

Automated Labelling by Facial Recognition
Visjonsdokument
Versjon 1.1

REVISJONSHISTORIE

Dato	Versjon	Beskrivelse	Forfatter
27.01.2023	1.0	1. Iterasjon	Markus Pedersen, Matias Raknes, Stian Aas Trohaug
21.02.2023	1.1	Siste iterasjon	Markus Pedersen, Matias Raknes, Stian Aas Trohaug

INNHALDSFORTEGNELSE

1 INNLEDNING	4
2 SAMMENDRAG PROBLEM OG PRODUKT	5
2.1 Problemsammendrag	5
2.2 Produktsammendrag	5
3 BESKRIVELSE AV INTERESSETER OG BRUKERE	6
3.1 Oppsummering interessenter	6
3.2 Oppsummering brukere	6
3.3 Brukermiljøet	6
3.4 Sammendrag av brukernes behov	6
3.5 Alternativer til produktet	6
4 PRODUKTOVERSIKT	8
4.1 Produktets rolle i brukermiljøet	8
4.2 Forutsetninger og avhengigheter	8
5 PRODUKTETS FUNKSJONELLE EGENSKAPER	9
6 IKKE-FUNKSJONELLE EGENSKAPER OG ANDRE KRAV	10
7 REFERANSER	11

1 INNLEDNING

I denne bacheloroppgaven vil gruppen først undersøke en eksisterende prototype for oppdaging og gjenkjenning av ansikter i videoopptak, som er utviklet av oppdragsgiveren til prosjektet. Deretter utvikler gruppen en egen prototype som fokuserer på problemstillingen for direktesendt TV. Den endelige prototypen vil være i stand til å oppdage ansikter på TV og tilordne dem en identitet basert på ulike faktorer, inkludert navn, alder, yrke eller utdanningsbakgrunn. Hovedoppgaven for prosjektet vil være å utvikle en prototype ved å analysere og evaluere den eksisterende løsningen, og deretter optimalisere og tilpasse den til å bedre imøtekomme kravene til prosjektets spesifikke anvendelsesområde.

Gruppens visjon er å kunne bruke denne prototypen til å automatisere og effektivisere kringkastingsprosessen ved å gi kringkastere muligheten til å lettere identifisere personer som vises på skjermen. Dette kan gjøre det lettere å skape mer relevant og engasjerende innhold for seere, samt øke effektiviteten i produksjonen.

Målet med denne oppgaven er å undersøke og implementere teknikker for oppdaging og gjenkjenning av ansikter, og evaluere deres effektivitet i forhold til problemstilling. Gruppen vil også se på et sett av forhåndsdefinerte kriterier for hva som skal avgjøre hvilken etikett en person skal få. Til slutt, skal prototypen testes ved hjelp av forhåndsdefinerte tester, samt innsamling av tilbakemeldinger fra brukere for å sikre at prototypen oppfyller oppdragsgiverens krav og forventninger.

Når grunnfunksjonene til systemet er utviklet og testet, planlegger gruppen å utvide programmet med ytterligere funksjonalitet for å muliggjøre anonymisering av personer ved å bruke en blur-effekt.

2 SAMMENDRAG PROBLEM OG PRODUKT

2.1 Problemsammendrag

Problem med	Å sette etiketter gjøres live, men det er ikke automatisert, og det gjøres ikke ved hjelp av ansiktsgjenkjenning.
berører	Problemstillingen er aktuell for Vizrt sine kunder.
som resultatet av dette	Personene som tar del i sendingen er planlagt, slik at etiketter klargjøres og legges til manuelt. Dersom det ikke er planlagt, vil det være nødvendig å lage etiketter fortløpende.
en vellykket løsning vil	Automatisere valg av etiketter og automatisk tilpasse plassering og tidspunkt.

2.2 Produktsammendrag

For	Kunder av oppdragsgiver
som	har behov for muliggjøring og automatisering av behandling av direktesendt tv.
produktet navngitt	implementasjon av prototype for ansiktsgjenkjenning
som	kan bestemme eller komme med anbefalinger for når og hvor etiketter skal plasseres.
I motsetning til	dagens metoder for å sette etiketter i direktesendt TV, som gjøres manuelt
Har vårt produkt	system for å automatisere dette.

3 BESKRIVELSE AV INTERESSENER OG BRUKERE

3.1 Oppsummering interessenter

Navn	Utdypende beskrivelse	Rolle under utviklingen
Vizrt	Utviklerselskap Vizrt som også er oppdragsgiveren.	Produsent av prototypen som skal utvikles videre. Sørger for nødvendig utstyr og teknisk hjelp.

3.2 Oppsummering brukere

Navn	Utdypende beskrivelse	Rolle under utviklingen
TV selskap	TV selskaper styrer i dag etiketter og andre grafiske elementer manuelt på direktesendinger.	Viser hva av funksjonaliteter som er aktuelt.

3.3 Brukermiljøet

Muligheten for redigering av direktesendinger i real-time består mest av enkle verktøy for å aktivere bilder eller animasjoner upåvirket av alt annet som vises. Dersom noe mer dynamisk er nødvendig, gjøres dette manuelt av en eller flere personer som styrer dette under direktesendingen.

3.4 Sammendrag av brukernes behov

Behov	Prioritet	Påvirker	Dagens løsning	Foreslått løsning
Automatisering av å sette passende grafikk på personer i direktesendinger.	1	Produsenter av direktesendinger. (TV selskap)	Manuelt styre forberedte endringer live.	System for å automatisk sette grafikk på enkeltpersoner i direktesendinger.

3.5 Alternativer til produktet

Per i dag har ikke gruppen klart å finne noen produkter som er et direkte alternativ til dette utviklingsprosjektet. Etter en samtale med oppdragsgiver, er det kommet frem til at de største konkurrentene på generell basis er Ross Video Ltd (Ross Video Ltd, i.d) og Grass Valley (Grass Valley, i.d). Dette er to canadiske selskap, som utvikler og produserer utstyr og programvare til kringkastingsindustrien. Ross Video Ltd har utviklet et system kalt VISION[Ai]RY Ft (VISION[Ai]RY, 2022), som ansees som et mulig alternativ til prosjektets produkt. Denne programvaren bruker kunstig intelligens til å oppdage, lokalisere og spore ansikter på en direkte videostrom fra kameraet. Detaljene som ekstraheres fra videostreamen

brukes til å automatisk styre panoretings-, tilt- og zoomakser til et robotkammersystem (VISION[Ai]RY, 2022). Selv om dette produktet er anvendt med andre funksjonaliteter enn prosjektets funksjonelle krav, kan det likevel trekkes likheter mellom de.

Det er funnet et lignende produkt som kalles for Automated Video Labelling (Automated Video Labelling: Identifying Faces by Corroborative Evidence, 2021), hovedforskjellen mellom prosjektets produkt og deres er at de bruker det på videoer som ikke er live.

4 PRODUKTOVERSIKT

4.1 Produktets rolle i brukermiljøet

Produktet vil være en stedfortreder for det arbeidet som gjøres når man skal identifisere mennesker på live-TV. Det skal gjøre det enklere og mer intuitivt ved bruk av automatiserte løsninger. Ansiktsgjenkjenning kan gjøre arbeidet med å identifisere personer raskere og enklere, noe som kan bidra til å effektivisere arbeidsflyten og øke produktiviteten til TV-selskaper. Det kan også samles data om hvem som dukker opp foran kamera og når, noe som kan bidra til å forstå hvem som er mer populære enn andre eller mer relevant for et gitt program eller begivenhet.

4.2 Forutsetninger og avhengigheter

Det er en nødvendighet at produktet lever opp til forventningene av brukerne, og at det anvendes riktig. Det er dermed viktig at produktet følger nøye presisjon. Prosjektet vil dermed avhenge av teknologi for ansiktsgjenkjenning, som for eksempel biblioteker eller APIer. Det er viktig at det brukes stabil teknologi og at den er oppdatert. Hvis noen av disse forutsetningene eller avhengighetene endres underveis i prosjektet, kan det påvirke dette dokumentet og kreve at det endres for å reflektere den nye situasjonen. Det er dermed ideelt å være forberedt på å tilpasse dokumentet etter riktige behov.

5 PRODUKTETS FUNKSJONELLE EGENSKAPER

1. Støtte for å vise video fra live kamera.
2. Automatisk gjenkjenning av ansikter foran kameraet
3. Påskrifter eller merking av gjenkjente personer på skjermen
4. Mulighet for manuell endring av automatisk gjenkjente personer
5. Lagring av historikk over gjenkjente personer
6. Vise antall ansikter som er på kamera
7. Kunne trykke på ansiktets rammer (velge personer i bildet for å gjøre endringer)
8. Produktet skal kunne kjøres som en desktop-applikasjon
9. Tilby knapper for pause og starte video
10. Programmet skal kunne aktivere etiketter manuelt og automatisk
11. Integrasjon med eksisterende kringkastingsutstyr og teknologi
12. Mulighet for å legge til eller endre ansikts biblioteket.
13. Mulighet for å sette opp integrasjoner med andre systemer.

6 IKKE-FUNKSJONELLE EGENSKAPER OG ANDRE KRAV

- **Brukervennlighet:**
Produktet skal ha et intuitivt og lettfattelig grensesnitt, som gjør det enkelt å sette opp og bruke funksjonene.
- **Modularitet:**
Produktet skal programmeres modulært, som gjør det enklere å teste og utvide. Ved å gjøre koden modulær, vil det også være enklere å integrere med andre systemer, og brukes på flere områder.
- **Ytelse:**
Produktet skal ha god ytelse, med rask gjenkjenning av ansikter og minimal forsinkelse i live TV-sending.
- **Skalerbarhet:**
Produktet skal kunne håndtere et stort antall ansikter og kameraer samtidig.
- **Dokumentasjon og support:**
Produktet skal ha god dokumentasjon og støtte for å hjelpe brukerne med å sette opp og bruke funksjonene.

7 REFERANSER

Ross Video Ltd (2023, 21.02). Hentet fra

<https://www.rossvideo.com/company/about-ross/company-overview/>

Grass Valley (2023, 17.02). Hentet fra

<https://www.grassvalley.com/company-information/about-us/>

VISION[Ai]RY (2023, 21.02). Hentet fra

<https://www.rossvideo.com/products-services/acquisition-production/robotic-camera-systems/studio-robotic-control-systems/visionairy-ft/>

Automated Video Labelling: Identifying Faces by Corroborative Evidence(2023, 14.02).
Hentet fra

https://www.robots.ox.ac.uk/~vgg/research/person_id_in_video/