Content-based retrieval of visual information

Oerlemans, A.A.J.

Nederlandse Samenvatting

In het huidige digitale tijdperk zijn zeer veel gegevens beschikbaar in de vorm van bijvoorbeeld foto's, video's en geluid. Deze hoeveelheid gegevens neemt dagelijks met onvoorstelbare snelheid toe. Een voorbeeld hiervan is YouTube, waar elke dag voor ongeveer zeven jaar aan video bijgeplaatst wordt. Ook op websites als Flickr zijn al miljarden foto's geüpload. Maar wat is het nut van al deze digitale informatie, als er niet op een handige manier in teruggezocht kan worden?

Dit proefschrift beschrijft een aantal onderzoeken die als doel hadden het terugzoeken van digitale afbeeldingen te vergemakkelijken. De technieken staan bekend als Content-Based Image Retrieval (CBIR) methodes, wat betekent dat de inhoud van afbeeldingen gebruikt wordt om de gebruiker te ondersteunen in zijn zoekproces.

De gangbare manier van zoeken op internet vindt plaats door middel van het invoeren van tekst. Als er gezocht wordt naar een afbeelding, dan is men hierbij afhankelijk van de bij de afbeeldingen geplaatste omschrijvingen. Helaas is het vaak zo dat niet de juiste omschrijving bij een afbeelding staat, of dat er zelfs helemaal geen omschrijving bij een afbeelding gegeven wordt. Denk hierbij bijvoorbeeld aan een serie vakantiefoto's die op internet gezet wordt. Er zal waarschijnlijk wel een omschrijving voor de gehele serie foto's zijn, zoals 'Vakantie in Frankrijk 2011', maar niet elke losse foto zal een specifieke beschrijving hebben, zoals bijvoorbeeld 'De Eiffeltoren'. In dit geval zullen de gangbare methoden van zoeken met tekst niet de gewenste resultaten kunnen geven, want er zijn immers geen tekstuele omschrijvingen beschikbaar aan de hand waarvan die ene specifieke afbeelding van de Eiffeltoren uit de serie gevonden kan worden.

Zoals hierboven al vermeld, onderzoeken de technieken uit dit proefschrift de inhoud van afbeeldingen en proberen hieruit informatie af te leiden die gebruikt kan worden voor zoekacties. Voor de gebruiker betekent dit dat hij of zij ook kan zoeken naar een afbeelding die lijkt op een afbeelding die hij of zij zelf al heeft, of dat er gezocht kan worden naar omschrijvingen die door de computer automatisch van afbeeldingen afgeleid zijn.

Een voorbeeld van de eerste manier, is het zoeken naar een afbeelding van een specifiek type auto. Als een gebruiker al een afbeelding heeft van het type auto, maar hij of zij wil er graag nog meer bekijken, dan kunnen CBIR technieken

uitkomst bieden. Deze technieken kunnen afbeeldingen met hetzelfde type auto terugvinden, zelfs als er verder geen informatie beschikbaar is bij de afbeelding. Er wordt dan gezocht op overeenkomsten tussen de inhoud van beide foto's. De computer kan zien of twee foto's op elkaar lijken.

Een voorbeeld van de tweede manier van zoeken op basis van beeldinhoud, is het zogenaamde 'visual concept detection', het automatisch herkennen van bepaalde visuele concepten of ideen op een afbeelding. Hierbij wordt een foto geänalyseerd door de computer en zullen er automatisch bepaalde woorden bij een foto geplaatst worden die uit deze automatisch analyse volgen. Hierbij kan gedacht worden aan 'gebouw', 'berg', 'zee', maar ook aan meer specifieke woorden als 'winkel', 'Mount Everest' of 'strand bij Noordwijk'. Uiteraard zullen deze concepten een keer door de computer geleerd moeten worden aan de hand van voorbeelden, maar hierna zal het geleerde concept gebruikt kunnen worden en zal er zonder tussenkomst van een mens bij elke afbeelding een lijst met woorden gemaakt kunnen worden. Hierna kan de gebruiker weer zoeken met tekst, zoals hij of zij gewend is. De foto van de Eiffeltoren uit de eerder genoemde serie vakantiefoto's kan dan toch gevonden worden.

In dit proefschrift worden drie wetenschappelijke bijdragen beschreven op het gebied van content-based image retrieval: het MOD paradigma, geconstruëerde textuur-patronen en de zogenaamde multi-dimensional maximum likelihood measure, een meerdimensionale aanpak voor het vergelijken van eigenschappen van afbeeldingen. Elk van deze bijdragen zal hieronder besproken worden.

Op het gebied van het terugzoeken van visuele informatie is een zeer uitdagend probleem van de laatste jaren het detecteren van visuele concepten, waarbij de computer gevraagd wordt om automatisch een beeld te voorzien van relevante steekwoorden. Op een fundamenteel niveau betekent dit dat de computer een vorm van begrip voor afbeeldingen heeft gekregen. Als de computer een strand, gebouw, gezicht of zonsondergang ziet, worden deze concepten herkend op basis van het beeld, net zoals een mens zou doen. De kleuren, vormen en andere eigenschappen (ook wel 'features' genoemd) worden hiervoor gebruikt. Het is tot zeer recent haast onmogelijk gebleken om dit probleem op te lossen.

De eerste bijdrage van dit proefschrift, te vinden in hoofdstuk 6, is een algoritme voor visuele concept detectie dat gebruik maakt van 'salient points', automatisch bepaalde punten in een afbeelding die interessant of opvallend zijn. In dit proefschrift wordt een nieuw paradigma voorgesteld dat MOD heet, wat staat voor Maximization Of Distinctiveness. Hierbij worden deze interessante punten geselecteerd op basis van hun onderscheidend vermogen. De MOD aanpak is getest op de meest uitdagende internationale wetenschappelijke test set en bleek een significante verbetering te geven ten opzichte van de beste methode uit de literatuur van visuele concept detectie. In tegenstelling tot de andere onderzoeken die salient points gebruiken voor het analyseren van beelden, generaliseert deze methode tot elk type beeldinhoud en elk type afbeeldingen.

In hoofdstuk 8 is de tweede bijdrage van dit proefschrift te vinden, op het gebied

van computationeel efficiënte textuur beschrijvingen. Eenvoudig omschreven, zegt een textuur beschrijving iets over het materiaal waar naar gekeken wordt, zoals bijvoorbeeld bakstenen of grind. Afbeeldingen die dezelfde textuur bevatten, kunnen teruggevonden worden als de door de computer bepaalde beschrijving van de textuur maar goed genoeg is. Textuur features zijn waarschijnlijk de meest gebruikte visuele eigenschappen in de computer vision. Op dit moment is de meest voorkomende textuur feature in de wetenschappelijke literatuur de 'local binary patterns' (LBP), een 256 dimensionale beschrijving van 3x3 patronen. Er is herhaaldelijk vastgesteld dat deze feature een hoge mate van accuraatheid heeft, maar dat het gebruik ervan ook een significante computationele rekenkracht vereist. In dit proefschrift wordt beschreven dat het mogelijk is om met grotere patronen dan 3x3 een vergelijkbare nauwkeurigheid te behalen als met LBP, maar dan met slechts 2 dimensies in plaats van 256. Dit verhoogt de computationale efficiëntie met een factor honderd en reduceert de hoeveelheid benodigd geheugen met een vergelijkbare factor.

De derde bijdrage van dit proefschrift, terug te vinden in hoofdstuk 7, is een beschrijving van de 'multi-dimensional maximum likelihood' (MDML) methode voor het vergelijken van eigenschappen van afbeeldingen. Op dit moment is op het gebied van computer vision de meest voorkomende manier van het vergelijken van eigenschappen van afbeeldingen de 'sum of squared differences' (SSD). Gebruikmakend van de theorie van meest aannemelijke schatters wordt in dit proefschrift beschreven dat het gebruik van SSD alleen optimaal is bij bepaalde aannames over de onderliggende kansverdeling van de ruis, in het bijzonder wanneer deze Gaussisch is. In dit proefschrift wordt het bekende computer vision probleem beschreven van het zoeken van overeenkomsten in afbeeldingen die uit twee zichtpunten gemaakt zijn, eenvoudig gezegd: een stereo-paar. Hierbij wordt voor elk punt uit de ene afbeelding een overeenkomend punt gezocht in de andere afbeelding. Met deze overeenkomsten kan dan de driedimensionale structuur van de beeldinhoud berekend worden of de beweging van de camera.

Dit proefschrift laat zien dat de optredende ruis in de analyse van stereo-paren niet Gaussisch is, dus dat in dit geval het gebruik van SSD niet optimaal is. Daarnaast worden er diverse methoden onderzocht om de werkelijke ruisdistributie te schatten en hierbij wordt gevonden dat de ééndimensionale aanpak een tweede fundamentele aanname heeft: het verschil tussen eigenschappen bevat voldoende informatie om de gelijkheidsdistributie te modelleren. Dit proefschrift toont aan dat dat niet zo is.

Op het gebied van computer vision is het probleem van het vinden van overeenkomsten tussen afbeeldingen van twee zichtpunten een zeer belangrijke en uitdagende geweest in de afgelopen twintig jaar. In dit proefschrift wordt beschreven dat de meerdimensionale aanpak de eerder genoemde tweede aanname overwint en significant betere resultaten geeft bij de meest geloofwaardige en gerespecteerde internationale test set in het vinden van overeenkomsten in stereo-paren. In het algemeen verbetert dit werk de theorie van computationale gelijkheid op een fun-

damentele manier, die mogelijk op alle gebieden van patroonherkenning en computer vision verbetering kan brengen.

De laatste hoofdstukken van dit proefschrift, 9 en 10, beschrijven onderzoeken die enigszins losstaan van de content-based image retrieval. Deze onderzoeken hebben te maken hebben met automatische video-analyse. De technieken die beschreven worden, zijn ontworpen voor videobeelden van stilstaande camera's, wat veelal neerkomt op beveiligingscamera's. Het doel van de technieken is om bewegende personen of object in beeld te onderscheiden en deze te volgen. Hierbij kan de gebruiker terugkoppeling geven aan de processen die objecten proberen te volgen, zodat bepaalde objecten altijd zichtbaar blijven als bewegend object, zelfs als ze lang stil staan. Ook is het mogelijk om bepaalde delen van het beeld juist aan te merken als achtergrond, waarmee het nooit als een bewegend object gezien zal worden.

Als laatste wordt er in dit proefschrift nog een programma beschreven dat voor zowel onderzoek als onderwijs gebruikt kan worden. RetrievalLab stelt de gebruiker in staat om complexe bewerkingen uit te voeren op databases van afbeeldingen, zonder dat daar iets voor geprogrammeerd hoeft te worden. Het geeft hiermee snel inzicht in de processen die schuil gaan achter content-based image retrieval en het detecteren van visuele concepten.