Analysing and interpreting data from medical wearables and digital questionnaires

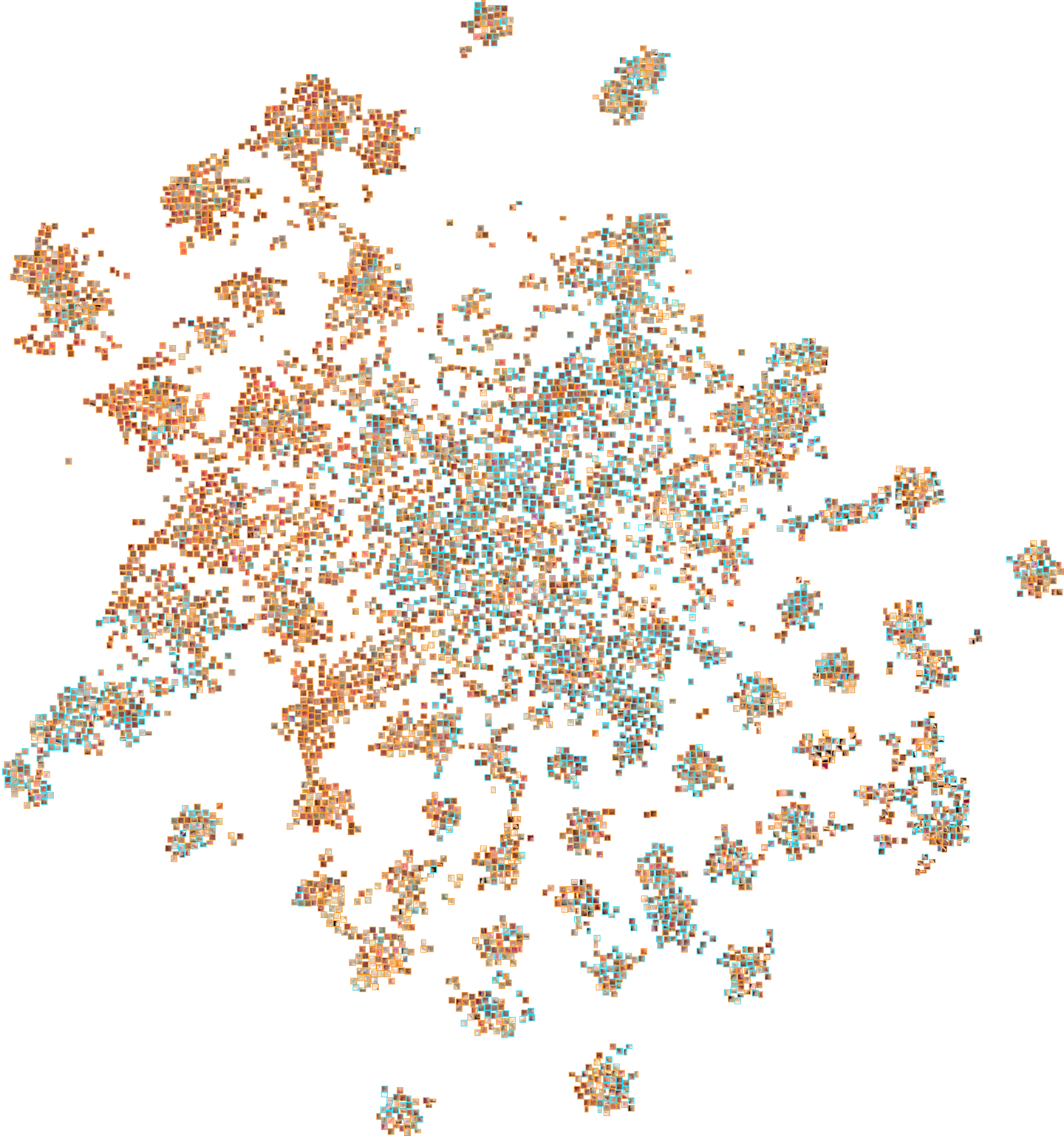
**JHFM Pinckaers, 0924121**

**Junior Research Project II**

May 2, 2016 to October 21, 2016

Supervisors: Robert Rissmann and Yuri Mejia Miranda

Departmental affiliation: Centre of Human Drug Research & Pharmacology, LUMC



Abstract

The abstract should provide the context or background for the study and should state the study’s purpose/ aim, basic procedures (study subjects, laboratory animals or cell lines, observational and analytical methods), main findings (giving specific effect sizes and their statistical significance, if possible), and principal conclusions. It should emphasize new and important aspects of the study or observations.   
Length: Max 250 words

Introduction

C

linical trials encompass time intensive repeated measurements of clinical parameters on their participants. Subjects stay in the clinical research unit for a prolonged period of time, during which they are assessed with a wide array of tests. Due to this process fragile participants are excluded or receive less tests.{Schwenk:2015wi} The clinical research unit is often located in an environment dissimilar from the normal environment of the participant. Measuring outside of the normal environment is likely influencing the results. This effect harms the external validity of a clinical trial. In addition, being observed during a measurement could also influence it, the Hawthorne effect.{McCambridge:2014dq} For example, the well‑known *white coat effect* in blood pressure measurements where blood pressure is higher in the clinic than at home.

Next to data gathered at the clinical research unit, parameters at interest at home are measured via paper diaries. However, this adds uncertainties. With paper diaries the researcher cannot be sure that the information the participant reports is actually filled in on mentioned day. Previous studies have shown that participants will not always fill in these questionnaires on the desired times, with patients ‘hoarding’ information until the last day, termed ‘parking lot compliance’.{Morren:2009ti, Stone:2003wi}

Parking lot compliance can be harmful in interpreting trial data. To evaluate the effect, safety and tolerability of a new compound in a clinical trial appropriate dosing is required to reach therapeutic concentrations.{Lee:2006tb} Compliance to the study protocol is necessary to reach and sustain those concentration. When adherence is measured with unreliable methods, such as a paper diary, it is generally assumed that adherence is nearly ideal in clinical trials.{Vrijens:2014hx} This assumption can result in incorrect assessments of clinical trials.{Feldman:2009cf} Czobor et al.{Czobor:2011un} re-analysed two trials utilizing their reported compliance. Originally without considering compliance, the trials reported no differences between the compound and placebo, however the reanalysis by Czobor et al. found statistical significant therapeutic effects.

During the last decade we have grown normal to carrying a computer device with us all the time; the smartphone.{Dediu:2014up} Therefore, we are able to receive and record information almost everywhere, all the time. This provides opportunities to use these devices in clinical trials. The paper diaries can be translated to ‘smart’ electronic diaries.

An electronic questionnaire, running on a handheld device, can store precise time stamps whenever users enter data. This provides an objective way to monitor the participant. In addition, these electronic diaries are able to send reminders to users, further improving compliance. Earlier studies with text messages (SMS) as reminders to apply medicine or score symptoms show an improved compliance.{Balato:2013hc}

Next to improving compliance, the electronic patient diary has advantages in data handling. A systematic review conducted in 2005 concerning the effectiveness of handheld computers versus paper methods suggests that the timeliness of data handling is improved. Furthermore, since the original data entry is digital, there are no errors when translating paper to digital.{Lane:2006in} One study done in 1997 reported an 80% reduction in time spent on data handling when using electronic devices.{Tiplady:1997vv}

There have been multiple trials reported with electronic diaries{Lane:2006in}, with only a few{Bryant:2013ig, Thriemer:2012bz, Hensel:2012ii, Knipe:2014ji, Allena:2012if, Hon:2013ia, Blake:2015kq, Ratcliff:2014ca, So:2013fb, Zhou:2014ce, Turk:2013bv, Priebe:2013em} after the introduction and general adoption of the smartphone (around 2012{Dediu:2014up}). However, none of these studies measured compliance using an objective measure. We performed dermatologic trials, in which mobile applications are used to ask participants questions about their complaints. Additionally, we asked for a picture of their skin lesion containing medicine as proof of compliance.

We report the results of the use of this framework in three dermatological trials. One of the mobile application was used in a phase 2 trial of a Topical Ionic Contra-viral Therapy (ICVT) comprised of digoxin and furosemide in cutaneous verrucae. Additionally, a mobile application was developed for a phase 1/2 trial of omiganan in patients with mild to moderate atopic dermatitis. Finally, a mobile application was used in a phase 2 trial of topical omiganan in patients with usual type vulvar intraepithelial neoplasia (uVIN). The trial results will be published elsewhere. This study evaluates the compliance in these trials, as well as the value-added of data gathered at home, using the electronic patient diary.

Furthermore, with increasing amount of data the complexity of analysis increases, the ‘big data’-problem. To help the future analysis of compliance using participant-provided pictures, we developed an experimental image analysis algorithm to identify applied topical medicine on these pictures. This algorithm could be used as a warning signal during the trial when a participant is found to be consistently incompliant to the protocol.

*maximum 1000 words*

Methods

A

n iOS application was developed for iOS 8 using Xcode and Objective-C. The application was installed on a fifth generation iPod Touch provided to the subjects by the researchers. Participants were instructed to charge the device on a regular basis. Subjects captured pictures using the on-board camera. A textual instruction was displayed on top of the image when capturing. Afterwards subjects could zoom in and crop the picture. Additionally, they could accept the picture, or replace it by taking a new one.

The application didn’t provide functionality to change data from other days. Only data of the current day could be changed. This way the recall-bias was reduced, and compliance data would be more trustworthy.

The application scheduled notifications at the participant-customizable time and three consecutive notifications each a half hour after the previous. If the subject completed all the information for that day further notification were cancelled until the next day. Data was saved on device and sent to the server using AES-256 encryption the following day. A PHP-server backend received the data and wrote the textual data to a MSSQL database and the pictures to disk. A script converted MSSQL data to CSV files to be imported into the clinical trial database and further use in the statistical analysis.

Compliance was measured by examining picture values in the electronic diaries. No compliance was presumed when no image on that day was available. When data was available completeness was evaluated. Analysis was done in Microsoft Excel 2016 and R studio. Figures were created using R, HTML and CSS-styling.

Topical Ionic Contra-viral Therapy (VV) Trial

During the topical ionic contra-viral therapy study all subjects were instructed to picture all the warts included in the study, every day. To track compliance, pictures had to be taken from warts containing the therapy. The application would remind the participants to apply the treatment on a specific time of the day, chosen by the participant.

Atopic Dermatitis (AD) Trial

During the atopical dermatitis study all 36 dosed subjects used the mobile application to track compliance with pictures, but also to record itch and pain scores twice daily. The application would remind the participants to supply itch and pain scores.

Vulvar Intra-epithelial Neoplasia (VIN) Trial

The vulvar intra-epithelial neoplasia study was not finished on time of writing. The currently 5 finished subjects used the mobile application to track compliance with pictures, by picturing the medicine on their finger before applying, but also to record itch and pain scores twice daily.

Questionnaire data

After the trials, questionnaires regarding the experience of using the electronic patient diary participants were administered. Questions were asked about user experience, annoyance of notifications, help with compliance. The whole questionnaire is included in the Supplemental Information.

Reminder phone calls

Since the patient-reported outcomes recorded with the electronic patient diary was sent over the internet, the researchers were able to precisely monitor the participants. In case of non-compliance the participant would receive a phone call to ask if there were problems with the electronic diary. [note: should we include this data at all? We get possible recall-bias/lying participants.]

Literature search for compliance in dermatological trials

To find papers on compliance of dermatological studies, we searched PubMed with the keyword: “dermatological”, “topical” and “compliance” together with the related diseases: “atopic dermatitis”, “verruca vulgaris”, “vulvar intraepithelial neoplasia” and its synonyms. This resulted in 109 papers on June 3, 2016. All papers describing a clinical trial mentioning the relevant dermatological diseases and relative compliance rates were gathered. The search query is included in the Supplemental Information.

Image analysis

T

he binary classifier was trained to discriminate between images that proved compliance and images that not. An image was classified as compliant if there was a wart present with applied medicine in the picture. For development Python 2.7.11 was used in combination with open source computer vision library OpenCV 3.1{Bradski:2000up} and scientific libraries SciPy 0.17.1{Jones:2001uv}, Numpy 1.11.0{vanderWalt:2011dp}, Scikit-learn{Pedregosa:2011tv}, Spearmint{Snoek:2012vl} and Matplotlib{Hunter:2007ih}. All source code is available online at https://github.com/HansPinckaers/wart-detection.

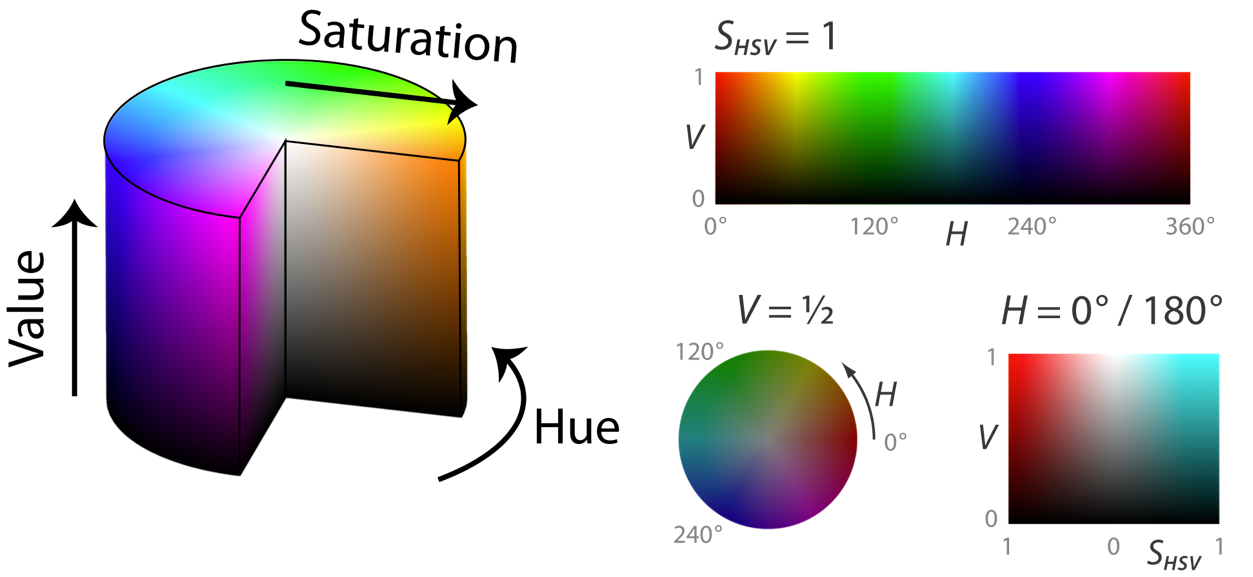
Image representation

To analyze an image, it needs to be represented in a mathematical way. For image analysis this is often done in matrices, but images can also be represented in other ways such as frequencies and wavelets. Matrices are useful since mathematical clustering algorithms can be used on them. A grayscale image of 20 by 20 pixels in size will have a matrix representation with 20 rows and 20 columns. The gray color of an individual pixel on a white-to-black scale can be represented between 0 – 255, where 0 is black and 255 is white. This means that at a certain value *v* in the matrix at row *y* and column *x* will be the grayscale value of the pixel at location *x, y* in the image. For color images multidimensional matrices (height x width x 3) are used. Each color channel (e.g. red, green or blue) will have its own dimension.

Naive algorithm to find skin lesions

To simplify the generation of a training set an algorithm to search for skin lesions was manually defined. The algorithm is depicted in textbox 1. This algorithm is naive since it is manually defined, not trained, and it will find any irregularities on skin.

The algorithm starts by applying a mean shift filtering on the image. Mean shift combined with hierarchical clustering provides a nonparametric way to cluster an image.{Comaniciu:2002ij} The average hue of the resulting clusters are compared to a predefined skin color (H 8, S 176, V 187). Skin-colored clusters are found based on their distance to this color. Here the HSV colorspace is used. Since hue in this space is a cyclic variable (the two ends of a hue color spectrum are similar, see *Figure 2*) the hue coordinated of the HSV colorspace is rotated such that the skin colors are in the middle, this means that both ends will be furthest away from resembling skin colors. The resemblance towards the average skin color was calculated as a Euclidean distance of the coordinates in the rotated HSV space.



**Figure 2.** The HSV colorspace (adapted from Wikipedia)

1: shifted 🡨 MeanShift(image)

2: clusters 🡨 HierarchicalClustering(shifted)

3: Keep a collection of clusters which are skin-colored *S*

4: Define standard skin color *skincolor*

5: **for** c **in** clusters **do**

6: hue 🡨 AverageHue(c)

7: distance🡨 norm(hue, skincolor)

8: **if** distance < threshold **do**

9: insert(S, c)

10: **end if**

11: end for

12: mergedcluster 🡨 MergeAligningClusters(S)

13: skincluster 🡨 GuassianBlur(mergedcluster)

14: edges 🡨 CannyEdgeDetection(skincluster)

15: regions 🡨 DilateAndErode(edges)

16: threshold 🡨 5, increment 🡨 1 *> threshold/increment in mm2*

17: **while** count(areas) > 3 **do**

18: threshold 🡨 threshold + increment

19: areas 🡨 FindRegionsAboveThreshold(regions, threshold)

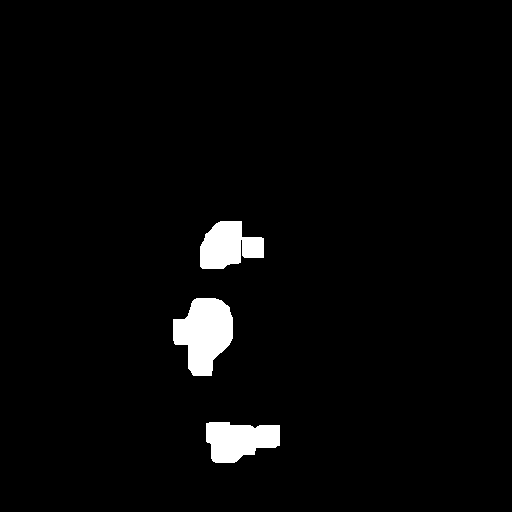
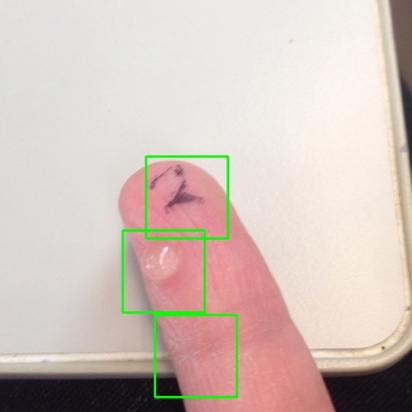
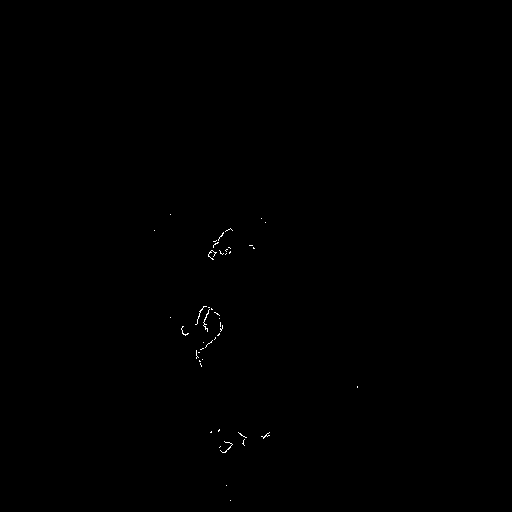
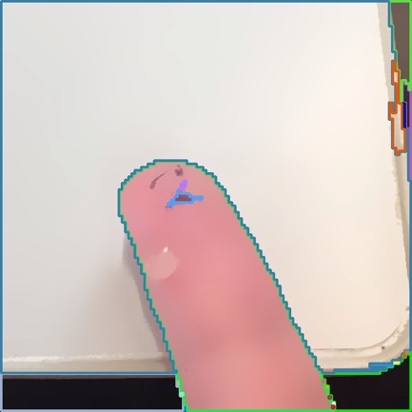
20: end while

**Textbox 1.** Naive algorithm for searching skin lesions

The clusters which are close to the skin color based on a predefined threshold are merged to one skin cluster. Subsequently, the skin cluster is blurred by applying a Guassian kernel to reduce noise. Canny edge detection is then performed on the blurred cluster. The edged are dilated and eroded to form regions between close edges. We then loop through these regions while raising the threshold for their size. Once we have less than three regions we define them as skin lesions. This means that the algorithm never finds more than three regions per image.

Mean shift et al.  
Before proceeding to develop the new algorithms the issue of the employed color space has to be settled. To obtain a meaningful segmentation perceived color differences should correspond to Euclidean distances in the color space chosen to represent the features (pixels**). An Euclidean metric, however, is not guaranteed for a color space [65, Secs.6.5.2; 8.4]. The spaces L u v and L a b were especially designed to best approximate perceptually uniform color spaces. In bothcases L the lightness (relative brightness) coordinate is deﬁned the same way, the two spaces differ only through the chromaticity coordinates.** The dependence of all three coordinates on the traditional RGB color values is nonlinear. See [46, Sec.3.5] for a readily accessible source for the conversion formulae. The metric of perceptually uniform color spaces is discussed in the context of feature representation for image segmentation in [16]. In practice there is no clear advantage between using L u v or L a b , in the proposed algorithms we employed L u v motivated by a linear mapping property [65, p.166].

The lesions determined by the algorithm were manually classified to contain be a wart, a wart with applied medicine, nothing or dubious when the region was only partially filled with a wart. The regions were saved as individual files and used as training data. 20% of the data was randomly separated to function as a test set. The random allocation was done per participant to prevent very similar looking pictures to be included in the training and test set.



**a**

**b**

**c**

**d**

**e**

**f**

**Figure 1.** Example of the algorithm. The wart with applied cream is found.   
However, also the number drawn on the finger and the prominent skin lines   
at the joint are identified.  
A: Original image. B: Mean shift segmentation. C: Find skin cluster.   
D: Canny edge detection. E: Dilate and erode. F: Final result.

Feature detectors and descriptors

For a model to detect objects in an image it compares small regions of an image with small regions in classified images. These small regions are called features. A classifier cannot compare all possible features of an image so it has to make decisions which features to compare. Which regions get defined as features is decided by a feature detector.

There are several feature *detector* algorithms, we looked at SIFT, SURF, AKAZE, KAZE, Agast, GFTT, and MSER. The algorithms choose regions based on different characteristics in the image, such as edges and contrast

After the selection of features the pixels within the regions have to be translated to a multidimensional vector or binary string to be compared with other features. *Descriptors* algorithms translate features to vectors. The descriptors also create an orientation vector per feature such that they are rotation invariant.

Most feature detectors such as SIFT and SURF have algorithms to find regions of interest and then describe them. However, there are dedicated descriptor algorithms such as BRIEF. These can be combined with a feature detectors like SIFT.

Bag of words

Since every image will have multiple feature and we deal with a few thousand images, comparing every feature description with another is unfeasible. To reduce the number of features to compare clustering techniques are used. We used k-means clustering for the feature descriptors that returned Euclidean vectors and we implemented the k-majority clustering algorithm for feature descriptors that returned binary features.

Using the clusters an histogram is created per image where the relative abundance per cluster is calculated. Every feature counts for the nearest cluster. Choosing the right number of clusters is important. If the number of bins is too small the visual words will not be representative of all features and thus maybe discriminative data will be lost. If the number of bins is too large there is a chance of overfitting (since every feature will have its own bin).

The resulting vector with the same length as the bag of words size can then be compared between images.

Testing detectors/descriptors

To test the different feature detectors and descriptors on our data we did an experiment with all of them, and with different parameters. We created a 2D scatter plot out of the multidimensional vectors using t-SNE to reduce the dimensions. Since t-SNE is optimized for similarity, similar images will be grouped together. This allows for inspection of the feature, because the discriminative power of the particular feature detector and descriptor can be judged by eye. Fig 2 shows an example of a t-SNE. The SIFT feature detector and descriptor was found to be the most discriminative with our data.

Naive model (k nearest neighbor)

Using the SIFT features and bag-of-words model we can classify images as containing medicine or being negatives using k-nearest neighbor. Since a model cannot be evaluated based on the images on which it was trained we used k-fold validation for evaluation. We divided the dataset into 5 folds, and over 5 iterations a model was trained based on 4 folds and tested on the remaining fold. Tabel .. shows the k-fold validation results of the k-nearest neighbor classifier (k = 15).

Support vector machine model

Bayesian optimization of hyperparameters

Explanation of C parameter: <http://stats.stackexchange.com/questions/31066/what-is-the-influence-of-c-in-svms-with-linear-kernel>

For the support vector machine we used a radial basis function kernel since our data will not be linearly separable, because we assume to have overlapping features for both classes. The RBF kernel function will result a smoothened fit in our multidimensional data.

Error/cost function

2.3.4 with hyperparameters generate svm and or random forest

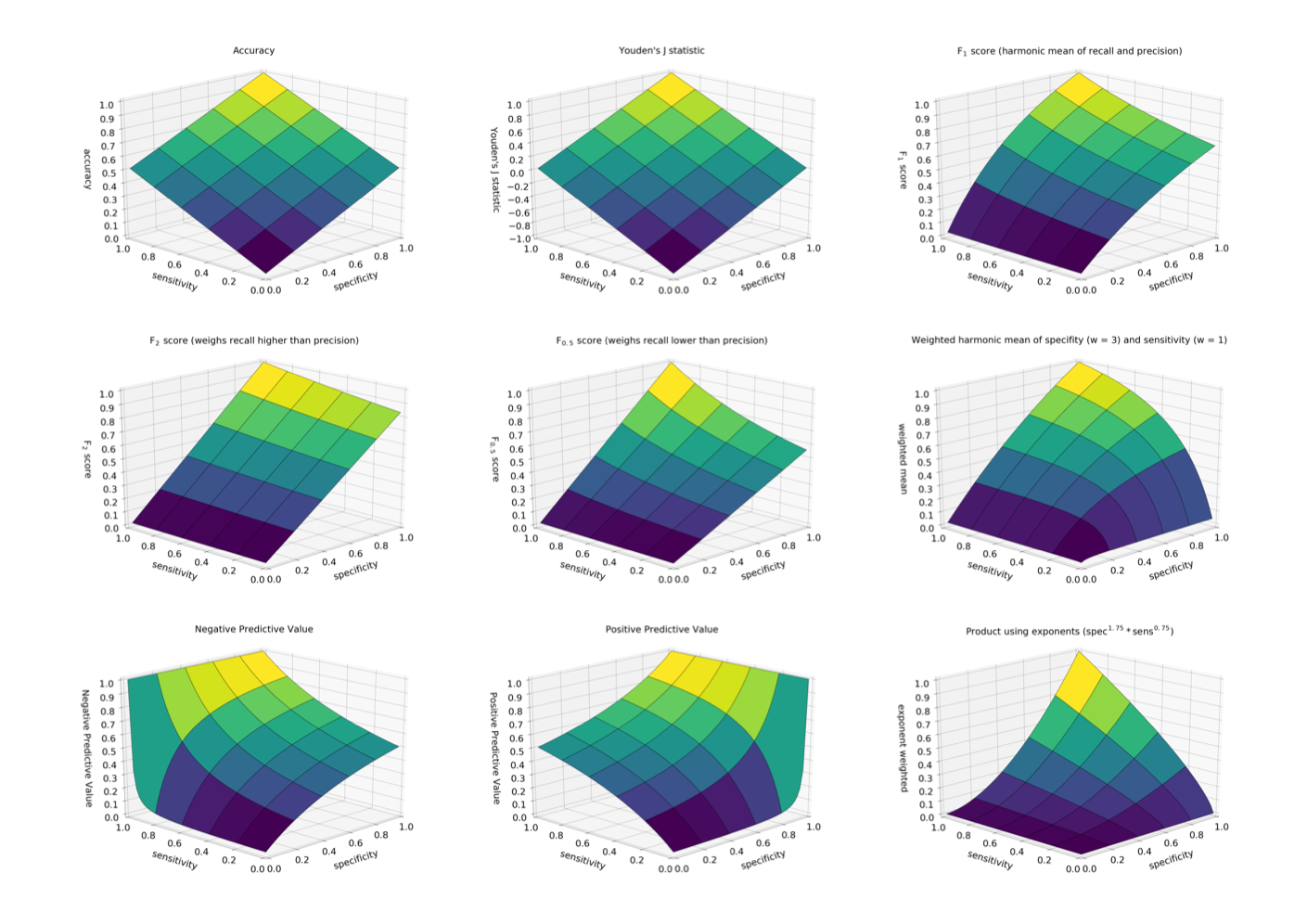
2.3.3 running on test set

2.3.4 Non maximum suppression on positive windows

2.3.5 Hard negative finding/mining (adding false positive windows to training set)

2.4 human versus model test

<http://imagelab.ing.unimore.it/imagelab/pubblicazioni/2013ElectronicImaging.pdf>



Results

80 participants were enrolled in de VV trial, 36 in the AD trial and, on time of submission, 5 finished in the VIN trial. All participants dosed finished the trial. The mean age of the VV and AD trial was relatively young, 25.8 and 24.9 respectively. The mean age of the participants in the VIN trial was older, since vulvar intraepithelial neoplasia occurs more often at older age.

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Topical Ionic Contra-viral Therapy Trial | Atopic Dermatitis Trial | Vulvar Intra-epithelial Neoplasia Trial |
| Age (SD) | 25.8 (10.6) | 24.9 (7.8) |  |
| Female | 49 (61%) | 27 (75%) |  |
| Male | 31 (39%) | 9 (25%) |  |
| BMI (SD) | 23.5 (3.2) | 22.5 (2.9) |  |

**Table 1.** Demographics of the different trials

Overall, the participants had high compliance rates. The VV trial showed 97.2% compliance, the AD trial 94.1%. The VIN trial showed lower compliance rate of 83.8%, this was mainly due to one participant having a low compliance rate (30%), the other 4 participants each had more than 90% compliance to the protocol.

|  |  |  |  |
| --- | --- | --- | --- |
| Study | Compliance | n | Days |
| Topical Ionic Contra-viral Therapy Trial | 97% | 80 | 41 |
| Atopic Dermatitis Trial | 94% | 36 | 28 |
| Vulvar Intra-epithelial Neoplasia Trial | 84% | 5 | 85 |
| Total | 92% | 121 | 154 |

**Table 2.**Compliance per study

All participants completed the questionnaire after the trial, **Table 3** shows an overview of relevant questions. The complete answers of the questionnaires can found in the Supplemental Information. 90% of the participants would rate the electronic patient diary as good to very good overall; 97% of the participants judged the user-friendliness of the electronic diary as good to very good. A majority of 79% said the diary took 5 minutes or less to fill in each day. 56% of the participants never encountered a technical error, 32% encountered an error one to two times in their use. The photo functionality was experienced by 79% of the participants as pleasant to very pleasant. 60% answered that their compliance would be worse without an electric patient diary, with 83% saying the burden of an electronic diary is less than of a paper diary. 94% would prefer an electronic diary over paper in a next trial.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **How user-friendly was the app?** | **Very good** | **Good** | **Average** | **Moderate** | **Bad** |  |
| VV trial | 30 (41%) | 40 (55%) | 2 (3%) | 1 (1%) | 0 (0%) |  |
| AD trial | 28 (74%) | 10 (26%) | 0 (0%) | 0 (0%) | 0 (0%) |  |
| VIN trial | 1 (20%) | 3 (60%) | 1 (20%) | 0 (0%) | 0 (0%) |  |
| *Total* | *51%* | *46%* | *3%* | *1%* | *0%* |  |
| **In general how would you rate the app?** | **Very good** | **Good** | **Average** | **Moderate** | **Bad** |  |
| VV trial | 14 (19%) | 50 (69%) | 8 (11%) | 1 (1%) | 0 (0%) |  |
| AD trial | 10 (26%) | 26 (68%) | 2 (5%) | 0 (0%) | 0 (0%) |  |
| VIN trial | 1 (20%) | 3 (60%) | 1 (20%) | 0 (0%) | 0 (0%) |  |
| *Total* | *22%* | *68%* | *10%* | *0.9%* | *0%* |  |
| **How much time did it take to use the app each day?** | **1-5 min.** | **5-10 min.** | **10-15 min.** | **15-20 min.** | **>20 min.** |  |
| VV trial | 56 (77%) | 16 (22%) | 0 (0%) | 0 (0%) | 1 (1%) |  |
| AD trial | 32 (84%) | 5 (13%) | 1 (3%) | 0 (0%) | 0 (0%) |  |
| VIN trial | 4 (80%) | 1 (20%) | 0 (0%) | 0 (0%) | 0 (0%) |  |
| *Total* | *79%* | *19%* | *0.9%* | *0.0%* | *1%* |  |
| **How often did technical problems occur?** | **Never** | **1-2 times** | **3-4 times** | **5-10 times** | **>10 times** |  |
| VV trial | 34 (47%) | 27 (38%) | 6 (8%) | 3 (4%) | 2 (3%) |  |
| AD trial | 28 (74%) | 10 (26%) | 0 (0%) | 0 (0%) | 0 (0%) |  |
| VIN trial | 3 (60%) | 0 (0%) | 2 (40%) | 0 (0%) | 0 (0%) |  |
| *Total* | *57%* | *32%* | *7%* | *3%* | *2%* |  |
| **How did you experience the photo function of the app?** | **Very pleasant** | **Pleasant** | **Neutral** | **Cumbersome** | **Very cumbersome** |  |
| VV trial | 11 (15%) | 43 (59%) | 16 (22%) | 2 (3%) | 1 (1%) |  |
| AD trial | 15 (42%) | 18 (50%) | 3 (8%) | 0 (0%) | 0 (0%) |  |
| VIN trial | 0 (0%) | 3 (60%) | 1 (20%) | 1 (20%) | 0 (0%) |  |
| *Total* | *23%* | *56%* | *18%* | *3%* | *1%* |  |
| **What would your compliance be with a paper diary?** | **Forget a lot more** | **Forget more** | **Neutral** | **Forget less** | **Never forget** | **Don’t know** |
| VV trial | 21 (29%) | 20 (27%) | 22 (30%) | 0 (0%) | 2 (3%) | 8 (11%) |
| AD trial | 17 (47%) | 5 (14%) | 6 (17%) | 1 (3%) | 0 (0%) | 7 (19%) |
| VIN trial | 2 (40%) | 3 (60%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| *Total* | *35%* | *25%* | *25%* | *0.9%* | *2%* | *13%* |
| **How do you estimate the burden of using the app compared to a paper diary?** | **A lot less work** | **Less work** | **No difference** | **More work** | **A lot more work** | **Don’t know** |
| VV trial | 42 (58%) | 16 (22%) | 5 (7%) | 5 (7%) | 2 (3%) | 3 (4%) |
| AD trial | 21 (58%) | 11 (31%) | 1 (2.8%) | 3 (8%) | 0 (0%) | 0 (0%) |
| VIN trial | 5 (100%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| *Total* | *60%* | *24%* | *5%* | *7%* | *2%* | *3%* |
| **What is your preference to use for subsequent studies?** | **Electronic** | **Paper** | **Don’t know** |  |  |  |
| VV trial | 66 (94%) | 0 (0%) | 4 (6%) |  |  |  |
| AD trial | 35 (92%) | 2 (5%) | 1 (3%) |  |  |  |
| VIN trial | 5 (100%) | 0 (0%) | 0 (0%) |  |  |  |
| *Total* | *94%* | *2%* | *4%* |  |  |  |

**Table 3.** Relevant answers of questionnaire in all trials.   
Due to rounding the sum of each category could be larger than 100%.

**Results**  
The Results section should present and illustrate your findings. Present your results in logical sequence in the text, tables, and illustrations, giving the main or most important findings first. Do not repeat all the data in the tables or illustrations in the text; emphasize or summarize only the most important observations. Extra or supplementary materials and technical detail can be placed in an appendix where they will be accessible but will not interrupt the flow of the text.

When data are summarized in the Results section, give numeric results as the absolute numbers from which any derivatives (for example, percentages) were calculated, and as the derivatives if appropriate. Restrict tables and figures to those needed to explain the argument of the paper and to assess supporting data. Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables.

Discussion

T

his study evaluated compliance in two finished clinical trials and an ongoing trial by using an electronic patient diary on a handheld device. The mean compliance reported over the trials was 92%. This suggests that an electronic diary is beneficial to the compliance itself, e.g. via reminders, and a good way to measure the resulting adherence. This is not the first study showing electronic patient diary can improve compliance{Walker:2000wn}. However, we believe this is the first study using recent technology in the smartphone era, with participants used to handheld devices. Futhermore, evaluating compliance by taking pictures of the lesion with topical medicine applied is novel.

The questionnaire data showed general acceptability of using an electronic diary. Although one or more technical problems were reported by 43% of the participant. Often reported problems were crashing of the application, difficulties in taking pictures (due to lighting/focusing), and bad network connectivity. Still a very high percentage of 94% preferred it over a paper diary. Additionally, a large majority of 83% estimated the burden of using a electronic diary as less than using a paper diary. This could be explained because the VV and AD trials had relatively young – more familiar with technology – participants. However, the VIN trial shows comparable results with older generation. We think this is due to the recent adoption of the smartphone, making participants more accustomed to handheld devices. A review before the smartphone era shows lower preference for electronic diaries, on average 59%{Lane:2006in}.

The high compliance rates reported could be explained by the reminders send by the electronic patient diary. Additionally, since reminder phone calls were made, frequent non-compliant participant could be influenced. Subsequently, participants may feel more involved with the trial when daily information is asked. This seems to agree with the answers to the questionnaires; 69% of the participants found the reminder function to aid in applying the medicine in time, with 60% saying that their compliance would be worse with a paper diary.

It is challenging to measure compliance without actually influencing it. Previously, trials have measured concentrations of drugs in blood, but this is never very precise, or possible due to short half-life of the substances. Pill counting and sensors in the caps of bottles{Conde:2008vv} have also been tried, but these methods are potentially easily gamed by the participants. Due to these limitations our results are hard to compare to what would have happened without an electronic diary, even though the participants claim their compliance would be worse. However, when comparing to dermatological trials with the same diseases (Table 4)the compliance reported in our trials seem higher than average. However, due to few trials reporting compliance and even fewer reporting standard deviation of the compliance rates, no valuable statistical testing could been done. [is this true? ask statistician?]

|  |  |  |  |
| --- | --- | --- | --- |
| Dermatological trials | Compliance | n | Method |
| *Atopic Dermatitis* |  |  |  |
| Kütting B et al.27 | 73.7% to 85.8% | 800 | Interviews |
| Wilson R et al.28 | 70% | 20 | MEMS (Aardex Corp) |
| Wahlgren CF et al.29 | 89.8% to 90.8% (<70% excluded) | 30 | Electronic diary with PC |
| Francis NA et al.30 | 81.8% (children) | 77 | Paper diary |
| Conde JF et al.26 | 70% (children) | 8 | MEMS (sensor in cap) |
| *Verruca Vulgaris* |  |  |  |
| Bruggink S et al.31 | 71% | 82 | Interviews |
| Cockayne S et al.32 | 77% (week 3) | 115 | Interviews |
| *VIN* |  |  |  |
| Tristram A et al.33 | 87% (first 6 weeks) | 180 | Paper diary |

**Table 4.** Papers with published compliance data from dermatological trials.

Additionally to providing pictures, the participants could also fill in comments accompanying their daily data. In these comments the volunteers often praised the reminder-function. Also, comments with several potential improvements for the electronic diary were contributed. Some participants wanted to take pictures from their lesions just after midnight. However, the application was configured such that after midnight the picture would count for the new day, and not the day before which the participants expected. Future applications could prevent this by honoring submissions just after midnight to count for the previous day.

Data gathered using electronic diaries can be marked with precise time stamps and frequently be surveyed. This allows for high temporal data. The diaries used in the AD and VIN trial asked the patients twice daily for their amount of pain and itch. This could potentially show pharmacodynamics effects of the drug previously invisible, because of recall bias or less temporal sampling,6 for example a compound only working in the morning, Examples of high temporal data in eczema has already been published some time ago,29 however integrating this in the electronic diary should be preferred. Recording data in real-time will more accurately represent symptoms.

[this paragraph could be unnecessary] Some authors doubt the conformity between digital questionnaire with paper questionnaires to be a problem to compare.34 Although a meta-analysis a year earlier concluded that patient and computer reported outcomes are equivalent.35 We speculate that the benefits of electronic methods to gather questionnaire data outweigh this problem. When all arms of the trial use the same data retrieval methods this problem should not exist.

Since the applications are relatively simple, containing a picture functionality and ability to answer questions, ways to automatically develop these electronic diaries using the protocol of the study could be possible. This has already captured the interest of the bigger technology companies, such as Apple’s framework “ResearchKit” and Google’s “Science Journal”36,37. These frameworks lower the development cost of electronic patient diaries and give the opportunity to capture data from sensors on the device.

Collecting information at home in clinical trials also provides the opportunity to completely include and follow participants virtually.38 For example, this could make it possible to capture lifestyle influences on chronic diseases. However, since a clinical trials has to be randomized, participants need to receive drugs and safety implications, including patients without seeing them will be troublesome. Notwithstanding, in rural areas it could replace follow-up visits when travel is expensive or time-consuming.11

As mentioned before, the high compliance rates reported suggest that daily reminders and capturing of data makes participants more involved. This effect could also be of interest outside a clinical trial setting. Studies with treatment reminders via sms seem to improve compliance.39 Next to the compliance, long-term data about symptoms in ambulatory care could be valuable. However, before such data could be used in the clinical practice further studies should be done to interpret and integrate them.

In conclusion, in dermatological trials the efficacy can be tightly coupled with compliance.1,3 We used electronic patient diaries, in which participants took pictures of applied medicine, to evaluate adherence. This, together with precise date stamps, provided an objective measure of compliance. The reported clinical trials participants showed a high overall compliance rate of 92%. Thus, we advice the use of electronic patient diaries to improve and evaluate compliance, and better judge compounds in clinical trials.

**Discussion**  
The discussion should answer the question posed in the Introduction. It should explain how the results support the answer. The discussion should emphasize the new and important aspects of the study and the conclusions that follow from them. Link the conclusions with the goals of the study. The discussion should consider how the research performed in the study contributes or adds to work done in that field. Strengths and limitations of the study should be stated. Implications for future research and clinical practice should be suggested. Avoid conclusions not adequately supported by the data.

Length: The Discussion should be between 900 and 1500 words.

Acknowledgements

We wish to thank the medical librarian Jan Schoones (Walaeus Library, Leiden University Medical Center, Leiden, the Netherlands) for his efforts with developing the search queries.

**Acknowledgements**  
The acknowledgements section should specify any substantial help received from organizations or individuals, whether they provided grants, materials, technical assistance, or advice. Concisely thank those who went out of their way to help, and describe their contribution.

**References**  
Direct references to original research sources should be provided whenever possible, rather than references to review articles that may not reflect original work accurately. Small numbers of references to key original papers often serve as well as more exhaustive lists.

**Appendix**Information or data that supports or supplements the research performed (but is not central) may be included in an appendix. Extra information may be deemed necessary by either the student or the supervisor. Examples of supplementary material are

* standard protocols
* supplementary data
* pilot studies

Supplemental Information

* Questionnaire about ediary!

Tables

Tables capture information concisely and display it efficiently. Number tables consecutively in the order of their first citation in the text. A brief title should be placed above the table. The Students should place explanatory matter in footnotes, not in the heading. Explain all non-standard abbreviations in footnotes. Identify statistical measures of variations, such as standard deviation and standard error of the mean. Be sure that each table is cited in the text.

**Illustrations (Figures)**Figures should be made as self-explanatory as possible. Titles and detailed explanations belong in the legends, not on the illustrations themselves. Figures should be numbered consecutively according to the order in which they have been cited in the text. If a figure has been published previously, acknowledge the original source.

* *Legends for Illustrations (Figures)*: Type legends for illustrations with Arabic numerals (1,2,3, etc) corresponding to the illustrations. When symbols, arrows, numbers, or letters are used to identify parts of the illustrations, identify and explain each one clearly in the legend. Explain the internal scale and identify the method of staining in photomicrographs.
* *Units of Measurement*: Measurements should be presented in metric units according to the International System of Units.

**Abbreviations and Symbols**Use only standard abbreviations. The spelled-out abbreviation followed by the abbreviation in parenthesis should be used on first mention unless the abbreviation is a standard unit of measurement.

{papers2\_bibliography}

Temporary ----

Generating training set

To recognize skin lesions with applied medicine, supervised machine learning algorithms can be used. Supervised means that these algorithms are trained on a predefined dataset. In our case this dataset should contain images of skin lesions with and without topical medicine applied. Since the differences between these two states will be slight, it is of importance to have a negative training set of skin lesions without medication.

We generated the training set using image analysis to fasten the process. First the image is segmented, in which the image is divided in regions based on similarity. Mathematically this is the same as clustering in a matrix. After the segmentation the segment containing a finger is recognized by averaging the color in the segment and looking for skin colors. Using the location of this segment the finger in the original image can be found. This finger is further analyzed on contrast and texture differences to detect a possible region of the skin lesion.

The region of interests the above-explained method will find are manually classified for containing warts and possible applied medication. This classified set of images will be used to train a supervised machine learning algorithm.

0. Image representation

To analyze an image, it needs to be represented in a mathematical way. This is most often done in matrices, but images can also be represented in other representations such as frequencies and wavelets. Matrices are useful since mathematical clustering algorithms can be used on them. A grayscale image of 20x20 pixels will have a matrix representation with 20 rows and 20 columns. The amount of gray on an individual pixel can be represented between 0-255. This means that at a certain value v in the matrix at row y and column x will be the grayscale value of the pixel at location x,y in the image. For color images multi-dimensional matrices are used. Each color channel (red, green or blue) will have its own dimension.

1. Image segmentation

We wanted to analyze the image to in a nonparametric manner, where the analysis makes no assumption about the underlying distribution. This has the advantage that in a nonparametric analysis a predefined number of clusters is not needed. Only a few algorithms could be used for image segmentation. We tested the watershed and mean shift algorithm.

* 1. Watershed algorithm

...

* 1. Mean-shift algorithm

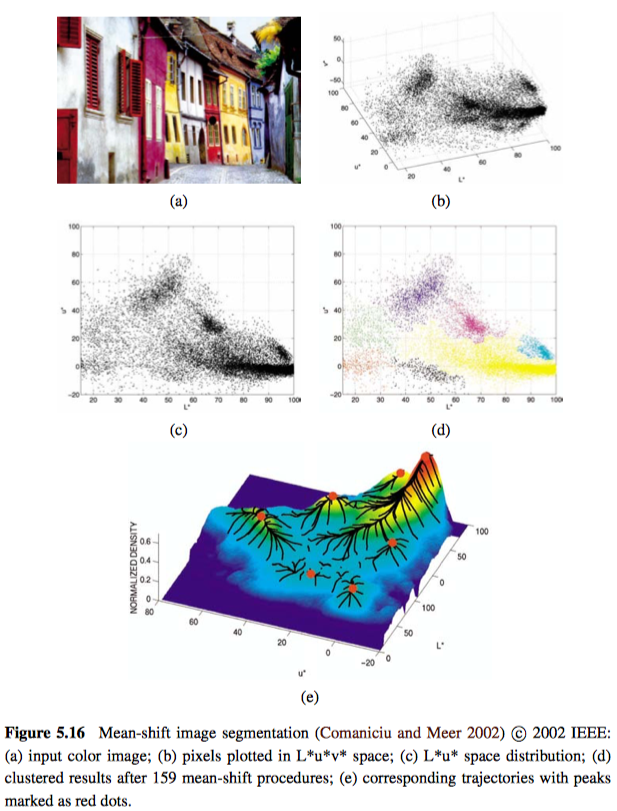
For the mean-shift algorithm the image representation is a multi-dimensional matrix by discarding spatial information (I think the algorithm also could take spatial information into account. Need to read the paper more carefully.) and only looking at color. This resulting representation is called the *feature space*. Mean shift considers the feature space to consist of a probability density function. The probability density function is a function in which the area under the curve describes the probability of a certain value range giving the supplied data points. For example, a well-known *probability density function* (p.d.f.) is the normal distribution.

The mean shift algorithm uses a sample of points (the window) in the feature space as a sample of a pdf, calculates the pdf on these discrete values with a supplied kernel and calculates the mean shift () this pdf. This *mean shift* is calculated by the weighted mean of the change of the p.d.f.: by summing the derivative’s value at every discrete of the point in the window and dividing it by the weights provided by the kernel. Another way of looking at it is calculating the direction of the p.d.f. at a distinct point; the direction towards the mean can be derived from the derivative. This is done equal to creating a ‘gradient’ vector, a vector of the partial derivatives (the derivate considering the different variables).

The window in the feature space (containing the analyzed sample points) of the algorithm then moves by this calculated direction mean shift, thereby moving the denser area of the p.d.f. This continues until a finite number of steps or after the shift is below a certain threshold, which mean a dense region is reached (local maximum of the p.d.f.)

Applied on an image, the ultimate value of each pixel is set to the mean of the underlying probability density function. Instead of using this algorithm on every pixel (value in matrix), random location samples can be taken, the paths of these mean-shifts toward the mean are tracked. If in the end there are pixels not contained in these paths they can be classified using the nearest path.

The segmentation can then be based on the spatial information and different hues of colors in the picture with a certain threshold of difference in hue to define a segment.



Source: Szeliski R. Computer Vision: Algorithms and Applications, 2010