



Universiteit
Leiden
The Netherlands

Large scale visual search

Wu, S.

Citation

Wu, S. (2016, December 22). *Large scale visual search*. Retrieved from <https://hdl.handle.net/1887/45135>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/45135>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/45135> holds various files of this Leiden University dissertation.

Author: Wu, S.

Title: Large scale visual search

Issue Date: 2016-12-22

Nederlandse Samenvatting

Met de al maar groeiende hoeveelheid visuele data op het web is er veel aandacht besteed aan het grootschalig zoeken naar afbeeldingen. Er zijn verschillende redenen waarom dit een van de grote uitdagingen op het gebied van computer vision is. Allereerst moet er rekening gehouden worden met visuele transformaties zoals perspectief, rotatie en schaling die zijn toegepast op afbeeldingen uit deze enorme hoeveelheid data. Ten tweede is het noodzakelijk de vereiste hoeveelheid geheugen en rekenkracht te minimaliseren die bij het genereren van de afbeeldingsrepresentaties nodig is. Tenslotte moet er een efficiënte index-ruimte en een geschikte afstands maat gecreëerd worden om wachttijd van de gebruiker te verkleinen. Deze scriptie beoogt robuuste afbeeldingsrepresentaties te leveren die minder gevoelig zijn voor de hiervoor genoemde visuele transformaties en geschikt zijn voor het grootschalig zoeken naar afbeeldingen.

Al langer bestaande technieken zoals het Bag-of-Words (BoW) model en haar varianten domineren het onderzoek naar grootschalig zoeken naar afbeeldingen. Het process van BoW bestaat voornamelijk uit drie stappen: (i) salient point feature extraction, (ii) visual vocabulary generation, (iii) het op BoW gebaseerde feature encoding. Veel onderzoek heeft zich gericht op het bereiken van de state-of-the-art prestaties voor grootschalig zoeken naar afbeeldingen.

Als een eerste stap onderzoeken we de sterke en zwakke punten van de bestaande salient point detectoren en descriptorren op diverse verdraaiingen van afbeeldingen. De gepresenteerde vergelijkende experimenten kunnen onderzoekers bijstaan in het selecteren van een passende detector en descriptor voor het genereren van een op BoW gebaseerde representatie van afbeeldingen.

NEDERLANDSE SAMENVATTING

Vergeleken met locale descriptoren van reële getallen hebben locale descriptoren bestaande uit een binaire getallenreeks het voordeel weinig geheugen nodig te hebben en efficiënt vergeleken te kunnen worden met behulp van de Hamming-afstand. Verder stellen we voor om de “K-majority” cluster methode met ANN-search te gebruiken voor het genereren van BoW afbeeldingsrepresentaties gebaseerd op locale descriptoren van binaire getallenreeksen. De evaluatie-resultaten laten zien dat deze aanpak weinig geheugen gebruikt voor vocabulaire opslag en competitief presteert vergeleken met BoW afbeeldingsrepresentaties gebaseerd op locale descriptoren van reële getallen.

Aangezien bestaande salient point methodes gevoelig zijn voor de kijkhoek en veranderingen in perspectief stellen we ook een nieuwe salient point descriptor voor, RIFF genaamd. RIFF wordt gegenereerd op basis van paarsgewijze intensiteitsvergelijkingen tussen selectie-patronen geïnspireerd door de patronen van het menselijke netvlies. De evaluatie hiervan laat zien dat de op RIFF gebaseerde BoW afbeeldingsrepresentaties andere kenmerkbeschrijvingen met betrekking tot schaal, rotatie en kijkhoek transformaties weet te overtreffen.

Recentelijk hebben afbeeldingsrepresentaties gegenereerd aan de hand van convolutional neural networks (CNNs) laten zien goed te presenteren ten opzichte van de state-of-the-art als het gaat om het zoeken naar afbeeldingen. In deze scriptie onderzoeken we reëel-waardige en binaire representaties van afbeeldingen gebaseerd op feature maps uit de lagen van de CNNs, en presenteren we een fusie-schema om de nauwkeurigheid bij het zoeken naar afbeeldingen verder te verbeteren. Daarnaast ontwerpen we een krachtigere CNN-architectuur om de robuustheid van CNN modellen te verbeteren.

Tenslotte, ondanks de substantiële bijdragen van deze scriptie op het gebied van het grootschalig zoeken naar afbeeldingen, presenteren we verdere uitdagingen en nieuwe onderzoeksrichtingen naar aanleiding van de bijdragen uit dit proefschrift.