmärken. På liknande sätt kan själva vattenmärkningen delas in i blind eller icke-blind, privat eller öppen samt bräcklig, halvbräcklig eller robust vattenmärkning.

Robusta vattenmärken skall vara svåra att ta bort och används exempelvis för identitetsinbäddning där information så som upphovsman, sensor-id och tidkod bäddas in i bilden. Vi har beskrivit olika kända angreppssätt för inbäddning såsom substituts-, additiv- och multiplikativ vattenmärkning samt kända metoder som spread spektrum- och patchworkvattenmärkning.

Hur robust en vattenmärkningsmetod för stillbilder är kan undersökas genom att utsätta bilder för olika attacker som förstörande kompression och införande av brus. När sedan vattenmärket utvinns går det att se om det har överlevt attackerna eller ej.

Bräckliga och halvbräckliga vattenmärken används för att verifiera att en bild inte har förändrat. Bräckligheten kan endera bero på att själva inbäddningen av märket är sedan att det lätt skadas om bilden förändras eller också beror det på att det inbäddade märket är en bräcklig signatur som beror på bilden. Det går dessutom kombinera dessa så båda inbäddningen och signaturen är bräckliga eller halvbräckliga.

Vissa förändringsindikationsmetoder kan lokalisera var i en bild som förändringar har skett och det finns dessutom metoder som helt eller delvis kan återskapa förstörda områden.

En form av vattenmärken kan dessutom vara inverterbar vilket betyder att originalbilden före den ursprungliga inbäddningen av vattenmärket kan återskapas. Detta kräver dock att bilden inte har blivit förändrad sedan inbäddning av vattenmärket samt oftast att originalbilden uppfyller vissa krav innan den ursprungliga inbäddningen.

Som synes finns en flora av metoder och kombinationer, och efter föreliggande inventering gäller det att matcha utbudet mot försvarmaktens (FM) behov. I dagsläget har FM:s säkerhetsarbete mer inriktats mot skydd mot obehörigt intrång och/eller kryptering av känsliga data, vilket är naturligt med tanke på de typer av dataströmmar som är/har varit vanligast. I takt med ökande datamängder och kanske framförallt ett breddat innehåll är det av vikt att FM:s behov av säkerhetsåtgärder av den typ som beskrivs i denna rapport klargörs.

Arbetet i projektet Vattenmärkning av sensordata fortsätter 2009 med fortsatt utvärdering av existerande metoder och i vissa fall utökning av dem eller konstruktion av ytterligare

metoder för de tre tillämningsområdena. Vidare bör, i dialog med FM, FM:s behov adresseras. Vi kommer dessutom titta mer på tillämpningen att använda vattenmärkning för att lokalisera källan för olovlig spridning av senordata, så kallade ”fingerprinting” eller individuella vattenmärken.

6 Referenser

[Barreto02] P.S.L.M. Barreto, H.Y. Kim and V. Rijmen, “Toward Secure Public-key Blockwise Fragile Authentication Watermarking”, IEE Proceedings on Vision, Image and Signal Processing, vol. 149, no. 2, pp. 57-62, Apr 2002.

[Bender96] W. Bender, D. Gruhl, and N. Morimoto, “Techniques for Data Hiding”, IBM System Journal, vol. 35 (3), pp.313-336, 1996

[Chang08] C. C. Chang, Y. H. Fan and W. L. Tai, “Four-Scanning Attack on Hierarchical Digital Watermarking Method for Image Tamper Detection and Recovery”, Pattern Recognition, vol. 41, no. 2, pp. 654-661, Feb 2008.

[Cox96] I. J. Cox, J. Kiliant, T. Leighton and T. Shamoon, “A Secure, Robust Watermark for Multimedia”, Lecture Notes in Computer Science, vol. 1174, pp. 185-206, 1996

[Goljan01] M. Goljan, J. J. Fridrich and R. Du, ”Distortion-Free Data Embedding for Images”, Lecture Notes in Computer Science, vol. 2137, pp. 27-41, 2001.

[Hartung99] F. Hartung and M. Kutter, “Multimedia watermarking techniques” In Proceedings IEEE: Special Issue on Identification and Protection of Multimedia Information, volume 87(7), pages 1079–1107, July 1999.

[He06] K. J. He, J. S, Zhang and H-M.Tai, “A Wavelet-Based Fragile Watermarking Scheme for Secure Image Authentication”, Lecture Notes in Computer Science, vol. 4283, pp. 422-432, 2006.

[Holliman97] M. Holliman and N. Memon, “Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes”, IEEE Trans.Image Processing, vol. 6, pp. 432–441, Mar. 1997.

[Hu03] J. Hu, J. W. Huang, D. Huang and Y. Q. Shi, “A DWT-Based Fragile Watermarking Tolerant of JPEG Compression”, Lecture Notes in Computer Science, vol. 2613, pp. 31-45, 2003.

[Huang04] J. W. Huang, J. Hu, D. Huang and Y. Q. Shi, ”Improve Security of Fragile Watermarking via Parameterized Wavelet”, Proceedings of the International Conference on Image Processing ICIP’04, vol. 2, pp. 721-724, Oct 2004.

[Kundur99] D. Kundur and D. Hatzinakos, ”Digital Watermarking for Telltale Proofing and Authentication”, Proceedings of the IEEE, vol. 87, no. 7 pp. 1167-1180, Jul 1999.

[Li04] C-T. Li, ”Digital Fragile Watermarking Scheme for Authentication of JPEG Images”, IEE Proceedings of Vision, Image and Signal Processing, vol. 151, no. 6, pp. 460-466, Dec 2004.

[Lin00] C-Y. Lin and S-F. Chang, “Semi-Fragile watermarking for authenticating JPEG visual content”, SPIE Security and Watermarking of Multimedia Contents II, vol.3971, pp. 140-151, May 2000.

[LinChen00] S. D. Lin and C-F Chen, “A Robust DCT-based Watermarking for Copyright Protection”, IEEE Transactions on Consumer Electronics, vol. 46, 2000.

[Lin05] P.L. Lin, C.K. Hsieh and P.W. Huang, “A Hierarchical Digital Watermarking Method for Image Tamper Detection and Recovery”, Pattern Recognition, vol. 38, no. 12, pp. 2519-2529, Dec 2005.

[Mukherjee04] D. P. Mukherjee, S. Maitra and S. T. Acton, “Spatial Domain Digital Watermarking of Multimedia Objects for Buyer Authentication”, IEE Transactions on Multimedia, vol. 6, no.1, pp. 1-15, 2004.

[Ni06] Z. Ni, Y. Q. Shi, N. Ansari and W. Su, ”Reversible Data Hiding”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 3, pp. 354-362, March 2006.

[Oostveen02] J. Oostveen, T. Kalker and J. Haitsma. “Feature Extraction and a Database Strategy for Video Fingerprinting”, Lecture Notes in Computer Science, vol. 2314, pp. 117-128, 2002.

[Puate96] J. Puate and F. Jordan, “Using Fractal Compression Scheme to Embed a Signature into an Image“, Proceedings of the SPIE Photonics East’96 Symposium, pp. 108-118, Nov 1996.

[Rivest92] R. L. Rivest, “The MD5 message digest algorithm,” Tech. Rep., RFC 1321, April 1992.

[Simmons83] G. J. Simmons, "The Prisoner's Problem and the Subliminal Channel" in Advances in Cryptology: Proceedings of CRYPTO '83 pp. 51-67, 1983.

[Stenborg09] K-G. Stenborg, "Scalable Distribution of Watermarked Media", Chapter XVIII in Handbook of Research on Secure Multimedia Distribution, Edited by S. Lian, ISBN: 978-1-60566-262-6, Information Science Reference, to appear march 2009.

[Sun02] Q. Sun and S-F. Chang. “Semi-Fragile Image Authentication using Generic Wavelet Domain Features and ECC”, Proceedings of the International Conference on Image Processing ICIP’02, vol. 2, pp. 901-904, 2002

[Sun05] Q. Sun and S-F. Chang, "Signature-based Media Authentication”, Chapter 21 in Multimedia Security Handbook, Edited by B. Furht and D. Kirovski, CRC Press, 2004.

[Tao06] P. Tao and A. M. Eskicioglu, “An Adaptive Method for Image Recovery in the DFT Domain”, Journal of Multimedia, vol. 1, no. 6, pp. 36-45, 2006

[Wolfgang97] R. B. Wolfgang and E. J. Delp, “A Watermarking Technique for Digital Imagery: Further Studies” In Proceedings of the International Conference on Imaging, Systems and Technology, 1997.

[Wong98] P. W. Wong, “A Public Key Watermark for Image Verification and Authentication”, Proceedings of the International Conference on Image Processing ICIP’98, vol. 1, pp. 455-459, 1998.

[Wong01] P.W. Wong and N. Memon, “Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification”, IEEE Transactions on Image Processing, vol. 10, pp. 1593-1601, Oct 2001.

[Yu04] D. Yu, F. Sattar and B. Barkat, ”Multiresolution Fragile Watermarking using Complex Chirp Signal for Content Integrity Verification”, Proceedings of the International Conference on Image Processing ICIP’04, vol. 5, pp. 3427-3430, Oct 2004.

[Yu06] D. Yu, F. Sattar and B. Barkat, ”Multiresolution Fragile Watermarking using Complex Chirp Signals for Content Authentication”, Pattern Recognition, vol. 39, no. 5, pp. 935-952, May 2006.

[Yuan03] H. Yuan and X-P. Zhang, “Fragile Watermark Based on the Gaussian Mixture Model in the Wavelet Domain for Image Authentication”, Proceedings of the International Conference on Image Processing ICIP’03, vol. 1, pp. 505-508, Sept 2003.

[Yuan06] H. Yuan and X-P. Zhang, “Multiscale Fragile Watermarking Based on the Gaussian Mixture Model”, IEE Transactions on Image Processing, vol. 15, no. 10, pp. 3189-3200, Oct 2006.