Automatic brightness control in near-infrared spectrum using approximate face region detection

Vugrin, Jurica

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Jurica Vugrin

AUTOMATIC BRIGHTNESS CONTROL IN NEAR-INFRARED SPECTRUM USING APPROXIMATE FACE REGION DETECTION

DOCTORAL THESIS

Zagreb, 2023

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Jurica Vugrin

AUTOMATIC BRIGHTNESS CONTROL IN NEAR-INFRARED SPECTRUM USING APPROXIMATE FACE REGION DETECTION

DOCTORAL THESIS

Supervisor: Professor Sven Lončarić, F.C.A.

Zagreb, 2023

Sveučilište u Zagrebu

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Jurica Vugrin

AUTOMATSKO UPRAVLJANJE SVJETLINOM U BLISKOM INFRACRVENOM PODRUČJU S APROKSIMATIVNOM DETEKCIJOM REGIJE LICA

DOKTORSKI RAD

Mentor: Akademik prof. dr. sc. Sven Lončarić

Zagreb, 2023.

The doctoral thesis was completed at the University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Electronic Systems and Information Processing.

Supervisor: Professor Sven Lončarić, F.C.A. Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb

The thesis has : 71 pages

Thesis number: _____

About the Supervisor

Sven Lončarić received Diploma of Engineering and Master of Science degrees in electrical engineering from the Faculty of Electrical Engineering and Computing in 1985 and 1989, respectively. He received Ph.D. degree in electrical engineering from University of Cincinnati, USA, in 1994. Since 2011, he has been a tenured full professor in electrical engineering and computer science at FER. He was a project leader on a number of research projects in the area of image processing and computer vision. From 2001-2003, he was an assistant professor at New Jersey Institute of Technology, USA. He founded the Image Processing Laboratory at FER and the Center for Computer Vision at University of Zagreb. Prof. Lončarić has been a co-director of the national Center of Research Excellence in Data Science and Cooperative Systems and the director of the Center for Artificial Intelligence at FER. With his students and collaborators he published more than 250 scientific papers. Prof. Lončarić is a Fellow of the Croatian Academy of Sciences and Arts. According to a Stanford University study published in 2022 he was ranked in the top 2% of the most cited world scientists in the category artificial intelligence – image processing. For his scientific work he received several awards including the National Science Award.

O mentoru

Sven Lončarić diplomirao je i magistrirao u polju elektrotehnike na Fakultetu elektrotehnike i računarstva, 1985. i 1989. godine. Doktorirao je u polju elektrotehnike na Sveučilištu u Cincinnatiju, SAD, 1994. godine. U zvanje redoviti profesor u trajnom zvanju u polju elektrotehnike i polju računarstva na FER-u izabran je 2011. godine. Bio je suradnik ili voditelj na brojnim istraživačkim i razvojnim projektima u području razvoja metoda za obradu slika i računalnog vida. Od 2001. do 2003. bio je Assistant Professor na Sveučilištu New Jersey Institute of Technology, SAD. Voditelj je istraživačkog laboratorija za obradu slike na FER-u. Osnivač je i voditelj Centra izvrsnosti za računalni vid na Sveučilištu u Zagrebu. Suvoditelj je nacionalnog Znanstvenog centra izvrsnosti za znanost o podatcima i kooperativne sustave i voditelj Centra za umjetnu inteligenciju FER-a. Sa svojim studentima i suradnicima publicirao je više od 250 znanstvenih i stručnih radova. Prof. Lončarić redoviti je član Hrvatske akademije znanosti i umjetnosti. Prema studiji Sveučilišta Stanford objavljenoj 2022. godine rangiran je u 2% najutjecajnijih svjetskih znanstvenika u kategoriji umjetna inteligencija i obrada slike. Za svoj znanstveni i stručni rad dobio je više nagrada uključujući Državnu nagradu za znanost.

Acknowledgments

The author thanks his supervisor Sven Lončarić, Ph.D., for the guidance and support throughout the author's doctoral studies, and all the colleagues and friends for every piece of advice and for every minute of their time spent helping him survive all of the Ph.D. challenges. The author also wants to thank his family, his mother Snježana, his father Saša, his sister Vedrana and his grandmother "baka" Marica, for unconditional love and support throughout this long thesis making marathon which brought upon a rollercoaster of emotions. Finally, the author wants to express his deepest gratitude to his life companion, his wife Sunčana, who was his biggest cheerleader through the whole process – the one who stayed positive even when the author had completely given up all hope of completing the Ph.D. thesis. She was the one who said never to give up, and she is one of the greatest reasons why this journey has been completed.

Abstract

This thesis investigates the automatic brightness control algorithm with approximate face region detection in the near-infrared spectrum that corrects the brightness of the face in the scene and prepares it for further facial analysis. This type of system is implemented on embedded systems, so it needs to be efficient, robust, and able to run in hard real-time. The proposed method combines the efficiency of the controller and the speed and precision of the foreground/background segmentation. For automatic brightness control, a split-range proportional-integral-derivative controller is used to reduce the complexity of the non-linear and multiple-input-single-output image acquisition system. Additionally, for approximate face region detection, spatio-temporal sampled skin detection is implemented to put the focus only on the face in the scene, which leads to reducing computational costs and preserving image acquisition system actuators. The thesis begins with an introduction to the face automatic brightness control problem and the motivation for tackling it, which is followed by the theoretical foundations needed for understanding the entire process, and a systematic overview of previous work related to the topic. Next, novel algorithms for approximate face region detection and automatic brightness control with face region detection are described and evaluated on a relevant public dataset and a real-time embedded system. The results demonstrate low execution time, with sufficient accuracy compared to the state-of-the-art methods suitable for real-time embedded systems. The final chapter provides a conclusion of the thesis.

Keywords: near-infrared images, automatic brightness control, face region detection, split range feedback control, skin detection, embedded systems, real-time image processing

Prošireni sažetak

Automatsko upravljanje svjetlinom u bliskom infracrvenom području s aproksimativnom detekcijom regije lica

Doktorski rad podijeljen je u šest poglavlja. U prvom poglavlju predstavljen je uvod u temu automatskog upravljanja svjetlinom lica i detekcije lica te motivacija za rješavanje navedenih problema uz opis glavne primjene u industriji i najvažnijih doprinosa doktorskog rada. U drugom je poglavlju objašnjena teorijska podloga za bolje i lakše razumijevanje doktorskog rada, dok je u trećem poglavlju napravljen pregled relevantne literature u područjima automatskog upravljanja svjetlinom i detekcije regije lica. Četvrto poglavlje opisuje novu metodu za aproksimativnu detekciju lica, dok je u petom poglavlju opisan cjelokupni algoritam za automatsko upravljanje svjetlinom uz detekciju regije lica. U posljednjem, šestom poglavlju obznanjeni su glavni zaključci doktorskog rada.

Prvo poglavlje - Uvod

Analiza lica jedno je od najaktivnijih područja u računalnom vidu. Analiza lica najčešće uključuje detekciju i praćenje lica, prepoznavanje lica te različite procjene poput starosti, spola i emocija. Može se naći u pametnim uređajima, ali i sigurnosnim i nadzornim sustavima te robotici i automobilskoj industriji. Takve slike lica potrebno je dohvaćati po danu i po noći. Slike u vidljivom spektru nisu optimalne u uvjetima smanjene svjetlosti, te se umjesto njih koriste slike u bliskom infracrvenom području.

Kako bi se dobili što bolji rezultati analize, ulazna slika mora biti visoke kvalitete. Unatoč tome što slike u bliskom infracrvenom području imaju svoje nedostatke, velika je prednost što je ljudska koža visoke reflektivnosti u nižem dijelu bliskog infracrvenog spektra. Obrada slika u bliskom infracrvenom području temeljito je istražena, no, ako je izvorna slika niske kvalitete zbog šuma ili zasićenja uzrokovanih vremenskim uvjetima ili različitim postavima scene, često je vrlo teško i čak nemoguće izvući bitne informacije iz slike. Potrebno je utjecati na samu scenu prije dohvata slike, a za to se koristi automatsko upravljanje svjetlinom. Glavni zadatak automatskog upravljanja svjetlinom jest upravljanje svjetlinom scene parametrima kamere i/ili vanjskim izvorom osvjetljenja što rezultira boljim rezultatima analize lica ako je svjetlost postavljena na ispravan način. Upravljanje svjetlinom je predobradba scene, stoga treba biti učinkovito i precizno u stvarnom vremenu kako bi se uspješno implementiralo na ugradbenom računalnom sustavu.

Automatsko upravljanje svjetlinom se koristi u različitim područjima, od medicine, preko agronomije pa sve do mobilne robotike. Predloženi algoritam za upravljanje svjetlinom radi

zadovoljavajuće, no smanjivanje veličine scene na regiju od interesa u kojoj se nalazi lice ga dodatno unaprjeđuje. Analiza ukupne scene ili statične regije usporava cijeli proces te ima značajan utjecaj na odnos prednjeg plana i pozadine u sceni, u kojem pozadina bitno utječe na svjetlinu scene ili regije te ga je potrebno reducirati. Najveći izazov u cijelom procesu je pomicanje lica te nestajanje istog iz scene, što je dodatna tema koju ovaj doktorski rad obuhvaća.

U ovom doktorskom radu proučava se spoj automatskog upravljanja svjetlinom i aproksimativne detekcije lica optimiranog za rad u stvarnom vremenu na ugradbenim računalnim sustavima. Predložena metoda sastoji se od PID regulatora razdvojenog područja za upravljanje svjetlinom scene te prostorno-vremenske metode zasnovanoj na uzorkovanoj detekciji kože. Predložene metode su učinkovito, robusno i pouzdano rješenje za prilagodbu svjetline scene za ugradbene računalne sustave u stvarnom vremenu. Glavni dijelovi znanstvenog doprinosa su sljedeći:

- Prostorno-vremenska metoda za aproksimativnu detekciju regije lica u bliskom infracrvenom području u stvarnom vremenu zasnovana na uzorkovanoj segmentaciji kože,
- Metoda za automatsko upravljanje svjetlinom zasnovanog na upravljanju razdvojenog područja s aproksimativnom detekcijom regije lica prilagođene ugradbenim računalnim sustavima.

Drugo poglavlje - Teorijska podloga

Automatsko upravljanje svjetlinom pripada području automatike. Glavni cilj gotovo svih automatskih sustava upravljanja je izvođenje bez poteškoća i intervencija operatera, a vrlo bitna stavka za uspješan razvoj sustava upravljanja je poznavanje upravljanog procesa. U ovom doktorskom radu glavni proces je sustav dohvaćanja slike s parametrima kamere i vanjskim infracrvenim osvjetljenjem kao ulazima te promjeni svjetline kao izlazima.

Kao što je već navedeno, u ovom radu koriste se slike u bliskom infracrvenom području umjesto slika u vidljivom području. Valna duljina bliskog infracrvenog područja nalazi se između 800 i 2500 nanometara, odmah iznad vidljivog područja te je sastavni dio sunčevog zračenja. Osim visoke reflektivnosti ljudske kože, dodatni razlog zašto se bliska infracrvena svjetlost koristi u ovom sustavu jest da ta svjetlost nije vidljiva ljudskom oku te ne dolazi do odvraćanja pažnje osobe čije se slike dohvaćaju, dok osjetila, posebno ona osjetljiva na blisku infracrvenu svjetlost, dobro prepoznaju navedenu svjetlost. Posebno zanimljiva je svjetlost valne duljine od 940 nanometara, koja je dovoljno udaljena od vidljivog spektra da ne bude uočljiva ljudskim okom, no vrlo blizu području gdje voda apsorbira sunčevu svjetlost u atmosferi, što rezultira znatnim umanjenjem energije oko navedene valne duljine.

Jakost infracrvenog svjetla najčešće se upravlja pulsno-širinskom modulacijom. Prilikom dizajniranja, treba uzeti utjecaj topline infracrvenog zračenja na oči i tkivo, pogotovo iz razloga što promjena jakosti infracrvenog zračenja ne izaziva promjenu veličine zjenice kao što to radi

vidljiva svjetlost te potencijalno velika količina infracrvene svjetlosti može ući u oko i djelovati štetno na njega.

Osim osvjetljenja, sustav dohvaćanja slike sastoji se od kamere, filtra te računalne jedinice za obradu signala i komunikaciju. Najvažniji dijelovi kamere su senzori i leće. Senzori mogu imate različitu rezoluciju, osjetljivost i brzinu osvježavanja. S druge strane, najvažnija svojstva leće su vidni kut i žarišna duljina. Filtar služi za propuštanje ili odsijecanje određenog dijela spektra. U ovom sustavu potreban pojasnopropusni filtar za blisku infracrvenu svjetlost.

Postoje nekoliko parametara kamere kojima se može upravljati. Najčešće se upravlja pojačanjem kamere i vremenom ekspozicije, dok je otvor blende najčešće nepromjenjiv za kamere široke potrošnje. Vrijeme ekspozicije je količina vremena otvorenosti blende te izloženosti osjetila svjetlosti. Veće vrijeme ekspozicije rezultira svjetlijom slikom, no za posljedicu može imati zamućenje slike uzrokovane pokretom. S druge strane, pojačanje, koje može biti analogno i digitalno, mehanizam je koji mijenja svjetlinu cijele scene, uključujući i šum te treba biti posebno oprezan prilikom korištenja istog. Izlaz iz sustava je svjetlina scene, koja je najčešće izražena srednjom svjetlinom svih piksela u sceni ili određenoj regiji.

Nakon što su navedeni svi glavni dijelovi procesa, potrebno je nešto reći i o samim sustavima upravljanja. Glavni cilj sustava upravljanja je reguliranje procesa pomoću upravljačkih petlji. Dizajn sustava upravljanja i samog regulatora obuhvaćen je teorijom upravljanja kako bi sustav došao u željeno stanje uz smanjenje vremena kašnjenja, prebačaja, statičke pogreške te povećanja stabilnosti sustava. Postoji nekoliko podjela automatskih sustava. Po broju ulaza, postoje sustavi s jednim ulazom i izlazom (engl. *SISO*), sustavi s jednim ulazom i više izlaza (engl. *SIMO*), sustavi s više ulaza i jednim izlazom (engl. *MISO*) i sustavi s više ulaza i više izlaza (engl. *MIMO*). Po tipu signala, automatski sustavi mogu biti kontinuirani i diskretni, u kojima je barem jedna veličina diskretna. Također, postoje sekvencijalni upravljački sustavi kao poseban slučaj diskretnih sustava koji nije upravljan vremenom, već slijedom različitih događaja, te sustavi upravljanja s neizrazitom logikom kao posebnom stavkom sustava upravljanja. Najvažnija podjela sustava upravljanja odnosi se na povratnu vezu, stoga postoje sustavi bez povratne veze (upravljanje u otvorenoj petlji) i sustavi s povratnom vezom (upravljanje u zatvorenoj petlji). Sustavi upravljanja u otvorenoj petlji nemaju informaciju o stvarnom stanju izlaza iz sustava, dok sustavi upravljanja u zatvorenoj petlji povezuju izlaz i ulaz sustava. Najvažniji signal sustava upravljanja u zatvorenoj petlji je signal pogreške: razlika između željene vrijednosti i trenutne vrijednosti izlaza dobivene s osjetila. Signal pogreške je i ulazni signal u regulator. Izlaz regulatora, pod nazivom i upravljački signal, spaja se na izvršne članove te tako mijenja stanje procesa. Cijeli se proces ponavlja dok sustav ne dođe do željenog stanja. Sustavi upravljanja u zatvorenoj petlji su poprilično skuplji i složeniji od sustava upravljanja u otvorenoj petlji, no puno precizniji te čine neizostavan dio svakog složenijeg sustava upravljanja gdje je potrebno osigurati da nema regulacijske pogreške.

Povratna veza koristi se za smanjivanje šuma i poremećaja u sustavu te za stabilizaciju nestabilnih i rubno stabilnih sustava. Kad se opisuje povratna veza, najčešće se misli na negativnu povratnu vezu, gdje se mjerena vrijednost oduzima od referentne vrijednosti. Sustavi s povratnom vezom podložni su kašnjenju i mrtvom vremenu, na što treba obratiti pažnju prilikom dizajna regulatora.

Regulator može imati nekoliko tipova promjene izlaza: binarni izlaz (uključeno/isključeno), što se koristi u vrlo jednostavnim sustavima, promjena s fiksnim korakom, što ne mora nužno rezultirati konvergencijom izlazne vrijednosti referentnoj zbog diskretne vrijednosti upravljačkog signala, te promjena s varijabilnim korakom, što rezultira konvergencijom ako je regulator dobro dizajniran. Najučestaliji regulator s varijabilnim korakom jest proporcionalno-integracijsko-derivacijski (PID) regulator. Proporcionalni član PID regulatora izravno je proporcionalan signalu pogreške. Integralni član je zbroj svih prethodnih signala pogreške u sustavu, uključujući i pogrešku u trenutnom koraku, dok je derivacijski član razlika između trenutne i prethodne pogreške. Svaki član PID regulatora ima vlastito pojačanje kojeg je moguće podesiti. Budući da se PID regulator u sadašnje doba najviše izvodi na računalu, potrebno ga je diskretizirati, u čemu veliku ulogu u ponašanju sustava ima vrijeme uzorkovanja, koje je najčešće uračunato u pojačanja članova PID regulatora.

Glavno osjetilo sustava u doktorskom radu je kamera, dok je glavni izvor informacija o svjetlini scene dohvaćena slika u kojoj se nalazi lice i pozadina. Potrebno je pronaći način kako efikasno odvojiti pozadinu od prednjeg plana u kojem se nalazi lice kako bi se izračunala samo svjetlina lica. Zbog visoke reflektivnosti kože u bliskom infracrvenom području te inverznoj kvadratnoj ovisnosti intenziteta svjetla i udaljenosti, lice u sceni je značajno svjetlije od pozadine.

Binarizacija slike je postupak podjele slike u dvije klase, prednjeg plana i pozadine. Za uspješno odvajanje prednjeg plana od pozadine, potrebno je izračunati prag koji će sve piksele više i jednake vrijednosti svjetline prepoznati kao prednji plan, a sve ostale kao pozadinu. Za izračun dinamičnog praga u svakoj slici koristi se Otsuova metoda koja odabire prag koji stvara maksimalnu varijancu između dviju klasa. Iako ta metoda nije nova, još se uvijek koristi često u raznim primjenama binarizacije.

Treće poglavlje - Pregled literature

U ovom poglavlju, predstavljeni su relevantni radovi u području automatskog upravljanja svjetlinom te detekcije lica u bliskom infracrvenom području. Sami pregled literature sastoji se od tri dijela: dva prethodno navedena područja te trećeg dijela koji predstavlja zajedničko djelovanje oba područja.

Automatsko upravljanje svjetlinom može se podijeliti na tri različita dijela: pristup zasnovan na modelu, strojno učenje i teorija upravljanja. Pristup zasnovan na modelu koristi određen

tip modela za predviđanje promjene svjetlosti, poput konveksnog/konkavnog modela ili procijenjene funkcije odziva kamere. Iako su rezultati upravljanja svjetlinom visoke kvalitete, potrebno je ispravno procijenjen model sustava i poremećaja, što može biti veliki izazov za nelinearne sustave i nepredvidiv poremećaj. Strojno učenje pokušava izgraditi i naučiti sustav kako reagirati na promjenu svjetline te se zasniva na prijašnjim situacijama koje je sustav vidio kroz skup za treniranje. Neke od metoda strojnog učenja, koje su se koristile za upravljanje svjetlinom, jesu metoda potpornih vektora, Bayesova optimizacija, konvolucijska neuralna mreža i Markovljev proces odlučivanja. Iako su rezultati upravljanja svjetlinom ponovno visoke kvalitete, potreban je velik skup podataka za učenje s označenim videozapisima te takav pristup nije optimiran za ugradbene računalne sustave u stvarnom vremenu. Teorija upravljanja još je uvijek najučestaliji pristup kod upravljanja svjetlinom. Jedna podgrupa teorije upravljanja su regulatori s neizrazitom logikom, koji, iako pogodni za nelinearne sustave, moraju imati vrlo precizan skup pravila, što nije zadovoljavajuće za sustave s nepoznatim modelom. Sljedeća podgrupa jesu sustavi upravljanja s povratnom vezom, temeljito opisanih u prethodnom poglavlju. Postoji nekoliko načina kako doseći referentnu vrijednost. Prvi način su pregledne tablice *engl. look-up tables, LUT*), u čijim se stupcima i redcima nalaze iznosi pojačanja i/ili vremena ekspozicije i svjetlina scene koju isti proizvode. Iako se radi o vrlo jednostavnom izračunu, takav pristup nije dovoljno robustan za različite poremećaje. Regulatori u zatvorenoj petlji s fiksnim korakom mijenjaju parametre kamere i/ili jakost osvjetljenja diskretnim koracima te, iako robusniji od preglednih tablica, nisu dovoljno robusni, ne moraju konvergirati, mogu uzrokovati oscilacije u sustavu te čak nestabilnost, slično kao i pregledne tablice. Regulatori u zatvorenoj petlji s promjenjivim korakom, predvođeni PID regulatorom te izvedenicom istog - PI regulatorom, najčešće se koristi kod sustava upravljanja kao brzo, učinkovito i precizno rješenje no s puno parametara za ugađanje.

Drugi dio pregleda literature odnosi se na detekciju lica u bliskom infracrvenom području, koju je moguće podijeliti na dva velika dijela: pristupi zasnovani na slici te zasnovani temeljeni na značajkama. Pristupi zasnovani na značajkama mogu se podijeliti na sljedeće podgrupe: analizu niske razine, analizu značajki i aktivne modele oblika lica (engl. *Active Shape Model*). Analiza niske razine uključuje analizu rubova, srednje vrijednosti i detekciju kože, najčešće uparenoj s Otsuovom metodom te vertikalnom i horizontalnom projekcijom. Iako navedene metode možda nisu najrobusnije i najpreciznije, mogu poslužiti za aproksimativnu detekciju regije lica. Analiza značajki zasniva se na traženju složenijih značajki na slici, poput očiju, nosa i usta. Postoje nekolicina metoda u ovoj podgrupi, poput Gaborovih značajki te model Gaussovih mješavina, no najvažnija metoda, koja je svojevremeno bila prekretnica u detekciji lica te koja se i danas vrlo često koristi za detekciju lica, jest Viola-Jones algoritam. Iako su navedene metode precizne što se tiče detekcije lica, vrlo su složene i potrebno im značajna količina računalnih resursa, što ih ne čini pogodnima za rad u stvarnom vremenu na ugradbenim

računalnim sustavima. Aktivni modeli oblika lica su precizniji od metoda analize značajke, no još uvijek nedovoljno brzi za izvršavanje u stvarnom vremenu. Pristupi zasnovani na slici pretražuju podregije slike kako bi svaki piksel identificirali kao dio koji pripada ili ne pripada licu. Najvažnije metode u ovom pristupu su neuronske mreže, od kojih se ističu dvije: višerazinska konvolucijska neuronska mreža s dvostrukim nadzorom (MDSCNN) i višezadaćna kaskadna konvolucijska neuronska mreža (MTCNN). Iako je preciznost navedenih metoda izrazito visoka, potrebno je vrlo moćno sklopovlje kako bi bio omogućen rad u stvarnom vremenu te veliki označeni skup podataka za učenje.

Posljednji dio pregleda literature odnosi se na metode u kojima je automatsko upravljanje svjetlinom upareno s detekcijom lica u bliskom infracrvenom području. Postoje tri rada koji uključuju obje stavke. U prvom radu, Viola-Jones algoritam za detekciju lica korišten je uz gamma korekciju svjetline slike. Kako je spomenuto prije, Viola-Jones algoritam nije prikladan za rad u stvarnom vremenu na ugradbenim računalnim sustavima gdje je potrebno ostvariti 30 brzinu osvježavanja od 30 sličica u sekundi, dok gamma korekcija nije pravi algoritam za promjenu svjetlinu scene, već je to metoda za predobradu slike. U drugom radu, za detekciju lica koristi se integralna projekcija dobivena segmentacijom slike te uspoređivanjem uzoraka, dok se za promjenu svjetline koristi dijagram odnosa osvjetljenja i napona. Iako je metoda za detekciju lica precizna, nije dovoljno brza za izvođenje na ugradbenim računalnim sustavima, dok je sustav upravljanja zasnovan na dijagramu, koji je zapravo pregledna tablica, nedovoljno robusno rješenje. Posljednji rad opisuje unaprijed izračunatu masku lica za detekciju lica, dok se za promjenu svjetline koristi sustav upravljanja s povratnom vezom s fiksnim korakom promjene. Iako unaprijed izračunata maska znatno ubrzava sustav i daje dobre rezultate ako se preklopi s licem, još uvijek je podložna greškama, dok sustav upravljanja s povratnom vezom s fiksnim korakom također nije dovoljno robusno rješenje za različite poremećaje u svjetlini scene zbog diskretnog koraka promjene.

Iako detekcija lica u navedenim radovima daje dobre rezultate, upravljanje svjetlinom je rudimentarno te ne mora rezultirati konvergencijom sustava upravljanja ili, ako dolazi do konvergencije, je vrlo spora. Iz navedenog razloga, potrebno je predstaviti rješenje koje daje zadovoljavajuće rezultate u oba područja.

Četvrto poglavlje - Prostorno-vremenska metoda za aproksimativnu detekciju lica u bliskom infracrvenom području u stvarnom vremenu

Cilj metode predložene u ovom poglavlju je učinkovita detekcija regije lica kako bi se algoritam za automatsko upravljanje svjetlinom mogao izračunati samo u navedenoj regiji. Metoda se sastoji od dva dijela: prostornog i vremenskog dijela. Bitan zahtjev za regiju lica jest da ne dolazi do naglih promjena granica regije, što i nije karakteristično za uobičajene slučajeve gdje

se lice ne pomiče previše između dvije slike, pogotovo u 30 sličica po sekundi.

Glavni zadatak prostornog dijela predložene metode je pronalazak regije lica u trenutnom koraku. Prostorni dio predložene metode započinje s dohvatom slike s osjetila, nakon čega se slika dijeli na stupce i retke. Kako se lice rijetko nalazi u rubnim ćelijama, određeni postotak vanjskih ćelija se ne uzima u obzir prilikom izračuna. U preostalim ćelijama izračunata je srednja vrijednost piksela iz histograma te dolazi do binarizacije ćelije pomoću Otsuove metode. Nakon izračuna binarizacijske maske, slijedi izvođenje jedne od četiri inačice algoritma prostornog dijela detekcije regije lica. U prvoj inačici, traži se najveći pravokutnik sa svim pozitivnim ćelijama u maski. U drugoj inačici traži se najveći pravokutnik s određenim postotkom pozitivnih ćelija u maski. Treća inačica je slična prvoj, no prethodno se negativne ćelije okružene većinom pozitivnim ćelijama zamijene pozitivnom vrijednošću te iste zatim ulaze u izračun svjetline regije lica. U četvrtoj inačici, stupci do detektirane regije ulaze u izračun svjetline regije lica ako postoji određen postotak pozitivnih ćelija u svakom stupcu pojedinačno. Uz navedene inačice, također postoji i opcija proširivanja regije za određen postotak ćelija. Izračunate granice regije lica predaju se vremenskom dijelu algoritma.

Glavni zadatak vremenskog dijela predložene metode je smanjivanje utjecaja naglih promjena granica regije lica uzrokovanih naglom promjenom položaja lica ili nestajanjem lica iz scene. Vremenski dio izvršava se u dva koraka. U prvom koraku, računa se medijan nekoliko prethodnih vrijednosti granica za svaku granicu zasebno. Na taj se način eliminiraju nagle promjene granice uzrokovane poremećajem u svjetlini scene. Nakon što se izračuna medijan za svaku granicu zasebno, slijedi drugi korak: histereza. Histereza uspoređuje trenutnu vrijednost izračunatih granica s prethodnom vrijednošću granica. Ako u novoj iteraciji regija ima tendenciju širenja za određen postotak ćelija, nova vrijednost granice će biti odbačena te će za granicu biti odabrana prethodna vrijednost. Na taj način, pokušava se zadržati konstantna veličina regije u slučaju ako se dijelovi lica djelomično nađu u novim ćelijama. Redoslijed koraka se može zamijeniti, no to može utjecati i na izračun granica. Isto tako, svaki od koraka se može izvesti zasebno. Sa svim mogućim postavkama, postoji ukupno 136 podinačica ove metode te su sve podvrgnute testiranju.

Predložena metoda testirana je na skupu podataka FADID, kako bi se izmjerila točnost detekcije, i ugradbenom računalnom sustavu Zynq UltraScale+ MPSoC ZCU104 Evaluation Kit, kako bi se izmjerila brzina izvođenja algoritama. U oba slučaja, predložena metoda je uspoređena s Viola-Jones algoritmom. Iako detekcija regije lica nije bila toliko precizna kao kod Viola-Jones algoritma, predložena metoda je višestruko brža od Viola-Jones algoritma s brzinom izvođenja manje od 2 milisekunde na ugradbenom računalnom sustavu, što je čini pogodnom za rad u stvarnom vremenu.

Peto poglavlje - Automatsko upravljanje svjetlinom zasnovano na upravljanju razdvojenog područja s detekcijom regije lica

Cilj metode predložene u ovom poglavlju je učinkovito automatsko upravljanje svjetlinom s detekcijom regije lica uz upravljanje nestankom lica iz scene. Sami algoritam za upravljanje svjetlinom lica zasniva se se na upravljanju razdvojenog područja.

Nakon što je eliminiran efekt namatanja integralnog člana, upravljački signal PID regulatora razdvojenog područja propagira se na četiri upravljiva parametra: vanjsko osvjetljenje, vrijeme ekspozicije te digitalno i analogno pojačanje. Budući da najveći utjecaj na svjetlinu scene mora imati vanjsko osvjetljenje, manji utjecaj mora imati vrijeme ekspozicije, dok najmanji utjecaj moraju imati pojačanja, sukladno tome se podešavaju i težinski faktori za svaki upravljivi parametar. Upravljačke vrijednosti parametara su ograničene na granice svog raspona ako izračunom izlaze iz istih. Nakon promjene svjetline scene uzrokovane promjenom upravljačkih signala parametara, prethodno opisana metoda za aproksimativnu detekciju regija lica pokušava pronaći lice u sceni. Ako uspije pronaći regiju lica, srednja vrijednost piksela se izračuna u regiji iz kojeg se, uz referentnu vrijednost svjetline, izračuna vrijednost signala pogreške te ulaza u PID regulator. Postupak se ponavlja dok se ne postigne željena svjetlina scene. Ako se regija lica ne detektira, aktivira se modul za upravljanje nestankom lica. Upravljački signal se postavi na nisku vrijednost da se što manje opterete parametri kamere i vanjsko osvjetljenje, no na dovoljno visoku da se regija lica pronađe ako se lice vrati u scenu. Nakon toga, izračuna se srednja svjetlina scene te se provjerava promjena srednje svjetline slike. Ako se srednja svjetlina slike ne dosegne vrijednost iznad postavljene vrijednosti praga, što ujedno označava povratak lica u scenu, postupak se ponavlja do trenutka dok navedeni uvjet nije zadovoljen. Nakon toga, nastavljan se normalan tijek algoritma.

Predložena metoda testirana je na ugradbenom računalnom sustavu Zynq UltraScale+ MPSoC ZCU104 Evaluation Kit u nekoliko iteracija. Brzina izvođenja cjelokupnog algoritma je manja od 10 milisekundi, što ju čini prikladnom i za sustave s većom brzinom osvježavanja od 30 sličica po sekundi, dok sam sustav konvergira od 10-15 koraka, što je manje od pola sekunde na navedenom sklopovlju. Detekcijom regije lica smanjuje se prisutnost pozadine, što uvelike utječe na upravljačke veličine, a samim time i na naprezanje izvršnih članova: parametara kamere i vanjskog osvjetljenja. Slični rezultati su dobiveni za slučaj kad je lice mirno u sceni ili se pomiče, dok upravljanjem nestankom lica ne dolazi do naprezanja do krajnjih granica parametara kamere i infracrvenog osvjetljenja.

Šesto poglavlje - Zaključci

Posljednje poglavlje donosi osvrt na postignute rezultate testiranja algoritama i doprinosa doktorskog rada. Predloženi sustav automatskog upravljanja svjetlinom s detekcijom regije

lica je učinkovit, robustan te prilagođen za rad u stvarnom vremenu na ugradbenim računalnim sustavima, što ga čini vrlo važnim čimbenikom u pripremi scene lica za daljnju analizu.

Ključne riječi: slike u bliskom infracrvenom području, automatsko upravljanje svjetlinom, detekcija regije lica, upravljanje razdvojenog područja, detekcija kože, ugradbeni računalni sustavi, obrada slike u stvarnom vremenu

Contents

Chapter 1

Introduction

Face analysis is one of the most active fields in computer vision research area. Face analysis often includes face detection and tracking, face recognition, and various estimations such as age, gender, and emotions. Face analysis is present in the majority of smart devices, from smartphones and personal computers through security and surveillance systems to robotics and automotive industry.

Images are captured in different parts of the electromagnetic spectrum. The part of the electromagnetic spectrum mostly used for capturing images is the visible light spectrum (VIS).

VIS images are good when there is high amount of ambient illumination. In real life scenarios, images are captured during both day and night, so, when VIS images are taken at night, their quality is poor despite the external illumination source. Instead of using VIS images, which are not optimal for nighttime, near-infrared (NIR) images are used. The near infrared spectrum is a part of the electromagnetic spectrum with higher wavelength than the VIS spectrum, as seen in Figure 1.1. An example of a NIR face image can be seen in Figure 1.2.

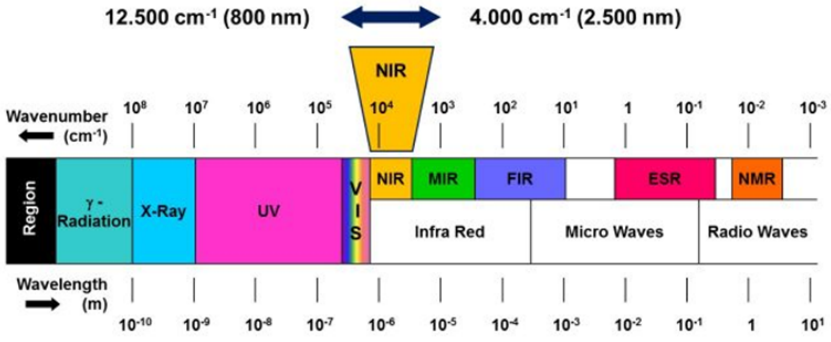

Figure 1.1: Electromagnetic spectrum with VIS and NIR spectrum highlighted [1]

The NIR spectrum is a part of the solar spectrum, together with the VIS, thus, no external illumination source is needed to capture the image during daytime [2]. However, a near-infrared illumination source is needed during nighttime to enhance sensor response.

To get the best possible analysis results, a high quality input image is needed. Although NIR

Figure 1.2: NIR face image example [1]

images have their downsides, such as sparsity, low signal to noise ratio (SNR), low contrast, and fuzzy smoothness of textures [3], a very useful property of lower NIR wavelengths is that the human skin is highly reflective in the lower part of the NIR spectrum [4]. While NIR image enhancement has been thoroughly explored [5], enhancing a poorly captured image can be a very difficult task.Thus, important image information often cannot be recovered, which is a result of the noise and the sensor's inability to capture a scene dynamic range bigger than its own [6], leading to overexposed and underexposed pixels. This is caused by various types of ambient illumination caused by solar radiance at different times of day, weather conditions, or different scene scenarios [7]. A brightness shift of the entire scene is needed, and this is where automatic brightness control is used.

The main focus of automatic brightness control is to properly adjust the scene brightness through the camera's parameters and/or external illumination source. When brightness is adjusted in an optimal way, it improves the face brightness and, consequently, face analysis tasks. Brightness control is a pre-processing task. Thus, it requires resourcefulness while providing good results in real-time, given that it is implemented on embedded systems.

1.1 Motivation

Automatic brightness control is an extensively covered research area used for different purposes: from medical applications such as endoscopy [8], [9], photoplethysmography [10], and arthroscopy [11], through leather industry [12] and agronomy [2], to mobile robotics, including simultaneous localization and mapping (SLAM) [13], visual odometry (VO) applications [14], self-driving cars [15], and robot vision [16], [17], where it is mostly used. The variety of applications indicates the need for automatic brightness control by moving and resizing the scene dynamic range. Changing the brightness greatly reduces the effort needed for image enhancement and restoration, thus, improving the image analysis process.

Adding face region detection improves automatic brightness control algorithms. Using a whole image or a static region for the calculation tends to slow down the analysis process because more resources are needed to execute the required analysis algorithms. Moreover, the region information contains both the face and background information. A significant background impact on the region brightness should be eliminated. A great challenge for automatic brightness control is the face moving in the scene or disappearing from it, which significantly affects the overall brightness in the scene or in the region of interest (ROI). Those are the main reasons why the dynamic ROI with face region included is needed. The approximate face region detection should also be a lightweight process because it needs to be used in embedded systems, in real-time.

Joining two different areas, control systems and information processing, is the research path of this thesis and a potential solution for real-life computer vision tasks, especially for face analysis. While some forms of solutions exist, they tend to lean towards control systems where face region information is missing, barely exists, or is not used at all, or towards the information processing part where the control algorithm is quite rudimentary and often neither robust nor precise. In this thesis, both approaches play an equally important role in the proposed algorithm in order to produce a proper method for successful pre-processing computer vision task, as will be shown later in the thesis. In order to successfully implement the combined method on an embedded system with all of its constraints and limited resources, the process requires lots of adjustments and modifications of both elements, starting from their basic forms. This question has not been researched thoroughly, thus, the findings can be very beneficial for automatic brightness control and, finally, for face analysis.

1.2 About this thesis

This thesis investigates a combination of an automatic brightness control algorithm and approximate face region detection. The proposed method consists of a split-range feedback

PID controller as an automatic brightness control element and a spatio-temporal method based on image statistics, binarization, and calculation of current and previous face region boundaries as an approximate face region detection element.

The presented novel methods form an efficient, robust, and reliable solution for properly adjusting the scene brightness, suitable for real-time embedded systems, as will be experimentally demonstrated. The main contributions of this thesis are as follows:

- Real-time spatio-temporal method for approximate face region detection in the NIR spectrum: approximate face region detection optimized for embedded systems which also includes previous region position;
- Automatic brightness control method based on the split-range feedback control with approximate face region detection optimized for embedded systems: the image acquisition in the system is a multiple-input single-output (MISO) system in which the proposed approximate face detection method is used.

To the best of the author's knowledge, the proposed methods and implementations have not been presented in previous research.

1.3 Organization of the thesis

The thesis is further organized as follows. Chapter 2 introduces the optics and electronics background, as well as control systems concepts used throughout the thesis. In Chapter 3, related work is introduced for both automatic brightness control systems and face region detection. The described algorithms are systematically analyzed to provide the reasoning behind the proposed methods in Chapters 4 and 5. Novel automatic brightness control methods using split-range feedback controller and spatio-temporal approximate face region detection are described in Chapters 4 and 5, respectively. The methods are experimentally verified and compared with real-time state-of-the-art embedded solutions on a benchmark dataset. Finally, Chapter 6 concludes the thesis.

Chapter 2

Theoretical foundations

Automatic brightness control is a branch of the control system research area. The main task of most control systems is to run the processes smoothly without much manual interference. To create a working control system, an understanding of how the process works is crucial. In this thesis, the image acquisition system is a central process with signals of the camera's parameters and external illumination as inputs and brightness change as an output. In the first part of this chapter, a brief explanation of high-layer optics and image acquisition is given to provide a better understanding of the system, which is followed by an explanation of control systems and, more specifically, feedback control. Finally, the process for segmenting background and foreground in the image used in the approximate face region detection method is described: the Otsu's method for threshold determination [18].

2.1 Optics and image acquisition system

Before diving into the topic of the optics and electronics in the system, we will discuss the NIR spectrum and its advantages over the VIS for the application. As stated in the introduction, the NIR spectrum is located between 800 and 2500 nanometers, just above the VIS. As seen in Figure 2.1, 53% of the solar spectrum consists of infrared spectrum; however, its spectral irradiance is lower compared to the visible spectrum.

Near the 940 nanometers, sunlight is absorbed by water in the atmosphere and the energy is quite lower around that wavelength [20], so it is used as an upper limit for the external NIR illuminator. While the human eye cannot see NIR light, cameras, especially the ones more sensitive to the NIR light, are able to capture the NIR light, as illustrated in Figure 2.2.

Another important reason why the NIR spectrum is used is the fact that the skin is highly reflective in the range from 800 to 1,100 nanometers, while the reflectance is low above 1,400 nanometers. There are two wavelengths from the proposed range of the NIR spectrum commonly used for the illuminator: 850 and 940 nanometers. Although the NIR light is invisible

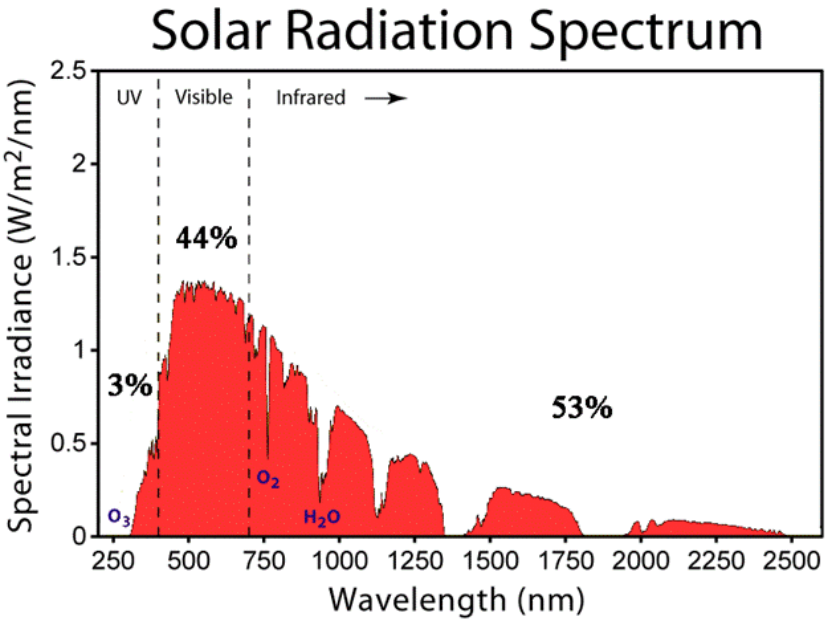

Figure 2.1: Solar spectrum as a function of wavelength [19]

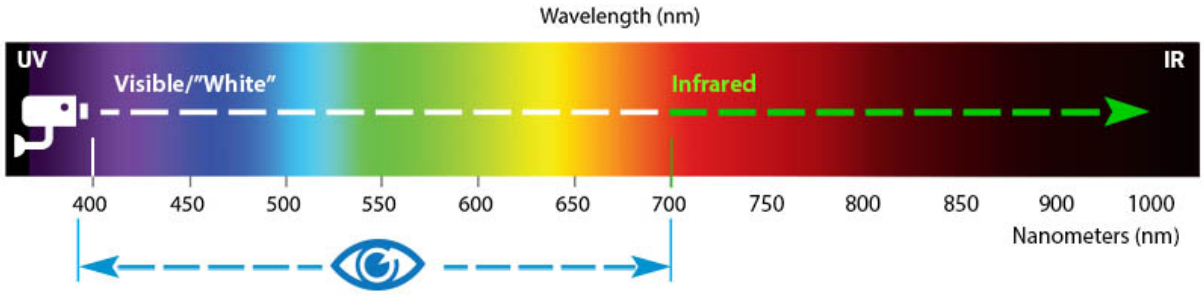

Figure 2.2: Visible wavelengths for a human eye and a camera [21]

to the eye, it does produce slight light-emitting-diode (LED) glow at the LED light source for the 850 nanometers source, while it is almost completely invisible for the 940 nanometer [21]. Thus, there are no distractions to the subject [22], whereas using an active 940 nanometers illumination source ensures that a great part of illumination changes due to ambient illumination being filtered away. NIR illumination sources are often controlled with pulse width modulation (PWM). NIR light, as every other heat source, can cause safety hazards for eyes and skin, mostly due to tissue heating caused by light absorption. The eyes are especially fragile as there is no pupillary light reflex, and a large amount of light comes through the pupil, as shown in Figure 2.3. When designing the NIR illuminators, the designer needs to be very careful to create a safe product in accordance with the recommendations [23].

Besides the external illumination sources, the image acquisition system consists of the ca-

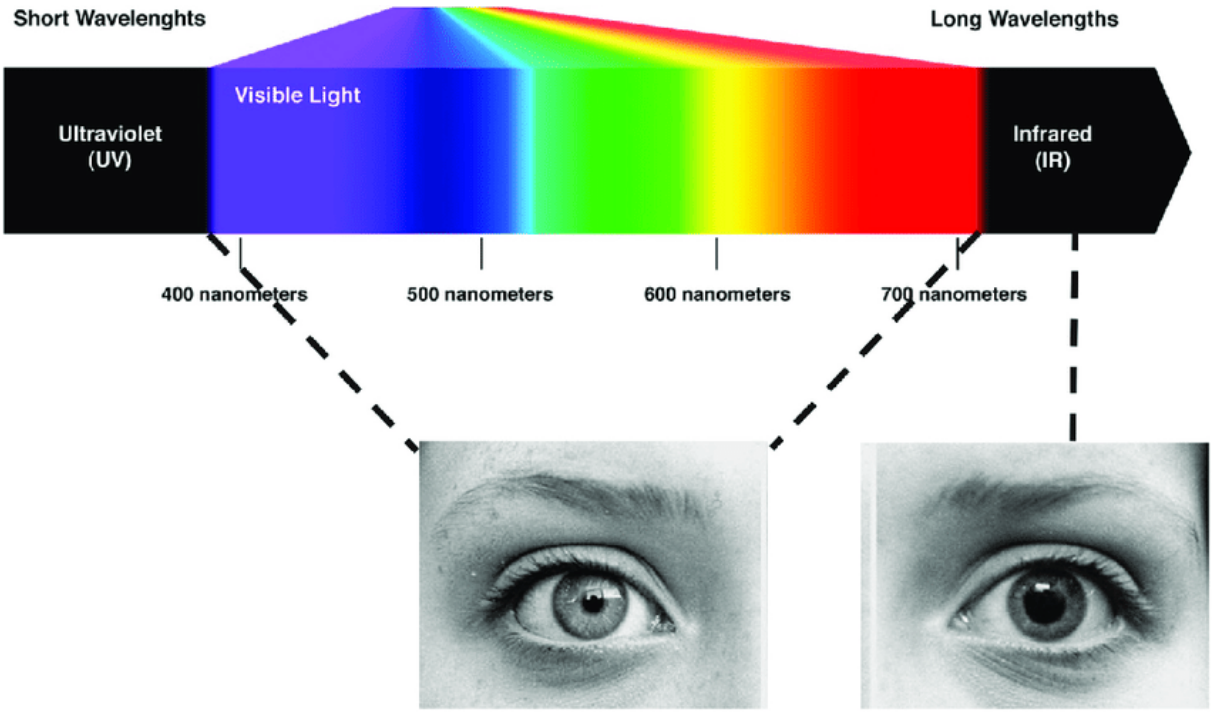

Figure 2.3: Effect of visible and infrared light on the size of the pupil [24]

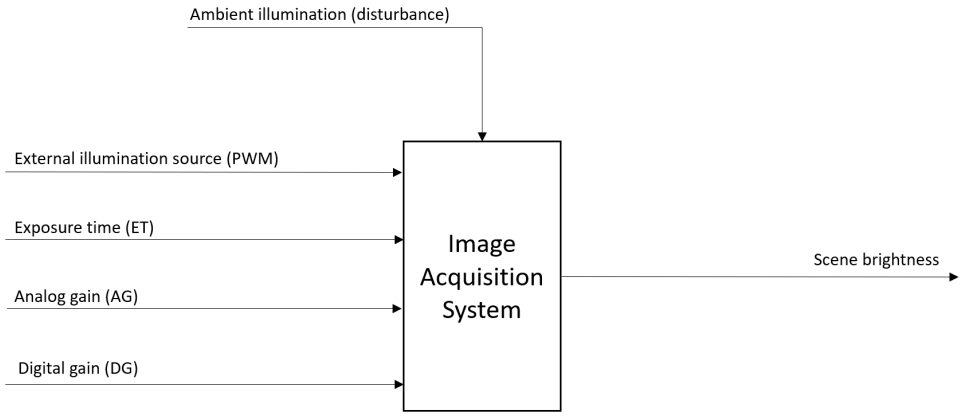

Figure 2.4: Block schematics of the image acquisition system

mera, the filter and the computer unit for signal processing and communication. The most important parts of the camera module are sensors and lenses. Sensors have different resolutions, sensitivity, and frame rate. The most important properties of the lens are the angle of view and focal length. The role of the filter is to pass or to cut off a certain part of the spectrum. The bandpass NIR filter on the camera allows only a narrow part of the NIR spectrum to pass and rejects further changes in ambient illumination [25].

Every system has its inputs and outputs. In the image acquisition system, the inputs are the camera's parameters and external illumination source, while the output is scene brightness, as seen in Figure 2.4. Besides the inputs and outputs, image acquisition systems are affected by disturbances in the form of ambient illumination which need to be considered when designing the controller.

According to [10], brightness can be controlled by changing aperture size, exposure time (automatic exposure control, AEC), and gain (automatic gain control, AGC). As in most applications, the aperture size of consumer cameras is fixed, while gain and exposure are controllable. Exposure time is the amount of time the image sensor is exposed to the light by the open shutter of the camera [10]. More exposure time will result in a brighter picture, but it can produce motion blur. Gain, which can be digital and/or analog, is an amplification mechanism that can produce a brighter image, but with the cost of noise amplification as well. The output of the system is scene brightness, as stated before. The captured image of a scene consists of pixels, whose number depends on the sensor's resolution. It would be very demanding to control every pixel independently, so the common measure is used as a representation of the pixel set. Median is calculated in [26], while variance is used in [27]. Skewness of the histogram is calculated as a measure in [28], and mean sample value (MSV) is calculated in [29]. While the proposed metrics calculation is fast, it might be a bit inferior to other metrics in terms of image quality assessment. Regarding entropy [30], it is a better indicator of image quality, but the drawback is that it presents an optimization problem in which a global maximum needs to be found to get the best results. This can be very demanding in terms of complexity and resource usage. Most methods that use entropy as a metric on an embedded system cannot be implemented in real-time. Gradient information can be used standalone as in [15], but it is often combined with some histogram metric [31] or entropy [17]. Gradient calculation can also entail a high demand of computing resources due to the calculation of the two-dimensional convolution in the space domain or its transformation back and forth into the frequency domain. Mean pixel value is the most used metric in which the average brightness of the whole image [32] or the specific region of interest (ROI) is calculated [9]. The mean pixel value in the region of interest is a good representation of histogram statistics for scene brightness. The image acquisition system is a non-linear MISO system with mutually dependent inputs, which makes it challenging to design an optimal controller. The next subsection includes more details about control systems.

2.2 Control systems

A control system is a set of mechanical or electronic devices that regulates the behavior of other devices or systems using control loops. The controller is designed by the control theory rules and recommendations. The controller drives the system to a desired state, while minimizing delay, overshoot, or steady-state error and ensuring a level of control stability, often with the aim of achieving a degree of optimality. There are several classification types of control systems, depending on the properties. Regarding the number of inputs and outputs, there are single-input single-output (SISO), single-input multiple-output (SIMO), MISO, and multiple-input multiple-output (MIMO) systems. Regarding the signal type, control systems can be

continuous or discrete. In a continuous control system, all the signals are continuous in time, whereas in a discrete control system, there are one or more discrete values present. Sequential control system is a special case of discrete control systems which are not time-driven, but comprise a series of different events taking place one after another, with the end of one event being the signal for the beginning of the next event. Fuzzy logic control is an approach to control systems that analyzes analog input values in terms of logical variables that take continuous values between 0 and 1, in contrast to classical or digital logic, and it does not belong to any specific category of the classic control approaches. The most important classification of controllers with regard to the feedback path are systems without feedback (open loop) and with feedback (closed loop). Feedback means connecting the output with the input by measuring it and including it in the controller output calculation. In open loop control systems, this connection does not exist- Thus, the control action is independent of the desired output. In closed loop control systems, the output is connected with the input, so the control action is dependent on the desired output. The controller sends signals to the actuators, so the controller output is also called actuating signal. In open-loop control systems, a direct input is given to the controller, which then calculates signals for the actuators to change the process output. In closed-loop control systems, the error term, the difference between the reference input and the measured output, is the input of the controller. Similar to the open-loop system, the actuator values are modified, and the process output is affected. The output is measured with sensors and the output signal is brought to the beginning where it is again compared with the input value, closing the infinite loop. The process in this system is often called plant. Both closed and open-loop systems have their advantages and disadvantages. While open-loop control systems are cheaper and easier to design, the control action is independent of the desired output, so it tends to be inaccurate. On the other hand, closed-loop control systems are more difficult and more expensive to design, because they entail additional equipment, such as sensors, as a part of feedback elements. However, the presence of an error signal makes the control action dependent on the desired output, and it is much more accurate and better for automatic systems, which is the main reason why it is used in most automatic systems. A block schematic of an open-loop control system is seen in Figure 2.5, while a schematic of a closed-loop control system is shown in Figure 2.6.

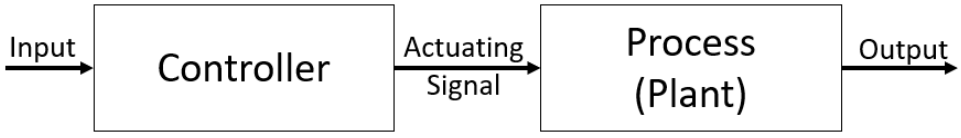

Figure 2.5: Block schematic of an open-loop control system

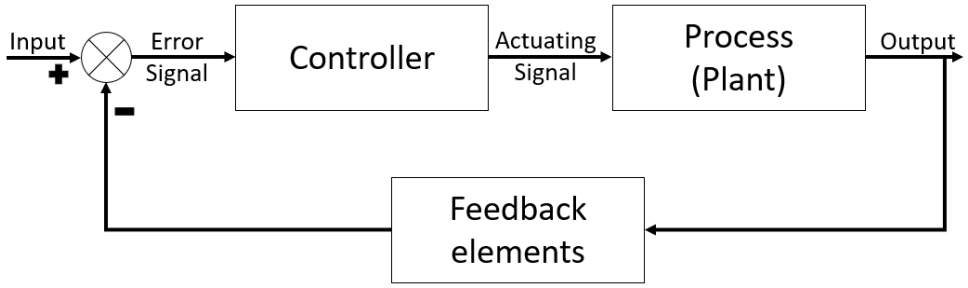

Figure 2.6: Block schematic of a closed-loop control system

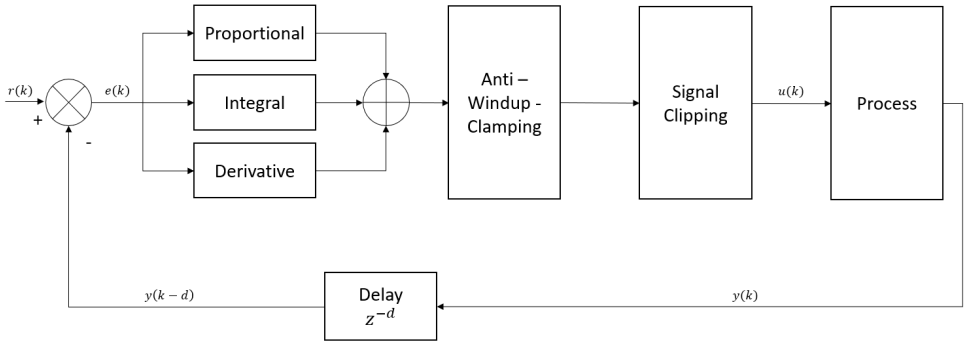

Figure 2.7: Block schematics of the PID control system with clamping and signal clipping

2.3 Closed-loop control systems

A closed-loop control system, or feedback control system, is a type of control system that measures the output of the system and compares it to the reference value in order to properly adjust the system behavior, and it is well-suited for automatic control. Feedback control is mostly used to reduce the noise and disturbance in the system, and to stabilize unstable systems. The feedback control described before is negative, which means that the error term is calculated by subtracting the output signal from the reference value. If the output signal and the reference value are added, this is a positive feedback loop, in which the event that had caused the error is increased. This type of system is unstable; it hits the limits very quickly, and it is not used in control systems. The real system is prone to dead time and delays, which also affects the system response, so it needs to be taken into account when designing a controller.

The controller can have different outputs. The on-off controller switches between two values: on and off. This is mostly used in temperature control in domestic heating systems. The change of the output signal can occur in a fixed or variable step. In a fixed step output change, only several output values can be achieved when the disturbance and the noise are not present, which can result in a non-converging system. In a variable step output change, the output of the controller can be any value in a certain range depending on the hardware and software limitations. The most common feedback control with a variable step controller output is the standard Proportional-Integral-Derivative controller, or the PID controller [9].

The main signal of the PID controller for adjusting brightness is error term e, which is

the difference between reference value r and the current measured mean pixel value y. There are three components to the PID controller: proportional (P), integral (I), and derivative (D) term. The proportional part, as its name states, is directly proportional to the error term. The integral part is the sum of all the previous errors that occur during the control process, while the derivative part is the difference between the current and previous error terms. A block schematic can be seen in Figure 2.7. An ideal PID controller discretized with the backward Euler method in parallel form is shown in the Equation 2.1, where u is the control variable in k-th step, K_p, K_i, and K_d are proportional, integral, and derivative gains, respectively, and T_s is sampling time. The sampling time is often embedded in the mentioned gains. Each part of the controller and their combinations can be used as a controller depending on the requirements. Proportional control adjusts the system according to error value, but there is an error in a steady state, and, with it, a difference in the output. Integral term eliminates the steady-state error, but it also has some drawbacks, such as slow response time, overcorrection, and instability for an oscillatory response. Furthermore, the integral part accumulates all the errors regardless of the control variable. If the control variable u hits the upper or lower physical limit, or saturation, the output control variable does not change, but the error is still added to the integral sum. This occurrence is called the integrator windup, and it results in poor performance of the controller, overshoot or undershoot, delays, and instability. Thus, this effect needs to be negated with some of the anti-windup mechanisms. The integral controller is rarely used alone, and the combined Proportional-Integral controller, or PI controller, is often used. The derivative control is a form of feed-forward control in which the error change is used to predict the process behavior. The controller tries to keep the system at a consistent setting by minimizing the change of error. The major benefit of derivative controllers is to resist changes in the system, such as oscillations and rapid response. However, it is highly sensitive to noise and derivative action wears out the equipment faster, so it needs to be set properly. Each part of PID controller has its gain with which the system response is manipulated – the process is called parameter tuning. There are numerous ways to tune the PID controller properly, especially for linear systems, regarding the wanted system response properties such as percentage overshoot, rise time, and settling time, such as [33]. However, the tuning methods are heuristic, and they are used for making rough estimates of the optimal gains, and additional fine tuning is needed to get the adequate system response.

$$u(k) = K_p e(k) + K_i T_s \sum_{i=0}^{k} e(i) + \frac{K_d}{T_s}[e(k) - e(k-1)]. \tag{2.1}$$

As seen in Figure 2.7, there are several steps before the controller output u reaches the process, first one being the anti-windup mechanism. The most used anti-windup mechanisms are back-calculation and integrator clamping. In back-calculation, difference between the unsatu-

rated and saturated controller is calculated and multiplied with the back-calculation gain that unwinds the integral accumulator by subtracting it from or adding it to the integral sum. Integrator clamping, or conditional integration, prevents the integral output from accumulating in the appropriate direction when the controller output is saturated. After the anti-windup, the output of the controller is clipped. If the controller output is out of the range, it is limited to the highest or the lowest value in that range, preventing the actuators from going outside of their range, and from forcing, breaking, and changing the system behavior.

The process in this control system is the image acquisition system whose primary sensor is the camera, and the main source of brightness information is the image of the scene. The whole image or the fixed region of interest contain both the face and the background. The focus of the image should be the face and its brightness. Thus, the background needs to be separated from the foreground in which the face is located.

2.4 Otsu's method for threshold determination

The image usually contains the background and the foreground, and, for face images, the face is a part of the foreground. Due to the inverse square law relation between the light intensity and distance [34], the foreground with the face is illuminated more than the background. Combined with NIR skin reflectance, it is easier to separate the foreground and the background in the image. Dividing the image into two classes, the background and the foreground, is called image binarization. For binarizing a grayscale image, a threshold value is needed. If the values are below the threshold, they are classified as the background, while the values greater than or equal to the threshold are considered as the foreground. It can be selected manually, it can be calculated once, or it can be adaptive in each step. Discriminant analysis is used for adaptive thresholding in [35]. However, one of the most common methods for adaptive threshold calculation is Otsu's method [18]. In Otsu's method, the threshold is calculated from the intensity histogram. This threshold is determined by minimizing intra-class intensity variance. The formula for finding the within-class variance ω_w^2 at any threshold t is given in Equation 2.2, where weights ω_{bg} and ω_{fg} are probabilities of the two classes separated by the threshold t, while the σ_{bg}^2 and σ_{fg}^2 are background and foreground variances, respectively. For the two classes, however, minimizing the intra-class variance is equivalent to maximizing inter-class variance [18], which is a little faster to calculate. The formula is shown in Equation 2.3, where weights are the same, and μ_{bg} and μ_{fg} represent the background and the foreground class mean, respectively.

$$\sigma_w^2(t) = \omega_{bg}(t)\sigma_{bg}^2(t) + \omega_{fg}(t)\sigma_{fg}^2(t) \tag{2.2}$$

$$\sigma_b^2(t) = \omega_{bg}(t)\omega_{fg}(t)[\mu_{bg}(t) - \mu_{fg}(t)]^2 \tag{2.3}$$

The weights are calculated by Equations 2.4 and 2.5, where P_{all} is the total count of pixels in an image, and P_{BG} and P_{FG} are the counts of background and foreground pixels at the threshold t, respectively. The class means are calculated by Equations 2.6 and 2.7, where i is the index of the current pixel value, $p(i)$ is the number of pixels with pixel value i and N is the maximum threshold value.

$$\omega_{bg}(t) = \frac{P_{BG}(t)}{P_{all}} \tag{2.4}$$

$$\omega_{fg}(t) = \frac{P_{FG}(t)}{P_{all}} \tag{2.5}$$

$$\mu_{bg}(t) = \frac{\sum_{i=0}^{t-1} ip(i)}{\sum_{i=0}^{t-1} p(i)} \tag{2.6}$$

$$\mu_{fg}(t) = \frac{\sum_{i=t}^{N-1} ip(i)}{\sum_{i=t}^{N-1} p(i)} \tag{2.7}$$

Figure 2.8 provides an example of image binarization with Otsu's method. From the input 6x6 image with minimum value 0 and maximum value 5, an intensity histogram is made. The inter-class variance $\sigma_b^2(t)$ is calculated with the Equations 2.3–2.7 and it is at its maximum when the threshold t equals 3. Finally, the binarized image is shown.

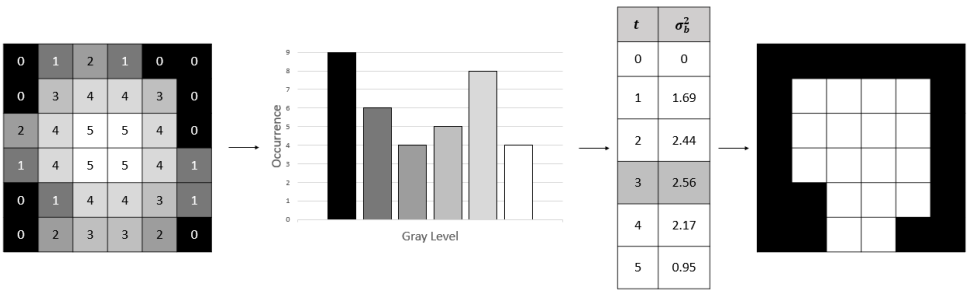

Figure 2.8: Image binarization example with threshold determined by Otsu's method

Although Otsu's method is not novel, it still has different applications as a resourceful and powerful solution for image binarization. It is suitable for real-time embedded system implementation and, accordingly, for the image automatic brightness control.

Chapter 3

Related work

In this chapter, the relevant prior work to the dissertation topic is presented. Automatic brightness control with face region detection is a part of two different research areas: control systems with regard to automatic brightness control, and information and image processing and analysis with regard to face region detection. The chapter is divided into three parts. The first part presents the work related to automatic brightness control, including control system strategies for controlling brightness. In the next part, research of face and face region detection in the NIR spectrum is described, focusing on the methods optimized for embedded systems. In the last part of this chapter, the related work on the combined effect of the two areas is outlined.

3.1 Automatic brightness control

Automatic brightness control is an extensively covered research area. There are three main ways to differentiate between various automatic brightness control approaches. The first approach is a model-based method [36] in which automatic brightness control is based on a convex/concave model of luminance-related function and the control parameter. The results, with the correction of overexposed and underexposed scenes by using a convex/concave model, can be seen in Figure 3.1. In [37] and [38], the brightness change is based on the estimated camera response function.

The second approach involves artificial intelligence, mostly represented by machine learning. In [39], Bayesian optimization exposure control based on entropy weighted image gradient through an internal prediction phase is presented. A block diagram of the algorithm can be seen in Figure 3.2. Gradient and entropy are calculated from the query image. Then, noise in the image gradient is reduced through weights calculated from the entropy, while simultaneously creating a saturation mask and using the activation function to feed the overall metric. The metric calculation is repeated until the Gaussian process finds the optimal value. In [11], support vector machines (SVM) with radial basis function are used to classify underexposed

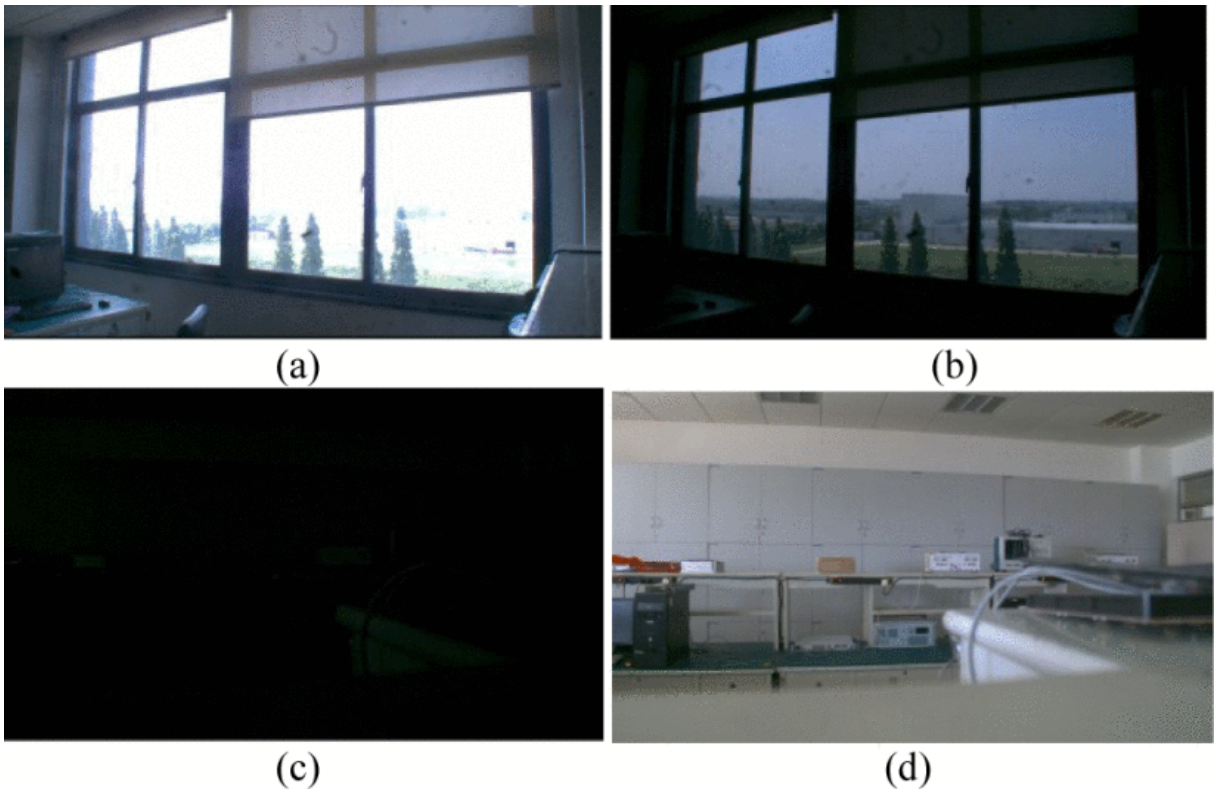

Figure 3.1: An overexposed scene before (a) and after adjustment (b), and an underexposed scene before (c) and after adjustment (d) [36]

and overexposed frames. In [40], a deep convolutional neural network (CNN) is trained to predictively adjust camera gain and exposure settings in real time, while in [41], semantic-aware exposure control system based on reinforcement learning is proposed based on Markov decision process (MDP). Both model-based and machine learning brightness control produce high quality results, but both approaches have their drawbacks.

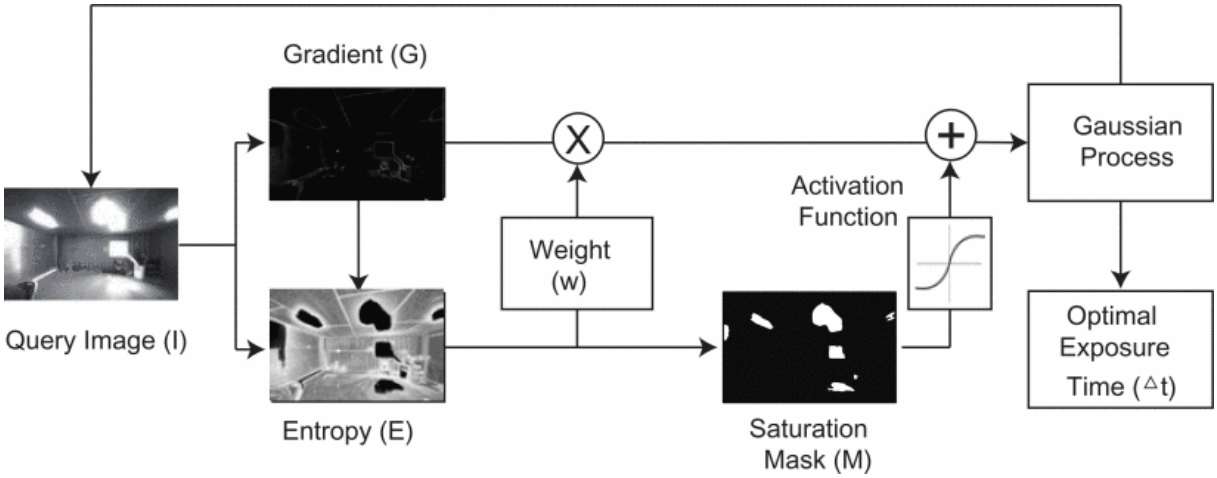

Figure 3.2: Block diagram for the Bayesian optimization exposure control based on entropy weighted image gradient through an internal prediction phase [39]

For the model-based approach, a properly estimated and accurate model of the system and

the disturbance is needed, which can be a very demanding task for a non-linear system and an unpredictable disturbance. On the other hand, machine-learning approach requires a large data-set with annotated videos and it is quite heavy to implement on real-time embedded systems.

The third, and still the most common approach to automatic brightness control is the control theory. The control theory methods can be divided into several subgroups. The first subgroup is fuzzy logic control [42]. Although it is often used for non-linear systems, it requires a very precise set of rules, which is not suitable for the types of system with an uncertain model. The second subgroup involves the feedback controller, the most common method used in the control theory. As stated before, it is based on a closed loop system in which the current state of the system is monitored, measured, and adjusted according to the reference value. There are several ways to reach the reference value, and the first way comprises the look-up tables (LUTs). In [43], LUT of different exposure settings is used while the gain setting is kept low, whereas in [44] there is a similar LUT, but for gain values. In [45], there are three respective look-up tables for analog gain, digital gain, and exposure time, which are used for adjusting the brightness of a scene. Although the calculation process is simplified, it is not robust enough for different disturbance values which are common in automatic brightness control systems. Next, there are feedback controllers with discrete step values. In [46], the exposure time is multiplied by the weight factor calculated from the brightness histogram; a flowchart of the algorithm is shown in Figure 3.3.

A similar approach is taken in [47], where the adjust step is added to or subtracted from the previous exposure time value depending on the scene brightness. Three different steps for adjusting exposure parameters are used in [48], while in [49], a control strategy with a fixed gain and exposure time step is used. If the brightness increases, exposure time is changed before the gain, and if the brightness decreases, the exposure time changes after the gain. Although the fixed step size feedback control is more robust than lookup tables, unwanted behavior still occurs due to discrete control values, and it leads to nonconvergence to the reference value, oscillations, and even system instability, similarly to lookup tables. Finally, the most common form of closed loop controllers is the variable-step size controller, with the main method being the PID controller and its variants, as a fast, resourceful, and accurate solution, but with a lot of parameters to tune and interact with. In [50], a PID controller with an adaptive target reference is used for an automatic exposure feature based on the evaluated overexposure ratio. In [51], two separate PI controllers for the gain and exposure parameters are used with the gain controller at a minimum value, if the exposure parameter does not reach a maximum. A position type PID is used in [52], while different PID controllers are used for gain and iris in [53]. PI controllers with experimentally obtained constants for exposure, gain, white-balance, and brightness are used in [54]. In [55], a dynamic coupled control system is used, which jointly controls a vector of exposure values with multiple PID controllers. A schematic of the system is

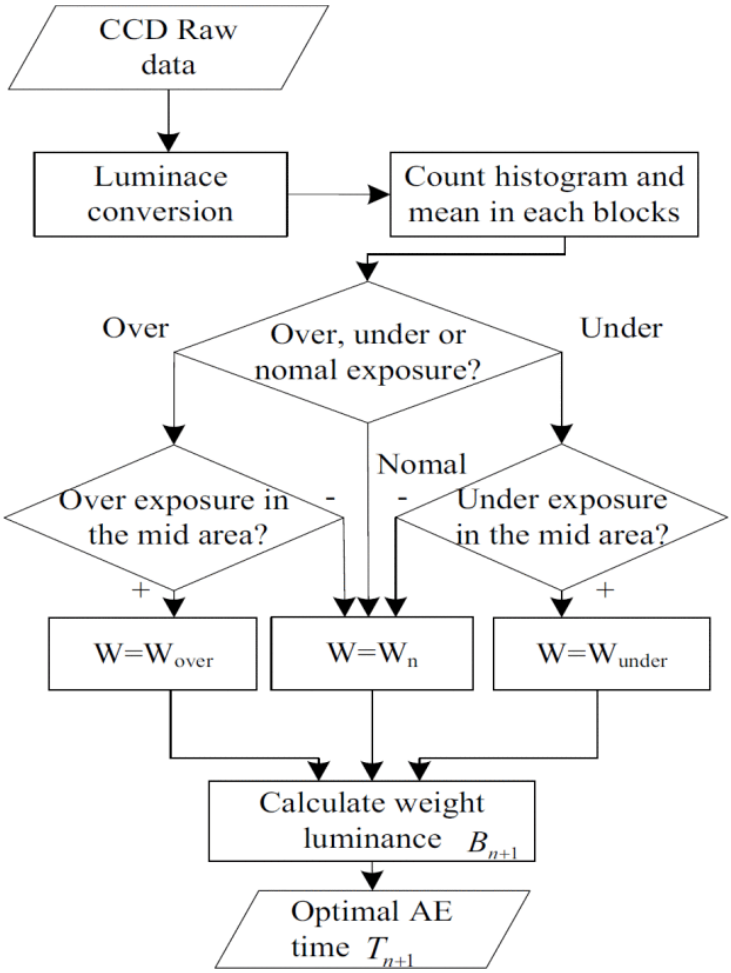

Figure 3.3: Flowchart of Jiang's algorithm for automatic exposure control with discrete steps by multiplying the weight factor calculated from the brightness histogram [46]

presented and described in Figure 3.4, where exposure control is combined with high dynamic range (HDR) compositing. Part (A) shows the automatic gain control, while part (B) describes HDR compositing. The system is improved in part (C), where multiple feedback control loops of exposures are introduced; they are coupled in part (D).

In [56] and [9], a PID controller is used for controlling the LED source in ROI-based histogram analysis with a fixed exposure level. Newly developed CMOS imaging system was created by [57], in which every pixel's exposure and frame rate are controllable in real-time with PI controllers. Table 3.1 provides a short overview of the methods' advantages and disadvantages described in the text, while an overview of automatic brightness control approaches is presented in Figure 3.5.

The next section offers an insight into the work related to face region detection.

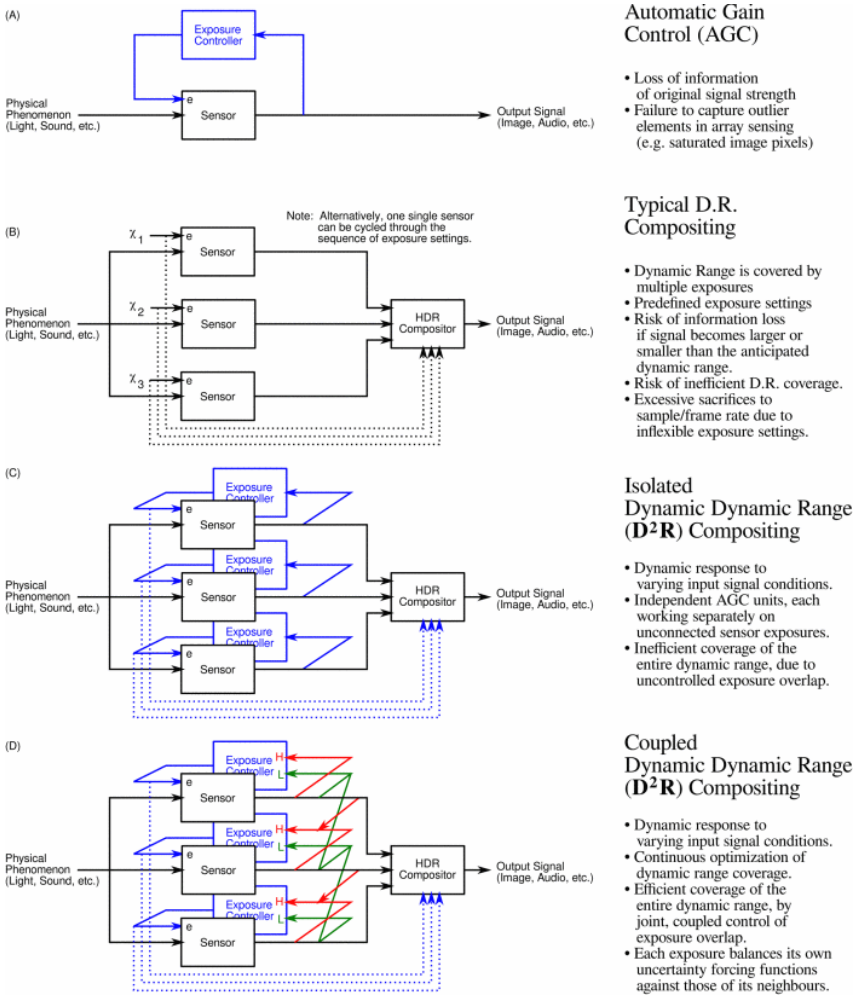

Figure 3.4: Schematic of Janzen's algorithm for multiple exposure feedback system that describes automatic gain control (A), HDR compositing (B), multiple feedback exposure time loops (C), and the coupling of the multiple feedback exposure time loops (D) [55]

Table 3.1: Advantages and disadvantages of the methods for automatic brightness control

Approach	Advantages	Disadvantages
Model-based method	High-quality results	Precise model needed
Machine-learning approach	High-quality results	Large dataset needed
Fuzzy logic control	Non-linear system friendly	Precise set of rules needed
Look-up tables (LUTs)	Simple to implement	Oscillations and instability
Fixed-step feedback controller	More robust than LUTs	Oscillations and instability
PID-variant controller	Fast, resourceful, accurate	Tuning parameters

3.2 Face region detection in the NIR spectrum

Face detection and face region detection are very important object detection tasks in the field of computer vision, and the first step in complex face analysis, face tracking, and facial

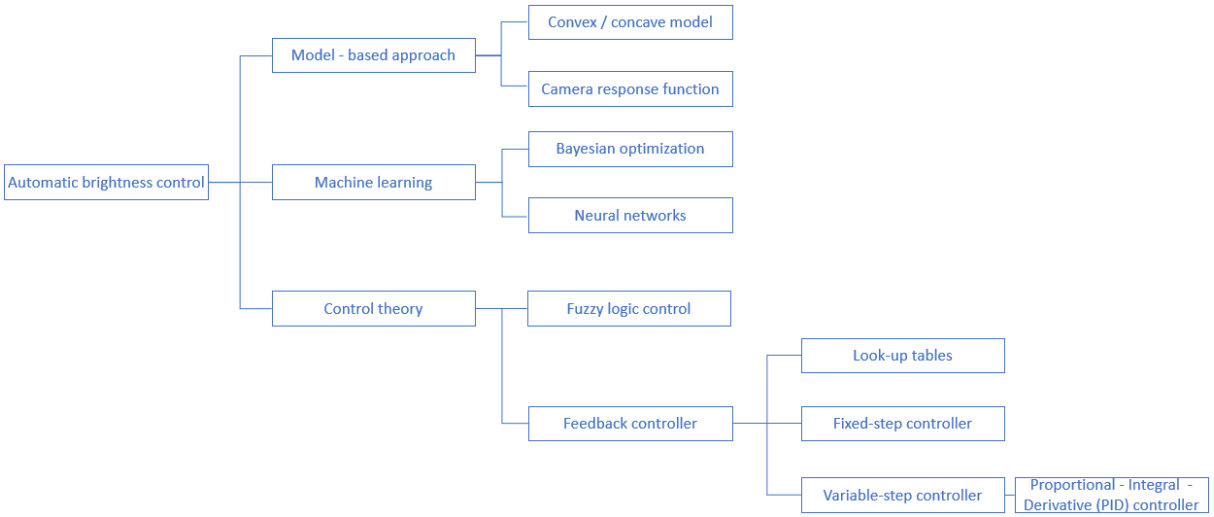

Figure 3.5: Automatic brightness control approaches overview

recognition. Face detection technology can be applied in various fields, from security [58] and law enforcement [59] to biometrics [60] and entertainment [61]. The amount of research that includes face detection in the VIS is enormous, but for face detection in the NIR spectrum the number of publications is significantly lower, and it further declines for applications that need to be in real-time. The NIR face detection methods can be divided into two major groups: feature-based and image-based methods. The feature-based methods are further divided into low-level analysis and feature analysis methods and active shape models. On the other hand, image-based solutions consist of neural networks and different statistical approaches. There are also several software developer kit (SDK) commercial solutions, such as Pittsburgh Pattern Recognition's SDK [62], Face++ [63], and OpenFace [25], whose face detection methods are not publicly available. In contrast, there are several open-source tools, such as *dlib* (used in [64] and [65]). However, very powerful hardware is needed to execute the face analysis algorithm, including face detection, and obtain the results in a decent amount of time.

Low-level analysis methods are the one discussed first. They are based on simple analyses, such as those of edges, gray level, and skin detection. In [66], skin detection is done by subtracting the images of two different infrared spectra, 940 nanometers and 970 nanometers. Although it is a very simple method to implement, the biggest disadvantage is the need for a multi-band NIR illuminant and diffusion filter. Skin detection method is often calculated from histogram analysis after the image threshold. In [67], a fixed threshold is set for image binarization in which the foreground represents the face, while the background is a non-face area. Although having a fixed threshold is also a very simple method, it is not robust enough for illumination changes in a scene. In [68], the dynamic threshold calculated with Otsu's method and horizontal and vertical projections are used in Figure 3.6, while in [35], the threshold is calculated automatically by using a linear discriminant analysis. Although the proposed methods may not be the most robust and precise solutions, they can certainly be used in approximate face region

detection as resourceful solution for embedded system implementation.

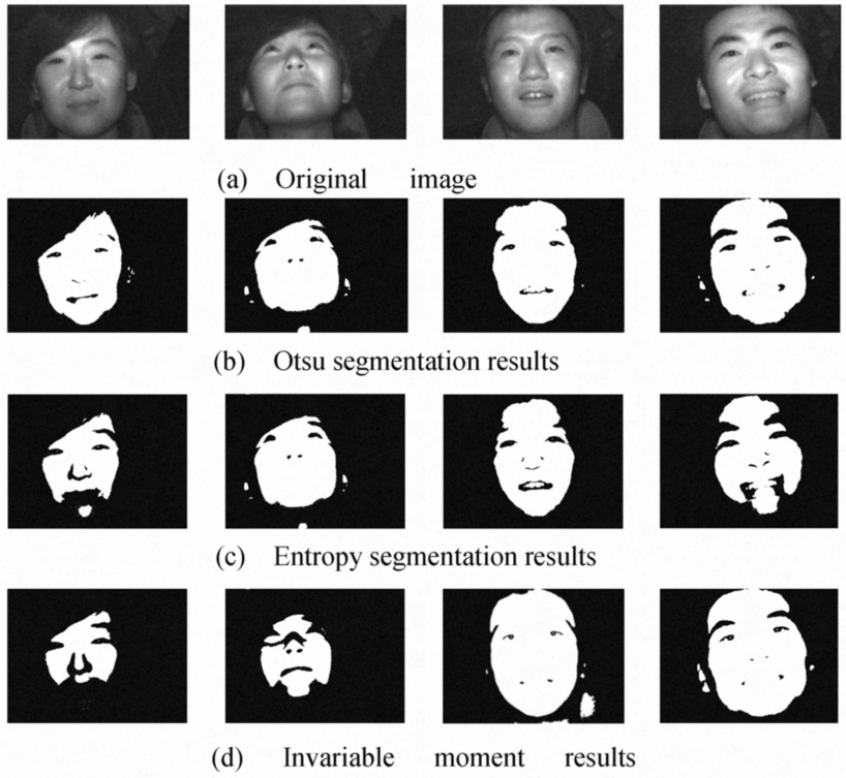

Figure 3.6: Image binarization with Otsu's method, and horizontal and vertical projections [68]

The second major feature-based face detection method is feature analysis. While low-level analysis is based on a simple analysis, feature analysis comprises a more complex analysis, such as searching for the eyes, nose, and mouth, which are further correlated to successfully finding the face in the image. In [69], optimal Gabor filters are created for facial feature extraction by using Gabor transform and a genetic algorithm. Successfully detected faces with optimal Gabor filters can be seen in Figure 3.7. The method is accurate, but it requires a significant amount of time (150 milliseconds) to perform face detection. Gaussian Mixture Model (GMM) face feature detector and enhanced appearance model are used in [70] and [71].

A breakthrough in the field of face detection happened in 2001, when the Viola-Jones algorithm was published [72]. There are three major contributions of that paper. The first one is integral image representation, which greatly reduces processing time and power. The integral image is derived from the original image by replacing the number in a certain location with the sum of the image pixels up to that location in the original image, including the location itself. Having the integral image results in a rapid summation calculation over the subregions in a constant time, regardless of the subregion size, and without the use of repeated actions to find and add an element. An example is provided in Figure 3.8. The upper left image is an original image, while the upper right image is an integral image. Both green and blue elements in the integral image are calculated by summing up all the values up to the element position in

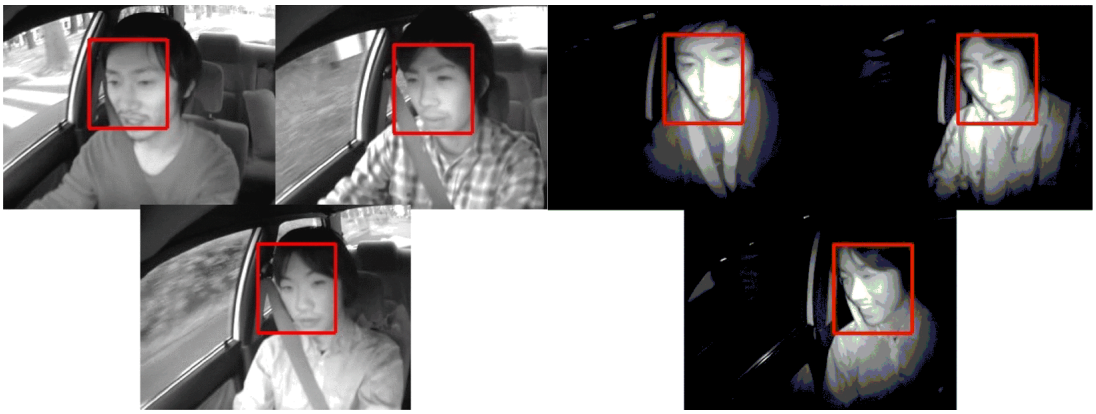

Figure 3.7: Successful face detection during daytime (left) and nighttime (right) with the optimal Gabor filters [69]

the lower part of the image; the example has been obtained by calculating the sum value over the purple area in the original image (lower left). This requires going through each value in the original image, and the speed of execution directly depends on the number of values in the region. In the integral image, however, it is just a single calculation in which the elements in red and yellow are subtracted from the elements in blue and green to calculate the same sum as in the original one (lower right).

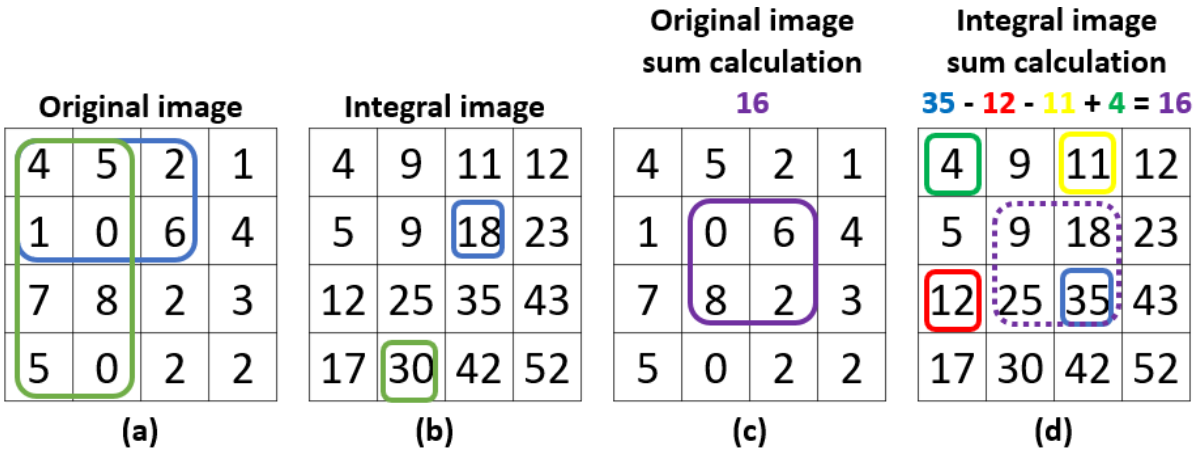

Figure 3.8: Original image (a), integral image (b), original image subregion sum calculation (c), and integral image subregion sum calculation (d)

The integral image is used with Haar-like features (classifiers) to categorize subparts of the image. A window is moved over the image in which Haar-like features are calculated. The difference between bright and dark regions is calculated and compared to a learned threshold to separate objects from non-objects. Haar-like features are divided into edge features, line features, four-rectangle features, and their variations, as seen in Figure 3.9, which also provides an example of the use of Haar-like features.

The number of features tends to be very large. For instance, a 24x24 pixel image has a total of 162,236 features [73], but not all features are equally relevant for face detection. The division

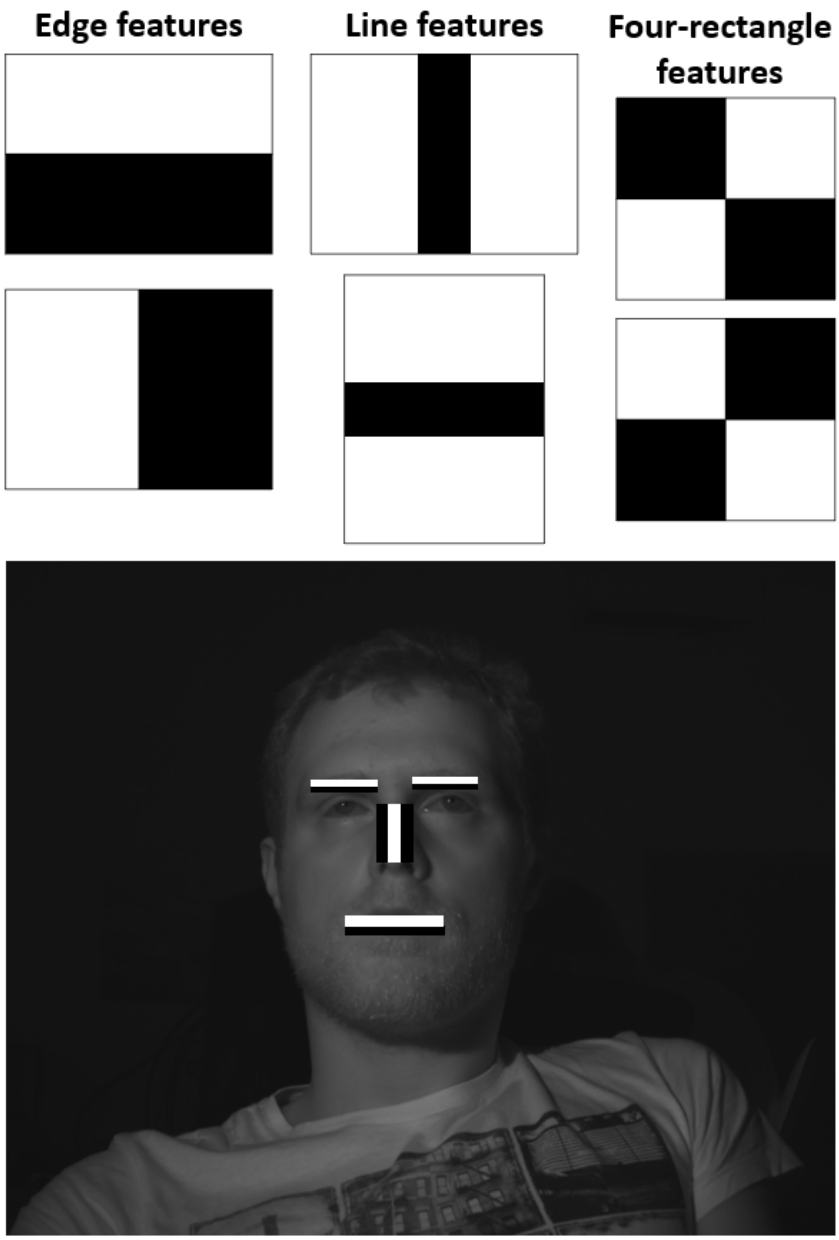

Figure 3.9: Types of Haar-like features (up) and an example of the use of Haar-like features (down)

between relevant and irrelevant features is the second major contribution of the paper, and it is done with Adaptive Boosting (AdaBoost) [74]. A single Haar feature classifier does not yield good results alone (but it is better compared with a random guess), and it is considered as a weak classifier/learner. Multiple weak classifiers are combined into a weighted sum to build a strong classifier. When a model is created, equal weights are given to all the data points, and, if the data points are wrongly classified, their weights are increased, which puts more emphasis on them in the next model iteration, and, eventually, results in lower error. In the Viola-Jones algorithm, AdaBoost is used for both selecting a small set of features and training the strong classifier. The last major contribution of the Viola-Jones algorithm is combining more complex classifiers in a cascade structure - the attentional cascade. Strong classifiers are formed into a

binary classifier. If a positive match occurs at one step, it is sent to the next feature; otherwise, the image is rejected, and the computation stops. If all of the feature checks are positive, an image subregion is classified as a face image subregion, as seen in Figure 3.10. This results in a significant decrease in computation time, while detection performance is increased.

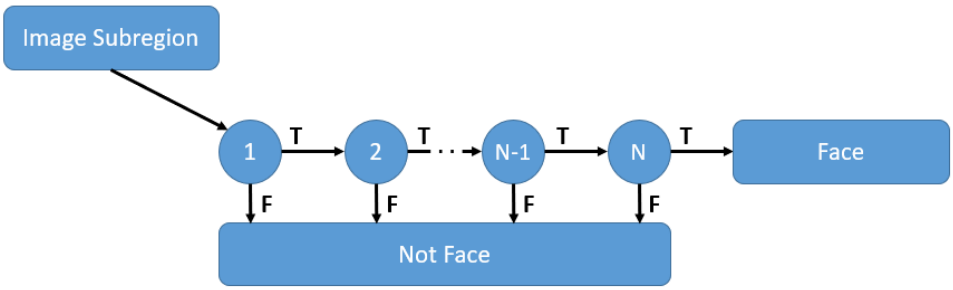

Figure 3.10: The attentional cascade illustration

Even today, it is still the most common method of face detection, even for NIR spectrum images ([75], [76]) implemented on embedded systems ([77], [78]). There are many derivations of the Viola-Jones algorithm. In [79], a part-based method for improving face detection robustness was used, with subwindow local binary pattern histograms (SLBPH) extracted from the NIR face image and AdaBoost learning, while in [80], FloatBoost learning is used instead of the AdaBoost in the Viola-Jones face detection algorithm. In [78], modified census transform (MCT) AdaBoost is used, while in [81], AdaBoost face detector and Haar features, trained using 50,000 NIR face examples, are used for face detection. Local binary pattern (LBP) histogram features and AdaBoost learning are used in [81] to detect facial features, while saliency map (SM) model is used with AdaBoost algorithm for face detection. In [82], the Viola-Jones algorithm is used for frames, with little motion on the foreground region in the frame. The example of face detection on an ambient frame, illuminated frame, and a difference frame can be seen in Figure 3.11, where the best detection results are in the difference frame.

In [83] and [84], a cascade random forest classifier (CaRF) with low-dimensional Haar-like features and oriented center symmetric-local binary patterns (OCS-LBP) are used for face detection in driver-state monitoring systems.

Although feature analysis methods are more accurate for face region detection than low level analysis, they are more computationally complex and require more resources, which makes them inadequate for real-time embedded system implementation.

The final feature-based face detection method is the active shape model (ASM). ASM is an iterative method which aims to find the best match location between the previously used model and the data in a new image. For NIR face images, however, there is not much research with ASM face detection, and most research uses the deformable part model (DPM). DPM creates a model of an object in which parts of the model are mutually constrained. If a face is modelled, the aim is to retain anthropomorphic configuration, and if there is an occurrence of geometri-

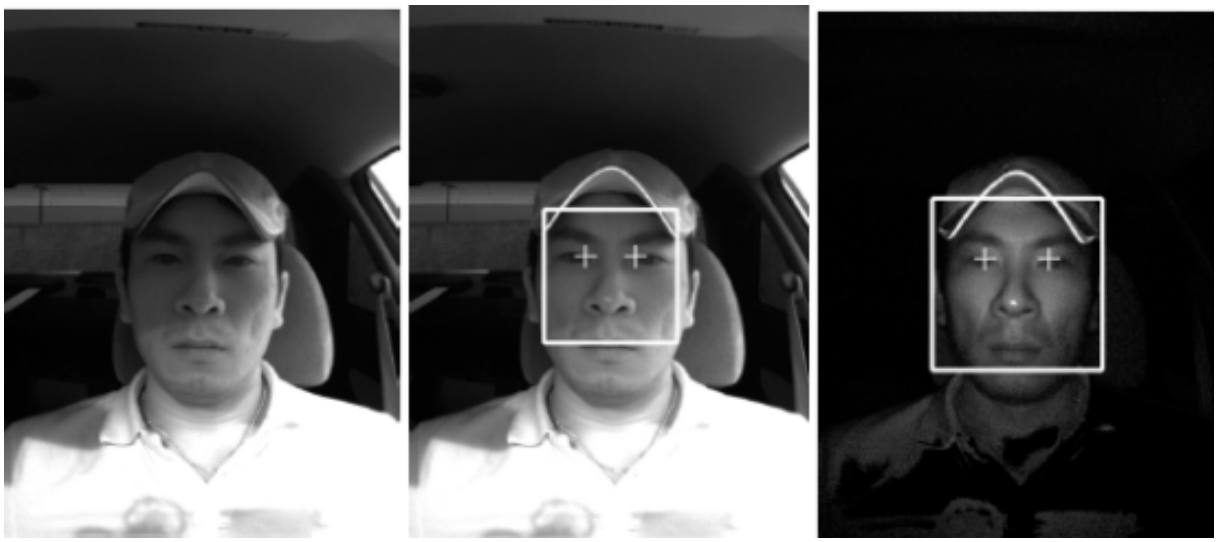

Figure 3.11: The example of the face detection on an ambient frame (left), illuminated frame (middle), and a difference frame (right) [82]

cally illogical configurations, the deviation from an expected configuration is penalized, which is called the spring model [85]. The deformable part model is used for detecting faces in [86], for cell phone usage violation detection in [87], and for cell phone usage and seat belt violation detection [88]. DPM is often more accurate than feature analysis methods, but still tends to be slow for real-time execution on embedded systems [89].

The second major approach used for NIR face detection is the image-based approach. The image-based approach uses subwindows scanned throughout the image, and every pixel is scanned to classify a face and non-face region. Although there is a subdivision of image-based face detection methods for VIS images, the most widely used method in NIR images are neural networks. Neural networks are trained to recognize the face pattern for face detection, and to predict the output for a new set of similar faces. In [90], partial face detection and facial features extraction are conducted with multiscale double supervision convolutional neural network (MDSCNN) from several multiscale patches cropped from the aligned images. MDSCNN workflow with patch-to-patch matching to find the minimum matching object can be seen in Figure 3.12, while the structure of the MDSCNN is shown in Figure 3.13, with four convolutional layers followed by max pooling and one full-connected layer.

Multi-Task Cascaded Convolutional Neural Network (MTCNN) is used for face detection and alignment in [91], [92], and [93]. Neural networks are a powerful tool to yield high accuracy in face detection; however, powerful hardware is needed in order for them to work properly in real-time, and big annotated datasets are needed to train them. That is why neural networks are rarely implemented on embedded systems. Table 3.2 gives a short overview of the advantages and disadvantages of the face detection method groups described in the text, while an overview of face detection approaches in the NIR spectrum can be found in Figure 3.14.

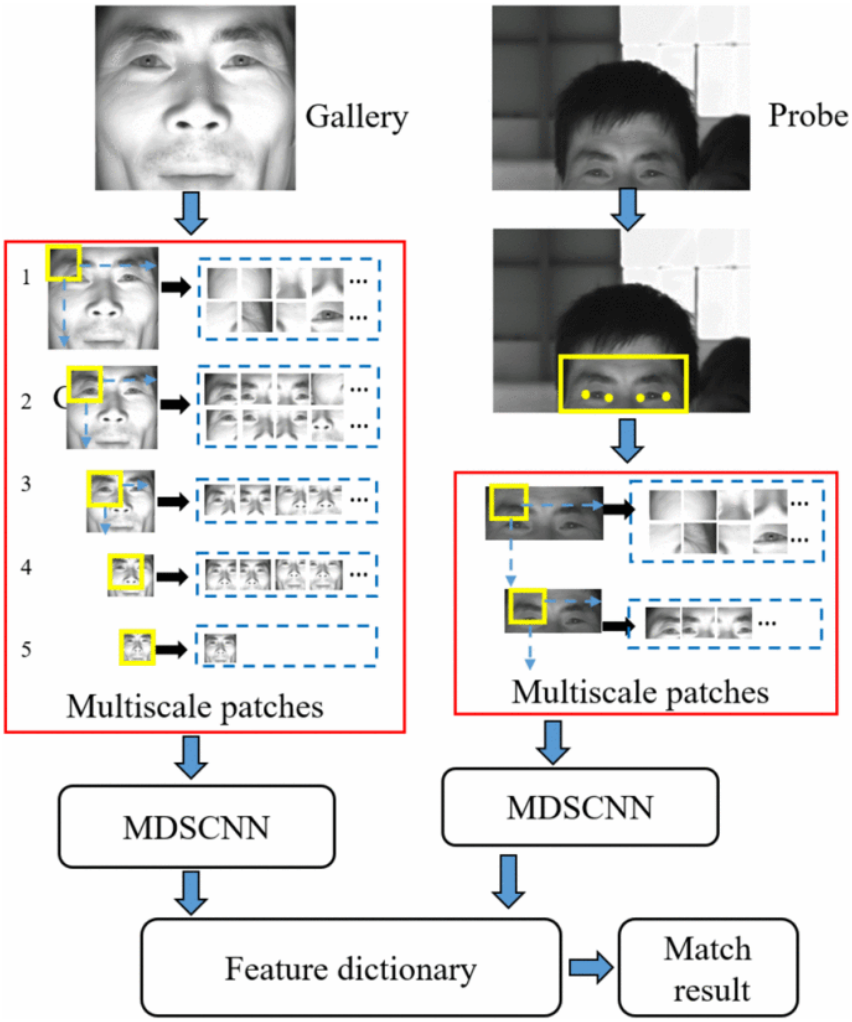

Figure 3.12: Multiscale double supervision convolutional neural network workflow with patch-to-patch matching to find the minimum matching object [90]

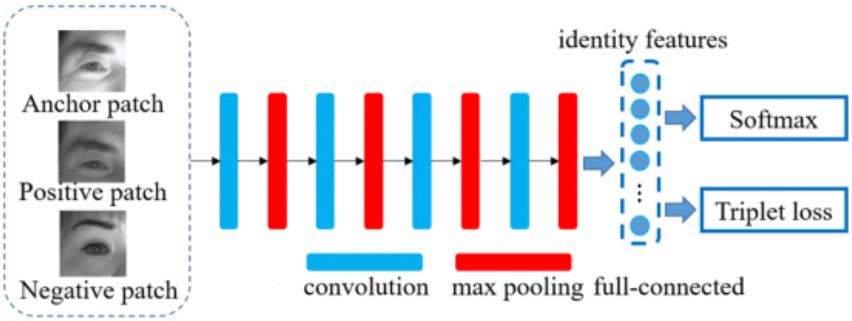

Figure 3.13: Multiscale double supervision convolutional neural network structure with four convolutional layers followed by max pooling and one full-connected layer [90]

In the next section, the related work on automatic brightness control combined with face region detection is presented.

Table 3.2: Advantages and disadvantages of the methods for NIR face detection

Approach	Advantages	Disadvantages
Low-level analysis	Very fast execution	Medium accuracy
Active shape model	High accuracy	Slow for real-time embedded system execution
Feature analysis	High accuracy	Slow for real-time embedded system execution
Neural networks	Very high accuracy	Powerful hardware needed, big datasets for training

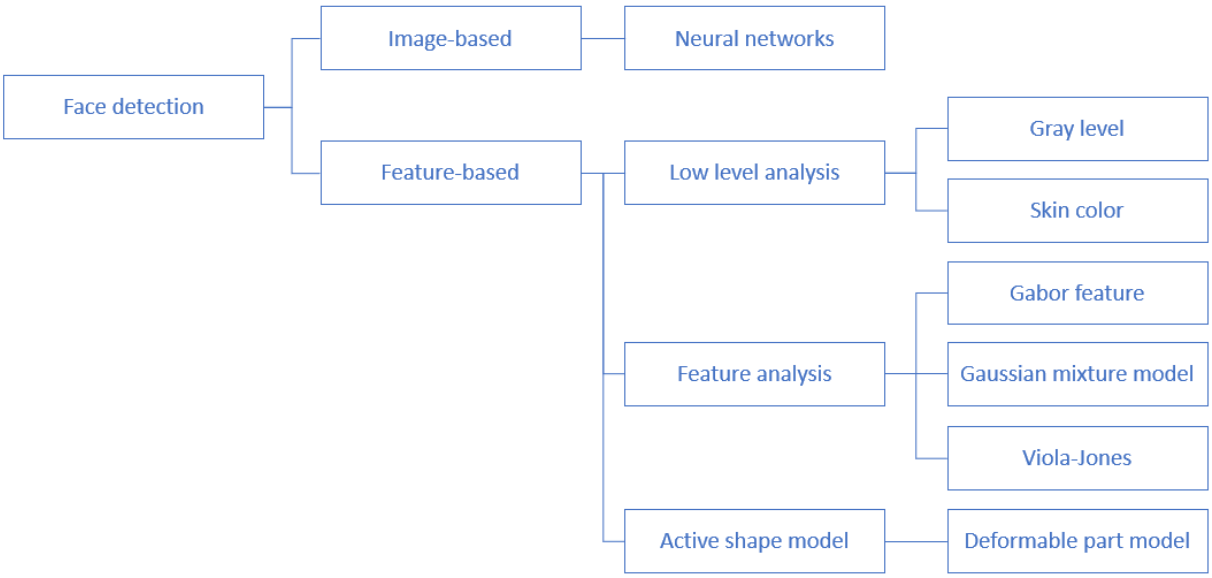

Figure 3.14: Face detection approaches in the NIR spectrum – overview

3.3 Automatic brightness control with face region detection

As seen in the previous subsections on related work, there are multiple individual approaches for automatic brightness control and face region detection in the NIR spectrum. To improve automatic brightness control, face or face region detection is combined with automatic brightness control algorithms to put the focus of the control algorithm on the face area. Automatic brightness control is still a pre-processing task, so, for common hardware, it needs to be fast and resourceful if real-time execution should be at least 30 FPS. However, there is not so much publicly available research which includes the combination of face region detection and automatic brightness control, especially in the NIR spectrum.

First, a pseudo-control algorithm is used in [94] for face spoofing detection. Gamma compression is applied to dark images and gamma expansion to bright images as a simple exposure correction. This is not an automatic brightness control task, it is an image pre-processing task. Face regions are detected with the Viola-Jones which, again, is a precise method, but slow for real-time embedded system implementation. The proposed face spoof detection algorithm is seen in Figure 3.15. Original NIR images without and with NIR lighting, NIR images, and

NIRL images, respectively, are sequentially captured and gamma corrected to detect face regions with an elliptical mask. From NIR and NIRL images, near-infrared differential (NIRD) images are constructed, in which elliptical mask is used to differentiate face and non-face areas. Context clues and lighting texture are employed for non-face and face areas to detect the spoofing.

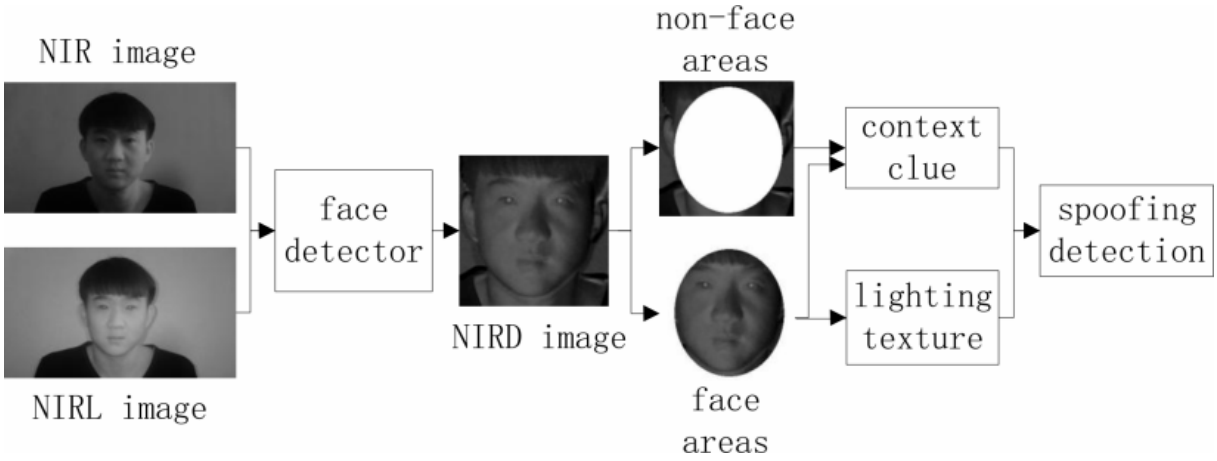

Figure 3.15: Face spoof detection framework [94]

The first proper automatic brightness control with face region detection is [95]. In the research, a tri-band camera system is used with lower NIR, higher NIR, and VIS band for comparative testing purposes. The algorithm diagram is shown in Figure 3.16. The first step after the frame acquisition is foreground-background segmentation based on frame differencing between the two NIR bands. The dynamic threshold for segmentation is calculated with Otsu's method. After that, the scene luminance is calculated for each NIR-band to adjust the scene brightness with Luminance-Voltage diagram [95]. The next step is skin detection, in which the binary image is created by weighted subtraction. A series of morphological operations is conducted on the binary image, which is then passed to the face detection module.

Face detection operates in two modes. In the first mode, multi-band integral projections are used to detect the eyes and the eyebrows, or, if the first mode fails, dynamic thresholding and template matching is used for facial feature detection. The example of the execution of the algorithm in lower (a) and higher (b) NIR band images can be found in Figure 3.16.

Although face region detection performs well, the major drawback is its automatic brightness control, in which a look-up operation at the Luminance-Voltage diagram, a one-dimensional LUT, is used to change the brightness in the lower and higher NIR band, which is not resistant to ambient illumination changes. Also, the performance of the algorithm is slow.

The second and final research [29] uses a pre-calculated weighted face mask in two stage exposure adjustment to emphasize the most important face regions. If the face region is not found, the whole image is used in the calculation. In the first stage, global exposure is adjusted based on a single shot to start the recognition process, while in the second stage, parameters are

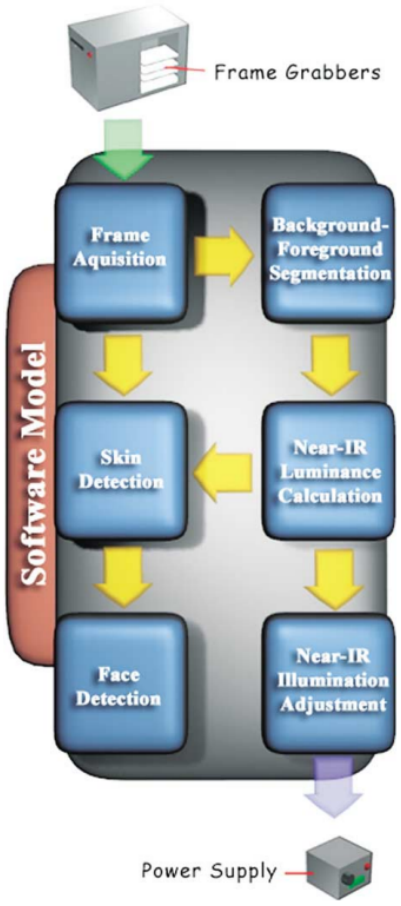

Figure 3.16: Algorithm diagram of the Dowdall for automatic brightness control and face region detection [95]

Figure 3.17: Algorithm execution for detecting eyes in lower NIR band image (a) and higher NIR band image (b) [95]

precisely adjusted to get the best image quality for the iris recognition. The exposure adjustment problem is modeled as a dependency between exposure time and the mean sample value (MSV) as a brightness level indicator. After the initial exposure time and gain guesses are made, predefined value of exposure time is added to or subtracted from the initial one until MSV reaches the value inside the wanted interval – a confidence zone. Figure 3.18 offers a visual comparison of different auto exposure methods.

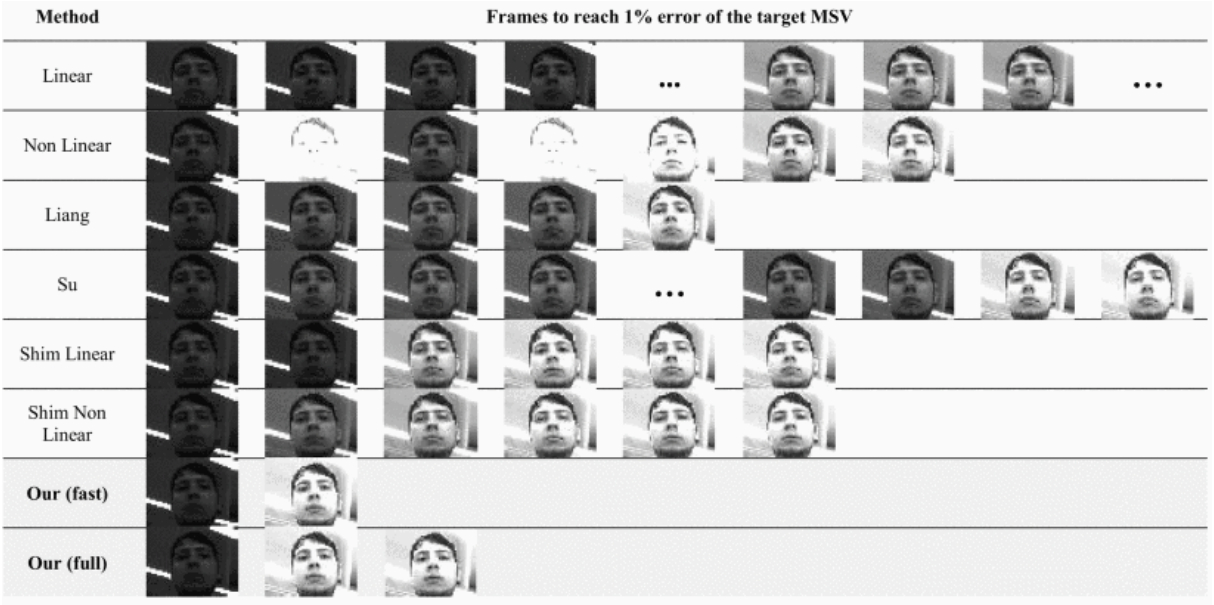

Figure 3.18: Visual comparison of different auto exposure methods [29]

Although this method tends to be fast, having a fixed step of exposure time and gain changes can result in oscillations, even for the mean sample value. Also, the pre-calculated mask can be prone to errors in different conditions, such as pose and ambient illumination changes.

A short overview of the automatic brightness control approaches with face region detection is shown in Table 3.3.

Table 3.3: Automatic brightness control with face region detection algorithms with automatic brightness control (ABC) and face region detection (FRD) elements – overview

Algorithm	Element	Overview
Sun [94]	ABC	Gamma corrected brightness, pseudo-brightness control
	FRD	Accurate, slow for real-time embedded systems
Dowdall [95]	ABC	Fast, look-up tables prone to oscillations
	FRD	Accurate, slow for real-time embedded systems
Gnatyuk [29]	ABC	Fast, fixed step for the exposure time and gain prone to oscillations
	FRD	Fast and accurate, but uses a pre-calculated face mask prone to error

Each of the mentioned publications has certain weaknesses, especially for embedded system implementation. Although face detection elements can yield good results, automatic brightness control is often rudimentary, which results in a slow or non-existent convergence. Thus, there is a need for a better solution, which should entail proper and robust brightness control with fast convergence in the region in which the face is located. The next chapter provides an insight into the spatio-temporal method for approximate face region detection in the NIR spectrum in real-time.

Chapter 4

Real-time spatio-temporal method for approximate face region detection in the NIR spectrum

Many methods for face and face region detection have been described in the previous chapter. Actually, the face is in only a part of the image; it does not occupy the whole scene, and that is why face detection is needed to isolate the part of the image which includes the face, which is then used for automatic brightness control. However, face detection is a complex and often slow task for real-time embedded system execution, especially for a neural network solution. Thus, a good representation of face brightness is needed to reduce the computational costs. Instead of using the exact face location for automatic brightness control, face region is used as a proper substitute. Face region detection is an auxiliary feature to automatic brightness control, which means that it needs to have similar properties as automatic brightness control: it needs to be fast, accurate, and adjusted for embedded system real-time implementation and execution. Also, it needs to have a good detection rate for NIR spectrum images. In most of the embedded system image signal processing (ISP) pipeline, the image is divided into x rows and y columns, which is often used for faster image analysis. While the exact face detection is not necessary for the proposed purpose, the major requirement for face region detection is stability. In common case, head movements are not significant between two frames at 30 FPS, so the face region should not change drastically between two frames. The rapid changes in region boundaries result in a rapid change of foreground/background ratio in an image, and in oscillatory behavior of the automatic brightness control system. Calculating the current face region boundaries is as important as taking the face region boundaries in previous steps into account during the region boundary calculation. For this reason, a real-time spatio-temporal method for approximate face region detection in the NIR spectrum based on sampled skin detection is introduced to separate the potential face region from the background. The proposed method consists of two elements:

the spatial and temporal element, which are described in the following subsections. After the description, the method is evaluated and compared to the still most commonly used method for face detection on embedded systems, the Viola-Jones, followed by a discussion. The spatial element of the method is described in the next section.

4.1 The spatial element of the spatio-temporal face region detection method based on sampled skin detection

The first part of the proposed spatio-temporal face region detection is the spatial element. In this step, the main task is to locate the face region detection in the current frame. The workflow of the spatial element of the proposed algorithm can be found in Figure 4.1.

The first step is acquiring an image through an ISP pipeline. After the acquisition, the frame is divided into x rows and y columns. The division of the frame is most often already implemented in the ISP, but the numbers of rows and columns are usually fixed and mutually equal. In most cases of the face being in the focus, the face is rarely located in the outer tiles of the grid. To reduce the computation time, edge tiles of the image are clipped from the calculation. The number of clipped tiles depends on the image sensor resolution and the total number of tiles, but, empirically, seen from the existing datasets, it should range from 12% to 18% of the total image width and height from each side. For instance, if one row and one column are removed from an 8x8 grid from each side, the area becomes a 6x6, which is a 43.75% reduction from the initial image, and, accordingly, the number of calculations is also reduced. In the remaining tiles, the mean pixel value is calculated, whether from counting the number of each pixel value manually or from the region histogram from the ISP. The calculated mean pixel value is the representation of the tile. In this step, the image is reduced to a $x-m$-by-$y-n$ matrix, where m and n are the respective numbers of clipped rows and columns. The new matrix is binarized, and the dynamic threshold is calculated with Otsu's method, as seen in the chapter "Theoretical foundations". The binarization mask with zeroes in the background area and ones in the foreground area calculated with Otsu's method is applied on the initial image, which is then propagated to the next step. As seen in Figure 4.1, binarization mask is of irregular shape, which is not common for face region detection algorithms. A regular shape, such as a rectangle, is often used to represent the face or face region. For this reason, a rectangular area layer is put at the top of the binarization mask to find the biggest rectangle area that satisfies certain conditions inside the rectangle area. There are four different variants in which the biggest rectangle area can be created: A, B, C, and D. The first variant, A, comes from a simple premise that a rectangle area should contain only the area with positive values in the matrix (ones - foreground). While this is the simplest approach, some parts may be excluded because of certain lighting conditions, shadows, or occlusions. The variant B expands the area

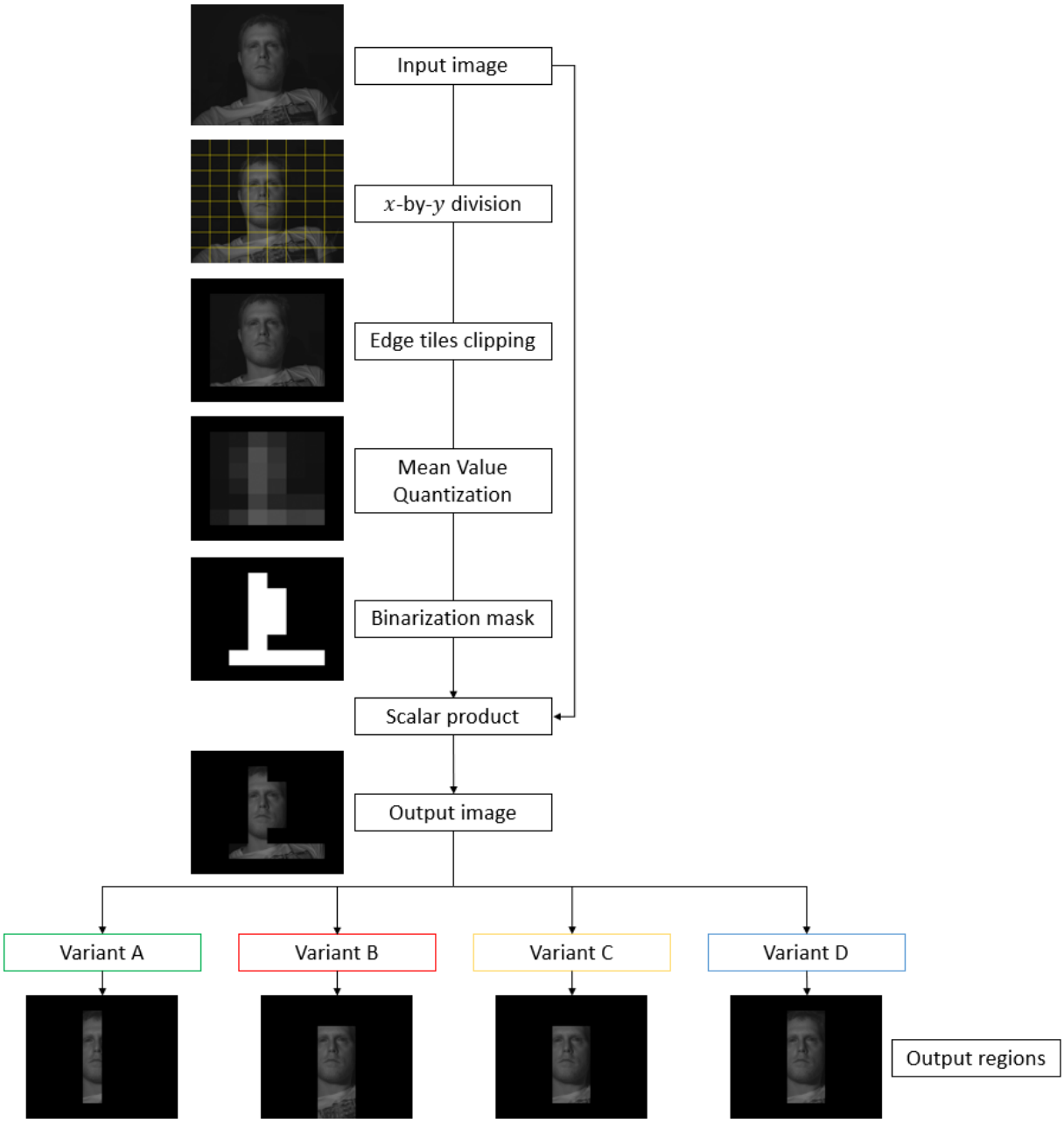

Figure 4.1: Flowchart of the spatial element of spatio-temporal sampled skin detection

of interest calculated in variant A by looking for the biggest rectangle in which there is a certain percentage of positive values; empirically, at least 75% of the rectangle are positive values. The third variant, C, is similar to variant A, but it includes an initial step before finding the biggest rectangle area. It searches for negative tiles with the highest number of surrounding positive tiles, and, if the number of positive tiles around a negative tile is over 50%, it is considered as a positive tile, and as a part of the rectangle. The fourth variant, D, is similar to variant B in terms of expansion, but the rectangle area can be expanded to columns only. The area calculated in variant A observes the columns left and right to the calculated area, with heights of the column equal to the area. If the number of positive values in a column is at least 50%, this column is

considered to be a part of the rectangle area, and it is included in further mean pixel calculation. There is an additional feature for the aforementioned variants, in which all calculated regions can be expanded by a certain number of rows and columns to reduce the possible error if the face region is located on region edges. For a small number of tiles (less or equal to 8), the number is often 1. However, this expansion increases the number of required computations and can be significant for a larger number of tiles. In Figure 4.2, the bounding boxes for each variant of the spatial step are shown, while in Figure 4.3, the area calculation of the variants is shown for three different artificial binarization masks. In both Figures 4.2 and 4.3, the green bounding box represents the results for variant A, red for variant B, yellow for variant C, and blue for variant D.

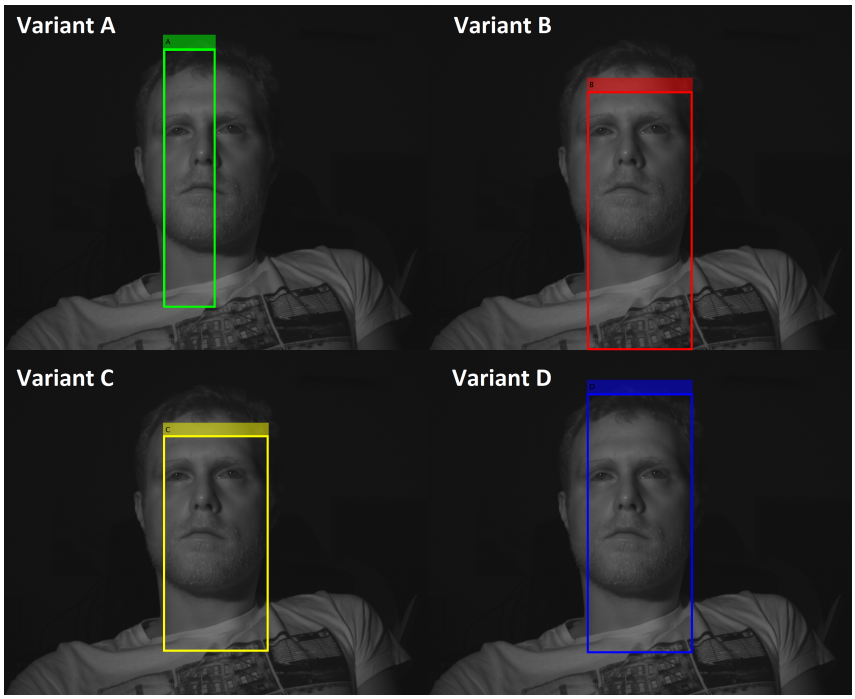

Figure 4.2: The bounding boxes of the spatial step variants: A (green), B (red), C (yellow), and D (blue) of the spatio-temporal sampled skin detection

The output boundaries of the rectangle area calculated from the spatial element are propagated to the next phase: the temporal element of the algorithm.

4.2 The temporal element of the spatio-temporal face region detection method based on sampled skin detection

For automatic brightness control, it is very useful and important to detect the face region to reduce the calculating area in each frame. However, as mentioned before, having a stable face region without major location changes over time is as important as having the face region location, because sudden and frequent region changes will cause oscillations in the control

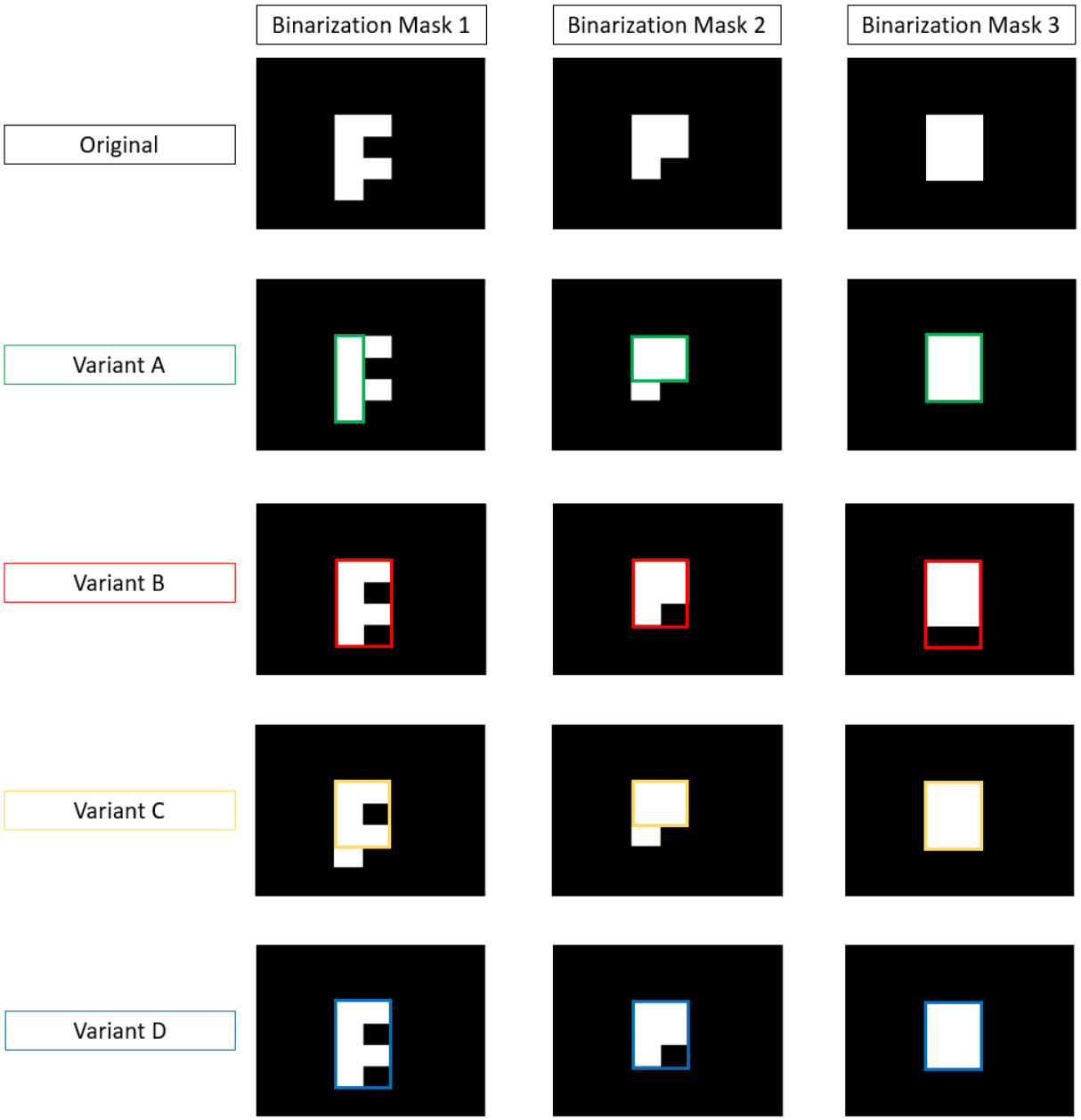

Figure 4.3: Approximate face regions obtained by the four variants for three different artificial binarization masks

system. For this reason, a temporal step of face region detection is introduced to reduce frequent changes in the detected face region, as a low-pass filter for region boundaries received from the spatial element of the algorithm. As seen in Figure 4.4, the temporal step consists of two parts: boundary time series median filtering and hysteresis.

Boundary time series median filtering is described as follows. The temporal element receives four different values for each of the region boundaries: the up boundary *upBoundary*, the down boundary *downBoundary*, the left boundary *leftBoundary*, and the right boundary *rightBoundary*. In further text, all boundaries are replaced with a single boundary value *xBoundary* for easier reading. For each of the four region boundaries, a new calculated boundary from the

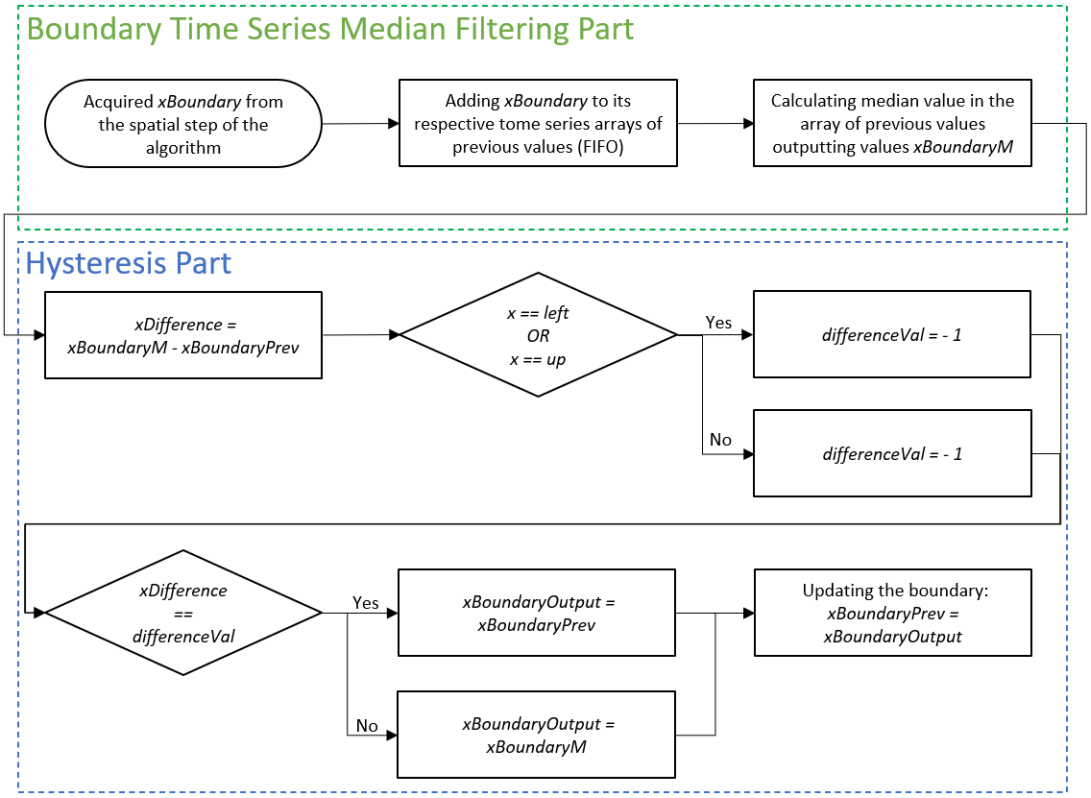

Figure 4.4: Flowchart of the temporal element for one boundary of the spatio-temporal sampled skin detection

spatial element *xBoundary* is added to the array of the $k-1$ previous boundary values according to the First-In-First-Out (FIFO) principle. The values of boundaries are sorted, and the median boundary value is used for further calculation, *xBoundaryM*. Median value of odd length array is chosen instead of the mean value to avoid dealing with the potential floating-point results obtained through division. As stated earlier, the face movement in common conditions between the two frames at 30 FPS is not significant, so the boundary changes should be minimal. The boundary time series median filtering is used to prevent the potential sudden and significant changes in boundaries due to extreme light conditions, face occlusions, or similar occurrences. The calculated median values are propagated to the second part of the temporal element, hysteresis. The movement of the face between frames is often subtle, but the face sometimes occurs in a small part of a new tile, while the rest of the tile is the background, but it is still used in the calculation. For this reason, the value *xBoundaryM* is compared with the previous value of that boundary, *xBoundaryPrev*. The difference between the two values *xDifference* is compared to the value *differenceVal*. *differenceVal* can have two values: -1 for the left and up boundary and 1 for the the right and down boundary. If those conditions are true, the hysteresis stops the region expansion, and the output boundary *xBoundaryOutput* is set to the previous value *xBoundaryPrev*; if they are not, it is set to the calculated median *xBoundaryM*. Finally, the boundary values are updated and *xBoundaryOutput* is stored to *xBoundaryPrev*, which ends

the current iteration of the whole face region detection algorithm. The example of the hysteresis element is shown in Figure 4.5.

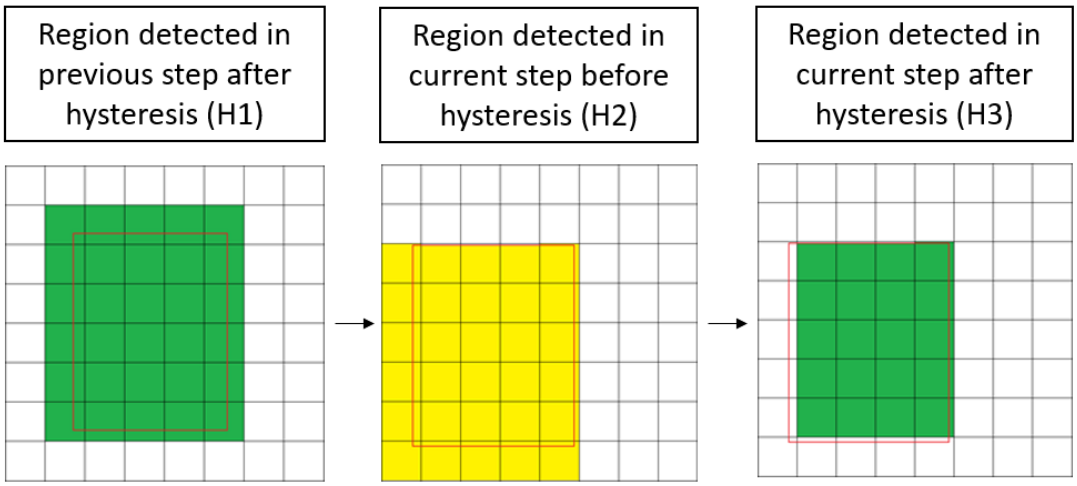

Figure 4.5: Hysteresis effect on a region expansion attempt

In H1, the red rectangle represents the actual face position, while the green tiles represent the calculated face region from the previous step. In H2, the face moves slightly to the left and down, and the calculated region is moved in the same direction. The face region is colored yellow. In H3, the hysteresis is calculated and the face region is reduced. The left and down boundary movement would enlarge the region, so the hysteresis reverts it to previous values. The right and up boundaries also change, but the region is not increased by it so no change in those boundaries occurs. The previous region in hysteresis does not need to be the same as the one in the boundary time series array due to hysteresis in the previous step. The order in the temporal step is interchangeable – boundary time series median filtering and hysteresis can switch places and be used independently. However, both boundary time series median filtering and hysteresis are non-linear operations so the result could differ from the described algorithm. In the next section, the performance of the algorithm is evaluated.

4.3 Experimental Results

After the description of the proposed method, this section offers an evaluation of its performance. Although there are many face detection algorithms for comparison, the proposed method is used with automatic brightness control on real-time embedded systems. The still most used face detection algorithm on embedded systems is the Viola-Jones, so a direct comparison with that algorithm is made. Face and face region detection are best evaluated with a benchmark dataset. The required dataset for this case should be comprised of annotated infrared face videos captured during daytime and nighttime under different illumination conditions and scenarios. The dataset that fulfils most of the requirements is the face alignment dataset used in

driving (FADID) [84], with examples from the dataset provided in Figure 4.6.

Figure 4.6: Example images from the face alignment dataset used in driving (FADID) [84]

The dataset was constructed for facial landmark detection under real driving conditions which consist of pose changes, illumination changes, and partial occlusions by hair or sunglasses. The resolution of the images is 720x480. The dataset contains the test and training set for both the daytime (3,327 images) and the nighttime (2,583 images). However, the proposed algorithm is not incorporated with a learning step, so the training sets can also be used in algorithm testing. The test set images are already divided by a single scenario, while the training sets need to be arranged. The daytime training set is divided into six batches: images 00001-00260, 00261-00517, 00518-00735, 00736-00971, 00972-01243, and 01244-01493, while the nighttime training set is divided into four batches: images 0001-00193, 00194-00397, 00398-00581, and 00582-00862. However, images 00796-00862 were not used in the calculation because of the sequence discontinuity at images 00796 and 00826. The facial landmarks are annotated together with the face location containing the top-left and bottom-right coordinates. However, the nighttime test set was not properly annotated, and it needed to be redone manually – the previous annotation procedures were followed to obtain similar face boundaries. Additionally, 3,327 daytime and 2,583 nighttime images were used for algorithm evaluation; faces with glasses appeared in 1,637 daytime and 581 nighttime images.

Two evaluations have been performed for the face region detection algorithm. In the first evaluation, statistical measures of the proposed algorithm's face region detection performance are calculated and compared with the Viola-Jones, while in the second evaluation, the execution time on a proper embedded system is tested. For the first evaluation, the whole dataset (except the nighttime training set images 00796-00862) is used, whereas for the second evaluation, only a small portion of the dataset is used for embedded system implementation. The embedded system used in this system supports grids up to 8x8, and this grid was used for both evaluations.

4.3.1 Statistical measures of quality for the proposed method

In this subsection, the performance of the proposed algorithm is tested and compared to the Viola-Jones algorithm. The Viola-Jones is an algorithm that determines the exact face location in the scene. However, in this thesis, it is the face region that needs to be located. Therefore, the face boundary box detected by the Viola-Jones is expanded to the whole region that contains the initial face boundary box, as seen in Figure 4.7, in which the red rectangle is the exact face location, and the green rectangle is the boundary box expanded to tile edges. This expansion is also applied to ground-truth images.

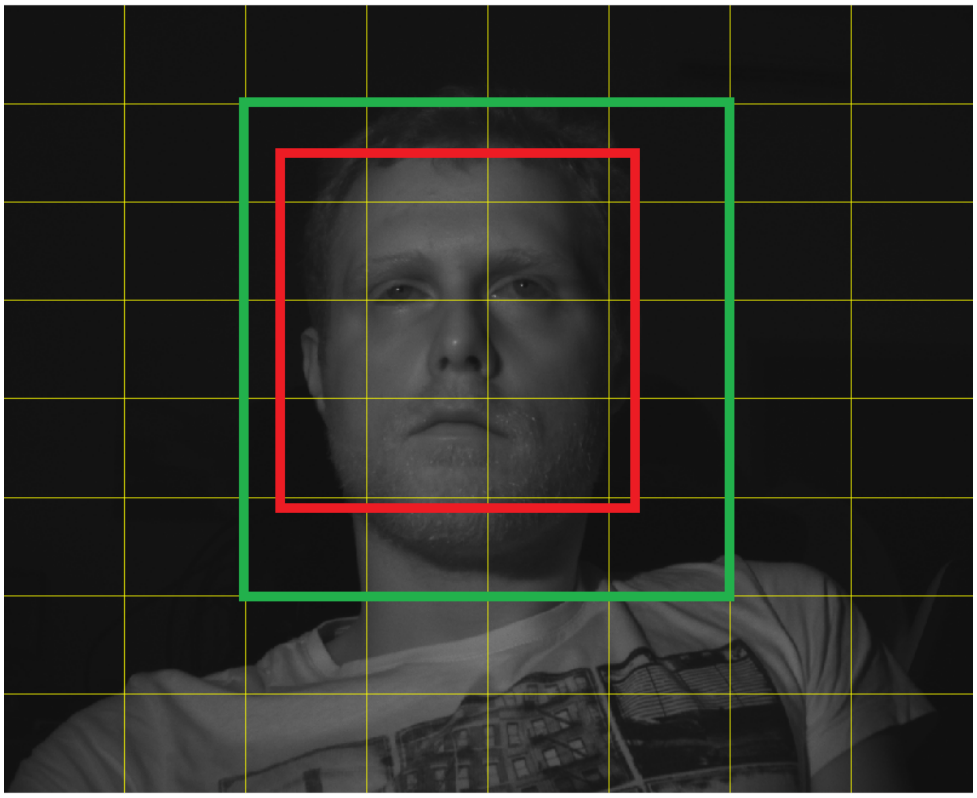

Figure 4.7: The face bounding box (red) expanded to the face region bounding box (green) over the 8x8 grid (yellow)

In this setting, the face regions are compared so the standard quality metrics could be calculated. The comparison is binary: the regions either overlap or not, and, consequently, there are four different outcomes of the comparison to the ground truth images: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). If the tiles marked as a part of the face region are overlapping for the ground truth and the algorithm, they are marked as a true positive. If the background tiles are overlapping for the ground truth and the algorithm, they are marked as a true negative. If the algorithm detects the face region outside the ground truth region, it is marked as a false positive, and if the algorithm does not detect the ground truth region, it is marked as a false negative. The example of the outcomes can be seen in Figure 4.8, in which the ground truth face region is marked in green, while the calculated face region is

marked in red.

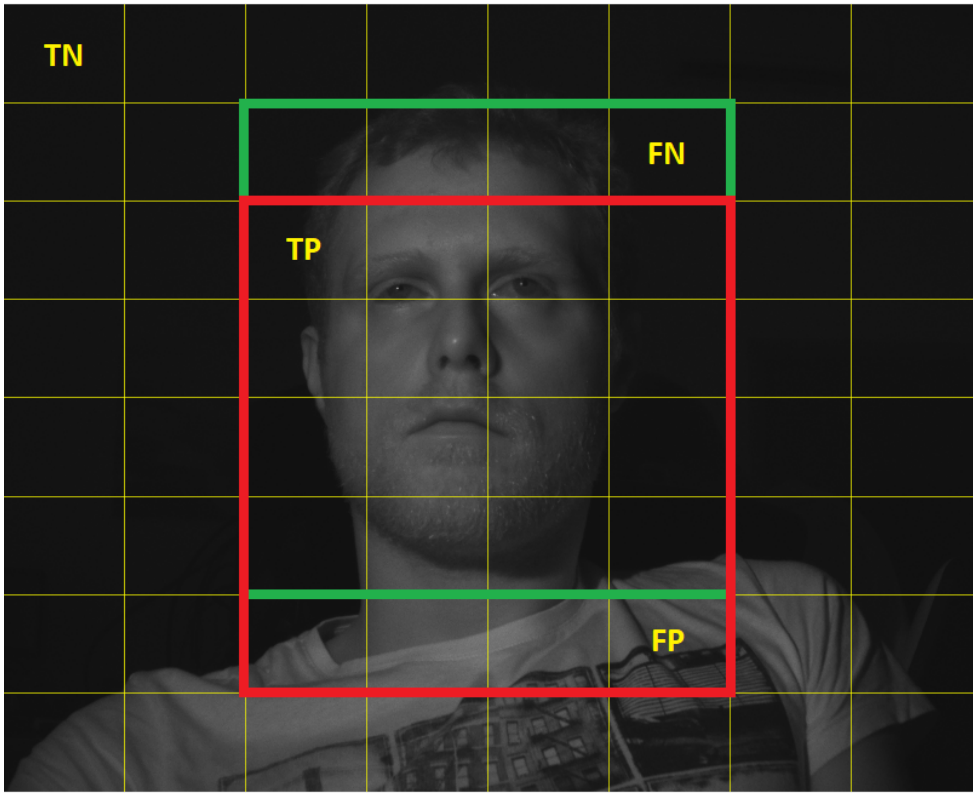

Figure 4.8: Overlapping ground truth face region (green) and calculated face region (red) with examples of true positive (TP), true negative (TN), false positive (FP), and false negative (FN)

Standard quality metrics are used for evaluating the binary classification tasks: accuracy, precision, recall, and F1-score. Accuracy shows the percentage of correctly identified tiles (true positive and true negative) among all tiles. Precision calculates the percentage of correctly identified positive tiles (true positive) among all tiles identified as positive (true positive and false positive). Recall or sensitivity is a measure in which the percentage of correctly identified positive tiles (true positive) among all real positive tiles (true positive and false negative) is calculated. F1-score is the harmonic mean between precision and recall. The formulas for calculating accuracy, precision, recall, and F1-score are shown in Equations 4.1, 4.2, 4.3, and 4.4.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.1}$$

$$Precision = \frac{TP}{TP+FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP+FN} \tag{4.3}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.4}$$

Regarding the number of the proposed method subvariants, each of the four variants (A, B, C, and D) was calculated with and without expansion, and with and without the temporal step. The temporal step itself can take four different forms: hysteresis only, boundaries time series median filtering only, hysteresis first and then boundaries time series median filtering, and vice-versa. If the window sizes for time series median filtering are chosen from 3, 5, 11, 15, and 29, there is a total of 136 tested subvariants of the spatio-temporal algorithm for approximate face region detection. In Table 4.1, accuracy, precision, recall, and F1-score are shown for the Viola-Jones algorithm and the five best-performing variants in the daytime and nighttime and in the overall dataset.

Table 4.1: Face region detection accuracy, precision, recall, and F1-score comparison between Viola-Jones face detection and the five best-performing proposed method variants in the daytime and nighttime and in total

	Accuracy (%)			Precision (%)			Recall (%)			F1-Score (%)		
Algorithm	Day	Night	Total	Day	Night	Total	Day	Night	Total	Day	Night	Total
Viola-Jones	89.02	89.13	89.07	55.87	59.29	57.37	89.89	90.05	89.96	66.89	69.96	68.24
A_Expn1_T00_HO	81.27	89.68	84.95	44.29	67.81	54.57	88.91	84.40	86.94	57.95	71.65	63.94
A_Expn1_T03_HT	81.09	88.96	84.53	43.96	66.75	53.92	89.30	84.82	87.34	57.80	70.80	63.48
A_Expn1_T03_TH	81.09	88.96	84.53	43.96	66.75	53.92	89.30	84.82	87.34	57.80	70.80	63.48
A_Expn1_T05_HT	80.93	88.84	84.39	43.81	65.75	53.40	89.72	85.30	87.79	57.75	70.65	63.39
A_Expn1_T05_TH	80.93	88.84	84.39	43.81	65.75	53.40	89.72	85.30	87.79	57.75	70.65	63.39

The variant name consists of one of the four variants (A, B, C, or D). The variant is followed by an expansion option separated with an underscore: *Expn0* means an expansion is not present, while *Expn1* means that the calculated region is expanded in every direction by 1. After another underscore, the number with the prefix *T* annotates the window size in time series median filtering. If the number is *00*, there is no time series median filtering. The last part consists of two letters and four possible combinations. *HO* means that only hysteresis is performed, while *TO* means only time series median filtering is performed. If the letter sequence is *HT*, hysteresis is performed before time series median filtering; if it is *TH*, the operation order is switched. The example *A_Expn1_T03_TH* means that variant A is used with region expansion *Expn1* in which time series median filtering is calculated first, with the window size 3 *T03*, followed by hysteresis *TH*.

The image consists of background tiles and face region tiles. On average, there are more background than face region tiles, which makes the overall data unbalanced. In lots of cases, most of the background tiles will be correctly detected (true negatives), but face region tile detection is more important. With a high number of true negatives, a reduced number of true

positives will not affect the accuracy calculation significantly. For this reason, accuracy is not recommended for unbalanced data, because it can lead to a wrong conclusion. Instead of using accuracy for an unbalanced dataset, precision and recall are used. F1-score is a combination of precision and recall, and it is chosen as the main quality measure for this evaluation. The subvariant with the best performance results is variant *A_Expn*1*_T00_HO*: variant A with the region expanded by one and hysteresis only as a temporal step, but the Viola-Jones has a better overall performance in F1-score. In Figure 4.9, a successful face region detection is presented, with the ground truth bounding box marked in yellow, overlaps with the proposed variant *A_Expn*1*_T00_HO* bounding box marked in green, and Viola-Jones bounding box marked in red. In the next subsection, the execution time of the proposed algorithm is evaluated.

Figure 4.9: Successful detection example with ground truth face region bounding box (yellow), proposed variant *A_Expn*1*_T00_HO* face region bounding box (green), and Viola-Jones face region bounding box (red)

4.3.2 Execution time of the proposed method

In this subsection, the execution time of the proposed algorithm on real-time embedded systems is tested and compared to the Viola-Jones algorithm. The execution time evaluation is performed on Zynq UltraScale+ MPSoC ZCU104 Evaluation Kit [96]. The images used for measuring execution time were from the FADID dataset [84], from the *test_ej_1* folder. Measurements and comparisons were conducted between the Viola-Jones algorithm and the worst-case scenario for every variant regarding time consumption, which includes region expansion and both hysteresis and time series median filtering calculation, with window size 29 for every variant. The average execution times in milliseconds are shown in Table 4.2. The Viola-Jones

is implemented from the open-source computer vision library - OpenCV [97].

Table 4.2: Average execution time comparison on the Xilinx ZCU104 between the Viola-Jones algorithm with a scale factor 2.5 and the worst-case scenario for every variant

	Average execution time (milliseconds)
Viola-Jones - scale factor 2.5	41.65
Worst case scenario (variant A)	1.62
Worst case scenario (variant B)	1.62
Worst case scenario (variant C)	1.65
Worst case scenario (variant D)	1.62

The scale factor is a parameter that specifies how much the image size is reduced at each image scale. A larger face image is iteratively resized to a smaller one, which is then sent to the Viola-Jones face detection algorithm. A higher scale factor results in faster execution [98], but with a higher chance of the face not being detected, and vice versa. The most common scale factor value used for the Viola-Jones algorithm for the best execution time/precision ratio is 1.1 [99], while the maximum scale factor value does not exceed 2 [98] in most cases. The scale factor was set to 2.5, which is an exaggerated value used in order to emphasize how the Viola-Jones algorithm is slow in comparison to the proposed method, even with an extreme value such as 2.5.

4.4 Discussion

Regarding detection performance evaluation, it can be seen from Table 4.1 that the Viola-Jones algorithm outperforms all variants in face region detection, but not at a significant rate. The biggest difference is seen in daytime recordings where the Viola-Jones performs better, but some variants outperform the Viola-Jones algorithm in nighttime. The best results are obtained with variant A with expansion using only hysteresis, combined hysteresis and time series median filtering, with window sizes 3 and 5. In this case, the results were the same for the subvariants with both hysteresis and time series median filtering, regardless of the order of execution. There are several cases in which the proposed algorithm does not perform well, as seen in Figure 4.10, where the ground truth rectangle is yellow, and the proposed variant rectangle is green. However, in these images, the Viola-Jones algorithm did not even recognize the face.

The biggest obstacle for face region detection performance is the color of clothing or the reflectance of clothes. However, the region does not shift often and it includes the face, which is beneficial for automatic brightness control systems where region position stability is as important as face region detection. It can also be seen that the proposed variant works well with

Figure 4.10: Unsuccessful detection example with ground truth face region bounding box (yellow) and proposed variant *A_Expn*1*_T00_HO* face region bounding box (green), with the Viola-Jones not detecting faces

different face poses, while different poses affect the Viola-Jones algorithm results significantly. Frontal faces with occlusions, such as glasses, are well detected by both the Viola-Jones and the proposed method. The FADID dataset lacks images with extreme lighting conditions and significant face occlusions, such as a hand covering the face, so no conclusions can be drawn for face region detection in this type of scene.

Regarding execution time evaluation, the data shown in Table 4.2 clearly indicates that the Viola-Jones algorithm is much slower than all proposed variants, and confirms statements in [100], [77], and [78] that the Viola-Jones algorithm is not able to perform over 20 FPS on embedded systems, which is not appropriate for a real-time embedded system task. The calculation time did not include image reading, mean pixel value calculation, and conversion to a grayscale image, as those tasks are often part of the ISP. While the mean pixel value execution is fast, the image loading/capturing can consume a lot of time if not implemented properly.

There are several potential improvements that can be tested and implemented in future work. The performance drop is seen in the daytime when the background can also be bright in some parts, which confuses the algorithm and mistakes those parts for a foreground. The reflectance of clothes is also a factor that can deteriorate approximate face region detection. This could be solved with the two-stage Otsu's multithreshold method [101]. To improve the behavior of the face region detection algorithm, the voting system between the variants of the algorithm can be implemented to make it more robust. If the detected regions exist in more than one variant, it is considered part of the face region. The second approach is the width and height face region ratio in which not all vertical rectangles are candidates for the face region, but only the ones that satisfy the criteria of the expected face region ratios between width and height. While this also reduces execution time, it can interfere with region expansion in case of a subject wearing bright clothes. Face tracking algorithms are often used in similar solutions. While the accuracy would be increased by using additional face tracking methods, the complexity of the algorithm would also increase, leading to a rise in hardware demands and execution time. Moreover, the temporal step of approximate face region detection behaves similarly to a face tracking algorithm by including the previous region values in the calculation of the current boundaries.

Another solution for performance increase is a convolutional neural network. Although it is stated in [69] that convolutional neural networks are not practical for real-time face detection, a lightweight and optimized neural network can be used for fast face detection. However, in that case, a large dataset is needed for training the classifier. The FADID dataset, as stated before, is not big enough and does not include difficult and extreme cases. Creating a new dataset or annotating the existing ones is an option, such as the MERL-Rice Near-Infrared Pulse (MR-NIRP) indoor and driving video dataset from [25], but that includes implementing the system in a vehicle to obtain fast light changes which can be achieved by driving towards, from, and by sunlight, through tunnels during the day, along tree lines, and towards car lights at night. Moreover, the process of capturing and/or annotating images is very time-consuming, and it is very complex to automate the face detection task; that can have a significant impact on the research conduction timeline. The temporal step of the approximate face detection algorithm can also be used as a separate algorithm and implemented along other face detection algorithms to make them more resilient to sudden region changes.

Although the face region detection variants outperform it in some segments, the Viola-Jones algorithm, as the most common embedded face detection algorithm, produces better results overall. This is to be expected because the Viola-Jones algorithm is face-oriented, whereas approximate face region detection is region-oriented. While this degrades detection quality, execution time is greatly improved, which is the main asset of the algorithm. Also, the region detected with the proposed method changes less frequently than the region from the Viola-Jones algorithm, which is beneficial for automatic brightness control systems. Moreover, the proposed method does not have face region detection problems with different face poses and glasses. The execution time of the algorithm allows it to perform well on real-time embedded systems. The best result is acquired when the temporal step is added to variant A with only hysteresis active, and time series median filtering can be omitted to further reduce execution time.

In the next chapter, the complete automatic brightness control algorithm with face region detection is presented.

Chapter 5

Automatic brightness control based on split-range feedback with face region detection

Although a great emphasis is put on approximate face region detection, it is just a part of a more complex automatic brightness control. Many methods for automatic brightness control have been described in Chapter 3: Related work chapter with different approaches, different controllable parameters, and different performance. As seen from related work, closed loop control with variable step change, PID, is the best solution for fast, precise, and robust automatic brightness control on real-time embedded systems. As stated before, the image acquisition is a non-linear MISO system. There are four inputs on the image acquisition system: analog gain, digital gain, exposure time from the camera, and external NIR illumination source controlled by PWM. This results in having four control signals: u_{PWM} is the infrared illumination source control signal, u_{AG} is the analog gain control signal, u_{DG} is the digital gain control signal, and u_{ET} is the exposure time control signal. In classic implementation, four different PID controllers would have to be built, which means tuning twelve gains, K_p, K_i, and K_d for each PID. The inputs are also coupled and they interfere with each other, which makes the controller design and tuning very challenging. Moreover, changing the inputs one by one and includes developing a proper strategy regarding the order and conditions of adapting the parameters, which increases the complexity of the control system. A simpler and faster method, instead of having four PIDs with four different control signals, would be having a single PID with a common control variable, which is the main premise of the split-range feedback controller. For this reason, automatic brightness control with the simultaneous control of camera parameters and external illumination source is presented with the introduction of approximate face region detection, which not only reduces the number of tuning parameters and the complexity of the system, but also results in faster convergence. The proposed approach consists of the following

parts: the control system for image acquisition, approximate face region detection, and the face absence handle. Split-range feedback controller is described in detail in continuation, while face region detection has been described in the previous chapter, so it will not be further elaborated. It appears only at the end of the description, where the complete method is shown. Finally, the face absence handle is a part of the automatic brightness control algorithm which deals with situations where the face disappears from a scene. After the detailed explanation of the complete automatic brightness control, the algorithm is implemented and evaluated on a real-time embedded system. Finally, a discussion will conclude this chapter. The detailed explanation of split-range feedback control with approximate face region detection and the face absence handle is presented in the next section.

5.1 Split-range feedback control with approximate face region detection and the face absence handle

As stated before, the design and tuning of a non-linear MISO system can be simplified by using a single PID with just one common control variable. As seen in Figure 2.7, control variable u is an output signal of the controller and an input to the process, which is, in this case, the image acquisition system. The control variable u affects the process and its output y with the system delay d. Then, the process output y is compared to the reference value r, and the difference is called error term e, which is also the input to the controller. The anti-windup is calculated, together with PID calculation, by clamping, to prevent the integral sum to accumulate if the controller output is saturated, as described Chapter 2: Theoretical foundations. The four control signals have similar, sigmoid-like [8] static characteristics, so, instead of having four different controllers for analog and digital gain, exposure time, and external illumination, only one controller can be used, just with a different weight factor for each control variable – this is the working principle of the split-range controller. As seen in Figure 5.1, the control variable u calculated from the PID controller is propagated to four control signals: u_{PWM}, u_{AG}, u_{DG}, and u_{ET}, which are calculated by multiplying the control variable u with different gains: K_{PWM}, K_{AG}, K_{DG}, and K_{ET}, respectively, which weigh the impact of each of the camera's parameters and illumination source on the image acquisition system. The control variable u is also called joint control variable.

The usage of split-range feedback controller results in having eight tunable parameters instead of twelve: gains K_p, K_i, and K_d and sample time T_s for the PID controller, and gains for the control signals K_{PWM}, K_{AG}, K_{DG}, and K_{ET}. The strategy for tuning control signal gains needs to be carefully planned. Produced infrared illumination needs to be dominant over ambient light, so the emphasis is put on K_{PWM}, while K_{AG} and K_{DG} need to be at lower values so as not to cause noise amplification. K_{ET} needs to be tuned in such a way that the ambient light

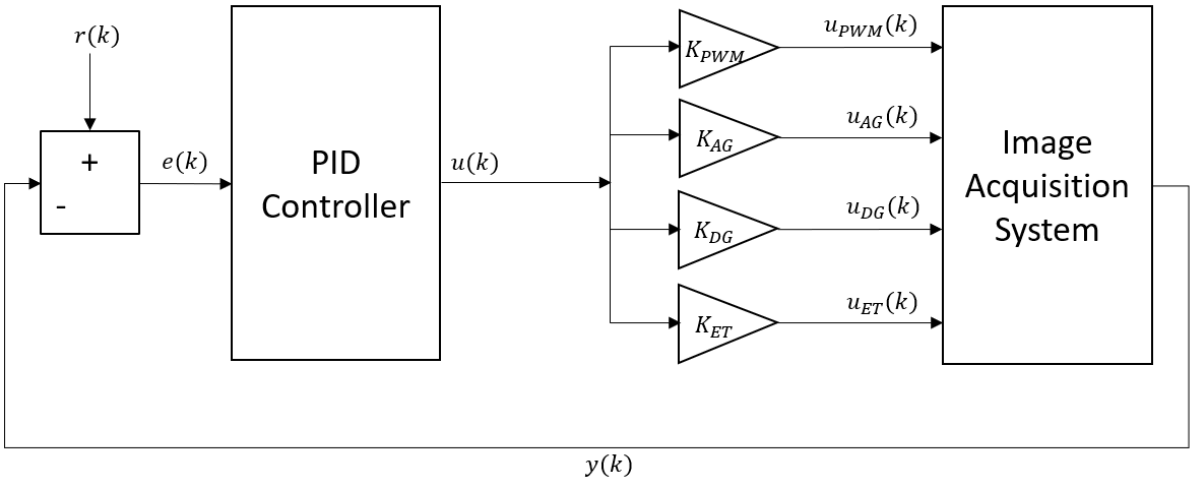

Figure 5.1: Simple split-range PID controller block scheme

is not very expressed, while preserving the overall brightness of the region of interest. If the calculated control signal values u_{PWM}, u_{AG}, u_{DG}, and u_{ET} are out of the actuator's proposed range, they are clipped and limited to the minimum or maximum value of the actuator's range. If the face region is detected successfully, a new iteration of the algorithm is executed until the face region detection reaches the wanted brightness. However, if the face region is not detected because of various reasons, such as camera covering or the face moving out of the scene boundaries, the background is put in the focus. In this case, the brightness of the background changes slowly, and the mean pixel value is reached barely or not at all, while the actuators are at their physical limits constantly, which wears them out faster. Moreover, if the face reappears suddenly, oscillations can occur while the system tries to compensate for face brightness. For this reason, the face absence handle is implemented to prevent such unwanted behavior of the control system. The difference between the brightest and darkest mean pixel value tiles d lower than the threshold value $ValueTH$ indicates that the difference between classes is too small to properly identify the background and the foreground. The joint control value u is set to a fixed lower value. That way, a lower mean pixel value is achieved, and the sudden potential ambient light change is compensated for, while being high enough to detect the reappeared face region. The joint control variable u remains at a low value until the face reappears, which increases the difference d and satisfies the conditions, resulting in the continuation of automatic brightness control in the common mode. In Figure 5.2, the workflow of the split-range PID controller with the face absence handle is presented. Finally, in Figure 5.3, the complete automatic brightness control algorithm with the control element is shown, including the split-range PID controller, clamping, and clipping marked in orange, and face region detection and the face absence handle marked in blue, while in Algorithm 5.1, pseudocode of the complete algorithm is presented. The evaluation of the complete automatic brightness control with face region detection on real-time embedded systems is presented in the next section.

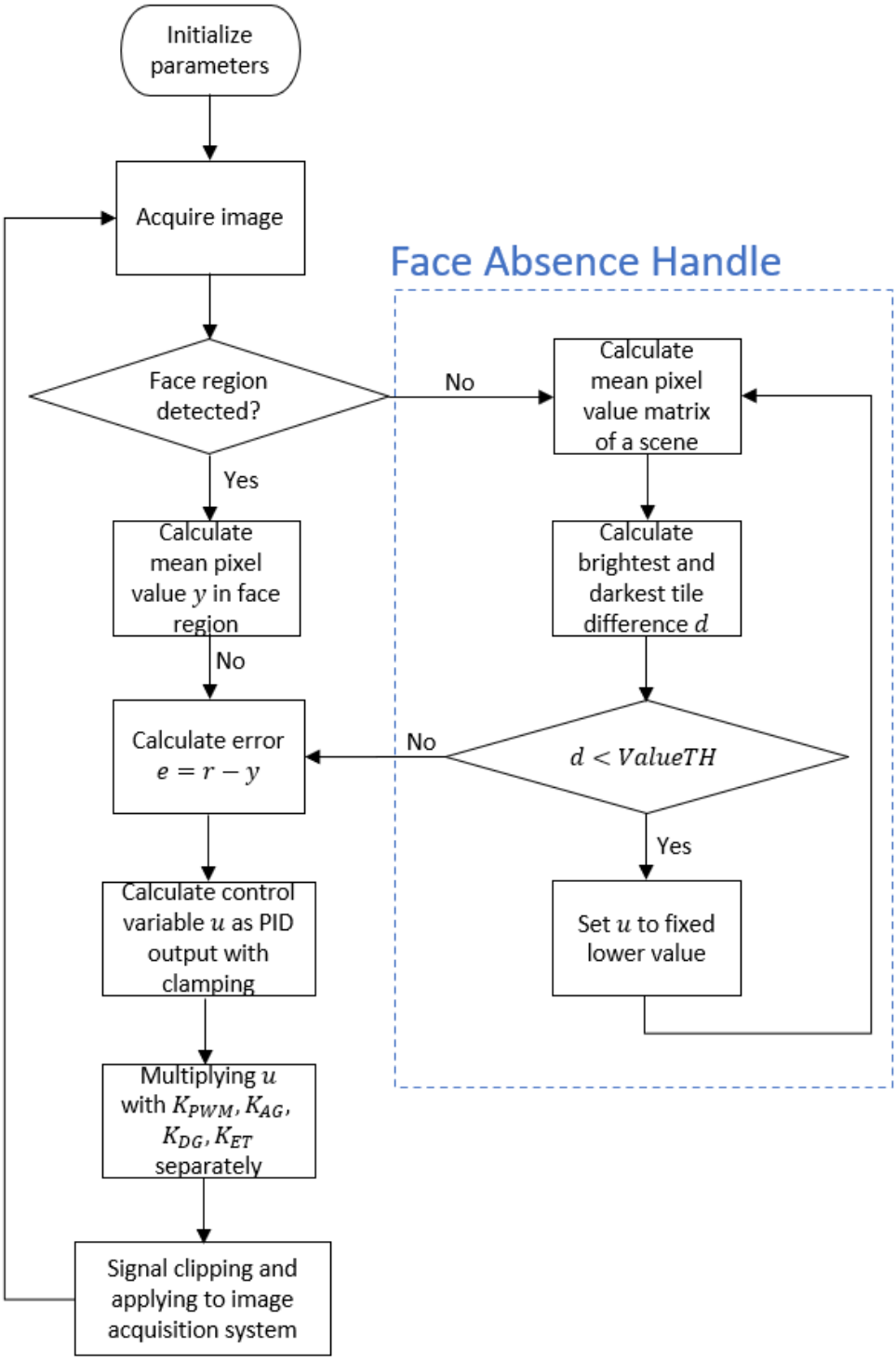

Figure 5.2: Workflow of the split-range PID controller with the face absence handle

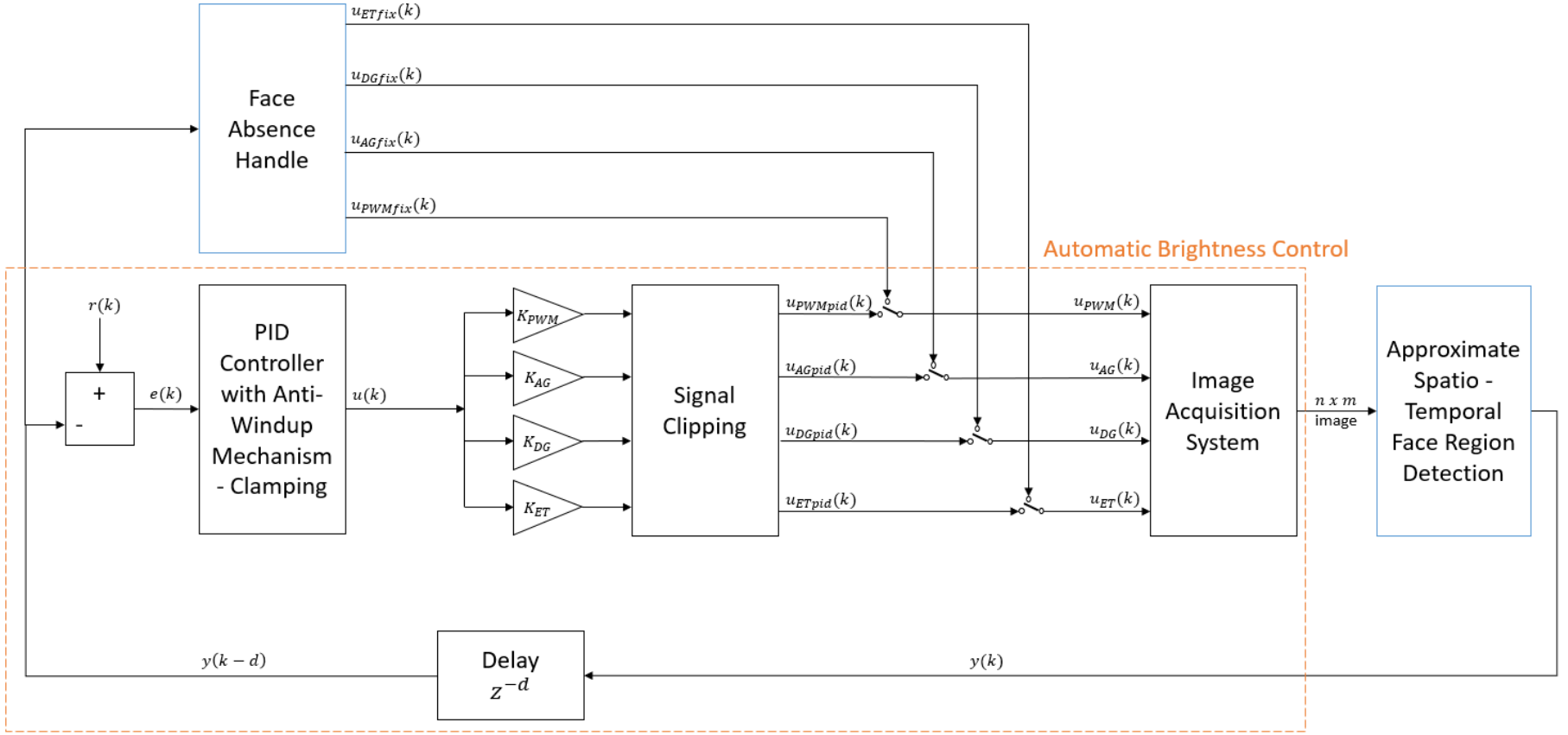

Figure 5.3: The complete automatic brightness control block scheme, with the split-range PID controller, anti-windup and clipping mechanism in orange, and face region detection and the face absence handle in blue

Algorithm 5.1 Automatic Brightness Control with the Split-Range Controller and Real-Time Spatio-Temporal Approximate Face Region Detection With Region Expansion, Time Series Median Filtering first and Hysteresis Calculation second

for Every time step **do**
  Acquire NIR image from the sensor
  Divide image in x-by-y
  Calculate the mean pixel value matrix of the scene
  Calculate brightest and darkest tile difference d
  if $d <$ Threshold value $ValueTH$ **then**
    Set joint control variable u to a fixed lower value
  else
    Clip edge tiles
    Calculate binarization mask with Otsu's method
    Choose the algorithm variant
    if Variant A **then**
      Find the biggest vertical rectangle or square of positive values
    else if Variant B **then**
      Find the biggest vertical rectangle or square positive values with $\geq 75\%$ of positive values
    else if Variant C **then**
      Change negative values to positive if surrounded by 50% of positive values
      Find the biggest vertical rectangle or square of positive values
    else if Variant D **then**
      Find the biggest vertical rectangle or square of positive values
      Expand to horizontal neighboring columns if $\geq 50\%$ of all values are positive
    end if
    Calculate the region boundaries $xBoundary$, x is $left$, $right$, up and $down$
    If possible, expand the region in every direction:
    if x in $xBoundary$ == $left$ or up **then**
      Decrease $xBoundary$ by 1
    else
      Increase $xBoundary$ by 1
    end if
    Calculate the median $xBoundaryM$ for every boundary of last $k-1$ previous boundary values
    Calculate hysteresis for every boundary
    if $xBoundary$ expands the face region by 1 **then**
      Keep the previous value of the boundary
    end if
    Calculate mean pixel value y in the region between the calculated boundaries
    Calculate error term $e = r - y$
    Calculate and output of the PID controller u
    Multiply u with K_{PWM}, K_{AG}, K_{DG} and K_{ET}, respectively and clip out-of-range signals
  end if
  Propagate the signals to the image acquisition system
end for

5.2 Experimental Results

After the description of the proposed method, performance evaluation of automatic brightness control with the split-range controller algorithm with spatio-temporal approximate face region detection is executed on an embedded system. The main goal of the evaluation is to examine convergence speed and the presence of oscillations in the mean pixel value signal while achieving a proper face luminance. The evaluation was conducted in a laboratory, where

real-world scenarios were simulated. The image acquisition system consisted of Xilinx Zynq UltraScale+ MPSoC ZCU104 Evaluation Kit, OmniVision OV2311 image sensor with a cut-off 940 nanometers filter, and two illumination sources with infrared light-emitting diodes (IR LED) with a wavelength of 940 nanometers. The setup can be seen in Figure 5.4.

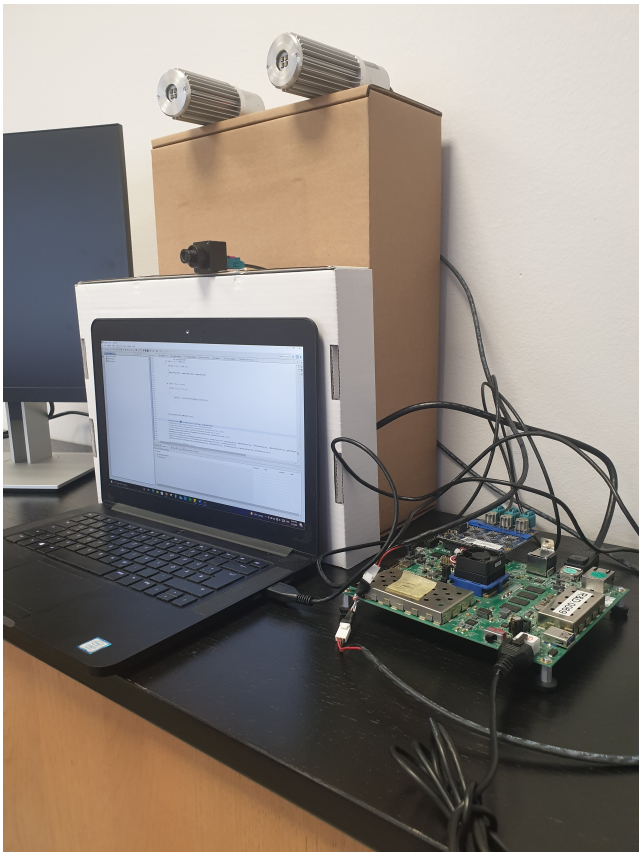

Figure 5.4: The image acquisition system setup including Xilinx ZCU104 board, OmniVision OV2311 image sensor, and two infrared illumination sources

As described in the evaluation part in the previous chapter, the framerate of the system is 30 FPS, and the ISP divides the frame into 8x8 equal tiles. Another major system constraint is system delay, which affects the brightness oscillations when a sudden illumination peak occurs. As stated before, there are eight different, tunable parameters (seven if sampling time is incorporated in integral and derivative gain). Proportional and integral gains of the PID controller are tuned in accordance with the recommendation in [9], where K_i is set relatively high to K_p, so the mean pixel value gradually reaches the target value in a system with frequent changes, while the derivative gain K_d is tuned heuristically. Regarding the control signal gains, most emphasis is put on the external illumination source. The exposure time control signal gain is lower, but still significant, and both digital and analog gain control signal gains are set to a low value. In laboratory conditions, there are no natural sources of NIR brightness disturbance, so they need to be produced artificially. The disturbance caused by ambient light is simulated with a second infrared illumination source. In the first part of the evaluation, for various light

condition simulations, the disturbance is changed in a step and ramp manner to see how well the control algorithm will adjust, as seen in Figure 5.5.

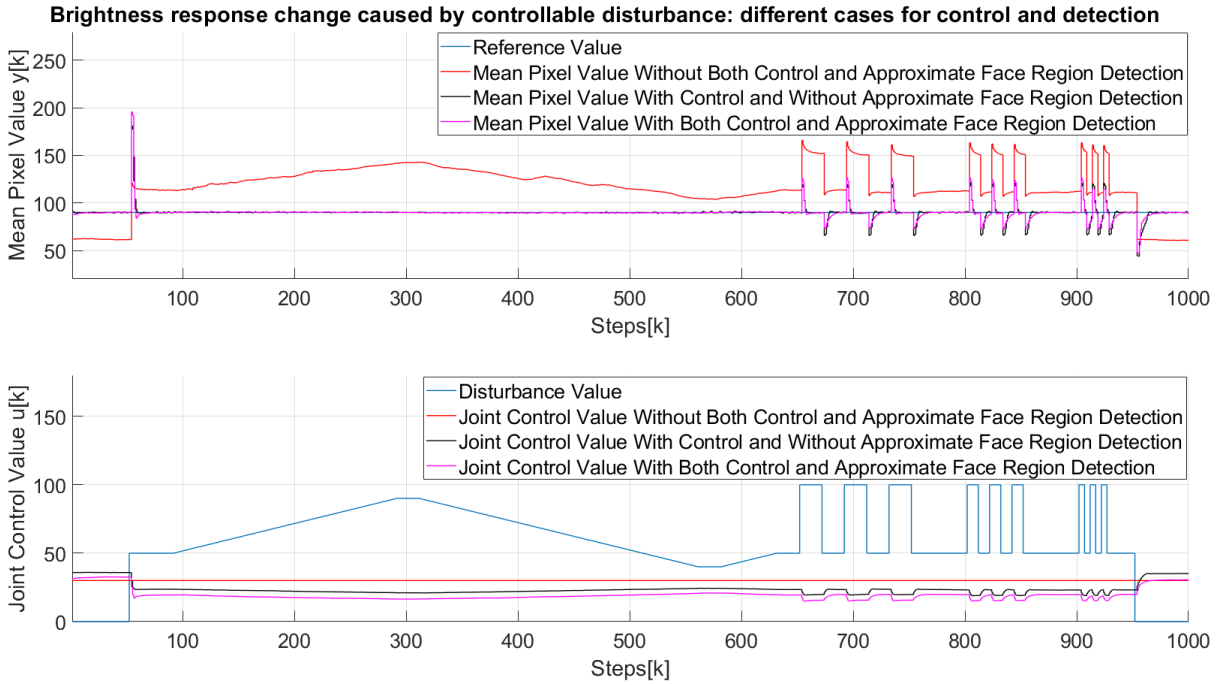

Figure 5.5: Brightness response change caused by a controllable disturbance with different control and region detection scenarios and with a still face

In the lower subplot, the disturbance value marked in blue is first changed in a step manner from 0% to 50%. Then, it rises in a ramp form to the maximum value, stays there for a short amount of time and starts to fall in ramp form. After a short period of constant value, it rises again from 40% to 50%. Subsequently, there are three periods with major oscillations in disturbance change, each with a different frequency to test the adaption of the algorithm to sudden changes. Finally, the disturbance value is reduced from 50% to the initial 0% in a step manner. The reference value is kept at a fixed level, as seen in the upper subplot. There are three different setups: the red line represents the setup with both the control and approximate face region detection disabled, the black line represents the setup with automatic brightness control enabled only, and the magenta line represents the setup with both the control and approximate face region detection enabled. The setups in which face region detection is disabled used a fixed region of interest with all the edge tiles excluded from the calculation, as seen in Figure 5.6, with the same reasoning as the proposed face region detection algorithm. In the upper subplot, the lines represent the mean pixel value, while in the lower subplot, the lines represent the joint control variable value. As expected, when both the automatic brightness control element and face region detection element are disabled, the mean pixel value changes according to the disturbance illumination changes. The convergence is similar for both setups where automatic brightness control is enabled, but the face region is, respectively, disabled and enabled. The face was present and not moving in the scene.

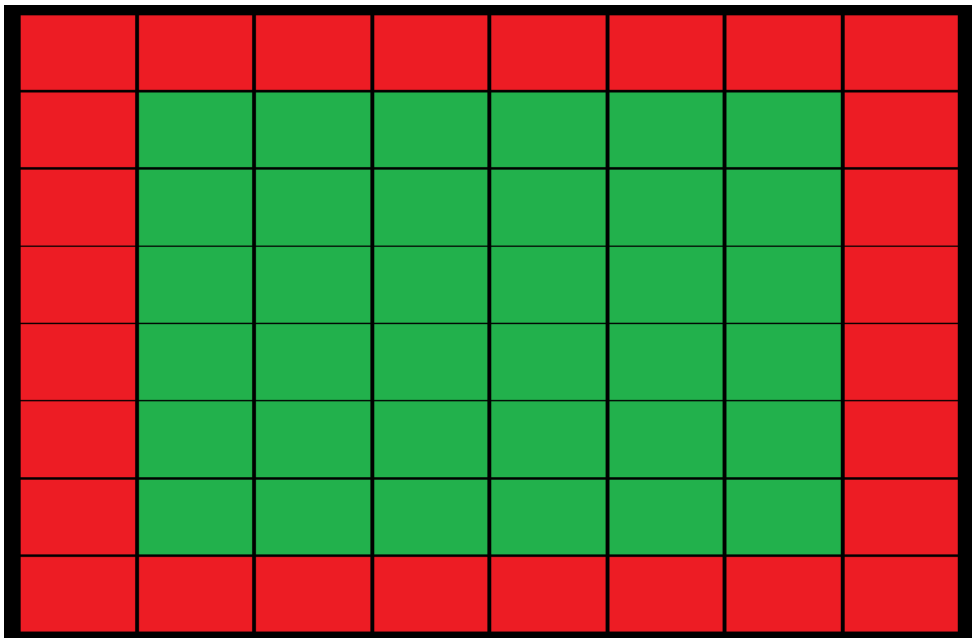

Figure 5.6: Fixed region of interest (green) in a scene used for setups with face region detection disabled, where outer tiles (red) are not included in the calculation

The second part of the evaluation consists of having a non-moving and moving face in a scene with the same disturbance and reference value as in the first part, as seen in Figure 5.7.

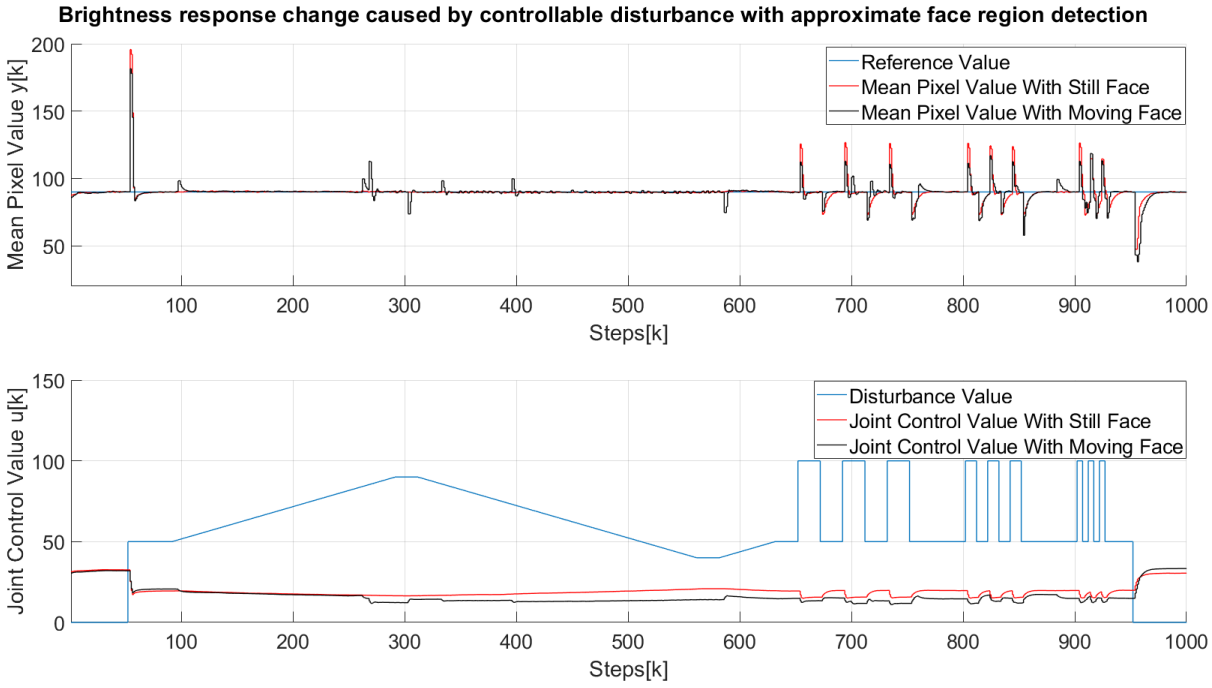

Figure 5.7: Brightness response change caused by a controllable disturbance with control and face region detection with a still and moving face

The red line represents the still face test, while the black line represents the moving face test. As in the first part, in the upper subplot, the lines represent the mean pixel value, while in the lower subplot, the lines represent the joint control variable value. Both tests for automatic

brightness control yield similar results in terms of convergence speed, and, although smaller peaks and valleys are visible because of the region boundaries change, the algorithm performs well in both cases. If the graph is slightly enlarged in the upper subplot, as seen in Figure 5.8, it is visible that the reference value is reached in 10–15 frames, which translates to 330–495 milliseconds in a 30 FPS system.

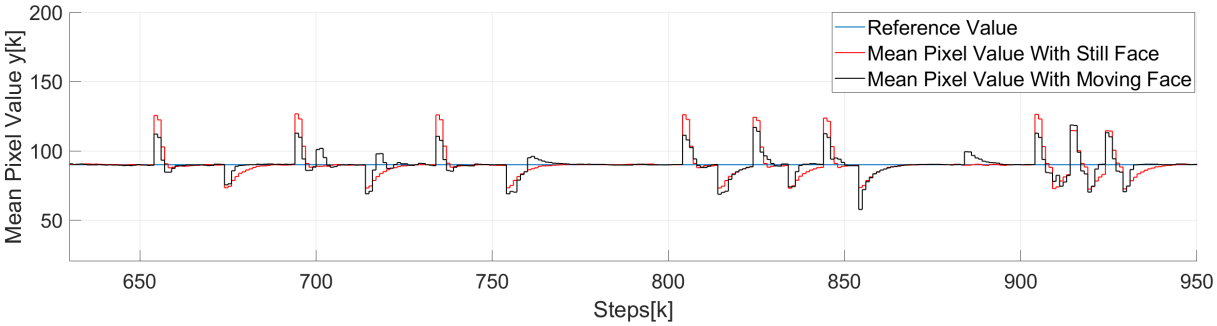

Figure 5.8: Enlarged brightness response change caused by a controllable disturbance with control and region detection with a still and moving face

The final part of the evaluation consists of testing the face absence handle, as seen in Figure 5.9.

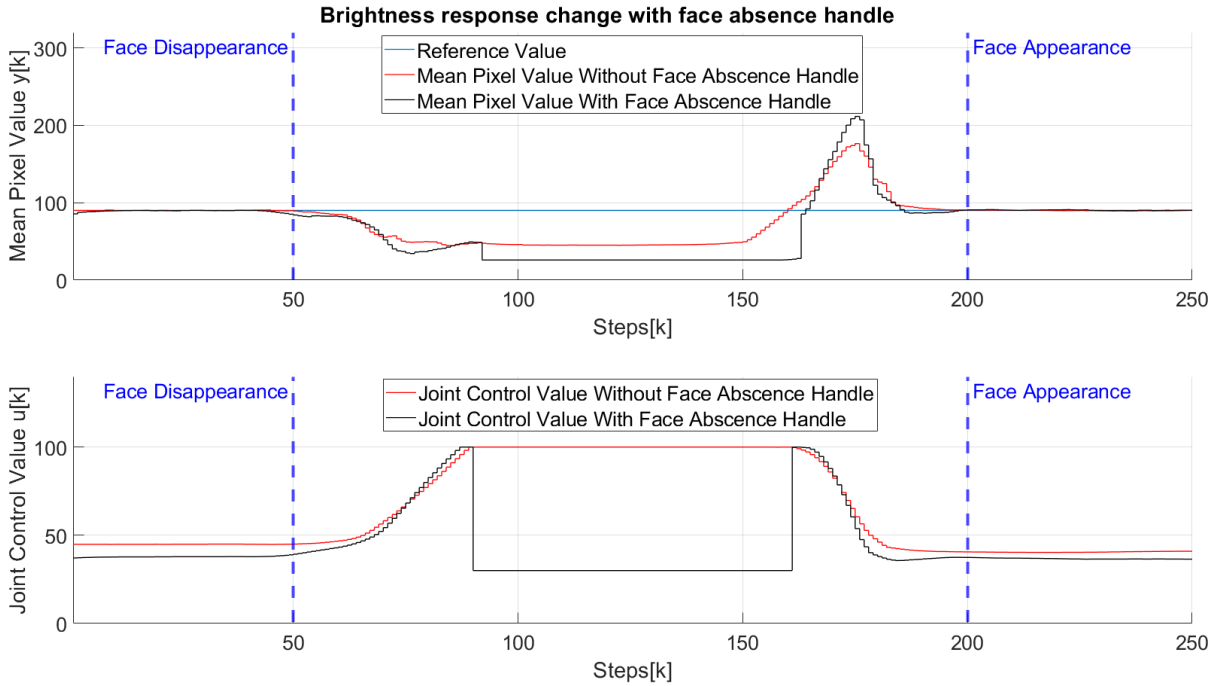

Figure 5.9: Face absence brightness response behavior without (red) and with (black) the face absence handle

In the upper subplot, the reference value (blue) and the brightness response change when the face absence module is disabled (red) and enabled (black) are shown. The reference value is constant, while the face first disappears from and later reappears in the scene, as marked with dark blue vertical lines. When the face is still present, automatic brightness control corrects

the brightness normally; when it disappears, the system without the face absence handle is struggling to reach the reference value, but without success, while all the control signals are at the maximum value. The face absence handle locks the joint control variable u to a lower value, but high enough to recognize the face region when the face reappears, and the mean pixel value enlarges. Finally, brightness control continues normally after the face reappears. Regarding the execution time of automatic brightness control with approximate face region detection, the average execution time of one step of approximate face region detection, automatic brightness control, and changing parameters is 9.13 milliseconds.

5.3 Discussion

In the first part, three different setups are tested: both control algorithm and approximate face region detection disabled, control algorithm enabled only, and both control algorithm and approximate face region detection enabled. It is shown that both algorithms that include control track the reference value well, while the joint control value is smaller in the case with approximate face region detection enabled. If there is no control algorithm, the main pixel value changes significantly and the face may not be illuminated properly.

As seen from Figures 5.5, 5.7, and 5.8 in the first and second part of the evaluation, brightness response change behaves similarly with a fixed region of interest and with face region detection when the face is present. The biggest difference occurs when the face goes outside of the scene or region of interest. In a control system with a fixed region, the joint control value reaches its peak without getting close to the mean pixel value which wears out the equipment faster. As seen in Figure 5.8, if approximate face region detection and the face absence handle are employed, the joint control value is set to a low value after the face disappears, but high enough to recognize the face region when the face reappears, which spares the equipment and prevents the brightness saturation when the face reappears. This is also seen in the lower subplots in Figures 5.5 and 5.7, where joint control variable u is lower when approximate face region detection is enabled. Regarding automatic brightness control, it is visible that the optimal brightness control fastly converges to the target value, which is an improvement in comparison with [95], [29], and [94]. The average execution time of the complete algorithm is under 10 milliseconds, which makes it convenient to implement on even higher frame rate systems than 30 FPS. Through several iterations of parameter tuning, it is concluded that the derivative term does not improve the control system response, so it can be omitted, and analog gain should be held fixed because it often causes oscillations around the reference value, which results in having a total of five tunable parameters. Regarding the values of those five parameters, for every camera system and external illumination source, they need to be tuned according to the sensor's response and power and wavelength of the illumination source individually, but the principle

is the same as stated before. Regarding the face absence module, it does a very good job of freezing the inputs in order to maintain the brightness. However, the control loop is broken at that moment, which is not very desirable. There is an alternative way to deal with this problem - by simultaneously monitoring the mean pixel values of both the foreground and background. When the face is present, the control algorithm performs normally by following the foreground reference value. However, when the face disappears from the scene, the control algorithm tries to maintain the background brightness as it was before the face disappearing, and a new reference value is followed until the face reappears. This approach will be tested in the next iteration. The future work mostly relies on the approximate face region detection element, which is discussed in the previous chapter. As the system tends to be robust, building and implementing a predictive disturbance model, or even a whole model of the image acquisition system and scene brightness change, would greatly improve the control system, but the disturbance model mostly depends on the type of system in which the proposed control algorithm is to be used. A major helpful addition would be creating a plug-and-play camera calibration and parameter tuning method based on an optimization algorithm for faster and easier commissioning of real-time embedded system optimization.

Chapter 6

Conclusion

Automatic brightness control with approximate face region detection is presented in this thesis, which is a complex system that stretches over image and signal processing, control systems, and embedded system optimization. The first chapter introduces automatic brightness as a pre-processing task for facial feature analysis, including approximate face region detection to calculate face brightness only in the face region. The importance of using both automatic brightness control and face region detection is highlighted, alongside the thesis motivation with an emphasis on speed, robustness, and computational efficiency. Both problems are dynamic tasks executed on embedded systems, and hard real-time performance is the main prerequisite. The theoretical foundation of the PID controllers for the control system solution and foreground/background detection for the face region detection solution are presented in the second chapter. The third chapter provides an overview of various methods for automatic brightness control and face detection in the NIR spectrum, which is used as a scientific basis for further decision making in the rest of the thesis.

Region detection method based on spatio-temporal sampled skin detection is introduced in chapter four. The region-based foreground/background segmentation is executed for face region boundaries valuation, with previous boundary values included in the calculation. In the fifth chapter, the complete automatic brightness control based on a split-range PID controller with approximate face region detection is described.

This thesis presents the automatic brightness control with approximate face region detection algorithm. The current state-of-the-art automatic brightness control algorithms with face detection often struggle with performance and/or execution time. The proposed algorithm is robust and resourceful, and the execution time under 10 milliseconds makes it faster than the state-of-the-art solution and convenient for real-time embedded systems. Although there is room for improvement, the current version of the algorithm performs satisfactorily for the pre-processing task, as it is an easily implementable and resourceful solution best suited for a real-world application with hardware and space limitations, for instance in the automotive industry.

Bibliography

[1]Bruker, "Why FT-NIR spectroscopy?", available at: https://www.bruker.com/en/products-and-solutions/infrared-and-raman/ft-nir-spectrometers/what-is-ft-nir-spectroscopy.html (July 29, 2022.).

[2]Montalvo, M., Guerrero, J. M., Romeo, J., Guijarro, M., Jesús, M., Pajares, G., "Acquisition of agronomic images with sufficient quality by automatic exposure time control and histogram matching", in International Conference on Advanced Concepts for Intelligent Vision Systems. Springer, 2013, pp. 37–48, doi: 10.1007/978-3-319-02895-8_64.

[3]Peng, Z., Zhao, B., "Novel scheme for infrared image enhancement based on contourlet transform", in Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Vol. 6. IEEE, 2011, pp. 3134–3137, doi: 10.1109/EMEIT.2011.6023751.

[4]Mendenhall, M. J., Nunez, A. S., Martin, R. K., "Human skin detection in the visible and near infrared", Applied optics, Vol. 54, No. 35, 2015, pp. 10 559–10 570, doi: 10.1364/AO.54.010559.

[5]Janani, V., Dinakaran, M., "Infrared image enhancement techniques—a review", in Second International Conference on Current Trends In Engineering and Technology-ICCTET 2014. IEEE, 2014, pp. 167–173, doi: 10.1109/ICCTET.2014.6966282.

[6]Kao, W.-C., Hsu, C.-C., Kao, C.-C., Chen, S.-H., "Adaptive exposure control and real-time image fusion for surveillance systems", in 2006 IEEE international symposium on circuits and systems. IEEE, 2006, pp. 4–pp, doi: 10.1109/ISCAS.2006.1692740.

[7]Westerhoff, J., Meuter, M., Kummert, A., "A generic parameter optimization workflow for camera control algorithms", in 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, 2015, pp. 944–949, doi: 10.1109/ITSC.2015.158.

[8]Shrestha, R., Mohammed, S. K., Hasan, M. M., Zhang, X., Wahid, K. A., "Automated adaptive brightness in wireless capsule endoscopy using image segmentation and sigmoid

function", IEEE transactions on biomedical circuits and systems, Vol. 10, No. 4, 2016, pp. 884–892, doi: 10.1109/TBCAS.2016.2546838.

[9]Sousa, R. M., Wäny, M., Santos, P., Morgado-Dias, F., Member, I., "Automatic illumination control for an endoscopy sensor", Microprocessors and Microsystems, Vol. 72, 2020, pp. 102920, doi: 10.1016/j.micpro.2019.102920.

[10]Laurie, J., Higgins, N., Peynot, T., Roberts, J., "Dedicated exposure control for remote photoplethysmography", IEEE Access, Vol. 8, 2020, pp. 116 642–116 652, doi: 10.1109/ACCESS.2020.3003548.

[11]Ali, S., Jonmohamadi, Y., Takeda, Y., Roberts, J., Crawford, R., Pandey, A. K., "Supervised scene illumination control in stereo arthroscopes for robot assisted minimally invasive surgery", IEEE Sensors Journal, 2020, doi: 10.1109/JSEN.2020.3037301.

[12]Kiran, B. R., Krishna, K. M., Basha, S. S., Vardhan, H., Raju, N., Kumari, V. V. *et al.*, "False position based auto exposure algorithm for properly exposing the leather samples in the leather industries", International Journal of Computer Applications, Vol. 83, No. 14, 2013, doi: 10.5120/14514-2895.

[13]Li, S., Handa, A., Zhang, Y., Calway, A., "HDRFusion: HDR SLAM using a low-cost auto-exposure RGB-D sensor", in 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016, pp. 314–322, doi: 10.1109/3DV.2016.40.

[14]Zhang, Z., Forster, C., Scaramuzza, D., "Active exposure control for robust visual odometry in HDR environments", in 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017, pp. 3894–3901, doi: 10.1109/ICRA.2017.7989449.

[15]Mehta, I., Tang, M., Barfoot, T. D., "Gradient-based auto-exposure control applied to a self-driving car", in 2020 17th Conference on Computer and Robot Vision (CRV). IEEE, 2020, pp. 166–173, doi: 10.1109/10.1109/CRV50864.2020.00030.

[16]Lu, H., Zhang, H., Yang, S., Zheng, Z., "Camera parameters auto-adjusting technique for robust robot vision", in 2010 IEEE International Conference on Robotics and Automation. IEEE, 2010, pp. 1518–1523, doi: 10.1109/ROBOT.2010.5509978.

[17]Shin, U., Park, J., Shim, G., Rameau, F., Kweon, I. S., "Camera exposure control for robust robot vision with noise-aware image quality assessment", arXiv preprint arXiv:1907.12646, 2019, doi: 10.1109/IROS40897.2019.8968590.

[18]Otsu, N., "A threshold selection method from gray-level histograms", IEEE transactions on systems, man, and cybernetics, Vol. 9, No. 1, 1979, pp. 62–66, doi: 10.1109/TSMC.1979.4310076.

[19]G-Energy, “Product line irbloc”, available at: https://www.g-energy.com.tw/product_irbloc.html (August 2, 2022.).

[20]Park, B., Keh, Y., Lee, D., Kim, Y., Kim, S., Sung, K., Lee, J., Jang, D., Yoon, Y., “Outdoor operation of structured light in mobile phone”, in Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops, Oct 2017, doi: 10.1109/ICCVW.2017.282.

[21]AXTON, “Infrared – 850nm vs 940nm”, available at: https://axtontech.com/infrared-850nm-vs-940nm-wavelength (August 2, 2022.).

[22]Magdalena Nowara, E., Marks, T. K., Mansour, H., Veeraraghavan, A., “SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared”, in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 1272–1281, doi: 10.1109/CVPRW.2018.00174.

[23]Renesas, “Eye safety for proximity sensing using infrared light-emitting diodes - application note”, available at: https://www.renesas.com/us/en/document/apn/an1737-eye-safety-proximity-sensing-using-infrared-light-emitting-diodes (August 2, 2022.).

[24]Maruthy, K., Padmavathi, R., Sowjanya, B., MaheshKumar, K. *et al.*, “Quantitative determination of pupil by dynamic pupillometry using infrared videography–role in evaluation of autonomic activity”, Clinical Epidemiology and Global Health, Vol. 8, No. 3, 2020, pp. 728–732, doi: 10.1016/j.cegh.2020.01.010.

[25]Nowara, E. M., Marks, T. K., Mansour, H., Veeraraghavan, A., “Near-infrared imaging photoplethysmography during driving”, IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 4, 2020, pp. 3589–3600, doi: 10.1109/TITS.2020.3038317.

[26]Vuong, Q. K., Yun, S.-H., Kim, S., “A new robust combined method for auto exposure and auto white-balance”, in Advances in Machine Learning and Data Analysis. Springer, 2010, pp. 165–178, doi: 10.1007/978-90-481-3177-8_11.

[27]Kehtarnavaz, N., Oh, H.-J., Shidate, I., Yoo, Y. F., Taluri, R., “New approach to auto-white-balancing and auto-exposure for digital still cameras”, in Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III, Vol. 4669. International Society for Optics and Photonics, 2002, pp. 268–276, doi: 10.1117/12.463431.

[28]Chen, W., Li, X., “Exposure evaluation method based on histogram statistics”, in 2017 2nd International Conference on Electrical, Automation and Mechanical Engineering (EAME 2017). Atlantis Press, 2017, pp. 290–293, doi: 10.1117/12.463431.

[29]Gnatyuk, V., Zavalishin, S., Petrova, X., Odinokikh, G., Fartukov, A., Danilevich, A., Eremeev, V., Yoo, J., Lee, K., Lee, H. *et al.*, "Fast automatic exposure adjustment method for iris recognition system", in 2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, 2019, pp. 1–6, doi: 10.1109/ECAI46879.2019.9042077.

[30]Ning, J., Lu, T., Liu, L., Guo, L., Jin, X., "The optimization and design of the auto-exposure algorithm based on image entropy", in 2015 8th International Congress on Image and Signal Processing (CISP). IEEE, 2015, pp. 1020–1025, doi: 10.1109/CISP.2015.7408029.

[31]Shim, I., Lee, J.-Y., Kweon, I. S., "Auto-adjusting camera exposure for outdoor robotics using gradient information", in 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2014, pp. 1011–1017, doi: 10.1109/IROS.2014.6942682.

[32]Qian, Y., Zhang, W., Liu, J., Chen, Q., Gu, G., "An auto-gain control algorithm for EMCCD based on dynamic gray-level", in Optoelectronic Imaging and Multimedia Technology III, Vol. 9273. International Society for Optics and Photonics, 2014, pp. 92732J, doi: 10.1117/12.2070216.

[33]Ziegler, J. G., Nichols, N. B. *et al.*, "Optimum settings for automatic controllers", trans. ASME, Vol. 64, No. 11, 1942, doi: 10.1115/1.2899060.

[34]Voudoukis, N., Oikonomidis, S., "Inverse square law for light and radiation: A unifying educational approach", European Journal of Engineering and Technology Research, Vol. 2, No. 11, 2017, pp. 23–27, doi: 10.24018/ejeng.2017.2.11.517.

[35]Ebisawa, Y., "Realtime 3D position detection of human pupil", in 2004 IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2004.(VCIMS). IEEE, 2004, pp. 8–12, doi: 10.1109/VECIMS.2004.1397176.

[36]Su, Y., Kuo, C.-C. J., "Fast and robust camera's auto exposure control using convex or concave model", in 2015 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2015, pp. 13–14, doi: 10.1109/ICCE.2015.7066300.

[37]Li, C.-C., Hsu, S.-W., Shen, P.-C., Guo, J.-I., "A single-camera high dynamic range technique by using contrast enhancement and exposure control", in Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. IEEE, 2014, pp. 1–4, doi: 10.1109/APSIPA.2014.7041720.

[38]Li, D., Tao, P., Wen, J., “High dynamic range video with synthesized gain control”, in 2012 IEEE International Conference on Emerging Signal Processing Applications. IEEE, 2012, pp. 29–32, doi: 10.1109/ESPA.2012.6152438.

[39]Kim, J., Cho, Y., Kim, A., “Exposure control using bayesian optimization based on entropy weighted image gradient”, in 2018 IEEE International conference on robotics and automation (ICRA). IEEE, 2018, pp. 857–864, doi: 10.1109/ICRA.2018.8462881.

[40]Tomasi, J., Wagstaff, B., Waslander, S. L., Kelly, J., “Learned camera gain and exposure control for improved visual feature detection and matching”, IEEE Robotics and Automation Letters, Vol. 6, No. 2, 2021, pp. 2028–2035, doi: 10.1109/LRA.2021.3058909.

[41]Yang, H., Wang, B., Vesdapunt, N., Guo, M., Kang, S. B., “Personalized exposure control using adaptive metering and reinforcement learning”, IEEE transactions on visualization and computer graphics, Vol. 25, No. 10, 2018, pp. 2953–2968, doi: 10.1109/TVCG.2018.2865555.

[42]Wang, Z., Yang, L., Huang, J., Zhang, Z., Xie, Q., Chen, Y., Liao, J., Zhang, Z., Zhang, H., Zhang, K., “The investigation of automatic exposure under extreme light”, in The International Conference on Photonics and Optical Engineering (icPOE 2014), Vol. 9449. SPIE, 2015, pp. 513–517, doi: 10.1117/12.2074827.

[43]Kao, W.-C., Cheng, L.-W., Chien, C.-Y., Lin, W.-K., “Robust brightness measurement and exposure control in real-time video recording”, IEEE Transactions on Instrumentation and Measurement, Vol. 60, No. 4, 2010, pp. 1206–1216, doi: 10.1109/TIM.2010.2087835.

[44]Luo, L., Chen, Q., He, W.-J., Lu, Z.-X., “The algorithm and implementation of EMCCD automatic gain adjustment based on fixed gray level”, in AOPC 2015: Optical and Optoelectronic Sensing and Imaging Technology, Vol. 9674. SPIE, 2015, pp. 575–582, doi: 10.1117/12.2201041.

[45]Park, S., Kim, G., Jeon, J., “The method of auto exposure control for low-end digital camera”, in 2009 11th International Conference on Advanced Communication Technology, Vol. 3. IEEE, 2009, pp. 1712–1714.

[46]Jiang, T., Kuhnert, K.-D., Nguyen, D., Kuhnert, L., “Multiple templates auto exposure control based on luminance histogram for onboard camera”, in 2011 IEEE International Conference on Computer Science and Automation Engineering, Vol. 3. IEEE, 2011, pp. 237–241, doi: 10.1109/CSAE.2011.5952672.

[47]Zhang, C., You, Z., Yu, S., “An automatic exposure algorithm based on information entropy”, in Sixth International Symposium on Instrumentation and Control Technology: Signal Analysis, Measurement Theory, Photo-Electronic Technology, and Artificial Intelligence, Vol. 6357. SPIE, 2006, pp. 152–156, doi: 10.1117/12.716772.

[48]Liu, J., Ren, D., Zou, J., Wu, Y., Li, S., “Study of automatic exposure algorithm based on hd ip camera”, in 2010 International Conference on Advanced Intelligence and Awarenss Internet (AIAI 2010). IET, 2010, pp. 265–268, doi: 10.1049/cp.2010.0766.

[49]Ying Liu, H., Feng, B., Wang, P., Li, Y., yun Wei, H., “Design and realization of an AEC&AGC system for the CCD aerial camera”, in 2015 International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Processing Technology, Vol. 9622. SPIE, 2015, pp. 200–205, doi: 10.1117/12.2182313.

[50]Chun, J.-B., Jung, H., Kyung, C.-M., “Dynamic-range widening in a cmos image sensor through exposure control over a dual-photodiode pixel”, IEEE transactions on electron devices, Vol. 56, No. 12, 2009, pp. 3000–3008, doi: 10.1109/TED.2009.2033327.

[51]Nourani-Vatani, N., Roberts, J., “Automatic camera exposure control”, in Proc. of the 2007 Australasian Conference on Robotics and Automation, Brisbane, Australia (December 2007), 2007.

[52]Yang, X., Zhong, S., “The implementation of fast light-adjusting based on PID system”, in 2015 Chinese Automation Congress (CAC). IEEE, 2015, pp. 1828–1833, doi: 10.1109/CAC.2015.7382801.

[53]Takahashi, Y., Nowak, W., Wisspeintner, T., “Adaptive recognition of color-coded objects in indoor and outdoor environments”, Lecture Notes in Computer Science, Vol. 5001, 2008, pp. 65–76, doi: 10.1007/978-3-540-68847-1_6.

[54]Neves, A. J., Cunha, B., Pinho, A. J., Pinheiro, I., “Autonomous configuration of parameters in robotic digital cameras”, in Pattern Recognition and Image Analysis: 4th Iberian Conference, IbPRIA 2009 Póvoa de Varzim, Portugal, June 10-12, 2009 Proceedings 4. Springer, 2009, pp. 80–87, doi: 10.1007/978-3-642-02172-5_12.

[55]Janzen, R., Mann, S., “Feedback control system for exposure optimization in high-dynamic-range multimedia sensing”, in 2016 IEEE International Symposium on Multimedia (ISM). IEEE, 2016, pp. 119–125, doi: 10.1109/ISM.2016.0031.

[56]Sousa, R. M., Wäny, M., Santos, P., Morgado-Dias, F., “Advanced illumination control algorithm for medical endoscopy applications”, in Image Sensing Technologies: Mate-

rials, Devices, Systems, and Applications II, Vol. 9481. SPIE, 2015, pp. 100–108, doi: 10.1117/12.2176928.

[57]Zhang, J., Newman, J. P., Wang, X., Thakur, C. S., Rattray, J., Etienne-Cummings, R., Wilson, M. A., "A closed-loop, all-electronic pixel-wise adaptive imaging system for high dynamic range videography", IEEE Transactions on Circuits and Systems I: Regular Papers, Vol. 67, No. 6, 2020, pp. 1803–1814, doi: 10.1109/TCSI.2020.2973396.

[58]Aydin, I., Othman, N. A., "A new iot combined face detection of people by using computer vision for security application", in 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE, 2017, pp. 1–6, doi: 10.1109/IDAP.2017.8090171.

[59]Sajjad, M., Nasir, M., Muhammad, K., Khan, S., Jan, Z., Sangaiah, A. K., Elhoseny, M., Baik, S. W., "Raspberry pi assisted face recognition framework for enhanced law-enforcement services in smart cities", Future Generation Computer Systems, Vol. 108, 2020, pp. 995–1007, doi: 10.1016/j.future.2017.11.013.

[60]Paliy, I., Sachenko, A., Kurylyak, Y., Boumbarov, O., Sokolov, S., "Combined approach to face detection for biometric identification systems", in 2009 IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. IEEE, 2009, pp. 425–429, doi: 10.1109/IDAACS.2009.5342946.

[61]Priadana, A., Habibi, M., "Face detection using haar cascades to filter selfie face image on instagram", in 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT). IEEE, 2019, pp. 6–9, doi: 10.1109/IDAACS.2009.5342946.

[62]Wheeler, F. W., Perera, A. A., Abramovich, G., Yu, B., Tu, P. H., "Stand-off iris recognition system", in 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems. IEEE, 2008, pp. 1–7, doi: 10.1109/BTAS.2008.4699381.

[63]Liu, J., Kumar, A., "Detecting presentation attacks from 3D face masks under multispectral imaging", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2018, pp. 47–52, doi: 10.1109/CVPRW.2018.00014.

[64]Yoon, H. S., Baek, N. R., Truong, N. Q., Park, K. R., "Driver gaze detection based on deep residual networks using the combined single image of dual near-infrared cameras", IEEE Access, Vol. 7, 2019, pp. 93 448–93 461, doi: 10.1109/ACCESS.2019.2928339.

[65]Naqvi, R. A., Arsalan, M., Rehman, A., Rehman, A. U., Loh, W.-K., Paul, A., "Deep learning-based drivers emotion classification system in time series data for remote applications", Remote Sensing, Vol. 12, No. 3, 2020, pp. 587, doi: 10.3390/rs12030587.

[66]Watanabe, T., Yamamoto, K., Kato, K., Kojima, S., Nakanisi, S., "Proposal of a method of skin region detection by two near wavelengths of near intrared", in SICE Annual Conference 2007. IEEE, 2007, pp. 291–294, doi: 10.1109/SICE.2007.4420993.

[67]Wu, C., Samadani, R., Gunawardane, P., "Same frame rate ir to enhance visible video conference lighting", in 2011 18th IEEE International Conference on Image Processing. IEEE, 2011, pp. 1521–1524, doi: 10.1109/ICIP.2011.6115734.

[68]Di, W., Wang, R., Ge, P., Cao, F., "Driver eye feature extraction based on infrared illuminator", in 2009 IEEE Intelligent Vehicles Symposium. IEEE, 2009, pp. 330–334, doi: 10.1109/IVS.2009.5164299.

[69]Tomita, Y., Suzuki, S., Fukai, H., Khosla, R., Mitsukura, Y., "The classification system with the evolutionary computation", in 2010 11th IEEE International Workshop on Advanced Motion Control (AMC). IEEE, 2010, pp. 750–755, doi: 10.1109/AMC.2010.5464033.

[70]Zou, X., Kittler, J., Messer, K. *et al.*, "Face recognition using active near-ir illumination.", in BMVC, 2005, doi: 10.5244/C.19.24.

[71]Hamouz, M., Kittler, J., Kamarainen, J.-K., Paalanen, P., Kalviainen, H., "Affine-invariant face detection and localization using GMM-based feature detector and enhanced appearance model", in Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. IEEE, 2004, pp. 67–72, doi: 10.1006/jcss.1997.1504.

[72]Viola, P., Jones, M., "Rapid object detection using a boosted cascade of simple features", in Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, Vol. 1. Ieee, 2001, pp. I–I, doi: 10.1109/AMC.2010.5464033.

[73]Wang, Y.-Q., "An analysis of the viola-jones face detection algorithm", Image Processing On Line, Vol. 4, 2014, pp. 128–148, doi: 10.5201/ipol.2014.104.

[74]Freund, Y., Schapire, R. E., "A decision-theoretic generalization of on-line learning and an application to boosting", Journal of computer and system sciences, Vol. 55, No. 1, 1997, pp. 119–139, doi: 10.1006/jcss.1997.1504.

[75]Kim, S., Hwang, B., Lee, M., "Gaze tracking based on pupil estimation using multilayer perception", in The 2011 International Joint Conference on Neural Networks. IEEE, 2011, pp. 2683–2689, doi: 10.1109/IJCNN.2011.6033570.

[76]Gupta, S., Gupta, N., Ghosh, S., Singh, M., Nagpal, S., Vatsa, M., Singh, R., "Face-surv: A benchmark video dataset for face detection and recognition across spectra and resolutions", in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp. 1–7, doi: 10.1109/FG.2019.8756510.

[77]Li, W.-C., Ou, W.-L., Fan, C.-P., Huang, C.-H., Shie, Y.-S., "Near-infrared-ray and side-view video based drowsy driver detection system: Whether or not wearing glasses", in 2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). IEEE, 2016, pp. 429–432, doi: 10.1109/APCCAS.2016.7803994.

[78]Baek, J. W., Han, B.-G., Kim, K.-J., Chung, Y.-S., Lee, S.-I., "Real-time drowsiness detection algorithm for driver state monitoring systems", in 2018 tenth international conference on ubiquitous and future networks (ICUFN). IEEE, 2018, pp. 73–75, doi: 10.1109/ICUFN.2018.8436988.

[79]Pan, K., Liao, S., Zhang, Z., Li, S. Z., Zhang, P., "Part-based face recognition using near infrared images", in 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007, pp. 1–6, doi: 10.1109/CVPR.2007.383459.

[80]Zou, X., Kittler, J., Messer, K., "Accurate face localisation for faces under active near-ir illumination", in 7th International Conference on Automatic Face and Gesture Recognition (FGR06). IEEE, 2006, pp. 369–374, doi: 10.1109/FGR.2006.18.

[81]Li, S. Z., Chu, R., Liao, S., Zhang, L., "Illumination invariant face recognition using near-infrared images", IEEE Transactions on pattern analysis and machine intelligence, Vol. 29, No. 4, 2007, pp. 627–639, doi: 10.1109/TPAMI.2007.1014.

[82]Kang, J., Anderson, D. V., Hayes, M. H., "Face recognition for vehicle personalization with near infrared frame differencing", IEEE Transactions on Consumer Electronics, Vol. 62, No. 3, 2016, pp. 316–324, doi: 10.1109/TCE.2016.7613199.

[83]Jeong, M., Kwak, J., Ko, B. C., Nam, J., "Facial landmark detection based on an ensemble of local weighted regressors during real driving situation", in 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016, pp. 2198–2203, doi: 10.1109/ICPR.2016.7899962.

[84]Jeong, M., Ko, B. C., Kwak, S., Nam, J.-Y., "Driver facial landmark detection in real driving situations", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 28, No. 10, 2017, pp. 2753–2767, doi: 10.1109/TCSVT.2017.2769096.

[85]Hasan, M. K., Ahsan, M. S., Newaz, S. S., Lee, G. M., "Human face detection techniques: A comprehensive review and future research directions", Electronics, Vol. 10, No. 19, 2021, pp. 2354, doi: 10.3390/electronics10192354.

[86]Artan, Y., Paul, P., Perronin, F., Burry, A., "Comparison of face detection and image classification for detecting front seat passengers in vehicles", in IEEE Winter Conference on Applications of Computer Vision. IEEE, 2014, pp. 1006–1012, doi: 10.1109/WACV.2014.6835994.

[87]Artan, Y., Bulan, O., Loce, R. P., Paul, P., "Driver cell phone usage detection from hov/hot nir images", in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2014, pp. 225–230, doi: 10.13140/2.1.4011.8409.

[88]Artan, Y., Bulan, O., Loce, R. P., Paul, P., "Passenger compartment violation detection in hov/hot lanes", IEEE Transactions on Intelligent Transportation Systems, Vol. 17, No. 2, 2015, pp. 395–405, doi: 10.1109/TITS.2015.2475721.

[89]Yan, J., Lei, Z., Wen, L., Li, S. Z., "The fastest deformable part model for object detection", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014, doi: 10.1109/CVPR.2014.320.

[90]He, L., Li, H., Zhang, Q., Sun, Z., He, Z., "Multiscale representation for partial face recognition under near infrared illumination", in 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2016, pp. 1–7, doi: 10.1109/BTAS.2016.7791187.

[91]Wu, F., You, W., Smith, J. S., Lu, W., Zhang, B., "Image-image translation to enhance near infrared face recognition", in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 3442–3446, doi: 10.1109/ICIP.2019.8804414.

[92]Kotwal, K., Marcel, S., "CNN patch pooling for detecting 3D mask presentation attacks in NIR", in 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020, pp. 1336–1340, doi: 10.1109/ICIP40778.2020.9191240.

[93]George, A., Mostaani, Z., Geissenbuhler, D., Nikisins, O., Anjos, A., Marcel, S., "Biometric face presentation attack detection with multi-channel convolutional neural network", IEEE Transactions on Information Forensics and Security, Vol. 15, 2019, pp. 42–55, doi: 10.1109/TIFS.2019.2916652.

[94]Sun, X., Huang, L., Liu, C., "Context based face spoofing detection using active near-infrared images", in 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016, pp. 4262–4267, doi: 10.1109/ICPR.2016.7900303.

[95]Dowdall, J., Pavlidis, I., Bebis, G., “Face detection in the near-ir spectrum”, Image and Vision Computing, Vol. 21, No. 7, 2003, pp. 565–578, doi: 10.1016/S0262-8856(03)00055-6.

[96]Xilinx, “Zynq ultrascale+ mpsoc zcu104 evaluation kit”, available at: https://www.xilinx.com/products/boards-and-kits/zcu104.html (March 16, 2023.).

[97]Bradski, G., “The OpenCV Library”, Dr. Dobb’s Journal of Software Tools, 2000.

[98]Dai, T., Dou, Y., Tian, H., Huang, Z., “The study of classifier detection time based on opencv”, in 2012 Fifth International Symposium on Computational Intelligence and Design, Vol. 2. IEEE, 2012, pp. 466–469, doi: 10.1109/ISCID.2012.286.

[99]Brousseau, B., Rose, J., “An energy-efficient, fast FPGA hardware architecture for opencv-compatible object detection”, in 2012 International Conference on Field-Programmable Technology. IEEE, 2012, pp. 166–173, doi: 10.1109/FPT.2012.6412130.

[100]Ou, W.-L., Shih, M.-H., Chang, C.-W., Yu, X.-H., Fan, C.-P., “Intelligent video-based drowsy driver detection system under various illuminations and embedded software implementation”, in 2015 IEEE International Conference on Consumer Electronics-Taiwan. IEEE, 2015, pp. 192–193, doi: 10.1109/ICCE-TW.2015.7216850.

[101]Huang, D.-Y., Lin, T.-W., Hu, W.-C., “Automatic multilevel thresholding based on two-stage otsu’s method with cluster determination by valley estimation”, International journal of innovative computing, information and control, Vol. 7, No. 10, 2011, pp. 5631–5644, doi: 10.1109/ICPR.2016.7900303.

Biography

Jurica Vugrin was born on September 13th, 1993 in Zagreb, Croatia. He attended secondary school in Kutina, after which he continued his education at the University of Zagreb, Faculty of Electrical Engineering and Computing, where he received his Master's Degree in 2017. After graduating, he started working in INETEC on a Horizon 2020 "CreepUT" project as a development engineer. In 2018, he started working at the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb as a researcher, where he participated in the scientific project KK.01.2.1.01.0136 (Research and development of the system for driver drowsiness and distraction identification – DFDM). He was involved in educational activities on the course called Digital Logic. Since January 2022, he has been working at Siemens d.d. as an automation and MES engineer. His research interests include image processing, image analysis, and control systems, and he is an author or co-author of one journal paper and six conference papers.

List of publications

Journal publications

1. **Vugrin, J.**, Lončarić, S., (2023) Real-time NIR camera brightness control using face detection, Automatika, 64:3, 593-605, doi: 10.1080/00051144.2023.2203554

Conference publicationns

1. **Vugrin, J.**, Lončarić S. "Automatic Brightness Control for Face Analysis in Near-Infrared Spectrum", in *2021 International Conference on Signal Processing and Machine Learning (CONF-SPML)*, November 2021, pp. 296-299,
   doi: 10.1109/CONF-SPML54095.2021.00064
2. Hrabar I., Veji ć P., **Vugrin, J.**, Bunčić N., Mamić A., Budimir M., Nageswaran C., Mažeika L., Arvanitis A., "An Ultrasound Subwavelength NDE Device for Detection of Subsurface Micrometer Defects Formation in Superheated Steam Systems", in *The 27th International Conference Nuclear Energy for New Europe*, September 2018

3.Veji ć, P., **Vugrin, J.**, Budimir, M., "Ultrasonic Inspection Scanner for Early Detection of Defects in Industrial Pipelines", in *12th International Conference of the Croatian Nuclear Society Nuclear Option for CO2 Free Energy Generation Former "Nuclear Option In Countries With Small and Medium Electricity Grids"*, June 2018

4.Razum, D., Šeketa, G., **Vugrin, J.**, Lacković, I., "Optimal threshold selection for threshold-based fall detection algorithms with multiple features", in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, pp. 1513-1516, doi: 10.23919/mipro.2018.8400272

5.Šeketa, G., Džaja, D., **Vugrin, J.**, Lacković, I. and Magjarević, R., "Signal Acquisition for Accelerometer-based Fall Detection", in *25th Mini-Symposium* January 2018, pp. 57-60

6.Šeketa, G., **Vugrin, J.**, Lacković, I., "Optimal threshold selection for acceleration-based fall detection", in *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017*, November 2021, pp. 151-155, doi: 10.1007/978-981-10-7419-6_26

Životopis

Jurica Vugrin rođen je 13. rujna 1993. u Zagrebu u Hrvatskoj. Srednju školu pohađao je u Kutini, a obrazovanje je nastavio na Sveučilištu u Zagrebu, na Fakultetu elektrotehnike i računarstva, gdje je diplomirao 2017. godine. Nakon diplomiranja zapošljava se u INETEC-u na Horizon 2020 projektu "CreepUT" kao razvojni inženjer. 2018. godine počinje raditi na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu na Zavodu za elektroničko inženjerstvo i obradbu informacija kao istraživač na znanstvenom projektu KK.01.2.1.01.0136 (Research and development of the system for driver drowsiness and distraction identification – DFDM). Uz istraživački projekt radio je kao asistent u nastavi na predmetu Digitalna logika. Od siječnja 2022. radi u Siemensu d.d. kao inženjer automatizacije i MES inženjer. Njegovi interesi uključuju obradu i analizu slike te sustave upravljanja, te je autor i koautor jednog časopisnog i šest konferencijskih članaka.