# Significantie keuze aantal means in k-means

## Namen en datum

Alexander Streng – vision groep 17 25-06-2014

## Doel

In dit meetrapport wil ik graag de significantie laten zien van de keuze van het aantal means in het bepalen en segmenteren van een afbeelding. Te weinig means kan betekenen dat je niet alles wat je wilt onderscheiden segmenteert en te veel means kan betekenen dat je te veel onderscheid maakt. Ik wil toelichten wat voor invloed de keuze van het aantal means op de segmentatie heeft. Volgens Wikipedia is een generieke regel voor het bepalen van het aantal clusters in een dataset:  k \approx \sqrt{n/2}  waar n het aantal elementen is. <http://en.wikipedia.org/wiki/Determining_the_number_of_clusters_in_a_data_set>

Voor een image van 1920 x 1080 is dus het aantal means 1018.

## Hypothese

Ik veronderstel dat Wikipedia ‘s rule of thumb niet van perse een goede keuze is voor het bepalen van het aantal means. Het aantal zal altijd redelijk hoog liggen, waardoor ook de performance in het geding komt. Tevens verwacht ik wel dat een hoger aantal means een afbeelding met veel verschillende objecten ten goede komt als je het wilt segmenteren. Tevens is het ook duidelijk dat de hoeveelheid means een negatief effect hebben op de performance.

## Werkwijze

Om hier een goede vergelijking te kunnen maken zal ik een afbeelding segmenteren beginnend met k = 1. Vervolgens blijf ik k incrementeren tot  k \approx \sqrt{n/2} . Om een goede onderscheiding te maken zal ik een afbeelding kiezen met veel detail, en 1 afbeelding met weinig detail. De afbeeldingen zijn 125 bij 125 (wat dus een totaal oplevert van 89 clusters. Vervolgens zal ik kijken bij welk aantal means de meest optimale segmentatie optreed.

Veel detail weinig detail

 

Ook zal ik het aantal iteraties bijhouden.

## Resultaten

Grafiek 1: aantal means ten opzichte van de tijd

|  |
| --- |
| F:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_2_means_1.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_8_means_2.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_8_means_3.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_7_means_4.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_21_means_5.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_19_means_6.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_14_means_7.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_25_means_8.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_23_means_9.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_21_means_10.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_25_means_11.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_25_means_14.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_25_means_13.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_25_means_12.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\detail\kmeans-detail-iterations_25_means_15.png |

Eerste 15 means van de image met veel detail

|  |
| --- |
| F:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_2_means_1.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_8_means_2.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_9_means_3.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_10_means_4.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_25_means_5.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_19_means_6.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_24_means_7.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_19_means_8.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_22_means_9.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_25_means_10.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_25_means_11.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_24_means_12.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_25_means_13.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_25_means_14.pngF:\Projects\School\Vision-team17\Meetrapporten\Week 3 - 4\Data Meetrapport kmeans\geen detail\kmeans-nodetail-iterations_25_means_15.png |

Eerste 15 means van image met weinig detail.

## Verwerking

Ik heb de meetresultaten verkregen door middel van de geleverde timer bij de practicaopgaven. Verder heb ik gebruik gemaakt van bestaande code van mijzelf om de images te genereren en op te slaan.

## Conclusie

Uit de afbeeldingen is te concluderen dat de rule of thumb een slechte keuze is als het aankomt op K-means. Na +- 20 means is het voor kleine afbeeldingen niet meer rendabel om meer means te selecteren.

## Evaluatie

De keuze voor het aantal means hangt af van het doel wat je er mee wilt bereiken. Het is wel duidelijk geworden dat hoe hoger het aantal means, hoe langer het duurt voordat de image is gesegmenteerd. Ook is duidelijk dat bij k-means een te hoog aantal means kan leiden tot een slechte segmentatie. Uit mijn tests is gebleken dat een klein aantal means ( 3 – 5 ) meestal voldoende is.

Omdat ik een beetje qua tijd verlegen zat heb ik geen ingewikkelder algoritme implementeren. Enkele goede voorbeelden zijn : “X-means: extending K-means with efficient estimation of the number of clusters” (<http://www.cs.cmu.edu/~dpelleg/download/xmeans.pdf>) en “Learning the K in K-means” (<http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2003_AA36.pdf>)