

UNIVERSITEIT TWENTE.

HOUDBAARHEID EN HERGEBRUIK BIOMETRIE

Rapportage

drs ir Florens de Wit en dr ir Luuk Spreeuwers

MANAGEMENT SUMMARY

Inleiding

Deze rapportage geeft de resultaten weer van het literatuuronderzoek door de Universiteit Twente (UT) zoals gevraagd door de ministeries van Buitenlandse zaken (BZ) en Binnenlandse Zaken en Koninkrijkrelaties (BZK).

De algemene aanleiding van dit onderzoek is de wens van het ministerie van Buitenlandse Zaken (BZ) om flexibilisering van paspoortaanvragen door te kunnen voeren, met name voor Nederlanders in het buitenland. Deze flexibilisering mag de veiligheid en betrouwbaarheid van identiteitsverificatie uiteraard niet nadelig beïnvloeden.

Daarom is de Universiteit Twente gevraagd een literatuuronderzoek te doen naar de invloed van leeftijd en biometrische ouderdom¹ op de betrouwbaarheid van geautomatiseerde identiteitsverificatie op basis van de vingerafdrukken en gezichtsafbeelding op Nederlandse paspoorten.

Er is, op verzoek van de ministeries, door de UT gericht onderzoek gedaan naar de volgende gebruikscasussen:

- Een aanvraag waarbij biometrische data² in een eerder stadium op een voor de houder geschikt moment is afgegeven en de aanvraag wordt doorgezet op het moment dat dit noodzakelijk is bijv. vanwege het verstrijken van de geldigheidsduur van het huidige paspoort.
- Een aanvraag waarbij de biometrische data² van het voorgaande paspoort wordt overgenomen op het nieuwe paspoort. Deze casus heeft met name betrekking op Nederlanders in het buitenland die vanwege hun hoge leeftijd niet in staat zijn te verschijnen voor een aanvraag.

Beide casussen zijn gericht op het vereenvoudigen van een hernieuwde paspoortaanvraag voor houders van een Nederlands paspoort door flexibilisering omtrent de verschijningsplicht.

Beperkingen onderzoek

Dit onderzoek is beperkt tot geautomatiseerde biometrische identiteitsverificatie op basis van vingerafdrukken en/of de gezichtsafbeelding op een Nederlands paspoort.

Identiteitsverificatie door bevoegd personeel is niet onderzocht vanwege het ontbreken van studies die relevant zijn voor de Nederlandse praktijk.

Dit is nadrukkelijk geen algemeen onderzoek naar de gevolgen van zogenaamde “aging” in geautomatiseerde biometrische identiteitsverificatie noch naar de huidige betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie in het algemeen en/of toepasbaarheid voor onderscheiden van tweelingen, deduplicatie etc. van biometrische herkenningmethoden.

Ook de gevolgen van leeftijd en biometrische ouderdom voor risico's met betrekking tot het uitbuiten van specifieke zwakheden van biometrische identiteitsverificatie, zoals bijvoorbeeld “morphing” en “spoofing”, zijn geen onderdeel van dit onderzoek.

¹ Voor zover niet nader gespecificeerd zijn de in dit rapport gebruikte begrippen te vinden in de “Lijst begrippen & afkortingen” op pagina 29.

² Naast vingerafdrukken en gezichtsafbeelding geldt ook de handtekening als biometrische data; deze blijft hier buiten beschouwing.

Vraagstelling

Naar aanleiding van de genoemde gebruiks-casussen en met de voornoemde beperkingen, probeert dit rapport een antwoord te geven op twee concrete vragen. De eerste vraag heeft betrekking op de aanvraag van een paspoort waarbij de biometrie in een eerder stadium al is afgegeven. Dit betekent een toename van de biometrische ouderdom op het einde van de geldigheidsperiode met de periode tussen afgifte van de biometrie en de aanvraag. De tweede vraag heeft betrekking op herbruikbaarheid van biometrische data. Het gevolg is een toename van de biometrische ouderdom met 10 of meer jaren.

De concrete vragen zijn voor de biometrische data (d.w.z. vingerafdrukken en gezichtsafbeelding) in paspoorten:

1. Tot welke toename van de biometrische ouderdom daalt de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie niet significant? Kan de daling in betrouwbaarheid met biometrische ouderdom gekwantificeerd worden in de vorm van toegenomen False Non-Match Rates (FNMR) en False Match Rates (FMR)?
2. Brengt een toename van de biometrische ouderdom met 10 of meer jaar een significante daling van de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie met zich mee? Kan deze daling in betrouwbaarheid gekwantificeerd worden in de vorm van toegenomen False Non-Match Rates (FNMR) en False Match Rates (FMR)?

De twee vragen hebben betrekking op de opgeslagen gelaatsfoto en vingerafdrukken op het paspoort. Verder heeft de 2e vraag m.n. betrekking op personen met een leeftijd boven de 65 jaren.

Gebruikte methode

Omdat eigen onderzoek een grote inspanning zou vergen op het gebied van verzameling van data en testen van algoritmen, is door de betrokken ministeries verzocht om een literatuuronderzoek. Bij het zoeken naar ter zake doende publicaties spelen drie criteria een rol. De publicaties moeten:

1. Een beeld geven van de effecten van biometrische ouderdom van 10 en meer jaar
2. Gebruik maken van biometrische samples (d.w.z. vingerafdrukken en/of gezichtsafbeeldingen) die vergelijkbaar zijn met de samples die worden gebruikt bij geautomatiseerde biometrische identiteitsverificatie op basis van een Nederlands paspoort (bijv. een vergelijking van een pasfoto met een ter plekke gemaakte opname van het gezicht van toonder)
3. Vergelijkbare State-Of-The-Art (SOTA) herkenningsalgoritmes als gebruikt bij identiteitsverificatie met biometrie van een Nederlands paspoort onderzoeken.

Indien studies niet geheel aan deze criteria voldoen zullen hun resultaten aan de hand van de criteria op relevantie worden beoordeeld.

Resultaten

Uit het literatuuronderzoek blijkt dat er weinig of geen voldoende recente publicaties zijn die de effecten van biometrische ouderdom van 10+ jaren beschrijven voor geautomatiseerde biometrische identiteitsverificatie op basis van vingerafdruk of gezicht. Er zijn wel onderzoeken die de effecten van biometrische ouderdom op kortere termijn presenteren (tot ca. 7 jaar). Deze onderzoeken gebruiken echter veelal data die niet goed vergelijkbaar is met de biometrische samples die in geautomatiseerde identiteitsverificatie met Nederlandse paspoorten kunnen worden gebruikt (m.n. geen ter plekke genomen opnames). Voor oudere onderzoeken zijn de

geteste algoritmes niet meer SOTA. De effecten van biometrische ouderdom verschillen per herkenningsalgoritme. Hierdoor zijn geen betrouwbare kwantitatieve uitspraken te doen over de veranderingen in de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie.

M.b.t. geautomatiseerde gezichtsvergelijking blijkt uit de literatuur dat er consensus is over dat de verificatienauwkeurigheid afneemt met een toename van de biometrische ouderdom en ook met de leeftijd van de persoon vanaf middelbare leeftijd. Dit komt met name door veranderingen in het gezicht met de leeftijd. Voor een toename van de biometrische ouderdom met 1 jaar zijn volgens de literatuur die effecten beperkt; voor een toename met 10+ jaren aanzienlijk.

M.b.t. geautomatiseerde vingerafdrukvergelijking blijkt uit de literatuur dat er consensus is over dat de verificatienauwkeurigheid afneemt met een toename van de biometrische ouderdom en ook met de leeftijd van de persoon vanaf middelbare leeftijd. Er is geen consensus over de oorzaak van de afname van de verificatienauwkeurigheid met biometrische ouderdom. Er is consensus over dat het papilairlijnenbeeld op de vinger niet verandert. De lagere betrouwbaarheid van geautomatiseerde vingerafdrukvergelijking bij opgroeiende kinderen (tot ± 12 jaar) wordt verklaard met de de kleinere vinger- en afdruk grootte waardoor algoritmen minder goed werken, en een lagere kwaliteit van afdrukken. De lagere betrouwbaarheid van geautomatiseerde vingerafdrukvergelijking bij ouderen (m.n. 65+) wordt geheel verklaard door de lagere kwaliteit van de afdrukken bijvoorbeeld door een minder elastische en over het algemeen drogere huid waardoor vingerscanners het papilairlijnenbeeld moeilijker kunnen vastleggen. Ook hier lijken de effecten, zoals gegeven in de literatuur, voor een toename van de biometrische ouderdom met 1 jaar beperkt en voor een toename met 10+ jaren aanzienlijk.

Conclusies

Op basis van de gevonden literatuur blijkt dat de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie daalt bij een toename van de biometrische ouderdom ten opzichte van de huidige 10 jaren geldigheidsduur. Deze daling is beperkt bij een toename van 1 jaar.

Op basis van de gevonden literatuur blijkt dat de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie daalt bij een toename van de biometrische ouderdom met 10 of meer jaar ten opzichte van de huidige 10 jaren geldigheidsduur. Deze daling is aanzienlijk.

De beperkingen in de representativiteit van gevonden studies voor identiteitsverificatie op basis van biometrie op Nederlandse paspoorten maken dat wij geen betrouwbare kwantificering van de toename van FNMR en FMR kunnen geven voor geautomatiseerde biometrische identiteitsverificatie in de Nederlandse praktijk.

Hierdoor is ook geen betrouwbaar kwantitatief onderbouwd advies met betrekking tot de houdbaarheid, en herbruikbaarheid mogelijk.

Aanbevelingen

Dit onderzoek heeft de vraagstelling niet volledig kunnen beantwoorden. Er is nader onderzoek nodig om houdbaarheid en herbruikbaarheid van de biometrie op Nederlandse paspoorten te bepalen, op basis van een voldoende grote dataset met data die representatief is voor de vingerafdrukken en gezichtsafbeeldingen op Nederlandse paspoorten.

Bij de dataverzameling en het onderzoek moet het volgende overwogen worden:

- Dataverzameling gaat, vanwege maximale geldigheidsduur van paspoort, 10+ jaren duren
- Data moet diversiteit van Nederlanders voor wat betreft leeftijd, geslacht en uiterlijk weergeven
- Er, voor het betrouwbaar bepalen van realistische error rates, ook voldoende paren met "look alike" zijn
- Gebruik voor de Nederlandse praktijk representatieve technologie en methoden

- Onderzoek de mogelijkheid voor maatwerk (bijv. houdbaarheid verschilt per leeftijdscategorie)

In algemene zin zijn er de volgende overwegingen:

- Vanwege de permanentie van het papilairlijnenbeeld zouden vingerafdrukken zeer lang houdbaar moeten zijn, en bovendien herbruikbaar. Door de kwaliteit van de afdrukken te verbeteren zou men deze situatie kunnen benaderen. Een mogelijkheid tot verbetering zijn het gebruik van scanner-technologie die de interactie tussen vinger en sensor optimaliseert.

De verbetering van de kwaliteit van gezichtsafbeeldingen (bijv. hoge resolutie digitale opnames) zal een beperkte(re) uitwerking hebben op de houdbaarheid daarvan, vanwege de lage permanentie van het gezicht.

- Het effect van leeftijd en biometrische ouderdom kan sterk verschillen van persoon tot persoon door verschillen in levensstijl, beroep, geslacht en uiterlijke kenmerken. Maatwerk, in de vorm van het aanpassen van technologie en procedure, kan een deel van de variatie ondervangen zodat de betrouwbaarheid van geautomatiseerde identiteitsverificatie zo veel mogelijk constant blijft gedurende het leven.

INHOUDSOPGAVE

Management summary	3
1 Introductie	8
1.1 Aanleiding van dit onderzoek	8
1.2 Initiële vraagstelling	8
1.3 Begrippenkader	9
1.3.1 Biometrische identiteitsverificatie	9
1.3.2 Identificatiefouten	9
1.3.3 Definities	10
1.4 Onderzoekskader	11
1.4.1 Beperkingen in scope	11
1.4.2 Onderzoeksmethode	11
1.5 Specificering Vraagstelling	11
1.6 Leeswijzer	12
2 Gezichtsvergelijking	13
2.1 Gezicht als biometrie	13
2.2 Invloed van biometrische ouderdom op betrouwbaarheid	13
2.2.1 Invloed biometrische ouderdom op mated scores	13
2.2.2 SOTA FNMR	15
2.3 Invloed biometrische ouderdom per leeftijd(scategorie)	16
2.3.1 SOTA waarden voor FMR en FNMR	16
2.3.1.1 FMR	16
2.3.1.2 FNMR	18
2.4 Deelconclusie Gezichtsvergelijking	18
3 Vingerafdrukken	20
3.1 Vingerafdrukken als biometrie	20
3.2 Invloed biometrische ouderdom op mated scores	20
3.3 Invloed leeftijd op mated scores	23
3.4 Deelconclusie vingerafdrukken	24
3.5 Aantekeningen Vingerafdrukken	24
4 conclusie & aanbevelingen	25
4.1 Conclusie	25
4.2 Aanbevelingen	25
4.2.1 Onderzoek/Bepalen houdbaarheid/herbruikbaarheid biometrie	26
4.2.2 Algemene overwegingen	26
Referenties	28
Lijst begrippen en afkortingen	29
Appendix	31

1 INTRODUCTIE

1.1 Aanleiding van dit onderzoek

Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) alsmede het ministerie van Buitenlandse Zaken (BZ) onderzoeken de (on)mogelijkheden van het flexibiliseren van de paspoort-dienstverlening.

Met name met betrekking tot Nederlanders in het buitenland wordt de optie onderzocht om opname van biometrie los te koppelen van de aanvraag van een paspoort, doordat iemand tot enige tijd voorafgaand aan het doorzetten van de aanvraag³ bij een daartoe toegeruste locatie⁴ zijn/haar biometrie af kan laten nemen voor later gebruik.

Deze optie impliceert dat de aanvraag van een paspoort tot enige tijd na afname van de biometrie, én online kan worden doorgezet. Daaruit volgt tevens dat de identiteitsverificatie - nu bij verschijnen en dus ongeveer elke 10 jaar – bij de eerstvolgende aanvraag pas na een langere tijd plaatsvindt.

Aanvullend wil men meer inzicht krijgen in de mogelijkheid voor hergebruik van de biometrie. Hierbij zou de biometrische data van een paspoort waarvan de geldigheidsduur bereikt is opnieuw gebruikt worden voor een nieuwe aanvraag.

Een bestaande use-case betreft mensen van hoge leeftijd die in het buitenland verblijven en niet fysiek kunnen verschijnen om biometrie af te nemen.

Er is dus inzicht nodig in de invloed van zowel de leeftijd van een persoon, als de ouderdom van de biometrie op de betrouwbaarheid van identiteitsverificatie.

1.2 Initiële vraagstelling

Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) alsmede het ministerie van Buitenlandse Zaken (BZ) hebben de Universiteit Twente (UT) verzocht hiernaar onderzoek te doen met als onderzoeksvragen (citaat cursief):

- I. *Wat zijn de effecten van de leeftijd van personen en ouderdom van de afgenomen biometrie (vingerafdrukken en gezichtsopnamen) voor betrouwbare identiteitsverificatie?*
- II. *Wat is de maximum te adviseren ouderdom van biometrie voor betrouwbare identiteitsverificatie per leeftijdscategorie?*

Hierbij moet in acht worden genomen dat ook in het laatste jaar van geldigheid de biometrie nog betrouwbaar dient te zijn, onder andere voor identiteitsvastelling bij een vervolgaanvraag van een document.

Aanvullend werd het volgende als relevant aangemerkt (citaat cursief):

- *Het formuleren van een eenduidige, en in deze context werkbare, definitie van de term “houdbaarheid” in relatie tot False Match Rate (FMR), False Non-Match Rate (FNMR), [...] ⁵ en de gekozen drempelwaarde.*
- *Inzicht in de houdbaarheidstermijn van de biometrie in relatie tot een aantal gekozen relevante normeringen voor de FMR*

³ De aanvraag zou formeel starten bij afname van de biometrie, en voortgezet worden op het moment dat de (toekomstige) houder in de huidige context zou verschijnen voor de aanvraag.

⁴ Te denken valt hier aan de ambassade, het consulaat-generaal of een externe dienstverlener, of de daarvoor aangewezen grensgemeente of Schipholbalie.

⁵ De in het originele citaat genoemde Failure to Enroll (FTE) is niet relevant voor dit onderzoek.

Om de vraagstelling nader te kunnen specificeren leggen we eerst het begrippenkader nader vast.

1.3 Begrippenkader

Hier geven we nadere invulling aan de relevante begrippen voor zover die niet algemeen bekend verondersteld mogen worden.

1.3.1 Biometrische identiteitsverificatie

Bij biometrische identiteitsverificatie worden twee biometrische samples vergeleken om te bepalen of de biometrische kenmerken van beide samples volgens de criteria zodanig op elkaar lijken dat deze van dezelfde persoon afkomstig zouden moeten zijn.

Met een biometrisch sample worden biometrische kenmerken op enigerlei wijze vastgelegd, bijvoorbeeld een gezichtsafbeelding, een afbeelding van de iris, of een vingerafdruk. Het principe van biometrische identiteitsverificatie is dat samples van dezelfde persoon meer op elkaar zullen lijken dan samples van verschillende personen.

Bij biometrische identiteitsverificatie met een paspoort besluit men dus of een persoon die het paspoort toont (de toonder) wel of niet degene is op wiens naam het paspoort staat (de houder) op basis van een vergelijking van de biometrische kenmerken van de pasfoto (afgedrukt dan wel digitaal) en/of vingerafdruk (digitaal) op het paspoort – afkomstig van de houder - en de biometrische kenmerken van het gezicht en/of vinger(s) van de toonder.

Samples van dezelfde persoon zullen in meer of mindere mate op elkaar lijken, net als samples van verschillende personen in meer of minder mate zullen verschillen. Uiteindelijk moet er dus bepaald worden of de samples in voldoende mate op elkaar lijken en/of niet te veel verschillen om het besluit te rechtvaardigen dat deze van dezelfde persoon afkomstig zijn.

Een systeem voor geautomatiseerde biometrische verificatie zal veelal een score geven die aangeeft in welke mate de twee samples op elkaar lijken. Het besluit of de samples van dezelfde persoon of verschillende personen afkomstig zijn wordt dan gereduceerd tot het bepalen of een score hoog genoeg is. Dit komt dus neer op het bepalen van een minimum drempelwaarde om een paar biometrische samples aan te merken als “van dezelfde persoon”.

Twee biometrische samples die afkomstig zijn van dezelfde persoon noemen we een *mated pair*; een *non-mated pair* is (uiteraard) het tegendeel. Een pasfoto en een opname van de houder zijn dus een *mated pair*, terwijl diezelfde pasfoto met een opname van een andere persoon een *non-mated pair* is.

Een score die door een (geautomatiseerd) vergelijkingsalgoritme wordt gegeven met een (*non-*)*mated pair* is een (*non-*)*mated score*.

Een oordeel of twee biometrische samples van dezelfde persoon zijn is echter niet altijd correct.

1.3.2 Identificatiefouten

Bij biometrische (identiteits)verificatie kunnen twee typen fouten optreden:

- De *verificatie faalt onterecht*; bijvoorbeeld de houder toont zijn paspoort en wordt als een andere persoon aangemerkt.
Dit wordt een fout-negatieve uitslag of een *False Non-Match* genoemd
- De *verificatie slaagt onterecht*; bijvoorbeeld een andere persoon als de houder toont het paspoort maar wordt als de houder aangemerkt.
Dit is een fout-positieve uitslag, of een *False Match*.

Het deel of percentage van de personen waarbij de verificatie onterecht faalt wordt de *False Non-Match Rate* (FNMR) genoemd; dit staat voor een geautomatiseerde identiteitsverificatie

gelijk aan het deel of percentage van de mated scores dat beneden de drempel valt. Omdat dit deel van de personen bijvoorbeeld de toegang tot een aanvraagprocedure of doorgang bij een grenscontrole onterecht geweigerd kan worden, is deze zogenaamde “error rate” een maat voor *ongemak*.

Het deel van de personen waarbij de verificatie onterecht slaagt heet overeenkomstig *False Match Rate* (FMR)⁶. Voor een geautomatiseerde identiteitsverificatie staat dit gelijk aan het deel van de non-mated scores dat boven de drempel ligt.

Dit is het deel van de personen die op basis van een biometrische identiteitsverificatie toegang of doorgang krijgen terwijl zij niet de houder van het paspoort zijn en is daarom een maat voor de *(on)veiligheid*

De hoogte van deze *error-rates* verandert dus met de criteria voor vergelijking; stelt men de criteria strenger dan zullen er minder fouten worden gemaakt met fout-positieven, en potentieel meer met fout-negatieve resultaten; de FNMR neemt toe en de FMR neemt af. Voor minder strikte criteria, bijvoorbeeld een lagere score-drempel, geldt het omgekeerde. De criteria geven dus de mogelijkheid om ongemak en onveiligheid tegen elkaar af te wegen.

Men kan daarbij zowel de mate van overlast als de *(on)veiligheid* als leidend nemen; kiest voor veiligheid – dit is gebruikelijk bij evaluatie van biometrische verificatiesystemen voor grenspassage - dan bepaalt men criteria die corresponderen met een acceptabele waarde voor de FMR en bepaalt men de daarmee corresponderende FNMR⁷.

1.3.3 Definities

Aan de hand van het voorgaande kunnen we nu een aantal voor het onderzoek centrale begrippen nader definiëren.

De **(biometrische) ouderdom** definiëren we als het verschil in tijdstip waarop de biometrische samples genomen zijn die gebruikt worden voor de geautomatiseerde biometrische identiteitsverificatie. Een paspoort is 10 jaar geldig dus is deze voor identiteitsverificatie bij de aanvraag ca. 10 jaar⁸.

De **houdbaarheid** van een biometrie, is de maximale biometrische ouderdom waarvoor de geautomatiseerde biometrische verificatie nog betrouwbaar⁹ werkt.

De **herbruikbaarheid** van een biometrie is het geschikt zijn van het biometrische sample (d.w.z. vingerafdruk(ken) en/of gezichtsafbeelding) op een Nederlands paspoort voor hernieuwd gebruik bij de aanvraag van een Nederlands paspoort. Dit betekent dat bij toename van biometrische ouderdom met 10 jaar – of bij meerdere keren herbruik met 20 jaar of meer – de geautomatiseerde biometrische verificatie nog betrouwbaar moet werken.

⁶ De FNMR wordt ook soms aangeduid met False Rejection Rate (FRR) en de FMR met False Accept Rate (FAR).

⁷ Als voorbeeld: bij de geautomatiseerde grenscontrole op Schiphol Airport geldt bij een FMR <0.15% een FNMR < 4% gehaald moet worden wil een biometrisch verificatiesysteem voldoen.

⁸ Hoewel de afdruk van de pasfoto recent moet zijn is er op dit moment geen mogelijkheid voorzien om dit te controleren en kan men niet uitsluiten dat de foto geruime tijd voor het moment van verschijnen is genomen.

⁹ Er is geen algemene definitie of algemeen criterium voor betrouwbaarheid te geven omdat deze af hangt van de context en de keuze van het uitgangspunt van normering van de error rates (zie paragraaf 1.3.2).

Dit komt erop neer dat de houdbaarheid van de biometrische data groter is dan twee of meer malen 10 jaar (d.w.z. de geldigheidsduur van een paspoort).

1.4 Onderzoekskader

1.4.1 Beperkingen in scope

Dit onderzoek is beperkt tot geautomatiseerde biometrische identiteitsverificatie op basis van vingerafdrukken en/of de gezichtsafbeelding op een Nederlands paspoort. Identiteitsverificatie door bevoegd personeel is niet onderzocht vanwege het ontbreken van studies die relevant zijn voor de Nederlandse praktijk.

Dit is nadrukkelijk geen algemeen onderzoek naar de gevolgen van zogenaamde “aging” in biometrische identiteitsverificatie noch naar de huidige betrouwbaarheid van biometrische identiteitsverificatie in het algemeen en/of toepasbaarheid voor onderscheiden van tweelingen, deduplicatie etc. van biometrische herkenningmethoden.

Ook de gevolgen van leeftijd en biometrische ouderdom voor risico's met betrekking tot het uitbuiten van specifieke zwakheden van biometrische identiteitsverificatie, zoals bijvoorbeeld “morphing” en “spoofing”, zijn geen onderdeel van dit onderzoek.

1.4.2 Onderzoeksmethode

Omdat eigen onderzoek een grote inspanning zou vergen op het gebied van verzameling van data en testen van algoritmen, is gekozen voor literatuuronderzoek. Bij het zoeken naar ter zake doende publicaties spelen drie criteria een rol. De publicaties moeten:

1. Een beeld geven van de effecten van biometrische ouderdom van 10 en meer jaar
2. Gebruik maken van biometrische samples (d.w.z. vingerafdrukken en/of gezichtsafbeeldingen) die vergelijkbaar zijn met de samples die worden gebruikt bij de biometrische identiteitsverificatie op basis van een Nederlands paspoort (bijv. een vergelijking van een pasfoto met een ter plekke gemaakte opname van het gezicht van toonder)
3. Vergelijkbare State-Of-The-Art (SOTA) herkenningalgoritmes als gebruikt bij identiteitsverificatie op basis van een Nederlands paspoort onderzoeken.

Indien studies niet geheel aan deze criteria voldoen zullen hun resultaten aan de hand van de criteria op relevantie worden beoordeeld.

1.5 Specificering Vraagstelling

Op basis van al het voorgaande wordt de initiële vraagstelling (uit paragraaf 1.2) nader gespecificeerd:

1. Tot welke biometrische ouderdom daalt de betrouwbaarheid van de geautomatiseerde biometrische identiteitsverificatie door middel van een Nederlands paspoort niet significant¹⁰? Kan de daling in betrouwbaarheid met biometrische ouderdom betrouwbaar gekwantificeerd worden in de vorm van toegenomen False Non-Match Rates (FNMR) en False Match Rates (FMR)?

¹⁰ Deze biometrische ouderdom staat gelijk aan de houdbaarheid.

2. Brengt een toename van de biometrische ouderdom met 10 of meer jaar een significante daling van de betrouwbaarheid van de geautomatiseerde biometrische identiteitsverificatie door middel van een Nederlands paspoort met zich mee? Kan deze daling in betrouwbaarheid betrouwbaar gekwantificeerd worden in de vorm van toegenomen False Reject Rates (FNMR) en False Match Rates (FMR)?

De twee vragen hebben betrekking op de opgeslagen gelaatsfoto en vingerafdruk op het identiteitsdocument. Verder heeft de 2e vraag m.n. betrekking op personen met een leeftijd boven de 65 jaren.

1.6 Leeswijzer

In dit rapport worden onze bevindingen per biometrie weergegeven; Hoofdstuk 2 geeft deze wat betreft *gezichtsvergelijking* en hoofdstuk 3 wat betreft *vingerafdrukken*.

2 GEZICHTSVERGELIJKING

2.1 Gezicht als biometrie

Het gebruik van het gezicht voor identiteitsverificatie berust in eerste instantie op het feit dat elke persoon een gezicht heeft. Het gezicht is echter niet immuun voor verandering; elk gezicht veroudert in een complex proces waarin genetische factoren, de omgevingsomstandigheden en levensstijl een rol spelen. Daarnaast kan een gezicht in korte tijd sterk veranderen doordat de weefsels van het gelaat beschadigd worden bijvoorbeeld door een ongeval of door medisch ingrijpen [1].

Hoewel gemeenschappelijke etniciteit of (voor)ouders kan leiden tot minder uitgesproken verschillen in biometrische eigenschappen van het gezicht [2] – de zogenaamde intervariabiliteit – kunnen omgevingsfactoren en levensstijl die verschillen weer groter maken.

Doordat het gezicht aan verandering onderhevig is gedurende het leven, veranderen de biometrische kenmerken daarin ook, en is het voor het betrouwbare gebruik van gezichtsafbeeldingen voor biometrie noodzakelijk deze periodiek te vernieuwen [3].

Gegeven controleerbare omstandigheden is het met de beschikbare moderne digitale camera's relatief eenvoudig om gezichtsafbeeldingen van goede kwaliteit te verkrijgen.

In welke mate deze algemene constatering de houdbaarheid en herbruikbaarheid van gezichtsafbeeldingen bij geautomatiseerde biometrische identiteitsverificatie beperkt, wordt in de hieropvolgende paragrafen bekeken aan de hand van de beschikbare literatuur over de invloed van biometrische ouderdom, zowel op zichzelf staand (paragraaf 2.2) als in combinatie met leeftijd (paragraaf 2.3).

2.2 Invloed van biometrische ouderdom op betrouwbaarheid

Er is consensus in de literatuur over dat de betrouwbaarheid van geautomatiseerde gezichtsvergelijking afneemt als de ouderdom van de biometrie toeneemt (zie bijv. [4]). Aan de hand van recente publicaties onderzoeken we hieronder in hoeverre dit blijkt uit recente resultaten, en in hoeverre het mogelijk is om de betrouwbaarheid in termen van FNMR en FMR te kwantificeren.

2.2.1 Invloed biometrische ouderdom op mated scores

In [5] maakt men een analyse van de mated score van een commercieel systeem¹¹ (Commercial Off The Shelf; COTS) als functie van (onder andere) de biometrische ouderdom.

In Figuur 1 (uit [5]) staat de gemiddelde mated score¹², als functie van de biometrische ouderdom, weergegeven als een zwarte lijn. De spreiding van de scores is weergegeven met blauwe en groene banden¹³. Het verloop van deze kenmerken is geëxtrapolerd naar een biometrische ouderdom van 25 jaren.

¹¹ Het systeem (hier "COTS-A") maakte deel uit van de top-3 van de Face Recognition Vendor Test (FRVT) van de National Institute of Standards and Technology (NIST) 2014 die bestond uit NEC, Morpho en Toshiba [17]; deze COTS is dus op moment van publicatie als SOTA aan te merken.

¹² De score is hier gestandaardiseerd; dit is een wiskundige bewerking waardoor het gemiddelde gelijk wordt aan 0, en de standaarddeviatie gelijk aan 1. Met deze bewerking kunnen scores van verschillende algoritmen eenvoudiger vergeleken worden; scores worden vaak op zeer verschillende manieren berekend, waardoor hun betekenis niet eenduidig en dus ook niet vergelijkbaar is.

¹³ Dat de spreiding niet (zuiver) toeneemt met de biometrische ouderdom is o.a. het gevolg van het feit dat de gemiddelde ouderdom van de hier gebruikte dataset ongeveer 5 jaar is. Voor die waarde is er het minste spreiding, hoe meer de biometrische ouderdom daar vanaf wijkt hoe groter de spreiding.

Tevens staan de score-drempelwaarden voor twee mogelijk relevante normeringen van de FMR afgebeeld in rood.

Deze figuur laat zien dat de gemiddelde mated score – en daarmee de gelijkheids - afneemt met de biometrische ouderdom.

Zoals vermeld in paragraaf 1.3.2, staat de FNMR gelijk aan het deel van de mated scores die onder de drempelwaarde valt. In Figuur 1 geldt bovendien dat het percentage mated scores dat, (voor een gegeven biometrische ouderdom) onder de onderste blauwe, respectievelijk groene begrenzing ligt gelijk is aan 2.5% resp. 0.5%.

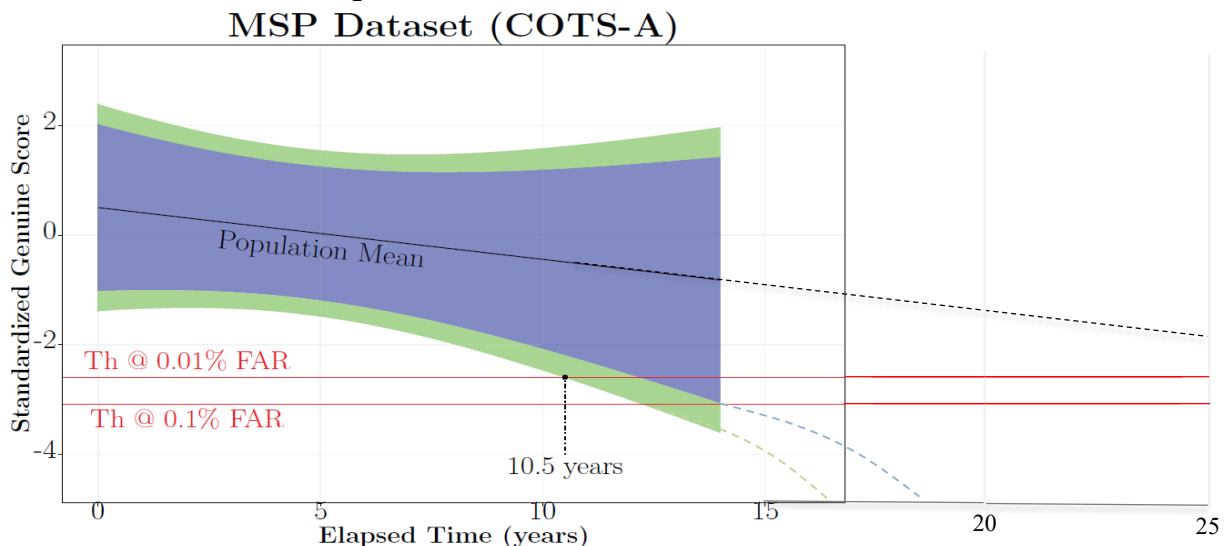
Als de rode lijn – de drempelwaarde – een van de ondergrenzen snijdt dan weten we dus de FMR, en is de FNMR gelijk aan de voornoemde percentages.

Een FMR van 0.01% betekent dat gemiddeld één op de 10.000 personen die een paspoort gebruiken waarvan zij niet de houder zijn (d.w.z. met een gezichtsafbeelding die niet met hen overeen komt) zal slagen in zijn opzet. Dit wordt in het kader van de evaluatie van biometrische verificatie bij grenspassage gezien als een relevante normering.

Bij de drempelwaarde voor de score die correspondeert met deze FMR, wordt de FNMR gemeten. Blijft deze onder een aanvaardbaar maximum, dan voldoet een systeem aan de voorwaarden voor gebruik.

Een FNMR van 2.5% wordt in deze context over het algemeen als aanvaardbaar beschouwd. Onder deze aannames, is de (biometrische) houdbaarheid, met dit algoritme en met deze dataset, te bepalen op maximaal 12 jaren.

Bekijken we de extrapolaties van gemiddelde en beide variatie-curves, dan zien we dat er bij een biometrische ouderdom van 20 jaren een FNMR te verwachten is die minimaal 15% bedraagt. Dit is een onaanvaardbaar hoge waarde.



FIGUUR 1: GESTANDAARDISEERDE MATED SCORE TEGEN OUDERDOM BIOMETRIE

(Bron: [5]) De gemiddelde score wordt weergegeven met de zwarte lijn. De spreiding is weergegeven met de gekleurde balken rond het gemiddelde. 2.5% respectievelijk 0.5% van de scores ligt onder het blauwe gebied resp. onderste groene rand.

Extrapolaties van gemiddelde score, en de spreiding zijn weergegeven in gestippelde lijnen van corresponderende kleur.

Extrapolatie is verricht door de auteurs van dit rapport.

Met de overwegingen uit paragraaf 1.4.2 zien we dat deze studie alle benodigde informatie levert om de houdbaarheid te schatten (via de FNMR en FMR).

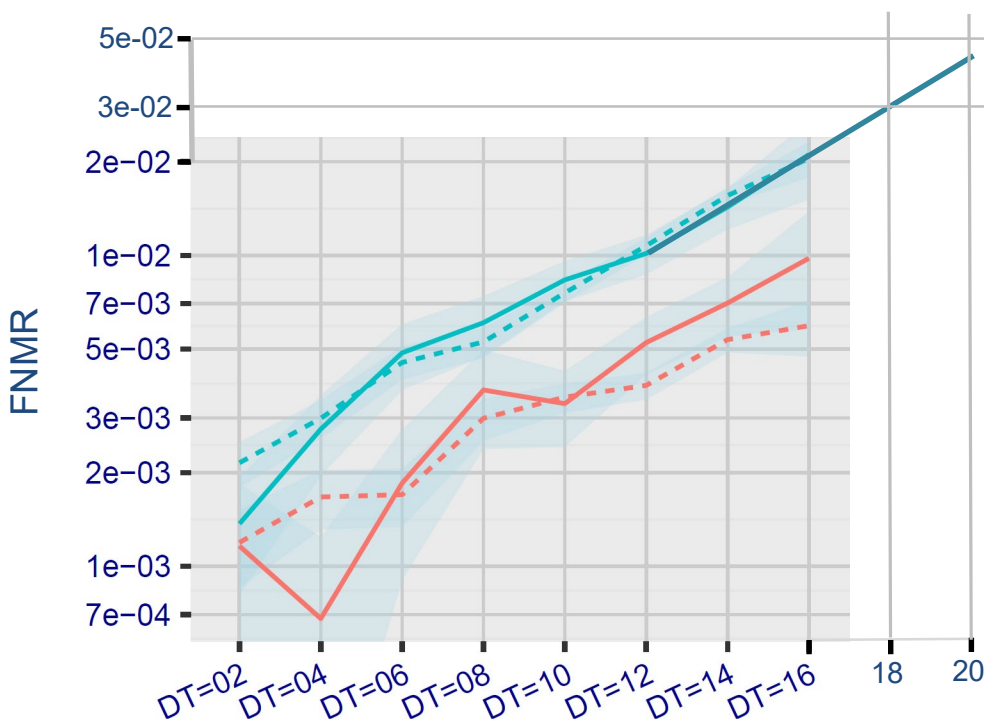
Dit is echter niet de huidige SOTA (maar die van 2014), er worden resultaten van slechts één algoritme getoond, en het type gezichtsafbeeldingen dat gebruikt wordt is niet representatief voor

gezichtsafbeeldingen die gebruikt worden voor geautomatiseerde identiteitsverificatie op basis van een Nederlands paspoort.

Men heeft hier namelijk gebruik gemaakt van “mugshots”; opnamen van verdachten bij aanhouding of in verzekeringstelling. Deze opnames zijn frontaal genomen onder gecontroleerde omstandigheden¹⁴. Dat zorgt voor lagere error-rates bij een vergelijking tussen dit type afbeeldingen, ten opzichte van een vergelijking tussen de gezichtsopname die van een paspoortchip is uitgelezen en een ter plekke genomen opname van de toonder zoals bij geautomatiseerde identiteitsverificatie plaatsvindt.

2.2.2 SOTA FNMR

Uit de gegevens van de meest recente rapportage van de NIST Face Recognition Vendor Test (FRVT) [6] blijkt dat de SOTA, bij een tien keer lagere FMR als in [5] (0.001% tegen 0.01%), en een hogere biometrische ouderdom (tot 16 jaar) tot een vergelijkbare maximale FNMR in staat is (tot ~ 2%); zie Figuur 2 voor een voorbeeld uitgesplitst naar geslacht en twee etniciteiten. We verwachten bovendien dat de FNMR lager is bij een (hogere) FMR zoals die bij toepassingen voor geautomatiseerde identiteitsverificatie wordt aangehouden.



FIGUUR 2: FNMR TEGEN BIOMETRISCHE OUDERDOM (FMR = 0.001%)

Een voorbeeld (uit [6]: figuur 155, pag. 188) van het verloop van de FNMR bij gegeven FMR voor een SOTA algoritme. Biometrische ouderdom (DT) wordt gegeven in jaren.

Hier is te zien dat de FNMR ongeveer met een factor 10 toeneemt over deze periode.

De vier grafieken geven de resultaten van verschillende sub-populaties:

Steenrood: African American Zeegroen: Blank

-----: Man _____: Vrouw

Extrapolatie gemaakt in vergelijkbare kleur en format; slechts de grafiek met het hoogste te verwachten eindpunt is geëxtrapoleerd.

Extrapolatie is verricht door de auteurs van dit rapport.

¹⁴ Er wordt meer variatie toegelaten in de stand van het gezicht ten op zichte van de camera (“pose”) en belichting dan is toegestaan voor paspoortfoto’s.

Bij extrapolatie van de “worst case” grafiek blijkt dat er voor een biometrische ouderdom van 20 jaren een FNMR te verwachten is tussen de 3.5% en 4.5%. Dit bereik aan waarden is over het algemeen niet aanvaardbaar voor geautomatiseerde identiteitsverificatie op basis van een paspoort.

Voor deze analyse is opnieuw gebruik gemaakt van mugshots. Dat maakt dat we er niet vanuit kunnen gaan dat de genoemde FMR- en FNMR-waarden (en hun verandering met biometrische ouderdom) representatief zijn voor de SOTA bij een vergelijking tussen een pasfoto en een ter plekke genomen opname.

Er is in de studie wel een grote dataset gebruikt om meerdere algoritmen van commerciële leveranciers te testen; dit is daadwerkelijk de SOTA voor gezichtsvergelijking.

2.3 Invloed biometrische ouderdom per leeftijd(scategorie)

De voorgaande resultaten geven slechts een “gemiddeld” beeld van de verandering in betrouwbaarheid van geautomatiseerde identiteitsverificatie door middel van gezichtsvergelijking met biometrische ouderdom. De gegevens die hiervoor zijn gepresenteerd waren niet uit te splitsen naar leeftijd of leeftijdscategorie.

Toch is het verschil tussen leeftijdsgroepen een belangrijk onderdeel van dit onderzoek; het gezicht is namelijk aan veranderingen onderhevig als gevolg van genetische en omgevingsfactoren en levensstijl. Die factoren zullen verschillen van persoon tot persoon, maar voor leeftijdscategorieën is er een grote lijn te verwachten.

Daarom bespreken we hieronder onderzoek dat leeftijd of leeftijdscategorieën heeft meegenomen als parameter in de analyse van de betrouwbaarheid van gezichtsvergelijking.

2.3.1 SOTA waarden voor FMR en FNMR

In het kader van NIST FRVT 2019 [7] is de invloed van leeftijdscategorie op FMR en FNMR uitgebreid bestudeerd [7]. In het rapport is de FMR binnen en tussen leeftijdsgroepen weergegeven als een heatmap (zie Figuur A1 in de appendix); de FNMR is weergegeven per leeftijdsgroep¹⁵.

2.3.1.1 FMR

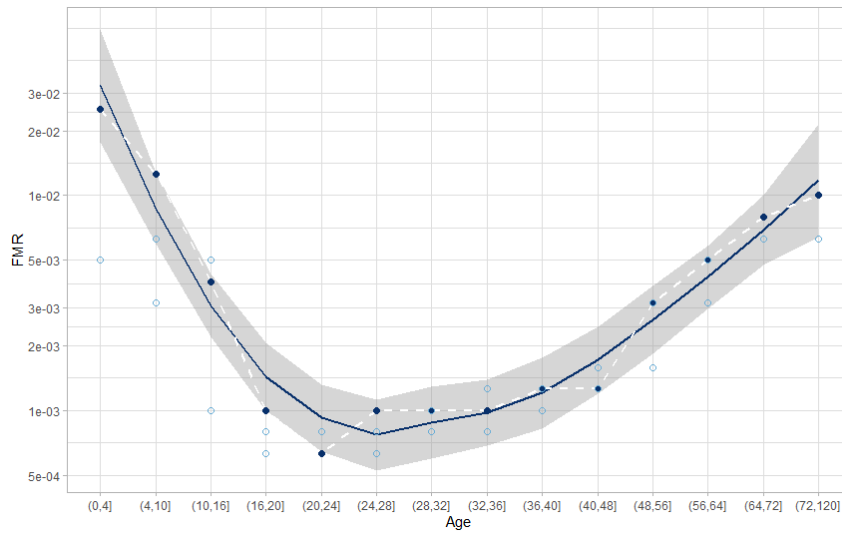
Voor de FMR-bepaling is de scoredrempel is zodanig gekozen dat de gemiddelde FMR gelijk is aan 0.01%¹⁶.

In Figuur 3 staat de FMR weergegeven per leeftijdscategorie zowel voor vergelijkingen binnen de groep als met direct aangrenzende categorieën, voor individuen van gelijk geslacht, en geografische origine.

De curve laat zien dat jonge kinderen relatief sterk op elkaar lijken, maar deze gelijkenis ook sterk afneemt met leeftijd. Voor jong volwassenen tot middelbare leeftijd is de FMR laag, om dan weer te stijgen voor hogere leeftijd. Met name de stijging voor een hogere leeftijd is onverwacht sterk in het licht van de hiervoor besproken resultaten van de FRVT van ± 5 jaar eerder.

¹⁵ Er zijn alleen FNMR waarden tussen leeftijdscategorieën mogelijk als van dezelfde personen op meerdere tijdstippen gezichtsafbeeldingen zijn gemaakt en opgenomen in de dataset.

¹⁶ Men heeft hiervoor alle afbeeldingen vergeleken met alle anderen, zonder onderscheid naar leeftijd, geslacht of etnische afkomst te maken. Dit onderschat de FMR zoals ook wordt geconstateerd in [7] paragraaf 4.2 .



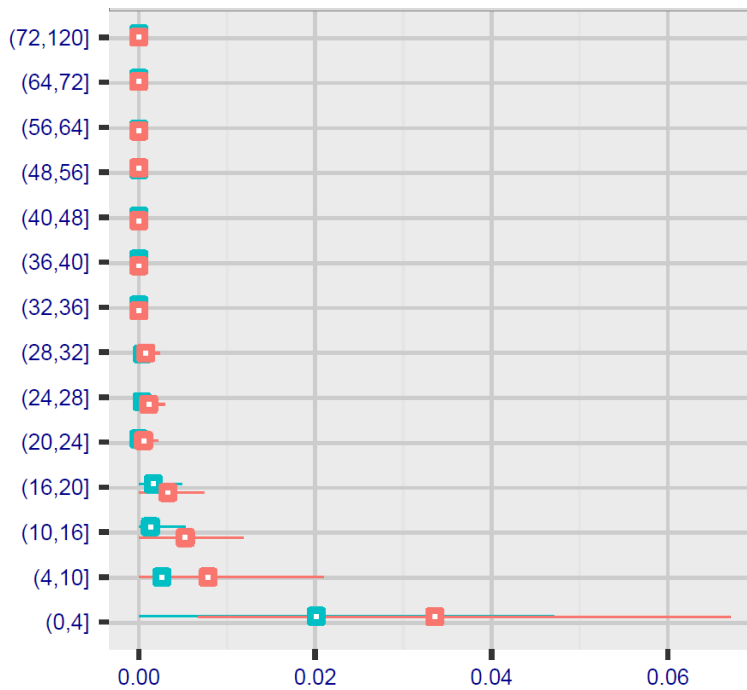
FIGUUR 3: FMR PER LEEFTIJDGROEP

(Naar Figuur A1; uit: [7])

Score-drempel ingesteld voor FMR 0.01% ¹¹

De FMR binnen de leeftijdscategorie is weergegeven als een donkerblauwe punt verbonden met een witte stippellijn. De FMR met aangrenzende leeftijdscategorieën staan weergegeven als een lichtblauwe cirkel.

Er is een gladde lijn getrokken die het verloop van de FMR binnen de leeftijdscategorie benadert, met in grijs de spreiding.



FIGUUR 4: FNMR NAAR LEEFTIJDGROEP

FNMR bij twee FMR-waarden:

Steenrood: FMR = 0.0001 (0.01%) Zeegroen: FMR = 0.001 (0.1%)

De FNMR voor leeftijden boven de 32 jaar zijn niet te onderscheiden op deze grafiek, en wijken niet zichtbaar af van nul.

Bron: [6] figuur 177 pag. 211

De FMR-waarden met aangrenzende leeftijdscategorieën (lichtblauwe cirkels) zijn, op vier uitzonderingen na, allen kleiner of gelijk aan de FMR binnen de corresponderende categorie (donkerblauwe punten). Voor de SOTA algoritmen maakt een leeftijdsverschil tussen de personen op de twee gezichtsopnamen die vergeleken worden het dus makkelijker om de misbruiker te onderscheiden.

Voor individuen boven de 20 jaren oud blijven de FMR-waarden met aangrenzende categorieën grotendeels binnen de spreidingsband; dit betekent dat ze op statistische gronden niet wezenlijk verschillend zijn. Boven de 48 jaren, vallen deze waarden incidenteel buiten de band, onder de 20 jaren vrijwel allemaal. Dit wijst erop dat leeftijdsverschil een grotere invloed heeft bij jongere personen; in vergelijking met 48+, zal de FMR bij hen ook sneller afnemen bij hetzelfde leeftijdsverschil.

2.3.1.2 FNMR

Figuur 5 geeft een voorbeeld van SOTA FNMR-waarden bij twee relevante FMR-normeringen voor dezelfde leeftijdsgroepen als in Figuur 3. Deze zijn afkomstig van de meest recent gepubliceerde NIST FRVT Ongoing rapportage [6].

De FNMR is voor leeftijden boven de 32 niet te onderscheiden van nul, en neemt daaronder zichtbaar toe tot waarden van enkele procenten voor de jongste kinderen.

Kanttekening bij deze gegevens is dat de afbeeldingen die hier vergeleken worden gezichtsafbeeldingen van visa aanvragen zijn. Deze afbeeldingen zullen over het algemeen van een betere kwaliteit zijn dan de ter plekke genomen opnamen die gebruikt worden bij geautomatiseerde identiteitsverificatie, maar vergelijkbaar van kwaliteit aan paspoort-opnamen. De FNMR zal dus waarschijnlijk onderschat worden in dit geval.

2.4 Deelconclusie Gezichtsvergelijking

De FNMR bij relevante normeringen van de FMR zal toenemen met de biometrische ouderdom (Figuur 2). Dat wil zeggen: bij een gelijk niveau van veiligheid (weinig fout-positieven, lage FMR) zal het ongemak toenemen (meer fout-negatieven; hogere FNMR).

De beperkingen van de beschikbare gegevens maken dat wij geen betrouwbare kwantitatieve uitspraak kunnen doen over de mate waarin de FNMR toeneemt met de biometrische ouderdom, voor SOTA gezichtsvergelijkingssystemen, bij gebruik van gezichtsafbeeldingen die representatief zijn voor geautomatiseerde biometrische identiteitsverificatie door middel van gezichtsafbeeldingen zoals die op Nederlandse paspoorten worden vastgelegd.

De FNMR bij relevante normeringen voor de FMR varieert met name voor kinderen en jong volwassenen met leeftijd. Dit betekent dat bij een gelijk niveau van veiligheid (klein aantal fout-positieven; kleine FMR) voor personen tot een leeftijd van ongeveer 12 jaren het ongemak (meer fout-negatieven; hogere FNMR) relatief groter zal zijn in vergelijking met volwassenen. Bij gebruik van SOTA gezichtsvergelijkingssystemen is de FNMR voor kinderen tot 4 jaar in de orde van maximaal enkele procenten, maar daalt deze snel met leeftijd om vanaf 32 jaar tot hoge leeftijd minimaal te blijven (Figuur 5).

De beperkingen in deze resultaten maken dat wij geen betrouwbare uitspraak kunnen doen over de mate waarin de FNMR toeneemt met biometrische ouderdom per leeftijdscategorie.

Ook de FMR per leeftijdscategorie, bij een relevante normering voor de gemiddelde FMR waarde, varieert met leeftijd. Zowel kinderen (jonger dan 12 jaar) als ouderen (vanaf 65 jaren) zullen bij een gelijk niveau van veiligheid voor alle groepen bij elkaar (weinig fout-positieven, lage FMR) een groter veiligheidsrisico geven (groter aantal fout-positieven, hogere FMR) dan volwassenen.

Voor SOTA gezichtsvergelijkers wordt de hoogste waarde van de FMR gevonden voor de

jongste kinderen, om dan scherp te dalen tot een minimum tussen 20 en 50 jaar, en dan te stijgen voor hogere leeftijden (Figuur 3).

De beperkingen in deze resultaten maken dat wij geen betrouwbare uitspraak kunnen doen over de mate waarin de FMR toeneemt met biometrische ouderdom per leeftijdscategorie.

Op basis van alle gegevens in samenhang blijkt dat er consensus is over dat de nauwkeurigheid voor geautomatiseerde identiteitsverificatie afneemt met een toename van de biometrische ouderdom en ook met de leeftijd van de persoon vanaf middelbare leeftijd. Er is ook consensus over dat dit voornamelijk veroorzaakt wordt door veranderingen als gevolg van natuurlijke ontwikkeling van het gezicht en veroudering.

Voor een toename van de biometrische ouderdom met 1 jaar zijn die effecten en als gevolg het risico beperkt. Voor een toename met 10+ jaren is er een aanzienlijk effect en risico.

Deze resultaten zijn echter verkregen uit vergelijkingen van gezichtsafbeeldingen die niet representatief zijn voor de vergelijking zoals die bij geautomatiseerde identiteitsverificatie op basis van een Nederlands paspoort - tussen een pasfoto en een tweede sample – plaatsvindt. Een deel van de resultaten zijn bovendien verkregen door toepassing van algoritmen die niet SOTA zijn.

Deze beperkingen van de gegevens maken dat wij geen betrouwbare kwantitatieve uitspraken kunnen doen over de invloed van biometrische ouderdom op FNMR en/of FMR, noch over de houdbaarheidstermijn of herbruikbaarheid van digitaal op Nederlandse paspoorten opgeslagen gezichtsafbeeldingen voor geautomatiseerde identiteitsverificatie op basis van een paspoort, in het algemeen noch voor leeftijdscategorieën in het bijzonder.

3 VINGERAFDRUKKEN

3.1 Vingerafdrukken als biometrie

Het patroon van papilairlijnen op de vingers blijft onveranderd gedurende het hele leven [8]. Zowel deze permanentie als het onderscheidend vermogen van vingerafdrukken is bevestigd in de praktijk over meer dan 100 jaar [9]. Bovendien bevestigt wetenschappelijk onderzoek dat het patroon niet verandert [10]¹⁷.

Problemen in de interactie tussen vinger en sensor zorgt echter voor vermindering van de kwaliteit van vingerafdrukken. Dit speelt met name bij jonge kinderen (onder ±12 jaren oud), en voor ouderen (65+).

Jonge kinderen hebben zodanig kleine vingers dat de papilairlijnen en daarmee de biometrische kenmerken door de gebruikte scanners niet goed kan worden vastgelegd.

Voor ouderen geldt dat de vingers van de juiste grootte zijn, maar de karakteristieken van de huid veranderen waardoor het vastleggen van de papilairlijnen en daarmee de biometrische kenmerken bemoeilijkt wordt. Met name neemt de elasticiteit af en wordt de huid droger.

3.2 Invloed biometrische ouderdom op mated scores

In de afgelopen 15 jaar hebben meerdere studies ([11] [12] [13]) geconstateerd dat de mated scores van geautomatiseerde vingerafdrukvergelijking afnemen als de tijd tussen de afbeeldingen (d.w.z. de biometrische ouderdom) toeneemt. Figuur 5 (uit [13]) geeft die bevinding duidelijk weer. Wat daarvan de oorzaak is, daarover bestaat minder consensus; men ziet over het algemeen een rol weggelegd voor leeftijd, biometrische ouderdom en kwaliteit van de afbeeldingen¹⁸. Er zijn zelfs auteurs die stellen dat biometrische ouderdom als oorzaak voor de afname van mated scores in het niet valt bij kwaliteit [14].

Analoog aan de beschouwing van Figuur 1 in paragraaf 2.2.1 zou men met Figuur 5 kunnen trachten een schatting te maken van de FNMR bij een FMR van 0.01%. Omdat een visualisatie van de spreiding ontbreekt beperken wij ons hier met de constatering dat de gemiddelde score de score-drempelwaarde bereikt bij een biometrische ouderdom van 30 jaar wat inhoudt in dat de FNMR gelijk is aan 50%.

Recent onderzoek [15] [16] analyseert de invloed van kwaliteit van vingerafdrukken, veroudering en biometrische ouderdom op mated scores.

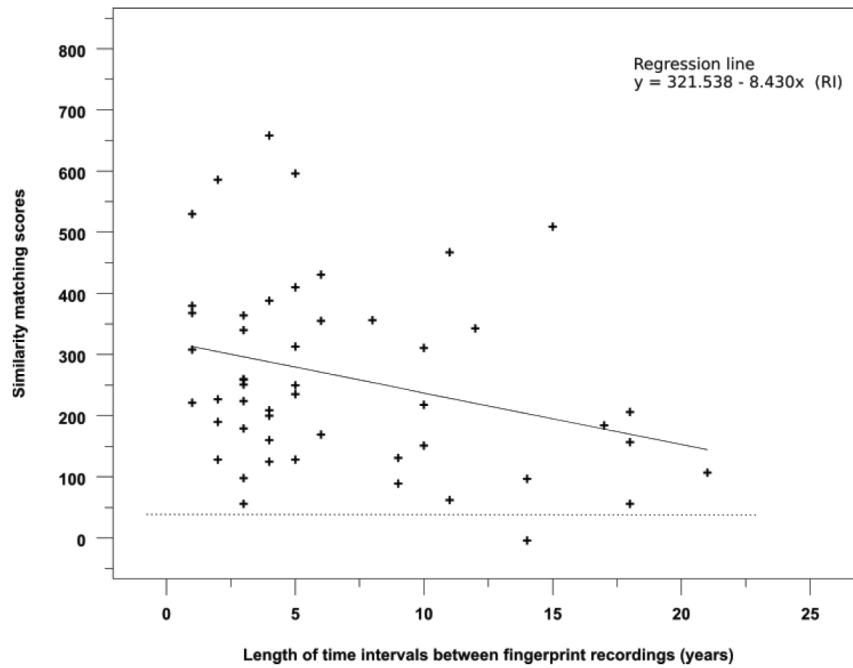
Er is bij dit onderzoek gebruik gemaakt van een dataset met vingerafdrukken die verzameld zijn door de Portugese overheid, in het bijzonder de Passport and Immigration Service (SEF). Deze vingerafdrukken werden afgenomen voor het uitgeven van biometrische paspoorten, waarvoor men de afdrukken van beide wijsvingers registreert¹⁹.

¹⁷ Een uitzondering is beschadiging van de zogenaamde germinatieve laag, waardoor na herstel littekens ontstaan die de papilairlijnen blijvend veranderen. Dit geeft dus een nieuw vanaf dan onveranderlijk kenmerk.

¹⁸ Bij kwaliteit refereert men aan metriecken als NFIQ2 een initiatief van NIST; zie: <https://www.nist.gov/services-resources/software/development-nfiq-20> en VERIQ

¹⁹ Als een wijsvinger ontbreekt wordt er een andere vinger gescand.

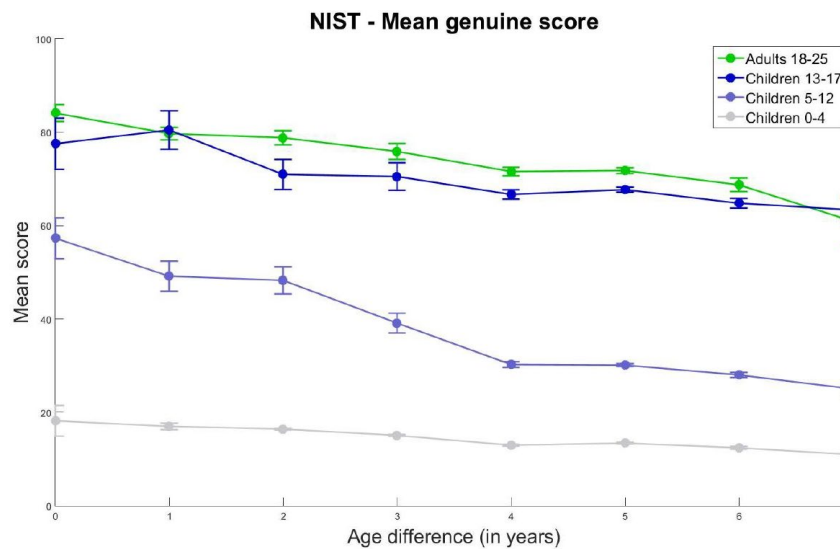
Voor een honderdtal personen zijn er vingerafdrukken beschikbaar die met enkele jaren tussenruimte genomen zijn met een maximum van 7 jaar.
 Er zijn verschillende leeftijdscategorieën gemaakt waarvoor de mated score is bepaald als functie van de biometrische ouderdom. Figuur 6 toont de grafiek voor de jonge leeftijdsgroepen met de volwassenen als referentie.



FIGUUR 5: GEMIDDELDE SCORE VERSUS TIJDSVERLOOP

Scatterplot van individuele waarden van score tegen biometrische ouderdom. Ononderbroken lijn is regressielijn van gemiddelde waarde, stippellijn is de score-drempelwaarde voor een FMR van 0.01%

Bron: [13]



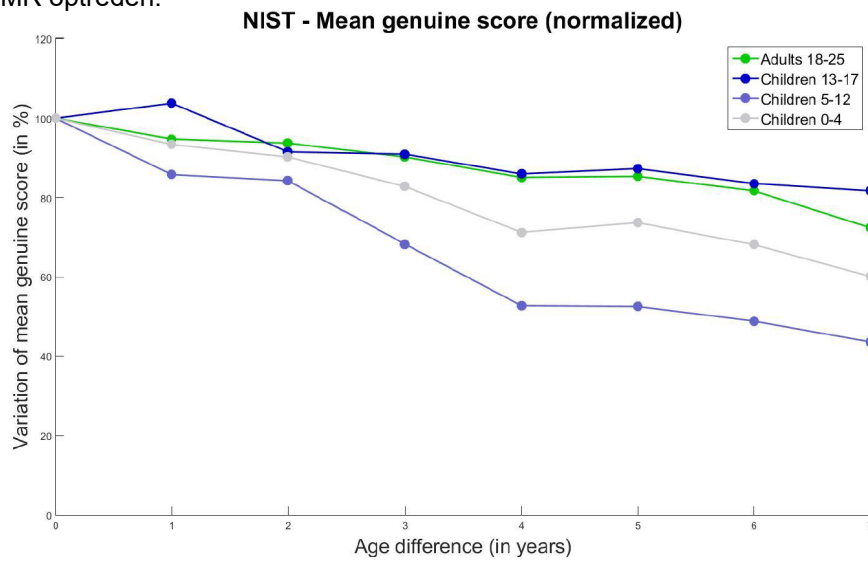
FIGUUR 6: MATED SCORES KINDEREN TEGEN BIOMETRISCHE OUDERDOM

Mated scores voor subgroepen van kinderen.

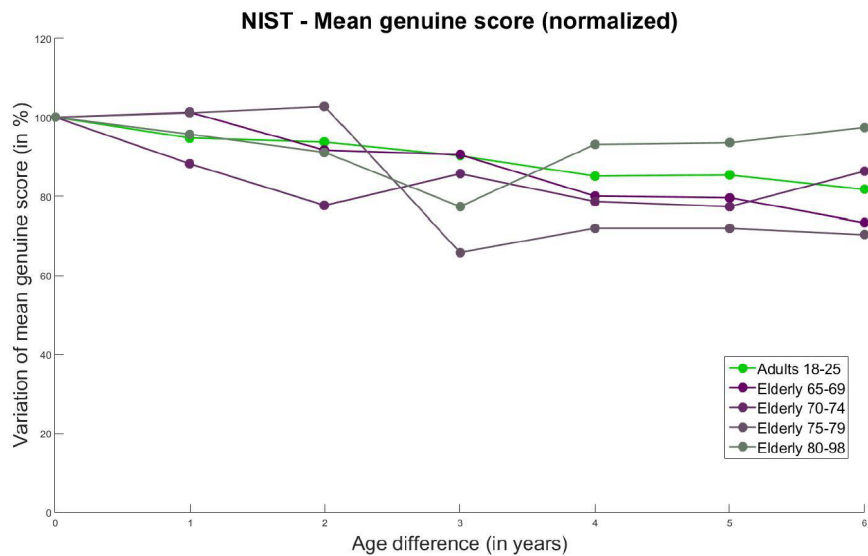
Mated scores van volwassenen staan in groen als referentie

De dalende lijn is beter zichtbaar als we de waarden normaliseren; alle waarden worden gedeeld door de score voor gelijktijdig genomen vingerafdrukken (op de verticale as; biometrische ouderdom nul). Nu is goed te zien dat de twee jongste categorieën een veel sterkere daling in score laten zien dan de 13 tot 17-jarigen, die sterk op de volwassen categorie (18-25) lijken.

De genormaliseerde scores van de oudere leeftijdscategorieën (Figuur 8; absolute scores zie Figuur A2) lijken een vergelijkbaar verloop te hebben met de biometrische ouderdom als die van volwassenen. Daarbij moet men wel voor ogen houden dat de absolute scores van 'ouderen' aanzienlijk lager zijn dan voor volwassenen; er zal bij dezelfde score-drempel nog steeds een hogere FNMR optreden.



FIGUUR 7: GENORMALISEERDE SCORES JONGERE LEEFTIJDSCATEGORIEËN
 Als Figuur 6 maar genormaliseerd met de waarde bij biometrische ouderdom nul



FIGUUR 8: GENORMALISEERDE SCORES OUDERE LEEFTIJDSCATEGORIEËN

3.3 Invloed leeftijd op mated scores

Ook de leeftijd zelf heeft invloed op de hoogte van mated scores. De variatie van gemiddelde mated score met leeftijd (Figuur 9) laat zien dat zowel jonge kinderen als de 65+ groepen een aanzienlijk lagere gemiddelde score hebben.

Zoals vermeld in paragraaf 3.1, blijft het patroon van papilairlijnen onveranderd gedurende het hele leven. Er moeten dus andere factoren zijn waardoor vingerafdrukvergelijking bij zowel kinderen als ouderen minder goed werkt.

Bij jonge mensen zijn de vingers kleiner en de papilairlijnen dus ook. Dit kan leiden tot kwaliteitsverlies bij het scannen van vingerafdrukken omdat de resolutie van de scanner voldoende hoog moet zijn om het patroon te kunnen vastleggen. Bovendien zijn vingervergelijkingsalgoritmen over het algemeen ontwikkeld op basis van afdrucken van volwassenen, waardoor ze minder makkelijk de kleinere afdrucken van kinderen kunnen verwerken.

Voor vingerafdrukken afgenomen op verschillende (jonge) leeftijden zullen de variaties in de afbeeldingen relatief hoog zijn in vergelijking met twee afbeeldingen op volwassen leeftijd waardoor de mated score afneemt met biometrische ouderdom.

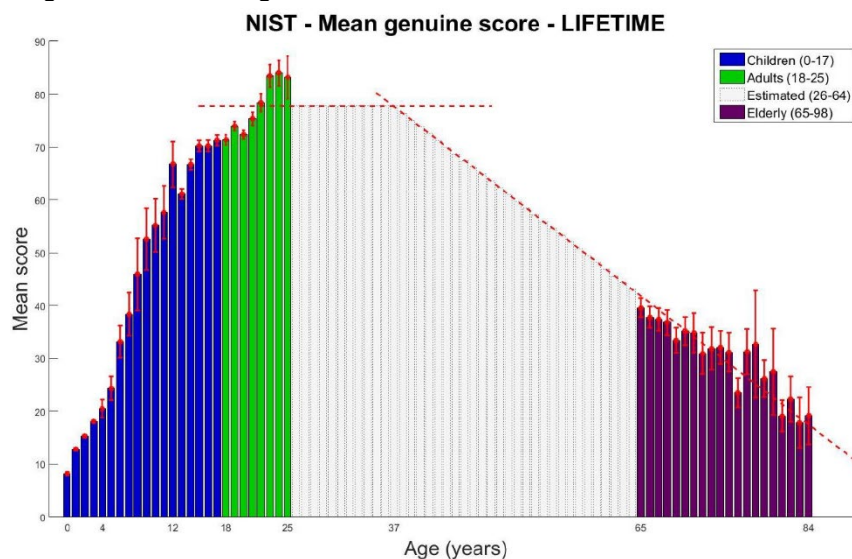
Omdat vingers natuurlijk groeien met de leeftijd verklaart dit ook dat de mated scores bij het naderen van volwassenheid ook de scores van volwassenen benaderen.

Mated scores kunnen verbeterd worden door voor jonge leeftijden de vingers op een hogere resolutie te scannen, en voor kwaliteitscontrole en vingervergelijking gebruik te maken van specifiek voor kleine afdrucken ontwikkelde software.

Voor ouderen geldt dat hun vingerafdrukken van voldoende grootte zijn. Bij hen speelt echter mee dat de huid van de vingers op hoge leeftijd minder elastisch en droger wordt, waardoor scanners de patronen moeilijker kunnen vastleggen en er dus een verlies in kwaliteit optreedt. Dit beïnvloedt de mated score dus in een negatieve zin en in toenemende mate met leeftijd.

Voor vingerafdrukken afgenomen op verschillende (oudere) leeftijden zullen de variaties in de afbeeldingen relatief hoog zijn in vergelijking met twee afbeeldingen op volwassen leeftijd waardoor de mated score eveneens afneemt met biometrische ouderdom.

Bij personen op hoge leeftijd (65+) zouden de mated scores verbeteren door scanners te gebruiken op basis van technologisch principes waarvoor geldt dat het vastleggen van patronen minder gevoelig is voor de vochtigheid en elasticiteit van de huid.



FIGUUR 9: MATED SCORE TEGEN LEEFTIJD

Bron: [15]

3.4 Deelconclusie vingerafdrukken

Op basis van de daling in mated scores (Figuur 5 - Figuur 8) is de verwachting dat de FNMR bij relevante normeringen van FMR zal gaan stijgen met biometrische ouderdom. Op basis van de gemiddelde mated scores (Figuur 9) zal de FNMR bij deze FMR-normeringen hoger zijn voor zowel kinderen (0- 12 jaar) als 65+.

Op basis van de gegevens in samenhang lijkt de afname in betrouwbaarheid van biometrische verificatie op basis van vingerafdrukken voor een toename van biometrische ouderdom van ongeveer 1 jaar beperkt, voor een toename van 10+ jaar aanzienlijk.

Deze resultaten zijn echter verkregen uit vergelijkingen van vingerafdrukken die onvoldoende representatief zijn voor de vergelijking tussen een digitaal opgeslagen afdruk en een ter plekke afgenomen vingerafdruk, zoals die bij geautomatiseerde identiteitsverificatie op basis van een Nederlands paspoort plaatsvindt. Een deel van de resultaten zijn bovendien verkregen door toepassing van algoritmen die niet SOTA zijn.

Deze beperkingen van de gegevens maken dat wij geen betrouwbare kwantitatieve uitspraken kunnen doen over de invloed van biometrische ouderdom op FNMR en/of FMR, noch over de houdbaarheidstermijn of herbruikbaarheid van vingerafdrukken voor geautomatiseerde identiteitsverificatie op basis van een paspoort, in het algemeen noch voor leeftijdscategorieën in het bijzonder.

3.5 Aantekeningen Vingerafdrukken

De lagere mated scores en de daling daarvan met biometrische ouderdom voor personen van leeftijdscategorieën van zowel 12- als 65+ kunnen grotendeels verklaard worden door de interactie tussen vinger en sensor, blijkt uit [15].

Bij personen tot 12 jaar oud zijn de vingers en daarmee de papilairlijnen en andere structuren, veel fijner; het is, met de gebruikte vingerscanners, daardoor lastig(er) om de voor biometrische identiteitsverificatie benodigde kenmerken vast te leggen. Ook zijn algoritmen voor vingervergelijking en kwaliteitsmeting vooral getraind zijn op volwassen samples waardoor de kleinere samples van personen tot 12 jaar minder goed kunnen verwerken waardoor er hogere error rates te verwachten zijn.

Voor vingerafdrukken afgenomen op verschillende leeftijden (onder de 12 jaar) zullen de variaties in de afbeeldingen relatief hoog zijn in vergelijking met twee afbeeldingen op volwassen leeftijd waardoor de mated score afneemt met biometrische ouderdom.

Voor de Nederlandse praktijk geldt echter dat bij deze leeftijdscategorie geen afdrukken worden verzameld.

Voor personen van 65 jaar en ouder neemt de elasticiteit van de huid af en wordt die droger waardoor sensoren gebaseerd op aanraking de voor geautomatiseerde biometrische identiteitsverificatie benodigde kenmerken veel minder goed kunnen vastleggen en de mated score afneemt met leeftijd. Voor vingerafdrukken afgenomen op verschillende leeftijden (boven de 65 jaar) zullen de variaties in de afbeeldingen relatief hoog zijn in vergelijking met twee afbeeldingen op volwassen leeftijd waardoor de mated score eveneens afneemt met biometrische ouderdom.

Voor beide leeftijdscategorieën kan een toegespitste technische oplossing gevonden worden, door aanpaste scanners en/of algoritmen te gebruiken.

4 CONCLUSIE & AANBEVELINGEN

4.1 Conclusie

Tot welke toename van de biometrische ouderdom daalt de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie niet significant? Kan de daling in betrouwbaarheid met biometrische ouderdom gekwantificeerd worden in de vorm van toegenomen False Non-Match Rates (FNMR) en False Match Rates (FMR)?

Op basis van recente studies blijkt dat de betrouwbaarheid van geautomatiseerde biometrische verificatie daalt bij een toename van de biometrische ouderdom ten opzichte van de huidige 10 jaren geldigheidsduur. Deze daling is beperkt bij een toename van enkele jaren.

De beperkingen in de representativiteit van genoemde studies voor geautomatiseerde identiteitsverificatie op basis van vingerafdrukken en/of gezichtsafbeeldingen op Nederlandse paspoorten maken dat wij geen betrouwbare kwantificering van de toename van FNMR en FMR kunnen geven voor geautomatiseerde biometrische identiteitsverificatie in de Nederlandse praktijk. Vanwege dezelfde reden kan eveneens de houdbaarheid niet betrouwbaar gekwantificeerd worden voor de Nederlandse praktijk.

Brengt een toename van de biometrische ouderdom met 10 of meer jaar een significante daling van de betrouwbaarheid van geautomatiseerde biometrische identiteitsverificatie met zich mee? Kan deze daling in betrouwbaarheid gekwantificeerd worden in de vorm van toegenomen False Non-Match Rates (FNMR) en False Match Rates (FMR)?

Op basis van recente studies blijkt dat de betrouwbaarheid van geautomatiseerde biometrische verificatie daalt bij een toename van de biometrische ouderdom met 10 of meer jaar ten opzichte van de huidige 10 jaren geldigheidsduur. Deze daling is aanzienlijk.

De beperkingen in de representativiteit van genoemde studies voor geautomatiseerde identiteitsverificatie op basis van vingerafdrukken en/of gezichtsafbeeldingen op Nederlandse paspoorten maken dat wij geen betrouwbare kwantificering van de toename van FNMR en FMR kunnen geven voor geautomatiseerde biometrische identiteitsverificatie in de Nederlandse praktijk. Vanwege dezelfde redenen kan ook geen betrouwbaar advies over herbruikbaarheid van de biometrie op Nederlandse paspoorten worden gegeven.

4.2 Aanbevelingen

Vanwege de beperkte mate waarin wij de vraagstelling hebben kunnen beantwoorden doen wij allereerst aanbevelingen met betrekking tot de mogelijkheden om deze alsnog volledig beantwoord te krijgen (paragraaf 4.2.1).

Naar aanleiding van het onderzoek doen wij ook enkele aanbevelingen voor onderzoek in het verlengde daarvan in paragraaf 4.2.2.

4.2.1 Onderzoek/Bepalen houdbaarheid/herbruikbaarheid biometrie

Voor het betrouwbaar kwantitatief bepalen van biometrische houdbaarheid aan de hand van representatief genormeerde FMR of FNMR is er een uitgebreid onderzoek vereist inclusief een grootschalige dataverzameling. Wij geven daarbij het volgende in overweging;

*De dataverzameling zal over 10+ jaren moeten plaatsvinden, omdat het de enige manier is om uit te vinden hoe de betrouwbaarheid van biometrische verificatie verandert.*²⁰

Tijdens de dataverzameling moet de *spreiding in leeftijden gehandhaafd* blijven; aangezien er beneden de leeftijd van 12 jaren geen vingerafdrukken worden afgenomen is het vooral van belang dataverzameling te borgen van personen vanaf 12 jaren tot een redelijk maximum.

Selecteer een afspiegeling van de diversiteit maar vooral ook goed gelijkende paren; de personen die in het onderzoek betrokken worden moeten de diversiteit van alle mogelijke Nederlandse ingezetenen vertegenwoordigen.

Echter zowel bij het doen van een aanvraag voor een paspoort als bij een grenspassage zal het aanzienlijk lastiger zijn – zowel voor een bevoegd ambtenaar als een automatische vergelijker – om gezichten te onderscheiden van mensen die qua geslacht, etnische afkomst en uiterlijke kenmerken op elkaar lijken.

Onderzoek met willekeurig geselecteerde paren (zogenaamde “zero effort pairing”) onderschat de False Match Rate (FMR) – onderscheiden van gezichten is makkelijker als ze meer verschillend zijn - en is ook niet realistisch; de kans dat iemand probeert een nieuw paspoort aan te vragen met behulp van een paspoort van iemand van een ander geslacht of een duidelijk ander uiterlijk is praktisch nul. Naast diversiteit is het dus nodig dat er ook voldoende paren zijn te maken die onderling vergelijkbaar zijn.

Gebruik representatieve technologie/methoden zowel voor de identiteitsverificatie als voor het vastleggen van de biometrie. De methoden voor identificatie – al dan niet geautomatiseerd – en bijvoorbeeld vingerscanners zoals de praktijk die gebruikt moeten ook in het onderzoek betrokken worden.

Onderzoek de mogelijkheid van maatwerk; ons onderzoek zocht naar een algemeen advies voor houdbaarheid en/of herbruikbaarheid, terwijl de biometrische houdbaarheid zeer waarschijnlijk niet gelijk is voor iedereen op dezelfde leeftijd, noch op dezelfde manier varieert over het leven. Bovendien verschilt de houdbaarheid mogelijk ook nog tussen gezicht en vingerafdruk.

Het is daarom aan te raden om de mogelijkheid van maatwerk te onderzoeken, uiteraard binnen redelijke grenzen en met inachtneming van de noodzakelijke kwaliteitsnormering en standaardisering.

4.2.2 Algemene overwegingen

Wij hebben in ons onderzoek gekeken naar de huidige Nederlandse praktijk bij hernieuwde paspoortaanvragen en controle bij grenspassage. Die praktijk kan aan verandering onderhevig zijn wat betreft organisatorische, juridische of technische aspecten. Hieronder geven wij een aantal mogelijkheden aan die wij in ons onderzoek zijn tegengekomen en die men zou kunnen overwegen ter flexibilisering en verbetering van het proces van geautomatiseerde identiteitsverificatie door middel van Nederlandse paspoorten.

Verbeter de kwaliteit van opnamen; de kwaliteit van vingerafdrukken is in hoge mate bepalend voor de betrouwbaarheid van geautomatiseerde vingervergelijking. Die kwaliteit wordt, op zijn beurt, in hoge mate bepaald door de interactie tussen vingers en sensoren; hoe beter de sensor het papilaairlijnenbeeld kan vastleggen hoe beter de afdruk zal zijn. De houdbaarheid en

²⁰ Eventueel kan men onderzoeken of gezichtsopnamen en vingerafdrukken van eerdere aanvragen langer bewaard kunnen worden als ze worden toegevoegd aan een anonieme dataset; op dit moment zijn er juridische beperkingen op de bewaartermijn.

herbruikbaarheid van vingerafdrukken kan dus geoptimaliseerd worden door de scanner-technologie en eventueel de algoritmen die de data verwerken zodanig te kiezen dat de kwaliteit van de vingerafdruk optimaal wordt.

Gezien de beperkte permanentie van het gezicht is er overeenkomstig slechts een beperkte verbetering te verwachten van een hogere kwaliteit gezichtsopname.

Onderzoek effect bij verschillende demografische kenmerken; Het effect van leeftijd en biometrische ouderdom kan sterk verschillen van persoon tot persoon door verschillen in levensstijl, beroep, geslacht en uiterlijke kenmerken. Door deze factoren als parameters in onderzoek mee te nemen kan men de factoren die de kwaliteit van het proces (bijvoorbeeld nauwkeurigheid van biometrische verificatie) beïnvloeden inzichtelijk maken. Factoren die specifiek voor bepaalde demografische groepen gelden kunnen dan met maatwerk – bijvoorbeeld aangepaste technologie of procedures - worden opgelost zodat de betrouwbaarheid van biometrische verificatie voor iedereen zo veel mogelijk gelijk is.

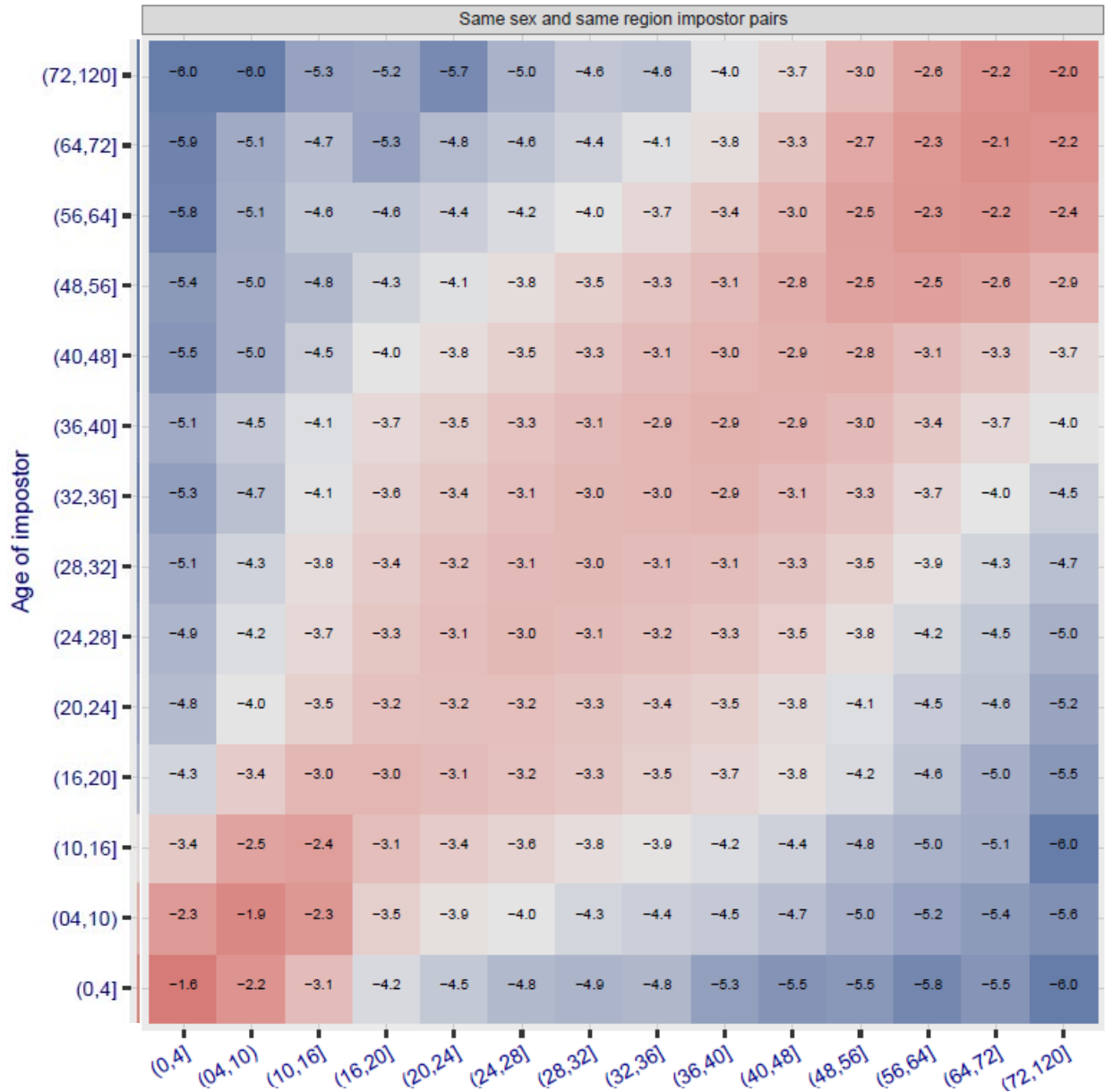
REFERENTIES

- [1] R. Khedgaonkar, M. Raghuvanshi en K. Singh, „Patch-based Face Recognition under Plastic Surgery,” in *First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar, 2018.
- [2] T. Valentine, M. B. Lewis en P. J. Hills, „Face-space: A unifying concept in face recognition research,” *The Quarterly Journal of Experimental Psychology*, vol. 60, nr. 10, pp. 1996-2019, 2016.
- [3] L. Best-Rowden en A. K. Jain, „Longitudinal Study of Automatic Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, nr. 1, pp. 148 - 162, January 2017.
- [4] Y. M. Liu, D. Bolme, B. A. Draper, J. R. Beveridge, G. Givens en P. J. Phillips, „A Meta-Analysis of Face Recognition Covariates,” in *Proceedings of the IEEE Third International Conference on Biometrics: Theory, Applications and Systems*, 2009.
- [5] D. Deb, L. Best - Rowden en A. K. Jain, „Face Recognition Performance Under Aging,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, 2017.
- [6] P. Grother, M. Ngan en K. Hanaoka, „Ongoing Face Recognition Vendor Test (FRVT) Part 1: Verification,” National Institute for Standards and Technology, 2020.
- [7] P. Grother, M. Ngan en K. Hanaoka, „Face Recognition Vendor Test (FRVT); Part 3: Demographic Effects,” National Institute of Standards and Technology, 2019.
- [8] D. Meuwly, „De Vingerafdruk,” in *Forensische Wetenschap*, A. Broeders en E. Muller, Red., Deventer, Kluwer, 2008, pp. 323-344.
- [9] C. Champod, C. J. Lennard, P. Margot en M. Stoilovic, *Fingerprints and other ridge skin impressions*, CRC Press, 2017.
- [10] H. Faulds, „On the skin-furrows of the hand,” *Nature*, vol. 22, p. 605, 1880.
- [11] S. K. Modi, S. J. Elliott, J. Whetsone en H. Kim, „Impact of Age Groups on Fingerprint Recognition Performance,” in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, Alghero, 2007.
- [12] S. Yoon en A. K. Jain, „Longitudinal study of fingerprint recognition,” *Proceedings of the National Academy of Sciences*, vol. 112, nr. 28, p. 8555–8560, 2015.
- [13] G. Praprotnik en N. Pavesic, „The Impact of Template Aging on the Performance of Automatic Fingerprint Recognition,” *Revija za kriminalistiko in kriminologijo/Ljubljana*, vol. 67, nr. 4, p. 371–388, 2016.
- [14] S. Kirchgasser, A. Uhl, K. Castillo-Rosado, D. Estévez-Bresó, E. Rodríguez-Hernández en J. Hernández-Palancar, „Fingerprint Template Ageing Revisited-It’s the Quality, Stupid!,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Los Angeles, 2018.
- [15] L. Beslay, J. Galbally en R. Haraksim, „Automatic fingerprint recognition: from children to elderly, Ageing and age effects,” Publications Office of the European Union, Luxembourg, 2018.
- [16] J. Galbally, R. Haraksim en L. Beslay, „A Study of Age and Ageing in Fingerprint Biometrics,” *IEEE Transactions on Information Forensics and Security*, vol. 14, nr. 5, pp. 1351-1365, 2019.
- [17] P. Grother en M. Ngan, „Face Recognition Vendor Test (FRVT): Performance of Face Identification Algorithms,” National Institute of Standards and Technology, 2014.

LIJST BEGRIPPEN EN AFKORTINGEN

Begrip	Afkorting	Betekenis*
Biometrische ouderdom		het verschil in tijdstip tussen de opname van twee <i>biometrische samples</i>
Biometrisch sample		Opname van biometrische kenmerken bedoeld voor <i>biometrische identiteitsverificatie</i>
Biometrische identiteitsverificatie		de vergelijking van twee <i>biometrische samples</i> om te bepalen of ze van dezelfde persoon afkomstig zijn
Commercial Off The Shelf	COTS	Aanduiding bij een product van commerciële technische dienstverlener dat zonder verder onderzoek of ontwikkeling inzetbaar is.
False Match		Vergelijking van <i>non-mated pair</i> waar <i>biometrische verificatie</i> onterecht slaagt
False Match Rate	FMR	Deel van de <i>non-mated pairs</i> die een <i>false match</i> geeft; maat voor (on)veiligheid van een systeem voor identiteitsverificatie.
False Non-match		Vergelijking van <i>mated-pair</i> waar de <i>biometrische verificatie</i> onterecht faalt
False Non-Match Rate	FNMR	Deel van de <i>mated pairs</i> die een <i>false non-match</i> geeft; mate van overlast die een systeem voor identiteitsverificatie veroorzaakt.
Herbruikbaarheid		Het geschikt zijn van het biometrische sample op een Nederlands paspoort voor hernieuwd gebruik bij de aanvraag van een Nederlands paspoort. Dit betekent dat bij toename van <i>biometrische ouderdom</i> met 10 jaar – of bij meerdere keren herbruik met 20 jaar of meer – de <i>biometrische verificatie</i> nog betrouwbaar moet werken.
Houdbaarheid		De maximale <i>biometrische ouderdom</i> waarvoor de <i>biometrische identiteitsverificatie</i> nog betrouwbaar werkt
Houder		Degene op wiens naam het reisdocument is gesteld en ten behoeve van wie het is uitgereikt (Bron: Paspoortwet Art.1 onder f)
Mated Pair		Twee biometrische samples die afkomstig zijn van dezelfde persoon
Mated Score		<i>Score</i> gegeven bij de vergelijking van een <i>mated pair</i>
Mugshot		Gezichtsafbeelding van een verdachte, opgenomen bij aanhouding of inverzekeringstelling
Non-mated Pair		Twee biometrische samples die afkomstig zijn van verschillende personen
Non-mated Score		<i>Score</i> gegeven bij de vergelijking van een <i>non-mated pair</i>
Score		Kwantitatieve maat, door geautomatiseerd biometrisch systeem gegeven, voor de gelijkenis tussen aangeboden twee biometrische samples.
State-Of-The-Art	SOTA	De op technische gronden best presterende applicatie, of het best presterende systeem.
Toonder		Degene die het reisdocument toont om zijn/haar identiteit te laten verifiëren.

APPENDIX



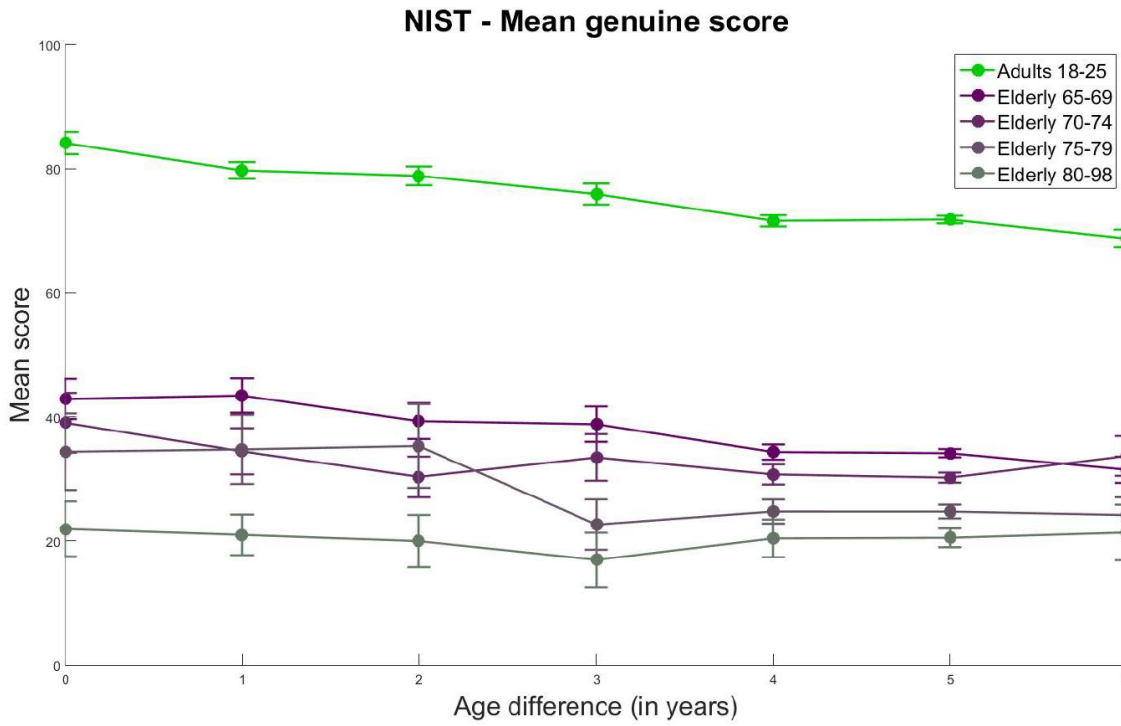
FIGUUR A1: HEATMAP VAN RELATIEVE FMR TUSSEN LEEFTIJDGROEPEN

De waarden in deze figuur zijn de $\log_{10}(FMR)$, oftewel de logaritme op basis 10 van de FMR; een waarde van -3.0 is dus 0.001. Hoe meer negatief het getal hoe kleiner de FMR.

De FMR voor alle vergelijkingen tussen de non-mated pairs is 0.0001; dit correspondeert met -4.0 in de figuur.

Elke cel geeft de waarde voor de vergelijking tussen afbeeldingen van de leeftijdscategorie op de horizontale as, tegen "impostors" van de leeftijdscategorie op de verticale as; de waarden op de diagonaal van links-onder tot rechts-boven, corresponderen met vergelijkingen binnen de leeftijdsgroep, die direct daaraan grenzend, van aangrenzende leeftijdsgroepen.

Bron: [7] figuur 15 pagina 50



FIGUUR A2: MATED SCORES VOOR OUDERE LEEFTIJDSCATEGORIEEN