HÖGSKOLAN SKÖVDE

Institutionen för kommunikation och information
Examensarbete i Datalogi 20p
C-nivå
Vårterminen 2007

Användarverifiering från webbkamera

Sami Alajarva

Användarverifiering från Webbkamera

Examensrapport inlämnad av Sami Alajarva till Högskolan i Skövde, för Kandidatexamen (B.Sc.) vid Institutionen för kommunikation och information. Arbetet har handletts av Fredrik Johansson.

Datum: 14 juni 2007

Härmed intygas att allt material i denna rapport, vilket inte är mitt eget, har blivit tydligt identifierat och att inget material är inkluderat som tidigare använts för erhållande av annan examen.

Signerat: _____

Användarverifiering från Webbkamera

Sami Alajarva

Sammanfattning

Arbetet som presenteras i den här rapporten handlar om ansiktsigenkänning från webbkameror med hjälp av principal component analysis samt artificiella neurala nätverk av typen feedforward. Arbetet förbättrar tekniken med hjälp av filterbaserade metoder som bland annat används inom ansiktsdetektering. Dessa filter bygger på att skicka med redundant data av delregioner av ansiktet.

Nyckelord: Biometri, Ansiktsigenkänning, Principal Component Analysis, Artificiella Neurala Nätverk, Feedforward, Filter

Innehållsförteckning

1 Introduktion

Datoranvändare har en tendens att glömma sina lösenord till de system som de arbetar med. Ett sätt att lösa det här problemet är att identifiera användaren med hjälp av biometri. Hietmeyer (2000) undersökte olika former av biometri för maskinläsliga resedokument så som pass och kom fram till att ansiktsdrag är enklast att använda jämfört med till exempel irisskanning och fingeravtryck. För att identifiera ett ansikte krävs det mindre ansträngning från enskilda individer, de behöver inte ens veta att de blir identifierade då det bara krävs en bild på deras ansikte.

Identifiering med hjälp av biometri sker som annan identifiering av personer, en jämförelse görs på de biometriska kännetecknen som läses in från personen ifråga och de sedan tidigare lagrade kännetecknen som identifierar personen.

Det finns många tillämpningar för ansiktsidentifiering och verifiering, bland annat som identifiering av brottslingar, tillträdeskontroll till byggnader och datorsystem samt övervakningssystem. Målet i alla dessa tillämpningar är att från en bild på en person kunna identifiera vem denne är eller verifiera att personens identitet är rätt. Många gånger har dessa identifieringssystem till uppgift att kunna identifiera ett stort antal användare. Allteftersom användarantalet i ett system växer så kommer det att innehålla många personer som är lika varandra och därmed ökar kravet på exakt identifiering.

Genom att utrusta ett datorsystem med webbkameror kan bilder fås som kan användas till identifiering av användare. Ett problem med webbkameror är att de har lägre upplösning jämfört med digitala stillbildskameror och denna låga bildkvalitet ökar svårigheten för användarigenkänning och verifiering.

Ansiktsigenkänning är ett högdimensionellt mönsterigenkänningsproblem och hör till områdena mönsterigenkänning och biometri. Ansiktsigenkänning har många olika problem så som olika ljussättningar på bilder, rotation av ansikten, ansiktsutryck, oskärpa och bakgrunder (Yokono och Poggio, 2005). Andra nämnvärda saker som kan försvåra igenkänning av ansikten är komprimeringsartefakter i bilder och brus från dåliga fotosensorer.

Den här rapporten behandlar problemet med säker verifiering av personer från bilder med låg kvalitet, främst bilder från webbkameror. Filtrerade metoder som idag används för att öka träffsäkerheten i ansiktsdetektering kommer att utvärderas för ansiktsigenkänning med hjälp av artificiella neurala nätverk. Som indata till de artificiella nätverken kommer data från webbkameror att användas. Bilderna från webbkameran kommer att bearbetas med Principal Component Analysis som beskrivs i sektion 2.2.

Metoden som används för att ta fram resultat är experimentering av två olika tekniker, den teknik som användes av Jiang (1996) och som beskrivs i kapitel två, samt en filterbaserad teknik som bygger på denna. Resultaten från experimentet presenteras i kapitel sju.

2 Bakgrund

Det här kapitlet börjar med att ge en bakgrundsbeskrivning av ansiktsigenkänning följt av en överblick av ansiktsdetektering och artificiella neurala nätverk.

2.1 Ansiktsigenkänning

Forskning om ansiktsigenkänning i bilder och video med hjälp av olika typer av maskininlärning började med tidiga system av Bledsoe (1966) och Kanade (1973). Ansiktsigenkänning hör till området mönsterigenkänning vilket är ett område som är komplext och därför svårt att hitta konventionella algoritmer för. Mönsterigenkänning kan användas på många typer av mönster så som nätverksövervakning och personidentifiering.

Efterhand har det tagits fram automatiserade och semiautomatiserade identifieringsmetoder. Dessa metoder är baserade på positionen på bland annat ögonen, näsan och munnen. Att låta en människa identifiera de kännetecken i ansiktet som igenkänning ska ske efter är enligt Turk och Pentland (1991) inte optimalt. Enligt dem ska datorn själv hitta de kännetecken den ska basera sin sökning på. De kan i slutändan vara de samma som människan intuitivt kommer fram till eller så kan de skilja sig från dessa.

Lu (2003) delar upp ansiktsigenkänningsscenarion i tre kategorier, *(i)* ansiktsverifiering, *(ii)* ansiktsidentifiering samt *(iii)* watch-list, där denna rapport fokuserar på *(i)*. Ansiktsverifieringsalgoritmer arbetar efter frågan ”Är du den som du säger dig vara?”.

2.1.1 Ansiktsverifiering

Varje dag måste människor verifiera sin identitet genom att exempelvis säga sitt namn, skriva sitt användarnamn eller betala med betalkort. För att andra personer ska kunna verifiera att personen är den hon hävdar sig vara kräver de att se någon form av identifikation. Vid inloggning till ett system hävdar en person att denne är en viss användare och verifierar sin identitet med ett lösenord eller biometri.

2.1.2 Ansiktsidentifiering

När personer umgås med individer i sin bekantskapskrets behöver de inte hävda sin identitet utan identifiering sker baserat på tidigare erfarenhet. Med biometri kan identifiering ske genom att användaren till exempel låter systemet skanna ett finger. Systemet söker sedan i en databas efter ett likadant avtryck och om skillnaden mellan det skannade avtrycket och avtrycket i databasen når under ett tröskelvärde har systemet identifierat användaren. Jämfört med verifiering är identifiering ett större problem eftersom systemet under identifieringen bara har tillgång till informationen från biometrin.

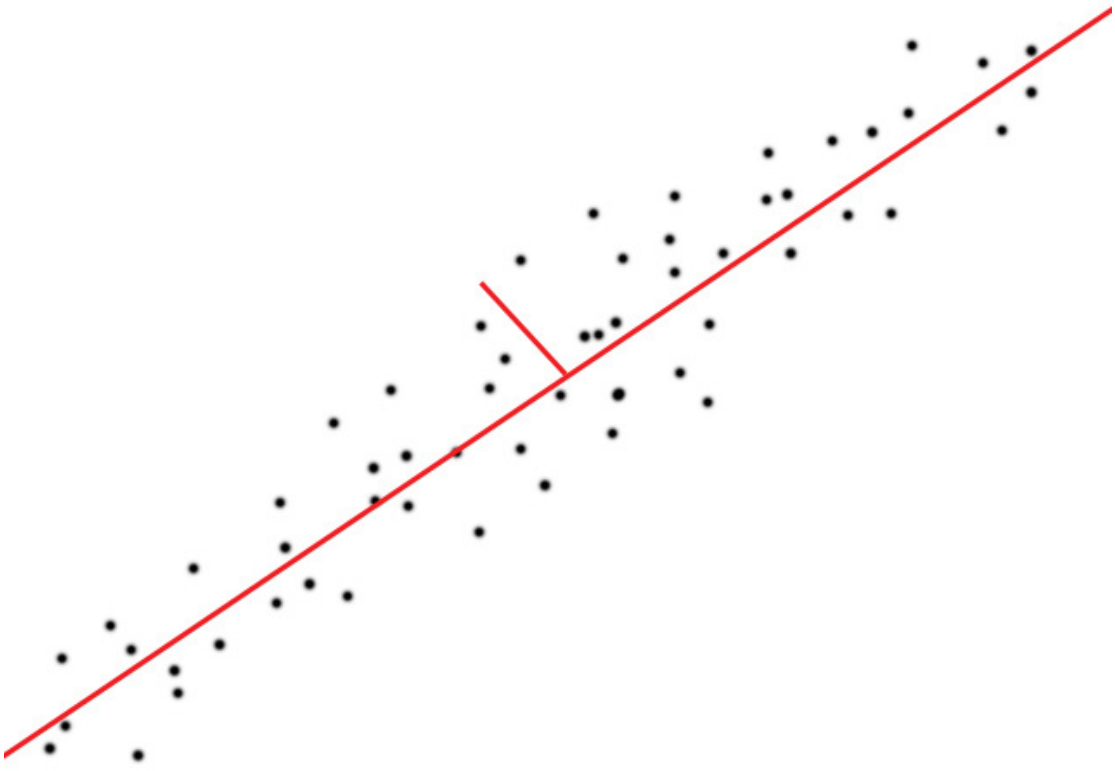
2.1.3 Watch-list

Watch-list arbetar i en öppen värld där de individer som testas kanske finns i systemet. Testindivider jämförs med lagrad information och rangordnas i en lista där de som är mest lika informationen i systemet kommer övers. Om testindividen faller inom ramen för felmarginal i jämförelsen med den lagrade informationen så larmar systemet. Denna teknik är riktad mot övervakningssystem och inte säkerhetssystem.

2.2 Principal Component Analysis

Turk och Pentland (1991) förespråkar användning av så kallade eigenfaces vid ansiktsigenkänning, eigenfaces beskrivs bland annat av Pissarenko (2002).

Om rå pixeldata[1] används erhålls mycket överflödig information. Enligt Russell och Norvig (2003) kan reduceringstekniker som principal component analysis öka hastigheten på igenkänning av till exempel ansikten. Om en bild på ett ansikte är av storleken 100x100 pixlar så kan denna ses som en punkt i ett 10000 dimensionellt plan. Med hjälp av principal component analysis kan denna dimensionalitet reduceras. Detta är möjligt eftersom bilder på ansikten oftast är väldigt lika varandra och hamnar då nära varandra i det högdimensionella planet. Genom att räkna ut eigenvektorer som kan beskrivas som axlar i en lägre dimension kan man fördela datan på ett mer optimalt sätt, figur 1. Denna datarymd kallas för eigenspace.



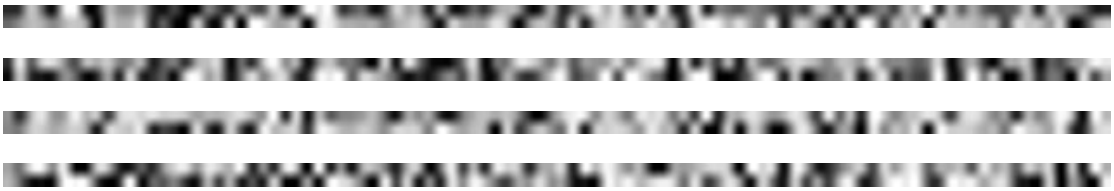
Figur 1 Exempel på vektorer som beskriver datarymden.

[1] Rå pixeldata avser data direkt från de pixlar som utgör bilden. Om bilden är 10x10 pixlar så blir der 100 pixlar data. Inom ansiktsigenkänning används oftast svartvita bilder på ansikten med 256 steg från svart till vitt.

Eigenface-algoritmen använder sig av principal component analysis för att få ner dimensionen på data. Principal component analysis bygger på statistiska metoder som är baserade på att om det finns en tillräckligt stor samling data så går det att analysera relationer mellan individuella punkter i samlingen. Tekniken har enligt Smith (2002) fått fotfäste inom ansiktsigenkänning och bildkomprimering.

Eigenface-algoritmen identifierar karakteristiska kännetecken i bilder på individer. Därefter jämförs de kännetecknen med kännetecknen för de individer som är lagrade.

I vanliga fall används statistiska metoder för att känna igen ett ansikte i eigenface form men det har utförts experiment som använder artificiella neurala nätverk som klassificeringsmetod av känslor i bilder (Padgett och Cottrell, 1996). Jiang (1996) beskriver en metod för att koppla samman principal component analysis med artificiella neurala nätverk för användning inom ansiktsigenkänning. Metoden går ut på att projicera varje bild i träningsdatan in i det eigenspace som tidigare skapats med träningsdatan och använda den vektor av reella tal som indata i artificiella neurala nätverk. Denna vektor av reella tal är av fast längd där längden bestäms av antalet eigenvektorer som används. Om endast de 100 eigenvektorerna med de högsta eigenvärdena används kommer vektorn att bestå av 100 reella tal. Denna vektor kan enligt Jiang (1996) användas som ett handtag till originalbilden, det vill säga att vektorn identifierar originalbilden på ett sätt som är mycket unikt för ansiktet. Visuella representationer av dessa vektorer kan ses i figur 1. Jiangs (1996) principal component analysis and neural network based face recognition metod, kallas från och med nu för bara för Jiangs teknik.



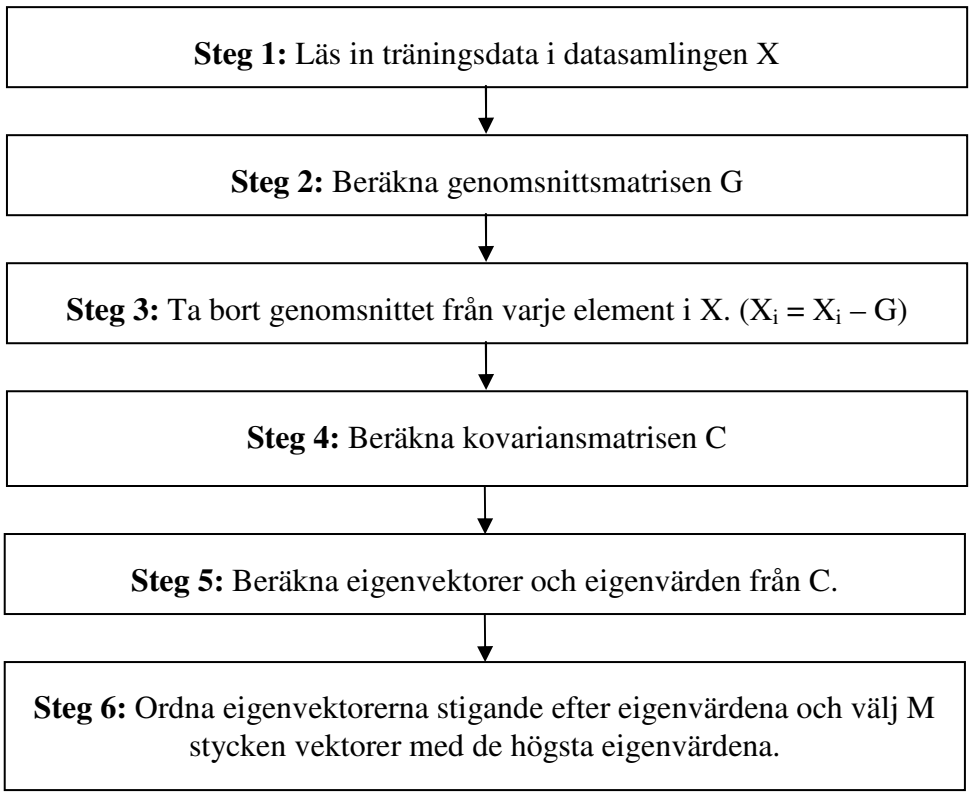
Figur 2 : Fyra exempel på visuella representationer av de två vektorer per bild som beskriver ansiktet+filtrering efter projecering in i eigenspace.

Ren pixelbaserad ansiktsigenkänning är också känslig för imperfekt data, och det är just imperfekt data som fås från exempelvis webbkameror på grund av den hårda komprimeringen som sker och allmänt höga nivåer av brus. För att få tillgång till de mest generella dragen från en samling bilder används de eigen-vektorer med de högsta eigen-värdena.

I korthet så fungerar principal component analysis som en teknik för att hitta relationer i data samt att den komprimerar data på ett sätt så att inte mycket information går förlorad.

2.2.1 Algoritmen för Principal Component Analysis

Steg för steg beskrivning enligt Smith (2002).



Flöde 1 Algoritmen för beräkning av eigenspace

Genom att välja endast de M vektorer med högst eigenvärden har vi kvar de vektorer med mest vikt, alltså principal component av det totala antalet vektorer. Denna delmängd av de totala antalet vektorer kallas för eigenspace.

För att reducera antalet dimensioner på varje enskild bild i träningsdatan projiceras bilderna ut på de M valda vektorerna. Detta ger M reella värden som beskriver bilden på dessa nya axlar.

2.3 Ansiktsdetektering

Ansiktsdetektering ur bilder är ett nödvändigt första steg för att kunna utföra ansiktsigenkänning. Ansiktsdetektering går ut på att ur en bild kunna skilja ut ansikten från bakgrund för att sedan kunna använda dem till bland annat ansiktsigenkänning. Ett sätt att utföra ansiktsdetektering är med hjälp av artificiella neurala nätverk där nätverken tränas att skilja på ansikten och bakgrund. Ett stort problem i ansiktsdetektering är falsk flaggning, vilket innebär att algoritmen hittar ett ansikte där det inte finns något ansikte.

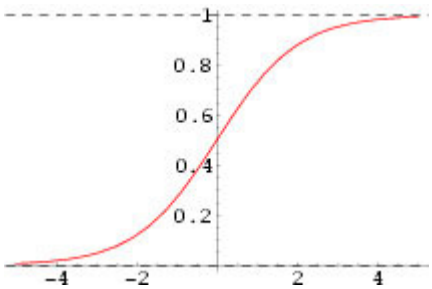
Figur 3. Exempel på indata till en ansiktdetektor samt indata + stripe filter.

Rowley, Baluja och Kanade (1998) föreslår en filterbaserad algoritm för att hitta ansikten i bilder. Deras metod gick ut på att skicka in en rektangulär region av en bild samt delregioner där de förväntade sig hitta till exempel ögonen till ett neuralt nätverk, exempel på filtrering kan ses i figur 2. Det sättet att skicka med både original data och filtrerad data minskade risken att identifiera ett område av bilden som ett ansikte fastän det inte var ett ansikte.

2.4 Artificiella Neurala Nätverk för Ansiktsverifiering

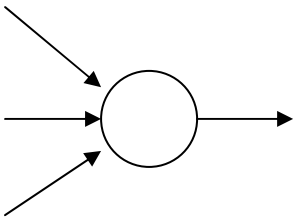
Informationsbehandling med algoritmer som försöker efterlikna hjärnans nervceller kallas för artificiella neurala nätverk. Den här typen av algoritmer kan lösa många problem som är svåra för vanliga datalogiska metoder så som klassificering av data. De fungerar efter principen att de kan lära sig att approximera svar. Den största skillnaden mellan neurala nätverk och vanliga algoritmer är att neurala lösningar är inlärda och inte programmerade (Callan, 1999).

I den biologiska hjärnan skickar neuroner elektriska impulser till varandra via synapser. Beroende på hur nära två neuroners synapser är kan den elektriska impulsen varieras. I den artificiella varianten av neuroner används så kallade vikter för överföring av impulser. Varje enskild vikt har ett värde bundet till sig som signalen från sändarneuronen multipliceras med. En neuron har också en aktiveringsfunktion som kan vara till exempel en stegfunktion eller en sigmoidfunktion. Sigmoidfunktionen kan ses i figur 3. Aktiveringsfunktionen aktiverar neuronen om det summerade värdet av inkommande signaler överstiger ett tröskelvärde. Det är aktiveringsfunktionen som bestämmer om en neuron ska skicka en signal till nästa lager i nätverket.



Figur 4: Aktiveringsfunktion sigmoid

Den enklaste formen av artificiella neurala nätverk har inga gömda lager och bara en neuron i sitt utdatalager, dessa nät kan endast lösa enkla linjära problem och kallas för perceptroner, se figur 4. Utöver de signaler som perceptronen får utifrån så har den en konstant input som kan ses som en basnivå för aktivering.



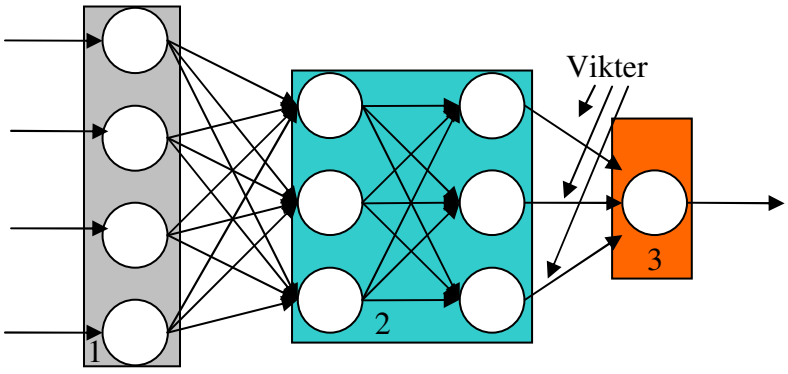
Figur 5 Enkelt artificiellt neuralt nätverk, så kallad perceptron.

För att kunna lösa ickelinjära problem krävs det att dessa neuroner kopplas samman till större nätverk. Beroende på hur neuronerna är sammankopplade inom nätverken passar näten till olika problemområden.

2.4.1 Arkitektur

Det finns ett stort antal olika nätverksarkitekturer inom artificiella neurala nätverk. De vanligast förekommande arkitekturerna är bland annat feedforward-nätverk och Hopfield-nät. Fokus i den här rapporten ligger på feedforward-arktitekturen som är lämpligast för den här sortens problem eftersom de kan tränas att endast avfyras då de känner igen ett mönster.

2.4.1.1 Feedforward-nätverk



Figur 6 Enkelt feedforward-nätverk med 2 gömda lager(2), 4 neuroner i indata lagret(1) samt en neuron i lagret för utdata (3)

Ett artificiellt neuralt nätverk baserat på feedforward-arkitekturen består av ett lager för indata, ett godtyckligt antal gömda lager samt ett lager för utdata. Varje lager skickar data framåt till nästa lager. Lagret för indata används för att fördela data till det följande gömda lagret så att varje neuron får data från alla neuroner i indatalagret.

Callan (1999) hävdar att ett gömt lager räcker för att lösa de flesta problem men genom att lägga till fler gömda lager så kan nätets effektivitet ökas samt träningsfasen snabbas upp.

2.4.2 Träning

> "Training neural networks is said to be more of an art than a science. While this claim could be made of many fields of science, it is true that there is no clear formula for successful network modeling" Rhode (2000)

Träning brukar delas upp i supervised, unsupervised samt så kallad reinforcement learning.

För att träna feedforward-nätverket till sin uppgift används en backpropagation-algoritm som propagerar fel tillbaka längsmed nätet och ändrar värden på de vikter som binder samman neuronerna för att minska felet (Rumelhart, Hinto och Williams, 1986). Backpropagation-algoritmen är en form av supervised learning. Vikterna som binder samman neuronerna i nätet ändras för att minimera det kvadratiska felet mellan verklig utdata och förväntad utdata från nätet. Träningen består av flera epoker som var och en går igenom all träningsdata och antalet epoker som träningen tar är beroende av vad det är för problem som ska lösas, vilket värde som variabeln learning rate har samt om momentum används. Variablerna learning rate och momentum tas upp senare i detta kapitel. Träningen avslutas när skillnaden mellan den förväntade utdatan och den verkliga utdatan från nätet ligger tillräckligt nära varandra. Om träningen pågår för länge kan ett nätverk få problem med generalisering av ny data, detta kallas för överträning och betyder att nätverket specialiserar sig på det data som används under träningen. För att få nätverket att generalisera på ett bra sätt kan antalet noder i de gömda lagren eller antalet epoker som träningen pågår ändras.

Feedforward-nät kan ha en till flera neuroner i output lagret, dessa kan tränas så att endast en av dem aktiveras av viss sorts indata. Detta kan användas till att klassificera data. På grund av att backpropagation-algoritmen går igenom neuronerna en i taget i utdatalagret kan det uppstå konflikter i hur vikter ska ändras och i värsta fall stabiliseras inte vikterna alls. Det har visats sig att den här typen av konflikter kan undvikas i speciella fall genom att öka antalet neuroner i indata lagret. Dessa extra neuroner kan vara statiska (Yamaguchi, 2004).

En nackdel med backpropagation-algoritmen är enligt Callan (1999) att det kan ta lång tid att träna ett nät men efter att träningen har avslutats så låses vikterna fast för att snabbt kunna klassificera okänd data. Storleken på indata har stor påverkan på träningstiden, därmed är det inte bra att använda sig av rå pixeldata som skulle ge en neuron per pixel i indatalagret, i en 100x100 bild skulle det bli 10000 neuroner. Eigenfaces är därmed en bra väg att gå för att den reducerar dimensioner på indata och samtidigt minimerar påverkan från brus och komprimeringsartefakter.

Batch-träning och mönster-träning är de två huvudkategorier för hur nätverk kan tränas med backpropagation.

Generalisering är ett sätt att mäta hur bra ett artificiellt neuralt nätverk är på att klassificera data som inte använts under träningen. Antalet neuroner i de gömda lagren samt längden på träningsfasen, det vill säga antalet epoker som näten tränas, kan påverka hur bra ett nätverk är på att generalisera (Callan, 1999).

Momentum och Learning Rate

Två variabler som har stor inverkan på hur effektivt ett feedforward-nätverk kan tränas är momentum och learning rate. Learning rate är den variabeln som styr hur mycket vikterna ska ändras varje epok. Ett stort värde på learning rate betyder att nätet lär sig snabbt men att det finns en risk för att nätets vikter aldrig stabiliseras eftersom de ändras för mycket varje epok.

Momentum används för att knuffa på viktändringen åt de håll som det börjat ändras åt. Eftersom vikterna ändras så att totala felet ska bli mindre för varje epok står det nära att anta att om en vikt börjar röra sig åt ett håll så kommer den att röra sig åt det hållet. Momentum beskrivs av bland annat Callan (1999). Det är svårt att hitta ett optimalt värde på både LR och momentum, de är väldigt beroende på det specifika problemet som skall lösas.

2.4.2.1 Mönster-träning

I mönster-träning utförs uppdatering av vikter stegvis direkt efter det att ett mönster har presenterats för nätverket. Det betyder att om det finns tillgång till ett stort antal mönster under träningen så kommer nätverkets vikter att uppdateras efter varje presentation och detta gör att träningen tar lång tid.

2.4.2.2 Batch-träning

I batch-träning samlas alla fel ihop för en hel epok och uppdateringen av vikterna sker först efter att hela epoken är klar. Detta kan enligt Callan (1999) spara tid när träning av nätverk sker för vissa typer av problem.

Det kan dock vara nödvändigt att plocka ut träningsdata i slumpvis ordning för varje epok för att undvika presentation av exakt samma batch för varje epok (Rhode, 2000).

2.5 Relaterat arbete

Många olika tekniker för att förbättra precisionen på ansiktsigenkänning har genom åren tagits fram.

2.5.1 3D-Modeller

3D-modellering av ansikten är en teknik som gör det möjligt att identifiera ansikten oberoende av dess position i och med att modellen kan roteras. Denna teknik kan bland annat användas för identifiering av skådespelare i filmer som ett led i att indexera dessa. Everingham och Zisserman (2004) utvärderade den här tekniken på gamla tv-serier och kom fram till att karaktärer kan identifieras i 75-95 % av alla bildrutor med 10 % falska identifieringar. Likheter mellan denna teknik och tekniken som presenteras i denna rapport är att båda använder sig av mellanlager för att representera ansikten. 3D-modellering representerar data som 3D-objekt medan principal component analysis representerar data som projecering i eigenspace.

2.5.2 Geometrisk igenkänning

I geometrisk igenkänning identifieras personer utifrån en databas som innehåller information som avstånd mellan olika punkter i ett ansikte. Dessa punkter har manuellt identifierats på alla bilder (Bledsoe, 1966). Fördelar med denna typ av igenkänning är att det inte är känsligt för olika ljussättningar eller brus. Nackdelen är att den kräver manuell markering av de punkter som igenkänningen är baserad på. Skillnaden mellan geometrisk inlärning och tekniken som presenteras i den här rapporten är att maskinen får identifiera de punkter som den ska basera ansiktsigenkänningen på genom användning av principal component analysis.

2.5.3 Convolutional Neural Network med Optimal Brain Damage

Convolutional neuralt nätverk är speciella feedforward-nätverk som bland annat använts för igenkänning av handskrivna tecken (Bengio, 1993). Det använder sig av speciella lager för att upptäcka speciella drag hos föregående lager. Optimal Brain Damage används för att skala bort onödiga parametrar från ett nätverk och höjer därmed nätets prestanda. Nystrand (1999) visade att användning av optimal brain damage i samband med convolutional neural networks höjer prestandan och därmed tillförlitligheten på befintliga system. Likheten mellan den här tekniken och den artificiella neurala nätverk baserade principal component analysis tekniken som presenteras i den här rapporten är att båda använder sig av artificiella neurala nätverk om än olika slags nät.

3 Problembeskrivning och problemställning

Inriktningen med det här arbetet är att implementera och utvärdera ofiltrerad eigenfacebaserad ansiktsigenkänning och filtrerad eigenfacebaserad ansiktsigenkänning med artificiella neurala nätverk.

3.1 Motivering

Ett sätt att lösa problemet med säker inloggning till system utan att behöva komma ihåg lösenord kan snabba upp arbetet med olika system. Ansiktsigenkänning är den form av biometri som enligt Hietmeyer (2000) kräver minst arbete från användare genom att de i bästa fall inte ens behöver vara medvetna om att de blir identifierade. Eftersom webbkameror räknas till standardtillbehör på dagens datorer är det av intresse att undersöka om de kan användas till ansiktsigenkänning.

Det stora problemet med webbkameror är att de endast producerar bilder av låg kvalitet. Bilderna som fås från webbkameror är ofta brusiga, suddiga och innehåller artefakter från bildkomprimering.

Eigenfaces använder sig av principal component analysis för att reducera dimensionen på data och gör det samtidigt möjligt att fokusera på de karakteristiska dragen hos en individ och därmed ignorera brus och artefakter från komprimeringsalgoritmer. Eigenfaces är därmed av intresse i den här sortens problem det har bland annat visats av Jiang (1996) som undersökte principal component analysis och artificiella neurala nätverk vid ansiktsigenkänning från webbkamera.

Inom ansiktsdetektering visade Rowley m.fl (1998) att det är möjligt att öka träffsäkerheten på ansiktsdetektering med hjälp av filtreringstekniker som beskrivs i kapitel 2.3. Eftersom denna teknik fungerar för ansiktsdetektering är det intressant att utvärdera denna typ av filtrering för ansiktsigenkänning från lågkvalitetsbilder.

Figur 7: Exempel på Eigenface och delregionsfilter (Bild tagna från experimentsystemet)

3.2 Problemprecisering

Problemet som denna rapport utreder består av:

Att undersöka om filterbaserade tekniker ökar träffsäkerheten på Jiangs (1996) teknik med principal component analysis och artificiella neurala nätverk.

3.3 Förväntade resultat

Resultatet från arbetet kommer att bli en undersökning om filtrerade tekniker kan öka träffsäkerheten i Jiangs ansiktsigenkänningsteknik. Resultatet från dessa jämförelser kommer att presenteras i rapporten.

4 Metod

I det här kapitlet identifieras den metod som kan användas för att få fram de resultat som beskrivs i sektion 3.3.

Experiment inom datalogi utförs ofta som en implementation av ett experimentsystem. Olika variabler undersöks för att se hur de påverkar det implementerade systemet (Berndtsson, Hansson, Olsson och Lundell, 2004). Grunden för genomförandes ges av Jiangs teknik. För att undersöka om filtrering förbättrar Jiangs teknik så kommer endast ett fåtal variabler att undersökas och ändras. Den data som undersökningen genomförs på kommer att vara den samma. Detta är enligt Berndtsson m.fl (2004) ett sätt som faller inom ramarna för ett Experiment inom datalogi.

För att kunna undersöka om den filterbaserade tekniken förbättrar pricksäkerheten i Jiangs teknik krävs det att ett experimentsystem implementeras.

Den metod som identifierats för denna rapport är experiment med tillhörande implementation.

4.1 Experiment

Eftersom målet med arbetet är att undersöka om den filterbaserade tekniken förbättrar pricksäkerheten för Jiangs teknik så krävs det att lösningen testas med bilder på ansikten som ej tidigare setts. Mätningar sker på hur de olika teknikerna klarar av att hantera tidigare osedda bilder på personen som ska få ett positivt svar, det vill säga rätt användare, samt bilder på personer som skall ge ett negativt svar, det vill säga intrångsförsök.

Implementationen i detta experiment innehåller två delar, en ansiktsdatabas med bilder tagna med en vanlig webbkamera på olika användare samt en implementation av ett testprogram baserat på Jiangs teknik och filterbaserade tekniker.

4.1.1 Ansiktsdatabas

Ansiktsdatabasen består av bilder tagna med en webbkamera på olika användare. Dessa bilder är tagna på ungefär samma avstånd och i liknande ljus. Bilderna kommer att genomgå en simulerad ansiktsdetektering, det vill säga att ur de bilder som fås från webbkameran så kommer en kvadratisk region att klippas ut där ansiktet är. Som ett sista steg innan ansiktena läggs in i ansiktsdatabasen så normaliseras varje bild. Normaliseringen består av rotation av de ansikten som inte är helt raka, korigering av ljuset i bilderna så att de får en jämn ljusstyrka över hela bilden samt att kontrasten blir ungefär samma i varje bild.

4.1.2 Implementation

Implemetationen försöker efterlikna Jiangs implemetation så långt det går. Den består av en modul för principal component analysis samt en modul för artficiella neurala nätverk. En detaljerad beskrivning på implemetationen finns i sektion 5.2.

5 Genomförande

I detta kapitel presenteras testmiljön och de olika testerna som använts till att testa Jiangs (1996) teknik samt den filterbaserade tekniken.

För att kunna jämföra Jiangs teknik med den filterbaserade tekniken var det första steget att implementera båda teknikerna i ett experimentsystem.

5.1 Insamling av data

Insamlingen av bilder till databasen sker genom att låta användare arbeta med en dator med en webbkamera. En mängd bilder, 15-20, tas på varje användare för träning av systemet. Av vissa användare tas ytterligare fem bilder för användning vid testning. Totala antalet olika individer i systemet var till slut 15.

Figur 8 Exempel på ansikten

5.2 Implementering av experimentsystemet

Den övergripande arkitekturen för de två delarna som finns i båda experimentsystemen syns i flöde 2.

5.2.1 Upprepning av Jiangs försök

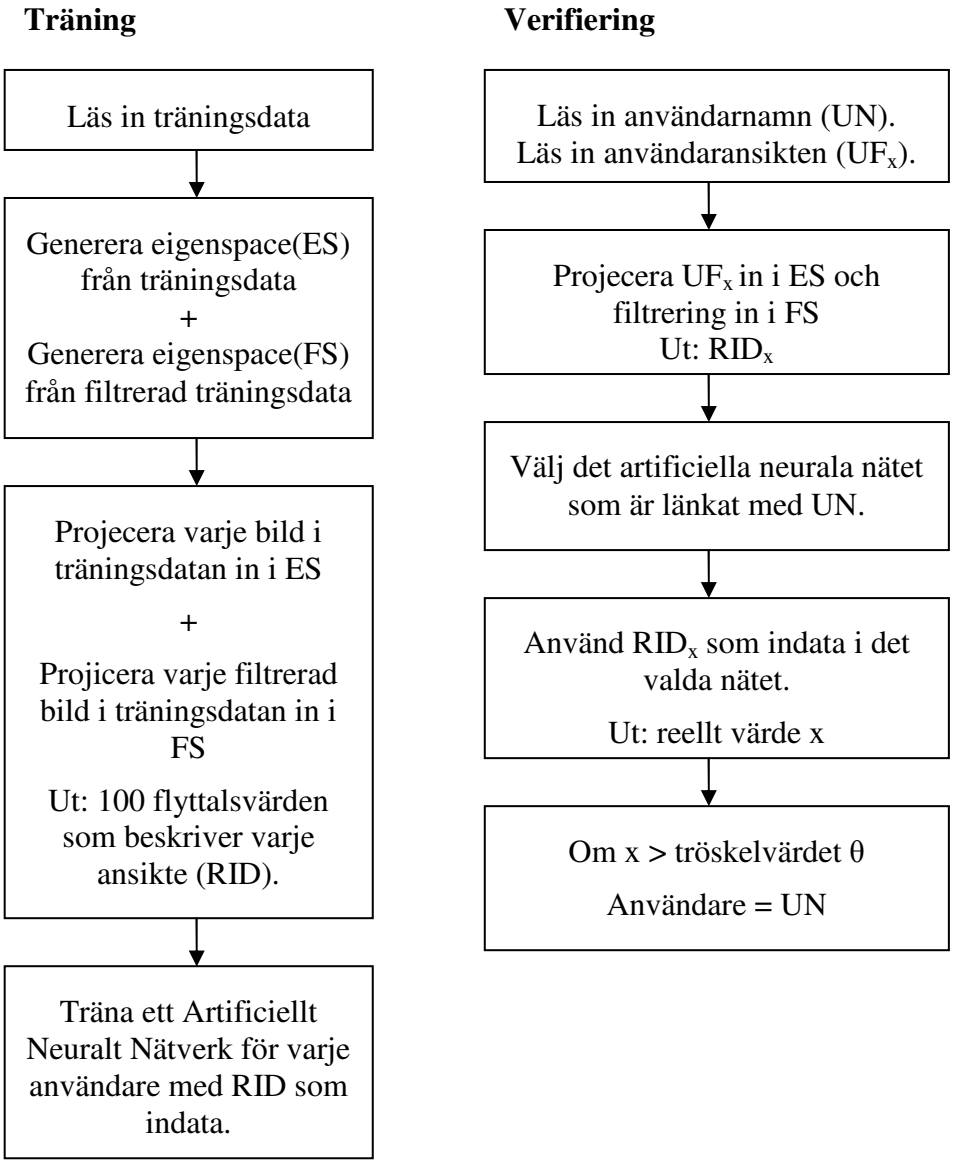
I Jiangs (1996) experiment användes ett feedforward nätverk med 100 neuroner i indata-lagret, 10 neuroner i det gömda lagret samt en neuron som utdata. Antalet neuroner i indata-lagret är baserat på antalet vektorer som beskriver eigenspace för ansiktena. Jiang (1996) använde sig av ett eigenspace som beskrevs med 100 eigenvektorer. Som aktiveringsfunktion i indatalagret användes en linjär funktion medan resterande lager använde en sigmoidfunktion.

Det framgår inte tydligt vilka värden som learning rate och momentum har i Jiangs system så värdena på dessa två variabler togs fram genom experimentering. Värdena framgår i tabell 1. Bildstorleken på ansiktsbilderna valdes till 46x46 pixlar för att det var denna storlek som användes i Jiangs teknik. Det framgår inte heller om Jiang system använde sig av bias i de artficiella neruala nätverken men eftersom bias allt som oftast är en självklarhet i nätverk av typen feedforward så antogs det att bias har använts.

5.2.2 Implementation av filterbaserad teknik

Den filterbaserade tekniken använder sig av samma antal vektorer för att beskriva eigenspace för hela ansikten som Jiang (1996), det vill säga 100 vektorer.

Det filtrerade området i varje bild är av storleken 46x13 pixlar och innehåller regionen runt ögonen, detta område omvandlas till en kvadratisk bild för att kunna projiceras i eigenspace. Eigenspacen för den filtrerade bilden, som från och med nu kallas för filterspace, beskrivs också den av 100 vektorer.



Flöde 2 Flödeschema över de två delar som utgör experimentsystemen.

I det första fallet för den filterbaserade tekniken används feedforward nätverk med 200 neuroner i indata-lagret, 20 neuroner i det gömda lagret samt en neuron i utdata-lagret. I andra fallet för den filterbaserade tekniken användes feedforward nätverk med 200 neuroner i indata-lagret, 75 neuroner i det första gömda lagret, 50 neuroner i det andra gömda lagret samt en neuron i utdata-lagret.

Valet på 200 neuroner i indata-lagret bestäms av de 100 vektorerna som definierar eigenspace för hela ansikten samt de 100 vektorerna som definerar filterspace. Valet av 20 neuroner i det gömda lagret i första fallet kom från Jiangs teknik, eftersom antalet indata neuroner är det dubbla så dubblas antalet neuroner i det gömda lagret. De gömda lagren i det andra fallet bestämdes med den 'tumregel' som ofta används innom feedforward nätverk. Tumregeln säger att antalet neuroner i ett lager ska vara ungefär hälften av neuronerna i det föregående lagret. Som det beskrivs i kapitel 2.4 så kan för många neuroner i ett nätverk göra det svårt för nätverket att generalisera och klassificera indata som inte ingår i träningssettet.

Learning rate och momentum variablerna sattes till samma värden i fallet med Jiangs teknik.

Värdet som fås ut från utdata-lagret är ett reellt tal som beskriver till vilken grad som indatan matchar den användare som nätet tränats för.

Tabell 1:

IN:	Antalet neuroner i indatalagret
G1:	Antalet neuroner i första gömda lagret
G2:	Antalet neuroner i andra gömda lagret
UT:	Antalet neuroner i utdatalagret
LR:	Värdet på variabeln LearningRate
Momentum:	Värdet på variabeln Momentum

Teknik	IN	G1	G2	UT	LR	Momentum
Jiangs Teknik	100	10	-	1	0.01	0.5
Filterteknik1	200	20	-	1	0.01	0.5
Filterteknik2	200	75	50	1	0.01	0.5

Tabell 1: Detaljer för feedforward nätverken

5.3 Testning

I detta delkapitel preseneteras de olika parametrarna som jämförelsen mellan de två olika teknikerna är baserad på.

I alla testfall har ett tröskelvärde på 0,5 använts. Detta betyder att om utdatan från de artificiella neurala nätverken är högre än 0,5 så flaggas bilden som verifierad. Det är okänt vilket värde som Jiang använde i sitt arbete. Värdet 0,5 valdes genom experimentering för att låta experimentsystemet få större marginaler att arbeta inom.

5.3.1 Variation på indata

I de olika testfallen varierades antalet användare, olika användare, olika värde på maximalt antal träningsepoker samt antalet bilder per användare.

Antalet användare i experimentsystemet varieras för att simulera system med två till åtta användare. För varje användare i ett testfall tränas ett artificiellt neuralt nätverk som har till uppgift att känna igen denna användaren. Om ett testfall har tre användare kommer alla tre näten att testas för att få resultat för testfallet.

Användare varieras från testfall till testfall för att testa teknikerna på så många olika slags ansikten som möjligt.

Maximalt antal träningsepoker varieras mellan 1200-12000 epoker. Det lägre antalet epoker är till för att testa teknikerna på ungefär samma ställe i träningen medan den högra gärnsen är till för att låta teknikerna tränas till sitt gränsvärde där de artificiella neurala nätverken anses vara färdigtränade.

5.3.2 Jämförbara parametrar

Felaktig verifiering

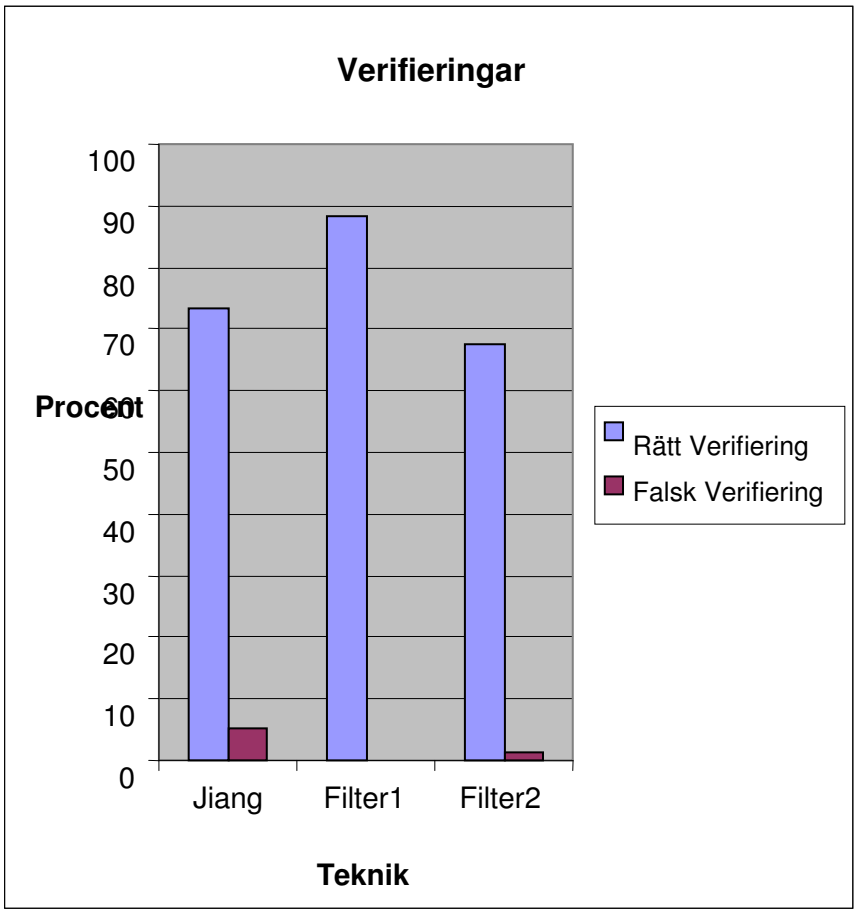
Räknas upp varje gång en bild på en användare passerar experimentsystemet och blir igenkänd av ett nät som inte hör till den användaren.

Användare ej verifierad

Räknas upp varje gång en bild på den rätta användaren passerar experimentsystemet utan att bli igenkänd av sitt eget nät.

6 Resultat

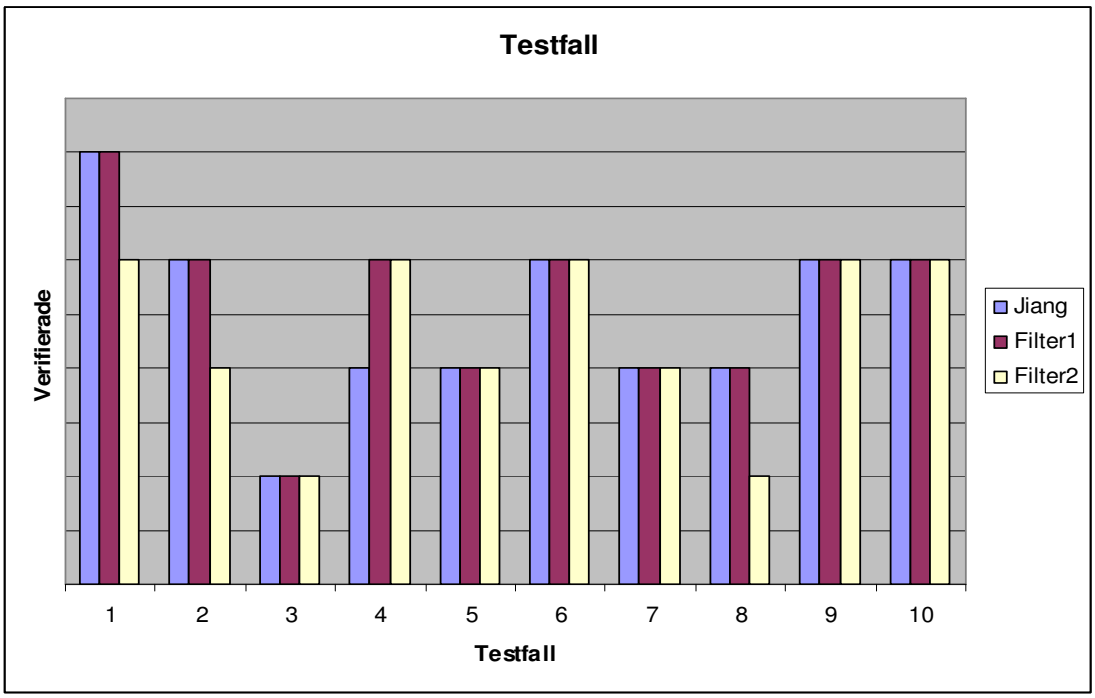
I det här kapitlet presenteras de resultat som kom fram under testningen av experimentsystemen. Totalt utfördes 10 olika testfall med olika antal användare, olika användare, olika värde på maximalt antal träningsepoker samt olika antal bilder per användare. Parametrarna beskrivs i kapitel 5.



Graf 1 Procentuell jämförelse av antalet korrekta verifieringar samt falska verifieringar mellan teknikerna.

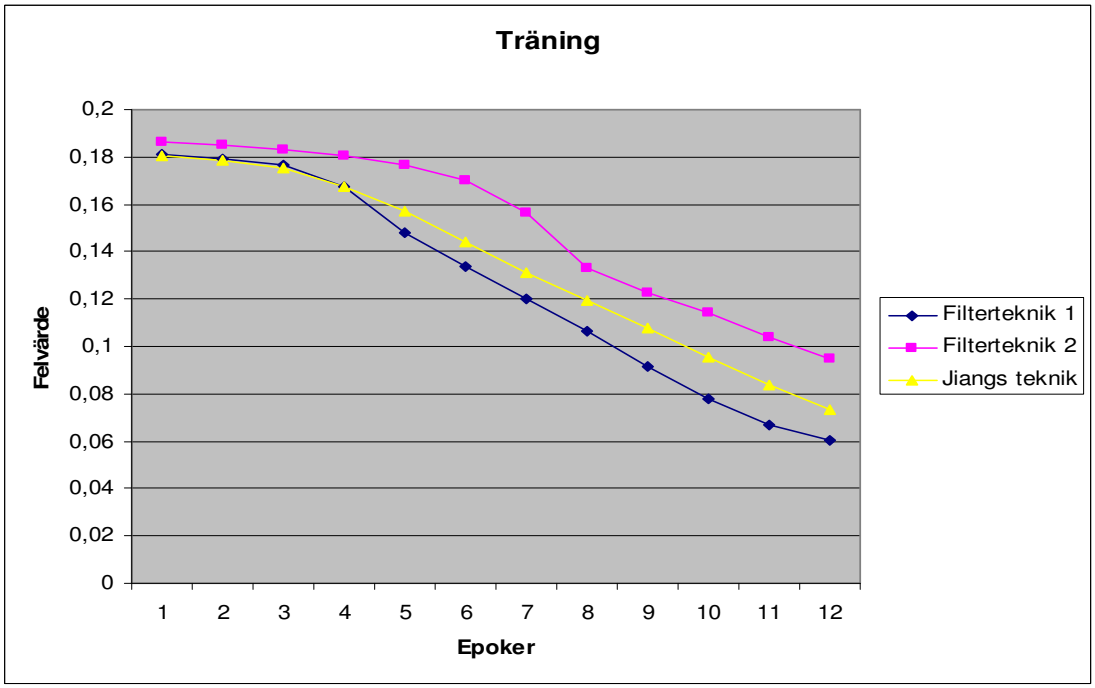
Graf 1 beskriver genomsnittliga resultaten från de olika teknikerna. Jiangs teknik har en genomsnittlig igenkänningsfrekvens på 73.5% av tiden. Filterteknik 1 kommer upp i 88.2% igenkänningsfrekvens på testdatan. Medan det större artificiella neurala nätverket i filterteknik 2 endast kommer upp i 67.6%.

Falsk verifiering sker i 5.1% av fallen med Jiangs teknik. Filterteknik 1 flaggar inte någon felaktig användare som verifierad medan filterteknik 2 gör det i 1.3% av fallen.



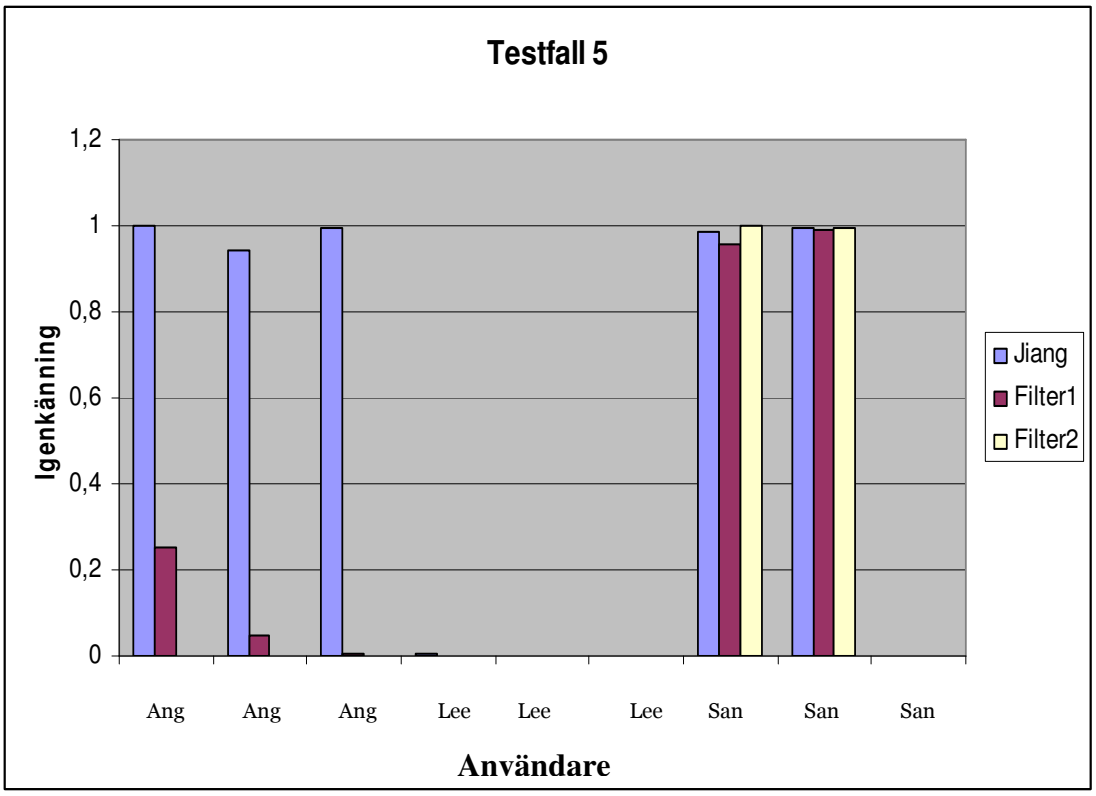
Graf 2 Jämförelse av antalet rätta verifieringar i de 10 testfallen

Graf 2 visar hur antalet rätta identifieringar relaterar mellan de olika teknikerna i de olika testfallen. Jiangs teknik klarar inte av att verifiera lika många av användarbilderna i testfall 4. Filterteknik 1 är lika bra eller bättre på att verifiera en användare från bilder i testdatan. Filterteknik 2 klarar inte av att generalisera klassificeringsproblemet lika bra som de två andra teknikerna i 3 av fallen.



Graf 3 Felkurvan i träningen av de artificiella neurala nätverken för användare Lee i testfall 8. Varje enhet i X-led motsvarar 100 epoker.

Graf 3 visar hur felkurvan närmar sig noll i träningen av artificiella neurala nätverken i testfall 8. Felkurvan för Filterteknik 1 och Jiangs teknik har ett liknande beteende medan kurvan för filterteknik 2 skiljer sig.



Graf 4 Igenkänningsgrad av användare i testfall 5 för ett nät som tränats att känna igen San.

Graf 4 visar graden av igenkänning av tre olika användarna i systemet. Alla tre teknikerna känner igen användare San i två av tre fall, det vill säga att de känner igen rätt användare i två av tre fall. Ingen av teknikerna förväxlar användare San med Lee medan Jiangs teknik tror att bilder på användare Ang är bilder på användare San till den grad att den skulle verifiera Ang som San. Filterteknik 1 skulle inte verifiera Ang om inte tröskelvärdet skulle vara satt extremt lågt. Med experimentens tröskelvärde på 0,5 kommer dessa inte flaggas som verifierade.

7 Analys

Det här kapitlet består av en analys av de resultat som presenterades i kapitel 6.

7.1 Avvikelser från Jiangs experiment

Systemet som utvecklades av Jiang (1996) kom fram till en genomsnittlig igenkänning på 96% av fallen. Den version som togs fram för experimenten i den här rapporten kom endast upp i 73.2%. Det står alltså klart att den version av Jiangs system som togs fram för den här rapporten inte är exakt likadan som originalet. Detta kan ha att göra med att Jiangs system använder sig av bilder direkt från en webbkamera under testning och träning medan systemet i den här rapporten endast har igenkänningsdelen av originalsystemet och tar bilderna från en databas på ansikten. Detta leder till färre antal bilder per användare och därmed mindre data för träning, Jiang använde i genomsnitt 140 bilder per person medan systemet i den här rapporten i genomsnitt har 20 bilder per person. I och med att experimentsystemet i den här rapporten implementerades efter beskrivningen av igenkänningsdelen i Jiang (1996) antas det att avvikelserna inte kommer ur själva implementeringen. Men utan tillgång till Jiangs implementation är det svårt att ta reda på om orsaken till denna avvikelse ligger i implementationen. En annan orsak till skillnaden kan vara bitmasken som Jiang använde för att klippa bort luft runt huvudet, i experimentsystemet som användes i den här rapporten ansågs denna bitmask inte nödvändig eftersom bilderna som användes var hårdare klippta och lämnade inget tomrum runt ansiktena.

Troligaste orsaken till avvikelsen är storleken på träningsdata och testdata.

7.2 Analys av den filterbaserade tekniken

De två olika filtreringstekniker som testats under experimentet i den här rapporten kom den ena fram till bättre igenkänningsfrekvens än Jiangs medan den andra fick en sämre frekvens. Båda filterbaserade teknikerna var bättre på att inte släppa igenom bilder på fel individer.

Filterteknik 1 som har ett artficiellt neuralt nätverk vars storlek är härledd ur Jiangs teknik, klarar av att generalisera problemet och känna igen 88,2% av testbilderna på rätt användare utan att flagga någon av de andra testbilderna som användaren. Jämför detta med Jiangs teknik som flaggar 5,1% av de andra testbilderna.

Filterteknik 2 som använder sig av ett artficiellt neuralt nätverk som är baserat på den tumregel som ofta används när artificiella neurala nätverk ska tas fram lär sig träningsdatan snabbare än Jiangs teknik och filterteknik 1, detta kan ses i tabell 2. Den snabba inlärningen av träningsdatan kommer av att det finns ett stort antal vikter i det artificiella neurala nätverket. Detta gör att det blir svårt för nätet att generalisera problemet och klassificera okänd data.

8 Slutsatser

Detta kapitel presenterar slutsatser på arbetet som presenterats i denna rapport.

8.1 Summering

Ansiktsigenkänning har visats vara den form av biometri som stör användaren minst. Webbkameror kan idag räknas som standardutrustning på datorer men problemet med bilder från webbkameror är att de allt som oftast lider av dålig kvalité jämfört med bilder tagna med vanliga digitala kameror. Jiang (1996) beskriver en teknik baserad på principal component analysis samt artificiella neurala nätverk som överkommer problemet med dålig bildkvalité. Filterbaserade tekniker har använts inom ansiktsdetektering för att öka träffsäkerheten på detekteringen (Rowley m.fl, 1998).

Den här rapporten har undersökt om dessa filterbaserade tekniker ökar träffsäkerheten på Jiangs teknik. Undersökningen har utförts på ett experimentsystem där jämförelser har kunnat göras mellan Jiangs teknik och den filterbaserade tekniken. Experimentet genomfördes på insamlade bilder från webbkameror på 15 personer.

Den slutsats som kan dras med hänsyn till analysen i kapitel 7 är att filterbaserade tekniker kan höja pricksäkerheten i experimentsystemet jämfört med den ofiltrerade teknik som Jiang skrivit om. Både den filtrerade och den ofiltrerade tekniken är bra på att identifiera den korrekta användaren medan de filtrerade teknikerna får bättre resultat i felidentifieringen. Det vill säga att de inte släpper igenom andra användare lika ofta som Jiangs teknik.

8.2 Diskussion

Det står klart att det experimentsystem som användes för att ta fram resultaten i den här rapporten

Ansiktsigenkänning med den teknik som presenterats i den här rapporten kan vara ett effektivt sätt att tillhandahålla säker inloggning i olika system med tanke på att när väl de artificiella neurala nätverken har tränats så sker verifieringen av användare snabbt. Ett av problemen med att använda sig av feedforward nätverk är att träningsfasen tar tid. Detta problem kringås med att använda enskilda artificiella neurala nätverk för varje användare och därmed krävs det inte att en omträning sker av de nätverken som redan är i systemet, i alla fall inte direkt. På lång sikt kan detta ändå vara en god idé.

Eftersom ansiktsigenkänningssystem är svåra att träna till en godtagbar nivå, det kan vi se genom att kolla på hur Jiangs teknik presterade med den träningsdata som fanns tillgodo i detta experiment, så krävs det att det utvecklas automatiserade träningsmetoder och datainsamling för att kunna säkerställa att träningsdatan håller viss kvalitét. Ansiktsigenkänning baserat på principal component analysis och artificiella neurala nätverk behöver fortfarande studeras grundligt innan de kan installeras för att hantera kritiska system med stora krav på säkerhet. Men redan idag kan de vara ett bra sätt för att lösa till exempel inloggning i hemdatorer.

8.3 Framtida arbete

Det vore intressant att se hur filterbaserade tekniker presterar i system som har tillgång till direkt data från webbkameror, det vill säga om filtrering ökar pricksäkerheten på samma sätt i dessa system som i experimentsystemet. Ett sätt skulle vara att implementera en kopia på Jiangs experimentsystem och sedan testa filtreringstekniken där för att verkligen kontrollera om tekniken höjer pricksäkerheten där.

En studie i hur dessa tekniker presterar på ett stort antal bilder på ansikten som inte finns i systemet skulle vara intressant för att se hur de olika teknikerna presterar mot intrångsförsök. Dessa bilder skulle kunna väljas efter olika parametrar som till exempel hur lika de är de användare som finns i systemet.

Studier i hur filtrering av olika delar av ansiktena och olika kombinationer av filter presterar. I denna studie användes endast filtrering på delregionen runt ögonen. Andra områden som skulle kunna vara intressanta kan vara regionen runt munnen eller näsan.

En analys av hur nätverken som används skulle kunna optimeras med till exempel optimal brain damage algoritmer för att korta ner träningstiderna och få bort onödiga neuroner ur de artificiella neurala nätverken.

9 Referenser

Bengio, Y., Le Chun, Y., Henderson, D. 1993. *Globally trained handwritten word recognizer using spatial representation, space displacement neural networks and hidden Markov models.* Advances in Neural Information Processing Systems 6, Morgan Kaufmann Publishers, Inc. pp. 937-944

Berndtsson, M., Olsson, B., Lundell, B., Hansson, J. 2004. *Planning and Implementing your Final Year Project with Success!* Springer

Bledsoe, W. W. 1966 *The model method in facial recognition*. Panoramic Research Inc., Palo Alto, CA, Rep PRI:15

Callan, R., 1999 *The Essence of Neural Networks*. Prentice Hall

Everingham, M.R., Zisserman, A. 2004. *Automated person identification in video*. In Proc. of the 3rd International Conference on Image and Video Retrieval (CIVR2004), Vol.1, pp. 289-298

Hietmeyer, R. 2000 *Biometric promises fast and secure processing of airline passenger*. ICAO Journal. Vol. 55, no 9, pp.10-11, 27-28.

Jiang, Q. 1996 *Principal Component Analysis and Neural Network Based Face Recognition*. Department of Computer Science, The University of Chicago.

Kanade, T., 1973. *Picture processing system by computer complex and recognition of human faces*. Dept. of Information Science, Kyoto University

Lu, X., 2003. *Image Analysis for Face Recognition*.

Nystrand, A. 1999. *Ansiktsigenkänning med Artificiella Neurala Nätverk*. Högskolan i Skövde

Padgett, C., Cottrell, G. 1996. *Representing face images for emotion classification*. Department of Computer Science, University of California, San Diego.

Pissarenko, D. 2002. *Eigenface-based facial recognition.*

Rhode, D. 2000. *Lens Manual* [online]. http://tedlab.mit.edu/~dr/Lens/. [Åtkomst 24 feb 2007]

Rumelhart, D.E., Hinto, G.E., Williams, R.J. 1986. *Learning Internal Representations by Error Propagation*. MA: MIT Press,

Russel, S., Norvig, P. 2003. *Artificial Intelligence - A modern Approach.* 2nd edition. Prentice Hall.

Smith, L. I. *A tutorial on Principal Component Analysis.* University of Otago. 2002

del Solar. J. R., Navarrete, P. *Towards a Generalized Eigenspace-Based Face Recognition Framework.* 2002

Turk, M., Pentland, A. 1991. *Eigenfaces for Recognition*. Journal of Cognitive Neuroscience. Vol 3, No. 1. 71-86

Yamaguchi, M. 2004. *Are multilayered backpropagation networks catastrophically amnesic?* Scandinavian Journal of Psychology 45, 357-361

Yokono, J. J., Poggio, T. 2005. *Boosting a Biologically Inspired Local Descriptor for Geometry-free Face and Full Multi-view 3D Object Recognition*.