

# Pattern recognition techniques used in data retrieval

# Image retrieval

<http://images.google.com>

[Het internet](#) **Afbeeldingen** [Maps](#) [Nieuws](#) [Video](#) [Gmail](#) [meer ▾](#) [sgorpi@gmail.com](mailto:sgorpi@gmail.com) | [Mijn account](#) | [Afmelden](#)



Afbeeldingen zoeken

[Geavanceerd zoeken naar afbeeldingen](#)  
[Voorkeuren](#)  
[Helpcentrum van Afbeeldingen zoeken](#)

De grootste afbeeldingenzoeker op het internet.

[Advertentieprogramma's](#) - [Bedrijfsoplossingen](#) - [Alles over Google](#)

© 2008 Google

# Key word based image retrieval

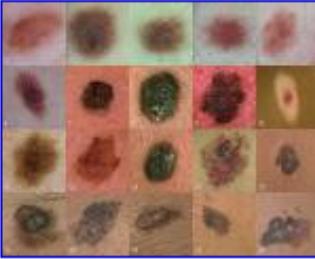
<http://images.google.com>

Het internet Afbeeldingen Maps Nieuws Video Gmail meer▼ sgorpi@gmail.com | Mijn account | Afmelden

**Google™** skin lesion Afbeeldingen zoeken Het internet

Gemiddelde SafeSearch is aan

Afbeeldingen Weergeven: Alle afbeeldingsformaten Resultaten 1 - 20 van circa 45.200 voor **skin lesion** (0,13 s)

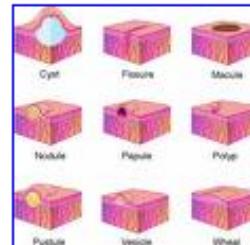
 Twenty images of **skin lesions**.  
1280 x 1024 - 274 kB - jpg  
[www.cs.wright.edu](http://www.cs.wright.edu)

 ... plaque-like **skin lesion**.  
680 x 521 - 175 kB - jpg  
[www.ispub.com](http://www.ispub.com)

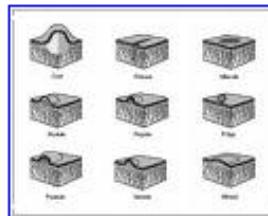
 Configuration of **Skin Lesions**  
263 x 400 - 10 kB - jpg  
[www.med-ed.virginia.edu](http://www.med-ed.virginia.edu)

 ... a slowly progressive **skin lesion** ...  
900 x 600 - 100 kB - jpg  
[www.afids.org](http://www.afids.org)

 ... infections on the **skin** do

 Skin Lesions

 ... infections on the **skin** do

 Various types of **skin**

# Content based image retrieval

- Search for images that are similar to an input image using image content, such as:
  - color
  - texture
  - shape
  - text tags
  - etc...
- See <http://en.wikipedia.org/wiki/CBIR>

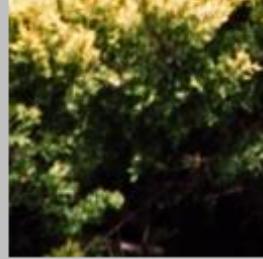
# IMEDIA project – select query

<http://www-rocq.inria.fr/imedia/>

[Settings](#) [HELP](#) [Other Demos](#) Visual search engine : project [IMEDIA](#) © 2001 INRIA

 Shuffle  Previous  Next  Regions mode  Points mode  Thesaurus mode  
 Feedback mode  Show keywords  
Keywords   
Feedback / Compose Send Keywords Send Image External jpeg image  Bladeren...

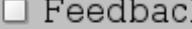
---


# iMedia project – query results

<http://www-rocq.inria.fr/cgi-bin/imedia/circario.cgi/v2std>

[Settings](#)    [HELP](#)    [Other Demos](#)    Visual search engine : project [IMEDIA](#) © 2001 INRIA

 Shuffle     Previous     Next     Regions mode     Points mode     Thesaurus mode  
 Feedback mode     Show keywords  
Keywords   
Feedback / Compose Send Keywords Send Image    External jpeg image  Bladeren...

---


# GIFT – select query

<http://viper.unige.ch/demo/php/demo.php>

(fetch a random set of images)     ( launch the query)     (Clear the query)

 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="rel"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>
 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>	 Similarity: 1.000000 <input type="button" value="neutral"/> <input type="button" value="top"/>

# GIFT – query result

<http://www.gnu.org/software/gift/gift.html>

(fetch a random set of images)     ( launch the query)     (Clear the query)

**Query:**



**Result:**

 Similarity: 1.000000 Query Image <a href="#">top</a>	 Similarity: 0.171830 <input type="button" value="neutral"/>	 Similarity: 0.160725 <input type="button" value="neutral"/>	 Similarity: 0.144877 <input type="button" value="neutral"/>	 Similarity: 0.141121 <input type="button" value="neutral"/>
 Similarity: 0.139827 <input type="button" value="neutral"/>	 Similarity: 0.137843 <input type="button" value="neutral"/>	 Similarity: 0.135897 <input type="button" value="neutral"/>	 Similarity: 0.135666 <input type="button" value="neutral"/>	 Similarity: 0.135661 <input type="button" value="neutral"/>

# Leiden 19<sup>th</sup>-century portrait database

<http://nies.liacs.nl:1860/cgi-bin/FindImage.pl>



**press OKE when image(s) are similar to the one you have in mind and  
press Search**

Search

# Leiden 19<sup>th</sup>-century portrait database

## (result examples)



**press OKE when image(s) are similar to the one you have in mind and  
press Search**

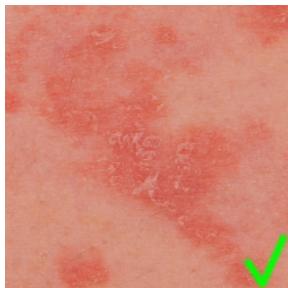
Search

# CBIR in dermatology



Query

Results



# CBIR in dermatology



Query

Results



# k-nearest neighbors retrieval (How CBIR is done)

# Feature vector computation

Reduce the image content to some feature vector (descriptor)



# Color feature vector representation of an image



$(R_h, G_h, B_h, R_l, G_l, B_l) = (89, 110, 41, 50, 172, 30)$   
Use of the colors of healthy and lesional skin

# MPEG-7 image descriptors

- <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>

# k-nearest neighbors for database retrieval

- 1) Compute vector representations for query object and each object stored in the database.
- 2) Compute distance between query **a** and each object **b** in the database:

$$L_2(a, b) = \left( \sum_{i=1}^n (a_i - b_i)^2 \right)^{1/2}$$

- 3) Sort distances from query to database objects.
- 4) Select the k database objects with the smallest distances to the query object.

# Other possible distance measures

Minkowski metric  $L_m(a,b) = \left( \sum_{i=1}^d |a_i - b_i|^m \right)^{1/m}$

Special cases:

$L_1$  norm (Manhattan or city block distance)  $L_1(a, b) = \sum_{i=1}^d |a_i - b_i|$

$L_2$  norm (Euclidian distance)  $L_2(a, b) = \left( \sum_{i=1}^d (a_i - b_i)^2 \right)^{1/2}$

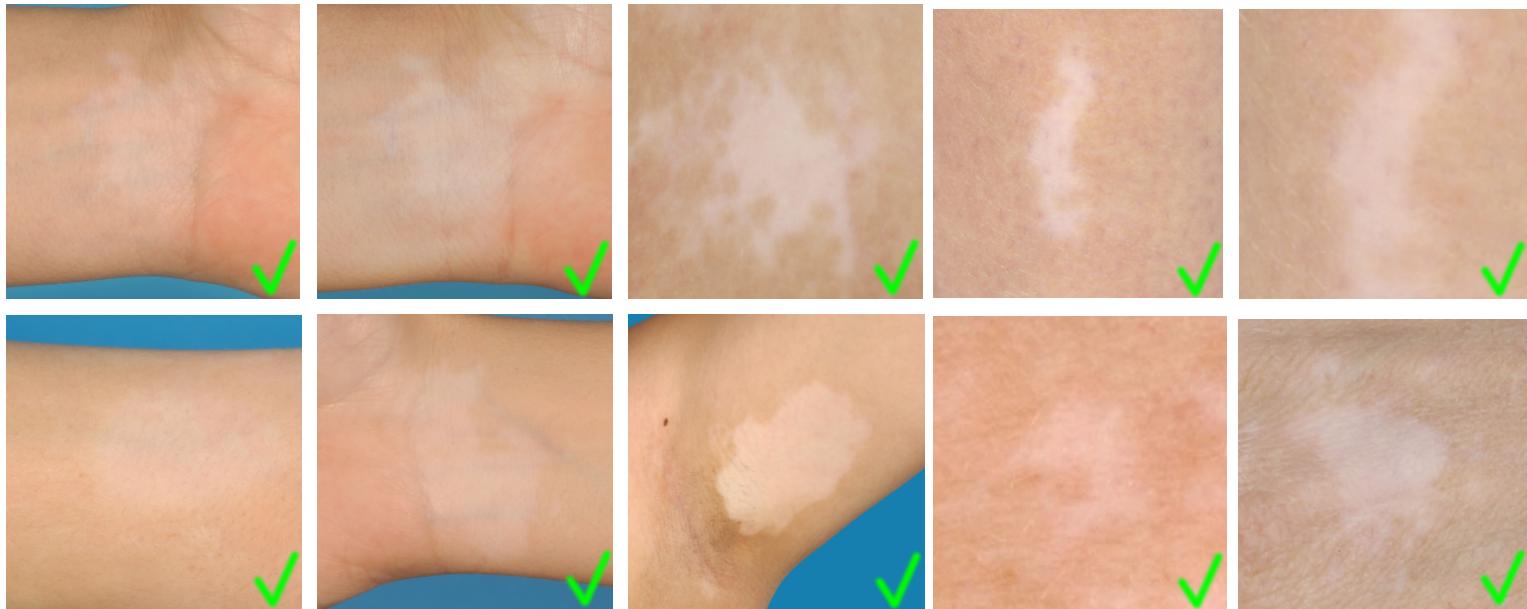
$L_\infty$  norm  $L_\infty(a, b) = \max_{i=1}^d |a_i - b_i|$

# CBIR in dermatology



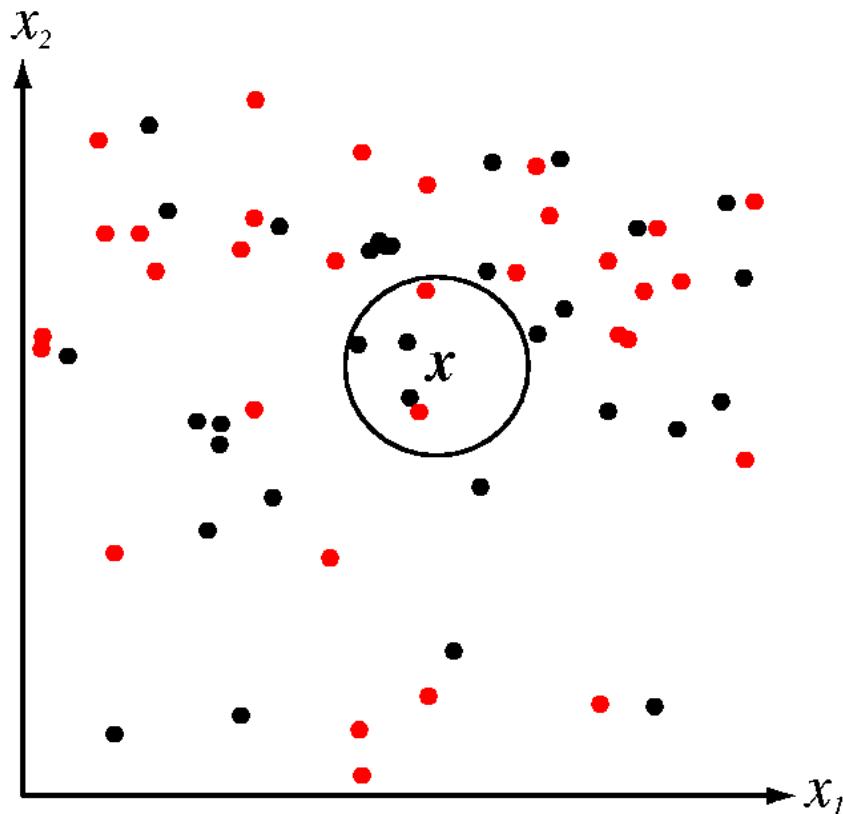
Query

Results



# $k$ -nearest neighbor retrieval

Retrieve the  $k$  database objects nearest to a query  $x$

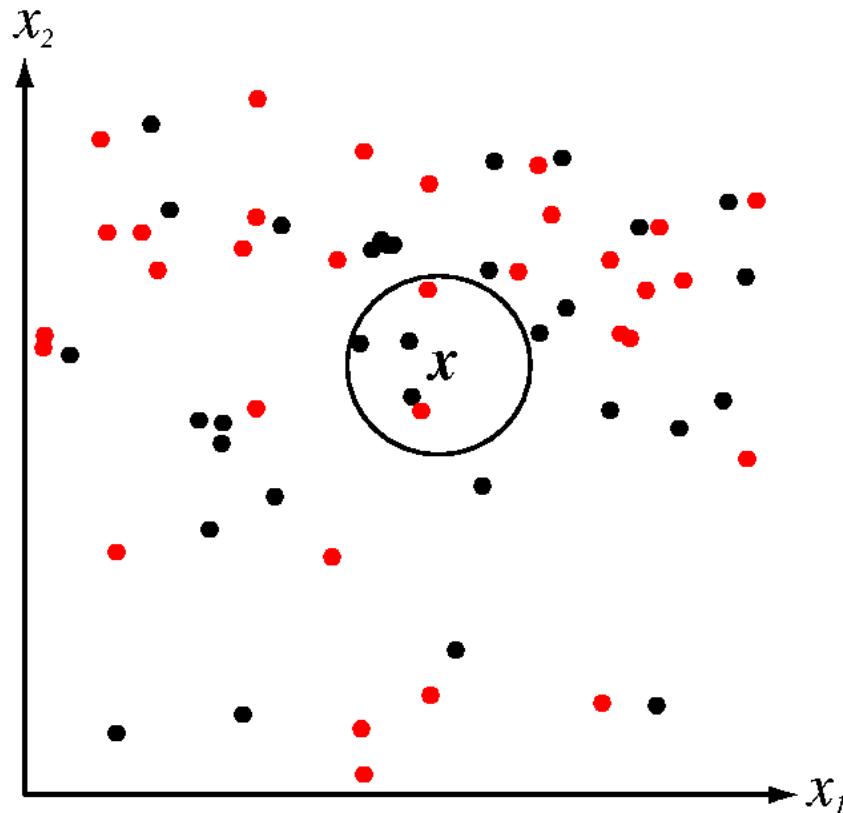


from Duda, Hart, Stork (2001) Pattern classification

# k-nearest neighbor classification

# k-nearest neighbor classification

Assign object  $x$  to the class most frequently represented among the  $k$  nearest objects from the training data set



Occurrence frequency  
is proportional to  
probability density.

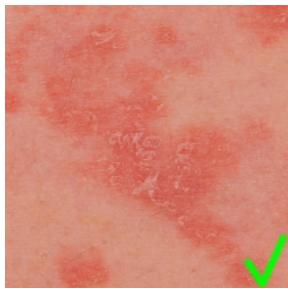
Therefore, k-nn is related  
to Bayesian classification!

# Example of k-NN classification



Query

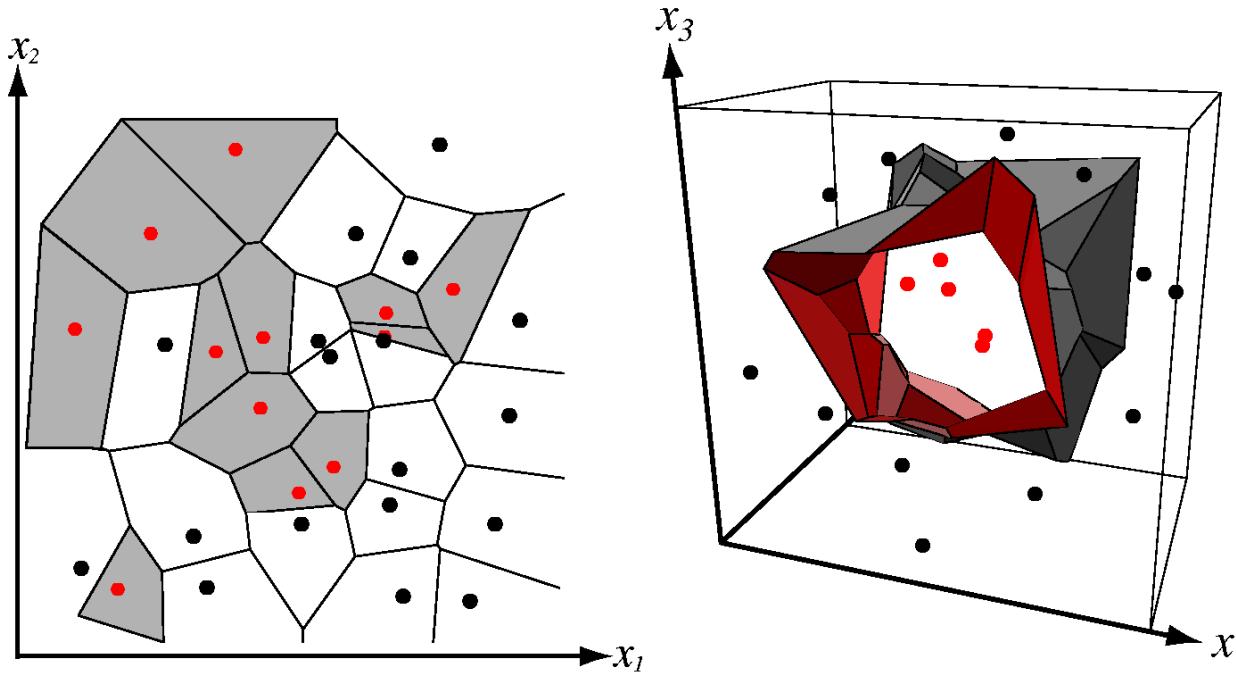
Results



8 of 10 nearest neighbor results belong to the class 'reddish lesion'

# Nearest neighbor classification ( $k = 1$ )

Assign a new feature vector  $\mathbf{x}$  to the class of the closest object from the training data set



from Duda, Hart, Stork (2001) Pattern classification

Voronoi tessellation (diagram) of the feature space.  
This is just an illustration of how irregular the decision boundary can be.  
In practice, no such tessellation of the feature space is computed.

# k-nearest neighbor classification (advantages and disadvantages)

## Advantages:

- 😊 no assumptions about the distributions (e.g. Gaussian)
- 😊 uses all available training data (no information reduction)
- 😊 has good theoretical foundation (related to Bayes theory)

## Disadvantages:

- 😢 Distances need to be computed to all training data points
- 😢 All training data needs to be stored

# Summary of concepts

- Content based image retrieval (examples)
- Example of color feature vector
- k nearest neighbor retrieval
- k nearest neighbor classification
- nearest neighbor classification
- distance measures
- Pro's and con's of k-NN classification