

Image Retrieval on Co-registered Confocal Microscopy Image Collections

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Edith Langer

Matrikelnummer 0725806

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eduard Gröller
Mitwirkung: Dipl.-Math. Dr.techn. Katja Bühler

Wien, 26.08.2014

(Unterschrift Verfasserin)

(Unterschrift Betreuung)

Image Retrieval on Co-registered Confocal Microscopy Image Collections

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieurin

in

Medical Informatics

by

Edith Langer

Registration Number 0725806

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Eduard Gröller
Assistance: Dipl.-Math. Dr.techn. Katja Bühler

Vienna, 26.08.2014

(Signature of Author)

(Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Edith Langer
Scheffelstraße 7/10/10 , 1210 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum)

(Unterschrift Verfasserin)

Acknowledgements

The time of my studies would not had been that enjoyable or even possible without some special people. I want to thank all of you who accompanied me in this time and made it unforgettable.

I want to thank Meister Eduard Gröller for supervising and reviewing this thesis. I also want to thank Katja Bühler for giving me the opportunity to write my thesis at the VRVis, for her input and for reviewing.

My thanks go to Florian and Martin who always had time to give me advices and participate in brainstorming processes.

I also want to say thank you to my colleagues at VRVis, who became friends, by and by, for listening and encouraging me.

Thanks to Christoph who supported me not only in technical issues.

I'm glad we founded the MIST (Medizinische Informatiker Stammtisch), although we have not meet as regularly as we should. It is always fun with you guys and talking to you showed me that finishing the thesis takes always longer as one thinks.

Thank you, Harri, for the motivating New Zealand calendar and also for helping me wherever and whenever you can. Thanks for cheering me up on the very last meters of finishing my thesis.

Special thanks go to Maria who went through thick and thin with me the last years.

Finally, I want to thank my grandmother, my sister and brother and especially my parents for their patience and their support in every possible way during my entire studies.

Abstract

Biologists at the Institute of Molecular Pathology (IMP) in Vienna scan brains of the species *Drosophila melanogaster* with a confocal microscope to find relations between genes, brain structure and behavior. The database contains now more than 40.000 volumetric images, which makes it time-consuming to search for an image of interest. For biologists it would be very helpful to have a method which can be used to search for specific images and works on the perceptual level of content.

The aim of this thesis is to develop a Content Based Image Retrieval (CBIR) method customized for 3D fly brain images. A biologist can choose an image which shows interesting gene expressions and as result images which are visually similar should be retrieved. Exhaustive literature research shows that in the biological field nothing comparable exists. However, CBIR plays an important role in the medical domain, which deals also with 3D images and therefore publications in this area can be seen as related.

The voxelwise comparison of two images would be on the one hand computationally expensive and on the other hand not practicable due to image registration errors and anatomical variations of neuronal structures. Creating maximum intensity projections from three directions and applying a principal component analysis on the gray values overcomes the before mentioned drawbacks and delivers satisfying results.

The fly brain can be divided into regions, so-called neuropils. The proposed method works on the basis of neuropils. This has, among others, the advantage that not only a global similarity can be computed, but also a comparison of images based on only some of the neuropils is possible.

An extensive evaluation of the developed method is given including a parameter space exploration. For example, different lengths of the feature vectors, which describe a fly brain in a lower dimensional space, are tried and also different distance measures are tested. The evaluation shows satisfying results and that the method facilitates the work of biologists when they are looking for similar images to create a hypothesis about the connection of genes and behavior.

Kurzfassung

Biologen des Forschungsinstitutes für molekulare Pathologie in Wien akquirieren 3D Bilder des Gehirns der Spezies *Drosophila melanogaster* mittels eines Konfokalmikroskops. Mit Hilfe dieser Bilder untersuchen sie den Zusammenhang zwischen Genen, Gehirnstruktur und Verhalten. Die Datenbank umfasst mittlerweile mehr als 40.000 solcher Volumenbilder. Diese große Datenmenge erschwert den Biologen die Suche nach bestimmten Bildern und ist sehr zeitaufwendig. Eine Methode zur Suche von Bildern basierend auf dem Inhalt würde ihre Arbeit sehr erleichtern.

Das Ziel dieser Arbeit ist die Entwicklung einer Methode, welche es ermöglicht nach Fliegenhirnbildern zu suchen, die aufgrund ihres Aussehens von Interesse sein könnten. Biologen sollen die Möglichkeit haben ein Bild, das interessante Gehirnstrukturen zeigt, auszuwählen und die Methode soll als Ergebnis visuell ähnliche Bilder zurückliefern. Eine ausgiebige Literaturrecherche hat ergeben, dass es noch keine Lösungsansätze für dieses Problem in der Biologie gibt. In der Medizin spielt die Bildsuche nach Ähnlichkeiten eine große Rolle und da auch in diesem Bereich mit 3D Bildern gearbeitet wird, kann man Publikationen aus diesem Bereich als relevant ansehen.

Voxelweise zwei Bilder zu vergleichen wäre einerseits sehr rechenaufwändig und andererseits nicht zielführend aufgrund von Registrierungsfehlern bei den Bildern und der anatomischen Schwankung neuronaler Strukturen. Um diese Nachteile zu umgehen, wird eine Dimensionsreduktionsmethode (PCA) auf die Grauwerte von drei Maximumintensitätsprojektionen des Bildes angewandt.

Das Fliegenhirn kann anhand einer Vorlage in verschiedene Regionen unterteilt werden, sogenannte Neuropile. Der entwickelte Ansatz arbeitet auf Basis dieser Neuropile. Das hat, unter anderem, den Vorteil, dass man nicht nur eine globale Ähnlichkeit berechnen kann, sondern, falls gewünscht, auch nur Teile des Bildes miteinander vergleichen kann.

In dieser Diplomarbeit werden die Ergebnisse einer umfassenden Evaluierung aufbereitet und unter anderem wird auch die Auswirkung verschiedener Parameter untersucht. Das ist zum Beispiel die Länge des Vektors, der ein Fliegenhirn in einem niederdimensionalen Raum beschreibt, oder die Methode mit der die Distanz zwischen zweier solcher Vektoren berechnet wird. Die Evaluierung liefert zufriedenstellende Resultate und zeigt, dass durch die entwickelte Methode die Arbeit der Biologen auf der Suche nach ähnlichen Bildern, um eine Hypothese über das Zusammenspiel von Genen und Verhalten aufzustellen, erleichtert werden kann.

Contents

1	Introduction	1
1.1	Problem Statement	1
1.2	Thesis Overview	3
2	Background	5
2.1	Drosophila Melanogaster	5
2.2	Data Acquisition	9
2.3	Existing Infrastructure	10
3	Related Work	13
3.1	Background of Image Retrieval	13
3.2	CBIR in Medicine and why is it related to this work	18
3.3	Retrieval of 3D Images	20
3.4	Retrieval through Dimension Reduction	25
3.5	Relation to this work	26
4	Methodology	27
4.1	Maximum Intensity Projection	27
4.2	Dimension Reduction	27
4.3	Distance between Feature Vectors	30
5	Implementation	37
5.1	Overview of Approach	37
5.2	Preprocessing	37
5.3	Dimension Reduction using PCA	39
5.4	Distance Measure	43
6	Evaluation and Results	45
6.1	Evaluation Measures for Retrieval Tasks	46
6.2	Global Evaluation	49
6.3	Local Evaluation	63
7	Conclusion and Future Work	79
7.1	Summary and Conclusion	79

7.2 Future Work	80
Bibliography	83

Introduction

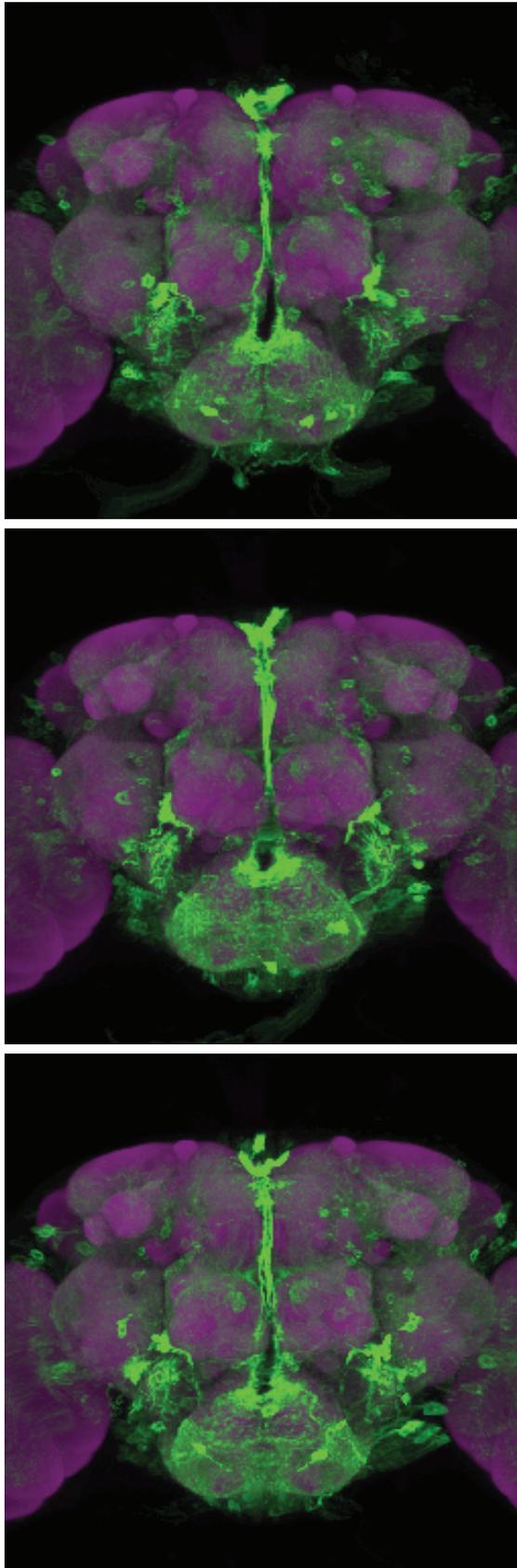
Researchers from the Institute of Molecular Pathology (IMP) in Vienna investigate the Central Nervous System (CNS) of the species *Drosophila melanogaster*, the so-called fruit fly. It is a popular model organism because it has many genes comparable to the human ones, but its genome is not so complex. It can be used, for example, to study behavior and try to answer the question how decisions are made or analyze genes responsible for diseases and find possible therapies.

The IMP has a collection of about 15.000 transgene *Drosophila melanogaster* lines. Flies with the same gene set belong to the same line and have similar expressions. Figure 1.1 shows an example of such a line consisting of three different flies as 2D preview images. More than 40.000 flies were scanned with a confocal microscope. The acquired volumetric images show gene expressions in neuronal structures and are stored in a database together with meta-information, like sex, tissue type and driverline. By means of these images they try to answer their research questions.

1.1 Problem Statement

Nearly everybody is familiar with the problem of having a lot of images stored digitally and not being able to find a specific image coming into ones mind because the name of the image or the path where it is stored can not be remembered. In this case it would be really helpful to search for images by content. For example, one wants to create a photo collage as birthday present and wants to use photographs which picture the presentee. The time needed to search for these pictures can be reduced if an image with the person's face on it and a retrieval method is available. With the picture as example the retrieval method delivers all available images containing the person.

The biologists at the IMP have a similar problem. They have a database containing thousands of images depicting fly brains and they want to filter out the ones which are relevant for their research problem. A manual screening of all images is not feasible because of the number of



2

Figure 1.1: Three brain images from the same line showing similar expressions.

available fruit fly scans. Therefore a computer-assisted method to browse the database based on image content would facilitate the scientific research work of the biologists. They are often only interested in images showing specific expressed neurons. A so called search by example retrieval method would be one way to solve this problem. If the biologists have an image showing the neuron they are interested in, they can use it as an example and the method retrieves images from the database with a similar content and therefore the chance is high that the specific neuron is visible.

The fly images have some special characteristics. They are volumetric, registered to a template and most of them are noisy, but at the same time contain small structures which may be of importance. Hence two questions have to be answered:

1. How can the similarity between two images be defined?
2. How can the image characteristics be exploited respectively what challenges do they involve?

A voxel by voxel comparison is not feasible because on the one hand it would be computationally too expensive and time consuming and on the other hand it is inflexible and can not take the image registration errors and the anatomical variations of neuronal structures into account. Therefore, to achieve efficiently good retrieval results based on image content, it is necessary to reduce the dimensions of the images and find the most outstanding expression patterns.

The contribution of this thesis is the development of a retrieval method which can deal with co-registered 3D images showing the brain of fruit flies. Additionally it should be possible to do a global or local similarity search. Sometimes biologists are not interested in globally similar images. However, they want to find images which are similar in specific regions and the similarities of the other brain regions do not matter. The retrieval has to be efficient, which can be achieved by a suitable dimension reduction.

1.2 Thesis Overview

The following chapter gives an overview of the anatomy of the model *Drosophila melanogaster* and how the biologists at IMP acquire the brain activity. Also a short outline on the existing project infrastructure is given. Chapter three introduces the concept of image retrieval and presents related work in the medical domain. The methods and approaches which are used for the practical part of the thesis are explained in chapter four. Chapter five depicts the workflow of the proposed method and describes the single steps in detail. In chapter six the most common evaluation measures for retrieval are listed and the evaluation methods for the global and the local case are discussed. The thesis concludes with a discussion about the contribution of this work and about possible future work.

Background

2.1 *Drosophila Melanogaster*

Most people know the *Drosophila melanogaster* because they come in hundreds to occupy spoiled fruits. Therefore the species is well-known as fruit fly. A drawing of a male and female fruit fly can be seen in Figure 2.1. However, it is a very important and well studied model for life science not only since the genome is deciphered. In comparison to other common model organism like *C. elegans* or zebra fish it has a more complex genome and a more extensive behavior repertoire. Fruit flies have the senses of sound, vision and smell and also higher level intelligence like learning and memory [80]. Mice can be seen as competitive models, but their brains consist of about 1000-fold more neurons, which makes it hard to understand how they

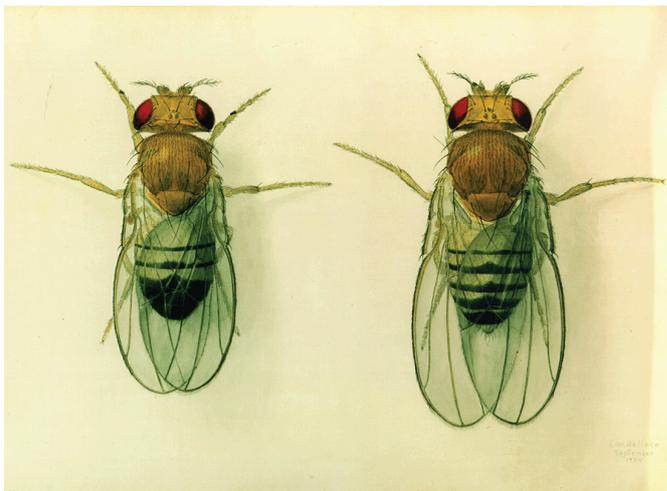


Figure 2.1: Drawing of *Drosophilas* - a male on the left and a female on the right [30].

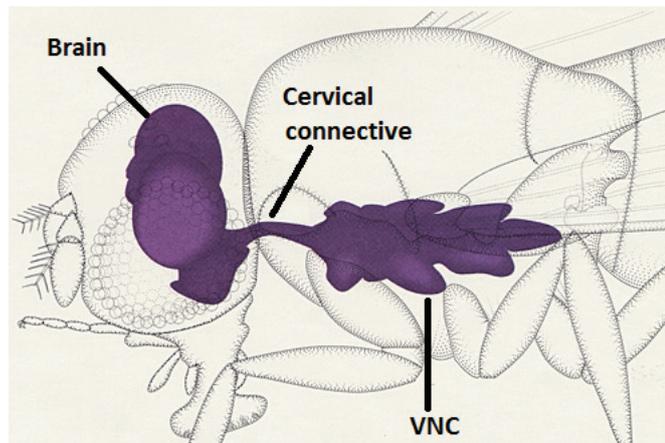


Figure 2.2: A drawing of a fruit fly with the CNS in violet [29].

work together. Therefore *Drosophila* is a compromise between the number of neurons and richness in behavior [65].

Another advantage is that the 1 to 7 mm small flies can be easily reared in a great number [105]. They can be maintained without great effort. In addition they have a very short generation time and a high reproductive rate. Within 14 days a larva evolves from the egg and transforms into a fly. This fly is then fertile within 12 hours. A fertilized female fly can lay 300 eggs.

The fact that more than 60 % of the genes responsible for human diseases have an ortholog in the *Drosophila* genome makes the *Drosophila* a suitable model to identify genes that effect complex traits [53]. Traits that are shown both in humans and *Drosophilas* are for example circadian rhythm, sleep, drug responses and locomotion. Mackay and Anholt [53] use the *Drosophila* model to study alcohol and drug dependence, sleep and neurodegenerative diseases like Alzheimer and Parkinson.

From the anatomical point of view humans and fruit flies have also commonalities. Both have a CNS consisting of the brain and the spinal cord respectively the Ventral Nerve Cord (VNC) as it is called for *Drosophila*. In Figure 2.2 the *Drosophila*'s CNS is schematically delineated. Another shared characteristic is that the brain is made up of two symmetric hemispheres as it can be seen in Figure 2.3 for the *Drosophila*.

Very important components of the nervous system are the nerve cells also known as neurons. They are responsible for processing and transmitting information and form together a neural network. The number of neurons in *Drosophila* is still uncertain, but estimates assume that around 200.000 neurons exist [65]. In contrast to vertebrates invertebrates have many neurons which are unipolar [9]. This means that a single process emanates from the cell body which later bifurcates into a pre- and post-synaptic part. A neuron consists of the following objects:

Cell body

Cell bodies are blob-like structures positioned within the cortex this means in the outer layer. They can differ in size and shape.



Figure 2.3: Schematic Drosophila brain.

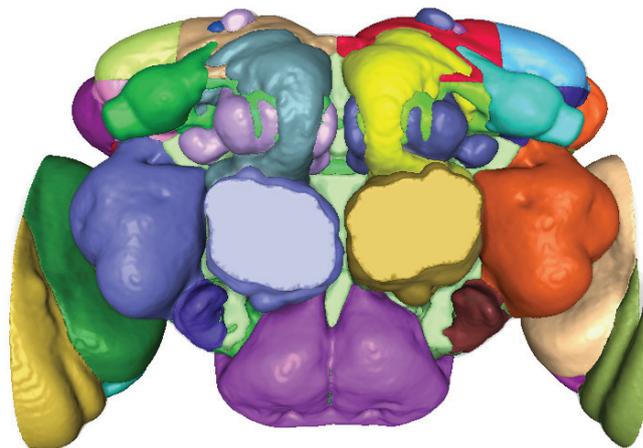


Figure 2.4: Drosophila brain divided into neuropils.

Projection

Projections are thin tubular structures which can reach a great distance in the brain. They bifurcate at some point into a dendritic branch and an axonal branch. Dendrites receive incoming signals and axons send signals.

Arborization

Arborizations are dense terminal branching structures. A neuron can consist of any number of arborizations. Unipolar neurons do not receive input signals on the cell body, therefore arborizations contain synapses which are needed to communicate with other neurons.

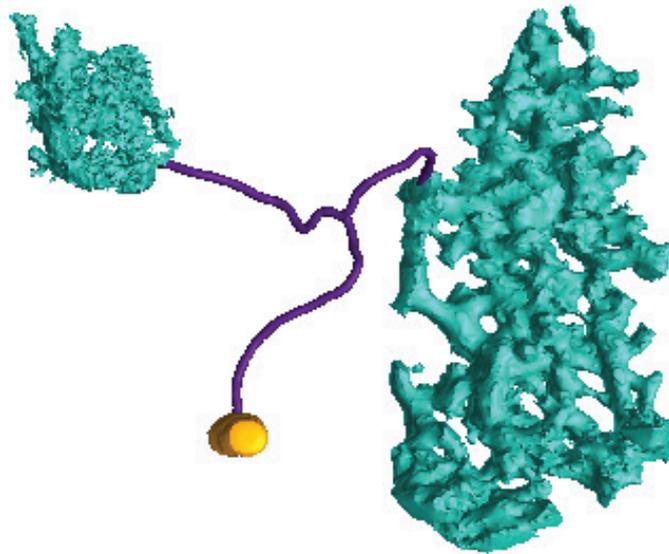


Figure 2.5: Neuron consisting of a cell body (yellow), a projection (violet) and two arborizations (turquoise).

The mentioned neuronal objects belong to exactly one neuron. An example of a neuron can be seen in Figure 2.5. The cell body is delineated in yellow and the projection in violet has two branches each terminating in an arborization shown in turquoise.

It is useful to have an atlas like it exists for the human brain. Therefore the fruit fly brain gets divided into regions which are called neuropils. Some of the neuropils can be easily separated because of their anatomy. Others have to be divided based on small landmarks [66]. The optimum would be a division into regions which correspond to certain functions. Recently Ito et al. [38] published a paper defining a nomenclature which should be the standard for future work.

The IMP defines in total 53 neuropils for the *Drosophila* brain based on Otsuna's and Kei's classification [66]. Figure 2.4 gives an overview of the neuropils indicating each neuropil in a different color. Not all neuropils can be seen in a single 2D view of the brain because of overlaps. In the ideal case, neurons located in a certain region in the brain fulfil specific tasks. This is true, for example, for neurons in the optical lobe which are responsible for vision.

A present research field in neuroscience is how information processing and storage work in brains. One of the research questions deals with the derivation of a decision from sensory input, internal states and individual experience [25]. Dickson [25] uses the *Drosophila* and their mating behavior to explore decision making and the behavior based on it. The goal is to find out which neurons build a functional entity, a so-called neural circuit, and how they are interconnected. As a result genetic manipulations can be examined to change the behavior. The challenge is to detect the neurons which are involved in a certain behavior. Most neurons do not respond to

only one stimulus. Therefore the question is which activity pattern of neurons produce which behavior [65]. However, even the output of a small neural circuit is the outcome of a complex interplay between its elements. Olsen and Wilson [65] give an overview of how to crack neural circuits if they are not too complex.

2.2 Data Acquisition

Brain and VNC have to be carefully dissected from the rest of the *Drosophila* body to be acquired and in further consequence visualized and explored. The IMP uses confocal microscopy to scan the CNS of *Drosophila*. Confocal microscopy is based on laser light, which can be focused on a particular 3D position. The reflected light of a certain wave length is measured at this position. A 2D image can be generated if the light is measured at every position at the same depth. Repeating this procedure at different depths results in a stack of 2D images which can be transformed into a 3D image. The volumetric images acquired by the IMP have a size of $420 \mu\text{m} \times 420 \mu\text{m} \times 165 \mu\text{m}$ and are sampled with a resolution of $768 \times 768 \times 165$ voxels.

For expressing targeted genes in the *Drosophila* brain the GAL4/UAS-system [11] is very common. GAL4 gets inserted into the genome and encodes yeast proteins. It can be expressed in different tissues, however in *Drosophila* it is typically a subset of neurons one is interested in. Several thousand genetic variations exist and are called GAL4 lines. The counterpart to GAL4 is an Upstream Activation Sequence (UAS). As the name indicates, UAS activates the gene next to it. But this transgene remains silent until GAL4 binds. Only then UAS gets activated and turns on the neighboring gene. The same applies for GAL4 - it has no effect on cells if there is no UAS to bind. Figure 2.6 gives a visual overview how the GAL4/UAS-system works. To study the cells which are expressed by GAL4 the Green Fluorescent Protein (GFP) is often used as target gene for the UAS. Due to its fluorescent characteristic it can be tracked with a microscope and the distribution in time and space is visible in a living organism. To achieve this two flies are needed. On the one hand a fly is needed with GAL4, controlled by a specific enhancer, inserted at a defined genomic location by site-specific integration to express the targeted neurons. Another fly is needed with the UAS-GFP transgene in its genome. Their offspring have then both, GAL4 and UAS-GFP, in their genes which bind and activate GFP in the targeted cells. They can then be analyzed with a microscope. High resolution 3D images can be generated showing the expressed patterns in one channel and the brain tissue marked by the nc82 antibody in another one [68]. Figure 2.7 shows such an example as 2D image using GFP to make cells visible.

The fact that brains of different flies underlie anatomical variations and the spatial differences during acquisition makes it necessary to co-register the scanned images. Only then it is possible to match and compare these images. Initially a manual alignment of the volume is necessary. Then a non-rigid transformation based on the parallel warping method developed by Rohlfing and Maurer [73] can be applied. As registration of 3D images is a complex problem they take advantage of parallel computations to achieve low computation times. Their method is a multiresolution approach based on B-spline transformations (see Rohlfing and Maurer [73] for more details). For registration the channel containing the brain tissue is used. It is registered against a standard brain template which is the outcome of averaging a set of representative brain images.

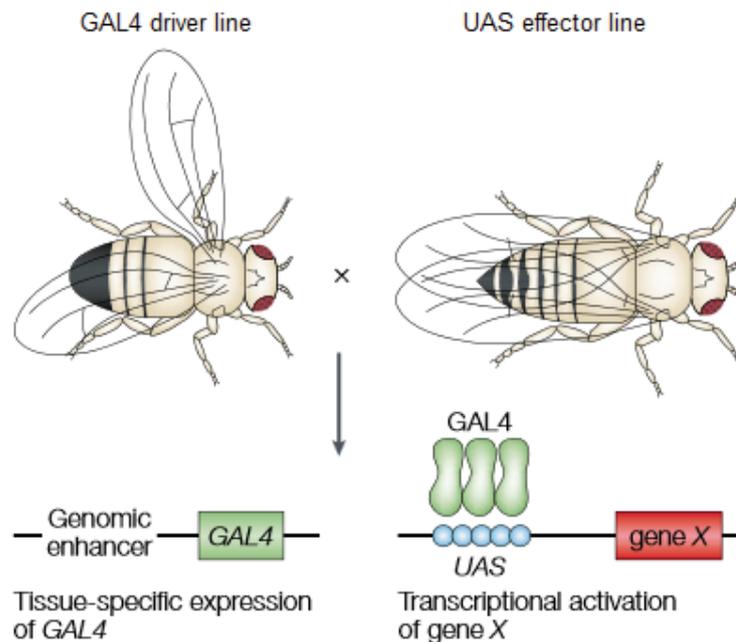


Figure 2.6: Schematic representation of the GAL4/UAS-system for Drosophila (adapted from [89]).

Figure 2.8 shows an example of a co-registered fruit fly brain as 2D image. The template in magenta and the expressed neurons in green can be seen. Blob-like cell bodies, tubular projections and arborizations, looking like dense branching structures, are visible. The point cloud all over the brain, which reminds of fog, is noise.

2.3 Existing Infrastructure

The infrastructure to manage and mine the data at IMP consists of several components. An early version was published by Bruckner et al. [12] in 2009. The infrastructure changed a lot since then. A short overview of the current components is given in this section.

BrainBase

BrainBase is a relational database holding the user management, relations between neurobiological objects, like cell bodies, arborizations and neuropils, and information about the image collection.

BrainBaseWeb

BrainBaseWeb is a web application to search and navigate through *BrainBase*. It is possible to do a textual search and retrieve objects and images with a certain name. From the fetched results some characteristics, a preview image and a 3D preview and for images

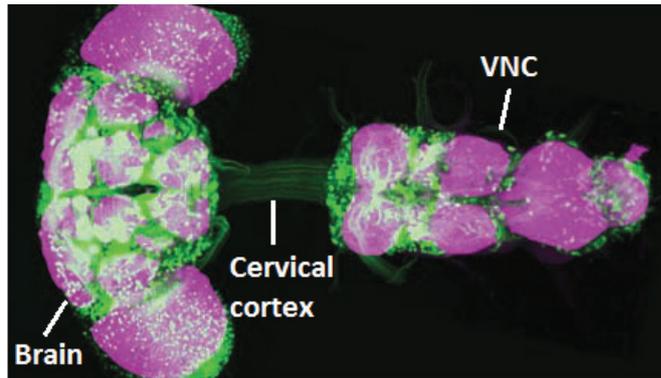


Figure 2.7: CNS of a *Drosophila* with the expressed, fluorescent cells in green and the neuropils in magenta acquired with the GAL4/UAS-system (adapted from [25]).

additionally a staining profile can be displayed. An alternative search for images is by line. It returns images belonging to the specified line. If one is interested in images with a certain amount of staining in specified brain regions the anatomical search is the right choice. One or more neuropils and the amount of staining a neuropil should contain at least respectively maximally can be selected. For more complex search requirements the brainbase query language can be used. It is a simple query language to find objects and images by using conditions that can be combined with AND and OR.

BrainGazer

BrainGazer is a desktop application that allows the user an interactive visualization and exploration of 3D images and 2D slice views. It fully integrates *BrainBaseWeb* and therefore searching for and loading objects respectively images into *BrainGazer* is straightforward. In contrast to *BrainBaseWeb* a semantic search by different visual queries is possible in *BrainGazer*. Depending on the type of visual query similar neurobiological objects or images can be retrieved. The user can draw a path on the displayed image and objects which are spatially close respectively images with a similar expression are retrieved. Hence, visual queries are suitable for local similarity search. Including the practical work that arises in the course of my thesis gives the user the possibility to find similar expression patterns globally respectively in interesting neuropils.

BrainWarp

BrainWarp is the component which is responsible for the preprocessing of new images. This includes the registration of the raw image to the template and creating an entry for *BrainBase*.

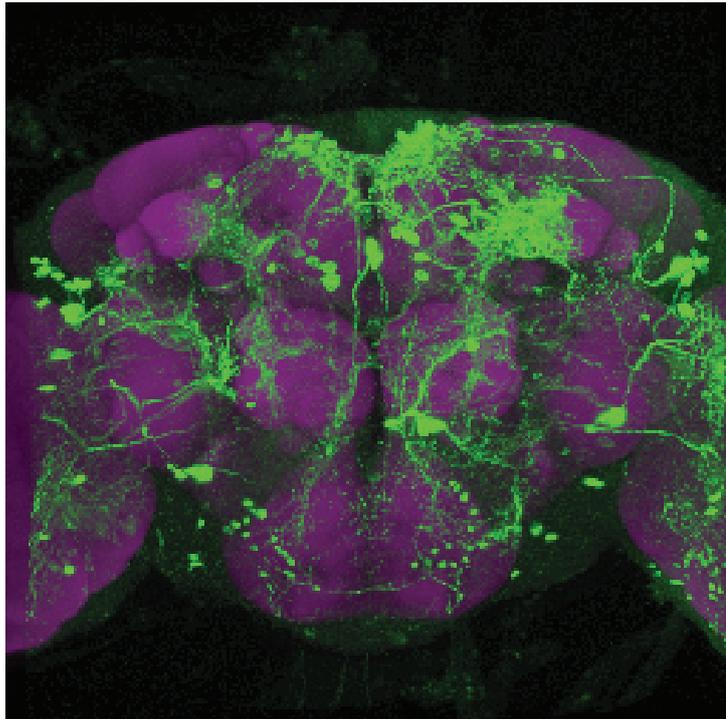


Figure 2.8: Example of a fruit fly brain (simplified to a 2D image) scanned with a confocal microscope and registered against a standard brain.

Related Work

3.1 Background of Image Retrieval

The number of digitally stored images increases rapidly. Therefore a technique to search efficiently for particular images is needed. The first approach to do so was a text-based search [28]. The user enters one or more keywords and retrieves images, which are tagged with these words or embedded in text containing the keywords. This method has some disadvantages. Every image has to be annotated. If this is done manually it is a time consuming task and if experts are needed, for example for medical images, also expensive. On the other hand the annotation result can be inaccurate if it is done automatically. Datta et al. [21] give an overview of annotation approaches. Another disadvantage of the text-based search is the existing discrepancy between the tags which are used to describe an image and the words which are used for searching because of subjectiveness. A further problem is that words are often not powerful enough. Describing, for example, a person's face with words is difficult.

To overcome these drawbacks Content Based Image Retrieval (CBIR) was invented in the early 1990s [74]. CBIR makes use of information extracted directly from images and tends to work on the perceptual level of image content. Perceptual properties are for example color, texture and shape [84]. The extracted features are then used to compute the similarity between images. Figure 3.1 shows the usual workflow of a CBIR system.

Instead of using text to perform an image search, a query image is used. Flickner et al. [28] implemented three possibilities in their system. The user can select an example image which looks similar to images that should be retrieved. Painting or sketching an image to describe color, texture, shape of objects and their spatial arrangement is another way of defining a query image. The user can also select color and texture patterns. If one is interested in outdoor scenes the color distribution can be chosen with 60 % green and 40 % blue and a grass texture can be selected. Prominent examples of CBIR systems are QBIR (Query by Image Content) from IBM [28], Virage [4], which are both commercial, and Photobook [69]. A famous example with the world wide web as data source is Google Images. Starting the service in 2001 the search was only text-based. Filename, link text and surrounding text are searched for the keywords the user

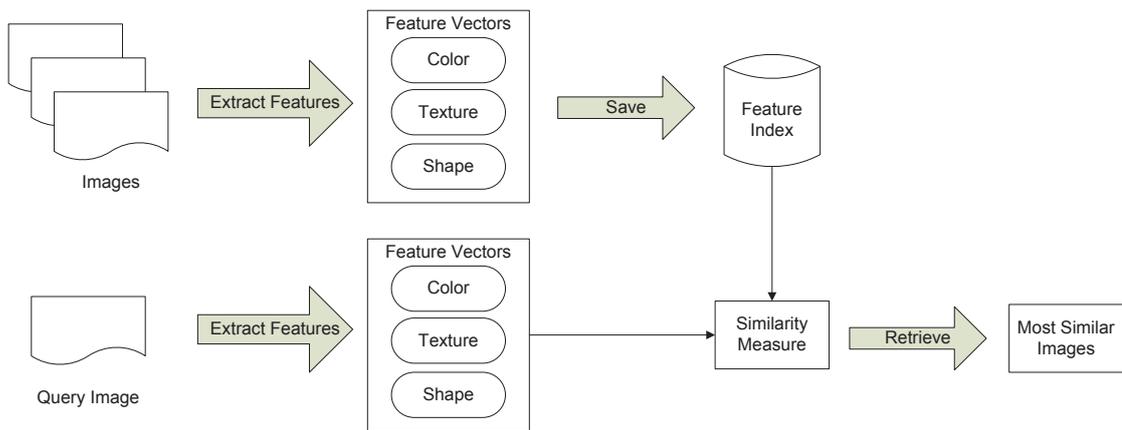


Figure 3.1: Usual CBIR workflow.

has entered before. Since 2009 it is possible to use an example image for the search [62]. The user has the possibility to refine the search by adding more example images for instance from the retrieved result set in the previous step. The combination of text-based search and CBIR seems promising [2].

In CBIR two gaps are to overcome, the sensory and the semantic gap [84]. The sensory gap deals with viewpoint, occlusion, clutter, and illumination of objects in images. Relevant domain knowledge can help to bridge this gap [84]. The semantic gap exists between low-level visual features, like color, texture and shape, and the objects and their meanings for a person in an image like table or chair, which are high-level semantic concepts. There exists several approaches to reduce this gap, but it has not been possible to fully eliminate it yet [33]. Yang and Zhu [108] and Min and Shuangyuan [58] give an overview about this challenge and possible solutions. Relevance Feedback is one way of reducing the gap between low level features and high level concepts [75]. It involves the user in the retrieval process and thereby improves the retrieval result in a way that fits more the needs of the user. Relevance feedback is an iterative process to refine the query. The user can indicate if an image is relevant or nonrelevant. Rui et al. [75] suggest to use five relevance levels from which the user can choose one. Zhou and Huang [112] give a comprehensive overview of relevance feedback in CBIR.

It depends on the image set which feature describes the images best and therefore should be chosen. The combination of low-level features can lead to a better description of images and to a more distinctive feature vector [103]. However, a high dimensional feature vector can lead to correlating and irrelevant elements as it is stated by Silva et al. [82]. A dimension reduction algorithm, for example Principal Component Analysis (PCA), can be performed to remedy this.

Features can be extracted globally for the whole image or locally for image parts. Global feature extraction can be done without any preprocessing but a lot of spatial information is lost. For extracting local features images have to be divided. The simplest way of doing this is to divide the image into small, disjunctive blocks of the same size. Superior results can be achieved if a segmentation algorithm is used and features are computed for segmented regions, however segmentation is still an open research area.

CBIR is a rapidly growing research area [20]. In the following sections an overview of the most popular feature extraction methods for color, texture and shape are given. The collection of mentioned methods does not claim to be exhaustive.

3.1.1 Color Features

Color is widely used for feature extraction. The retrieval results strongly depend on the choice of the color space. Several different color spaces exist [96]. The probably best known is the RGB space. RGB does not correspond to human perception and is only useful for image retrieval if images are taken under the same circumstances [84]. A color space that overcomes this disadvantages is for example HSV. It has three channels for hue, saturation and lightness. Non-linear transformations are necessary to switch between RGB and HSV [86].

The color histogram [92] is one of the most popular color features for image retrieval. It represents the color distribution of an image. The color space has to be quantized and pixels are assigned to the bin they lie in. Histograms from two different images can then be compared for example with histogram intersection. Ma and Zhang [52] compare among other things the retrieval performance of color histograms in different color spaces and with various numbers of bins. Color histograms have the advantage of being invariant to rotation and translation and the change in the histogram is small if the scale differs slightly or partial occlusion occurs. An essential disadvantage is that all spatial information is lost. A simple approach to obtain at least some spatial information is to divide images into blocks and compute for each block a color histogram.

Stricker and Orengo [90] suggest to compute the first three moments, that are mean, standard deviation and skewness, of each color channel of an image. In the HSV space this leads to nine features. It is possible to assign different weights to features. For example, it is often more important that hue is more similar than saturation and value. The authors claim that the approach generates more robust results than color histogram does and also the retrieval is faster.

Dominant Color Descriptor (DCD) [55] is also a more compact feature based on color than the color histogram. Initially the colors of an image have to be clustered. A widely used approach is the generalized Lloyd algorithm [51] which results at most in eight dominant colors. The structure of the extracted feature F is defined as

$$F = \{\mathbf{c}_i, p_i, v_i, s\}, \quad i = 1, 2, \dots, N \quad (3.1)$$

For each of the N dominant colors four parameters are saved. The vector \mathbf{c}_i is the color representation of the dominant color i and p_i is the percentage of this color in the image. The color variance v_i is an optional parameter and defines the variance of the dominant color and the color of the pixels corresponding to this dominant color. The parameter s is a scalar describing the spatial coherency of the dominant colors. It is a term for homogeneity of the image. Like color histogram DCD lacks also of spatial information.

Huang et al. [37] introduce the color correlograms. This feature describes the local spatial color correlation and also the global distribution of it. It defines the probability of finding a pair of pixels with a distance d where the one pixel has the color i and the other one has the color j . To get only a subset of the correlogram Huang et al. [37] propose to compute an auto-correlogram



Figure 3.2: Left: part of a gray-level image; right: Corresponding co-occurrence matrix with distance 1 and orientation 0° .

which considers only pixels at a distance d that have the same color. One has to decide what is the optimal set of distances to use for computing the correlogram. The color correlogram has the big advantage that it is more stable concerning changes in the appearance than color histograms.

Color features described in this section are just some examples which are very prominent. Several adaptations exist and many more completely different features can be found in the literature. For example, Schettini et al. [78] list several possible color descriptors.

3.1.2 Texture Features

It exists no unambiguous definition for texture. Smith and Chang [85] give the following definition: “Texture refers to a visual pattern that has properties of homogeneity that do not result from the presence of only a single color or intensity“. Computing a texture feature takes always the neighborhood of a pixel into account. It does not make sense to talk about the texture of tiger stripes while looking at just one pixel. The choice of the size of the neighborhood is very important and depends on the structure that should be extracted.

Already in 1973 Haralick et al. [34] introduced the co-occurrence matrix or gray-tone spatial-dependence matrix, as they called it, which is still commonly used to extract texture features nowadays. The dimension of the matrix corresponds to the different gray values in the image respectively the image part one is interested in. Often the gray values are quantized to keep the matrix compact. It holds the frequency of two pixels with a distance d in a direction θ whereby one pixel has the gray value i and the other the gray value j . Several of these matrices can be generated with different distances between pixels and with four different orientations (0° , 45° , 90° , 135°). Co-occurrence matrices are always symmetric. An example of a co-occurrence matrix is shown in Figure 3.2. On the left is an image part with gray values from 0 to 3 and on the right is the corresponding co-occurrence matrix with a distance of 1 and an orientation of 0° . Haralick et al. [34] propose 14 different statistical measures, for example energy, correlation, variance and entropy, which can be calculated from such co-occurrence matrices and represent the extracted texture features.

Some of the measures suggested by Haralick et al. [34] can not be explained in terms of human visual perception. Tamura et al. [95] intended to develop texture features which are

visually meaningful and work well with all kinds of textures. They did some psychological experiments and the outcome was compared to the computational measures. A definition of the proposed features, namely, coarseness, contrast, directionality, line-likeness, regularity, and roughness, can be found in [95].

Gabor filters are another way of extracting texture features which correspond to human perception [22]. A 2D Gabor Filter [22] can be represented as a Gaussian modulated sinusoid in the spatial domain. A set of filters with varying parameters is used to analyze an image. Frequency, orientation, Gaussian standard deviation and frequency and angular bandwidths have to be chosen properly to extract the texture contained in an image. Filter responses are generated by the convolution of an image with the filter set as masks. Apart from the possibility to use these responses directly as texture features, Clausi and Jernigan [16] and Andrysiak and Choraś [1] state several feature extraction methods based on the Gabor filter responses. Gabor filters are most suitable for images showing textures with highly specific frequency and orientation properties [10].

This is just an extract how texture can be described. Readers interested in more texture features can have a look at the surveys of Rui et al. [74] and Datta et al. [20].

3.1.3 Shape Features

Besides color and texture also the shape can be used to describe an object. Shape features should optimally be translation, rotation and scale invariant. They can be broadly organised in the two categories boundary based and region based depending if the method takes only the contour into account or also the interior and therefore for example holes [57]. For many shape extraction approaches an initial challenge is the segmentation of an image into foreground, which corresponds to objects, and background. Satisfying shape descriptions rely on good automatic segmentation algorithms which still struggle with complex, heterogeneous image content [20].

Objects can be described for example by their edges. The most common edge extraction methods are Canny [15], Sobel [26], Prewitt [70] and Roberts [72]. Edges can then be used directly to build an edge direction histogram [40]. These histograms can then be compared with histogram intersection. This very simple approach has the disadvantages of not being rotation and scale invariant per se. Jain and Vailaya [40] suggest approaches to attenuate these drawbacks. But they state that even with the improvements the retrieval accuracy for rotated images is only 63 %.

A well-known shape descriptor from the boundary based category is the Fourier descriptor [17]. It has the advantage of being invariant to translation, rotation and scaling [42]. One drawback of this method is that boundary information is needed and that objects exist with no clear contour. Another drawback is that two shapes with the same contours are considered to be equal although the shape content is different. When the boundary of the shape is extracted, N samples are taken. Based on these points a signature function is computed. Different approaches for this task exist. For example, Zhang and Lu [111] suggest to compute for each of the N boundary points the distance to the centroid of the shape. Zahn and Roskies [109] propose to use the tangent angle difference between two successive contour segments as function. With

Equation 3.2 Fourier transformed coefficients are computed.

$$a_n = \frac{1}{N} \sum_{t=0}^{N-1} s(t) \exp\left(\frac{-j2\pi nt}{N}\right), \quad n = 0, \dots, N - 1 \quad (3.2)$$

where N is the number of sample points and $s(t)$ is the signature function. The normalized magnitudes can then be used as feature vector. The similarity between two shapes can be easily computed with the Euclidean distance.

An example of the region based category are moment invariants [36]. Moment invariants take the shape boundary and also the interior region into account. They are invariant to affine transformations like translation, scaling and rotation. Moment invariants are based on the regular moments, which are defined in Equation 3.3 for the discrete case:

$$M_{pq} = \sum_X \sum_Y x^p y^q f(x, y) \quad (3.3)$$

p and q are natural numbers and their sum specifies the order of the moment M . $f(x, y)$ is the intensity function. To be translation invariant the moment invariants have to be normalized with respect to the shape center and are then called central moments. These central moments have to be divided by a factor depending on the order values of the moments to be scale invariant [44]. Hu [36] developed seven invariant moments based on the normalized central moments which are also rotation invariant and with an order of two or three. Because some values for the moment invariants can get quite small, it is recommend to normalize them into the range from 0 to 1 [57]. The seven normalized values represent a feature vector which can be compared using the Euclidean distance.

Fourier descriptor and moment invariant approaches can also be combined and deliver better retrieval results than each of them alone [57]. Many more shape description methods and possibilities to compare shapes have been developed [74], [57], [20].

3.2 CBIR in Medicine and why is it related to this work

Bioimage informatics is quite a new research area [67]. Recently published papers comparable to this thesis can be found mainly in the medical domain using images from different medical imaging techniques and also histological images from microscopic studies.

A useful CBIR system in the medical domain is needed because a vast number of images is generated. Already in 2003 Montagnat et al. [59] stated that a radiology department in an industrialized country hospital generates images of about 10 Terabytes in total in a year. The Radiology Department of the University Hospitals in Geneva produced 80.000 images a day in 2008 [24]. Different image modalities exist, for example, Computer Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), X-ray and ultrasound. Every modality generates images with different characteristics and therefore for each of these images other features are suitable and should be extracted. Domain knowledge is important for a satisfying medical retrieval system. That means it should take into account the image modality and from which body part the image is taken. The domain knowledge can be extracted, if

provided, for example, from the Digital Imaging and COmmunications in Medicine (DICOM) header [7]. DICOM is a standard for an interchange protocol and file format for biomedical images. The DICOM header contains metadata and can provide information about the patient, a case description, information about the acquisition or the diagnosis. The DICOM standard was a prerequisite for the development of a system to store and access digital medical images. Such a system is called Picture Archiving and Communication System (PACS) [47]. It is possible to use the DICOM header for a text-based image retrieval, but it involves the same problems as described in Section 3.1. Two reasons for developing good CBIR systems for the medical domain are the support for physicians in finding medical diagnosis and educational purposes [93]. For a physician it is helpful to find similar cases with proven pathologies. In education it is also important to find images which are visually similar but have a different diagnosis. A lot of research has been done in the last decade in medical CBIR. However, most of the approaches can handle only a single image type and are designed for a specific body part. Müller et al. [60] give an overview which medical CBIR system can handle which image modality. Kumar et al. [45] published a review paper in 2013 that lists about 50 studies from the medical image retrieval area assigned to several imaging modalities. Other challenges are the fact that pathologies are often local and not global, the difference between health and pathology is often small and the variability of anatomical structures between patients.

Why are medical images comparable to fruit fly brain images in the context of CBIR?

Monochromatic

Medical images differ from daily life images. They are monochromatic and therefore color and other features based on color are not expressive. On the other hand lighting conditions do not have to be taken into account [60]. The same holds for the images in the fruit fly brain dataset as information is only held in one channel.

Noise

Another characteristic of medical images is the high amount of noise [32]. Noise is also existent in the fly brain images because of the acquisition with a confocal microscopy as it is explained in Section 2.2 and visible in Figure 2.8.

Registration to template

After scanning a fly brain the image is registered to a template (see Figure 2.3). Medical images used in a CBIR system are often aligned to an atlas, for example brain images are registered to the Talairach atlas [94]. This is also necessary for multi-modality images like the popular PET-CT overlapping. A great advantage, that justifies the effort of registration, is a less complex comparison of registered images.

Local feature selection

Due to the complexity and heterogeneity of medical images it is difficult to describe the whole image. Therefore it is useful to segment the image into regions, which can for example correspond to anatomical structures. Extracting features only from regions a user is interested in is more accurate. This could be for example the lung for lung tumor

research. Also for the fly brain images the features are not computed for the whole image, but for neuropils (see Chapter 5 for details).

A lot of work was invested in the research of CBIR in the medical domain [60], [104], [14], [45]. Therefore a restriction on papers mentioned as related work has to be made. Hence the following sections concentrate on works which emphasize that they are using 3D images or dimension reduction methods for feature extraction because the fly brain images are 3D and the proposed image description method is based on PCA.

3.3 Retrieval of 3D Images

Several data acquisition methods support 3D imaging, for example CT and MRI. Two approaches in CBIR exist to deal with 3D images. Either one dimension is removed which simplifies the task or feature descriptors designed for 2D images are adapted to the 3D space. In Section 3.4 works dealing with CBIR in medicine and using PCA as feature descriptor are discussed.

3.3.1 Medical CBIR for 3D images transformed to a 2D problem

The most common way to get rid of the third dimension is to build a 2D stack and compute the features for each slice or for a key slice. However, a problem of this approach is that interlaced data are missed. This are objects or structures, which are spread in the third dimension

One of the most popular medical image retrieval systems is ASSERT [81]. The image database holds high-resolution CT lung images. A physician has to select the most outstanding slice of each 3D image. This slice is called key slice and saved in an extra database for retrieval purposes. On each key slice the Pathology Bearing Regions (PBRs) have to be delineated by a physician and the lung is extracted automatically. Figure 3.3 shows an example query in ASSERT. The image on the top left is the query image and the four images on the bottom are the retrieved images which match the query image best. Next to the query image is an enlarged version of a retrieved image and on the right are the adjacent slices of it. This is helpful for the physician to see how the disease spreads over the slices. For each PBR a 26 dimensional feature vector is computed. 14 of these 26 features are perceptual features like thickness of the bronchi walls or average nodule size. The remaining 12 features are selected from 255 additional extracted features from each PBR including the gray scale mean and deviation inside the region, gray-level histogram values, and the texture measurements entropy, energy homogeneity, contrast, correlation, and cluster tendency. Shape attributes are described with the length of the axis and their orientation and with measures using Fourier descriptors and moments. Besides the edge distribution is computed. Some of the features to describe the relationship between the PBRs and the entire lung are the Mahalanobis distance from the centroid of a PBR to the nearest lung boundary point and the difference of gray-scale mean respectively deviation of a PBR and that of the lung. From the 255 features only 12 are selected with sequential forward search to reduce computational overload. For evaluation purposes only precision was taken into account. The number of retrieved images is set to four. A retrieved image is correct when it shows the

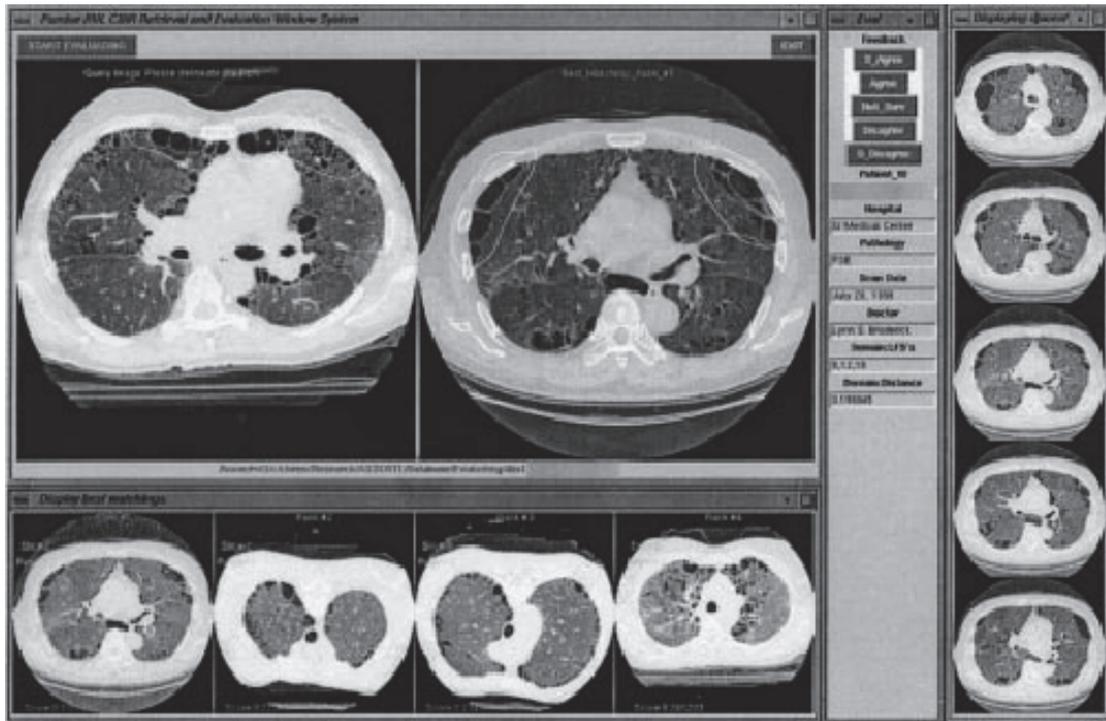


Figure 3.3: Screenshot from ASSERT – a computer aided diagnosis system [81].

same disease as the query image. ASSERT is one of few CBIR systems which were evaluated concerning their benefit in computer assisted diagnosis. The voluntary physicians could improve their answers from 29 % correctly given diagnoses to 62 %.

Song et al. [87] describe an approach based on 2D slices for 3D PET-CT thoracic images showing lung cancer. Initially the slices are preprocessed to extract roughly lung and mediastinum. For each slice and each modality 24 Gabor filters are applied. The slices are divided into 4x4 pixel patches and the Gabor responses for each patch are combined into one for PET respectively CT. From these combined patches a feature vector can be generated with six elements. It contains mean, minimum and maximum of both modalities. To preserve local spatial information the first-order local auto-correlation is calculated. This leads to a feature vector length of 294. The feature vectors of each patch are then used to generate a visual bag of words. Due to the effect that spatial information is lost with the bag of words approach a hierarchical partition model was designed. It consists of three levels and the half of the image containing most probably the pathological patch is divided into smaller blocks in each level. In each level for every block a bag of words histogram is generated. All the histogram results are concatenated and represent the feature vector. It is possible to describe the whole 3D image by extracting the feature vectors from the 2D slices and combine them using a weight according to the estimated distance to the pathological region. The method was tested with 40 PET-CT image sets with visible lung tumors. For each of them the ground truth was determined by labelling the other image

sets as similar or not similar. Two experiments were performed. In the first one each image set was used as query and average precision and recall were computed. For the other experiment five key slices from each image set were selected. With each key slice a query was performed and the average precision after 3, 5 and 7 retrieved images was computed.

The basis for Batty et al.'s work [5] are brain images recorded with PET. The images are registered to the Talairach atlas [94] and segmented into anatomical regions. In this case CBIR works only with a 2D slice as query image which has to be selected by the user besides the anatomical regions he is interested in. For each selected region the mean intensity and Gabor texture features are computed. The Gabor filter was applied with six different scales and four rotations. Mean magnitude and standard deviation of the filter responses build a feature vector with 48 elements. The similarity measure is based on the sum of squared differences. Not much is said about evaluation. The retrieved images are in the order they expected and the ones with small distances are diagnostically close to the query image.

The approach from Glatard et al. [32] uses also Gabor functions, but with six orientations and based on MR images. They are interested in the cardiac cycle which can be acquired with eight to ten 3D images. Each image shows the thorax with the heart near the center. One query possibility could be to retrieve images with the same slice number and therefore showing the same anatomical structures. For that purpose features are extracted globally. The heart occupies only a small part of the image and therefore the heart state does not have a great influence on the retrieval results. Feature vectors are compared using the Euclidean distance. To retrieve images from a specific state of the heart, the heart as Region Of Interest (ROI) has to be selected. A segmentation algorithm which uses the Gabor texture features divides the snippet into the three regions blood, myocardium and background with the help of the k-nearest neighbor algorithm. As the contraction of the myocardium gives information about the state of the cardiac cycle the texture features of the center from the k-nearest neighbor algorithm from this class is extracted. The magnitude responses are used to compute the coarseness of the myocardium and this measure can be taken for retrieval. The authors give only results for the first query case. They define the accuracy of their approach by comparing the slice number of the query image to the slice numbers of the eight best retrieved images and compute a mean weighted by the ranks of the retrieved images in the result set.

Li et al. [48] deal with traumatic brain injuries scanned with CT. Lesions are identifiable as pixels with high intensity. Each slice from a CT volume has to be preprocessed. The skull region is segmented with a simple threshold and from this result the intracranial region is extracted. Translation and rotation transformations are performed so that the region is in the center. Further steps are the normalization of intensities and the removal of cupping artefacts, which lead to pixels near the skull with high intensities. The last preprocessing step is to resize the image to a quadratic bounding box. Then a threshold segments the lesion pixels. To generate a binary feature vector a circle divided into N angular and N annular regions, resulting in N^2 equally sized bins, is used as mask. If more than half of the pixels in one bin correspond to lesion pixels, the bin is set to 1. The feature vectors of two slices can be compared with the Jaccard-Needham measure [110]. The similarity of two CT volumes is computed as the weighted sum of the similarity measure of corresponding slices. From 500 CT studies 30 were chosen as queries for evaluation. An expert scored the retrieved cases with points from zero to three, which is used as

ground truth for computing the normalized discounted cumulated gain [41].

MR brain images are used to analyze the different accuracies of intensity and texture features and which impact spatial information has [100]. Initially the images are registered, intensities are normalized and the brain tissue is segmented. From each volume four landmark slices showing the same anatomical structures are selected manually. As intensity feature a simple intensity histogram with 256 bins is used. Local Binary Pattern (LBP) [64] are chosen as texture feature and result also in a 256 bin histogram. Due to the fact that spatial information is lost with ordinary histograms a spatial histogram is generated with the help of a grid in the shape of a circle divided into several regions. As similarity measure for two slices the Euclidean distance of two histograms is used. To evaluate the methods they use each landmark slice as query image and the retrieval task is to find in every MR volume the corresponding landmark slice. Therefore the performance measure depends on the rank of the relevant slice. The sum of the difference of the slice numbers of the images, which are ranked better than the relevant slice, and the number of the relevant slice multiplied by the slice thickness gives the query error. The evaluation shows that the spatial texture histogram delivers the best results. In a follow-up work [101] they compare LBP, Kanade-Lucas-Tomasi (KLT) feature points [97] and a spatial method based on edge points identified by the Sobel operator [26]. In contrast to their previous work the only preprocessing steps are segmenting the brain tissue and finding the mid-sagittal plane for rotation, however the images do not get registered and no normalization of the intensities is applied. It is again shown that LBP with spatial information is superior to LBP without this information. The KLT feature tracker selects feature points representing corners in a slice. These feature points can be used to find matching points on another slice. The ratio of the number of matched feature points to the number of points found in the query image is the similarity measure. Due to the computational effort of feature point matching a grid can be used like the one for the spatial LBP method to capture just the spatial distribution of feature points. However, this method performs worse. Selecting only the dominant patterns when using the LBP histogram with spatial information delivers the best results. The retrieval task for evaluation is defined as in the previous work.

A paper that deals with 3D CBIR in the biomedical field was published by Berretti et al. [6]. They are interested in intracellular structures and deal with 3D cell images obtained from a confocal microscopy. The images are defined as an ordered stack of 2D images. For each of these images four statistical measures from the co-occurrence matrix are computed. To retain the information about the sequence of the feature vectors a Hidden Markov Model (HMM) is trained for each cellular structure one is interested in. It is now possible to compute for each HMM the likelihood that an image belongs to the class the HMM is representing. This results in a vector containing the likelihoods for every HMM. With this vector two tasks can be solved. The image can be classified by taking the class which belongs to the HMM with the highest likelihood. For image retrieval the reciprocal value of the sum of squared differences of the likelihood vectors give the similarity between two image stacks. The authors claim that the accuracy using HMMs is higher than with features from only one image per stack. The paper does not say if and how the cellular structures are segmented. Precision and recall for 40 query image stacks are computed, but it is not stated how a true positive image is defined.

3.3.2 Medical CBIR for 3D images adapting 2D feature descriptors

Less papers can be found that deal with CBIR of 3D medical images as a whole and do not only extract features from slices.

Liu et al. [49] developed their approach for neuroimages. These are images showing the brain and ideally help to understand how it works. Initially the images have to be registered based on a template and anatomical regions according to the Tzourio-Mazoyer atlas [99] have to be labelled. It represents 116 3D functional regions. Masks can be defined which contain several regions that are relevant for a disease. Different diseases infect different regions and therefore the search area can be constrained to the user's interests. To extract features from the image part that is covered by the mask the grey-level co-occurrence matrix for volumetric data [46] is computed. A voxel has 26 neighbours. Besides the distance two angles have to be defined. Only 13 directions are required to cover the data with no information loss. From the co-occurrence matrix eleven Haralick features [34] are computed and normalized to be comparable. For each of these feature spaces the top 5 % are eliminated because they are probably noise. The remaining values are divided by the maximum value of this space. The extracted image features are compared with the Euclidean distance. The approach was tested on PET images taken from patients with suspected dementia. For evaluation the leave-one-out strategy is used. As similarity measure serves a relevance score depending on if the retrieved image belongs to the same disease as the query image.

The distinction between malign and benign lung nodules is very important to detect lung cancer. The physician can be supported by finding similar images and their diagnoses. Kawata et al. [43] introduce an approach based on 3D CT thoracic images. These images do not have to be registered. As an initial step the lung area has to be extracted from the images. Then the ROI including the nodule one is curious about has to be selected interactively. As the last segmentation step the nodule in the ROI is automatically segmented by a deformable surface approach. The internal structure of a nodule is described by two joint histograms. In both histograms the distance to the nodule center is one dimension. The other dimension is either the CT density value or the scale index, which is computed from two principal curvatures. Two nodules can then be compared by computing a correlation coefficient based on their joint histograms. It is not described how the method gets evaluated.

Liu et al. [50] focus on the semantic image retrieval of CT brain data. For this purpose expert-labelled data are needed. Their goal is to extract features and find a subspace which suites best for classification by pathology. They give an overview of three different feature space reduction methods so that it is still possible to discriminate the data. Features which are redundant or irrelevant can be discarded. They explore memory-based learning, classification trees and discriminant analysis which gives the best trade-off between the dimension of the reduced feature space and the improvement of the classification rate. For retrieval evaluation 48 CT volumes showing a normal brain, stroke or blood cases are used. Due to the fact that registration of brains with lesions is challenging only the midsagittal plane is computed and the images are rotated accordingly to have the same alignment. As features multiscaled statistical properties like mean, standard deviation and gradient in two directions are extracted from different regions of the original image. They exploit the fact that healthy brains are symmetric for

additional feature extraction. The difference image which is computed between the two brain halves divided by the midsagittal plane, the thresholded difference image and the original image filtered with it are potential features. They authors use `Vizier` [79] for computing the best feature subset and a kernel regression classifier simultaneously. Depending if one is interested in a classifier that can discriminate between normal state and pathology or in a three-class classifier the reduced feature vector contains five to nine elements. How this feature vectors are compared is not stated. Although they developed their framework for 3D image retrieval, they evaluate it with 2D slices as query images. Depending on which classifier is chosen, a retrieved image is correct if it has the same label as the query image respectively for the two-class classifier no distinction between the pathologies is made. To show the accuracy of the method only precision is computed.

3.4 Retrieval through Dimension Reduction

The CBIR approach developed by Bucci et al. [13] is based on a set of numerical attributes extracted with PCA from each image. They evaluate their method with tomographic images. As these images show variabilities in anatomy, geometry and grey levels a computationally expensive preprocessing step is necessary to obtain satisfying results with PCA. To compensate these variabilities images are scaled by a factor that keeps the ROI constant, a registration method is applied and a histogram equalization is performed. The images are also filtered with a Gaussian filter because the brain is centered in the image and thus the regions one is interested in are emphasized. From the preprocessed images a representative set is selected to compute a lower dimensional space onto which the images can be projected. The lower dimensional feature vectors of the images can then be simply compared with the squared Euclidean distance. They generated software-simulated MR images from the brain and added Gaussian noise. With this image set they show that with an appropriate number of eigenvectors the method is robust against noise. As performance measure of their method they use the comparison of the slicing level of the query image to the mean level of the first seven retrieved images weighted by their similarity to the query. Using real MR data 13 eigenvectors are used to project the images. In 77 % the correct level was computed and in the other cases a difference of only one level was achieved.

The work from Bucci et al. [13] serves as basis for Sinha et al. [83]. They propose the same preprocessing steps to compensate image variabilities, but they compute the covariance matrix used for PCA differently. Instead of using the difference of the image intensities and the mean image intensities, the logarithm of the ratio is used. This leads to less sensitivity concerning contrast and intensity scaling and is superior to histogram equalization. This is especially important if the image database consists of images obtained with different acquisition methods, for example T1 and T2-weighted MR images, which leads to different contrast and intensity scaling. For evaluation purposes an expert assigned classes to all training images. Images showing the same anatomical structures refer to the same class. For every query image the expert selects the best matching image from the training set. A query is counted as accurate if this image is retrieved or one that is maximally three slices apart from it. Hence, for their evaluation only the first image which gets retrieved counts. They see their use case in automatically listing the structures visible

in an image. This can be done by assigning the query image the class of the best matching image and a mapping of class and structures exist. Another use case is to retrieve images showing specific anatomical structures by selecting a query image of an appropriate class.

Breast density is a risk index for developing breast cancer. Therefore MammoSys [23] was developed which should assist the physician when looking at a mammography. The texture of the breast is described with a variation of ordinary PCA called 2DPCA [107]. It is easier to compute because it can be applied directly to an image and does not expect the image to be transformed into a 1D vector. The ROI, which contains only breast tissue, is automatically segmented and from this region the features are extracted with 2DPCA. An experiment shows that using the first five principal components provides the best result. A Support Vector Machine (SVM) is used for classification and also retrieval. Because the density is divided into four classes the conventional SVM is not suitable and the one-against-one method was used. This means for every pair of classes an SVM is trained.

3.5 Relation to this work

Most of the previously mentioned methods simplify the retrieval task of 3D images by using only 2D slices. The method developed in the course of this thesis also works just on a 2D image. However, it is not simply one slice extracted of the image volume, but the result of a maximum intensity projection (see Section 4.1 for details).

Liu et al. [49] and Batty et al. [5] register their images based on an atlas and segment them into anatomical regions. Comparable to this work the features are computed for each of the regions separately and the user can decide which regions are of interest and should be taken into account for the retrieval.

The related works, which state which distance measure they use, mainly use Euclidean distance [32], [100], [49], [13]. It is one of the two distances applied in this work.

Many methods use statistical measures or Gabor filter responses as features. These features are not suitable for the fruit fly brain images because of the high variability of the brain structures and the many falsifying pixel which contain just noise. Therefore a feature selection based on Bucci et al.'s work [13] is used. With a set of representative images a lower dimensional space is computed with the principal component analysis (see Section 4.2.3 for details).

For the evaluation of their methods Shyu et al. [81] and Liu et al. [50] compute just the precision, which alone is not a sufficient measure. Berretti et al. [6] state precision and recall. An extensive evaluation is performed by Song et al [87]. For one experiment they compute the average precision and recall and for the other one they state *precision at k* for three different *ks*. To indicate the accuracy of the proposed method the precision-recall curve, the ROC-curve, the average precision and the F-measure are computed for each query. To get an overview of the whole test run the mean average precision can be stated as single-figure measure and the 11-point interpolated average precision curve can be plotted for an overview.

Methodology

In this section concepts are explained which were used in the practical work in the course of the thesis. How and during which processing steps they are applied is described in Chapter 5.

4.1 Maximum Intensity Projection

Maximum Intensity Projection (MIP) is a simple volume rendering method which is used since the late 1980s especially for medical images [102], [27]. It simplifies a 3D image in the sense that it gets projected onto a 2D image. For a 3D image three possible axis aligned projection directions exist: back to front, left to right and up to down. Projections in the opposite directions result in the same output. The size of the resulting 2D image is the same as when looking on the image from the chosen direction. For each output pixel a ray is casted through the 3D image in the projection direction and the highest value of a voxel along the ray is the resulting pixel value in the projection image at this position:

$$I = \max_{k=1..N}(v_k), \quad (4.1)$$

where I is the resulting intensity value, v_k is the value of the pixel at position k along the ray and N is the number of pixels sampled by the ray. Figure 4.1 demonstrates on an example how MIP works. The dimension of the original image is $4 \times 4 \times 3$. It can be seen that projecting the image along the z -direction the dimension of the resulting image is the same as the dimensions in x - and y -direction of the original image.

One disadvantage of removing a dimension is the loss of the information if an object in the resulting image is near or far away from the observer.

4.2 Dimension Reduction

Dimension reduction is a critical aspect when dealing with high dimensional data like images. A relative small image with a quadratic dimension of 768 pixels consists of already 589824

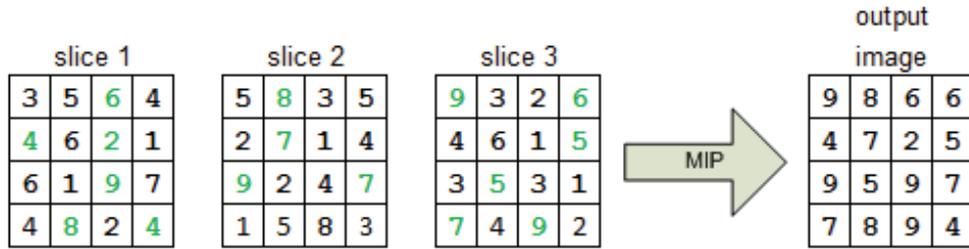


Figure 4.1: A 3D image with dimension $4 \times 4 \times 3$ sliced along the z-axis. The first three rasters represent the voxel values of the three slices and the most right raster is the projection image after MIP is applied in the z-direction.

pixels in total. Pixelwise comparison of images, especially bigger ones, is very costly. Also in some cases it is not practical to compare images pixel by pixel, but to extract the most important structures. Therefore the aim of dimension reduction is to reduce the amount of data and keep as much information as possible at the same time. PCA is explained in detail on the following pages as it is the chosen method for dimension reduction for this thesis.

4.2.1 Covariance and Covariance Matrix

Covariance is a statistic measure that provides information about the monotonic correlation between two random variables [71]. It is the expected value E of the product of two mean-centered random variables X and Y given by Equation 4.2. The covariance is positive if the values change together, that means that a high/low value of one variable goes along with a high/low value of the other variable. If the two variables have an opposite behavior (if a variable has a high/low value the other has a low/high value) the covariance is negative and if no monotonic correlation exists at all the covariance is zero and the two random variables are uncorrelated. The covariance informs only about the direction of the relationship between two variables, however it does not state anything about the strength.

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] \quad (4.2)$$

Considering n samples with k variables a covariance matrix \mathbf{S} with dimension $k \times k$ can be estimated. The sample covariance σ_{ij} between two variables i and j can be calculated as

$$\sigma_{ij} = \frac{1}{n-1} \sum_{l=1}^n (x_{li} - \mu_i)(x_{lj} - \mu_j), \quad (4.3)$$

where n is the number of samples, x_{li} respectively x_{lj} are the i^{th} respectively j^{th} element of the l^{th} sample vector and μ is the mean of a variable. If $i = j$ the covariance is identical to the variance. Computing the covariance between all variables leads to a symmetric matrix because multiplication is an associative operation. The matrix dimension corresponds to the number of

variables and its diagonal consists of the variances of the variables. Another characteristic of the covariance matrix is that it is positive semi-definite [88].

4.2.2 Eigenvectors and Eigenvalues

Let \mathbf{C} be an $k \times k$ matrix, then maximally k eigenvectors with their corresponding eigenvalues can be extracted from \mathbf{C} . An eigenvector e has to be a non-zero vector with k elements fulfilling Equation 4.4 for a scalar value λ which is the corresponding eigenvalue [31].

$$\mathbf{C}e = e\lambda \quad (4.4)$$

Eigenvectors can only be computed from square matrices. The direction of an eigenvector is unique. Eigenvectors with the same direction but a different length count only as one eigenvector as the length does not matter. Therefore it is very common to normalize the eigenvector length to one [71]. A very important characteristic of eigenvectors is that they are orthogonal to each other and uncorrelated.

On the other hand *eigenvalues* are estimations of how well eigenvectors can explain the data. They represent the variance of the transformed variables [31]. The value of such an eigenvalue provides information about how much the corresponding eigenvector contributes to the total variance. Eigenvalues are positive or zero if they are computed for a positive semi-definite matrix [71].

4.2.3 Principal Component Analysis

PCA is a very common method for dimensionality reduction. A well known application example is face recognition [98]. The gray values of an image are saved as a vector. It does not matter if the pixel values are saved row by row or column by column. It just has to be the same order for each image which is used to build a statistical model. The covariance matrix (4.2.1) of these image samples is calculated and an eigenvalue decomposition of the covariance matrix is performed. The fact that covariance matrices are always positive semi-definite leads to non negative eigenvalues [31]. However, if less samples than variables exist ($n < k$), only k definite eigenvectors and k non-zero eigenvalues exist.

The eigenvectors are the axes of the new coordinate system. The dimensions can be reduced by reducing the number of eigenvectors. As the value of the eigenvalues is proportional to the explained variance, only the eigenvectors with high corresponding eigenvalues are kept. The discarded eigenvectors with low eigenvalues hold only little information about the data. If the anisotropy is high, adjusting the coordinate system to the eigenvectors can compress the data a lot while keeping most of the information.

Transforming an image \mathbf{x} to the new coordinate system (often called eigenspace) can be done by mean-centering the image and multiplying it by a matrix \mathbf{E} of k eigenvectors (see Equation 4.5). The eigenvectors make up the transformation matrix because a multiplication by an orthogonal matrix has always the effect of rotating the axes [71]. This transformation preserves the Euclidean distances [31].

$$\mathbf{x}_{transformed} = \mathbf{E}[k]^T(\mathbf{x} - \boldsymbol{\mu}) \quad (4.5)$$

It often would be helpful to plot high dimensional data to get an overview, for example about how they build clusters. With PCA it is possible to reduce the data to the two or three most descriptive dimensions, which can then be easily plotted and analyzed. Figure 4.2 and Figure 4.3 show a simple example on a 2D data set and the affect of dimension reduction by using PCA. The former one shows 30 random data points which are mean-centered and have a high variance in one direction. As the data points have two dimensions, two eigenvectors with corresponding eigenvalues can be calculated. The first eigenvector with the highest eigenvalue points in the direction of the highest variance. It is plotted in red and scaled by the eigenvalue. The green arrow depicts the second eigenvector scaled with its corresponding eigenvalue. It is perpendicular to the first eigenvector. The green arrow is much shorter than the red one because of the smaller eigenvalue. The reason is that the data is not that much spread out in this direction and therefore contains less information. The aim of PCA is to reduce one-dimension and get a more compact data representation with the least information loss. For this purpose the data are transformed into a new coordinate system with the eigenvectors as axes as Figure 4.3 shows. The black lines show for two example points how they would be projected when using only one dimension and how much information would be lost. Reducing the dimension with the lowest eigenvalue depicted by the green arrow leads to a one dimensional data set lying on the first eigenvector which can be seen in Figure 4.4.

4.3 Distance between Feature Vectors

A measure is needed which defines how close respectively similar two feature vectors are. A feature vector can be seen as a point in a multidimensional space and its elements representing the coordinates. Several different distance measures exist. In this section the Euclidean distance and Mahalanobis distance are explained as they are used and compared in the practical work.

4.3.1 Euclidean Distance

The Euclidean distance, also called L2 distance, is a very common distance measure because it is very simple to compute [63]. It is based on the Pythagorean theorem as Figure 4.5 shows for the two-dimensional case. The two points \mathbf{p} and \mathbf{q} have the coordinates (p_1, p_2) respectively (q_1, q_2) . The orange line indicates the distance $d(\mathbf{p}, \mathbf{q})$ between these points and builds the hypotenuse of a triangle. The remaining two sides of this triangle are represented as dashed lines and have a length of $|p_1 - q_1|$ respectively $|p_2 - q_2|$. Matching the Pythagorean theorem $a^2 + b^2 = c^2$ to this triangle results in $(p_1 - q_1)^2 + (p_2 - q_2)^2 = d(\mathbf{p}, \mathbf{q})^2$. Taking the square root of this result leads to the Euclidean distance between the two points and correspond to the Equation 4.6 which gives the Euclidean distance for the general case, whereby $\mathbf{p} = (p_1, p_2, \dots, p_N)$ and $\mathbf{q} = (q_1, q_2, \dots, q_N)$ are two points in the N-dimensional space.

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(\mathbf{p} - \mathbf{q})^T (\mathbf{p} - \mathbf{q})} = \sqrt{\sum_{i=1}^N (p_i - q_i)^2} \quad (4.6)$$

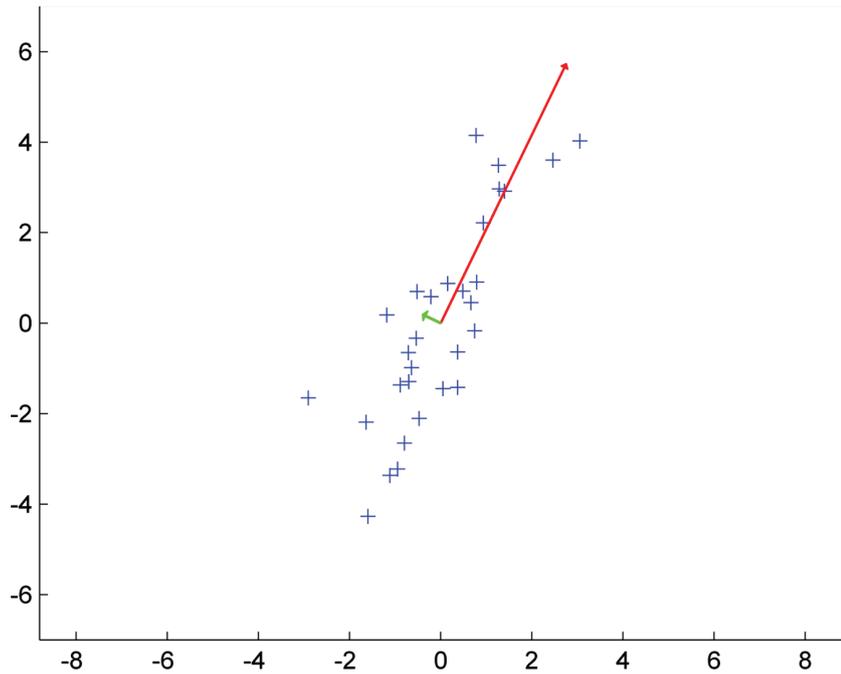


Figure 4.2: Random, mean-centered data points with a high variance in one direction. The red arrow depicts the first eigenvector and the green one the second eigenvector. Both vectors are scaled with their eigenvalues.

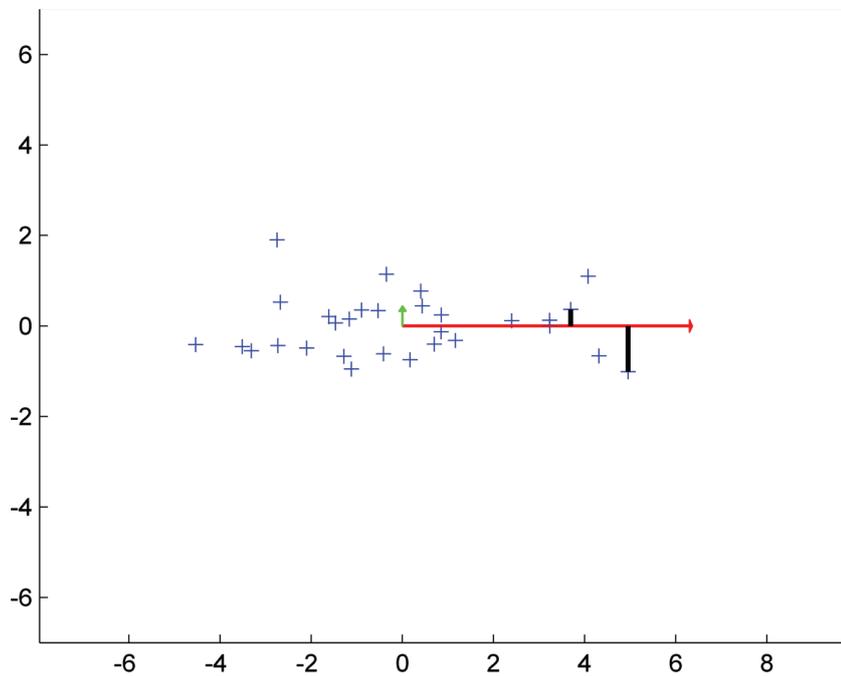


Figure 4.3: Data points from Figure 4.2 rotated by the eigenvector matrix to a new coordinate system. The black lines demonstrate how much information would get lost if the dimension with the lowest variance would be discarded.

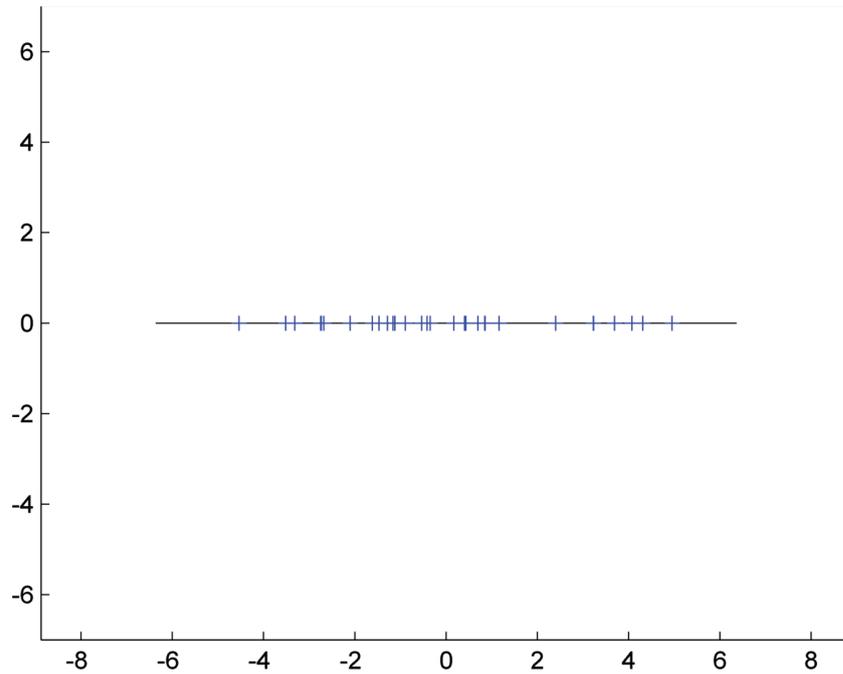


Figure 4.4: Data points from Figure 4.2 reduced to one dimension by PCA.

To summarize, the Euclidean distance is the square root of the sum of the squared differences between the corresponding coordinates of two points. In other words, it measures the length of a straight line between two points as one would do it with a ruler. Two characteristics that can be easily derived are the facts that the distance between a point and itself is always zero and that the distance between \mathbf{p} and \mathbf{q} is the same as between \mathbf{q} and \mathbf{p} .

Each coordinate of a point is treated equally when computing the Euclidean distance between points. In the two-dimensional case this leads to the fact that all points which have the same distance to one specific point lie on a circle. This specific point is then the center of the circle. This characteristic is not true, for example, for the Mahalanobis distance (see 4.3.2).

Data Normalization

The biggest disadvantage of the Euclidean distance is that it does not take variance into account. For example, given a two-dimensional vector with values between zero and one in the first dimension and between 200 and 1000 in the second dimension the maximum contribution to the Euclidean distance would be one for the first dimension and 640 000 for the second dimension. This example shows that computing the Euclidean distance for dimensions with different scales leads to higher influences of dimensions with a broader range.

Normalizing the data is the standard way to get rid off the different value ranges. A very

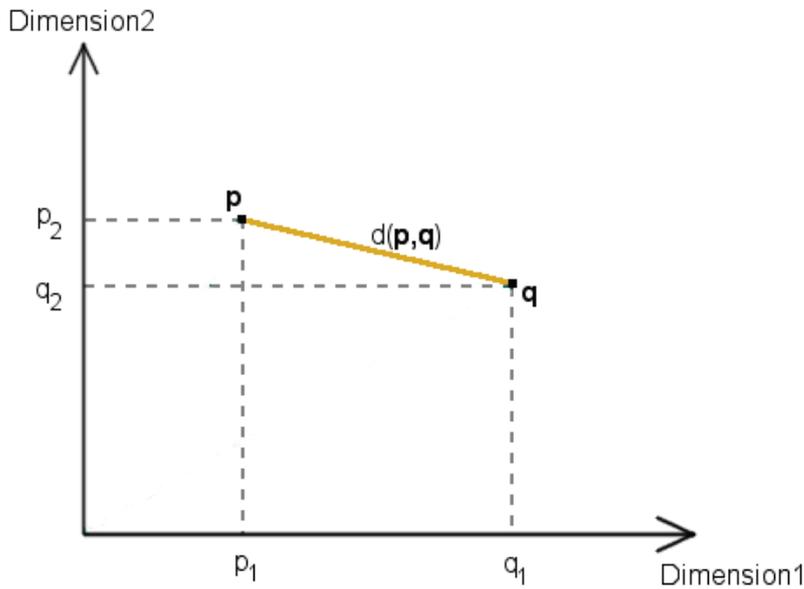


Figure 4.5: Drawing of two points \mathbf{p} and \mathbf{q} and the distance $d(\mathbf{p},\mathbf{q})$ between them in orange showing that the Euclidean distance is based on the Pythagorean theorem.

easy possibility to transform the values between zero and one is the min-max-normalization [39]

$$v'_k = \frac{v_k - \min_k}{\max_k - \min_k}, \quad (4.7)$$

where k ranges from one to the number of vector elements, v_k is the original value from vector \mathbf{v} at position k , \min_k is the lowest value at position k of all data and \max_k is the highest value at position k of all data and v'_k is then the normalized value. The values are scaled by a factor, but the original distribution remains. This normalization method is not robust against outliers. One outlier can be the reason that many transformed values are squeezed into a small range or even get assigned to almost the same value.

Another very common way to normalize data is the standard score, also called z-score, normalization [39]. The distribution of the data has to be known, otherwise the mean and standard deviation have to be estimated and then it is called Student's t-statistic, however it is computed analog to the z-score given by Equation 4.8. The mean μ gets subtracted from the original data value v_k and the result is then divided by the standard deviation σ . The normalized data have mean zero and standard deviation one. Transformed values are negative if the original value is smaller than the mean. In contrast to the min-max-normalization the normalized values are not in a specific range and the distribution of the data is not retained if the data are not normally

distributed. This method is also sensible to outliers.

$$v'_k = \frac{v_k - \mu}{\sigma}, \quad (4.8)$$

4.3.2 Mahalanobis Distance

An alternative measure to the Euclidean distance is the Mahalanobis distance which was invented 1936 [54]. It can be used to compute the distance between two points from the same distribution and therefore having the same covariance. In contrast to the Euclidean distance the Mahalanobis distance takes the variance of each coordinate as well as the covariance between the coordinates into account. It ensures that coordinates with a high variance and also highly correlated coordinates do not contribute as much as low variance coordinates and coordinates which are less correlated [71]. This is achieved by multiplying with the inverse covariance matrix:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(\mathbf{p} - \mathbf{q})^T \mathbf{S}^{-1} (\mathbf{p} - \mathbf{q})}, \quad (4.9)$$

where $\mathbf{p} = (p_1, p_2, \dots, p_N)$ and $\mathbf{q} = (q_1, q_2, \dots, q_N)$ are two points in the N-dimensional space, $d(\mathbf{p}, \mathbf{q})$ is the Mahalanobis distance and \mathbf{S} is the covariance matrix.

With the Euclidean distance points in 2D depict a circle when they have the same distance from a point. However, as Figure 4.6 shows, points with the same Mahalanobis distance represent an ellipse in the two-dimensional space. The data points are randomly selected from a normal distribution and the plotted ellipses represent the points with a distance of one, two respectively three standard deviations away from the center. The major axes of the ellipses point in the direction with the highest variance. As a logical consequence, in directions with small variances the distance between two points must be smaller to be similar than in directions with high variances.

The Mahalanobis distance corresponds to the Euclidean distance if and only if the covariance matrix is equal to the identity matrix (a matrix which elements are all zero, except the diagonal elements which are one).

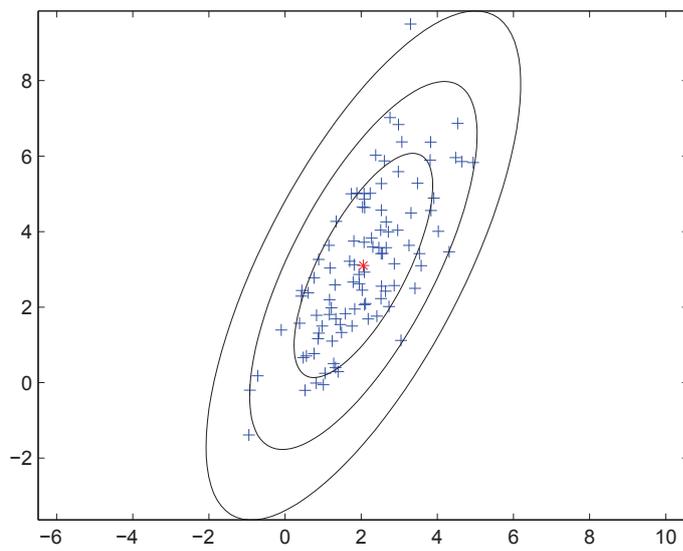


Figure 4.6: 100 random generated points from a normal distribution are plotted in blue and their mean in red. The three ellipses define the points which are one, two respectively three standard deviations away from the mean point when using the Mahalanobis distance.

Implementation

In this chapter the single steps of the proposed approach are explained. The left mushroom body is used as example for explanation and supportive images. Biologists are very interested in this neuropil and it is used for a lot of studies.

5.1 Overview of Approach

Figure 5.1 gives an overview of the processing steps, shows their outcomes and additionally provides information whether a step is performed for each image respectively each neuropil. Initially, for each neuropil of each image three MIPs are generated. This step is conclusive as biologists themselves do not look at images layer by layer, but use projected images if they browse through an image set. The gray values can be extracted from the MIPs and for each neuropil a PCA can be performed resulting in eigenvectors and eigenvalues. These outcomes can then be used to transform each image neuropilwise to feature vectors in a lower dimensional space. The content of two images can be compared by computing the differences between the feature vectors of each neuropil and sum up the distances. The smaller the distance is the more similar the images look like. The dissimilarity is computed between all images leading to a symmetric dissimilarity matrix, which can be used for retrieval.

5.2 Preprocessing

As the method has to deal with 3D images some preprocessing is necessary to simplify the problem. Initially, one dimension gets removed. This step is very common and related work from the medical domain can be found in Section 3.3. Generating MIPs (see Section 4.1) results in 2D images while losing information on one direction. The information loss should be as small as possible. This is one reason why it makes sense to divide the images into neuropil regions (see Figure 2.4). It has the advantage of being able to keep more information of the reduced dimension than performing the MIP method on the whole image, because the regions overlap.

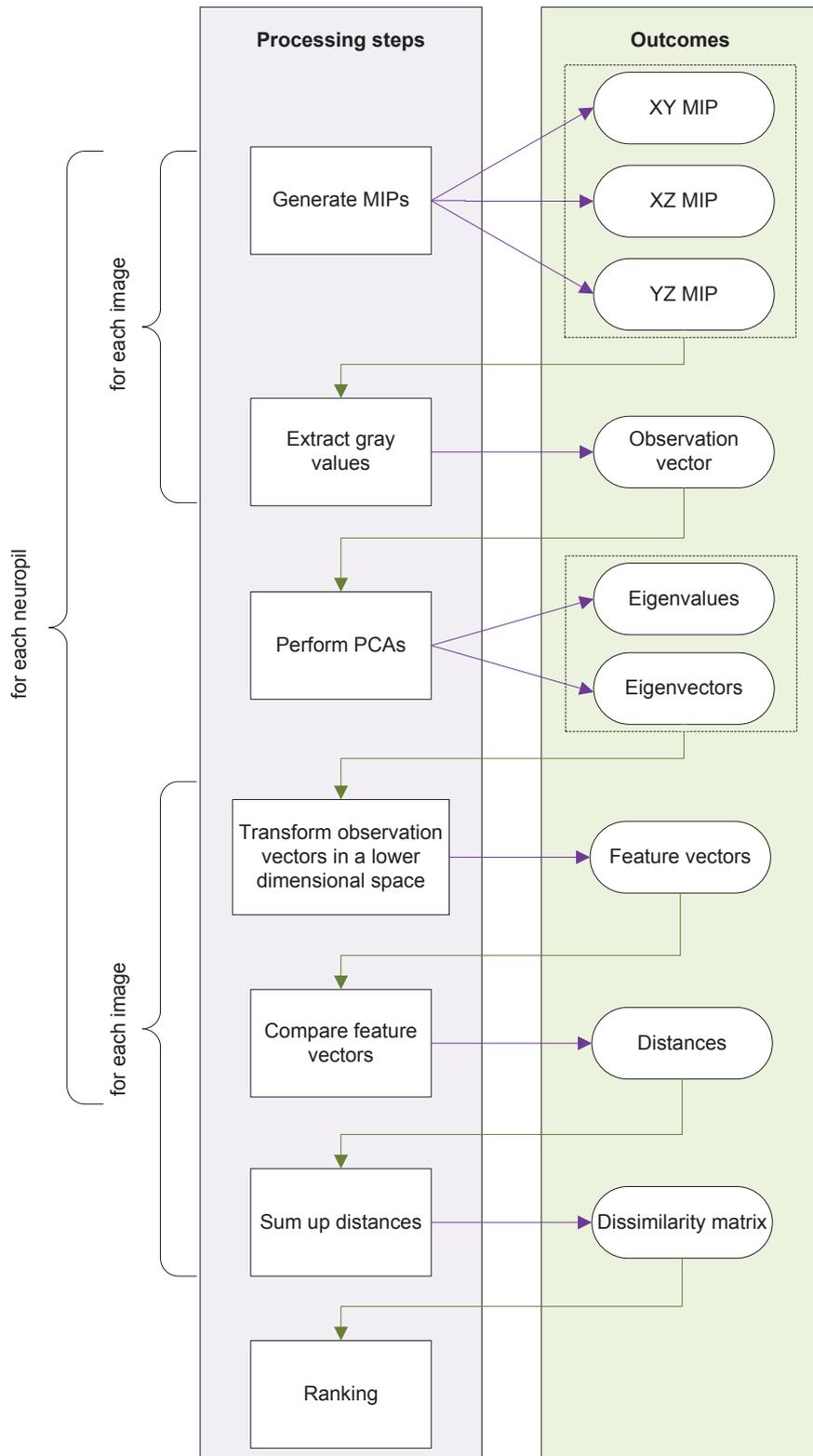


Figure 5.1: Workflow of proposed method.

The fact that the brain images are aligned to a template makes it easy to extract the neuropils. For each neuropil a 3D mask has to be provided. The part of an image which matches the neuropil can be extracted by setting all voxels to black except the ones where the mask value is set. For each image these extracted volumes, where only a part of the voxels is considered, are produced and three different MIPs are generated by reducing either the x-, y- or z-direction. The three created images correspond to the side view, the view from above and the front view. As mentioned in Section 2.2 the acquired images have a size of $768 \times 768 \times 165$ voxels. Reducing the z-direction leads to a 2D image with a dimension of 768×768 , reducing the y-direction to an image with 768×165 pixels and reducing the x-direction also to 768×165 pixels. Table 5.1 shows how a volume looks like when removing one of the three dimensions using the left mushroom body as example. The first image column presents the MIPs created from the neuropil volume mask. The last two columns show the MIPs for the left mushroom body for two example images. The images in Table 5.1 are plotted with dimensions being a fifth of the original size. Computing a distance measure based on the MIPs in the original size would be still too expensive. Therefore the MIPs are downsampled to a size of 64×64 respectively 64×14 pixels.

5.3 Dimension Reduction using PCA

Even with the downsampled MIPs approximately 20.000 pixels need to be compared. Therefore further dimension reduction is necessary. For this purpose PCA is suitable for several reasons. It is a very efficient dimension reduction method and it can find patterns in high dimensional space [3]. The only disadvantage of PCA is the amount of data needed for computing meaningful eigenvectors. On the other hand it is an unsupervised method and no annotations are necessary. PCA fits perfectly the needs of the retrieval task. It can be easily applied to the brain images because they are co-registered and therefore anatomical regions have the same size and position in each image. Eigenvectors and eigenvalues are computed for each neuropil based on the grey values of the generated MIPs. The pixel values are extracted from the positions where the pixels of the MIP masks are set. The grey values are concatenated to a vector. A neuropil is described by three MIPs. Therefore the grey values of these three MIPs have to be merged to one vector - the so called observation vector. A collection of such observation vectors from different brain images but from the same neuropil is the input for the PCA. It is important that the order of the grey values is the same. The order used in this thesis is the front view in the beginning, then the view from above and at last the side view. The pixels from the MIPs are read from left to right and from top to bottom. Supposing the observation vector has n elements, applying PCA delivers n eigenvectors with n elements and n corresponding eigenvalues.

Figure 5.2 shows nine representations of eigenvectors with the highest eigenvalues using the left mushroom body as example. They characterize the variation of this neuropil. The range of values for eigenvectors is not between 0 and 255 as it is for the brain images. Values can also be negative. The image contrast can be enhanced when using the whole grey value range. For each image the values are mapped to 256 grey values individually. This has to be borne in mind when comparing eigenvector representations in Figure 5.2 - same grey value in different images does not mean same value in this case.

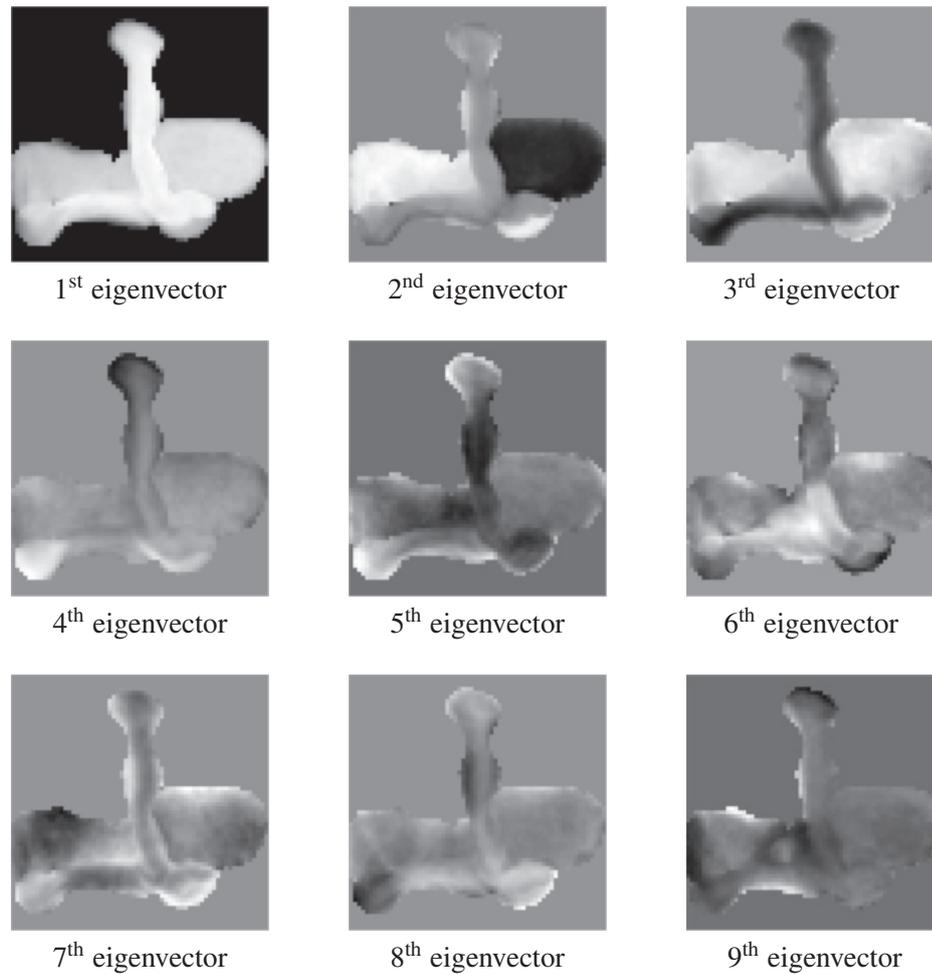


Figure 5.2: The first nine eigenvector representations of the left mushroom body.

The aim is to reduce the dimensions and at the same time keep most of the information. Most of the variance can be explained with the first few eigenvectors. Figure 5.3 shows the eigenvalues for the first 20 eigenvectors of the left mushroom body and how rapidly the eigenvalues drop. Already the fourth eigenvector can only explain 6 % of the variance the first eigenvector can explain.

The number of eigenvectors for the transformation into a lower dimensional space is determined by defining the percentage of variance they should explain. For each neuropil the eigenvalues have to be summed up and from this sum the defined percentage is calculated. With the descendingly ordered eigenvalues a cumulative sum is computed until the value calculated before is reached. The eigenvectors corresponding to the eigenvalues which were summed up are then used for the transformation. The number of these eigenvector-eigenvalue-pairs gives the number of elements of the feature vector. This means that it is not fixed how many elements the feature vector has after dimension reduction. It differs from neuropil to neuropil and depends on

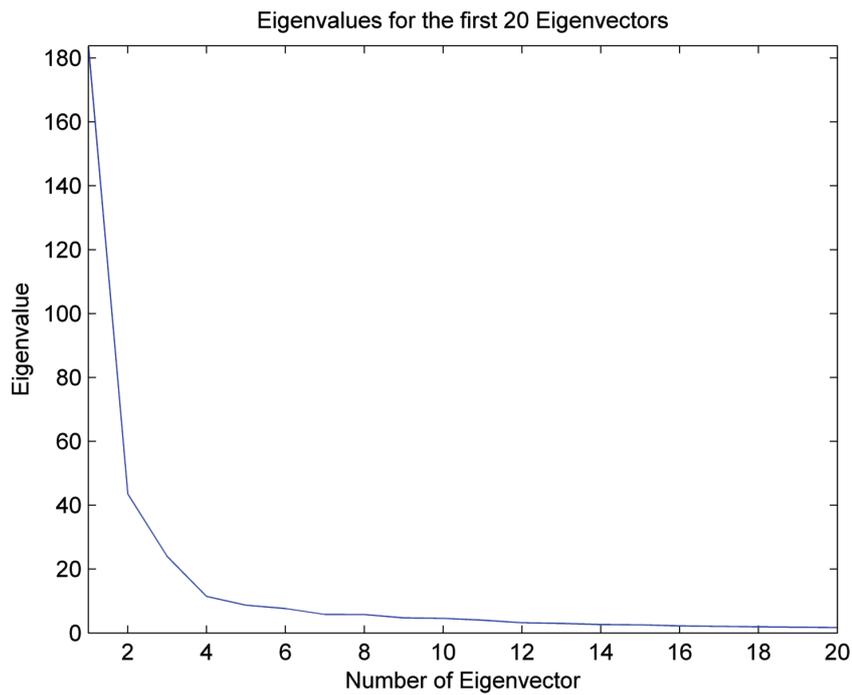


Figure 5.3: For the 20 first eigenvectors the corresponding eigenvalues are plotted for the left mushroom body.

the one hand on how long the observation vector is and on the other hand how well eigenvectors can describe the variance of a neuropil.

Table 5.2 shows for each neuropil how many elements the observations vector has and how many elements the feature vector has after dimension reduction and with 60 %, 70 %, 80 % and 90 % of the variance explained. The last row provides information about the number of elements that have to be compared for global retrieval. This is the length of the vector if the feature vectors from all neuropils would be concatenated.

Name	100 %	90 %	80 %	70 %	60 %
right optic tubercle	150	33	13	6	3
right middle superiorlateral protocerebrum	129	22	8	4	2
subesophageal ganglion	743	129	41	14	6
left optic tubercle	155	34	14	6	3
right medulla	310	25	6	3	2
left mushroom body a' lobe	27	7	3	2	1
right middle inferiorlateral protocerebrum	148	39	16	7	4
right lateral horn	166	39	15	6	3
right ventrolateral protocerebrum	465	90	33	14	6
right mushroom body a' lobe	30	9	4	2	1
brain cortical layer	4377	720	340	159	70

Name	100 %	90 %	80 %	70 %	60 %
left antennal mechanosensory and motor center	135	28	11	6	3
left ventrolateral protocerebrum	480	95	35	14	7
fan-shaped body	239	28	8	4	2
left posterior superiormedial protocerebrum	196	48	20	10	6
left unclassified brain neuropil structure	97	21	8	4	2
left lateral horn	160	37	14	6	3
left posterior inferiormedial protocerebrum	181	53	23	11	6
left middle superiormedial protocerebrum	255	49	18	8	5
left lobula	437	57	19	8	4
right antennal mechanosensory and motor center	130	27	12	5	3
right posterior inferiormedial protocerebrum	188	54	24	12	6
right lobula	529	57	18	8	4
left mushroom body y neuron innervation	26	6	3	1	1
right posterior superiormedial protocerebrum	193	48	20	10	5
brain internal cavity	1795	385	161	67	29
left posterior superiorlateral protocerebrum	140	29	12	5	3
right ventral body	427	89	34	15	8
protocerebral bridge	132	26	9	3	1
ellipsoid body	132	12	3	2	1
right mushroom body y neuron innervation	27	4	2	1	1
left middle inferiormedial protocerebrum	271	58	22	11	6
right medial posterior	519	144	63	30	15
right ventral	696	167	70	32	16
left medial posterior	547	149	65	30	15
right middle inferiormedial protocerebrum	274	59	23	11	6
right lobula plate	347	33	10	5	3
left ventral	716	157	62	27	13
left ventral body	446	101	42	18	9
right middle superiormedial protocerebrum	237	44	17	8	5
left posterior inferiorlateral protocerebrum	159	42	18	8	5
left medulla	200	28	9	3	2
left middle superiorlateral protocerebrum	139	26	10	4	2
right antennal lobe	273	32	10	4	1
right posterior inferiorlateral protocerebrum	137	38	16	8	4
right posterior superiorlateral protocerebrum	126	24	9	4	3
right unclassified brain neuropil structure	94	21	9	4	2
noduli	32	9	4	2	1
left lobula plate	212	24	8	4	2
right mushroom body	375	46	14	5	3
left mushroom body	392	49	15	6	3
left middle inferiorlateral protocerebrum	149	39	16	8	4
left antennal lobe	286	35	11	4	1

Name	100 %	90 %	80 %	70 %	60 %
numbers of elements to compare	19226	3625	1470	659	322

Table 5.2: For each of the 53 neuropils the length of the observation vector (= 100 %) and the length of the feature vectors with 60 %, 70 %, 80 % and 90 % of the variance explained is given.

5.4 Distance Measure

Depending on the interest of the biologist the similarity between images can be computed for the whole image or just for some neuropils which are of interest. Both use cases can be easily achieved by comparing the feature vectors of neuropils and sum up the vector distances. In this thesis two methods of computing vector distances are implemented and evaluated. The Euclidean distance is a very simple distance measure (see Section 4.3.1) summing up the quadratic differences of each vector element. The Mahalanobis distance (see Section 4.3.2) takes the variance of the different dimensions into account by multiplying the Euclidean distance by a covariance matrix. This covariance matrix can be reused from the PCA computation. With the distances between all images a dissimilarity matrix can be generated. It is a symmetric matrix with zeros in the diagonal. For a retrieval task just the row or column corresponding to the query images has to be extracted and ordered ascendingly.

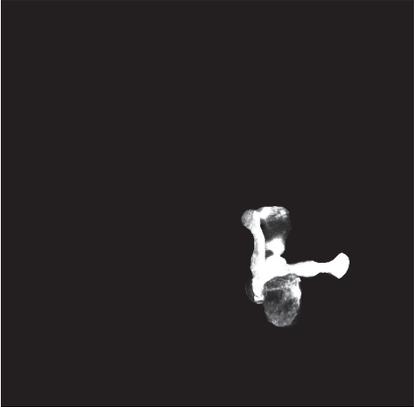
Removed dimension	Mask	Example image	Example image
z			
y			
x			

Table 5.1: The first column states the dimension which got reduced to obtain the MIPs. In the second column the three different views on the left mushroom body mask are shown. In the last two columns the MIPs from two different example images are depicted.

Evaluation and Results

At the beginning of retrieval research text documents were used. Later on also image retrieval was an issue. As the thesis deals with images, the term images will be used in this chapter, but can be substituted with the term document as the evaluation techniques are the same for both. In order to compare different retrieval systems it is necessary to evaluate their performance. It is possible to ask a user to rate the retrieved results. However, the satisfaction of a person is hard to measure and to compare and very subjective. An easier and more objective way is to compute some measures based on the retrieved objects and a ground truth. Three components are needed for an objective comparison between systems and have to be the same for each system for the performance evaluation:

1. Image collection

No constraints exist which images the collection can contain, but it makes sense that they are somehow related. For example, an image collection could consist of only medical images.

2. Query images

A subset of the image collection represents the query images. For each of these query images, images from the image collection are retrieved, which are relevant to the query image according to the tested retrieval system. These retrieval results are the basis for the evaluation.

3. Ground truth

For each query image it has to be defined which images of the image collection are relevant and should be retrieved and which are not. If this can not be done automatically with some meta-information, it has to be done manually. For big image collections this is an impossible task. However, it is advisable that more than one person does the annotation to reduce the subjectiveness.

According to the retrieved images and the ground truth several measures can be computed to indicate how well a retrieval system performs. Other factors, like speed and layout, can also be

important for a user. In the following section the most important and well-known measures for retrieval are presented based on the work of Manning et al. [56].

6.1 Evaluation Measures for Retrieval Tasks

For evaluation purposes every retrieved image has to be categorized according to Table 6.1. An image which is in the retrieval result and is relevant for the query is True Positive (TP). On the other hand relevant images not in the retrieved image set are False Negative (FN). All images which are not classified as relevant and are not retrieved are True Negatives (TN), but if they are retrieved they are False Positives (FP).

		retrieved	
		yes	no
relevant	yes	True Positive (TP)	False Negative (FN)
	no	False Positive (FP)	True Negative (TN)

Table 6.1: An overview of how images are usually classified in retrieval.

With this classification it is possible to define some performance measures. The two, which are most often used, are *precision* and *recall*. They are defined in Equation 6.1 and in Equation 6.2. The precision measure tries to answer the question about the relevance of the retrieved result. Recall provides information whether the system retrieved many of the relevant images. The value of these measures lie between 0 and 1 respectively 0 and 100 when dealing with percentages. The higher the value the better is the retrieved result.

$$\text{Precision (P)} = \frac{\text{number of relevant retrieved images}}{\text{number of retrieved images}} = \frac{TP}{TP + FP} \quad (6.1)$$

$$\text{Recall (R)} = \frac{\text{number of relevant retrieved images}}{\text{number of relevant images}} = \frac{TP}{TP + FN} \quad (6.2)$$

Retrieval can be seen as a classification task because images can be classified as relevant and not relevant. Therefore it is reasonable to compute the *accuracy* of a system with Equation 6.3. However, most of the time this is not a very useful measure for retrieval because many images are non relevant and are classified as such. Precision and recall are a better choice as they have a reference to the TPs and this is what users are interested in.

$$\text{Accuracy} = \frac{\text{number of correctly classified images}}{\text{number of all images}} = \frac{TP + TN}{TP + FP + TN + FN} \quad (6.3)$$

A meaningful evaluation of systems is only possible when both measures, recall and precision, are stated. A good system must have high values for both. It is easy to develop a retrieval system with either high precision or high recall. Retrieving all images leads to a recall of 1

because each relevant document gets retrieved, but leads usually to a low precision. On the other hand retrieving only the first few results can lead to a high precision because the possibility that many TP are in the beginning is high, but some relevant objects will be neglected. Therefore it is necessary to consider both values, whereby a weighting of these two measures can make sense depending on the retrieval task. *F-measure* can be used to quantify the trade-off between precision and recall. It is the weighted harmonic mean of precision and recall and the most common way to use it is to equally weight them like in Equation 6.4. Equal to precision and recall the value for F-measure is between 0 and 1 respectively 0 and 100 % and a high value is desirable.

$$\text{F-measure} = \frac{2PR}{P + R} \quad (6.4)$$

The measures introduced so far are for unranked retrieval results. They do not take into account if relevant images are retrieved in the beginning or at the end, but only if they are retrieved at all. However, for most applications it is important for the user, that relevant results are retrieved in the beginning. The following measures take this into account.

A *precision-recall curve* gives an overview of how precision and recall change with the size of the retrieved image set. After the first, second, third, k^{th} retrieved image the two measures are computed and the values are plotted and connected in a diagram. This results in a curve with a saw-tooth shape. For a good system the curve reaches to the upper right corner. Figure 6.1 shows an example of a precision-recall curve in blue.

A precision-recall curve can be easily smoothed by using the *interpolated precision*. This means that for a recall level r the maximum of the precision values of recall levels that are equal or greater than r is taken (see Equation 6.5, where $P(r)$ is the precision of the recall value r). The result of such an interpolated precision-recall curve can also be seen in Figure 6.1.

$$\text{interpolated precision for recall level } r = \max_{r' \geq r} P(r') \quad (6.5)$$

Receiver Operating Characteristics (ROC) curves are often used to evaluate classifiers and can therefore also be used in retrieval. The true positive rate, also called sensitivity (which corresponds to recall, therefore see Equation 6.2 for the formula), is plotted against the false positive rate (1-specificity) stated in Equation 6.6. The curve always starts at the bottom left and goes to the top right and is optimally very steep in the beginning. The curve is not very expressive if the image set contains many non-relevant images because the false positive rate is then always almost 0. An example of an ROC-curve can be seen in Figure 6.2.

$$\text{false positive rate} = 1 - \text{specificity} = \frac{FP}{FP + TN} \quad (6.6)$$

Müller et al. [61] suggest several alternative graphs for performance evaluation for retrieval systems.

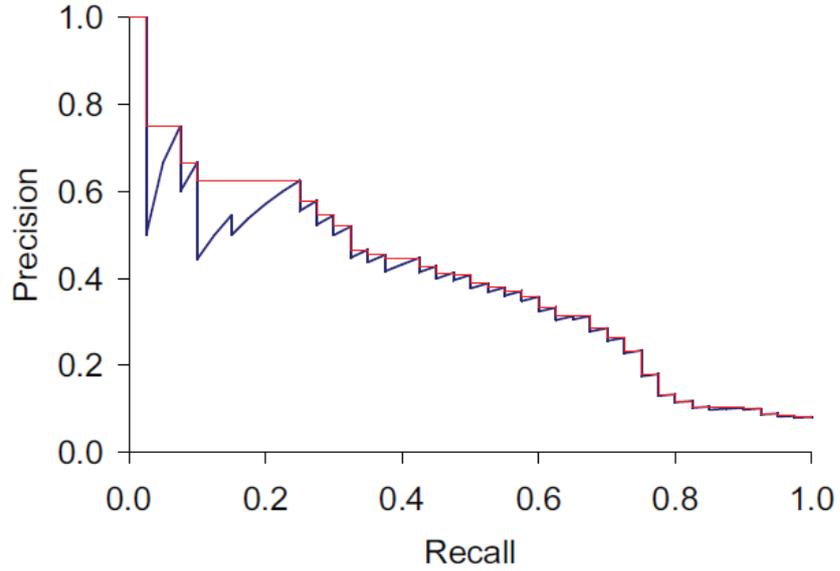


Figure 6.1: Example of a precision-recall curve in blue and the interpolation in red [56].

For a single query the *Average Precision (AP)* can be computed with

$$AP = \int_0^1 P(r) dr \approx \sum_{k=1}^N P(k) * \Delta R(k), \quad (6.7)$$

where $P(r)$ gives the precision for the recall r , N is the size of the image collection and $\Delta R(k)$ is the difference between the recall value after the $k-1^{th}$ and the k^{th} retrieved images. AP defines the average of the precision values after each retrieved relevant image. It approximates the area under the precision-recall curve as the integral in Equation 6.7 indicates.

For most systems it is not interesting how they perform for one single query, but for several queries. A very common single-figure measure for this purpose is the *Mean Average Precision (MAP)*. For Q queries it is defined as the mean over their APs.

$$MAP = \frac{1}{Q} \sum_{q=1}^Q AP(q) \quad (6.8)$$

A way to visualize the query results from a system in only one curve is the *11-point interpolated average precision curve*. For every query the interpolated precision (see Equation 6.5) for the eleven recall values 0, 0.1, 0.2,...,1 has to be calculated. The interpolated precision for the recall value 0 is always 1. If the query does not retrieve all relevant images and therefore no interpolated precision value exists for a specific recall value (and consequently for non of the higher recall levels) a precision of 0 is assumed. The interpolated precision values are then

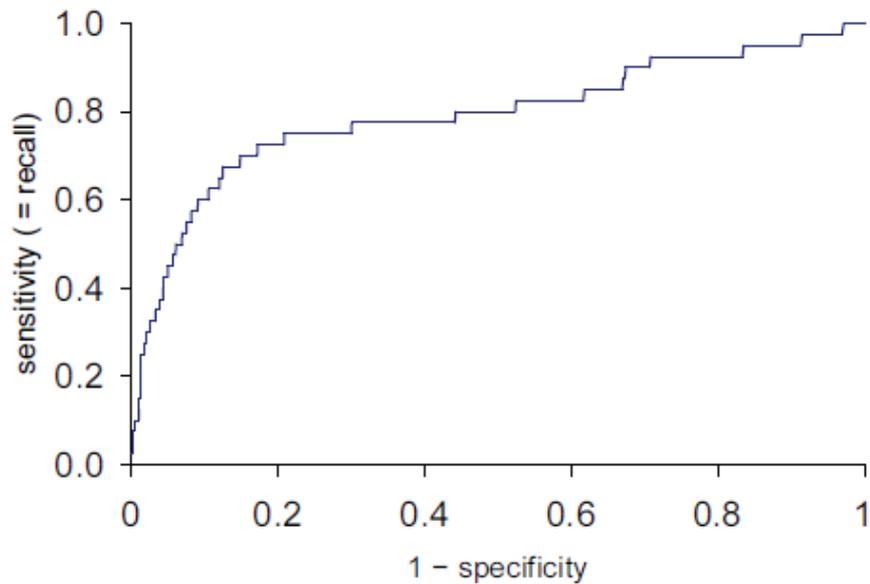


Figure 6.2: An ROC-curve corresponding to the precision-recall curve in 6.1 [56].

averaged over all queries. Figure 6.3 shows an example of an 11-point interpolated average precision curve depicting the eleven recall/interpolated precision-pairs. The area below and above the curve shows the standard deviation.

Precision at k is also a very common measure. The precision is computed after the first k retrieved images. Precision at 10 refers to the precision computed after 10 retrieved images. A disadvantage of this measure is its instability. It strongly depends on the total number of relevant images. Also a very good system can score badly if the precision at 20 is computed and only five relevant documents exist. The precision would then be only 0.25.

Another suggested measure for evaluation is the rank of the first relevant image. However, in our particular case this information is not very useful because the user is not satisfied with just one good result. The user is not looking for a specific target image, but wants to retrieve all images showing similar content to the query image.

6.2 Global Evaluation

The global evaluation provides information about the suitability of the proposed approach if the similarity of the whole image is of interest. The lack of ground truth data for the fly brain images is a heavy restriction for evaluation purposes. After defining the data set used for global evaluation, two different evaluation strategies are provided in this section.

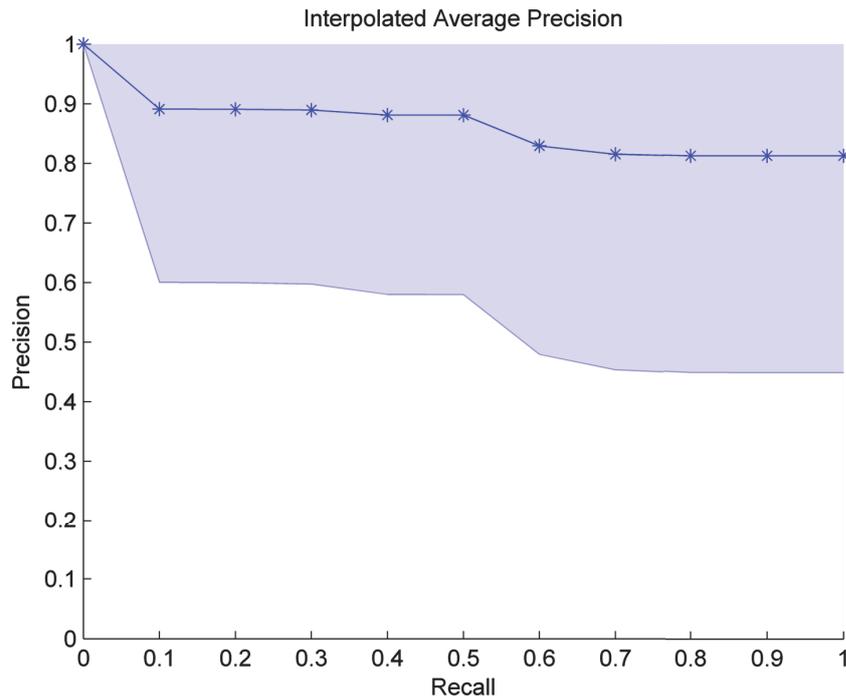


Figure 6.3: 11-point interpolated average precision curve of a test run with 3809 queries. The blue area depicts the standard deviation.

6.2.1 Image Dataset

For the evaluation flies with a database entry fulfilling the following requirements are appropriate:

- flies bred at IMP and therefore be part of the Vienna Tiles (VT) library
- available at the beginning of the thesis
- volumetric image acquired and stored
- images showing the brain (and not the VNC)
- sex has to be male (only a few females with bad quality are available)

In total 11169 images from 6063 different lines are used. Table 6.2 gives the distribution of how many different gene sets with a specific number of scanned flies exist.

6.2.2 Evaluation based on meta-data

It would be too time-consuming and expensive to generate a ground truth for the retrieval task, therefore the information that images from the same line should have a similar appearance is

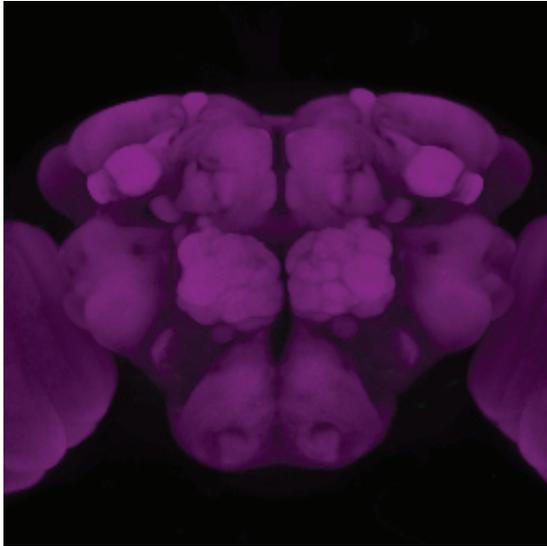
x	Number of lines containing x images
1	2310
2	2652
3	910
4	153
5	26
6	5
7	4
8	2
9	1

Table 6.2: The integers x in the first column represent the number of images belonging to one line. In the second column the number of lines consisting of x images is given.

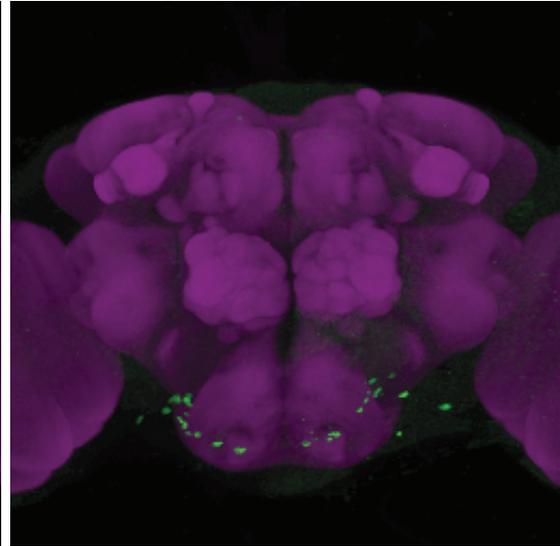
exploited. An image is only chosen as query image if at least one other image belongs to the same line. Hence, at least one true positive exists. Another criterion for an image to become a query image is the amount of staining. A lot of images have no visible expression, contain only glia cells (do not participate in synaptic interaction and have no axons and dendrites) or are overstained (the image contains so much noise that no patterns can be recognized). A biologist would never search for such images as they hold no information about expressed neurons. An example for each of these three cases as preview image can be seen in Figure 6.4. Such images have to be excluded from the query image set. This is done by simple empirical thresholding. For each image all pixels within the template are extracted and the values of the green channel are summed up if they are higher than 120. This value was chosen because most pixel which show only noise have a lower value. If the image has a bad quality it contains a lot of noise and therefore it is very important to exclude these pixel, otherwise they falsify the result. A sum between 303600 and 2160170 indicates an image suited as query image. This range was picked by visual exploration. After filtering 3809 images remain as query images.

For computing the eigenvectors and eigenvalues of PCA only a subset of the available images is used. Basically it is the set of query images with the restriction that only one image per line is selected. The reason is to avoid biased results because otherwise lines containing a lot of images would have a higher impact. Therefore only 2724 images are used for PCA.

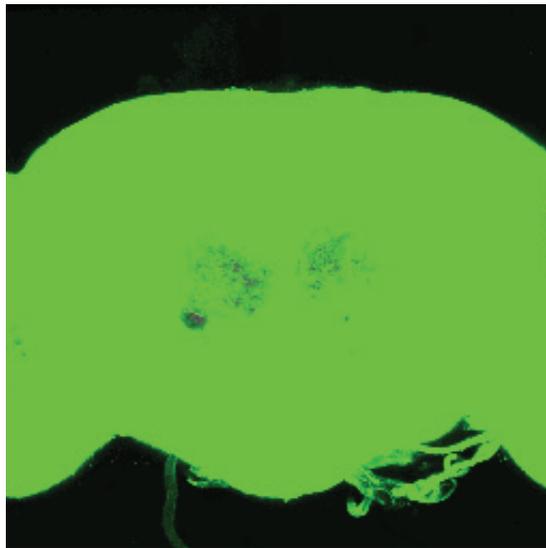
After determining the eigenvectors and eigenvalues the images can be reduced and for each query image a retrieval can be performed. The query image itself is excluded from the retrieved image set. It always has a distance of 0 to the query image and is retrieved as the first image. Additionally the biologists are not interested in retrieving the query image. For each query a precision-recall curve, an ROC-curve, the average precision and the F-measure are computed to evaluate the results. An HTML page with the query images, the corresponding retrieved images (both as 2D previews) and the computed measurements is generated in order to have also the possibility to visually evaluate the results. As the image set consists of more than 11000 images, it would be unreasonable to list them all in the retrieval result because only the most similar images are of relevance and nobody looks at all of them. Therefore only the 30 most



(a) Fly brain with no staining



(b) Fly brain with no staining except glia cells and noise



(c) Overstained fly brain

Figure 6.4: Example images which are not suitable as query images.

similar images are displayed for each query on the HTML page. Figure 6.5 and Figure 6.6 show two examples. Due to the lack of space only the first four retrieved images are displayed here. The query images can be identified by the turquoise background. Below the picture the name and information about how many images belong to the same line, and therefore how many true positives exist, are stated. The retrieved images are listed below the query image. Again some information can be found below each retrieved image. In particular, this is the name, the distance to the query image computed by the approach and the four most similar neuropils compared between the query image and the retrieved image. How much information content the smallest neuropil distances have is discussed in Section 7.2. If a retrieved image is a true positive one the information is highlighted in orange. For the query in Figure 6.5 two out of three possible images from the same line are retrieved in the beginning. However, also the next two images look very similar to the query image. They count as false positives, nevertheless for biologists this could be very informative. The second example query shown in Figure 6.6 has one true positive in the beginning. The following images do not match the query image perfectly, however each of them is partly similar.

On the right next to the query image the average precision and F-measure values for the query are written. Additionally, on the very right two graphs visualize the retrieval result for the 30 images displayed on the HTML page. If not all true positives are retrieved within the first 30 images the graphs are extended by as many images as necessary to achieve a recall of 1. The precision-recall curve shows that a recall of 1 is achieved with a precision of around 0.6 for the query shown in Figure 6.5. It can be seen that the precision is 1 for the first two images and drops for the next two images to 0.5. The fifth point of the curve signals that the fifth retrieved image is a true positive again (which can not be seen in Figure 6.5 as the results are cut away) and that it is the last one of the line that was missing. The recall value is now constant at 1 and precision drops significantly because no more true positives can be retrieved. The precision-recall curve for the query shown in Figure 6.6 is completely different. Precision for the first retrieved image is 1 as it is a true positive and recall is 0.5 because the line consists of three images. Whereby one of them is the query image, one is retrieved and one is missing. The precision value drops rapidly in the beginning and the second true positive image is retrieved as 30th image. Precision is only 0.05 then. The ROC curve is not very meaningful for a retrieval task with a lot of non-relevant images as it is the case here. The problem is that most of the time the false positive rate is always around zero because the number of true negatives is part of the divisor. Only if true positive images are retrieved very late the graph differs.

The mentioned measurements and curves are computed for every single query. To get an overview of the whole test run an 11-point interpolated average precision curve, like the one in Figure 6.3, is created.

6.2.3 Parameter Space Exploration

The evaluation based on meta-data is done for different parameters and for combinations of these parameters. Table 6.3 gives an overview of parameters for which it makes sense to try out different values. In total 48 different combinations of parameter values exist.

Changing the variance that should be explained affects on the one hand the length of the feature vector and therefore the time it takes to compare two images. On the other hand it has

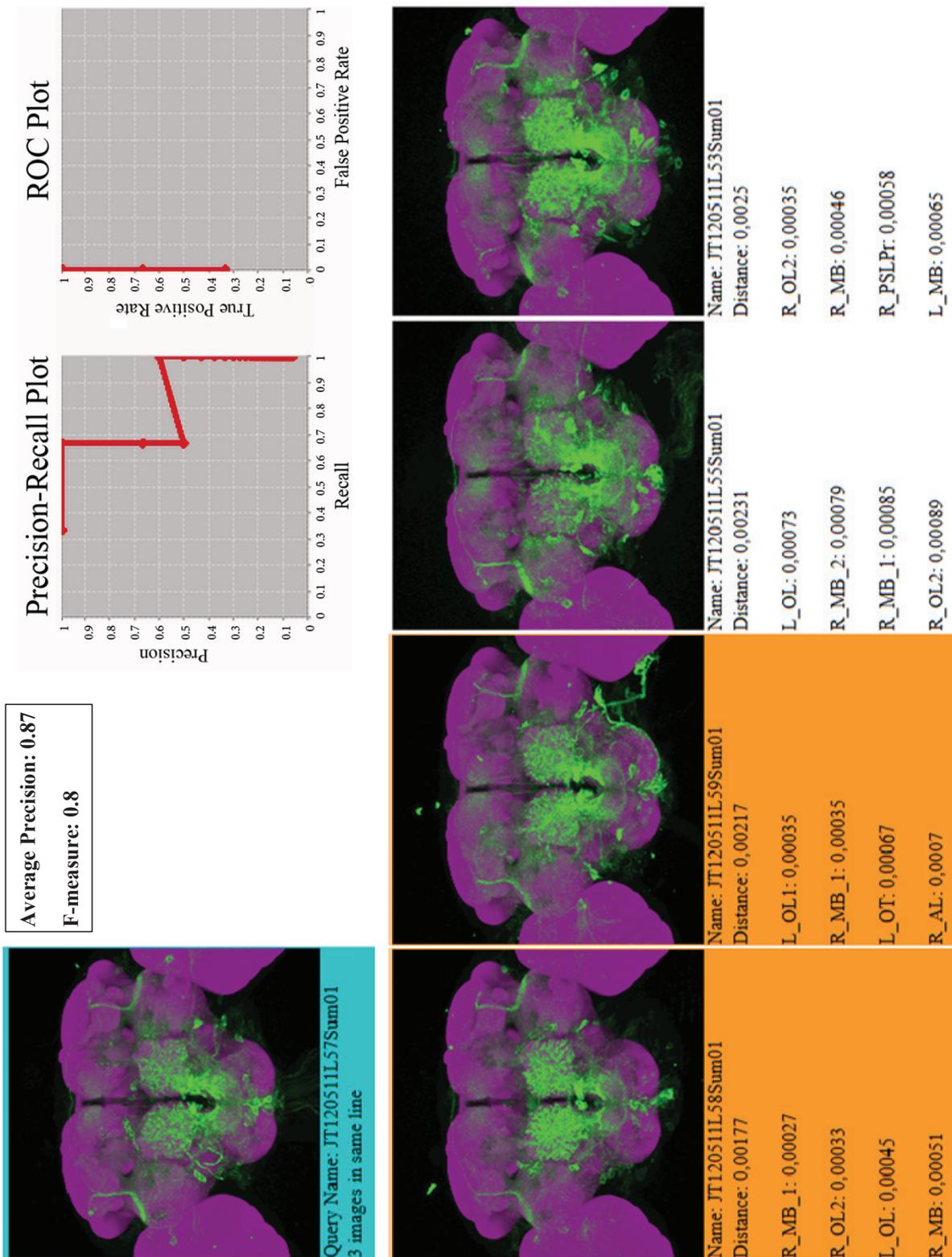


Figure 6.5: An example query with query image in turquoise and the true positives in orange. Additionally two measurements concerning the result are given and two graphs are plotted.

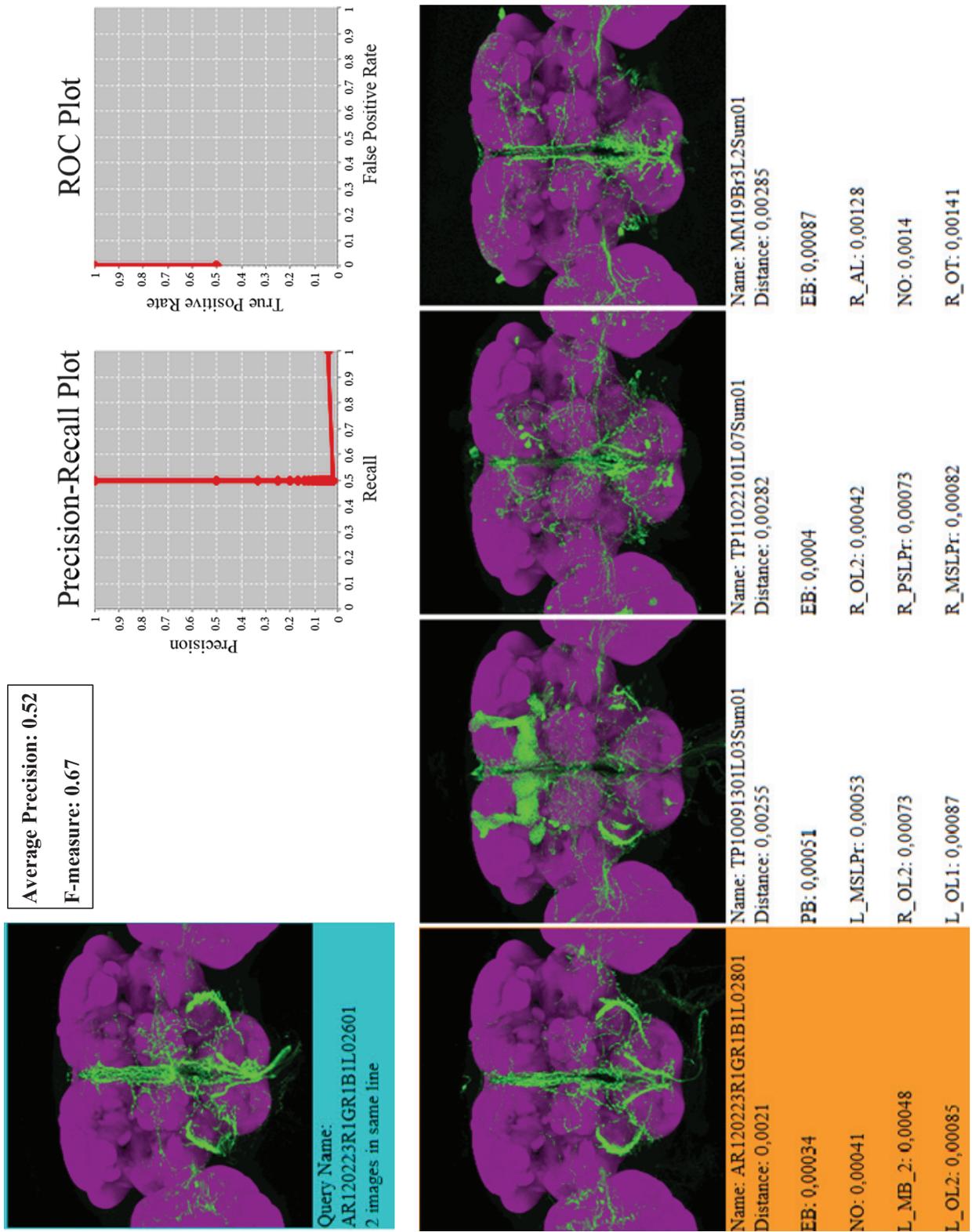


Figure 6.6: Another example query with query image in turquoise and the true positives in orange. Additionally two measurements concerning the result are given and two graphs are plotted.

Parameters	Possible values
Variance explained by PCA in %	60-85 in steps of 5 (both values inclusive)
Distance measure	Euclidean or Mahalanobis distance
Ignore first eigenvector	yes, no
Take neuropil size into account	yes, no

Table 6.3: Parameters to explore and their possible values.



(a) Mean of left mushroom body



(b) First eigenvector of left mushroom body

Figure 6.7: Comparison of mean and first eigenvector on the example of the left mushroom body.

consequences on the staining patterns that can be described. Short feature vectors can be too unspecific, however long ones can be too specific and inflexible.

Several possibilities to compare points in a high dimensional space exist. In the course of this thesis Euclidean and Mahalanobis distances are compared. As explained in Section 4.3.1 the Euclidean distance does not take the different dimension variances into account and therefore the data need to be normalized to get meaningful results. Before the distance between two feature vectors is computed, the elements of the feature vector get normalized by the min-max normalization as it is explained in Section 4.3.1. The minimal and maximal values which are possible are computed separately for each neuropil. Multiplying all positive values of an eigenvector by 255 and all negative ones by 0 and sum this up results in the highest possible value that can theoretically be achieved when transforming an observation vector. This is done for each eigenvector of a neuropil and results in the feature vector with the highest possible values. Getting the minimal values follows the same scheme, but negative values are multiplied by 255 and positives by 0.

At first glance it does not make sense to ignore the first eigenvector, but Yambor et al. [106] state that it could be that the first eigenvector does not contain relevant information. In contrast to the other eigenvectors the first one does not show any structure and is very homogeneous (see Figure 5.2). Figure 6.7 shows that it looks very similar to the mean image.

The last parameter that is tested is the influence of the neuropil size. The relative number of voxels that a neuropil has compared to the whole fly brain is taken into account. The computed distance between two feature vectors is multiplied by this value. This means that bigger neu-

ropils have more influence. This scaling is explainable by the fact that larger neuropils contain significant structures and are therefore more relevant.

The parameters are explored based on the image set and the approach described in Section 6.2.2. Figure 6.8 shows the evaluation results when exploring the different values for the explained variance and the distance measures. If additionally the first eigenvector is ignored the results can be seen in Figure 6.9. If the distances are weighted according to the neuropil size the results are given in Figure 6.10. Finally, Figure 6.11 shows the results for the case that the last two mentioned options are combined. On the x-axis the variance explained by the PCA in percent and on the y-axis the MAP is depicted. The red asterisks depict the results computed with the Euclidean distance and the blue ones with the Mahalanobis distance. It can be seen at a glance that the results computed with the Euclidean distance are always stable but the more variance is explained the worse are the results if the Mahalanobis distance is used. In the two cases, which take the size of the neuropils into account, the results for the Euclidean measure is better than in the other two, however the opposite is true for the Mahalanobis distance. The best MAP that is obtained is 0.85636 (see Figure 6.10). It can be achieved if explaining 70 % of the variance, weight the distances according to the neuropil sizes and use Euclidean as distance measure. The F-measure values are not plotted in the figures because then they would get too cluttered. However, the maximum is 0.8732 and is also achieved with the before mentioned parameter combination.

The k-fold cross-validation [35] is one method to show that it was not a coincidence that this parameter combination delivers good results. The image collection is divided into k blocks. The images are randomly assigned to a block resulting in equally sized blocks. k-1 blocks are used for training and the images from the block, which was left out, are used for validation. The blocks are cycled k times so that each block is used for validation once. In the case of the thesis the training images correspond to the images used for PCA and the images from the validation block are the query images. This approach is illustrated in Figure 6.12 for k=5, which is aside from k=10 the recommended value [35]. For the thesis k=10 is chosen and not the images themselves are divided but the lines. Therefore each block consists of about 600 lines, but has a different number of images as the number of images per line varies. To get unbiased results only one image per line is used to compute the PCA and also for validation.

Figure 6.13 shows the MAP and F-measure for the ten-fold cross validation with the best parameter combination. It can be seen that the results are stable and the two measurements correlate. The mean for MAP is 0.69 and for F-measure 0.72. The corresponding standard deviations are 0.025 and 0.024. The results are a little bit worse than the ones achieved while exploring the parameter space. This has two reasons. First, the whole image collection is used for the k-fold cross validation and therefore also images which are not suitable are used as query images (see Figure 6.4 for such examples). And second, dividing the lines into blocks and leave some out for validation could lead to the case that not all patterns are captured by PCA if a specific pattern is only visible on images in the validation block.

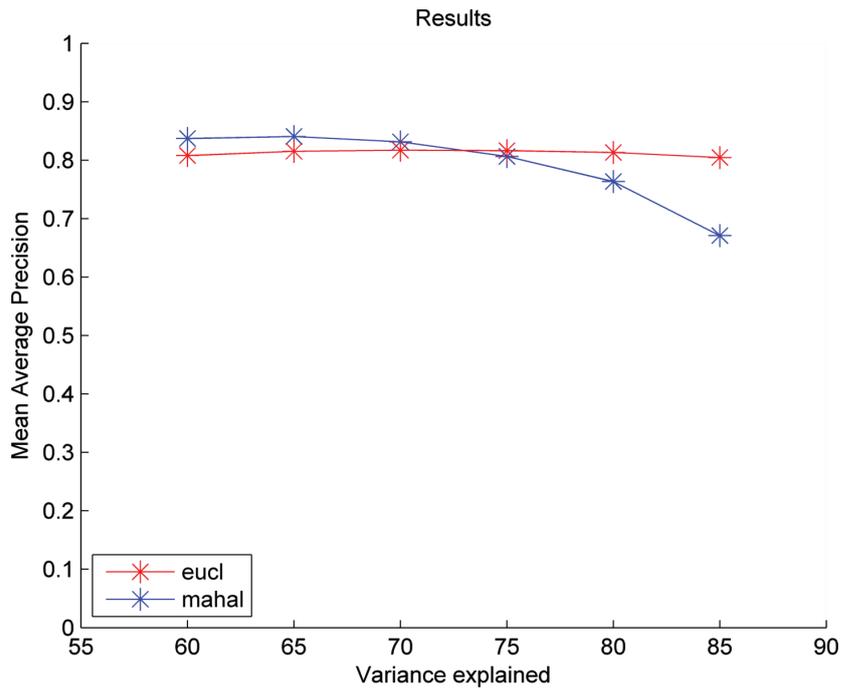


Figure 6.8: MAP for the different variance values explained by PCA using Euclidean (red) respectively Mahalanobis (blue) distance.

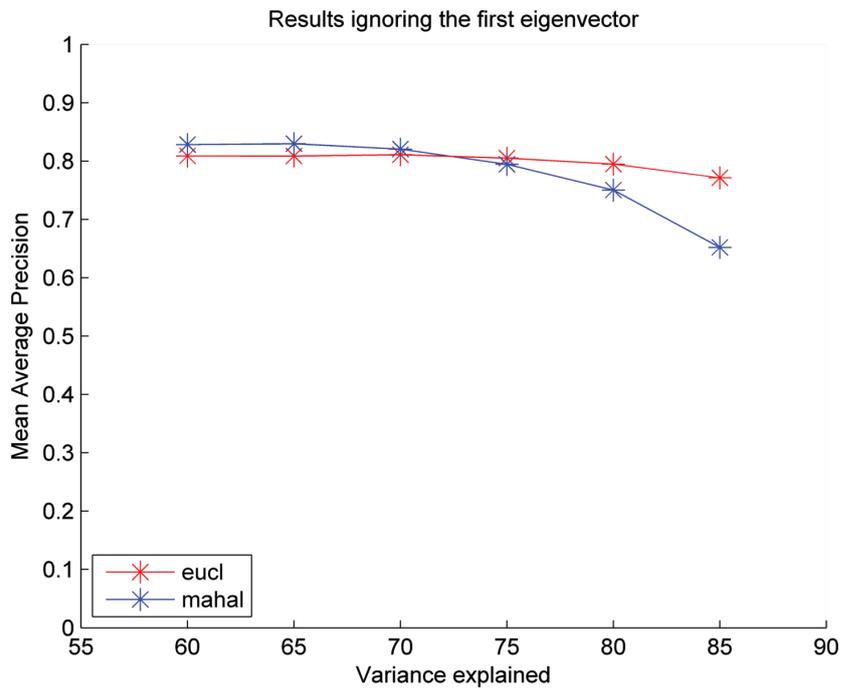


Figure 6.9: MAP for the different variance values explained by PCA when ignoring the first eigenvector using Euclidean (red) respectively Mahalanobis (blue) distance.

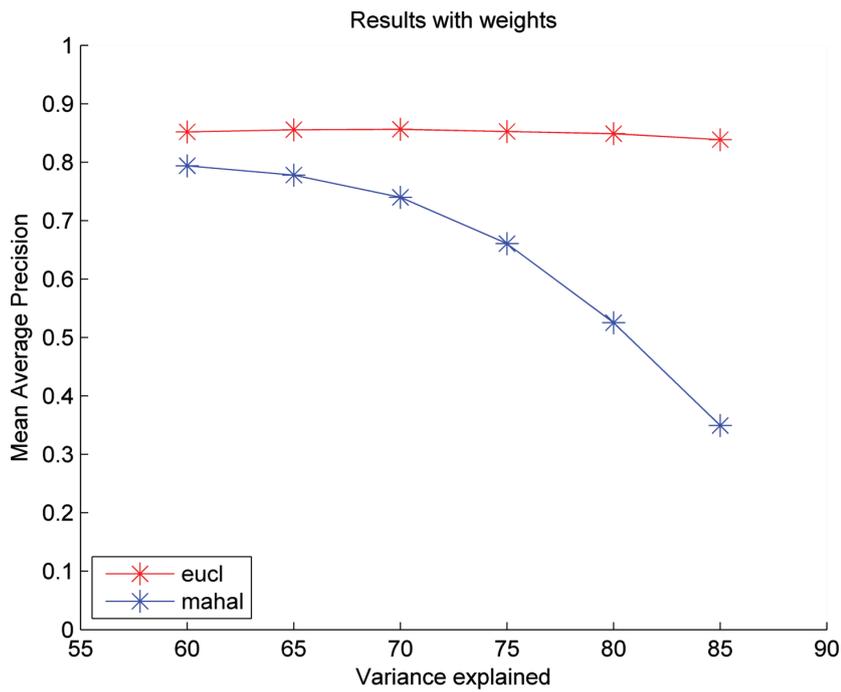


Figure 6.10: MAP for the different variance values explained by PCA when taking the neuropil size into account using Euclidean (red) respectively Mahalanobis (blue) distance.

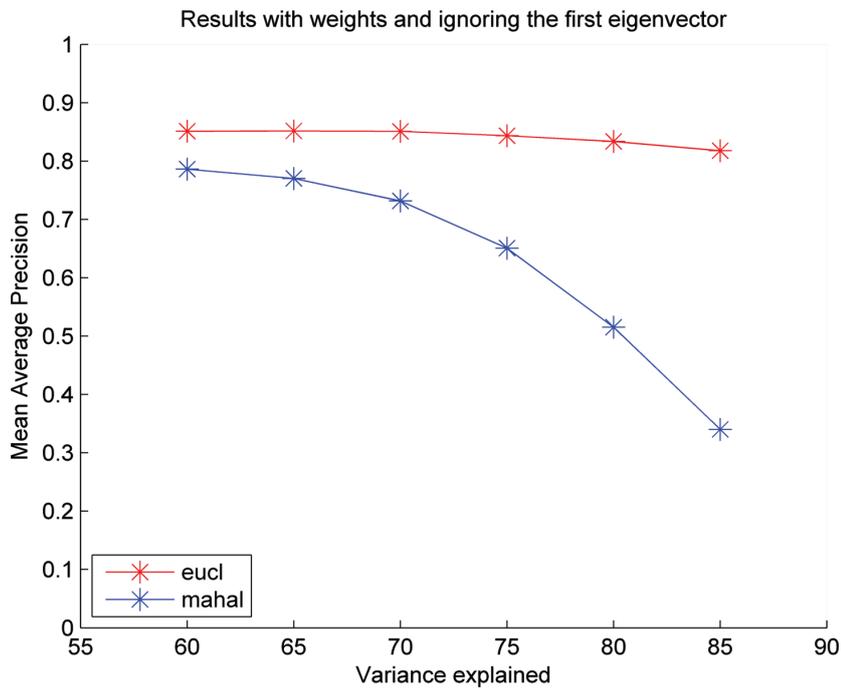


Figure 6.11: MAP for the different variance values explained by PCA when ignoring the first eigenvector and weight the distances according to neuropil size using Euclidean (red) respectively Mahalanobis (blue) distance.

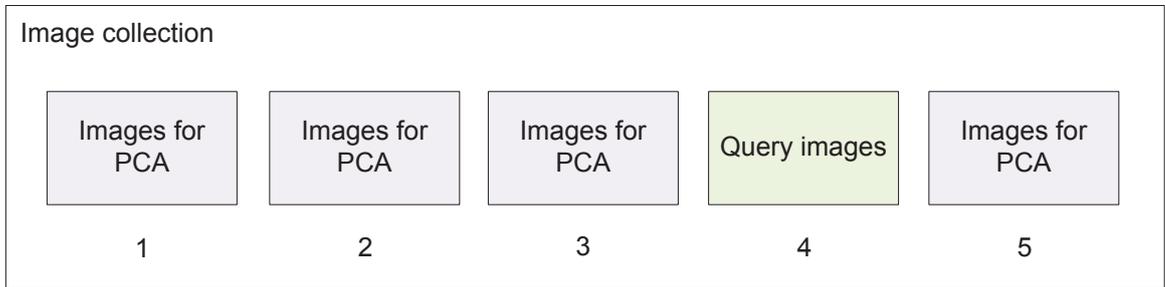


Figure 6.12: k-fold cross-validation with k=5 on the example of image retrieval using PCA.

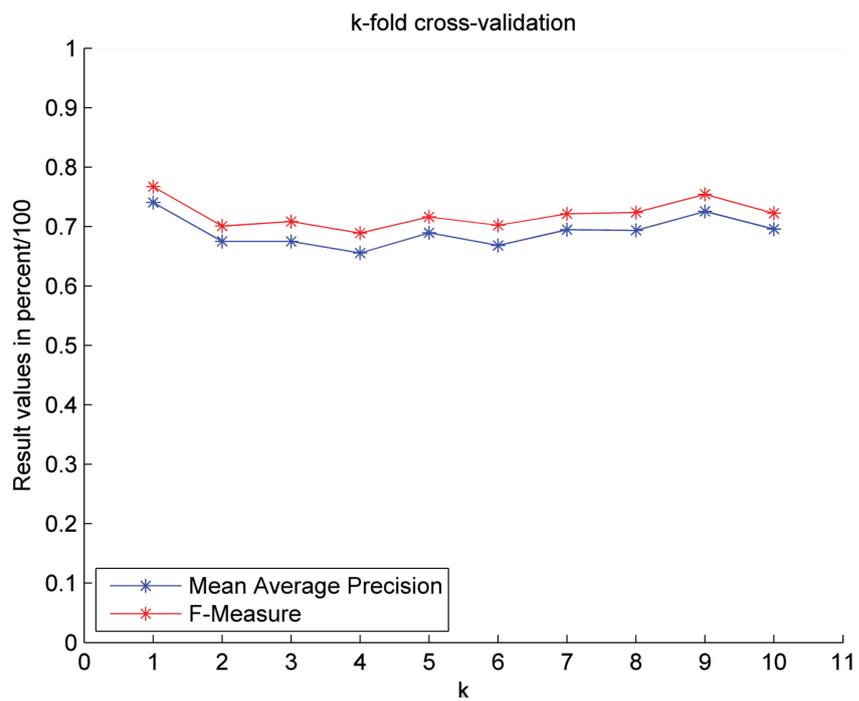


Figure 6.13: MAP (blue) and F-measure (red) for the ten cross-validation runs.

6.2.4 Evaluation by an expert

The evaluation by a domain expert is independent from a ground truth. However, it is very time-consuming because the expert has to look at every retrieved image and make a decision if it is a true positive or not. Therefore not all possible parameter combinations are tested when doing the evaluation by an expert. The parameters space exploration in Section 6.2.3 gives already an estimation which combination delivers the best results. Hence, this knowledge is used to generate queries. A trade-off between the effort for the expert and significant results has to be found. 5 % from the available query images are selected, which are 135 images. For each query the 30 most similar images computed with the presented approach are retrieved. This means that in total the expert has to compare $135 * 30 = 4050$ images to the corresponding query image.

The expert is assisted in the evaluation task by a tool using HTML and JavaScript. Figure 6.14 shows the user interface of the evaluation tool. For each query an HTML page is generated. On the top of such a page a short description of the task is given. The domain expert has to decide for each retrieved image if it is relevant or not compared to the query image. The query image is delineated on the left and the retrieved images are arranged in a table layout with two columns on the right. The images are positioned from left to right and from top to bottom according to their computed distances to the query image. The most similar image is in the top left and the image with the greatest distance is on the bottom right. To see all retrieved images one has to use the scroll bar at the very right. The expert has to decide for every image if it matches the query image totally or just partially or if it is not similar at all. Clicking on an image once gives it an orange border and indicates a partial match. Clicking again on the image changes the color of the border to turquoise, which stands for total match. A third click lets the border disappear again and has the same meaning as not clicking on the image. No border means that the image is not interesting in respect to the query image. The expert sees the images only as 2D previews, but clicking on the text below the image links to the image volume. The text area below the query image leaves room for comments concerning the query image, the retrieved image or decision making. The *Next* button on the very bottom right directs to the next page with another query. When the expert is through with all queries the evaluation results can be sent by mail.

The content of such a mail can be parsed and used for computing meaningful measurements. For example, in this case it does not make sense to compute the recall because no information about the number of relevant images exists. Both, total matches and partial matches, are rated as true positives as it is suggested by Su [91]. This makes sense because it is possible that biologists can also derive information from partial matching images which are therefore relevant results. For future analysis a more detailed measurement, which distinguishes between these two groups can be computed. Sakai [77] and Järvelin et al. [41], for example, suggest two different ways of analysing graded evaluations.

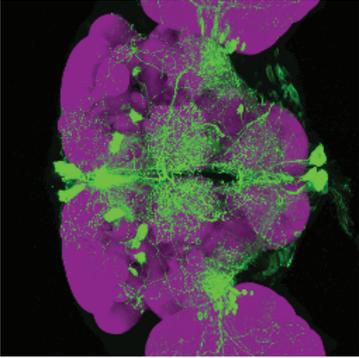
A useful measurement in this case is the precision at k , whereby $k = 1, 2, 3, 4, 5, 10, 15, 20, 25, 30$. This measure provides information about how many total or partial matches are in the first k images. For each query the values for the different k s are computed and averaged over the queries. This leads to ten measuring points. Figure 6.15 shows the results from the evaluation by an expert. The values for the ten k s are marked with an asterisk and the area above and below the curve depicts the standard deviation. The graph shows that for nearly every query the first

Survey

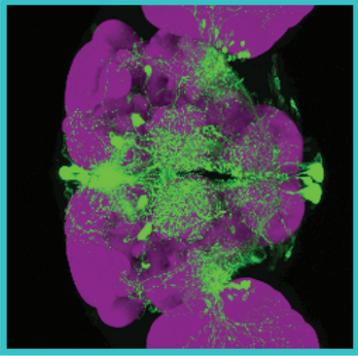
Check those images that are similar to the query image.

Click on an image if you think it is **partially** similar (the image gets an **orange** border).
If the image matches **totally** the query image click on the image once more (the image gets a **turquoise** border).
If you change your opinion and the image does **not match** at all, click again on the image until it has **no border**.

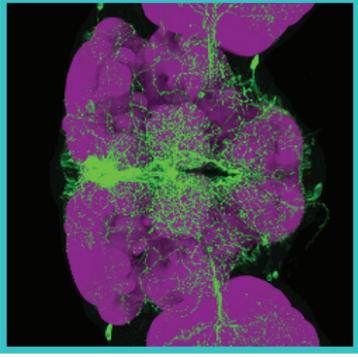
Query image
TP11082501L33Sum01.am



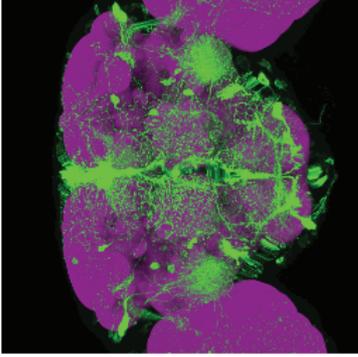
TP11082501L34Sum01.am



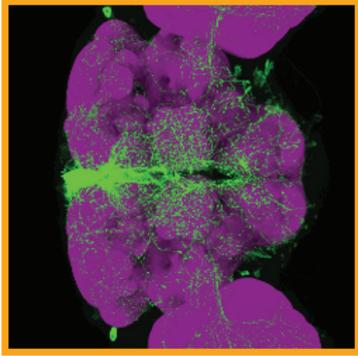
TP12031201L28Sum01.am



TP11050501L32Sum01.am



TP12031201L29Sum01.am



TP11082501L33Sum01.am



Next

Figure 6.14: An evaluation tool supporting the expert.

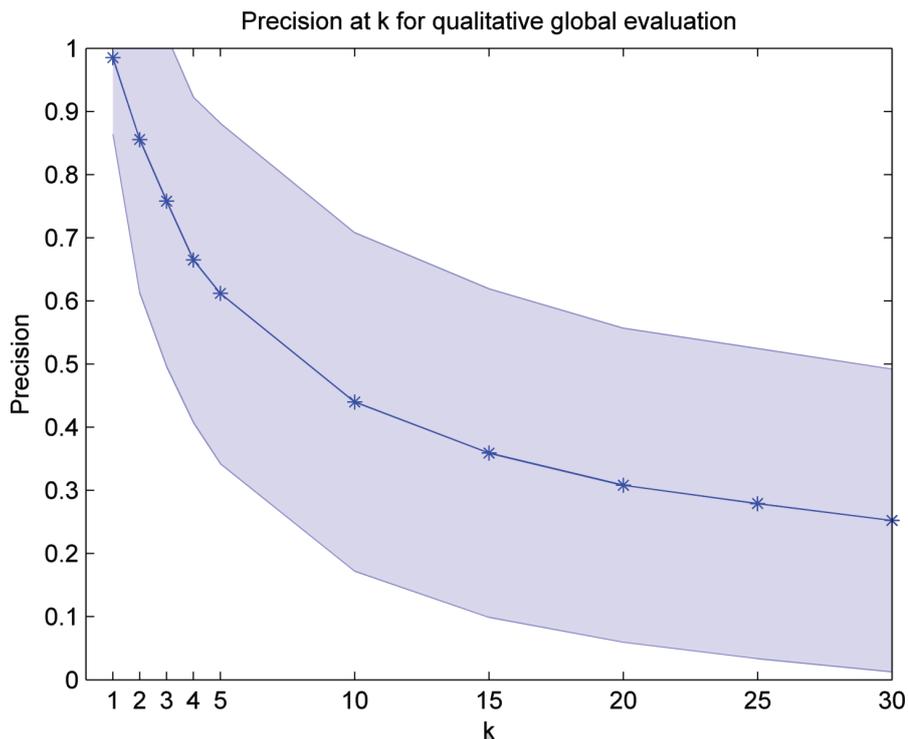


Figure 6.15: Graph showing *precision at k* values for ten *ks* with the standard deviation depicted as the area around the curve. The values are from an evaluation by an expert.

image was rated as a match. In the top five around two thirds of the images are true positives and in the top ten every second image is relevant on average.

6.3 Local Evaluation

The aim of a local search is to find images which are similar to the query image in some specific neuropils and which are independent of the similarity between other - not interesting- neuropils. The evaluation of the results of a local search is way more difficult than to perform a global evaluation. On the one hand no ground truth data for only a set of neuropils exist and on the other hand it is very tedious for an expert. It is not possible to make a decision just by looking at the preview image as it is often the case for global similarity because the expert has to estimate where the neuropils are on the image and they can overlap. Therefore, from an image where the third dimension is removed it can not be said in which neuropil staining is visible. Hence, in the course of this thesis an evaluation with quantitative measurements is not feasible.

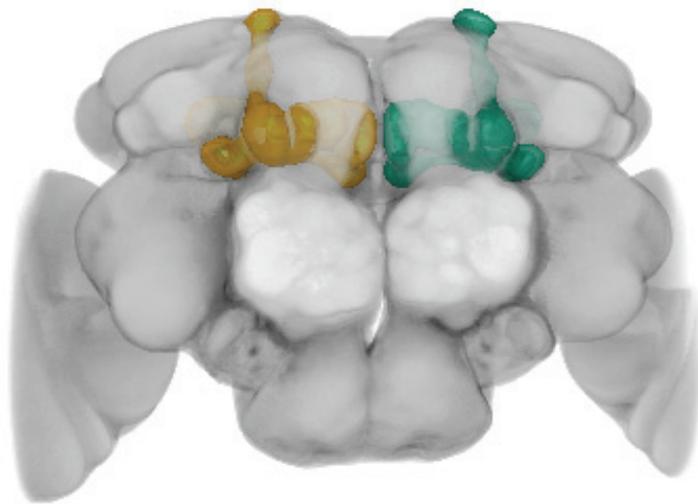
Instead a use case simulating the similarity search based on a set of neuropils is shown. The left and the right mushroom body are chosen for the local search because they are interesting for biologists and have a very outstanding shape and it is an example for which it is possible to visually get a feeling of the quality of the result. Figure 6.16 shows the two neuropils in

combination with the brain template from the front and at an angle from above. In Figure 6.17 an example for a local search with the mushroom body neuropils is presented. Figure 6.18 and Figure 6.19 show further retrieved images for this query. The query image is highlighted in turquoise and from left to right and from top to bottom are the most similar images represented. It can be seen that all retrieved images present the same structure as the query image for the left and right mushroom body. The similarity of the other neuropils does not matter as for example the fourth image shows. Another example query with a different staining of the two mushroom body neuropils can be seen in Figure 6.20 and is continued in Figure 6.21 and Figure 6.22. Again the results look very good. Nearly the whole neuropils are stained in the query image, therefore some overstained images are retrieved. For the third query depicted in Figure 6.23 and continued in Figure 6.24 and Figure 6.25 the neuropil set is extended by the ellipsoid body, which is the circular neuropil with the center close to the hemisphere and which can be seen clearly in the query image. In all retrieved images the staining in the mushroom body neuropils is similar to the one in the query image. However, the ellipsoid body is only partial visible or not visible at all in some result images. This retrieval result suggests that a feature vector with only two elements, as it is the case for the ellipsoid body, is not suitable to compare staining in a neuropil.

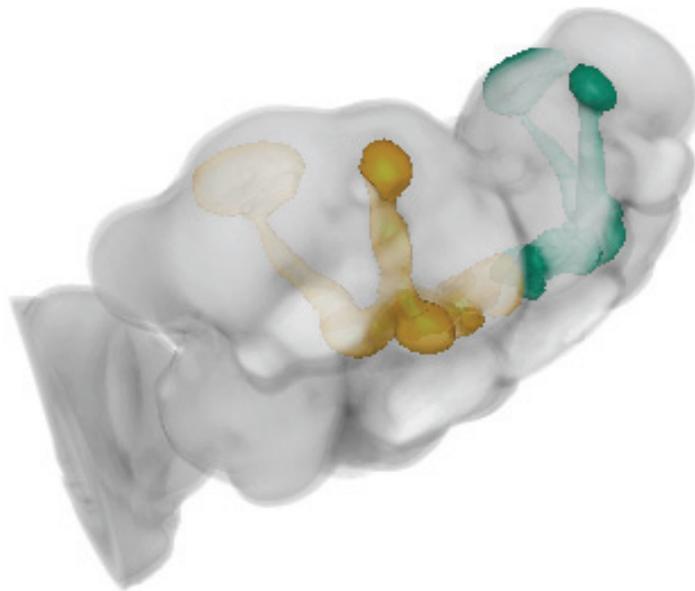
Another use case is to find images which are similar in some specific neuropils but differ in the remaining ones. This could be useful to find lines which show comparable neuronal structures in some regions and therefore possibly allow a formulation of a hypothesis about which genes are responsible for which staining. To achieve this the feature vectors are computed as usual and the differences of the neuropils between the query image and the other images are computed. The sum of the computed differences of the neuropils that should be similar is added to the sum of the differences of the remaining neuropils subtracted from one. The second term can be weighted by a suitable factor. It is more important that the similarity in the chosen neuropils is higher than the dissimilarity for the others. A corresponding formula is given in Equation 6.9, whereby d_i is the difference of the image i and the query image, SN stands for the set of neuropils that should be similar and DN is the set of neuropils which should differ. Therefore d_i^n indicates the difference of neuropil n between image i and the query image. α is a weighting factor between 0 and 1.

$$d_i = \sum_{n \in SN} d_i^n + \alpha * \sum_{n \in DN} (1 - d_i^n) \quad (6.9)$$

An example of such a query is shown in Figure 6.26 and is continued in Figure 6.27 and Figure 6.28. The same query image as for the local query search example in Figure 6.17 is used. This time the retrieved images should not only be similar in the left and right mushroom body, but should additionally differ in the remaining neuropils. The similarity for the differing neuropils is weighted by $\alpha = 0.5$ in this case so that the similarity of the mushroom body neuropils has a stronger impact. The weight can be chosen according to the context and the needs of the user. For this kind of queries overstained images constitute a problem if the query image has less staining in the neuropils which should differ.



(a) Front view



(b) View at an angle from above

Figure 6.16: Brain template with the left and right mushroom body highlighted.

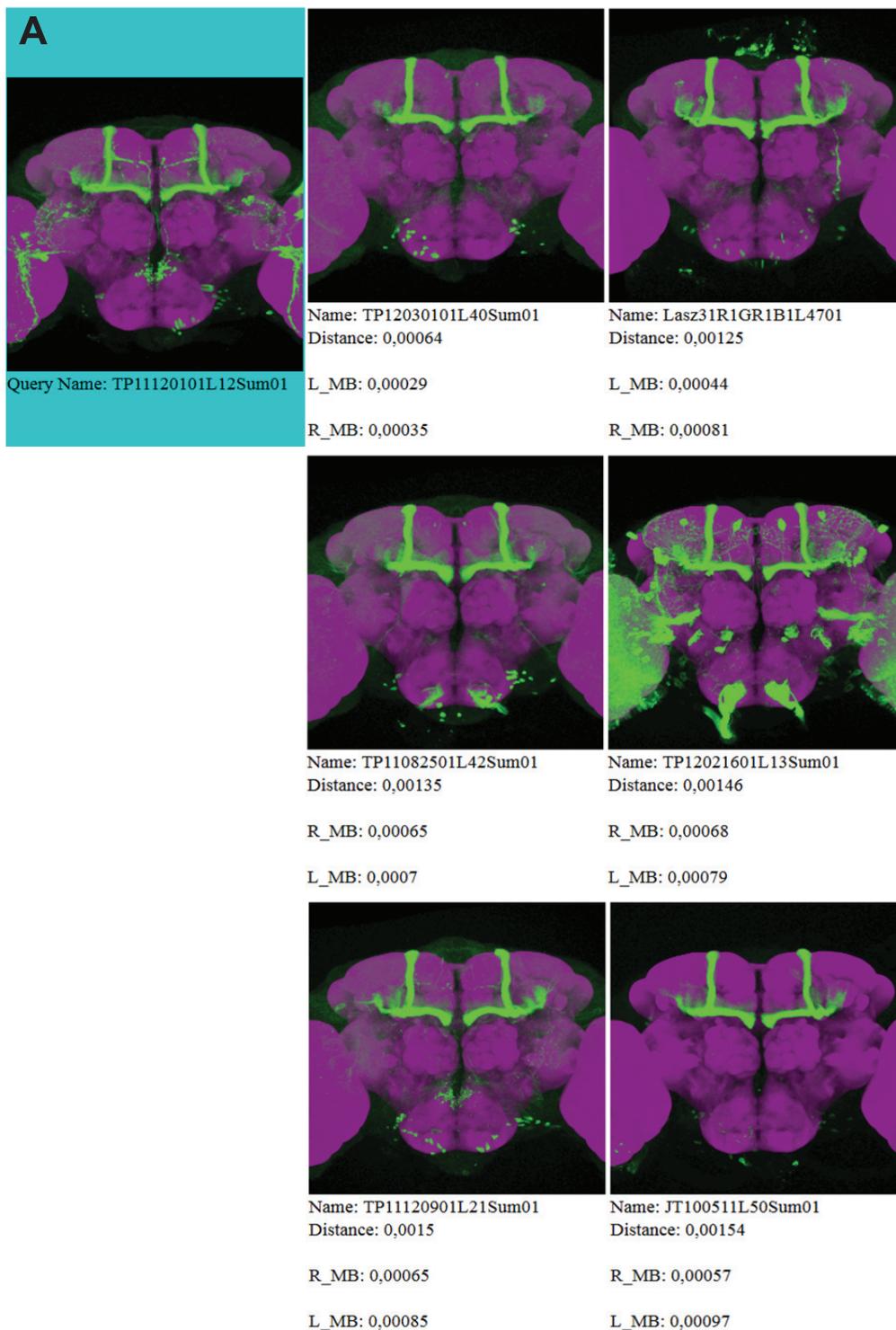


Figure 6.17: Local query search example A using left and right mushroom body showing the first six retrieved images (continued in Figure 6.18 and Figure 6.19).

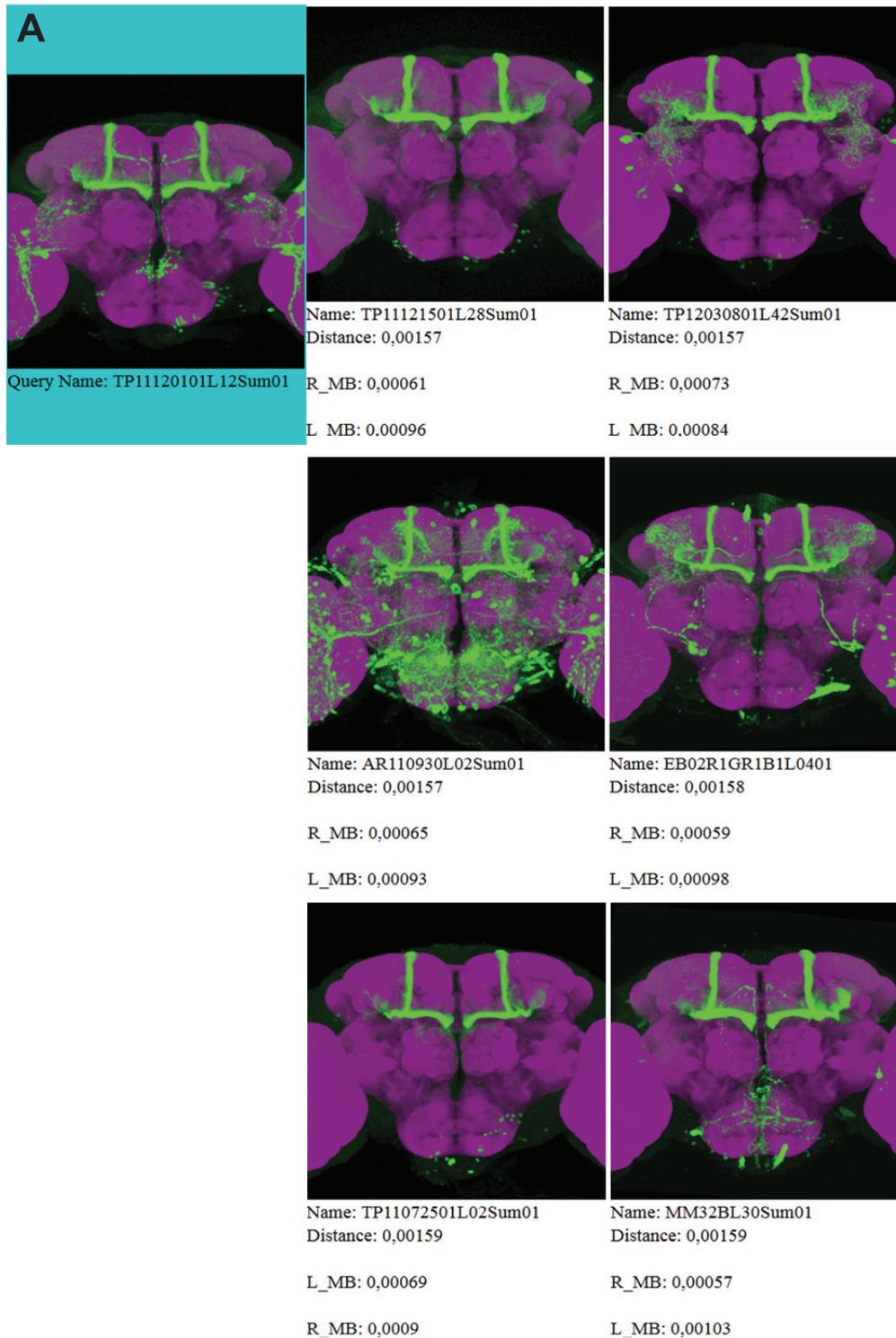


Figure 6.18: Local query search example A (continuation of Figure 6.17) using left and right mushroom body showing rank seven to twelve of the retrieved images.

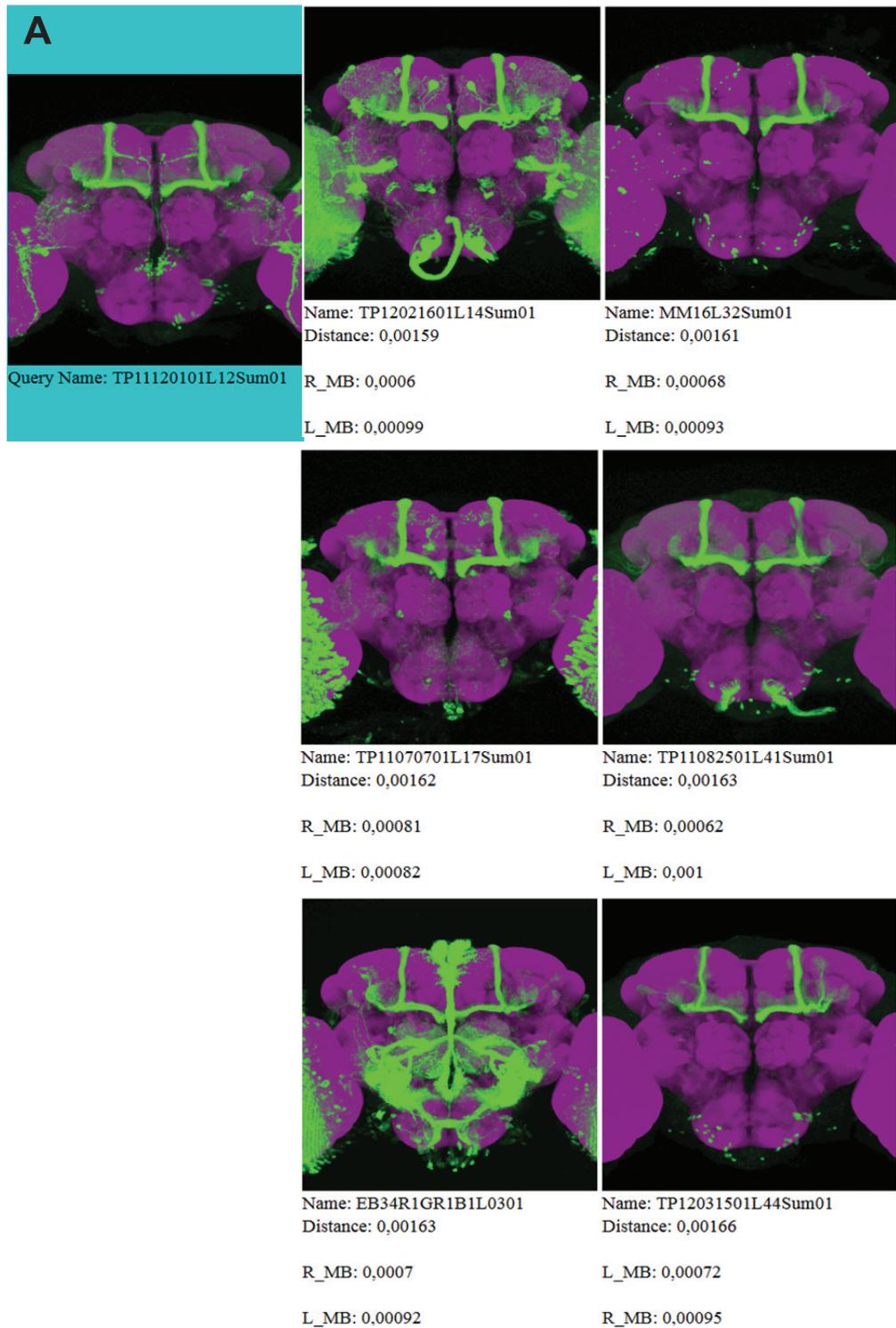


Figure 6.19: Local query search example A (continuation of Figure 6.18) using left and right mushroom body showing rank 13 to 18 of the retrieved images.

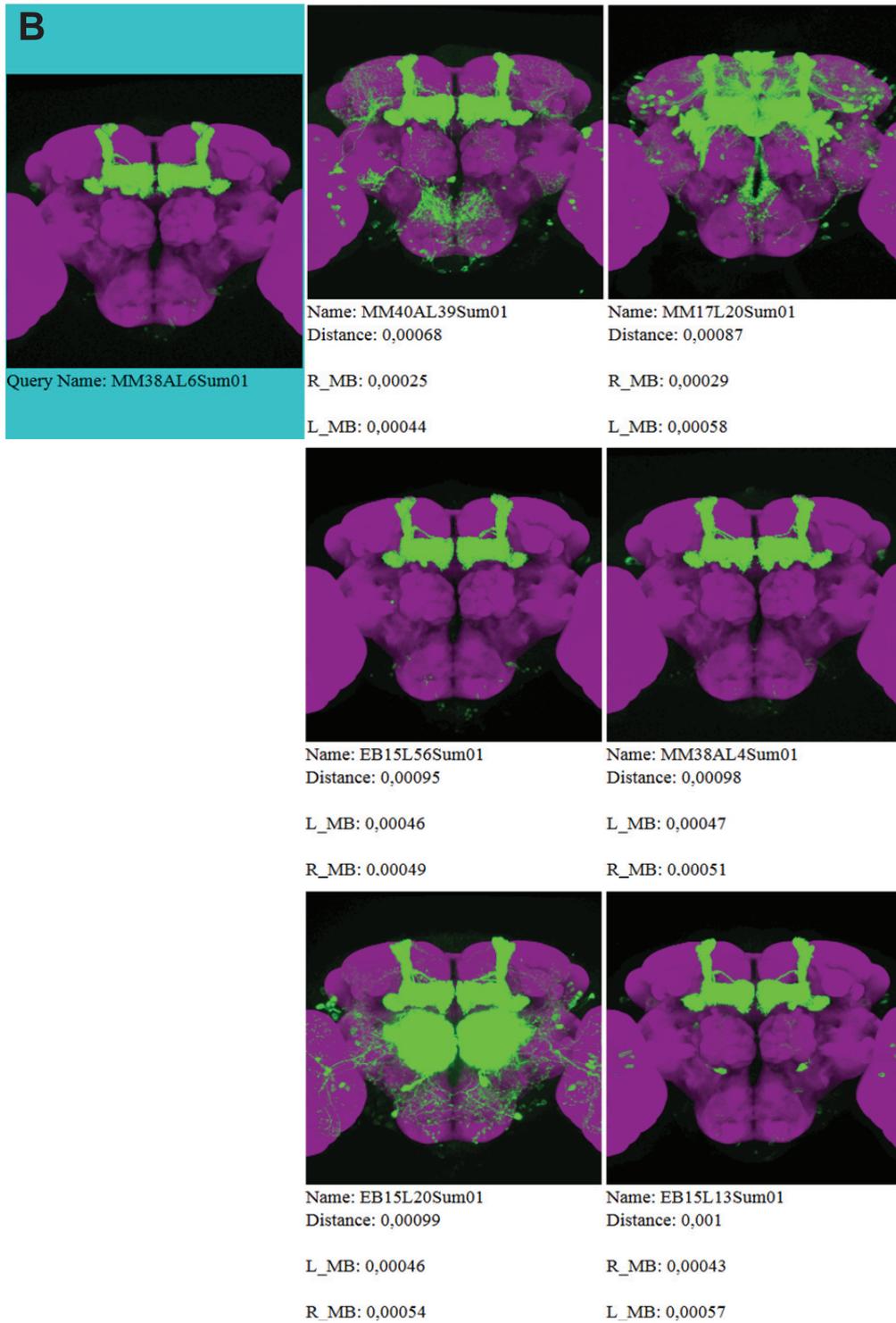


Figure 6.20: Local query search example B using left and right mushroom body showing the first six retrieved images (continued in Figure 6.21 and Figure 6.22).

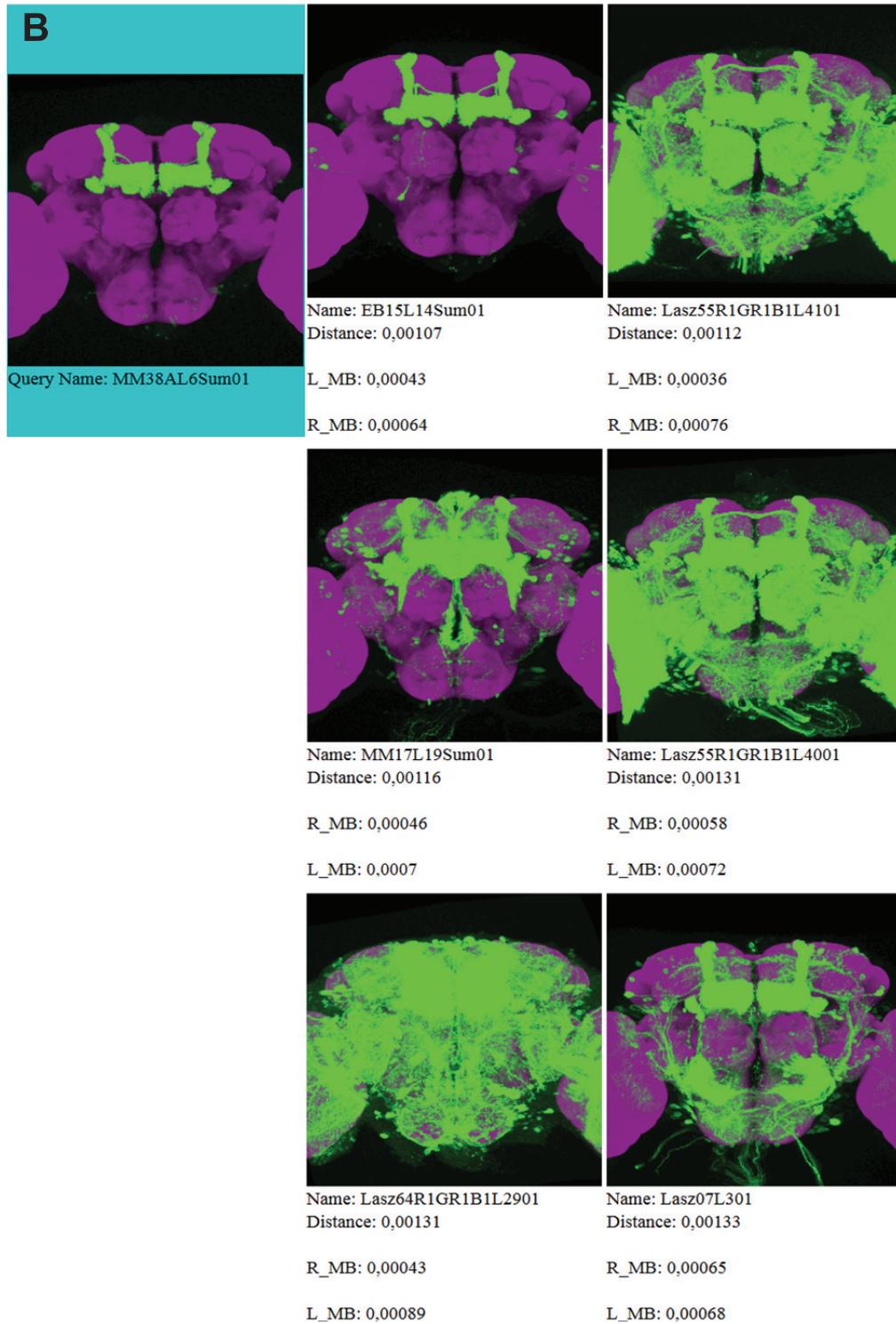


Figure 6.21: Local query search example B (continuation of Figure 6.20) using left and right mushroom body showing rank seven to twelve of the retrieved images.

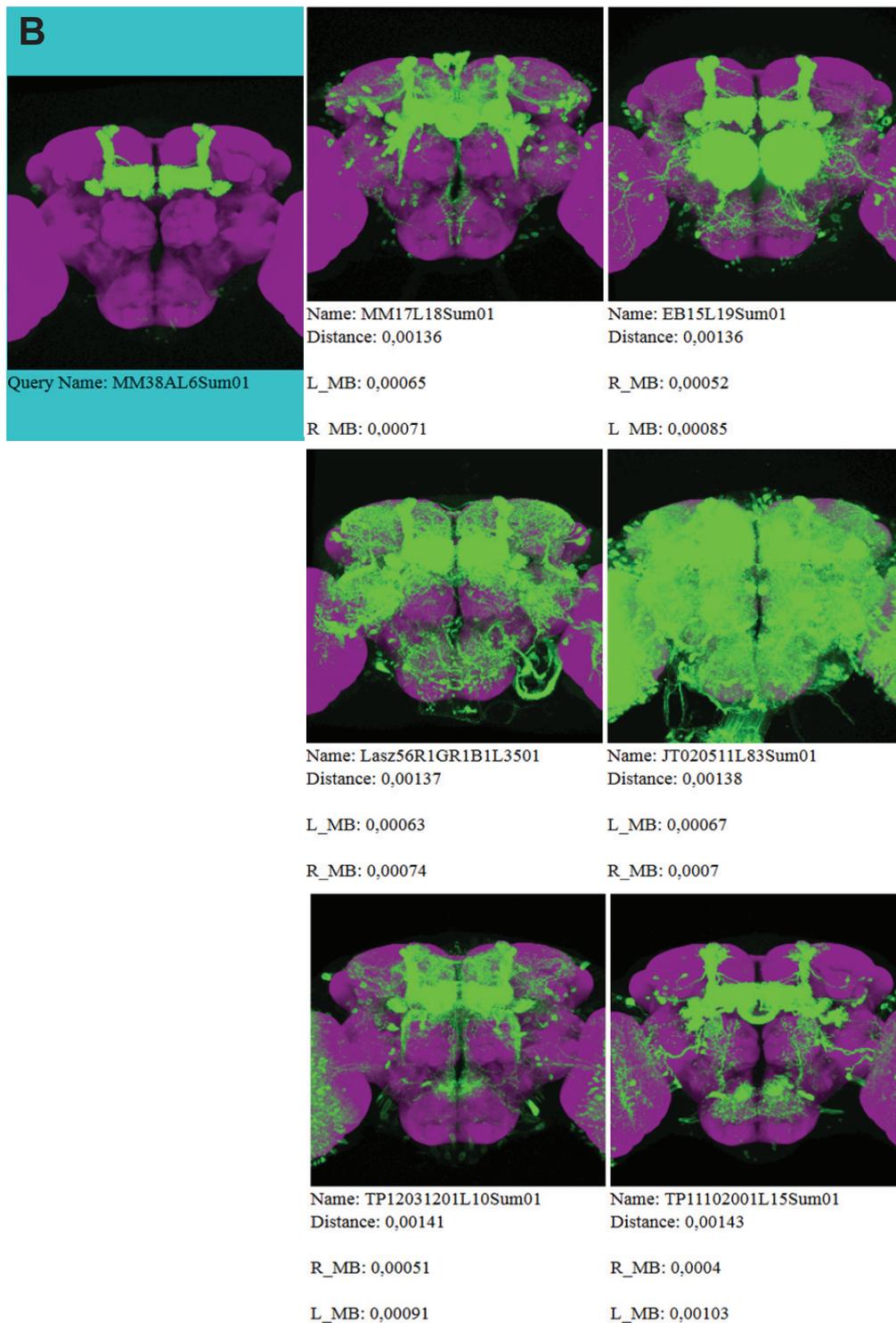


Figure 6.22: Local query search example B (continuation of Figure 6.21) using left and right mushroom body showing rank 13 to 18 of the retrieved images.

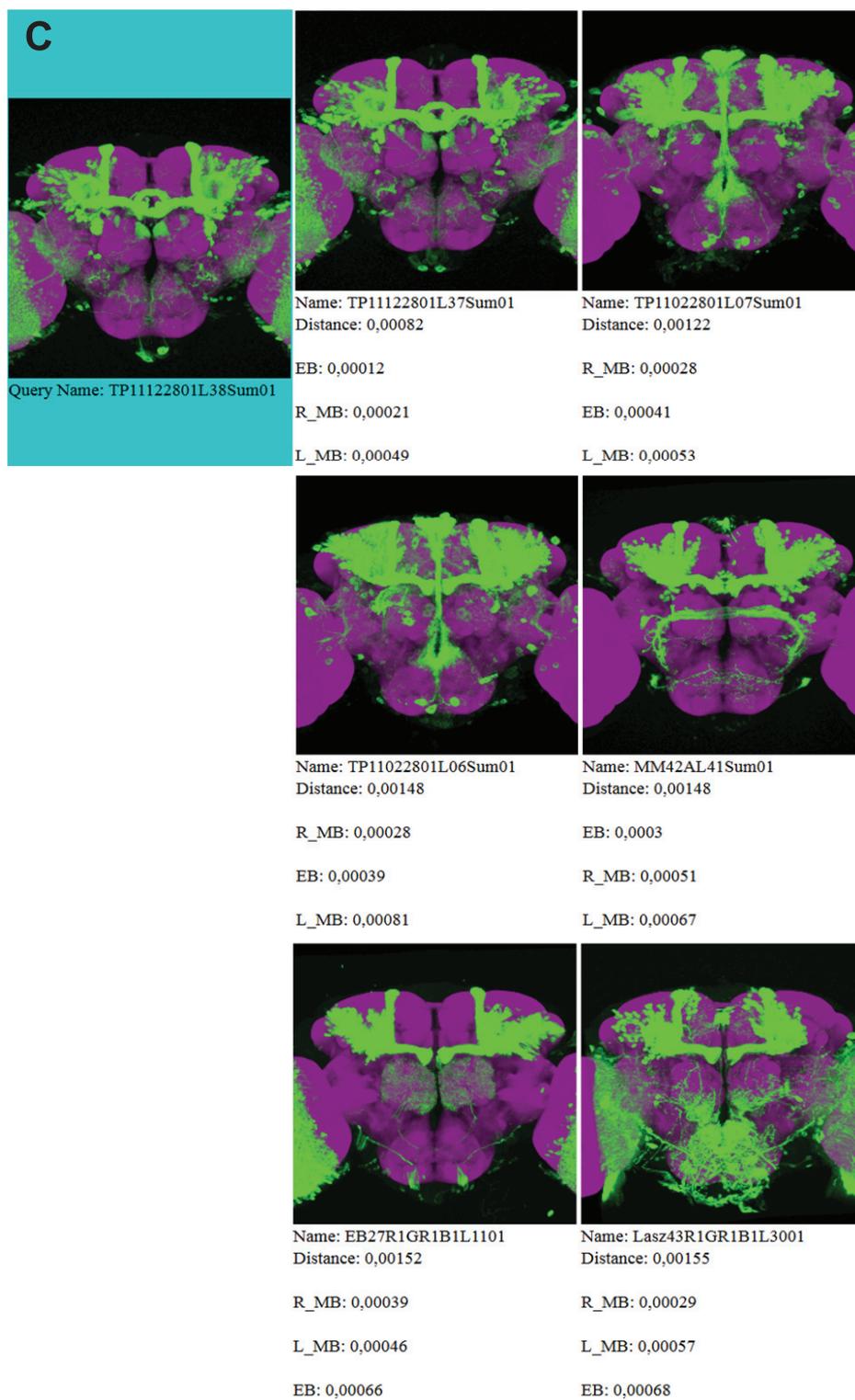


Figure 6.23: Local query search example C using left and right mushroom body and the ellipsoid body showing the first six retrieved images (continued in Figure 6.24 and Figure 6.25).

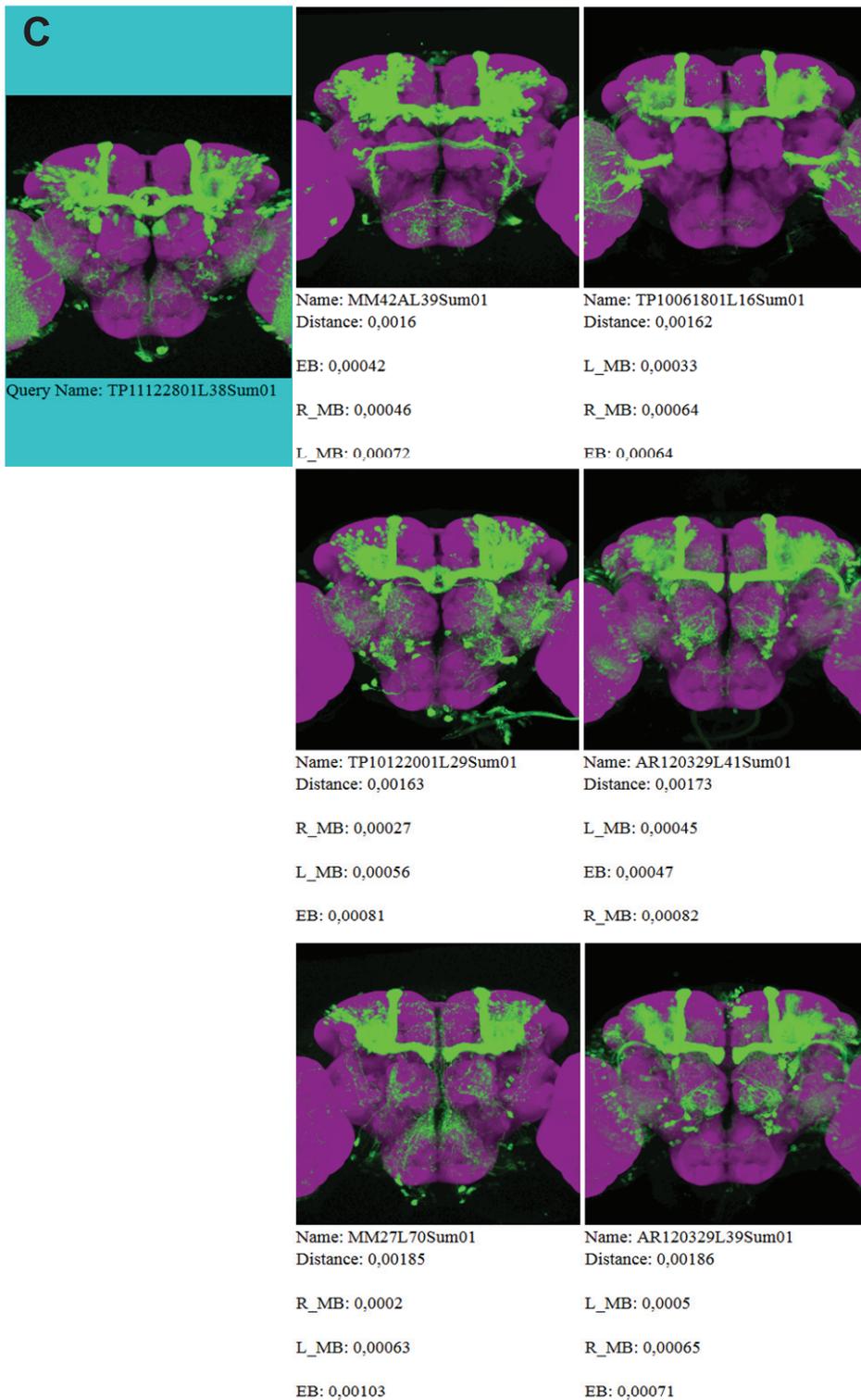


Figure 6.24: Local query search example C (continuation of Figure 6.23) using left and right mushroom body and the ellipsoid body showing rank seven to twelve of the retrieved images.

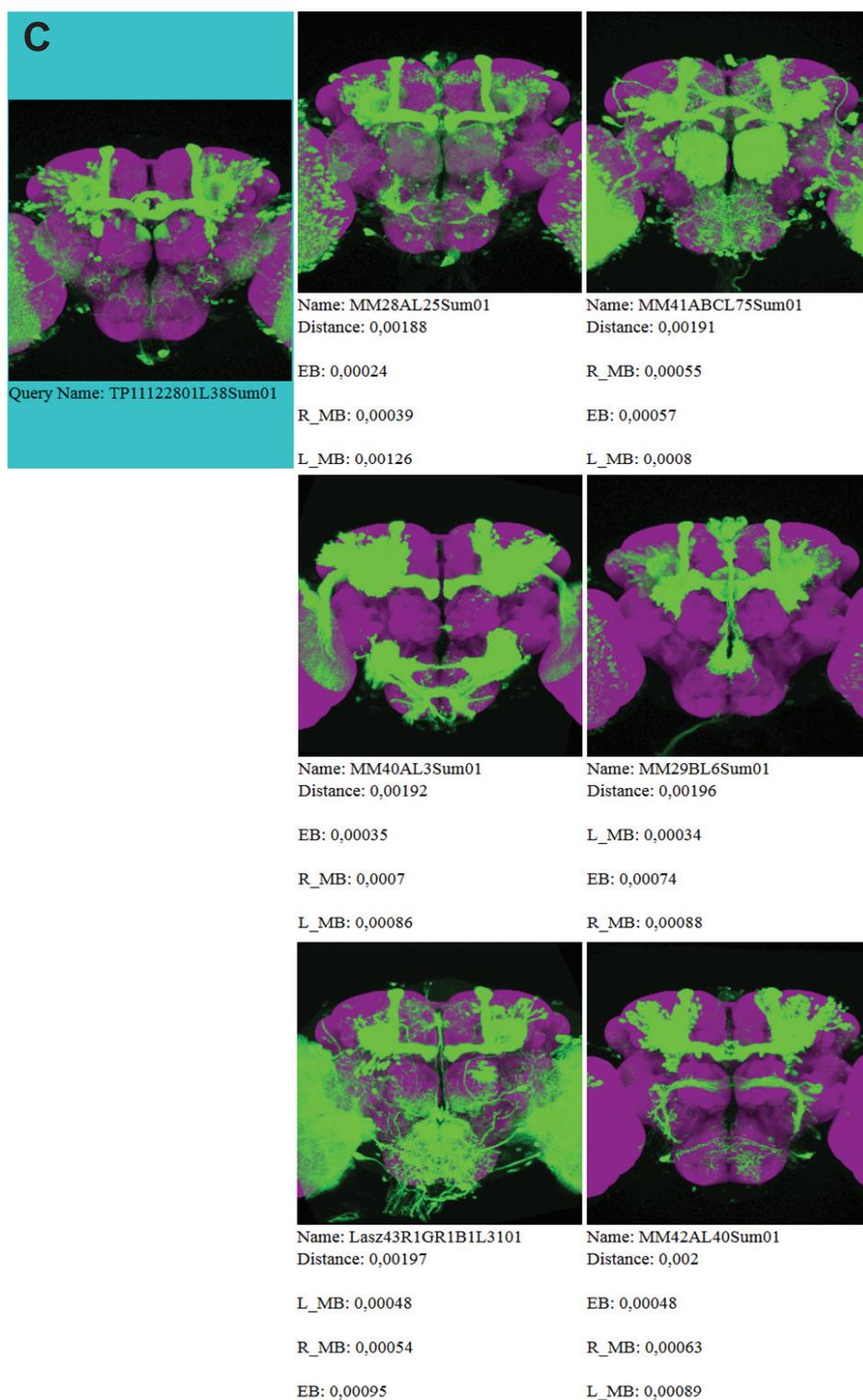


Figure 6.25: Local query search example C (continuation of Figure 6.24) using left and right mushroom body and the ellipsoid body showing rank 13 to 18 of the retrieved images.

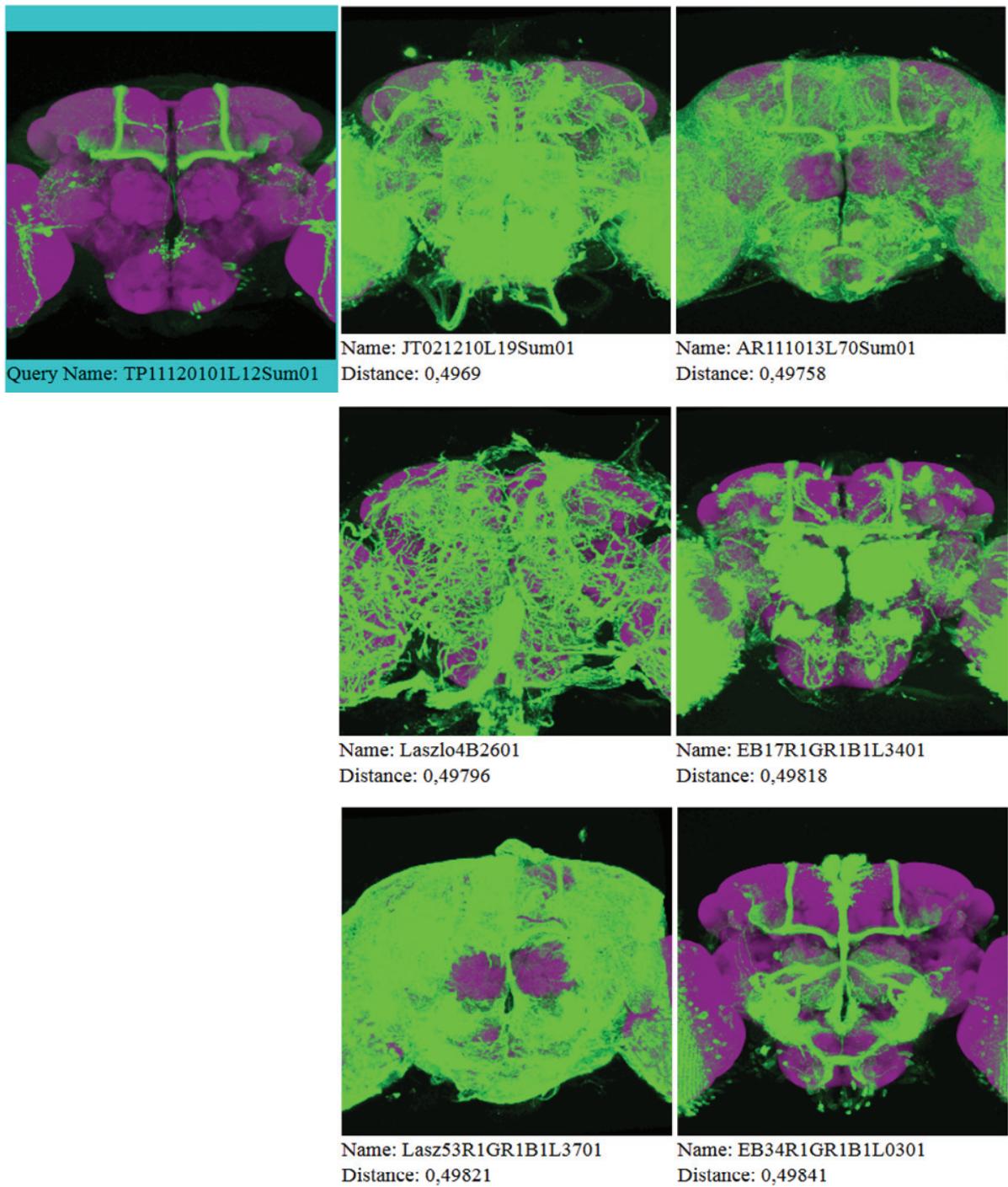


Figure 6.26: A complex query search where the staining in the left and right mushroom body should be similar and different in the remaining neuropils. The first six retrieved images are shown (continued in Figure 6.27 and Figure 6.28).

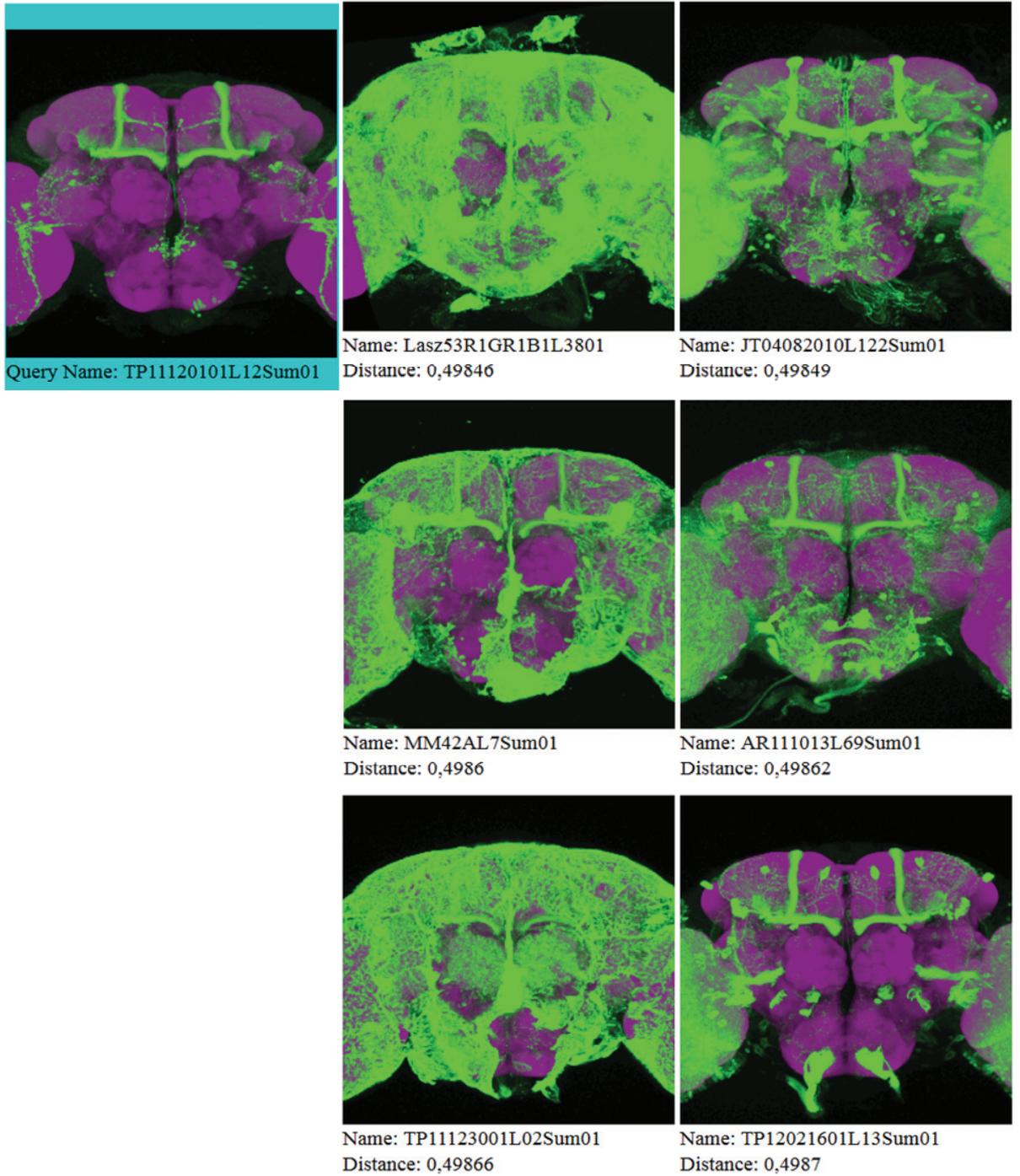


Figure 6.27: A complex query search (continuation of Figure 6.26) where the staining in the left and right mushroom body should be similar and different in the remaining neuropils. Rank seven to twelve of the retrieved images are shown.

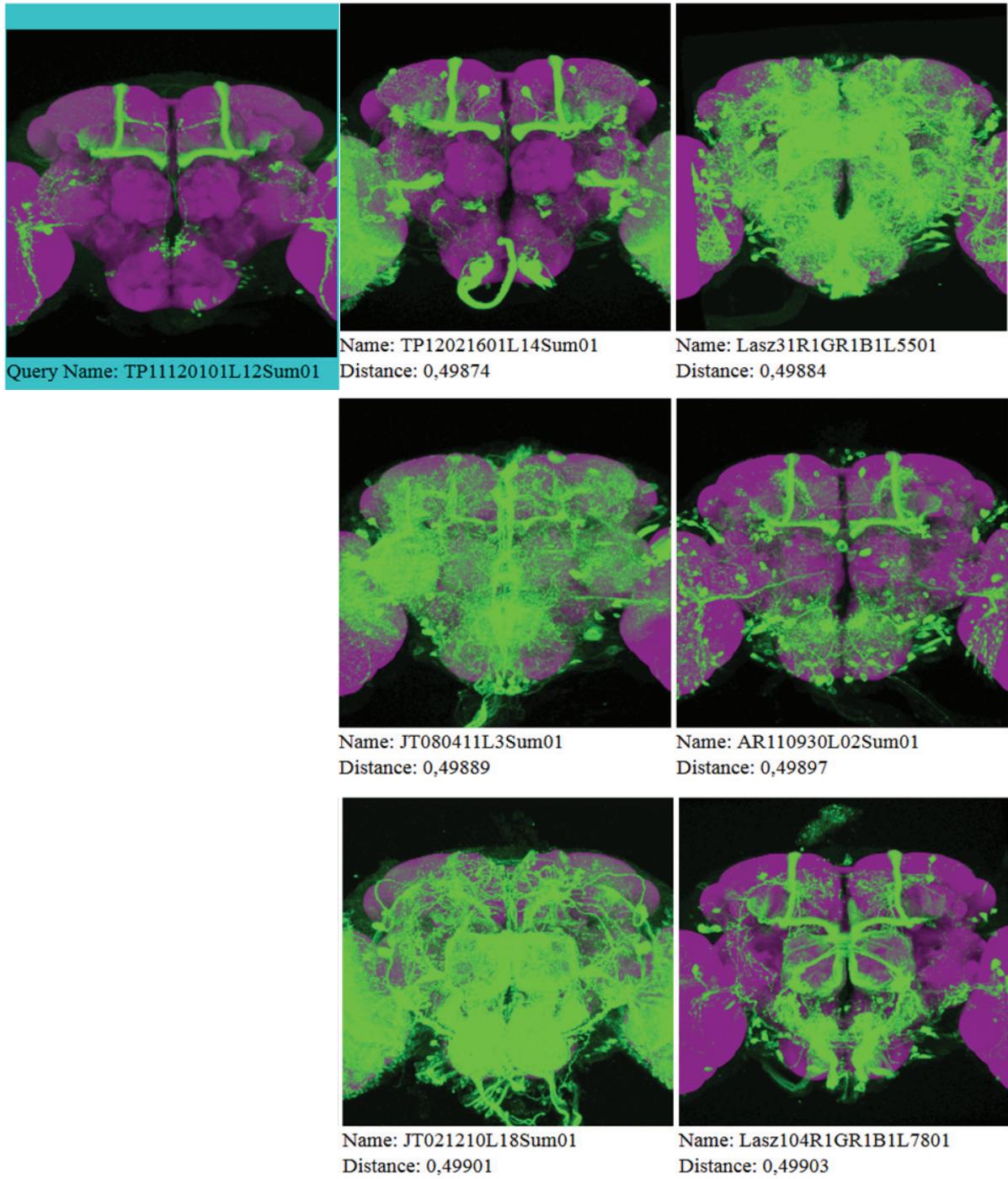


Figure 6.28: A complex query search (continuation of Figure 6.27) where the staining in the left and right mushroom body should be similar and different in the remaining neuropils. Rank 13 to 18 of the retrieved images are shown.

Conclusion and Future Work

This chapter sums up the invented approach and gives an overview of possible extensions and improvements.

7.1 Summary and Conclusion

To get an insight into how the brain works the species *Drosophila melanogaster* is used as model system. It is still unclear which neurons are connected, how they interact and which neurons control which behavior. Scientists approach this problem by performing high throughput screenings of large genetic line collections and subsequently search manually in the resulting image database for those which possibly can be used in experiments and might help reveal the mechanics of the investigated brain function.

The aim of this thesis is to significantly reduce biologists work load by giving them the computer-assisted possibility to search for visually similar images and therefore only have to look at a subset of images. Hence, a content based image retrieval approach for co-registered 3D images of fly brains was developed. According to a defined template the brain can be divided into several regions, so called neuropils. For each image and each neuropil three MIPs along the main axes are generated and the gray values are extracted. For each neuropil a PCA based on the gray values is performed for dimension reduction purposes. Each neuropil from each image is then transformed into a lower dimensional space while keeping most of the information. In this spaces the neuropils can be compared in a computationally efficient way. Summing up the distances between the neuropils of two images leads to an overall distance.

The comparison of two images with this approach is much faster than a voxel by voxel comparison and is more robust against slight variations of neuronal structures. The distances are computed separately for each neuropil, which is a big advantage because the needs of the biologists can be addressed in a more flexible way. They can decide which regions are of interest for their research and in which neuropils the images should be similar.

An extensive evaluation was carried out to show the effects of different parameter values on the results. The parameter combination which provides the best retrieval results was found

by an evaluation using meta-data. An expert of the *Drosophila melanogaster* research field was asked to rate the retrieved images of more than 130 queries. The results indicate that the proposed approach is one very satisfying way to search for fly brain images based on the present expressed neuronal structures. The only prerequisites are a set of images to perform a reliable PCA and the biologist needs an image showing interesting neurons as query image.

7.2 Future Work

The possibility to show the top most similar neuropils for each image compared to the query image has to be improved. Most of the time very small neuropils are listed with a distance of nearly zero. The reason for this is that they often contain no staining. Therefore the results are mathematically correct and it is sometimes good to know that in both neuropils there is no staining. However, it is desirable that the results make sense in a semantic way. The user should have the possibility to decide that neuropils with no or less staining should not be listed. It is context-sensitive where the threshold should be set. Additionally the length of the feature vector has to be taken into account because comparing two vectors with few elements leads to a smaller distance than vectors with many elements.

Using PCA on gray values extracted from MIPs is only one out of several ways to achieve the goal of finding similar images. It would be possible to use PCA on 3D data like Russ et al. [76] described it for the use of face recognition. This approach requires more memory to hold the data needed to compute PCA and also the computation itself is more expensive.

Testing other feature descriptors like Gabor filter, Frangi filter, Sato filter or tensor structures is worth a try. The outcomes can then be reduced by PCA again if the feature vectors are very long.

Random projection [18] could be used as an alternative to PCA. Both are linear mapping methods, but random projection is computationally less expensive. The high dimensional data are projected into a lower dimensional space using a random matrix R consisting of columns with unit length. The elements of R can be chosen randomly and should be uniform and independent. More details, examples and a comparison between PCA and random projection can be found in the work of Bingham and Mannila [8] and in Dasgupta's work [19]. It would take again some time to evaluate the results to find out the optimal dimension of R .

The approach developed in the course of this thesis has the big disadvantage that very small neuropils have only a few gray values as input for PCA after downsampling the MIPs. Therefore structures get lost and results are fuzzy because the feature vector contains only very few elements to describe this neuropil. One possibility to preserve the structure is to cut out the neuropil according to its bounding box from the original MIP and downscale it afterwards to a specified size. The drawback of this approach is that PCA takes very long because all neuropils consist of many gray values and therefore have a big covariance matrix. Additionally the feature vector length increases dramatically which raises the time to compare two images.

A further development is to do the image retrieval based on semantic knowledge. For this a finer subdivision than neuropils, for example three-dimensional blocks, is necessary. A suitable block size has to be determined. A classification of the staining within each block should be done with an appropriate feature descriptor. Then a biologist can search for example for images with a projection or cell body at a specific position.

The information about image similarity can be used to build clusters. For example, the average distance between lines can be computed and clustered. This requires the extraction of the similarity information of the images belonging to one line compared to the images belonging to another line. These values have to be averaged by computing for example the mean or the median.

Bibliography

- [1] T. ANDRYSIAK and M. CHORAŚ. Image retrieval based on hierarchical Gabor filters. *International Journal of Applied Mathematics and Computer Science (AMCS)*, 15:471–480, 2005.
- [2] S. ANTANI, L.R. LONG, and G.R. THOMA. A biomedical information system for combined content-based retrieval of spine X-ray images and associated text information. In *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2002.
- [3] D.S. ASADI, CH.D.V.S. RAO, and V. SAIKRISHNA. A comparative study of face recognition with principal component analysis and cross-correlation technique. *International Journal of Computer Applications (IJCA)*, 10(8):17–21, 2010.
- [4] J.R. BACH, C. FULLER, A. GUPTA, A. HAMPAPUR, B. HOROWITZ, R. HUMPHREY, R. JAIN, and C.-F. SHU. Virage image search engine: an open framework for image management. In *Electronic Imaging: Science & Technology*, pages 76–87. International Society for Optics and Photonics, 1996.
- [5] S. BATTY, J. CLARK, T. FRYER, and X. GAO. Prototype system for semantic retrieval of neurological PET images. In *Medical Imaging and Informatics*, pages 179–188. Springer, 2008.
- [6] S. BERRETTI, A. DEL BIMBO, and P. PALA. Content based retrieval of 3D cellular structures. In *Proceedings of the Second International Conference on Multimedia and Exposition (ICME)*, pages 1096–1099. IEEE, 2001.
- [7] W. D. BIDGOOD, S. C HORII, F.W. PRIOR, and D.E. VAN SYCKLE. Understanding and using DICOM, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association (JAMIA)*, 4(3):199–212, 1997.
- [8] E. BINGHAM and H. MANNILA. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 245–250. ACM, 2001.
- [9] J.E. BLANKENSHIP and B. HOUCK. Nervous system (invertebrate). *AccessScience*, 2012.

- [10] A.C. BOVIK, M. CLARK, and W.S. GEISLER. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 12(1):55–73, 1990.
- [11] A.H. BRAND and N. PERRIMON. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, 118(2):401–415, 1993.
- [12] S. BRUCKNER, V. SOLTÉSZOVÁ, M. E. GRÖLLER, J. HLADŮVKA, K. BÜHLER, J. Y. YU, and B. J. DICKSON. Braingazer - visual queries for neurobiology research. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1497–1504, 2009.
- [13] G. BUCCI, S. CAGNONI, and R. DE DOMINICIS. Integrating content-based retrieval in a medical image reference database. *Computerized Medical Imaging and Graphics (CMIG)*, 20(4):231–241, 1996.
- [14] T.W. CAI, J. KIM, and D.D. FENG. *Biomedical Information Technology*. Academic Press, 2011.
- [15] J. CANNY. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, PAMI-8(6):679–698, 1986.
- [16] D.A. CLAUSI and M.E. JERNIGAN. Designing Gabor filters for optimal texture separability. *Pattern Recognition*, 33(11):1835–1849, 2000.
- [17] R.L. COSGRIFF. Identification of shape. *Ohio State Univ. Res. Foundation, Columbus, OH, Tech. Rep. ASTIA AD*, 254:792, 1960.
- [18] S. DASGUPTA. Learning mixtures of Gaussians. In *Proceedings of the 40th IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, pages 634–644. IEEE, 1999.
- [19] S. DASGUPTA. Experiments with random projection. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 143–151. Morgan Kaufmann Publishers Inc., 2000.
- [20] R. DATTA, D. JOSHI, J. LI, and J.Z. WANG. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 40(2):5, 2008.
- [21] R. DATTA, J. LI, and J.Z. WANG. Content-based image retrieval: approaches and trends of the new age. In *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*, pages 253–262. ACM, 2005.
- [22] J.G. DAUGMAN. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A: Optics and Image Science (JOSA A)*, 2(7):1160–1169, 1985.
- [23] J.E.E DE OLIVEIRA, A. MACHADO, G.C. CHAVEZ, A.P.B. LOPES, T.M. DESERNO, and A.D.A. ARAÚJO. Mammosys: A content-based image retrieval system using breast density patterns. *Computer Methods and Programs in Biomedicine*, 99(3):289–297, 2010.

- [24] A. DEPEURSINGE, A. VARGAS, A. PLATON, A. GEISSBUHLER, P.-A. POLETTI, and H. MÜLLER. 3D case-based retrieval for interstitial lung diseases. In *Medical Content-Based Retrieval for Clinical Decision Support (MCBR-CDS)*, pages 39–48. Springer, 2010.
- [25] B.J. DICKSON. Wired for sex: the neurobiology of *Drosophila* mating decisions. *Science*, 322(5903):904–909, 2008.
- [26] R.O. DUDA and P.E. HART. *Pattern classification and scene analysis*, volume 3. John Wiley & Sons, New York, 1973.
- [27] K. ENGEL, M. HADWIGER, J.M. KNISS, C. REZK-SALAMA, and D. WEISKOPF. *Real-time volume graphics*. AK Peters, Ltd., 2006.
- [28] M. FLICKNER, H. SAWHNEY, W. NIBLACK, J. ASHLEY, Q. HUANG, B. DOM, M. GORKANI, J. HAFNER, D. LEE, D. PETKOVIC, D. STEELE, and P. YANKER. Query by image and video content: The QBIC system. *Computer*, 28(9):23–32, 1995.
- [29] FlyBase. Hartenstein (1993) central nervous system adult. <http://flybase.org/reports/FBim0000656.html>. Online; Accessed: June, 16th 2014.
- [30] FlyBase. Wallace (1934) drawing of male and female. <http://flybase.org/reports/FBim0000001.html>. Online; Accessed: June, 16th 2014.
- [31] K. FUKUNAGA. *Introduction to statistical pattern recognition*. Academic Press, New York, 1990.
- [32] T. GLATARD, J. MONTAGNAT, and I.E. MAGNIN. Texture based medical image indexing and retrieval: application to cardiac imaging. In *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*, pages 135–142. ACM, 2004.
- [33] M. HAO, Z. JIANKE, M.R.-T. LYU, and I. KING. Bridging the semantic gap between image contents and tags. *IEEE Transactions on Multimedia*, 12(5):462–473, 2010.
- [34] R.M. HARALICK, K. SHANMUGAM, and I.H. DINSTEN. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics (SMC)*, 3(6):610–621, 1973.
- [35] T. HASTIE, R. TIBSHIRANI, and J. FRIEDMAN. *The elements of statistical learning*, volume 2. Springer, 2009.
- [36] M-K HU. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.
- [37] J. HUANG, S. R. KUMAR, M. MITRA, W.-J. ZHU, and R. ZABIH. Image indexing using color correlograms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 762–768. IEEE, 1997.

- [38] K. ITO, K. SHINOMIYA, M. ITO, J.D. ARMSTRONG, G. BOYAN, V. HARTENSTEIN, S. HARZSCH, M. HEISENBERG, U. HOMBERG, A. JENETT, H. KESHISHIAN, L.L. RESTIFO, W. RÖSSLER, J.H. SIMPSON, N.J. STRAUSFELD, R. STRAUSS, L.B. VOSSHALL, and Insect Brain Name Working Group. A systematic nomenclature for the insect brain. *Neuron*, 81(4):755–765, 2014.
- [39] A. JAIN, K. NANDAKUMAR, and A. ROSS. Score normalization in multimodal biometric systems. *Pattern recognition*, 38(12):2270–2285, 2005.
- [40] A.K. JAIN and A. VAILAYA. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, 1996.
- [41] K. JÄRVELIN and J. KEKÄLÄINEN. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.
- [42] M. JIE, Z. ZHIWEI, T. HONGMEI, and Z. QUANMING. Fast Fourier descriptor method of the shape feature in low resolution images. In *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, pages 1–4. IEEE, 2010.
- [43] Y. KAWATA, N. NIKI, H. OHMATSU, M. KUSUMOTO, R. KAKINUMA, K. MORI, H. NISHIYAMA, K. EGUCHI, M. KANEKO, and N. MORIYAMA. Three-dimensional CT image retrieval in a database of pulmonary nodules. In *Proceedings of the 2002 International Conference on Image Processing (ICIP)*, volume 3, pages 149–152. IEEE, 2002.
- [44] L. KEYES and A. WINSTANLEY. Using moment invariants for classifying shapes on large-scale maps. *Computers, environment and urban systems*, 25(1):119–130, 2001.
- [45] A. KUMAR, J. KIM, W. CAI, M. FULHAM, and D. FENG. Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. *Journal of Digital Imaging (JDI)*, pages 1–15, 2013.
- [46] A.S. KURANI, D.-H. XU, J. FURST, and D.S. RAICU. Co-occurrence matrices for volumetric data. In *7th IASTED International Conference on Computer Graphics and Imaging (CGIM)*, pages 447–452, 2004.
- [47] T.M. LEHMANN, M.O. GULD, C. THIES, B. FISCHER, D. KEYSERS, M. KOHNEN, H. SCHUBERT, and B.B. WEIN. Content-based image retrieval in medical applications for picture archiving and communication systems. In *Medical Imaging 2003*, pages 109–117. International Society for Optics and Photonics, 2003.
- [48] S. LI, T. GONG, J. WANG, R. LIU, C.L. TAN, T.Y. LEONG, B.C. PANG, C.C.T. LIM, C.K. LEE, Q. TIAN, and Z. ZHANG. Tbidoc: 3D content-based CT image retrieval system for traumatic brain injury. In *Proceedings of the SPIE Medical Imaging*, volume 7624, pages 762427–762427–10, 2010.

- [49] S. LIU, W. CAI, L. WEN, S. EBERL, M.J. FULHAM, and D. FENG. A robust volumetric feature extraction approach for 3D neuroimaging retrieval. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5657–5660. IEEE, 2010.
- [50] Y. LIU, N.A. LAZAR, W.E. ROTHFUS, F. DELLAERT, A. MOORE, J. SCHNEIDER, and T. KANADE. Semantic based biomedical image indexing and retrieval. In *Trends and Advances in Content-Based Image and Video Retrieval*. Springer, 2004.
- [51] S. LLOYD. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- [52] W.-Y. MA and H.J. ZHANG. Benchmarking of image features for content-based retrieval. In *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers*, volume 1, pages 253–257. IEEE, 1998.
- [53] T.F.C. MACKAY and R.R.H. ANHOLT. Of flies and man: *Drosophila* as a model for human complex traits. *Annual Review of Genomics and Human Genetics*, 7:339–367, 2006.
- [54] R.C. MAHALANOBIS. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2:49–55, 1936.
- [55] B.S. MANJUNATH, J.-R. OHM, V.V. VASUDEVAN, and A. YAMADA. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.
- [56] C.D. MANNING, R. RAGHAVAN, and H. SCHÜTZE. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008.
- [57] B.M. MEHTRE, M.S. KANKANHALLI, and W.F. LEE. Shape measures for content based image retrieval: a comparison. *Information Processing & Management (IP&M)*, 33(3):319–337, 1997.
- [58] H. MIN and Y. SHUANGYUAN. Overview of content-based image retrieval with high-level semantics. In *3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, volume 6, pages V6–312. IEEE, 2010.
- [59] J. MONTAGNAT, H. DUQUE, J.-M. PIERSON, V. BRETON, L. BRUNIE, and I. MAGNIN. Medical image content-based queries using the grid. In *Proceedings of the HealthGrid Conference*, pages 138–147, 2003.
- [60] H. MÜLLER, N. MICHOUX, D. BANDON, and A. GEISSBUHLER. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International Journal of Medical Informatics (IJMI)*, 73(1):1–23, 2004.

- [61] H. MÜLLER, W. MÜLLER, D.M. SQUIRE, S. MARCHAND-MAILLET, and T. PUN. Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recognition Letters*, 22(5):593–601, 2001.
- [62] E. MURPHY-CHUTORIAN and C. ROSENBERG. Similar images graduates from google labs. <http://googleblog.blogspot.co.at/2009/10/similar-images-graduates-from-google.html>. Online; Accessed: June, 16th 2014.
- [63] M.S. NIXON and A.S. AGUADO. *Feature Extraction and Image Processing for Computer Vision*. Academic Press, New York, 2012.
- [64] T. OJALA, M. PIETIKÄINEN, and T. MÄENPÄÄ. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 24(7):971–987, 2002.
- [65] S.R. OLSEN and R.I. WILSON. Cracking neural circuits in a tiny brain: new approaches for understanding the neural circuitry of *Drosophila*. *Trends in Neurosciences*, 31(10):512–520, 2008.
- [66] H. OTSUNA and K. ITO. Systematic analysis of the visual projection neurons of *Drosophila melanogaster*. i. lobula-specific pathways. *Journal of Comparative Neurology*, 497(6):928–958, 2006.
- [67] H. PENG. Bioimage informatics: a new area of engineering biology. *Bioinformatics*, 24:1827–1836, 2008.
- [68] H. PENG, P. CHUNG, F. LONG, L. QU, A. JENETT, A.M. SEEDS, E.W. MYERS, and J.H. SIMPSON. BrainAligner: 3D registration atlases of *Drosophila* brains. *Nature methods*, 8(6):493–498, 2011.
- [69] A. PENTLAND, R.W. PICARD, and S. SCLAROFF. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision (IJCV)*, 18(3):233–254, 1996.
- [70] J.M.S. PREWITT. *Object enhancement and extraction*, volume 75. Academic Press, New York, 1970.
- [71] A.C. RENCHER. *Methods of Multivariate Analysis*. John Wiley & Sons, New York, 2002.
- [72] L.G. ROBERTS. Machine perception of three-dimensional solids. Technical report, DTIC Document, 1963.
- [73] T. ROHLFING and C.R. MAURER. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*, 7(1):16–25, 2003.

- [74] Y. RUI, T.S. HUANG, and S.-F. CHANG. Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation (JVCIR)*, 10(1):39–62, 1999.
- [75] Y. RUI, T.S. HUANG, M. ORTEGA, and S. MEHROTRA. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.
- [76] T. RUSS, C. BOEHNEN, and T. PETERS. 3D face recognition using 3D alignment for PCA. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1391–1398. IEEE, 2006.
- [77] T. SAKAI. New performance metrics based on multigrade relevance: Their application to question answering. In *Proceedings of NTCIR-4*, 2004.
- [78] R. SCHETTINI, G. CIOCCA, and S. ZUFFI. A survey of methods for colour image indexing and retrieval in image databases. *Color Imaging Science: Exploiting Digital Media*, pages 183–211, 2001.
- [79] J. SCHNEIDER and A.W. MOORE. *A locally weighted learning tutorial using vizier 1.0*. Carnegie Mellon University, the Robotics Institute, 2000.
- [80] T. SHIMADA, K. KATO, A. KAMIKOUCHI, and K. ITO. Analysis of the distribution of the brain cells of the fruit fly by an automatic cell counting algorithm. *Physica A: Statistical Mechanics and its Applications*, 350(1):144–149, 2005.
- [81] C.-R. SHYU, C.E. BRODLEY, A.C. KAK, A. KOSAKA, A.M. AISEN, and L.S. BRODERICK. ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. *Computer Vision and Image Understanding (CVIU)*, 75(1):111–132, 1999.
- [82] S.F. SILVA, M.X. RIBEIRO, J.E.S. BATISTA NETO, C. TRAINA JR., and A.J.M. TRAINA. Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decision Support Systems*, 51(1):810–820, 2011.
- [83] U. SINHA and H. KANGARLOO. Principal component analysis for content-based image retrieval. *Radiographics*, 22(5):1271–1289, 2002.
- [84] A.W.M. SMEULDERS, M. WORRING, S. SANTINI, A. GUPTA, and R. JAIN. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(12):1349–1380, 2000.
- [85] J.R. SMITH and S.-F. CHANG. Automated binary texture feature sets for image retrieval. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 2239–2242. IEEE, 1996.
- [86] J.R. SMITH and S.-F. CHANG. Tools and techniques for color image retrieval. In *Electronic Imaging: Science & Technology*, pages 426–437. International Society for Optics and Photonics, 1996.

- [87] Y. SONG, W. CAI, and D. FENG. Hierarchical spatial matching for medical image retrieval. In *Proceedings of the 2011 International ACM Workshop on Medical Multimedia Analysis and Retrieval (MMAR)*, pages 1–6. ACM, 2011.
- [88] M. ŠONKA, V. HLAVÁČ, and R. BOYLE. *Image processing, analysis, and machine vision*. Chapman & Hall Computing, 1993.
- [89] D. ST. JOHNSTON. The art and design of genetic screens: *Drosophila melanogaster*. *Nature Reviews Genetics*, 3(3):176–188, 2002.
- [90] M.A. STRICKER and M. ORENGO. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III*, pages 381–392. International Society for Optics and Photonics, 1995.
- [91] L.T. SU. The relevance of recall and precision in user evaluation. *Journal of the American Society for Information Science*, 45(3):207–217, 1994.
- [92] M.J. SWAIN and D.H. BALLARD. Color indexing. *International Journal of Computer Vision (IJCV)*, 7(1):11–32, 1991.
- [93] H.D. TAGARE, C.C. JAFFE, and J. DUNCAN. Medical image databases: A content-based retrieval approach. *Journal of the American Medical Informatics Association (JAMIA)*, 4(3):184–198, 1997.
- [94] J. TALAIRACH and P. TOURNOUX. *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. Thieme, 1988.
- [95] H. TAMURA, S. MORI, and T. YAMAWAKI. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics (SMC)*, 8(6):460–473, 1978.
- [96] M. TKALCIC and J.F. TASIC. Colour spaces: perceptual, historical and applicational background. In *The IEEE Region 8 EUROCON 2003. Computer as a Tool*, volume 1, pages 304–308. IEEE, 2003.
- [97] C. TOMASI and T. KANADE. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ., 1991.
- [98] M.A. TURK and A.P. PENTLAND. Face recognition using eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–591. IEEE, 1991.
- [99] N. TZOURIO-MAZOYER, B. LANDEAU, D. PAPATHANASSIOU, F. CRIVELLO, O. ETARD, N. DELCROIX, B. MAZOYER, and M. JOLIOT. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

- [100] D. UNAY and A. EKIN. Intensity versus texture for medical image search and retrieval. In *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 241–244. IEEE, 2008.
- [101] D. UNAY, A. EKIN, and R.S. JASINSCHI. Local structure-based region-of-interest retrieval in brain MR images. *IEEE Transactions on Information Technology in Biomedicine*, 14(4):897–903, 2010.
- [102] J.W. WALLIS, T.R. MILLER, C.A. LERNER, and E.C. KLEERUP. Three-dimensional display in nuclear medicine. *IEEE Transactions on Medical Imaging (T-MI)*, 8(4):297–230, 1989.
- [103] X.-Y. WANG, Y.-J. YU, and H.-Y. YANG. An effective image retrieval scheme using color, texture and shape features. *Computer Standards & Interfaces*, 33(1):59–68, 2011.
- [104] C. WEI, C. LI, and R. WILSON. *A content-based approach to medical image database retrieval*. Database Modeling for Industrial Data Management, 2006.
- [105] B. WIPFLER, K. SCHNEEBERG, A. LÖFFLER, F. HÜNEFELD, R. MEIER, and R.G. BEUTEL. The skeletomuscular system of the larva of *Drosophila melanogaster* (Drosophilidae, Diptera) – a contribution to the morphology of a model organism. *Arthropod Structure & Development*, 42(1):47–68, 2013.
- [106] W.S. YAMBOR, B.A. DRAPER, and J.R. BEVERIDGE. Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures. *Empirical Evaluation Methods in Computer Vision*, pages 1–15, 2002.
- [107] J. YANG, D. ZHANG, A.F. FRANGI, and J. YANG. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 26(1):131–137, 2004.
- [108] J. YANG and S.-J. ZHU. Narrowing semantic gap in content-based image retrieval. In *International Conference on Computer Distributed Control and Intelligent Environmental Monitoring (CDCIEM)*, pages 433–438. IEEE, 2012.
- [109] C.T. ZAHN and R.Z. ROSKIES. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers (CT)*, 100(3):269–281, 1972.
- [110] B. ZHANG and S.N. SRIHARI. Properties of binary vector dissimilarity measures. In *Proceedings of International Conference on Computer Vision, Pattern Recognition, and Image Processing (CVPRIP)*, volume 1, 2003.
- [111] D. ZHANG and G. LU. Shape-based image retrieval using generic Fourier descriptor. *Signal Processing: Image Communication*, 17(10):825–848, 2002.
- [112] X.S. ZHOU and T.S. HUANG. Relevance feedback in image retrieval: A comprehensive review. *Multimedia systems*, 8(6):536–544, 2003.