



FullFokusBild
Maximera skärpa med hjälp av focus stacking

Bildteknik

TNM089

Max Eriksson & Tim Myllylä & Rikard Vestergaard

Projektarbete (PRA1)

Linköpings Universitet
26 oktober 2023

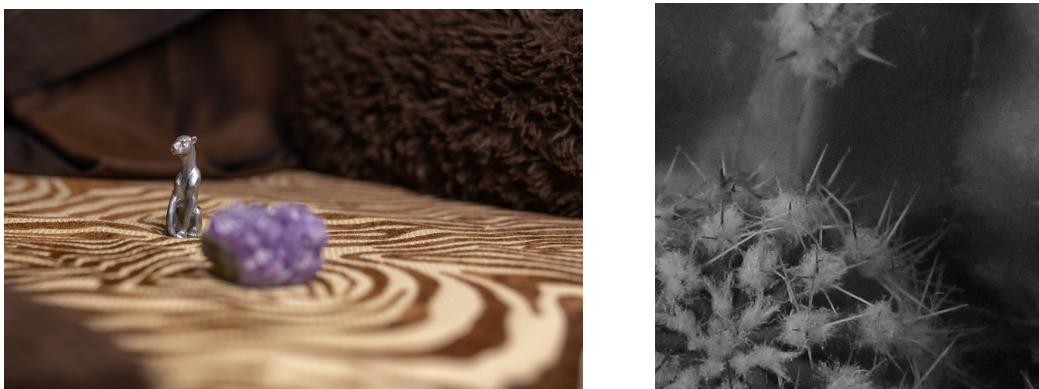
1 Introduktion

I den digitala bildbehandlingsvärlden är uppnåendet av fullständig skärpa och fokus i ett enda fotografi en utmanande uppgift, särskilt när man hanterar scener eller objekt med olika djup. För att överkomma detta problem har olika tekniker utvecklats för att skapa fullfokus bilder genom att kombinera flera bilder av samma scen som har olika fokuspunkter, även kallat ”focal stack”. Detta arbete utforskar användningen av Laplacian pyramider för konstruktion av ”focal stacks” med fler än två bilder och bygger på arbetet som Wencheng Wang och Faliang Chang diskuterar i ”*A Multi-focus Image Fusion Method Based on Laplacian Pyramid*” där de föreslår att Laplacian Pyramider kan användas för att konstruera en fullfokusbild av två bilder.

Laplacian pyramider är en hierarkisk representation av en bild som har varierande upplösning och innehåller information om skillnaden mellan två på varandra följande nivåer i pyramidstrukturen, medan fusion är processen där informationen från flera bilder med olika fokuspunkter kombineras för att skapa en enda bild med ökad skärpa och fokus. För konstruktionen undersöks två olika fusionstekniker för att generera fullfokusbilder från en uppsättning bilder. Den ena fokuserar på att kombinera endast två bilder från stacken åt gången och den andra bearbetar alla bilder samtidigt.

2 Metod

Focus stacking är en teknik inom fotografi och bildbehandling som används för att öka eller maximera skärpedjupet i en bild och är särskilt användbart i situationer där det är svårt att uppnå tillräckligt skärpedjup genom att använda en enda exponering. Under projektet har två olika ”focal stacks” använts, en bestående av bilder över en högtexturerad miljö, och en i gråskala, i Figur 1a respektive Figur 1b visas exempelbilder från dataseten.



(a) Exempelbild från datasetet med högtexturerad scen. (b) Exempelbild från datasetet i gråskala

Figur 1: Exempelbilder från två dataseten.

2.1 Laplacian pyramid-dekomposition

Inom bildbehandling är pyramid dekomposition en vanlig och välkänd teknik för åstadkomma en fusion av flera bilder. Inom detta område anses Laplacian pyramider vara en särskilt effektiv metod för att förbättra både konturer och kontrast i de resulterande sammanslagna bilderna [3]. Metoden utgör en flerskalig bildrepresentation där en bild bryts ner i en serie av bandpassbilder. Varje bild i serien fångar olika nivåer av detaljer och möjliggör därmed en flexibel och effektiv hantering av olika frekvenskomponenter.

De grundläggande stegen för att utföra en Laplacian pyramid är som följande:

1. Börja med att skapa en Gaussisk pyramid av den ursprungliga bilden g_0 genom att successivt applicera ett Gaussiskt filter och sedan ned-sampla bilden. (Varje nivå i pyramiden representerar en "blurrad" och ned-samplad version av g_0)
2. För varje nivå i den gaussiska pyramiden (förutom den högsta), "uppsampla" genom att infoga nollar mellan varje rad och kolumn och interpolera de saknade värdena med samma lågpassfilter som används i den gaussiska pyramiden. (Detta skapar en expanderad version av den akutella nivån).
3. Subtrahera den expanderade nivån från nästa högre nivå av den Gaussiska pyramiden. (Resultatet blir en nivå som fångar högfrekventa detaljer.)

Upprepa ovanstående steg för varje nivå av den Gaussiska pyramiden och skapa motsvarande nivåer i en Laplacian pyramid [2].

Arbetet som presenteras i denna rapport utforskar tre olika nivåer för Laplacian pyramiderna: 3, 5 och 7.

2.2 Fusionsmetoder

Enligt Wang & Chang finns det generellt sett två fusionsmetoder, en pixelbaserad och en regionsbaserad. Fördelen med pixelbaserad fusion är att algoritmerna är relativt lätt att implementera och tidseffektiva. Den mest grundläggande metoden innebär att ta medelvärdet eller den pixeln med högst intensitet. Men detta resulterar i begränsade resultat eftersom det inte tar hänsyn till sambandet mellan pixlarna och behandlar varje enskild pixel oberoende av de övriga. Wang och Chang presenterar i deras rapport, *A Multi-focus Image Fusion Method Based on Laplacian Pyramid*, en alternativ, regionsbaserad metod. Denna metod utmärker sig genom att välja områden som innehåller rikligare information som ett resultat.

För det översta lagret tillämpas en särskild metod för att välja område. Indikatorer som återspeglar informationsnivån i dessa områden inkluderar huvudsakligen avvikelse, D , som beräknas enligt Ekvation 1 och entropi, E , som definieras enligt Ekvation 2. Avvikelsen hänvisar till variationen i pixelvärdet inom en region, så områden med högre variation anses vara mer detaljerade. Å andra sidan är högre entropi en indikation på ökad komplexitet inom ett område. [2]

$$D = \sum_{m=1}^J \sum_{n=1}^K \frac{(X(m, n) - \bar{X})^2}{J \times K} \quad (1)$$

$$E = - \sum_{i=0}^{L-1} P_i \log P_i \quad (2)$$

där:

$X(m, n)$: Pixelvärdet vid koordinaterna (m, n)

\bar{X} : Medelvärdet av pixelvärdet

J, K : Storleken på den lokala regionen

L : Gråskalenivåer i bilden

P_i : Sannolikheten för gråskalenivåen i

I övriga lager tillämpas en regel som baseras på den maximala energin i regionen enligt Ekvation 3.

$$RE_l = \sum \omega^l(m', n') [LP_l(m + m', n + n')]^2 \quad (3)$$

där:

LP_l : den l:e Laplacian pyramiden
 $\omega^l(m', n')$: Viktmatriks-operator
 J, K : Storleken på den lokala regionen

2.3 Fusionstekniker

I Wang & Changs arbete låg fokus på att behandla endast två bilder åt gången (TÅG). I detta arbete, med ett dataset som omfattar över 20 bilder innebär detta att resultatet av varje fusion används som indata tillsammans med nästa bild i datasettet. Denna procedur upprepas successivt tills alla bilder har genomgått behandlingen. Vidare har fusionen tagits ett steg längre genom att utforska en alternativ metod, där alla bilder i datasettet behandlas samtidigt (AS). I denna metod skapas Laplacian pyramid för varje bild, och fusionen utförs sedan på varje nivå i pyramiderna. Denna metod tillåter en parallell bearbetning av alla bilder i datasettet.

I AS dekonstrueras alla bilder i bildserien till respektive Laplacian pyramid. Fusionssteget väljer därför delar från den bäst presterade kandidaten i varje pixel för att bygga upp en resulterade Laplacian pyramid med information från alla bilder i bildserien. Den stora skillnaden mellan TÅG och AS är att för AS sker det bara en konstruktion från Laplacian pyramid till bild medan i TÅG konstrueras det en bild för varje ihopslagning och sedan dekonstrueras varje ihopslagning med nästa bild i bildserien.

2.4 Utvärdering

De genererade bilderna utvärderas både på subjektiv och objektiv nivå där tre kvalitetsmått används, BRISQUE, NIQE och PIQE. Gemensamt för de tre måtten är att ett lågt värde är att föredra. Anledningen till att dessa kvalitetsmått valdes grundar sig i att samtliga är referensfria mått [1]. Det vill säga att dessa mått ej kräver en referensbild för att utvärdera kvaliteten hos en bild. Då det är svårt att hitta bildserier bestående av bilder med olika fokus och en tillhörande referensbild med totalt fokus som inte blivit konstruerad med hjälp av en "focus stacking"-metod krävdes det referensfria objektiva kvalitetsmått.

3 Resultat

Nedan visas de genererade fullfokus bilderna. Figur 2 visar resultatet från den pixelbaserade metoden där fem nivåer har använts i kombination med de två fusionsteknikerna som diskuteras i Sektion 2.3. Vidare visas resultatet från den regionsbaserade metoden i övriga figurer. För Figur 3 och 4 har alla bilder bearbetats samtidigt där tre olika nivåer i Laplacian pyramiderna har använts: tre, fem och sju. Resultatet i Figur 5 visar samma nivåer som föregående figurer men i detta fall har endast två bilder bearbetats åt gången. För den högtexturerade bilden i Figur 6 gjordes endast en rendering med fusionstekniken att bearbeta två bilder åt gången, denna använde sig av fem nivåer i Laplacian pyramiden. Slutligen i Figur 7 visas resultatet där metoden som presenteras i andra stycket av Sektion 2.2 även inkluderas.

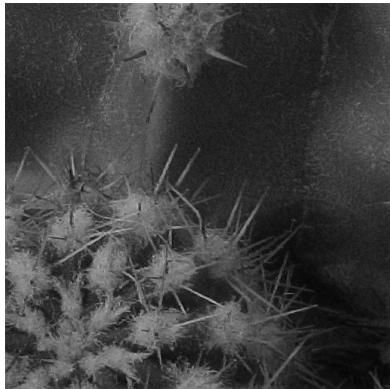


(a) Två åt gången, 5 nivåer, pixelbaserat



(b) Alla samtidigt, 5 nivåer, pixelbaserat

Figur 2: Resultat från pixelbaserad fusion.



(a) Tre nivåer



(b) Fem nivåer



(c) Sju nivåer

Figur 3: Resultat från regionsbaserad fusion där alla bilder bearbetats samtidigt (gråskala).



(a) Tre nivåer



(b) Fem nivåer

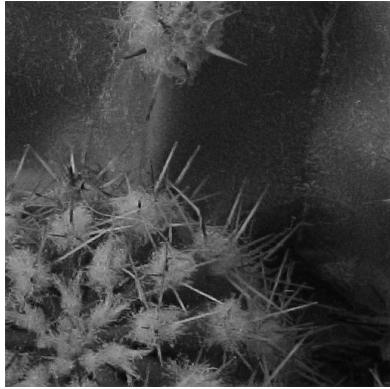


(c) Sju nivåer

Figur 4: Resultat från regionsbaserad fusion där alla bilder bearbetats samtidigt.



Figur 6: Resultat från regionsbaserad fusion där två bilder bearbetats åt gången, fem nivåer.



(a) Tre nivåer



(b) Fem nivåer



(c) Sju nivåer

Figur 5: Resultat från regionsbaserad fusion där två bilder bearbetats åt gången (gråskala).



Figur 7: Resultat med hänsyn till avvikelse och entropi i det översta lagret där alla bilder bearbetats samtidigt.

Tabell 1 redovisar de objektiva kvalitetsmåtten för de genererade gråskalebilderna, grönamarkerade och rödmarkerade värden är de bästa respektive sämsta för de olika kvalitetsmåtten.

Tabell 1: Tabell med kvalitetsmått för de genererade gråskalebilderna.

Metod	brisque	niqe	piqe
AS*, 3 nivåer	29.4934	4.6545	4.3255
AS*, 5 nivåer	30.4808	4.7327	4.0426
AS*, 7 nivåer	30.7139	4.7068	4.1739
TÅG**, 3 nivåer	32.0874	4.6375	3.8866
TÅG**, 5 nivåer	32.3536	4.6869	4.0009
TÅG**, 7 nivåer	32.5222	4.8005	4.2957

*AS: Alla bilder bearbetas samtidigt

**TÅG: Två bilder bearbetas åt gången

4 Diskussion

I detta kapitel diskuteras och utvärderas två huvudsakliga aspekter av fusionsmetoder som används i projektet: pixelbaserad och regionsbaserad fusion samt fusionsteknikerna TÅG (Två

åt gången) och AS (Alla samtidigt). Vidare undersöks och jämförs olika antal nivåer i Laplacian pyramiden samt diskuteras de objektiva och subjektiva kvalitetsmått som används för att utvärdera bildkvaliteten.

4.1 Pixelbaserad- och regionsbaserad fusion

Under arbetet testades två fusionsmetoder, pixelbaserad och regionsbaserad. Genom att okulärt besiktiga resultatet av båda metoderna gick det att fastslå att den reginonsbaserade fusionen presterade markant bättre. Fördelen med den pixelbaserade metoden är liten men genom att ha en mycket snabb beräkningstid och relativt enkel implementation så finns det användningsområden för denna metod. Den pixelbaserade metoden resulterade i grynpiga bilder med mycket artefakter, detta syns tydligt i Figur 2, vilket inte skulle kunna bidra till projektets mål att konstruera en fullfokusbild och därfor användes den regionsbaserade metoden vid fortsatt implementation.

4.2 Utvärdering av fusionstekniker

För att slå ihop bildserierna testades två olika implementationer, TÅG och AS. Baserat på de objektiva kvalitetsmåtten och gruppens subjektiva bedömning så presterade AS bättre. Vid användning av TÅG noterades artefakter i form av färgskiftningar, haloeffekt och dåligt fokus i vissa områden, i en betydligt större utsträckning i förhållande till AS, se Figur 6. En anledning till att TÅG presterar sämre kan grunda sig i den upprepade dekonstruktionen och konstruktionen. Då varje ihopslagning behandlas likvärdigt kan artefakter tolkas som detaljer och därfor ärvas till nästkommande steg. Artefakter som uppstår kan därfor i nästa steg behandlas som viktig information och efter flera upprepningar till och med förstärkas. Vidare kan därfor områden med kraftiga artefakter väljas över områden med mer fokus och på så sätt ge ett resultat med områden utan fokus.

Denna problematik är något som AS inte delar. När AS används så sker endast en konstruktion och artefakter bör således inte uppstå mer än en gång. Detta resulterar i en bild med till synes bättre totalt fokus då tekniken väljer delar ur den bild i bildserien som har bäst fokus (mest detaljer) i just det området, se Figur 4 för jämförelse mot Figur 6.

4.3 Jämförelse av olika nivåer i pyramiden

I dekonstruktionen av en bild till en Laplacian pyramid är en av parametrarna vi har kontroller över hur många lager denna pyramid ska bestå av. Varje lager innehåller information om grövre och grövre detaljer, det översta lagret innehåller de finaste detaljerna. I teorin är det detta lager vi är mest intresserade av då detta lager bör visa oss var fokusområdet i bilden är. I resultatet kan vi se att endast använda sig utav 3 nivåer medför mer haloeffekt vilket beror kan bero på att vi inte har tillräckligt mycket information för att skapa en naturlig gradient av fokus i det området, se Figur 4a. Det vill säga att vi går från ett område med väldigt skarp kant (mycket detalj) till ett område med mycket suddighet och vi har inte brutit ner bilden tillräckligt för att ge en trovärdig övergång. Effekten av detta är tydligare i högupplösta bilder på grund av att information i de lägre nivåerna fortfarande kan innehålla viktig information. Under projektet testades tre olika antal nivåer i pyramiden, 3, 5 och 7. Den subjektiva bedömningen är att fler nivåer resulterade i en bättre bild men att det var marginell skillnad mellan 5 och 7 nivåer. Genom att studera Tabell 1 kan vi se att de objektiva kvalitetsmåtten inte stämmer överens med den subjektiva bedömningen. Enligt samtliga objektiva mått ger 3 nivåer i Laplacian pyramiden en bild med bättre kvalitet. Anledningen till att dessa bilder skulle vara kvalitetsmässigt bättre är svårt att fastslå. De objektiva kvalitetsmåtten som användes är samtliga relativt känsliga för

brus och anser att det ger en bild sämre kvalitet. Genom att använda fler nivåer kan vi notera starkare kontrast i de resulterade bilderna se figur 4a och 4c för exempel. Denna kontrastskillnad skulle potentiellt kunna uppfattas som brus av kvalitetsmåtten och därmed bedöma dessa som bilder med lägre kvalitet.

4.4 Alternativ metod för översta lagret

I Wang & Changs rapport[2] diskuterades en alternativ metod för fusion i det översta lagret av Laplacian pyramiderna. För att undersöka om denna föreslagna metod även skulle gynna resultatet vid fusion av fler än två bilder implementerades denna vid slutskedet av projektet. I figur 7 kan vi se resultatet av en fusion som använder sig av denna metod. Bilden visar tydligare artefakter, mer specifikt tydliga haloeffekter och viss grynhet. Bilden har även lägre kontrast än motsvarande bild utan denna metod. I helhet presterar den här metoden sämre än tidigare genomförda fusioner. Resultatet är i bästa fall okej men det som verkligen talar emot användandet av denna metod för fusion av fler än två bilder är beräkningstiden. För att konstruera figur 7 tog det programmet ca 9 timmar, som referens tog samma "focal stack" fast utan att använda denna metod ca 300 sekunder att beräkna.

4.5 Kvalitetsmått

För att objektivt mäta bildkvaliten användes tre automatiserade metoder: BRISQUE, NIQE och PIQE. Trots deras förmåga att kvantifiera bildkvalitet har samtliga projektmedlemmars subjektiva observationer avvikat från dessa mätningar, och det föreligger en inkonsekvens mellan de automatiserade mätningarna och nivåerna i pyramiden.

Det finns flera möjliga orsaker till detta resultat

- Algoritmen för att bestämma kvalitén kan potentiellt ha svårt för att korrekt bedöma bilder med avancerat mönster, texturer och detaljer vilket är fallet för bilderna som användes i projektet.
- Om algoritmen tränas på ett dataset som inte är representativt för skillnader i fokus kan det leda till inkonsekventa resultat. Det är möjligt att algoritmerna är mer anpassade för att avgöra brus och distortioner i bilderna snarare än skärpa.
- Mänskliga bedömningar av bildkvalitet är subjektiva och varierar mellan individer. Automatiserade algoritmer kan ha svårt att replikera detta, vilket resulterar i skillnader i bedömningarna.

5 Sammanfattning

Denna rapport har utforskat och implementerat olika metoder för att fusionera bilder med varierande fokuspunkter för att generera bilder med maximalt skärpedjup. Användningen av Laplacian pyramidér har undersökts, och olika antal nivåer i pyramiderna har testats. Resultaten tyder på att regionsbaserad fusion överträffar pixelbaserad fusion och att en simultan behandling av alla bilder ger överlägsna resultat jämfört med att bearbeta endast två bilder samtidigt. Dessutom har en jämförelse av olika antal nivåer i Laplacian-pyramiden visat att en ökning av nivåerna kan förbättra kontrasten i bilderna.

Aspekter av metoden har diskuterats och utvärderats, inklusive användningen av objektiva och subjektiva kvalitetsmått. Trots vissa inkonsekvenser mellan dessa mått har slutsatser kunnat dras om vilka metoder och parametrar som resulterar i bättre prestanda. Slutligen har en

föreslagen metod från tidigare arbete undersökts, men dess otillfredsställande resultat och långa beräkningstid har observerats.

Denna rapport bidrar till förståelsen av hur olika tekniker för bildfusion kan tillämpas för att skapa bilder med maximalt skärpedjup och ger insikter om de faktorer som påverkar resultatet.

Litteraturförteckning

- [1] The MathWorks Inc. Statistics and machine learning toolbox, 2022.
- [2] Faliang Wang, Wencheng Chang. A multi-focus image fusion method based on laplacian pyramid. *Journal of Computers*, 2011.
- [3] Zhaobin Wang, Zijing Cui, and Ying Zhu. Multi-modal medical image fusion by laplacian pyramid and adaptive sparse representation. *Computers in Biology and Medicine*, 123:103823, 2020.