Joachim Hornegger, Kurt Höller,
Martin Spiegel, Hans-Peter Niedermeier
(Eds.)

HSS-Cooperation Seminar

"Engineering Science Contributions for a more efficient Healthcare System"

Proceedings of the
HSS-Kooperationsseminar
"Ingenieurwissenschaftliche Beiträge
für ein leistungsfähigeres Gesundheitssystem"
Kloster Banz, Bavaria, April 11 to 13, 2010

Friedrich-Alexander-Universität
Erlangen-Nürnberg
TECHNISCHE FAKULTÄT

Volume Editors:

Joachim Hornegger
Kurt Höller
Martin Spiegel
Erlangen, Germany

Hans-Peter Niedermeier
Munich, Germany

Technical Support:
Eva Kollorz
Arne Militzer
Sebastian Bauer
Jan Paulus
Alexander Brost

Printed in Erlangen, Germany

ISBN 978-3-00-031916-7

Preface

Computer science is a major driver of innovation in medical and healthcare engineering. It is impressive to see that software updates providing additional functionality in highly complex medical systems are released nowadays within innovation cycles of less than a year. In particular pattern recognition technology speeds innovation in many medical applications and allows for the efficient processing, analysis and visualization of large-scale medical data. Main drivers for competitive research and the rapid transfer of research results to industry are considered to be

- excellently qualified, highly motivated, and fair-minded collaborating scientists,
- interdisciplinary cooperations, and
- the existence and maintenance of a network with excellent researchers and market leaders in industry.

The present proceedings impressively show that these driving forces do exist at the Lehrstuhl für Mustererkennung (LME) at the University of Erlangen-Nuremberg. The following pages give an overview of major research results achieved in 2010 by the members of LME. The contributions are clustered in groups. The reader will hopefully find inspiring research projects in the exciting fields of computer vision, segmentation and navigation, signal processing, CT reconstruction, eye imaging, time-of-flight imaging, and image registration.

The success of this workshop is the success of many people that helped us to organize the meeting. At this point we want to thank all the restless people who contributed to the organization of the workshop and in particular we acknowledge the Hanns-Seidel-Foundation which was an excellent host at Kloster Banz.

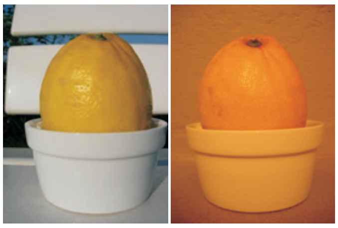
July 2010

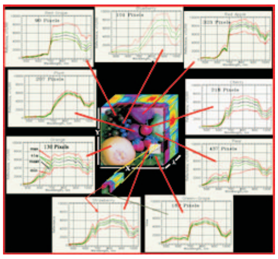
Joachim Hornegger
Workshop Initiator

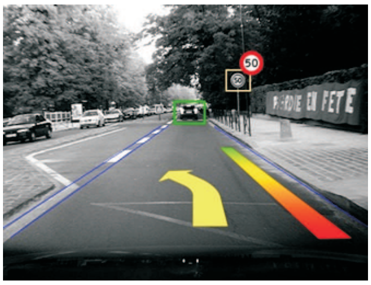
Table of Contents

1

Computer Vision







A New Benchmark for Evaluations in Image Forensics

J. Jordan, C. Riess, and E. Angelopoulou

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

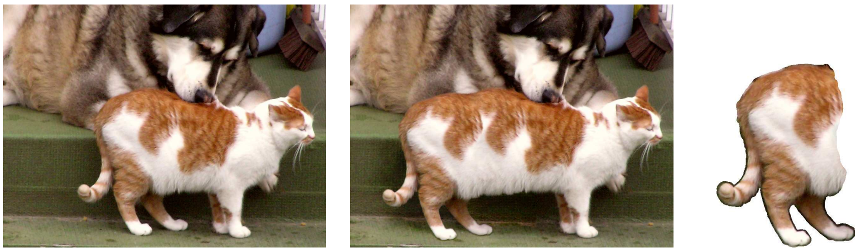
Abstract. The lack of common data, ground truth and standardized evaluation metrics hinders the comparative analysis of image forensics algorithms for the detection of tampering artifacts. With that in mind, we create a large, flexible database, that facilitates their performance evaluation. We introduce a protocol for testing different methodologies under various common assumptions and challenges in image forensics.

Image forensics is a relatively new research direction that aims to assess the authenticity and originality of digital images. In blind image forensics we concentrate on images from unsecured sources. This area quickly evolved in a very active community in the last few years. A diverse set of techniques has been developed, including (among others) the exploitation of camera hardware properties, statistical approaches, detection of inconsistencies in the scene and image processing artifacts. Please refer to [1,2] for an overview.

A large portion of the published work concentrates on the class of *copy-move forgery*. This manipulation scenario features a simple workflow by the forger, and prominent real world examples exist. A manipulation is concentrated around the duplication of regions from the same image to either add information to an image, or hide it. While there are many publications that propose a wide range of features to classify duplicated regions, little is known about their detection performances. Existing image databases in the field either consist of authentic images (e.g. to test algorithms exploiting camera properties), or provide spliced images, that were created automatically. These do not suffice to test for real-world applicability or discriminate different forgery scenarios.

In our work, we created a dataset of 96 image manipulations that are well controlled according to specific modifiers determining detection difficulty. These are: relative size of the forgery, texture of copied regions and number of manipulations per image. Also, the images come from four different sources (the web and three single camera sources, one of which providing RAW images). For each forged picture we provide the original image, the tampered image and pixel-wise labeling. We also provide data necessary to recreate the manipulation. Figure 1 gives an example manipulation from the dataset, as well as its corresponding image cutouts and ground-truth maps. The source region of the copy (Fig. 1c) is vital for recreating the manipulation. This allows testing for algorithmic performance under different circumstances, e.g. noise or compression introduced during the copy. For example, to test detection of double JPEG compression, JPEG compression may be introduced only on the region to be copied, resulting in the same effect as if the information introduced to the image was drawn from another compressed image source. On the other hand, copy-move forgery detection (CMFD) methods can be evaluated using a specific ground-truth map (Fig. 1e).

Unlike previously available data, the supply of pixel-wise ground-truth segmentation allows a fine grained quantitative performance evaluation. We propose the use of per-pixel false negative/positive measurements as a simple, yet effective metric for detection performance. Early tests showed that our benchmark is capable of revealing specific strengths and weaknesses of state-of-the-art CMFD methods.



(a) Original Image (b) Manipulated Image (c) Cutout

(d) Ground-truth Insert, Edit (e) Ground-truth for CMFD (f) Visible Copy

Fig. 1. Example data as available in our manipulation dataset. The original and tampered image are shown in **(a)** and **(b)**, respectively. In **(c)** the original region used in the copying process is shown, that may also be needed to recreate the forgery, while **(f)** shows the duplicated pixels present in the manipulated image. In **(d)** and **(e)** ground-truth masks are depicted that can be used to evaluate different detection scenarios.

References

1. Sencar, H. and Memon, N. Overview of State-of-the-art in Digital Image Forensics Algorithms, Architectures and Information Systems Security pp. 325-344, Nov. 2008.
2. Farid, H. A Survey of Image Forgery Detection Signal Processing Magazine vol. 26, no. 2, pp. 16-25, Mar. 2009.

On Physics-based Illuminant Color Estimation

C. Riess, J. Jordan, and E. Angelopoulou

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. We present our recent work on illuminant color estimation. We adopted the inverse-intensity chromaticity space by Tan et al. and examined its behavior and various assumptions with the goal of extending it for improved illuminant color estimation on natural images.

The color of an illuminant heavily influences the appearance of digitally captured objects. Color features for high level computer vision task like recognition, tracking and detection of objects are therefore inherently dependent on the illumination conditions. The problem of neutralizing the effects of the illumination color is known as color constancy. It is typically assumed that the knowledge of the color of the dominant light source suffices to correct the image for the most disturbing illumination effects. One physics-based approach is to use the dichromatic reflectance model [4] to express the relationship between incident light, surface material and geometry. In our group, we examine ways to relax widely-used, yet constraining, assumptions in order to make physics-based methodologies applicable to natural images. As a first result, we present a novel illumination estimation method that takes different votes on the illuminant color and makes a consensus decision afterwards.

1 Physics-Based Illuminant Color Estimation as an Image Semantics Clue

We investigated the behavior of the inverse-intensity chromaticity (IIC) illuminant estimation by Tan et al. [5] on real-world images downloaded from the web. The method consists of two parts: first, a specularity segmentation, and second, the estimation of the illuminant color from the segmented pixels. We found that the specularity segmentation might complicate the estimation. Furthermore, we did experiments towards a confidence measure for the illuminant color estimation. In IIC-space, the illuminant color is determined as the prevalent y-axis (ordinate-) intercept of intersecting lines which are derived from the distribution of specular pixels in this space. Tan et al. proposed to find this y-intercept via Hough transform. We examined the distributions in Hough space as a means of automated self-assessment. Indeed, the shape of the prominent mode in this distribution seems to provide information appears on how well the specularity-segmented pixels fit the underlying physics-based model [2].

In future work, we aim to expand on the analysis of these distributions towards a more robust confidence measure for this illumination estimation in IIC space.

2 A Common Framework for Ambient Illumination in the Dichromatic Reflectance Model

The dichromatic reflectance model is the core of most physics-based algorithms for illumination color estimation. Maxwell et al. [1] recently proposed an extension of this model, the Bi-Illuminant Dichromatic Reflectance (BIDR) model. This extension contains a term for a second illuminant. Interestingly, this extension can be seen as a theoretical basis for several algorithms that deal with two illuminant cases (for example, see Fig. 1). Most prominently, shadow detection algorithms often model ambient illumination as a second global light source. In our paper [3], we show connections between different shadow detection methods via the Bi-Illuminant Dichromatic Reflectance Model. Our ultimate goal is to find a way to exploit the BIDR model directly, in order to obtain more general methods for handling two-illuminant scenarios.

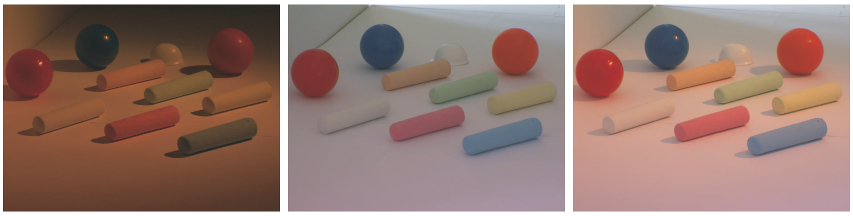


Fig. 1. A scene with illumination that comes from only one direct source (left), only ambient illumination (middle), and both ambient and direct sources (right).

References

1. Maxwell, Bruce, Friedhoff, Richard and Smith, Casey, A Bi-Illuminant Dichromatic Reflection Model for Understanding Images, IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA, June 2008.
2. Riess, Christian and Angelopoulou, Elli, Physics-Based Illuminant Color Estimation as an Image Semantics Clue, IEEE International Conference on Image Processing, Nov. 2009, Cairo, Egypt, pp. 689-692.
3. Riess, Christian, Jordan, Johannes and Angelopoulou, Elli, A Common Framework for Ambient Illumination in the Dichromatic Reflectance Model, IEEE Workshop on Color and Reflectance in Imaging and Computer Vision, Oct. 2009, Kyoto, Japan, pp. 1939-1946.
4. Shafer, Steven, Using Color to Separate Reflection Components, Color Research and Applications, vol. 10, no. 4, August 1985, pp. 210-218.
5. Tan, Robby, Nishino, Ko and Ikeuchi, Katsushi, Color Constancy through Inverse-Intensity Chromaticity Space, Journal of the Optical Society of America A, vol. 21, no. 3, March 2004, pp. 321-334.

Skin Segmentation and the Compensation of Illumination Color

E. Eibenberger[1,2,3] and E. Angelopoulou[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] International Max Planck Research School for Optics and Imaging
[3] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany

Abstract. Computer vision applications addressing the detection and recognition of humans often require the segmentation of skin regions as a pre-processing step. Compared to structural information, color cues have the advantage to be relatively invariant regarding scale and orientation. However, as the appearance of the skin color is dependent on the illumination color, skin segmentation approaches have to compensate for these variations. In this paper, a small review of skin segmentation approaches is given, whereas the focus is on the adaptation to the illuminant color.

Many computer vision applications dealing with the detection and recognition of humans and their activities require the segmentation of skin regions as a pre-processing step. Examples for such systems can be found in the fields of surveillance, human computer interaction and face recognition [2]. Although, color cues have several advantages, the utilization of color information is often a challenging task, as in images the appearance of the skin color is dependent on different factors, such as illumination, ethnicity, camera characteristics etc. Figure 1 shows examples of the appearance change due to illumination variations.

The reviews by Kakumanu *et al.* [2] and Vezhnevets *et al.* [7] illustrate that an extensive body of work on skin segmentation exists. A small set of techniques applies explicitly defined color thresholds, which are empirically determined. These approaches are often too inflexible and cannot easily adapt to illumination variations. The largest group of approaches uses machine learning techniques to estimate the probability of skin pixels. Although these approaches are more flexible, their degree of illumination adaptability is sensitive to the training set. Illuminant colors outside the training data can not be handled well.

Some methods exist that are explicitly designed for handling illumination variations. They can be separated into two categories: a) color-constancy techniques and b) dynamic adaptation techniques. For the first group an estimate of the illuminant color is used to create a new normalized image, where colors are independent of the incident illuminant. In the second group the skin-color model is transformed according to the measured variations in the image colors [2]. Some illuminant adaptive methods have restrictive requirements, like the method developed by Do *et al.* [1], which is constrained to frontal face images. Other techniques depend on training data, like the approach presented by Sun [6], and do not explicitly use the illuminant color.

Another group of physics-based approaches analyzes the change of the skin locus [4] under varying illumination. With respect to normalized RGB space, Martinkauppi *et al.* [3] showed that for different illuminant colors the skin pixels lie inside a region which can be bounded by two quadratic functions. Furthermore, Störring *et al.* [5] showed that for a known camera and known color temperature of the illuminant the search area for skin color can be adjusted accordingly. The normalized RGB space has also been used for illuminant-invariant tracking and locating image-specific skin clusters. All prior research on skin locus uses the normalized RGB space which is mathematically more complex and which often makes the explicit knowledge of the illuminant color temperature mandatory. However, color temperature is more difficult to extract than the illuminant color.

The goal of this project is the development of a robust illumination-adaptive skin segmentation approach. Thereby, the focus is on physics-based skin models. Particular attention should be paid to the reliable detection of different ethnicities and the handling of atypical illumination colors. For quantitative evaluations, a database with ground truth segmentations has to be created. The database needs to contain skin and non-skin objects and a variety of different ethnicities and illumination colors.

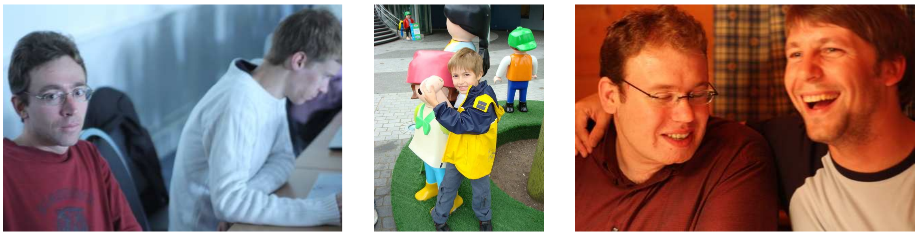


Fig. 1. Examples of skin under different lighting conditions.

Acknowledgments

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative. Eva Eibenberger is supported by the International Max Planck Research School for Optics and Imaging.

References

1. Do, H.-C.; You, J.-Y.; Chien, S.-I.: Skin Color Detection through Estimation and Conversion of Illuminant Color Under Various Illuminations. IEEE Transactions on Consumer Electronics, Vol. 5, No. 3, pp. 1103–1108, May 2007.
2. Kakumanu, P.; Makrogiannis, S.; Bourbakis, N. G.: A Survey of Skin-Color Modeling and Detection Methods. Pattern Recognition, Vol. 40, No. 3, pp. 1106–1122, March 2007.
3. Martinkauppi, J. B.; Soriano, M. N.; Laaksonen, M. H.: Behavior of Skin Color under Varying Illumination Seen by Different Cameras at Different Color Spaces. Proceedings of SPIE, Machine Vision Applications in Industrial Inspection IX, pp. 102–113, San Jose, USA, January 2001.
4. Störring, M.; Andersen, H. J.; Granum, E.: Skin Colour Detection under Changing Lighting Conditions. Symposium on Intelligent Robotics Systems, pp. 187–195, Coimbre, Portugal, July 1999.
5. Störring, M.; Andersen, H. J.; Granum, E.: Physics-Based Modelling of Human Skin Colour under Mixed Illuminants. Robotics and Autonomous Systems, Vol. 35, No. 3-4, pp. 131–142, June 2001.
6. Sun, H.-M.: Skin Detection for Single Images Using Dynamic Skin Color Modeling. Pattern Recognition, Vol. 43, No. 4, pp. 1413–1420, April 2010.
7. Vezhnevets, V.; Sazonov, V.; Andreeva, A.: A Survey on Pixel-Based Skin Color Detection Techniques. International Conference GraphiCon, pp. 85–92, Moscow, Russia, September 2003.

Body Sensor Networks in Sports and Healthcare

P. Kugler[1], U. Jensen[1], and H. Schlarb[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] adidas innovation team (a.i.t.), adidas AG, Germany

Abstract. Body sensor networks are networks of small sensors that can be directly worn on the body. The goal of this project is to explore the application of methods from pattern recognition to biomechanical and medical problems using such body sensors. A main aspect is hereby he investigation and development of embedded support systems in the area of sports and health care.

1 Body Sensor Networks

Embedded systems are omnipresent in many aspects of daily live, e.g. in modern cars, mobile phones or medical implants. Due to the increasing miniaturization such embedded sensor systems can be built in very small packages containing various sensors, a wireless transmitter and a microcontroller.

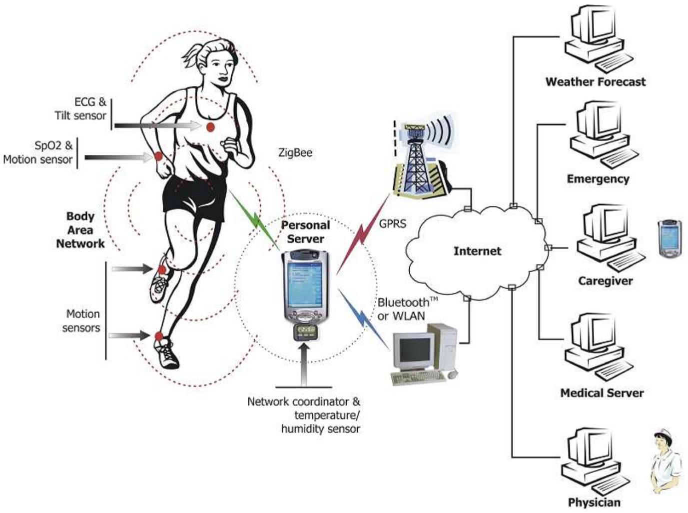


Fig. 1. Schematic overview for a typical Body Sensor Network (BSN). Various wearable sensors transmit information to a smartphone, which aggregates the data, performs classification and gives feedback to the user. Additionally the data can be transmitted to a central data storage and to health care providers or trainers. (Image from [2])

Today there are sensor nodes availiable that can measure heart rate, muscle activity or motion, using various sensors like accelerometers, magnetometers or micromechanical gyroscopes. These nodes individually record and process the sensor data and then transmit it via wireless to a local central node (e.g. a PC or a smart phone), which further analyses the data, provides feedback to the user or transmits the data to a remote server [1].

2 Applications

Within the scope of this project different applications of body sensor networks shall be explored and improved. This includes medical applications [2], as well as novel systems in the digital sports domain. With such small systems it is imaginable to directly integrate such sensors into intelligent sports equipment, where they might directly support or help athletes. Today it is already possible e.g. to measure the heart rate using sensors directly embedded into a shirt and provide this information to the athlete. Another example for such an intelligent sports equipment is the adidas 1 shoe, which is able to adapt the shoe cushioning automatically to changing surface conditions and the current needs of the athlete [3].

Fig. 2. The adidas 1 intelligent running shoe. It consists of a magnetic sensor element that can measure heel compression, a embedded microcontroller that performs classification and a motor which can adjust the stiffness of the sole.

3 Project Outlook

To achieve this goal, methods from pattern recognition will be employed to develop novel ways of processing and classifying data from wearable sensor networks. The focus will be on the analysis of human movements using kinematic and kinetic parameters, as well as the assessment of muscle activity using electromyography (EMG). Subsequently the integration of different sensors into sportswear and potential embedded processing methods of the measured signals will be studied. The concepts from pattern recognition will then be used on the embedded systems to extract information of interest from the data and provide it to the user. Examples for such information include the degree of fatigue or assessing the health aspect of certain motion sequences, e.g. running. This will lead to the invention of lightweight, mobile support systems, which can directly support, guide or motivate athletes and might help to prevent possible injuries.

4 Acknowledgements

We acknowledge the financial support from adidas AG, Germany and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie.

References

1. L. F. Akyildiz, W. Sun, Y. Sankarasubramaniam, E. Cayirci, “A Survey on Sensor Networks” *IEEE Communications Magazine* 51(4), 102–114, 2002.
2. E. Jovanov, A. Milenkovic, C. Otto, P. C. de Groen, “A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation”, *Journal of NeuroEngineering and Rehabilitation* 2(6), 2005
3. B. Eskofier, M. Oleson, C. DiBenedetto, J. Hornegger, “Embedded Surface Classification in Digital Sports” *Pattern Recognition Letters* 30(16), 1448–1456, 2009.

A Discussion on the Sample Positioning Problem of Particle Filters

A. Linarth[1,2], J. Jordan[1], P. Nordhus[1], and E. Angelopoulou[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Elektrobit Automotive GmbH, Erlangen, Germany

Abstract. In this work we discuss the problem of efficiently positioning samples on Particle Filters. The concentration of particles resulting from the resampling step of these filters very often leads to under- and over-sampling problems. An optimal compromise between computational effort and precision is directly dependent on the amount of samples and their position in space. In our approach, we tackle these problems by aligning the samples to an adaptive grid. Particles are deterministically allocated, while their movement is still determined by the system dynamics and standard resampling strategies. Concentration of particles, as well as volume coverage, is intuitively controlled by limiting the voxel sizes. The benefits of such an approach are discussed in this work.

The overall objective of tracking is to estimate a state $\mathbf{x}_t$ given all the measurements $\mathbf{z}_{1:t}$ up to a given time t. In other words, the goal is to construct the probability density function (pdf) $p(\mathbf{x}_t \mid \mathbf{z}_{1:t})$. A Particle Filter (PF) recursively estimates this function by alternating two steps: the *prediction* step estimates the new state given the dynamic model of the system and the previous estimations; in the *update* step, the posterior probability is calculated based on the predicted state density and the current measures. These steps are recursively applied over discrete samples, representing the pdf. The true pdf is approached when the number of samples approximates infinity. Since this is not feasible in practice, the idea is to move particles with low likelihood in one iteration, to regions with higher probability in the next iteration. As the result, a good estimation of the pdf may be reached with a relative small number of samples. Despite the achieved benefits, this strategy presents also some side effects. As samples converge to a given region, large areas of the parameter space start lacking coverage. In tracking, new modes may be completely lost. Furthermore, over-sampling may also occur in some regions, which leads to a waste of computational resources. These problems can be seen in Fig. 1, in a lane boundary tracking application. The left hand side image shows three lane boundaries being followed. Due to the low contrast on the shadow areas, all particles tend to move to good visible markers in the upcoming frames. The result is shown in the right hand side image. The system would not be able to recover the middle and left boundaries due to the lack of samples, while an excessive amount of particles would be allocated in tracking the right boundary. Previous approaches

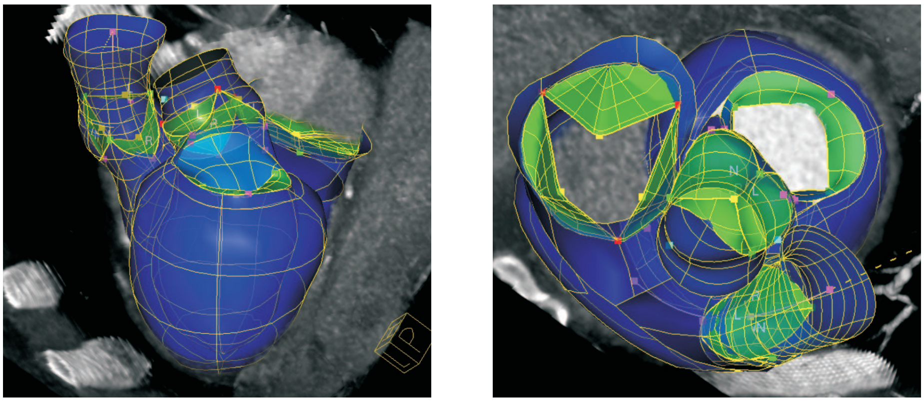
Fig. 1. Mode Miss Example in Lane Boundary Tracking

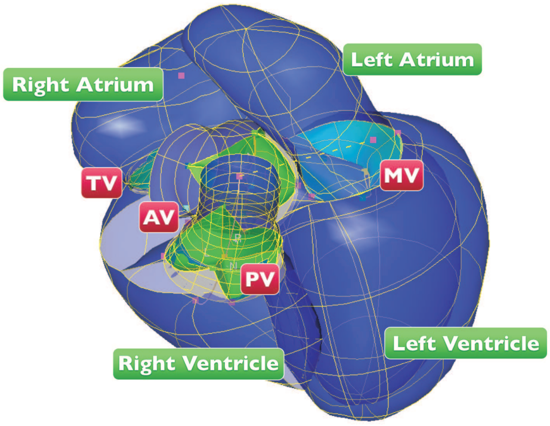
tackling the positioning problem focus on deriving a better proposition for the resampling step. For a work in this direction the reader is referred to [1]. Attempts to adapt the number of particles cal also be found in the literature (see [2] for a good reference). Unlike previous systems, our approach tackles the positioning problem by simply aligning particles to an adaptive grid. The first expected advantage is on the evaluation of the likelihoods (i.e. update step). It can be accomplished departing from the parameter estimate, like in standard particle filters. However, due to the deterministic positions, it may be more efficient, for some applications, to find which particles are affected by a given input sample (i.e. like in Hough transform). The adaptation of the grid is achieved through the same principles applied in sequential Monte Carlo methods. Each voxel, represented by a particle in its center, is resampled and a new state is predicted. A new grid is then calculated as follows: starting from a representation with the maximum voxel size, assuring coverage, voxels are subdivided within an octree according to the amount of particles present in its volume. Over-sampling is then easily avoided by limiting the depth of subdivisions. The grid representation also allows different resampling approaches. The volume of the voxel is associated with the information provided by the neighbors. It, thus, allows particles to be positioned in more restricted regions. Calculation of discrete marginals are also made possible and that may lead to different resampling strategies. Clustering and mode finding can also be achieved more efficiently through the use of a grid representation. We are currently evaluating our method, to quantitatively and qualitatively support the proposed approach. Further work on resampling, temporal constraints, as well as on the adaptation of particle set and voxel sizes, are planned for the near future. Early results already showed promising results for some subset of applications.

References

1. Schön, T. ; Gustafsson, F. ; Nordlund, P.-J. : Marginalized Particle Filters for Mixed Linear/Nonlinear State-space Models IEEE Transactions on Signal Processing V. 53, Nr. 7, Juli 2005, 2279-2289
2. Fox, D.: Adapting the Sample Size in Particle Filters Through KLD-Sampling International Journal of Robotics Research V. 22, Nr. 12, Dec. 2003, 985-1003

Segmentation and Navigation





Personalized Pulmonary Trunk Modeling for Intervention Planning and Valve Assessment Estimated from CT Data

D. Vitanovski and J. Hornegger

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Pulmonary valve disease affects a significant portion of the global population and often occurs in combination with other heart dysfunctions. Emerging interventional methods enable percutaneous pulmonary valve implantation, which constitute an alternative to open heart surgery. As minimal invasive procedures become common practice, imaging and non-invasive assessment techniques turn into key clinical tools. In this paper, we propose a novel approach for intervention planning as well as morphological and functional quantification of the pulmonary trunk and valve. An abstraction of the anatomic structures is represented through a four-dimensional, physiological model able to capture large pathological variation. A hierarchical estimation, based on robust learning methods, is applied to identify the patient-specific model parameters from volumetric CT scans. The algorithm involves detection of piecewise affine parameters, fast centre-line computation and local surface delineation. The estimated personalized model enables for efficient and precise quantification of function and morphology. Experiments performed on 50 cardiac computer tomography sequences demonstrated the average speed of 202 seconds and accuracy of 2.2mm for the proposed approach.

1 Introduction

Valvular Heart Disease (VHD) is an important cardiac disorder that affects a large number of patients and often requires operative intervention. Until recently, the pulmonary valve replacement has been performed surgically on open heart, with all associated risks. Percutaneous pulmonary valve implantation (PPVI) is a new developed technique for transcatheter placement of a valve stent. The main difficulties of PPVI are: the assessment of the pulmonary trunk and the right ventricle outflow track (RVOT < 22mm) before treatment, the classification of patients suitable for the procedure and identification of the exact location for anchoring the stent. Hence, precise assessment of the morphology and dynamics is crucial for the pre-procedural planning and successful intervention of PPVI.

2 Model Representation

Initially, the pulmonary trunk is represented by five anatomical landmarks: Trigone, RVOT, Main-Bifurcation, Left-Bifurcation and Right-Bifurcation.

Constrained by the landmarks, the artery center line form the second abstraction layer (see Fig. 1(b)). The Main Center Line is bounded by the RVOT and Bifurcation. The Left and the Right Center Lines extend from the main bifurcation landmark to the Left-Bifurcation and Right-Bifurcation landmark, respectively.

The main pulmonary artery along with its bifurcations is modeled as parametric closed surface similar to a deformed cylinder. NURBS are used as the parametric representation due to the compact mathematical formulation and convenient properties.

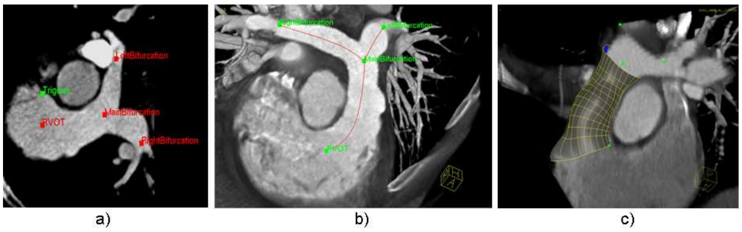


Fig. 1. Model representation

3 Model Estimation

The segmentation task is performed by fitting a 4D generic model to the patient specific image data [2]. The proposed workflow, from modeling to quantification, involves three stages. In the first place, a generic model of the pulmonary artery (see Section 2) is computed from a comprehensive training set. This is fitted to the available volumetric scans using robust machine-learning methods in order to obtain a patient specific physiological model. Finally, the morphology and function of the artery is efficiently quantified from the personalized abstract representation.

4 Results

The performance of the proposed approaches is evaluated on 50 4D CT (503 volumes) studies, associate with a manual annotation considered to be the ground truth. Included data of patients with various ages (from 5 months infant to 65 years old adult) lead to significant differences in morphology and pathology of the pulmonary trunk. CT data is acquired using different protocols, resulting in volume sizes between $153 \times 153 \times 80$ and $512 \times 512 \times 350$ and voxel resolutions from 0.28mm to 2.0mm. The accuracy of our detection framework is evaluated by using a three-fold cross validation and an average precision of 2.2 mm and computation time of 202 sec on a standard 2.0GHz Dual Core PC.

References

1. Vitanovski, D., Ionasec, R.I., Georgescu, B., Huber, M., Taylor, A.M., Hornegger, J., Comaniciu, D.: Personalized pulmonary trunk modeling for intervention planning and valve assessment estimated from ct data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), London, UK (2009), pp. 17-25

Learning Discriminative Distance Functions for Valve Retrieval and Improved Decision Support in Valvular Heart Disease

I. Voigt and D. Vitanovski

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Disorders of the heart valves constitute a considerable health problem and often require surgical intervention. Recently various approaches were published seeking to overcome the shortcomings of current clinical practice, that still relies on manually performed measurements for performance assessment. Clinical decisions are still based on generic information from clinical guidelines and publications and personal experience of clinicians. We present a framework for retrieval and decision support using learning based discriminative distance functions and visualization of patient similarity with relative neighborhood graphs based on shape and derived features. We considered two learning based techniques, namely learning from equivalence constraints and the intrinsic Random Forest distance. The generic approach enables for learning arbitrary user-defined concepts of similarity depending on the application. This is demonstrated with the proposed applications, including automated diagnosis and interventional suitability classification, where classification rates of up to 88.9% and 85.9% could be observed on a set of valve models from 288 and 102 patients respectively.

1 Learning discriminative distance functions

Generally we address two tasks: retrieval of similar cases using a learned distance function, which measures the similarity of two particular shapes, and a binary classification problem, based on geometric valve models and derived features as described in the previous section. The second task can actually be formulated using the first[3,5]. So instead of learning a classifier directly, we choose learning distance functions instead and divide the learning process into two sequential steps, i.e. distance learning followed by classification or clustering, where each step requires search in a less complex functional space than in the immediate learning[3]. The labels for classification $y \in \{-1, +1\}$ are chosen depending on the application. Each case is represented with a parameter vector C containing the N 3D points $\mathbf{P_i}$ of the respective model and a set of M additional features F_i derived from the model, depending on the application:

$$\hat{y} = \underbrace{\text{argmax}}_{y \in \{-1,+1\}} (p(y|C)) \quad C = ((P_1^x, P_1^y, P_1^z), \cdots, (P_N^x, P_N^y, P_N^z), F_1, \cdots, F_M) \tag{1}$$

2 Application to heart valves

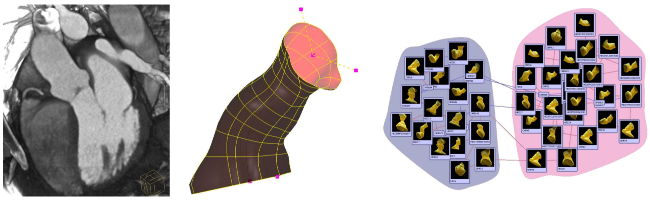


Fig. 1. Workflow of our framework for the application of PPVI selection.

The basic workflow within our framework is visualized in Fig. 1: A 4D acquisition (Fig. 2) is loaded into the modeling component, which then generates the personalized 4D models. These are passed on to the retrieval component, which finds similar valves and classifies them depending on the application (PPVI suitability in this case). A neighborhood graph is generated and presented to the user, who can inspect the similar valves and patient records to then find a diagnostic or treatment decision. Such a graph is an expressive way of presenting relevant information. Image thumbnails allow for easily identifying the different cases and their shape characteristics. From the graph, one may easily comprehend patient distribution according to the studied similarity context and see patient groupings, identify outliers, easy to classify cases and the borderline cases classification for which is likely to be uncertain. In our framework shape similarity can be defined arbitrarily and depending on the purpose and enables for different applications in field of decision support for VHD such as shape-based diagnosis and patient-suitability selection for particular interventions such as percutaneous pulmonary valve implantation.

Acknowledgements

Part of this work has been done in the framework of the EU project Health-e-Child (IST 2004-027749) and the MEDICO Project in the THESEUS Program, which is funded by the German Federal Ministry of Economics and Technology under the grant number 01MQ07016. The responsibility for this publication lies with the authors.

References

1. Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Hornegger, J., Navab, N., Comaniciu, D.: Personalized modeling and assessment of the aortic-mitral coupling from 4d tee and ct. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), London, UK (2009), pp. 767-775

2. Vitanovski, D., Ionasec, R.I., Georgescu, B., Huber, M., Taylor, A.M., Hornegger, J., Comaniciu, D.: Personalized pulmonary trunk modeling for intervention planning and valve assessment estimated from ct data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), London, UK (2009), pp. 17-25
3. Tsymbal, A., Zhou, S.K., Huber, M.: Neighborhood graph and learning discriminative distance functions for clinical decision support. In: Proc. IEEE International Conference on Engineering in Medicine and Biology (EMBC). (2009) pp. 5617-5620
4. Jaromczyk, J.W., Toussaint, G.T.: Relative neighborhood graphs and their relatives. In: Proc. IEEE. (1992) 1502–1517
5. Tsymbal, A., Huber, M., Zhou, S.K.: Discriminative distance functions and the patient neighborhood graph for clinical decision support. In: Advances in Computational Biology. Springer (2010)

Free-Breathing Cardiac MRI: Efficient and Robust Self Navigated Coronary MRI

D. Piccini[1], A. Littmann[2], and M. O. Zenge[2]

[1] Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander
University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] MR Applications and Workflow Development, Healthcare Sector, Siemens AG,
Allee am Röthelheimpark 2, 91052 Erlangen, Germany

Abstract. An efficient and robust approach for respiratory self-navigation is introduced in the field of coronary MRI. This approach features the uniform distribution and intrinsic interleaved arrangement of the projections of an innovative 3D spiral phyllotaxis trajectory and a new method for the suppression of the unwanted signal, based on the combination of local coil elements. During interleaved, ECG-triggered examination the relative displacement of the heart due to the respiratory motion is extracted directly from the SSFP echoes used for cardiac imaging. Thus, retrospective, rigid-body, motion correction is performed online.

1 Introduction

Respiratory motion is the major source of artifacts in ECG-triggered coronary MRI [1]. Free breathing techniques featuring hardware or software devices, such as respiratory belts or pencil beam navigators, can minimize respiratory artifacts and reduce the need for patient cooperation. These techniques, however, can be affected by hysteretic effects, non ideal positioning of the navigator and temporal delays. Irregularities in the breathing pattern usually lead to prolonged scan times and reduced acquisition efficiency, thus resulting in poor image quality. Self-navigated, free-breathing, whole heart 3D radial coronary MRI potentially overcomes these drawbacks. While the radial 3D trajectory proves to be robust against motion and radial undersampling, scan efficiency can be increased by the adaptation to self-navigation. Moreover user-dependency is reduced, since the preparation for locating individual anatomical details and the placement of the navigator are no longer needed.

2 Methods

Whole heart 3D radial acquisition was performed with a new interleaved 3D spiral phyllotaxis trajectory, recently described in [2]. The trajectory intrinsically features: 1) a uniform distribution of the projections and 2) reduction of the eddy current effects. The former property allows the application of a simple density compensation algorithm, while the latter leads to an improved image quality. An extra radial projection, oriented along the superior-inferior (SI) direction [3] was added at the beginning of each interleave and used for respiratory navigation. The contribution of the unwanted signal to the SI projection coming from external anatomical structures, such as arms and chest wall, was reduced using a new coil combination technique, based on the Siemens local array coil technology [4]. After normalization, a cross correlation algorithm was then applied in order to calculate the relative respiratory motion for each interleave. This value was directly used for in-line motion correction. The new approach was compared to a 3D radial navigator gated acquisition in 8 healthy volunteers.

3 Results and Outlook

The total acquisition time was reduced from an average of 15 minutes to 7 minutes. Robustness and efficiency of the new method were proved even in case of volunteers presenting a high variability in the breathing pattern. An improved image quality was achieved in all self-navigated acquisitions and an example is displayed in Fig. 1. More extensive clinical studies and further improvements in the motion correction algorithm are planned.

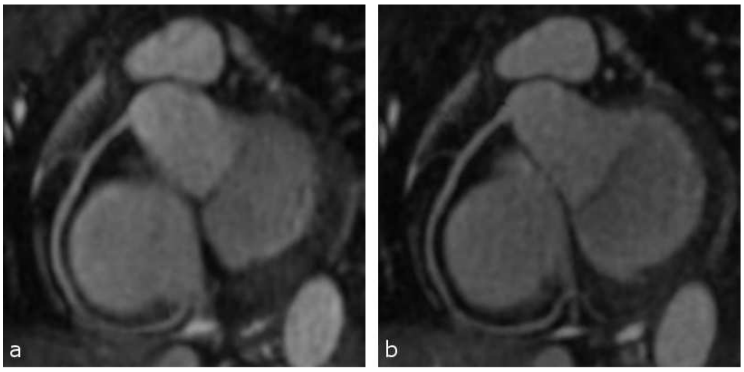


Fig. 1. Detail from a curved reformat of the right coronary artery (RCA) for one of the volunteers. The acquisition was performed with: (a) a 3D radial navigator gated technique, as reference, and (b) the new self-navigated technique. The acquisition time of (b) was halved and an improvement of the image quality can be noticed particularly in the distal part of the vessel.

References

1. McLeish K, Hill LGD, Atkinson D, Blackall JM and Razavi R: A study of the motion and deformation of the heart due to respiration. IEEE Transactions on Medical Imaging 21(9) (September 2002) 1142–1150

2. Piccini D, Littmann A, Schmitt P and Zenge MO: Spiral phyllotaxis: a better way to construct a 3D radial trajectory in MRI. Proceedings of the 18th Annual Meeting of ISMRM, Stockholm, Sweden (May 2010)
3. Stehning C, Börnet P, Nehrke K, Eggers H and Stuber M: Free-breathing whole-heart coronary MRA with 3D radial SSFP and self-navigated image reconstruction. Magnetic Resonance in Medicine 54 (Aug 2005) 476–480
4. Reykowski A and Blasche M: Mode matrix - A generalized signal combiner for parallel imaging arrays. Proceedings of the 12th Annual Meeting of ISMRM, Kyoto, Japan (May 2005) 1587

Advanced Real-Time Visualization and Monitoring of MR-Guided Thermal Ablations

E. Rothgang[1,2], W. D. Gilson[2], C. H. Lorenz[2] and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Center for Applied Medical Imaging, Siemens Corporate Research, Baltimore (MD), USA

Abstract. In modern oncology, thermal ablations are increasingly used as a regional treatment option to supplement systemic treatment strategies such as chemotherapy and immunotherapy. The goal of all thermal ablation procedures is to cause cell death of disease tissue while sparing adjacent healthy tissue.

1 Introduction

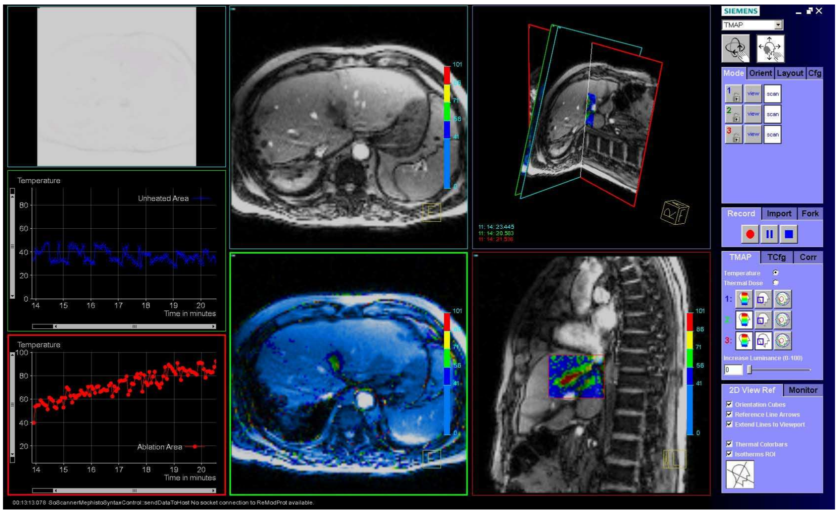
There is growing interest in thermal ablation therapies as minimally invasive alternatives to surgery, particular for patients that are poor surgical candidates. Real-time assessment of thermal damage is the key to therapeutic efficiency and safety of such procedures. Magnetic resonance thermometry is capable of monitoring the spatial distribution and temporal evolution of temperature changes during thermal ablations. In this work, we present an advanced monitoring system for MR-guided thermal ablations that includes multiplanar visualization, specialized overlay visualization methods, and additional methods for correcting errors resulting from magnetic field shifts and motion. The primary purpose of this work is to provide clinicians with an intuitive tool for accurately visualizing the progress of thermal treatment at the time of the procedure. Importantly, the system is designed to be independent of the heating source. The presented system is expected to be of great value not only to guide thermal procedures but also to further explore the relationship between temperature-time exposure and tissue damage. The software application has been validated with clinical data.

2 Methods

The presented thermotherapy guidance tool, referred to as TMAP, was implemented in C++ within the eXtensible Imaging Platform and integrated into the Interactive Front End (IFE) [2], a prototype for advanced visualization and real-time parameter control of image acquisition. Seamless integration with the IFE supports the thermal ablation workflow from placing the ablation device to online monitoring of the progress of ablation. The full application runs on an independent PC connected to the standard Siemens MR scanner network via

Ethernet and receives magnitude and phase images acquired by the MR system in real-time.

In our system, we used the temperature dependence of the proton resonance frequency (PRF) [1] for spatial and temporal temperature measurements during the intervention. The TMAP tool has been developed to be compatible with gradient echo, single shot EPI and segmented EPI PRF sequences. To provide an intuitive and clear user interface, the presented TMAP application features visualization of up to three multiplanar slices during the thermal ablation (Fig. 1). Systematic on-line quality control of temperature measurements is carried out online to ensure the reliability of the displayed temperature maps.



```
00:13:13.078 SoScannerMephistoSyntaxControl::sendDataToHost No socket connection to ReModProt available.
```

Fig. 1. Screen capture of the monitoring and guidance system for minimal-invasive thermal therapy showing a respiratory belt triggered laser-induced thermal ablation.

3 Conclusion

We have created a promising new system to guide and monitor thermal ablations with the purpose of facilitating accurate temporal and spatial control of heat deposition during the intervention.

References

1. Ishihara, Y, et al: A precise and fast temperature mapping using water proton chemical shift. Magnetic Resonance in Medicine, 2005 Dec., 34(6): 814-823.
2. Lorenz, CH, et al: Interactive Front End (IFE) - a platform for graphical MR scanner control and scan automation. Proc. ISMRM, 2005 May, 2170.

Quantitative 3-D Imaging of the Coronaries using C-Arm CT

M. Polyanskaya[1,2], G. Lauritsch[2], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. Providing surgeons with 3-D reconstructed cardiac volumes intra-operatively is a significant step forward in cardiac interventions. However, visual information often is insufficient for interventions aimed at treatment of coronary vessel disease. In these cases quantitative data, such as vessel diameter, tortuosity, calcification and spatial orientation are also required. This work states its purpose to provide 3-D coronary vessel segmentation which is a necessary basis for further quantitative estimations.

1 Introduction

3-D imaging of the coronary tree from a single 5-second rotational angiography (C-arm CT) is challenging due to the cardiac motion. Recent techniques achieve 3-D reconstruction of the coronary tree by estimating and compensating the coronary motion field which results in good visual quality of the volume, but nevertheless blurred and not appropriate for quantitative measurements. However, topology can be extracted from this 3-D reconstructed volume and numerical parameters of vessels can be estimated in 2-D projection images. In this work we focus on the first step - segmentation of coronary vessel tree in a 3-D C-arm CT reconstructed volume.

2 Methods

Reconstructed and motion compensated 3-D cardiac volumes are illegible and have streak artefacts. Therefore, vessel enhancement filtering [1] is applied as a preprocessing step.

The vessel enhancement filtering considers vessels as tubular structures and is based on the analysis of a Hessian matrix. For each voxel of a volume its likeliness of being a vessel, called vesselness, is computed. It depends on the relation of eigenvalues of a Hessian matrix and can be represented in different ways. In this work we used the representation of Sato [2], which is strictly speaking a likeliness measure of a voxel to be a part of a line structure in 3D volume:

$$V(x) = \begin{cases} \lambda_2 \cdot e^{-\dfrac{\lambda_1^2}{2(\alpha\lambda_2)^2}} & , if \lambda_2 \neq 0 \\ 0 & , otherwise \end{cases} \tag{1}$$

where $V(x)$ is a vesselness measure, λ_1, λ_2 denote the eigenvalues of a Hessian matrix at a current voxel x, α is an adjustable parameter. Eigenvalues are sorted in a descending order: $|\lambda_1| > |\lambda_2| > |\lambda_3|$.

The Hessian matrix is computed at each voxel of the image, which was previously smoothed with Gaussian filter. To enable enhancement of vessels of different width the size of Gaussian kernel is varied. The final vessel-enhanced image represents a collection of the best responses of the vesselness filter at each voxel.

Fig. 1. 3-D Coronary vessel segmentation: left - original data volume, right - result after the segmentation algorithm

The preprocessed image is then binarized with the thresholding operator in order to determine seed points. These seed points are used for an initialization of the region growing algorithm. Region growing algorithm is used for the final segmentation of the vessel tree from the enhanced volume.

3 Conclusion

The proposed method was evaluated on eight C-arm CT real-data sets. One of the cases is presented in Figure 1. The results show that the algorithm is able to segment large branches of coronary tree confidently.

Segmentation of smaller vessels needs further improvements of the proposed algorithm. The reason of omitting smaller vessels is vanishing of vesselness measure in the points of bifurcations. In future work this problem is supposed to be solved by adding the information about vessel direction, which can be extracted from eigen vectors of Hessian matrices. Additionally it is planned to create a proper evaluation scheme to assess the algorithm.

References

1. A. Enquobahrie, L. Ibanez, E. Bullitt, and S. Aylward, "Vessel enhancing diffusion filter," 2007. [Online]. Available: http://hdl.handle.net/1926/558
2. Sato Y., Nakajima S., Shiraga N., Atsumi H., Yoshida S., Koller T., Gerig G., Kikinis R. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. Med Image Anal. 1998 Jun;2(2):143-68.

Clustering for Ear Impressions

K. Sickel[1] and V. Bubnik[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Siemens Hearing Instruments, Inc., NJ, USA

Abstract. Feature detection for ear impressions is a challenging task due to the high variability of organic shapes. A further drawback is the high amount of noise in the data due to the acquisition procedure. To improve accuracy and robustness we apply clustering as a preprocessing step to identify regions of interest. This additional information allows the analytical algorithms to perform better and suppresses outliers.

1 Introduction

The manufacturing of customized hearing aids is a tedious and error-prone process. To eliminate the dependency on human skill and expertise and automation framework was developed by Sickel et al.[1]. The framework utilizes an expert system, which stores the design process in its knowledge base. The encoded rules are using features detected on an ear impression, see Fig. 1. Therefore, the accuracy and robustness of the detected feature is crucial for the final result.

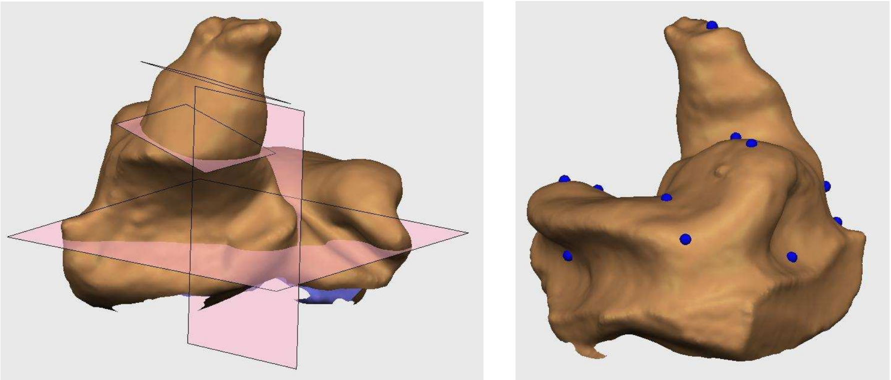


Fig. 1. On the left a subset of the point features and on the right a subset of the plane features is depicted.

2 Method and Results

The algorithms for the feature detection are described in [2]. The detected features are comprised of point, plane, ridge and area features. They describe the important anatomical features of an ear.

However, the algorithms suffer from the fact, that no previous knowledge is used to guide the detection. Our approach will add this knowledge by identifying regions of interest for every feature. This is achieved by applying the following steps:

- Clustering a dataset of about 400 ear impressions.
  - Clustering is done by registering the ear impressions, with a specifically adapted ICP algorithm. The adaptations include a pre-registration using the centerline of the impressions. The ICP algorithm uses the point-to-plane error metric [2].
  - Hierarchical clustering is applied since the number of natural clusters is unknown.
  - Heuristics, analyzing the similarity level are used to identify the natural number of clusters [4].
- Cluster centers are labeled by an expert.
- For each new impression the best cluster is identified and the labeled features are used as starting point for a local feature detection.

The first prototype results of the system proved the effectiveness of the idea by correcting the feature detection for a known problematic ear impression.

3 Outlook

The next steps will be the evaluation of the clustering results and labeling of the cluster centers. Furthermore the current feature detection algorithms have to be modified to utilize the additional information.

References

1. Konrad Sickel; Sajjad Baloch; Vojtech Bubnik; Rupen Melkisetoglu; Sergei Azernikov; Tong Fang; and Joachim Hornegger: Semi-Automatic Manufacturing of Customized Hearing Aids Using a Feature Driven Rule-based Framework. In *Proceedings of the Vision, Modeling, and Visualization Workshop (VMV)*, pp. 305–312, 2009, Braunschweig, Germany.
2. Sajjad Baloch; Rupen Melkisetoglu; S. Flory; Sergei Azernikov; Greg Slabaugh; and Tong Fang: Automatic Detection of Anatomical Features on 3D Ear Impressions for Canonical Representation. In *Proceedings of International Conference, Medical Image Computing (MICCAI) 2010*, 2010 (accepted).
3. Szymon Rusinkiewicz and Marc Levoy : Efficient variants of the ICP algorithm. In *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, pp. 145–152, 2001, Quebec City, Canada.
4. R. O. Duda; P. E. Hart; and D. G. Stork: Pattern Classification: Pattern Classification Pt.1, 2nd edition 2000, John Wiley & Sons.

Texture Analysis of 3-D Ultrasound Data

E. Kollorz[1,2], J. Hornegger[1], D. Schmidt[2], and T. Kuwert[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Department of Nuclear Medicine, University Hospital Erlangen, Krankenhausstr.
12, 91054 Erlangen, Germany

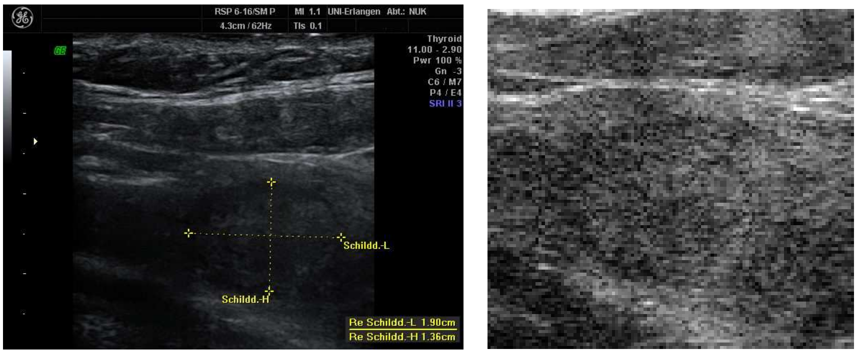
Abstract. The goal of this work is to develop methods that classify nodule types within the thyroid gland to get a 3-D segmentation. This facilitates therapy planning and allows to study the growing behavior and the shape of thyroid nodules. Our approach is based on texture features and support vector machine.

1 Motivation

The standard examination of thyroid nodules with ultrasound is done in 2-D. To get the volume of the observed nodule the three major axes of the nodule are measured with two orthogonal 2-D slices. The measured distances are multiplied by a fixed factor of 0.52 to get the “ellipsoid” volume. To track the growth of the nodule the patient will revisit the hospital approximately every third month. Fig. 1 (a) shows that sometimes no clear boundary of the nodule exists and therefore measurements can differ. The physician also looks at the echogenicity pattern of the nodule. We work with 3-D ultrasound (US) volumes to get a precise 3-D segmentation. Fig. 1 (b) shows a cropped 2-D slice of a 3-D volume.

2 Methods

First, the 2-D slices of the volume were divided into subregions for computing gray level co-occurrence matrices (GLCM) of each subregion [2]. The GLCMs were computed with four different orientations (0°, 45°, 90°, 135°) and selected neighbor distances $d \in 1, \ldots, 4$. Based on these 16 GLCMs, 14 original features described by Haralick *et al.* [1] were used as well as some additional ones, e.g. energy, dissimilarity, autocorrelation, GLCM mean, intensity mean, $\ldots$. For each distance, the values were averaged over the orientations to achieve rotation invariance. Hence, we got a feature vector with 82 components in total for each subregion. A support vector machine (SVM) with different kernels was then used to classify these subregions.



(a) Volume estimation of the nodule with one of the two 2-D slices

(b) Hypoechoic, cystic nodule in the thyroid gland (a cropped 2-D slice of the 3-D volume)

Fig. 1. 2-D B-mode image and 2-D slice of a 3-D US volume of a hypoechoic, cystic thyroid nodule.

3 Results

Our preliminary results were generated using one volume containing a hypoechoic and cystic nodule. However, a bigger evaluation on our current database, including 28 patients with 49 nodule volumes, will follow. These volumes represent the first examination of a long-term observation. The follow-up examinations will be in approximately three month cycles. Different nodule types are included in the database: hypoechoic and cystic, complex echoic, and isoechoic.

We did a balanced 10-fold cross validation on the subregions of this one volume. Accuracy of $80\% - 97\%$ was achieved with a linear kernel (Table 1) for each subregion size ($8 \times 8 - 32 \times 32$). In addition, normalized cross-correlation (NCC) was computed to identify highly correlated features. The 22 features for one distance can be reduced to approximately 10-15 in subregions.

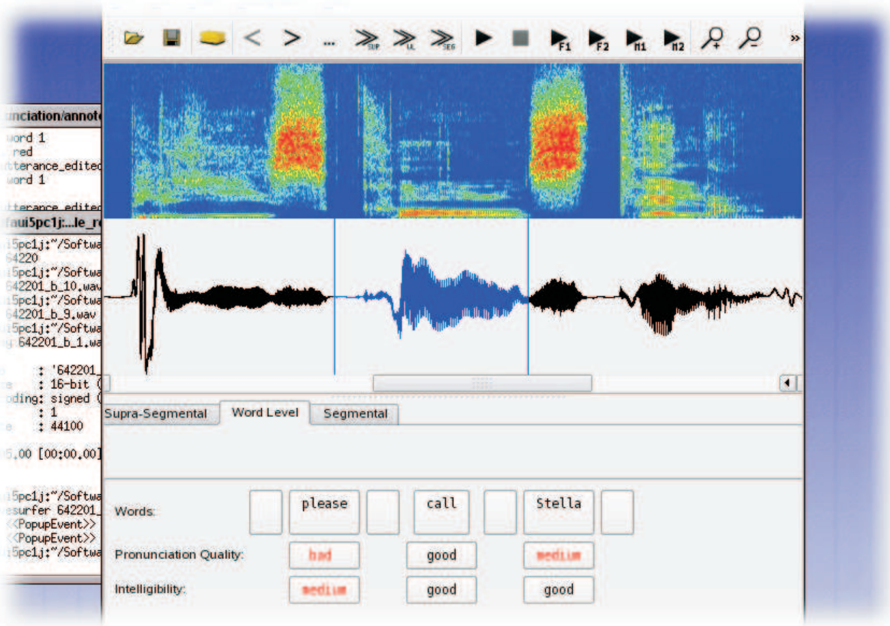
Boxsize	**Number of samples (+/-)**	**Linear kernel**	**Polynomial kernel**	**RBF kernel**	**Sigmoid kernel**
8	2190	80.61%	77.53%	80.18%	79.22%
16	510	86.86%	76.86%	83.13%	78.92%
24	190	88.42%	76.84%	85.26%	81.57%
32	80	97.50%	82.50%	97.50%	98.12%

Table 1. Accuracy results with penalty term $C = 150$ for four different kernels. For a subregion of 8×8, 2190 positive and 2190 negative samples were extracted from selected slices.

References

1. Haralick, R. M.; Shanmugan, K.; Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, Vol. (SMC-3), No. 6, Nov. 1973, 610–621
2. Liu, B.; Cheng, H.D.; Huang, J.; Tian, J.; Tang, X.; Liu, J.: Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recognition, Vol. 43, No. 1, Jan. 2010, 280–298

Signal Processing



```
nciation/annot
word 1
red
tterance_edited
word 1

tterance_edited
faui5pc1j:...le_r
i5pc1j:~/Softwa
64220
i5pc1j:~/Softwa
642201_b_10.wav
i5pc1j:~/Softwa
642201_b_9.wav
i5pc1j:~/Softwa
642201_b_1.wa

      : '642201_
      : 16-bit (
oding: signed (
      : 1
      : 44100

5.00 [00:00.00]

i5pc1j:~/Softwa
esurfer 642201_
<<PopupEvent>>
<<PopupEvent>>
i5pc1j:~/Softwa
```

Words:	please	call	Stella
Pronunciation Quality:	bad	good	medium
Intelligibility:	medium	good	good

Interaction in the ISA-House — On-Talk vs. Off-Talk

W. Spiegl

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. In this article we present our experiments on the automatic detection of focus of attention of a speech-controlled home-assistant system: On-/Off-Talk. Based on prosodic features we achieved a error rate of approx. 25 % on a SVM (support-vector-machine).

1 Introduction

The intelligent, senior-adapted House (ISA-House) is a living room with an integrated assistance system which can be controlled by spontaneous speech. The goal of such a system is to support the everyday life of elderly people and to elongate their independency. To achieve this, usability studies were performed to find out the needs of elderly people and to record speech data. The acquired speech data is pooled in the FAU IISAH Corpus [1]. Next to transcribed speech for training a speech recognizer adapted to the voice of elderly people the FAU IISAH Corpus contains labeled dialogs whether the user talks to the system (On-Talk) or not, for example to someone else (Off-Talk).

2 Experiments and Results

The here presented work refers to experiments to decide automatically between On- and Off-Talk. An initial approach is based on the assumption, that people change the intonation of their utterance to indicate an interaction. Therefore prosodic features (187) of each utterance in the FAU IISAH Corpus are calculated over voiced segments. Subsequently the features are modeled by a simple average/mean over all segments per turn, so that each utterance is represented by a feature vector of 187 elements. Based on correlation the best features were selected. After separating the data in training and test (66 %/33 %) three classifiers were utilized: nearest neighbour, bayes and SVM (support vector machine). The SVM gave the best results with an error rate of 25.36 %.

With a closer look on the selected features nine out of the ten highest ranked features of the 187 prosodic features are energy based. With respect to this an inspection of the recorded speech data supported this. In figure 1 a representative sample wave form for On-Talk and Off-Talk is presented. One clearly can see the difference in energy in the wave forms.

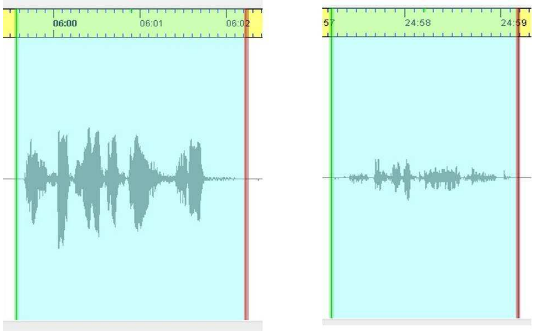


Fig. 1. Energy-based difference between On- and Off-Talk: representative sample wave form for On-Talk (left image:"Dann schalten wir *des Licht ein, oder?") and Off-Talk (right image: "*Bin g'spannt, welcher Termin er *etsa löscht."). The x-axis represents the timeline and the y-axis the amplitude of the sample sounds.

3 Outlook

The here presented approach for On-/Off-Talk detection is based on pre-segmented utterances. This does not reflect the reality where a system must decide in realtime if the user talks to him or not. Therefore the utterance-based analysis must be extended to an online-based: Each frame of the short-term analysis has its own estimation of On-/Off-Talk (probability) processed during talking.

Acknowledgement

In the research association FitForAge[1] of the Bavarian Research Foundation a team of scientists and engineers of 13 chairs of the four Bavarian Universities Erlangen-Nuremberg, Munich, Regensburg and Würzburg with 25 industrial partners is cooperating to develop products and services for an aging society. The aim of the research association is the development of technology-based solutions for aging people in their future life, to assist them at home, in their professional life, in the communication with their surroundings and in their participating of road traffic. Finally, not only elderly people should profit from this solutions, but all groups of society.

References

1. Spiegl, Werner; Riedhammer, Korbinian; Steidl Stefan; Nöth Elmar: FAU IISAH Corpus — A German Speech Database Consisting of Human-Machine and Human-Human Interaction Acquired by Close-Talking and Far-Distance Microphones. in Proceedings of the seventh international conference on Language Resources and Evaluation (LREC), S. 2420 – 2423, Malta, 2010

---

[1] www.fit4age.org

The Multimodal SKT

S. Soutschek[2], S. Steidl[1], E. Nöth[1], J. Hornegger[1], H. Erzigkeit[2], and
J. Kornhuber[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany,
[2] Department of Psychiatry and Psychotherapy, University Hospital Erlangen,
Schwabachanlage 6, 91058 Erlangen, Germany

Abstract. The aim of this work is to reduce influences of examiners at psychometric tests for the example of the "Syndrom-Kurztest"' (SKT). This means to increase the intersubjective conformance of the examiners by acquiring validity increasing variables with the help of multimodal sensors. Microphones, Eyetracker or Touch-Screen monitors are examples for such sensors.

1 Introduction

Particularly with regard to diagnosis and therapy control of age-related diseases, e.g. dementia, the measurement of cognitive impairments takes on an increasingly significant role. Within informARTik, it is investigated, whether an interactive test system for the measurement of cognitive impairments can be realized which is appropriate for the use in the field of telemedicine and which can be independently used by the patient at home. Although there are also other tests available [1] in the field of dementia, the basis for this investigation is the SKT [2], an international validated performance test for assessing deficits of memory and attention. The original material as well as the computerized version [3] is shown in figure 1.

Fig. 1. Left: Original SKT, Right: Computerized SKT

2 Integration of Multimodal Sensors

In an intelligent home environment one huge factor to execute a reliable test is missing compared to the daily routine of a hospital: The surgeon with his professional expertise. Independent of the place where the test is executed, a reliable analysis of the test results strongly depends on additional information about this test environment. In the clinical environment this data is already given by the assessment of the situation through the supervisor. In case of an intelligent home this information needs to be estimated by the analysis of acquired sensors data. Sensor data used in this work to solve this task are:

- Physiological signals
- Speech
- Visual focus

Future evaluation studies with the focus group will show if the acquired data from these sensors is sufficient to estimate the test situation and to provide the surgeon all necessary information for a correct and reliable interpretation of the test results.

Acknowledgement

In the research association FitForAge[3] of the Bavarian Research Foundation a team of scientists and engineers of 13 chairs of the four Bavarian Universities Erlangen-Nuremberg, Munich, Regensburg and Wrzburg with 25 industrial partners is cooperating to develop products and services for an aging society. The aim of the research association is the development of technology-based solutions for aging people in their future life, to assist them at home, in their professional life, in the communication with their surroundings and in their participating of road traffic. Finally, not only elderly people should profit from this solutions, but all groups of society.

References

1. B. Cullen, B. O'Neill,J. Evans R. Coen, B. Lawlor, *A review of screening tests for cognitive impairment.* Journal of Neurology, Neurosurgery, and Psychiatry, 78, 2007, 8, p.790-9, ISSN: 1468-330X
2. H. Erzigkeit *SKT. A short cognitive performance test for assessing deficits of memory and attention* User's Manual, 23rd completely revised edition. Geromed GmbH, 2001.
3. S. Soutschek, W. Spiegl, M. Gropp, S. Steidl, E. Nöth, J. Hornegger, H. Erzigkeit, J. Kornhuber *Validierter SKT als Multimodele Telemedizinische Applikation.* VDE Verlag GmbH, 2009. 2. Deutscher AAL (Ambient Assisted Living)-Kongress. ISBN 978-3-8007-3138-1.

---

[3] www.fit4age.org

A Snappy Radio Signal Classification Framework for KAIMAN

D. Kolb[1,2] and E. Nöth[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] MEDAV GmbH, Graefenberger Str. 32–34, 91080 Uttenreuth, Germany

Abstract. The strong need for a robust and real-time transmission classification can be found both in civil and non-civil applications. In this paper, we propose a new technique based on modern and low-complexity object detection approaches. The transmission scanner uses a combination of Haar-like, cepstral and spectral features to detect and classify different, co-occurring narrowband transmissions within wideband signals. The transmission scanner is evaluated with recorded real world data in order to fulfill real conditions. The evaluation shows that this system performs very well with up to 99.9% accuracy.

1 Introduction

To know what is on air has always been a strong interest of civil and non-civil institutions. The civil area of research has the aim to detect white spaces in frequency bands in order to establish secondary usage, as it is done in Cognitive Radio (CR) [2]. In contrast to this, the non-civil institutions try to identify every signal or a special type of signal. Both benefit from increasing instantaneous receiver bandwidths, made possible by Software-Defined Radio (SDR). The wider the bandwidth, the higher the knowledge gain can be. The drawback of such a wideband receiver is that the requirements for a real-time signal processing are higher than before, as the search space increases. For the project KAIMAN, we propose an adaptive classification framework, which is able to classifiy different transmission standards.

2 Classification Framework

For classification it is necessary, to process different input data, delivered by a wideband receiver. In case of a scanning receiver, the classifier operates in the scan mode and receives only spectral information. If this is not the case, the classifier obtains both spectral and time information. Both operating modes are shown in Fig. 1. For classification a transmission class adapted 2-D mask $\mathbf{X}_{\mathrm{M}}$ or 1-D mask $\mathbf{x}_M$ is swept over the spectrogram or the spectrum. For every position of the mask, features will be extracted and classified with an AdaBoost

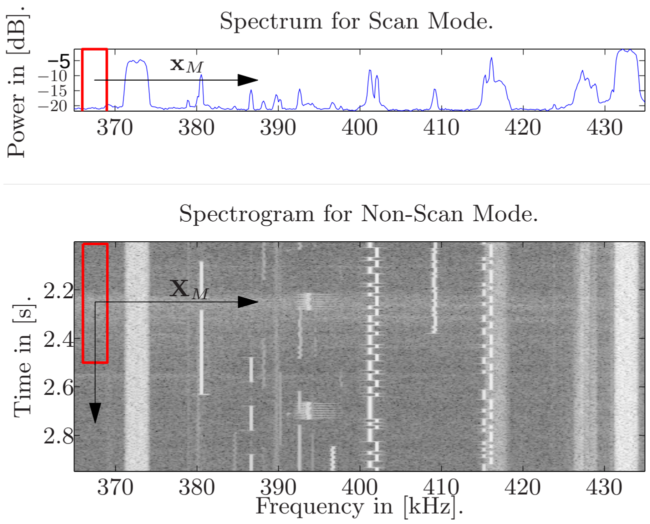


Fig. 1. An absolute, logarithmic spectrogram, containing several transmissions, e. g., STANAG-4285 and BAUDOT.

classifier. We use Haar-like features [3], cepstral features [1] and spectral features to describe the different transmission standards. Table 1 shows the results of the evaluation with real world signals. We obtained impressive classification rates up to 99.9% and performed about 13.000 classifications per second, which is a great step to real-time applications.

	Scan Mode		Non-Scan Mode	
Transmission	TP Rate	FP Rate	TP Rate	FP Rate
BAUDOT	0.998	0.109	0.901	0.099
F3E SPEECH	0.929	0.004	0.945	0.003
J3E SPEECH	0.800	0.011	0.815	0.003
POCSAG	0.999	0.001	0.791	0.008

Table 1. Accuracy of the classifier with 30 iterations.

3 Acknowledgements

We acknowledge the financial support from MEDAV GmbH, Germany and the Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie.

References

1. Bogert, B. P., Healy, M. J. R., Tukey, J. W.: The frequency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking.

Proc. Symp. Time Series Analysis, Ed.: M. Rosenblatt,John Wiley, pp. 209–243, 1963.

2. Mitola III, J., Maguire, Jr.,G. Q.: Cognitive radio: Making software radios more personal. *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13–18, Aug 1999.
3. Viola, P., Jones, M. J.: Robust Real-Time Face Detection. *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

Automatic Detection of Edentulous Speakers with Insufficient Dentures

T. Bocklet

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Dental rehabilitation by complete dentures is a state-of-the-art approach to improve functional aspects of the oral cavity of edentulous patients. We introduce a dataset of 13 edentulous patients that have been recorded with and without complete dentures in situ. There are also recordings of these patients that have been rated an insufficient fit. We performed ASR experiment with an state-of-the-art system and showed an performance increase of 27 % when sufficient dentures are worn.

1 Dataset

The dataset was first introduced in [1]. The original version contains 28 edentulous speakers (see Fig. 1) with an average age of 64 ± 10 years. 13 out of these 28 patients have been rated an insufficient fit of dentures. New dentures have been produced for these patients. Three recordings are available for each speaker: without any dentures, with sufficient dentures, with insufficient dentures. In each recording the text “Der Nordwind und die Sonne” (NWS) was read. These speakers represent the basis of this paper.

Fig. 1. Edentulous person

2 Experiments and Results

Based on the 13 speaker dataset described in Sec. 1 we performed different automatic speech recognition (ASR) experiments. Our speech recognition system is based on

dataset	mean WA	std. dev	min WA	max WA
without	60.06	± 10.35	39.48	77.78
insufficient	64.35	± 9.64	45.37	80.56
sufficient	70.91	± 6.04	57.51	80.56

Table 1. Word accuracy (WA) result, according standard deviation and minimum/maximum WA value for the three different subsets: Without wearing dentures, with insufficient dentures, with sufficient dentures

semi-continuous Hidden Markov Models (HMM). Phones are modeled in a variable context, the so-called polyphones. Details on the system can be found in [2]. As recognition measure we calculate the word accuracy (WA)

$$\mathrm{WA} = [1 - (n_{\mathrm{sub}} + n_{\mathrm{del}} + n_{\mathrm{ins}})/n_{\mathrm{all}}] \cdot 100 \tag{1}$$

w.r.t. the NWS text. Here, n_{sub} denotes the number of wrongly substituted words, n_{del} the number of deleted words and n_{ins} the number of wrongly inserted words.

Table 1 shows the WA results we achieved on the three different subsets, i.e., without, with insufficient and with sufficient dentures. The mean WA on recordings without dentures (60.06 %) was lower then recordings with insufficient dentures (64.35 %). In the case of sufficient dentures a WA of 70.91 % was measured. This is an improvement by 18 % compared to the recordings without dentures and 10 % compared to the recordings with insufficient dentures. The standard deviation on recordings without any dentures was ±10.35. The value decreased to ±9.64 and ±6.04 when wearing insufficient dentures or sufficient dentures respectively. This is an improvement of 37 % when comparing the values of insufficient and sufficient dentures. This improvement is also visible when focusing on the minimum word accuracy values. Insufficient dentures improved the minimal WA from 39.48 % to 45.37 %. This is an improvement by 15 %. Wearing sufficient dentures again improved the value by 27 % to a WA of 57.51 %.

3 Summary

n this work we performed ASR experiments on a dataset of 13 edentulous patients without, with insufficient and sufficient dentures. Our experiments showed an improvement of the mean WA of 10 % between recordings without any dentures and recordings with insufficient dentures. Sufficient dentures improved these results by another 18 %.

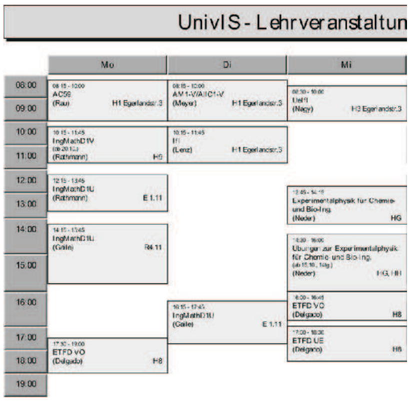
4 Acknowledgement

This work was supported by the Wilhelm Sander-Foundation, Germany (AZ: 2007.100.1). The authors are responsible for the content of this article.

References

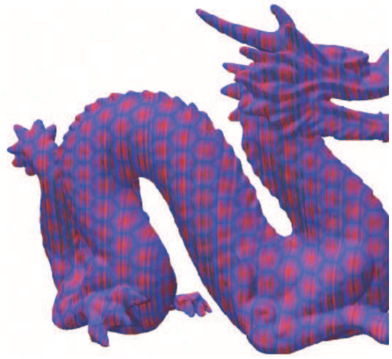
1. Haderlein, T; Bocklet, T; Maier, A; Nöth, E; Knipfer, C, and Stelzle, F: Objective vs. Subjective Evaluation of Speakers with and without Complete Dentures. Proc. TSD 2009, Pilsen, Czech Republic, September 2009, Springer LNAI (2009) 170-177
2. Stemmer, G: Modeling Variability in Speech Recognition. Logos Verlag, Berlin, Germany (2005)

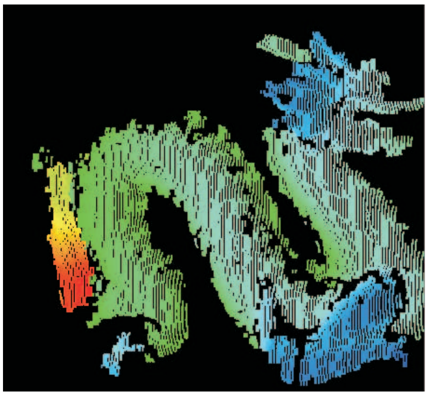
Mixed Session



UnivIS - Lehrveranstaltun

	Mo	Di	Mi
08:00	08:15 - 10:00 AC59 (Rau) H1 Egerlandstr.3	08:15 - 10:00 AM1-V/AIIC1-V (Meyer) H1 Egerlandstr.3	08:30 - 10:00 UeIfI (Nagy) H3 Egerlandstr.3
09:00			
10:00	10:15 - 11:45 IngMathD1V (ab 20.10.) (Rathmann) H5	10:15 - 11:45 IfI (Lenz) H1 Egerlandstr.3	
11:00			
12:00	12:15 - 13:45 IngMathD1U (Rathmann) E 1.11		12:45 - 14:15 Experimentalphysik für Chemie- und Bio-Ing. (Neder) HG
13:00			
14:00	14:15 - 15:45 IngMathD1U (Gaile) R4.11		14:30 - 16:00 Übungen zur Experimentalphysik für Chemie- und Bio-Ing. (ab 15.10., 14tg.) (Neder) HG, HH
15:00			
16:00		16:15 - 17:45 IngMathD1U (Gaile) E 1.11	16:00 - 16:45 ETFD VO (Delgado) H8
17:00	17:30 - 19:00 ETFD VO (Delgado) H8		17:00 - 18:30 ETFD UE (Delgado) H8
18:00			
19:00			





ASR Performance for Emotional Speech

S. Steidl

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Automatic speech recognition (ASR) systems trained exclusively on emotional speech are compared to a baseline system trained on neutral speech. ASR performance is evaluated on speech of the four emotion categories *motherese*, *neutral*, *emphatic*, and *anger*.

1 ASR Scenario

ASR systems that are trained exclusively on speech data of one emotion category are compared to a baseline system that is trained on neutral speech only. Comparisons of this kind are difficult since the emotional speech corpora have to be comparable with respect to the amount of data, the acoustic recording conditions, the speakers, etc. In these experiments, the FAU Aibo Emotion Corpus [1] is used. Thus, the same set of speakers and the same acoustic conditions are guaranteed. Emotional subsets of equal size (in terms of the number of words) are created by down-sampling the corpus. The same vocabulary is used in all ASR experiments. However, the vocabulary that is actually used by the speakers in one particular emotional state depends on the emotional state. The ASR system is based on semi-continuous hidden Markov Models sharing the same set of Gaussian densities. The number of densities is reduced from 500 to 50 in order to train acoustic models on such small amounts of data. Another scenario where an ASR system trained on large amounts of neutral speech is adapted to emotional speech by adding emotional data to the neutral training data is described in [2].

2 ASR Experiments

The evaluation of the baseline system shows that *motherese* speech (43.6 % WA) is clearly recognized worse than *neutral* speech (60.3 % WA). Acoustic and linguistic realizations obviously differ strongly from the *neutral* ones. However, *emphatic* and *angry* speech are even recognized better than *neutral* speech: 61.3 % and 64.9 % WA, respectively. This is certainly due to the fact that the corpus contains only slight forms of anger. In this corpus, *emphatic* can be considered to be a pre-stage of anger. The children speak in a pronounced, accentuated, sometimes hyper-articulated way, which is obviously recognized better than normal *neutral* speech.

If the ASR system is trained on emotional speech, the performance on speech of the very same emotion in general increases. The results show that training on *emphatic* speech also helps to increase the performance on *angry* speech and vice versa, supporting the assumption that *emphatic* is a pre-stage of *anger* in our type of data. Fig. 1 shows the results of the re-training of the ASR system on *neutral* speech. Acoustic and linguistic models can be re-trained independently from each other. Whereas the performance increases significantly for *emphatic* and *angry*, it decreases clearly for *motherese* as well as for *neutral*.

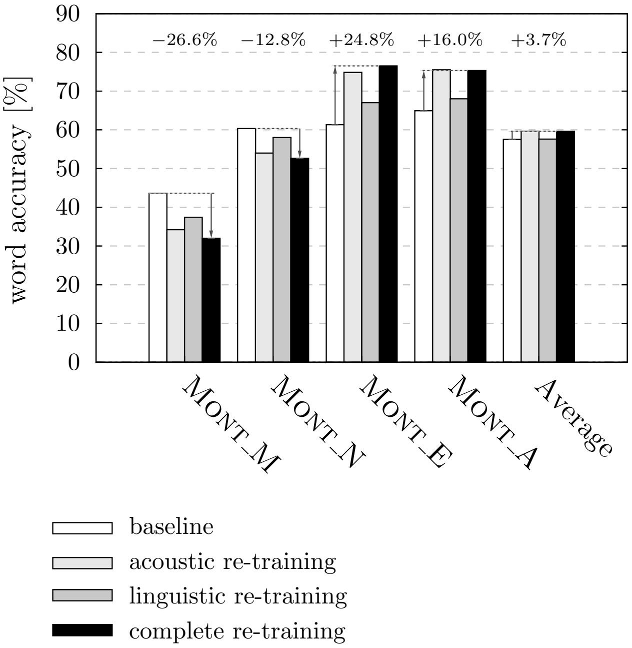


Fig. 1. ASR system trained on *emphatic* speech compared to the baseline system trained on *neutral* speech and its effect on speech of the emotion categories *motherese*, *neutral*, *emphatic*, and *anger*

References

1. Steidl, S.: Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech. Logos Verlag, Berlin, 2009.
2. Steidl, S.; Batliner, A.; Seppi, D.; Schuller, B.: On the Impact of Children's Emotional Speech on Acoustic and Language Models. EURASIP Journal on Audio, Speech, and Music Processing, Volume 2010, Article ID 783954, 14 pages, doi:10.1155/2010/783954

Automatic Evaluation of Voices after Partial Laryngectomy – Intelligibility on the Phone

T. Haderlein[1 2], A. Maier[1], E. Nöth[1], and F. Rosanowski[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Department of Phoniatrics and Pedaudiology, University Clinics Erlangen,
Bohlenplatz 21, 91054 Erlangen, Germany

Abstract. The intelligibility of patients after partial laryngectomy was measured automatically using automatic speech recognition and prosodic analysis. Correlations of up to r=0.66 were found between subjective and objective evaluation of close-talk and telephone recordings.

For voice and speech rehabilitation of partially laryngectomized patients, speech intelligibility is an important criterion. Especially the communication on the telephone should be considered since it is an important aspect in modern society and even more for elderly people whose physical abilities are restricted. 82 persons (68 men and 14 women) read the German version of the tale “The North Wind and the Sun” which consists of 108 words. They were recorded simultaneously by a close-talk microphone (16 kHz, 16 bit) and via a landline telephone (8 kHz, 16 bit). Five experienced raters evaluated the intelligibility subjectively on a 5-point scale (1 = “very good”, 5 = “very bad”).

Objective evaluation was performed by using the word accuracy (WA) of a speech recognition system and a set of 95 word-based and 15 phrase-based prosodic features [1,2]. For the close-talk recordings, the recognizer was trained with 16 kHz recordings. For the telephone recordings, the same training data were down-sampled to 8 kHz since no sufficiently large amount of real telephone speech was available.

Intelligibility was rated lower on the telephone recordings by the speech recognizer (47.0% WA; see Table 1) and the raters (score 3.3) than on the close-talk recordings (51.8% WA, score 2.9). The correlation between WA and the human evaluation was r=–0.51 for the close-talk recordings and r=–0.62 for the telephone recordings. The correlation between single prosodic features and the human intelligibility ratings revealed correlations of up to r=0.66 (see selected results in Table 2). Obviously, intelligibility is related to speaking rate and voice quality since the highest correlations were found for duration-based features and features measuring energy. A distorted voice shows higher noise level, i.e. more portions of high frequencies. This may be the reason why there is a difference in the results for close-talk and telephone speech for the energy-based features: The telephone cuts off all frequencies above 3400 Hz and with it a lot of noise. This may also hold for the difference in the correlations between WA and human rating. For the duration-based features, no such differences between both qualities of the recordings were noticed.

Better correlations between human rating and automatic evaluation had been achieved with other voice and speech pathologies, e.g. for total laryngectomees [3] and children with cleft lip and palate [4]. These pathologies, however, showed a much broader range of distortions than partial laryngectomees. Hence, the single features

Table 1. Word accuracy (WA) in percent for 82 partially laryngectomized persons

	recognizer	μ(WA)	σ(WA)	min(WA)	max(WA)
headset	*16kHz*	51.8	19.2	0.0	82.4
telephone	*8kHz*	47.0	19.6	–2.7	79.6

Table 2. Correlation r of subjective intelligibility scores from 5 raters against single prosodic features for close-talk and telephone recordings (selected results)

	headset	telephone
ratio of durations: voiceless sections/signal	0.56	0.65
duration of silent pause before word	0.57	0.57
duration of word-pause-word interval	0.66	0.64
normalized energy (word-pause-word interval)	0.66	0.51

examined in this study are not able to capture the complex phenomenon of intelligibility alone. Future work will incorporate the combination of features to find a subset that forms a better representation of the human evaluation.

The data obtained in this study allow for the following conclusions: There is a significant correlation between subjective rating of intelligibility and automatic evaluation. Hence, the method can serve as the basis for more research towards an automatic, objective system that can support voice and speech rehabilitation.

This work was partially funded by the German Cancer Aid (Deutsche Krebshilfe) under grant 107873.

References

1. Batliner, A., Buckow, A., Niemann, H., Nöth, E., Warnke, V.: The Prosody Module. In: Wahlster, W., ed.: Verbmobil: Foundations of Speech-to-Speech Translation. Springer, Berlin (2000) 106–121
2. Batliner, A., Fischer, K., Huber, R., Spilker, J., Nöth, E.: Desperately Seeking Emotions: Actors, Wizards, and Human Beings. In: Cowie, R., Douglas-Cowie, E., Schröder, M., eds.: Proc. ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research. Newcastle, Northern Ireland (2000) 195–200
3. Haderlein, T.: Automatic Evaluation of Tracheoesophageal Substitute Voices. Logos Verlag, Berlin, Germany (2007)
4. Maier, A.: Speech of Children with Cleft Lip and Palate: Automatic Assessment. Logos Verlag, Berlin, Germany (2009)

How to Find Optimal Time Tables

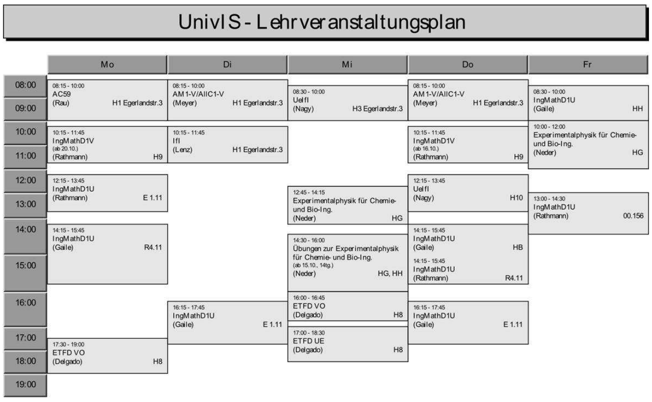
P. Wilke

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. We describe how optimal time tables are found using various algorithms provided by the EATTS Erlangen Advanced Time Tabling System.

1 Introduction

A time table or schedule or plan is an organized list, usually set out in tabular form, providing information about a series of arranged events. The cost of a time table is the sum of all penalty points for violations of soft and hard constraints. A typical time table is shown in fig. 1. An optimal time table is the time table with the lowest costs. Because time tabling problems are NP-complete efforts have to be reduced to finding sub-optimal solutions.



UnivIS - Lehrveranstaltungsplan

	Mo	Di	Mi	Do	Fr
08:00	08:15 - 10:00 AC59 (Rau) H1 Egerlandstr.3	08:15 - 10:00 AM1-V/AIIC1-V (Meyer) H1 Egerlandstr.3	08:30 - 10:00 UeIfI (Nagy) H3 Egerlandstr.3	08:15 - 10:00 AM1-V/AIIC1-V (Meyer) H1 Egerlandstr.3	08:30 - 10:00 IngMathD1U (Gaile) HH
09:00					
10:00	10:15 - 11:45 IngMathD1V (ab 20.10.) (Rathmann) H9	10:15 - 11:45 IfI (Lenz) H1 Egerlandstr.3		10:15 - 11:45 IngMathD1V (ab 16.10.) (Rathmann) H9	10:00 - 12:00 Experimentalphysik für Chemie- und Bio-Ing. (Neder) HG
11:00					
12:00	12:15 - 13:45 IngMathD1U (Rathmann) E 1.11		12:45 - 14:15 Experimentalphysik für Chemie- und Bio-Ing. (Neder) HG	12:15 - 13:45 UeIfI (Nagy) H10	
13:00					13:00 - 14:30 IngMathD1U (Rathmann) 00.156
14:00	14:15 - 15:45 IngMathD1U (Gaile) R4.11		14:30 - 16:00 Übungen zur Experimentalphysik für Chemie- und Bio-Ing. (ab 15.10., 14tg.) (Neder) HG, HH	14:15 - 15:45 IngMathD1U (Gaile) HB 14:15 - 15:45 IngMathD1U (Rathmann) R4.11	
15:00					
16:00		16:15 - 17:45 IngMathD1U (Gaile) E 1.11	16:00 - 16:45 ETFD VO (Delgado) H8	16:15 - 17:45 IngMathD1U (Gaile) E 1.11	
17:00	17:30 - 19:00 ETFD VO (Delgado) H8		17:00 - 18:30 ETFD UE (Delgado) H8		
18:00					
19:00					

Fig. 1. A typical time table

2 EATTS

EATTS Erlangen Advanced Time Tabling System is a software build to design, implement and run time tabling algorithms. It is a client-server solution written in Java and uses a web-based interface. The performance of some algorithms solving a school time table problem are shown in fig. 2.

EATTS is currently is use at our university to schedule several courses and events.

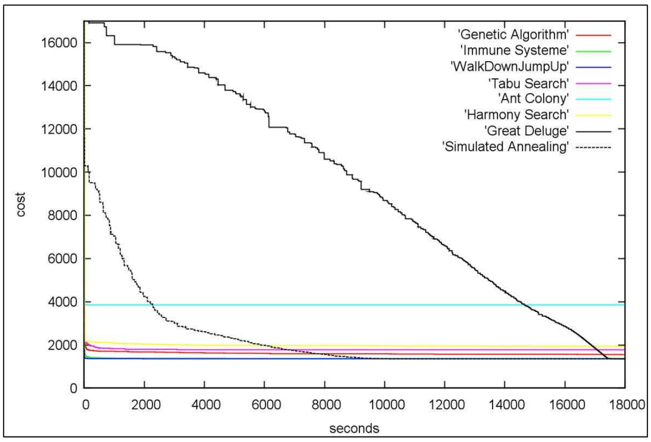


Fig. 2. A typical time table

3 Outlook

EATTS will be extended from time to time providing new algorithms and improved performance.

References

1. Wilke, P.: "The Erlangen Advanced Time Tabling System (EATTS) Version 5", Proceedings PATAT 2010, Belfast 10.-13. Aug. 2010, Springer, 2010
2. Ostler, J.; Wilke, P.: "The Erlangen Advanced Time Tabling System (EATTS) Unified XML File Format for the Specification of Time Tabling Systems", Proceedings PATAT 2010, Belfast 10.-13. Aug. 2010, Springer, 2010
3. Killer, Helmut; Wilke, P.: "Walk Up Jump Down - a new Hybrid Algorithm for Time Tabling Problems ", Proceedings PATAT 2010, Belfast 10.-13. Aug. 2010, Springer, 2010
4. Killer, Helmut; Wilke, P.: "Comparison of Algorithms solving School and Course Time Tabling Problems using the Erlangen Advanced Time Tabling System (EATTS)", Proceedings PATAT 2010, Belfast 10.-13. Aug. 2010, Springer, 2010

A Unified Format for Defining Timetabling Problems

J. Ostler

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. We describe an unified format for defining different timetabling problems, like university, curriculum based or rostering timetabling problems.

1 Introduction

Timetabling problems are a part of our daily life. In some cases the effort to find a good time table is quite high. For example the organisation of the lectures of an university or of the shifts of a factory takes a lot of time. So automatic generation of the time tables should be very useful.

There are some tools to solve special timetabling problems. These tools use a set of optimization algorithms for finding optimal time tables. To compare the different tools and algorithms a benchmark format would be helpful.

We introduce an unified XML format for defining a large quantity of timetabling problems. Benchmarks specified using the EATTS framework can be solved by different optimization algorithms. So the algorithms and their results can be compared using various input data sets.

2 Resources, Events and Constraints

Our basic concept for defining timetabling problems is based on resources, events and constraints. A time table, respectively a plan, is the assignment of resources of different resource types to a list of events. The resource types define a set of attributes. The resources are determined by the corresponding attribute values. Resources of the same type can be grouped into resource collections. For every event and resource type a collection of resources must be declared which is assigned fixed to the event and a set of resources which could be assigned.

A feasible solution of a timetabling problem must satisfy the hard constraints. Soft constraints should be satisfied. The quality of a plan is given by a penalty point value. Every violation of a constraint generates penalty points. The sum of the penalty points of all constraint violations is the quality of the solution.

3 School Timetabling

In the case of curriculum based school timetabling the resource types are typically teacher, class, student, subject and room. The resource types and the corresponding attributes are shown in fig. 1.

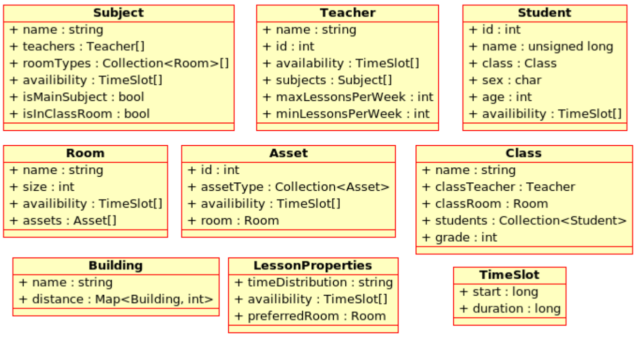


Fig. 1. Resource Types of School Timetabling Problem

The resources are the teachers, classes and students of the school. The events are the lessons. Examples of typical constraints are: a teacher can give only one lesson at the same time or that the teacher must be legitimated to give the subject of the lesson.

Figure 2 shows the definition of an event, here the math lesson of class 5b.

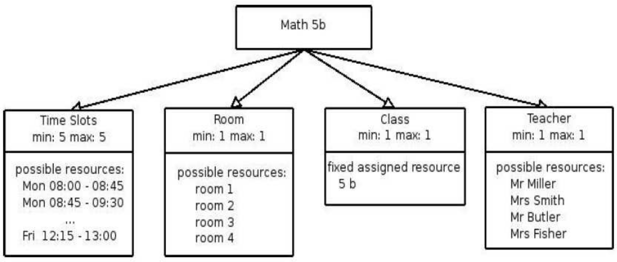


Fig. 2. Declaration of Event "Math 5b"

The associated class 5b is assigned fixed. The algorithm can select one teacher from the teacher list and one room. Math should be given five times per week, so five time slots should be assigned to the event.

4 Outlook

We will use out XML format to describe many different time tabling problems. This will lead to a library of benchmark problems and will also detect the issues where our format has to be extended to ensure it still is capable to describe all sorts of time tabling problems.

References

1. Ostler, J.; Wilke, P.: "The Erlangen Advanced Time Tabling System (EATTS) Unified XML File Format for the Specification of Time Tabling Systems" Proceedings PATAT 2010, Springer, 2010
2. Wilke, P.: "The Erlangen Advanced Time Tabling System (EATTS) Version 5" Proceedings PATAT 2010, Springer, 2010
3. Killer, Helmut; Wilke, P.: "Walk Up Jump Down - a new Hybrid Algorithm for Time Tabling Problems " Proceedings PATAT 2010, Springer, 2010
4. Killer, Helmut; Wilke, P.: "Comparison of Algorithms solving School and Course Time Tabling Problems using the Erlangen Advanced Time Tabling System (EATTS)" Proceedings PATAT 2010, Springer, 2010

Robust Single-Shot Structured Light

C. Schmalz[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Siemens CT HW2, Munich, Germany

Abstract. Structured Light is a well-known method for acquiring 3D surface data. Single-shot methods use only one pattern and require only one image of the scene. Thus, they make it possible to measure even moving objects with simple and compact hardware setups. We present a new graph-based algorithm for decoding images of a scene illuminated by a single-shot color stripe pattern. It is more robust than previous approaches and works reliably on traditionally difficult scenes with strong textures and low contrast.

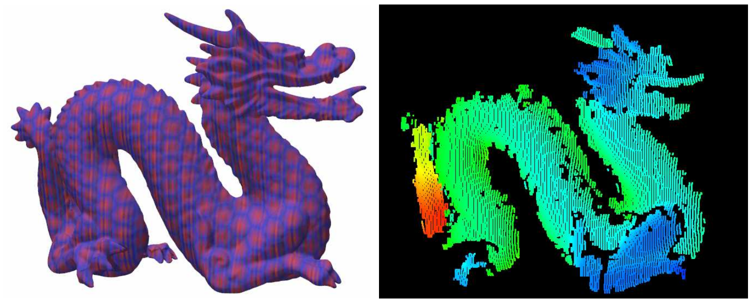


Fig. 1. Example input image and color coded depthmap result

1 Introduction

The underlying idea of Structured Light methods is to project a known illumination pattern on a scene. Range data is then generated by triangulating the camera ray and the projection ray to a point in the scene. This requires solving the correspondence problem: determining which pairs of rays belong together. In single-shot variants, the necessary information is encoded in a spatial neighborhood of pixels [1].

One image of a scene illuminated by such a pattern suffices to generate 3D data. This makes it particularly applicable to moving scenes. The need for only a single pattern also allows the use of simpler hardware, which in turn allows easy miniaturization. This can then lead to Structured Light-based 3D video endoscopes for medical, as well as industrial applications.

2 Graph-based Decoding

The robust pattern decoding starts by generating a superpixel representation of the input image. This is achieved with a watershed transformation. We then build a region adjacency graph. Each basin of the watershed segmentation corresponds to a vertex of the graph. Its color is determined by a statistically robust rank-filter over the image pixels covered by the basin. Neighboring basins are connected by edges in the graph. Edges usually connect vertices of different color. Given the knowledge about the projected pattern, there is only a finite number of possibilities of how the color can change between connected vertices. We compute the probabilities for each color change given the observed color difference. To solve the correspondence problem we look for unique paths of edges. Once found, the correspondences are propagated through the region adjacency graph.

3 Conclusion

We have implemented an improved algorithm for decoding images of Single-Shot color stripe patterns. It generates up to 10^5 data points from an input image with a resolution of 780x580 pixels. It runs at 20Hz and works even under adverse conditions, like highly textured or volume-scattering objects and low contrast illumination.

References

1. Salvi, Joaquim, Pags, Jordi and Batlleet, Joan, Pattern Codification Strategies in Structured Light Systems, Pattern Recognition, 37, 827-849 (2004).

An Epipolar Approach for Visual Patient Detection: An Overview

P. W. Mewes[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Workflow and Solutions, Special Products, Healthcare Sector, Siemens AG, Allee
am Röthelheimpark 2, 91052 Erlangen, Germany

Abstract. Tracking of the patient position in imaging devices such as MRIs or CTs can mainly be done using two different approaches. One uses the imaging device itself to localize the patient's position and get an overview on relevant regions. Another approach involves the use of external trackers, such as TOF cameras or electromagnetic tracking systems that feed their output to the imaging devices. The use of external trackers has certain advantages such as: no exposure to additional radiation and usually instant tracking feedback, which reduces the procedure time. Nevertheless, there are problems that need to be addressed, when using external tracking techniques: a) Some trackers are affected by the strong magnetic fields or radiation. b) Most of the time imaging devices do not offer a lot of space. Thus, there is typically a very constraint viewpoint to most of the patients body. We propose a vision based approach which uses two off-the-shelf cameras that combine stereo vision and structure-from-motion to recover the shape and the position of the patient. Within this work we are addressing the issues and technical limitations arising from this kind of approach.

Our technique combines two methods of computer vision to assure a reliable reconstruction of the patient shape and his position in the related coordinate system of an imaging device. Our approach consists of four major steps: Calculation of corresponding feature points from multiple time-adjacent frames; coarse estimation of partial structures of the patient by solving the epipolar equation using structure from motion (SfM); estimation of the remaining structure of the patient using dense stereo techniques; and combination of different partial structure pieces to a full-body shape recognition. These steps should lead to an exact shape, orientation and position estimation, which can be used in further processing. Our setup consists of two cameras installed inside the bore and looking towards its center. Both cameras fulfill the same two tasks: SfM and dense stereo vision. In the SfM task each camera works independently, while in stereo the cameras are treated as part of a single setup.

SfM has been a central problem of computer vision for at least 20 years. Early examples of 3D scene recovery and motion-based reconstruction can be found in Poggio's and Ullman's work [3] and [4]. This work can be described as the origins of SfM. As the field evolved, epipolar geometry played a major role in structure recovery. SfM techniques based on the 8-point algorithm were described very early in [2]. The work of Hardley [5] and Faugeras [6] contributed to a deeper understanding of such mathods.

Our work lies on multi-image SfM in contrast to two-image algorithms that are the subject of most early papers. Nevertheless, both approaches need to be optimized in order to meet quality criteria such as accuracy and reliability. Furterhmore, in order to obtain reliable results one needs to analyze the error introduced by different parts of the process like sensor noise, geometry constraints, systematic errors in the image quality, etc. For two perspective views, error analysis has been addressed by Weng in [7], shortly after the 8-point algorithm became popular. For multi-image SfM, a large number of optimization methods, e.g. [9] and [8], and filtering methods, e.g. [10] and [11], have been proposed. More complete reviews about such techniques can be found in [14], [13] and [12]. We are currently adopting these theories in order to address the error issues raised by our particular sensor setup.

References

1. Juyang Weng, Thomas S. Huang and Narendra Ahuja, "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," *IEEE Trans on Pattern Analysis and Machine Intelegence*, 1989, 11:451–476.
2. Longuet-Higgins, H.C., "A computer algorithm for reconstructing a scene from two projections," *Readings in computer vision: issues, problems, principles, and paradigms*, 1987. 60.
3. Poggio, T. and Marr, D., "Cooperative computation of stereo disparity," *Science*, 1976. 194:283–287.
4. Ullman, S., "The interpretation of visual motion," *MIT press Cambridge, MA*, 1979.
5. Hartley, R., Gupta, R. and Chang, T. "Stereo from uncalibrated cameras" *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1992. 761–764
6. Hartley, R., Gupta, R. and Chang, T., "What can be seen in three dimensions with an uncalibrated stereo rig?" *Computer VisionECCV*, 1992. 563–578
7. Weng, J., Huang, TS and Ahuja, N., "Motion and structure from two perspective views: algorithms, erroranalysis, and error estimation," *IEEE transactions on pattern analysis and machine intelligence*, 1989. 11:451–476
8. Soatto, S. and Brockett, R., "Optimal structure from motion: local ambiguities and globalestimates," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1989. 11:282–288
9. Soatto, S. and Brockett, R., "A new structure-from-motion ambiguity," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on.*, 1999. 1
10. McLauchlan, P.F., "The variable state dimension filter applied to surface-based structure from motion," *Technical ReportVSSP-TR-4/99, University of Surrey*, 1999.
11. Soatto, S. and Perona, P., "Reducing Structure from Motion: a general frameworkfor dynamic vision. 2. Implementation and experimental assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998. 20:943–960
12. Oliensis, J., "A critique of structure-from-motion algorithms," *Computer Vision and Image Understanding*, 2000. 80:172–214
13. Jebara, T., Azarbayejani, A. and Pentland, A., "3D structure from 2D motion," *IEEE Signal processing magazine*, 1999. 16:66–84
14. Huang, T.S. and Netravali, A.N., "Motion and structure from feature correspondences: A review," *Proceedings of the IEEE*, 1994. 82:252–268

Reconstruction CT

Noise Reduction Techniques in CT Volume Data

M. Balda[1], B. Heismann[1,2], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Siemens Healthcare, Erlangen, Germany

Abstract. Image noise has a huge impact on the diagnostic quality of CT images. The noise in the measured data is directly related to patient dose. The dose of a CT acquisition is chosen in a way that noise may not impair the diagnostic quality. If a suitable noise reduction method lowers the noise-level without impairing other image quality factors, such as sharpness, the diagnostic quality can be maintained at lower patient dose.

1 Sources of noise

The two main sources of noise in the CT projection data is Quantum noise and electronics / detector noise. Quantum noise is directly related to the number of incoming X-ray photons. It follows a Poisson distribution. For a mean number m of incoming photons its standard deviation is $\sqrt{m}$. It can be shown, that the standard deviation in the measured attenuation values is approximately $\frac{1}{\sqrt{m}}$. See [1] pp. 397 for the complete derivation.

Detector noise has different sources such as quantization or signal amplification. It is partly independent of the signal power thus it is dominant for low input signal and negligible for high signal power / incoming photon energy.

2 Noise reduction in CT

Besides the patient dose, the reconstruction kernel influences the noise and the image quality in general. Soft kernels are used for diagnostic tasks which do not require a high image resolution. These kernels suppress noise at the loss of image detail. Sharp kernels maintain small structures but do not reduce image noise effectively. Figure 1 shows a comparison of soft hard kernels on a reconstruction of a high contrast phantom with fine resolution structures. More sophisticated noise reduction methods try to maintain as much image detail as possible while reducing image noise. This process requires analytic techniques to assess the noise and image properties in order to locally adapt noise filters to the local properties [2]. Well known examples from image processing are bilateral filtering [3] or anisotropic diffusion filtering [4]. These methods can be adapted to CT data [5].

In CT, two filtering strategies are possible: Filtering can be applied in the measurement / projection domain (pre-reconstruction) or in the image / volume domain (post-reconstruction).

Pre-reconstruction filtering has the advantage that noise properties in the measurement domain can be described analytically and the influences of the measurement

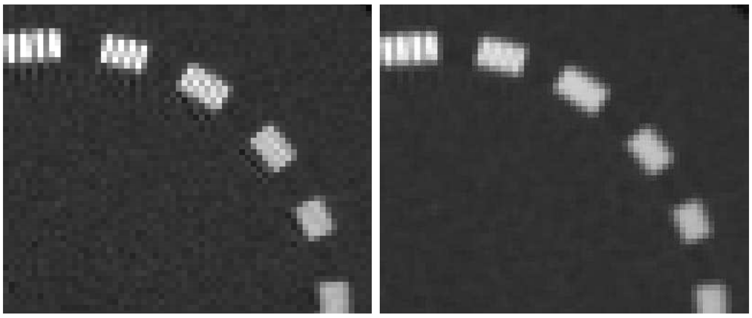


Fig. 1. Hard- (left) and soft (right) reconstruction kernel examples.

system are well known, however, the signal-to-noise (SNR) in the projection values is rather bad, so the signal properties can hardly be exploited.

Post-reconstruction methods suffer from the opposite phenomena: The noise properties in the image domain are complex. The noise is inhomogeneous and non-stationary. The reconstruction process introduces various noise correlations which are heavily dependent on the properties of the reconstruction process, so noise adaption becomes complex as well.

Acknowledgement

This research project was funded by Siemens Healthcare, Forchheim, Germany.

References

1. Buzug, T. M.: Einführung in die Computertomographie: Mathematisch-physikalische Grundlagen der Bildrekonstruktion, Springer, Berlin, 2003.
2. Borsdorf, A., Kappler, S., Raupach, R., Hornegger, J.: Analytic Noise Propagation for Anisotriopic Denoising of CT Images, 2008 IEEE Nuclear Science Symposium Conference Record., Dresden, 2008, 5335–5338
3. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images, Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, 1998, 839–846
4. Weickert, J.: Anisotropic Diffusion in Image Processing, Teubner-Verlag, Stuttgart, 1998
5. Wang, J., Lu, H., Li, T., Liang, Z.: Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters, Volume 5747., SPIE Medical Imaging, San Diego, USA, 2005, 2058–2066

CT Reconstruction Using Compressed Sensing

H. Wu[1,2], and J. Hornegger[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany

Abstract. In computed tomography, rapid scanning with a reduced X-ray dose delivered to the patient can be achieved by reconstructing image from few views. There are challenges due to the under-sampling and insufficient data. To solve this problem, compressed sensing based methods are under development. The algorithm reconstructs images by solving a constrained optimization problem.

1 Introduction

Computed tomography (CT) is considered as a radiation-intensive procedure, yet it becomes more and more common. The risk of radiation induced disease makes a strong need to reconstruct CT images with practically useful quality using as low radiation as possible[1]. One strategy is to use limited number of projections. Mathematically, the procedure of the CT reconstruction can be described as a linear problem

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{1}$$

where $\boldsymbol{A}$ is the system matrix which describes each X-ray path, $\boldsymbol{x}$ is the image we want to reconstruct and $\boldsymbol{y}$ is the observed value on the detector. To reduce the radiation dose, the projection images can be only acquired at few views. The major challenge is that the number of observations is much less than the number of unknowns.

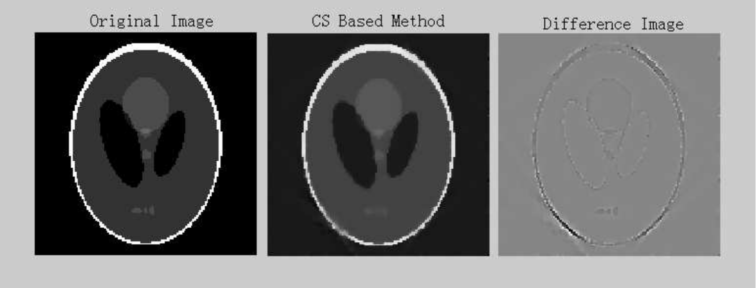


Fig. 1. Reconstruction results

2 Compressed Sensing Based CT Reconstruction Method

Recently, Candes, Romberg and Tao proposed a new mathematical theory, compressed sensing(CS), which can perform nearly exact image reconstruction with few measurements[2]. Based on this theory, we aim to develop a CT reconstruction method requiring as few projection images as possible. According to the framework of compressed sensing, the image should firstly be sparsified and then be reconstructed by a constrained L1 norm minimization method. We used total variation as the sparsifying transformation and tested the algorithm using the Shepp-Logan phantom. The reconstructed image using only 20 projection images is almost the same compared to the original image[Fig 1]. Since, firstly, the image used to reconstruct is sparse after the total variation transform; secendly, the L1 norm has a sparsity promoting property.

3 Outlook

Although images reconstructed by the compressed sensing based method have a high image quality, the method is quite time consuming. In our experiments, half hour is needed to reconstruct the results.(The algorithm is implemented with matlab running on the computer with 2G ram and 2.0 GHZ cpu) In the future, to meet the needs of practical application, a fast optimization scheme will be investigated by us.

4 Acknowledgements

The first author is finaced by Chinese scholarship council. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

References

1. Cho, S., Sidky, E. Y., Bian, J., Pelizzari, C. A., Pan, X.: A preliminary investigation of using prior information for potentially improving image reconstruction in few-view CT, Medical Imaging, vol. 6913 of Proc. SPIE, p. 69132C.
2. Candes, E., Romberg, J., Tao, T.: Robust Uncertainty Principles : Exact Signal Reconstruction from Highly Incomplete Frequency Information. IEEE Transactions on Information Theory In Information Theory, IEEE Transactions on, Vol. 52, No. 2. (23 February 2006), pp. 489-509.

Tissue Perfusion Quantification with C-arm CT

A. Fieselmann[1,2,3], Y. Deuerling-Zheng[3], and J. Hornegger[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), University
Erlangen-Nuremberg, Germany
[3] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. Interventional perfusion imaging using a C-arm CT could optimize the clinical workflow during catheter-guided stroke therapy.

For a successful treatment of stroke a fast and accurate diagnosis is important. Perfusion imaging using computed tomography (CT) or magnetic resonance imaging (MRI) is an established method to quantify the regional cerebral blood flow for stroke diagnosis.

In this project we investigate the use of a C-arm angiography system (C-arm CT) to measure cerebral blood flow. A C-arm CT can acquire tomographic images during an intervention while the patient is positioned on the treatment table. This technique has a huge potential benefit in the clinical workflow. For example, it would be possible to monitor the current state of perfusion immediately before or during the intervention without moving the patient to a CT or MRI scanner.

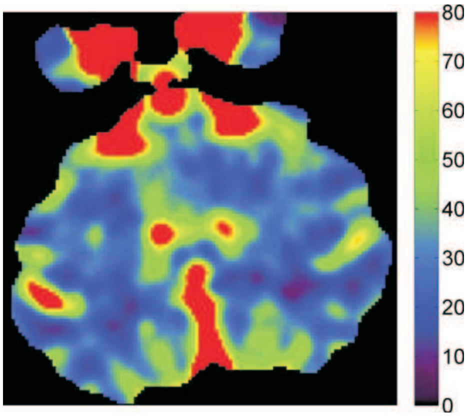


Fig. 1. Example cerebral blood flow parameter map with units in ml/100g/min

During perfusion imaging a region of interest is imaged several times at short intervals to study the flow of a previously injected bolus of contrast agent. A blood flow parameter map (Fig. 1) can be computed from the time attenuation curves. Using a C-arm CT for perfusion quantification faces certain challenges. While a conventional CT has a rotational speed of the order of 0.5 s / 360° the rotational speed of the C-arm CT X-ray source and detector pair is about 4-5 s / 180°. Thus inconsistencies in the projection images due to the changing attenuation values from the contrast agent can cause reconstruction artifacts.

One focus of this project is the development of dynamic image reconstruction algorithms. Fig. 2 shows an illustration of a dynamic reconstruction approach [1]. A consistent set of projection data at time t_{est} is estimated by interpolation of data from neighboring sweeps.

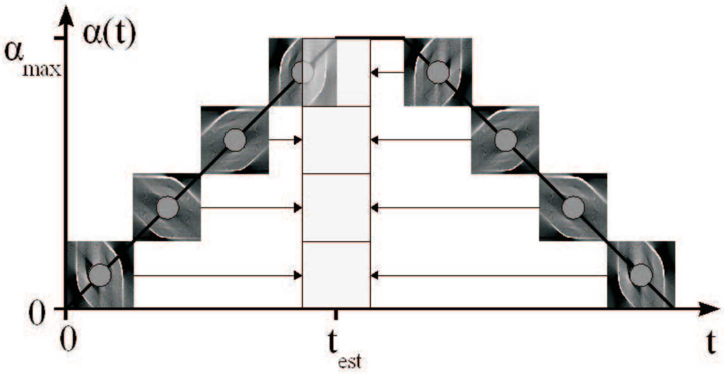


Fig. 2. The X-ray source angle $\alpha(t)$ is plotted against time t. In order to approximate consistent data at a certain time the data from neighboring sweeps is used to estimate a complete set of data.

References

1. Fieselmann A., Ganguly A., Deuerling-Zheng Y., Zellerhoff M., Boese J., Hornegger J. and Fahrig R. A Dynamic Reconstruction Approach for Cerebral Blood Flow Quantification With an Interventional C-arm CT In: Proc. IEEE ISBI, April 2010, Rotterdam/The Netherlands, p. 53-56
2. Fieselmann A., Ganguly A., Deuerling-Zheng Y., Zellerhoff M., Marks M., Boese J. and Fahrig R. Volume Cerebral Blood Flow (CBF) Measurement Using an Interventional Ceiling-Mounted C-arm Angiography System In: Proc. ECR, March 2010, Vienna/Austria, p. 127
3. Ganguly A., Fieselmann A., Boese J., Rohkohl C., Hornegger J. and Fahrig R. Evaluating the Feasibility of C-arm CT for Brain Perfusion Imaging: An in vitro Study In: Proc. SPIE Medical Imaging, February 2010, San Diego/CA, p. 76250K

Interventional Imaging using C-arm Systems

C. Rohkohl[1], G. Lauritsch[2], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. The vision of our work is to develop methods that open up the possibility to visualize highly dynamic (moving) vital parts of the human body as the heart. This provides one key towards the development of novel diagnostic tools and medical treatment procedures. This paper provides an overview of our work in this research area.

1 Introduction

During the recent years 3-D imaging has become a standard tool for diagnosis and planning in the clinics using well-known imaging modalities as CT or MRT. The integration of those image during an intervention is difficult and requires various registration techniques. A solution that has emerged in recent years is the use of fluoroscopic C-arm systems which are highly common in the interventional field for reconstructing 3-D images [Hett 09]. The workflow can be seen in figure 1 from one of our clinical cooperations (Klinikum Coburg GmbH):

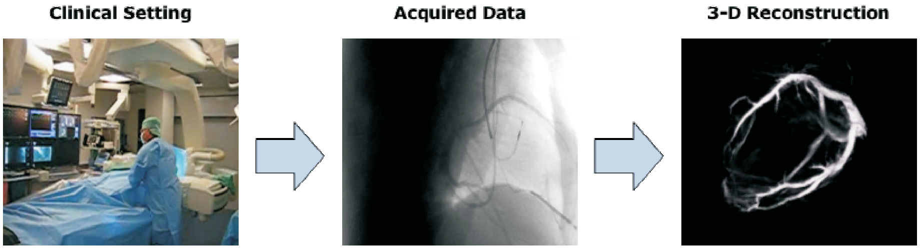


Fig. 1. Clinical Workflow

The general goal of my research topics is to enrich that workflow for mainly cardiac applications by providing high quality, fast and reliable algorithms for the 3-D reconstruction of challenging subjects. Of course this work is highly interdisciplinary and requires techniques and concepts from the complete field of medical image processing.

2 Non-Periodic Motion Estimation and Motion Compensation Algorithms

The primary goal of our research activities in this project is the development of methods for estimating and correcting cardiac motion in order to increase the image quality of cardiac vasculature reconstruction. In interventional environments patients often do have arrhythmic heart signals or cannot hold breath during the complete data acquisition. This important group of patients cannot be reconstructed with current approaches that do strongly depend on a high degree of cardiac motion periodicity for working properly. In this project we try to develop novel algorithmic approaches to cardiac vasculature reconstruction and therefore address the following questions [Rohk 09b,Rohk 09c,Rohk 09d]:

- Development of algorithms for motion estimation and reconstruction without periodicity assumption or ECG information.
- Analysis and development of models for the description of cardiac and breathing motion.
- Development of algorithms for estimatating the optimal heart phase for phase-correlated reconstruction algorithms.
- Development of methods for the segmentation and functional assessment from 3-D plus time image data.

References

[Rohk 09b] C. Rohkohl, B. Keck, H. Hofmann, and J. Hornegger. “RabbitCT - an open platform for benchmarking 3D cone-beam reconstruction algorithms”. *Medical Physics*, Vol. 36, No. 9, pp. 3940–3944, 2009.

[Rohk 09c] C. Rohkohl, G. Lauritsch, M. Prümmer, and J. Hornegger. “Interventional 4-D Motion Estimation and Reconstruction of Cardiac Vasculature without Motion Periodicity Assumption”. In: G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, Eds., *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, pp. 132–139, Springer, London, UK, September 2009.

[Rohk 09d] C. Rohkohl, G. Lauritsch, M. Prümmer, and J. Hornegger. “Towards 4-D Cardiac Reconstruction without ECG and Motion Periodicity using C-arm CT”. In: *10th Int. Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, pp. 323–326, Beijing, China, September 2009.

[Hett 09] H. Hetterich, T. Redel, G. Lauritsch, C. Rohkohl, and J. Rieber. “New X-ray imaging modalities and their integration with intravascular imaging and interventions”. *The International Journal of Cardiovascular Imaging*, doi: 10.1007/s10554-009-9529-z, 2009.

On Hardware Acceleration Techniques for 3-D Cone-Beam Reconstruction

B. Keck, H.G. Hofmann, C. Rohkohl, and J. Hornegger

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Fast 3-D cone beam reconstruction is mandatory for many clinical workflows. For that reason researchers and industry work hard on hardware-optimized 3-D reconstruction. Backprojection is a major component of many reconstruction algorithms which requires a projection of each voxel onto the projection data, including data interpolation, before updating the voxel value. This step is the bottleneck of most reconstruction algorithms and the focus of optimization in recent publications.

1 Introduction

In a collaboration between the Pattern Recognition Lab, Siemens Healthcare (CV Division) and Intel Corp., innovative developments of computed tomography (CT) algorithms and their implementation on different hardware architectures have been investigated. The main interest lies on (iterative) algorithms for 3-D reconstruction in C-arm systems including necessary pre- and post-processing steps.

2 Hardware acceleration techniques and comparison

Reconstruction methods in CT require an immense computing power of the reconstruction system, because of the amount of projection data to be processed and also because of the number and size of the slice images to be computed. In order to meet the requirements for the computing time of productive systems, it is still necessary to use high performance hardware components. In our research project, we focus on parallel implementations of different reconstruction algorithms on multi processor systems [1,2], graphics accelerator boards [3] (NVIDIA CUDA, Intel Larrabee), specialized processors [4] (IBM Cell), as well as on reconfigurable accelerator hardware based on Field Programmable Gate Array (FPGA) technology [5].

So far the optimized implementation of iterative reconstruction algorithms and the Feldkamp method on GPUs using CUDA have been developed [6]. Since a comparison of published results in this field is hardly possible a standardized data set and corresponding performance measurement software [7] were released. We provide an open platform for worldwide comparison in backprojection performance and ranking on different architectures using a specific high resolution C-arm CT dataset of a rabbit (Figure 1). This includes a sophisticated benchmark interface, a prototype implementation in C++ and image quality measures. The performance of developed highly optimized many-core CPU implementations was published on the accompanying website "www.rabbitct.com". Furthermore, the achieved insights were transferred to Intel's new hardware architecture, Larrabee, and extended by techniques to exploit its novel features.

Fig. 1. Volume rendering of the reconstructed rabbit using RabbitCT

3 Results & Summary

We compared optimized implementations of 3-D reconstruction algorithms for C-arm CT systems. The performance of optimized CPU [2] and GPU [3] implementations of the backprojection was compared using the RabbitCT[1] benchmark [7]. Its publicly available dataset consists of 496 projections with 1240 × 960 pixels each. The CPU-based system used for comparison was a workstation equipped with a Intel Core2 Extreme QX9650 at 3.00GHz. For the GPU measurements we utilized a NVIDIA QuadroFX 5600 graphics card with CUDA 2.0. The graphics card performed the reconstruction of a 512^3 volume in $14.69s$, 13.5 times faster than an the CPU.

An hardware accelerated version of the Simultaneous Algebraic Reconstruction Technique (SART) [6] was used as a representative of the class of iterative reconstruction algorithms. We used 228 projections representing a short-scan from a C-arm CT system to perform the iterative reconstruction with a projection size of 256×128 pixels. The reconstruction yields a $512 \times 512 \times 350$ volume. A workstation equipped with two Intel Xeon QuadCore processors at 2.33 GHz was compared against a NVIDIA QuadroFX 5600 graphics card using CUDA 2.0. In this case, the SART implementation on the GPU can outperform the optimized CPU implementation by a factor of 8. To summarize, we have shown that GPUs are a viable accelerator platform for C-arm CT reconstruction.

---

[1] http://www.rabbitct.com

References

1. H. G. Hofmann, B. Keck, and J. Hornegger, “Accelerated C-arm Reconstruction by Out-of-Projection Prediction” in Bildverarbeitung für die Medizin 2010 (T. M. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, eds.), pp. 380-384, Berlin (Germany), Mar 2010.
2. H. G. Hofmann, B. Keck, C. Rohkohl, and J. Hornegger, “Putting 'p' in RabbitCT - Fast CT Reconstruction Using a Standardized Benchmark”. In: GI/ITG-Fachgruppe PARS (Eds.) Mitteilungen - Gesellschaft fr Informatik e.V., Parallel-Algorithmen und Rechnerstrukturen, PARS (22. PARS - Workshop), Vol. 26, pp. 91-100, Parsberg (Germany), Jun 2009.
3. H. Scherl, B. Keck, M. Kowarschik, and J. Hornegger. Fast GPU-Based CT Reconstruction using the Common Unified Device Architecture (CUDA). In Eric C. Frey, editor, Nuclear Science Symposium, Medical Imaging Conference 2007, pages 4464-4466, Honolulu, Hawaii (USA), Nov 2007.
4. H. Scherl, M. Koerner, H. Hofmann, W. Eckert, M. Kowarschik, and J. Hornegger. Implementation of the FDK Algorithm for Cone-Beam CT on the Cell Broadband Engine Architecture. In J. Hsieh and M. J. Flynn, editors, Proceedings of SPIE, volume 6510, San Diego (USA), Feb 2007.
5. B. Heigl and M. Kowarschik, “High-speed reconstruction for C-arm computed tomography” in Proc. Int. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Fully 3D) and Workshop on High-Performance Image Reconstruction (HPIR), pp. 25-28, Lindau (Germany), Jul 2007.
6. B. Keck, H. Hofmann, H. Scherl, M. Kowarschik, and J. Hornegger, “GPU-accelerated SART reconstruction using the CUDA programming environment,” in *Proceedings of SPIE*, E. Samei and J. Hsieh, Eds., vol. 7258, Lake Buena Vista (USA), Feb 2009.
7. C. Rohkohl, B. Keck, H. G. Hofmann, and J. Hornegger: “RabbitCT — An Open Platform for Benchmarking 3-D Cone-beam Reconstruction Algorithms” Medical Physics, vol. 36 (9), pp. 3940-3944, Sep 2009.

Accelerated C-arm Reconstruction by Out-of-Projection Prediction

H.G. Hofmann, B. Keck, and J. Hornegger

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. 3-D image reconstruction from C-arm projections is a computationally demanding task. However, for interventional procedures a fast reconstruction is desirable. The presented method reduces the number of actually processed voxels and eventually the runtime by 20–30% without affecting image quality. The proposed method detects and skips subvolumes that are not visible in the current view. It works with projection matrices and is thus capable of handling arbitrary geometries. It can easily be incorporated into existing algorithms and also into the back-projection step of iterative algorithms.

1 Introduction

Interventional procedures require fast 3-D reconstruction of tomographic data. The state of the art algorithm for reconstruction of cone-beam data in a clinical environment is the FDK method by Feldkamp et al.

Yu et al. [1] proposed to reconstruct only a region of interest, approximating the Field-of-View (FOV) of the scan, to reduce the number of back-projected voxels. Their method, however, requires a priori knowledge about the trajectory and offline computation of the FOV. They further suggested data partitioning, similar to the subvolumes approach used e.g. by Scherl et al. [2] and Kachelriess et al. [3], to fit the problem into the 256 KB-sized local stores of the Cell processor. They also proposed that subvolumes improve performance on CPUs due to better caching. However, they did not actually skip subvolumes as in this work.

Our method adapts dynamically to the FOV of each projection during reconstruction. It uses projection matrices from the scanner and does not require prior knowledge about the acquisition trajectory. It is easy to integrate into existing reconstruction algorithms and does not affect image quality.

2 Materials and Methods

In cone-beam CT, the source and the corners of the detector define a pyramid, the current FOV. Areas of the object that are outside of the FOV are not captured by the detector. Hence, the current view does not contain information about them.

Algorithm 1 shows our modified back-projection step. First, we added the partitioning into subvolumes. Second, before a subvolume is back-projected it is tested whether it is inside the current FOV. Therefore, the subvolume's corner voxels are projected and its shadow determined. If the shadow overlaps with the detector we perform a normal back-projection. Otherwise, the whole subvolume is skipped.

Fig. 1. Modified back-projection step, called once for each projection image I_n.

```
Partition volume $f_{\mathrm{FDK}}$ into subvolumes $f_{\mathrm{FDK},i}$
for all subvolumes $f_{\mathrm{FDK},i}$ do
  // Test if $f_{\mathrm{FDK},i}$ is visible
  Compute shadow of $f_{\mathrm{FDK},i}$ by projecting corners
  if $\emptyset == \mathit{shadow} \cap I_n$ then
    Skip $f_{\mathrm{FDK},i}$
  else
    for all voxels $\underline{x} \in f_{\mathrm{FDK},i}$ do
      Project voxel onto detector plane
      Check if point lies on detector
      Fetch projection values
      Bilinear interpolation
      Update $f_{\mathrm{FDK}}(\underline{x})$
    end for
  end if
end for
```

Since the proposed method skips only subvolumes not visible in the current projection the result is identical to the reference method. The reconstruction benchmark RabbitCT [4] was used to test our implementation for correctness and to measure its performance. Obviously, the performance of our method depends on the acquisition geometry. The method was evaluated on the the public RabbitCT dataset and two other clinical datasets (cmp. table 1). All datasets were acquired with real C-arm systems using clinical protocols.

The new method was incorporated into our vectorized and multi-threaded FDK implementation for CPUs [5] to that it also improves the performance of highly optimized code. Performance was tested on a multi-core system with four hexa-core Intel Xeons processors (X7460) at 2.66 GHz and 32 GB of RAM. The system used in this report was a preproduction system. We expect production hardware to deliver similar performance levels.

3 Results

The reconstructed volume had a size of 512^3 voxels. 48 different subvolume sizes were tested and 128×64×4 performed overall best, only 5% slower in the worst case. The large size in x-direction is justified by the fact that neighboring voxels in this direction are stored successively in memory and therefore supportive for the hardware prefetching mechanism.

The reconstruction times of two different implementations are shown in table 1. The baseline ("base" in tbl. 1) is defined by our previous optimized implementation [5]. The results of the method proposed in this work are presented in the rightmost column ("skip"). The reconstruction time was reduced to about 53–58% compared to the baseline version (1.73–1.89× speedup) on this bandwidth starving system.

4 Discussion

The savings of the proposed method depend on the acquisition geometry, namely the ratio between volume and FOV sizes. To show its relevance the method was tested

Dataset	Number of projections	Projection size	X7460 base	X7460 skip
A	496	1240 × 960	60.53	31.96
B	543	1240 × 960	66.04	35.08
C	414	1024 × 1024	50.59	29.21

Table 1. Description of the three datasets. Further, reconstruction times (seconds) are shown for two implementations (“base” and “skip”).

using three different clinical datasets. The proposed method delivers a considerable speedup even to highly optimized implementations. Although all datasets used here were circular trajectories the method is suitable for arbitrary trajectories. It requires no prior knowledge other than the projection matrices. Since it works on-the-fly it can easily be incorporated into existing algorithms.

References

1. Yu, R. ; Ning, R. ; Chen, B.: High speed cone beam reconstruction on PC. Sonka, M. ; Hanson, K.M. ; eds.: Medical Imaging 2001: Image Processing. Proceedings of the SPIE. Volume 4322., San Diego (2001) 964–973
2. Scherl, H. ; Koerner, M. ; Hofmann, H. ; Eckert, W. ; Kowarschik, M. ; Hornegger, J.: Implementation of the FDK algorithm for cone-beam CT on the cell broadband engine architecture. Hsieh, J. ; Flynn, M.J. ; eds.: SPIE Medical Imaging. Volume 6510., SPIE (2007) 651058
3. Kachelrieß, M. ; Knaup, M. ; Bockenbach, O.: Hyperfast parallel-beam and conebeam backprojection using the cell general purpose hardware. Medical Physics 34(4) (2007) 1474–1486
4. Rohkohl, C. ; Keck, B. ; Hofmann, H.G. ; Hornegger, J.: Technical note: Rabbitct—an open platform for benchmarking 3d cone-beam reconstruction algorithms. Medical Physics 36(9) (2009) 3940–3944
5. Hofmann, H.G. ; Keck, B. ; Rohkohl, C. ; Hornegger, J.: Putting ’p’ in RabbitCT – Fast CT reconstruction using a standardized benchmark. PARS-Mitteilungen (ISSN 0177-0454), GI, Gesellschaft für Informatik e.V. (2009) 91–100

PHACT - Phase-Contrast Tomography

W. Haas[1,2], P. Bartl[2], F. Bayer[2], J. Durst[2], T. Michel[2], A. Ritter[2], and T. Weber[2],
G. Anton[2], J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] ECAP, Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Abstract. X-rays contain not only information about the absorption properties of a sample but also about the influences of the object on the phases of the X-ray beams. In the last decades many different approaches have been invented to measure this information. One of these approaches is called the Talbot-grating interferometry which is subject of our research. It is one of the first phase-contrast imaging systems which has the potential to be applied in a medical environment.

1 Introduction

Phase-contrast imaging is an approach to gain information about a sample by measuring the local phase of X-rays passing through it. Nowadays experimental techniques have become applicable not only for optical wavelengths but also for wavelengths in the X-ray range [1]. One of them is the grating based interferometry [2], [3], which resolves both the phase and the absorption of the investigated sample.

2 Grating-based Interferometry

A Talbot-grating interferometer [2] (fig. 1) consists of three gratings the so called source grating G0, phase grating G1 and analyzer grating G3. Each grating has a specific period p. Without an object in the path of the X-ray beams the generated wave is modulated by the phase grating G1 so that a periodic rectangular interference pattern downstream towards G2 is formed. This effect is called the *Talbot-effect* or *self imaging*. Since this pattern is too small to be resolved directly by a detector a second grating G2 is needed [2]. This grating is required for a process called *phase stepping* [4]. Phase stepping means the grating G2 is moved a fraction of its own periodicity. This will be repeated until G2 was shifted by a whole number of its periodicity. During phase stepping each detector pixel observes an intensity modulation which corresponds in an ideal case to a triangle function. Figure 2 shows as solid graph an ideal intesity modulation with respect to the grating position G2.

If an object is placed in the path of the X-ray beams, the interference pattern will be distorted. This distortion is given by a local angular deviation of the pattern. For X-rays the angle is typically in the order of a few micro-radians. Again, during the phase stepping the pixels observe a triangle function but in this case the triangles are laterally shifted (dashed graph in fig. 2). The shift between the images without and with an object is the desired measure of the angular deviation.

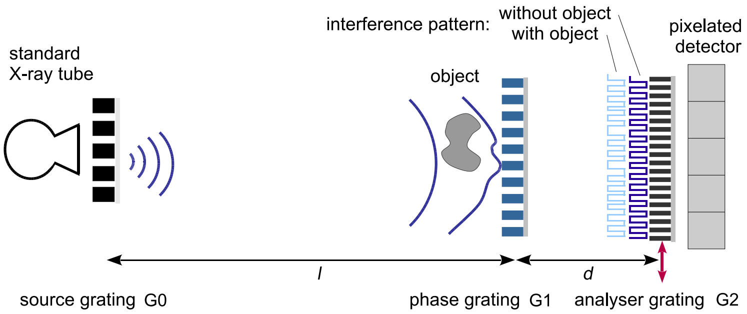


Fig. 1. Schematic drawing of the Talbot-grating interferometer.

In order to get enough coherent X-rays an additional grating G0 [2] is needed. G0 splits up the large source into many small sources which are small enough in order to produce interference pattern. G0 allows to use an ordinary X-ray tube to acquire phase-contrast images.

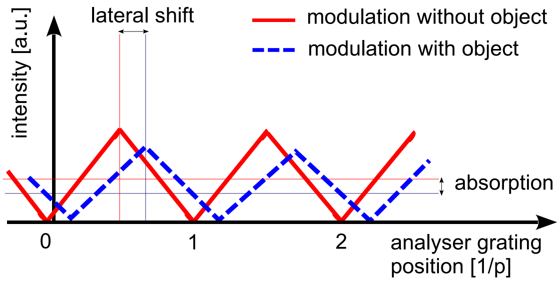


Fig. 2. Triangle function as the ideal intensity modulation resulting from phase stepping.

3 Result

Figure 3 shows the first result of a small fish taken at our lab at the ECAP Institute. This data was acquired with a 80 keV tungsten spectrum and the Medipix 2 detector. It was a symmetric setup where each grating had the period of 5 μm. G1 was designed for 40 keV. The image shows from left to right the differential phase-contrast image, absorption image and a so called dark-field image.

4 Outlook

Grating base phase-contrast imaging approach is a relatively young topic with many unknows. To understand the physics behind the scene we are going to develop a simu-

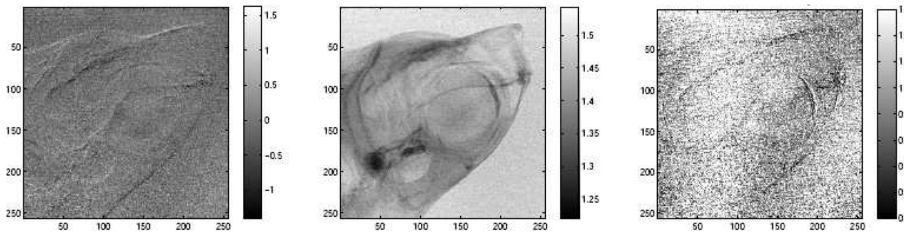


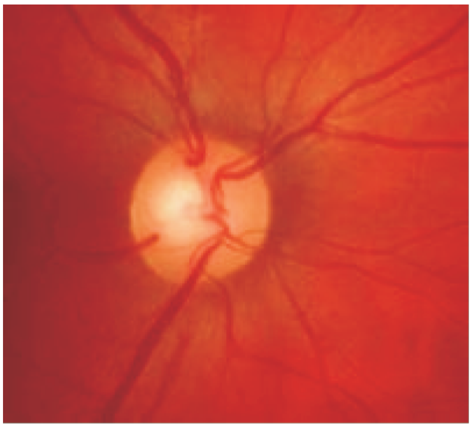
Fig. 3. First result taken at our lab showing from left to right the differential phase image, absorption image and dark-field image of a fish.

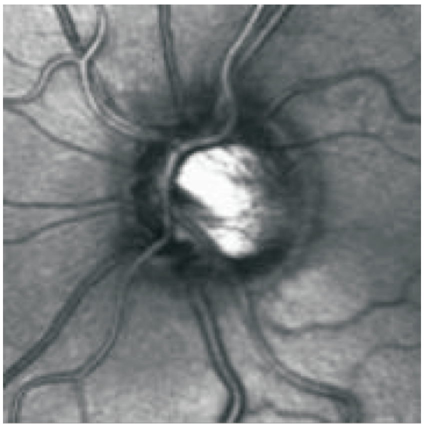
lation framework which additionally enables us to simulate all possible configurations and setups to get a better idea of how all the components work and influence the quality of the signal. Hot topic is to prove the applicability in a medical environment and also compare the phase-contrast with the ordinary absorption based imaging approach.

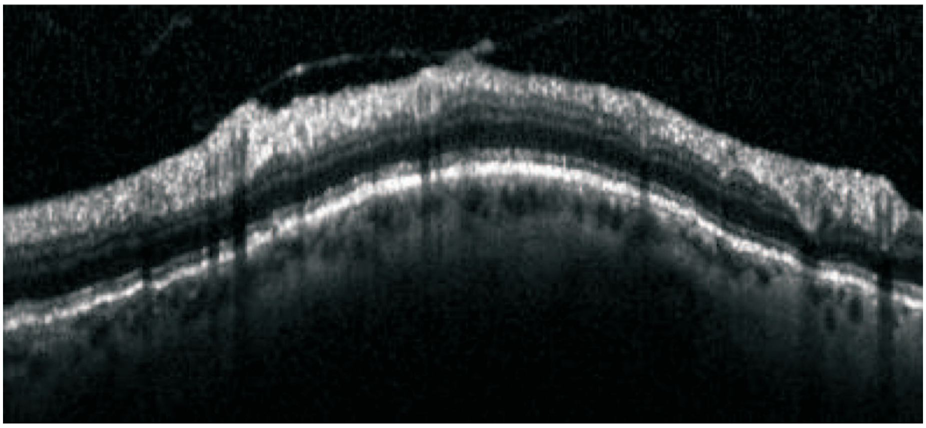
References

1. S.W. Wilkins, T.E. Gureyev, D. Gao, A. Pogany, A.W. Stevenson, “Phase-contrast imaging using polychromatic hard X-rays,” *Nature* 384, 335–337, 1996.
2. F. Pfeiffer, T. Weitkamp, O. Bunk, C. David, “Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources,” *Nature* Phys. 2, 258–261, 2006.
3. A. Momose, W. Yashiro, Y. Takeda, Y. Suzuki, T. Hattori, “Phase tomography by X-ray Talbot interferometry for biological imaging,” *Jpn. J. Appl. Phys.*, 45 (6A), 5254–5262, 2006.
4. Weitkamp, T., et al., “X-ray phase imaging with a grating interferometer,” *Optics Express* Vol. 13, No. 16, 6296–6304, 2005.

Eye - OCT, DTI and Fundus Imaging







Statistical Deformation Modeling of the Optic Disk

R. Bock[1,2], J. Hornegger[1,2], and G. Michelson[2,3,4]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
[3] Department of Ophthalmology
[4] Interdisciplinary Center of Ophthalmic Preventive Medicine and Imaging (IZPI),
Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Abstract. A statistical deformation model provides a dense and detailed representation of the main modes the optic nerve head inter-subject variability.

1 Introduction

Quantitative glaucoma indices commonly rely on geometric parameters of the optic nerve head (ONH). However, the complex structure of the ONH cannot be accurately analyzed by geometric morphometry because of (i) sparse sampling of ONH and (ii) lack of biologically interpretable ONH descriptions [1]. Inspired by deformation based morphometry, we provide dense descriptions of the ONH variability in glaucomatous cases presented by tissue deformations.

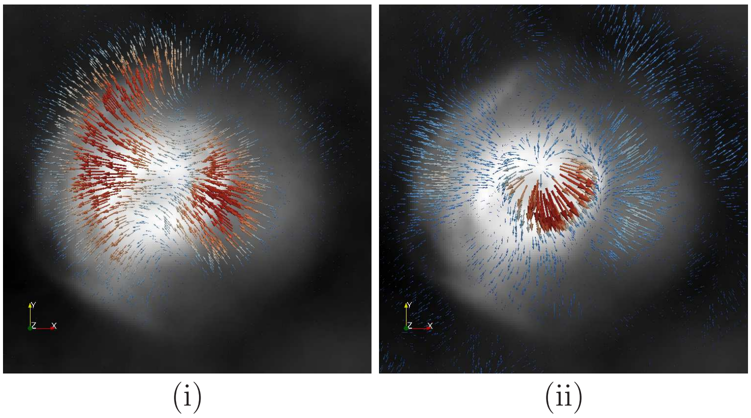
2 Methods

In a preprocessing step, disease independent variations such as illumination homogeneities, ONH position and vessel structures are compensated from the color fundus images [2]. The inter-subject ONH variability is then characterized as a dense deformation field between a fixed reference image and the current sample that is calculated by non rigid registration [3]. This fixed reference image is calculated iteratively by minimizing the residual of the deformations over the sample set. From glaucomatous and control subjects a statistical deformation model (SDM) is generated by principal component analysis (PCA).

3 Results

The gold standard diagnosis of the data set (149 glaucoma, 246 controls) was given by a glaucoma specialist based on an elaborate ophthalmological examination with ophthalmoscopy, visual field, IOP, FDT, and HRT II.

The constructed deformation-based average of the ONH appears naturally shaped and is not as smooth as it is the case for simple averaging. The SDM modes best discriminating between controls and glaucomatous cases illustrate two main types of the typical ONH variations in glaucoma: (i) Simultaneous altering of neuroretinal rim and the cup and (ii) isolated local extension of the cup (Fig. 1).



(i) (ii)

Fig. 1. Selected PCA modes clearly discriminating between control and glaucoma providing a mathematically derived and biologically interpretable description of ONH deformations: (i) Simultaneous deformation of the cup and rim, (ii) local extension of the cup (dark/red: high, bright/blue: low magnitude of a deformation vector).

4 Conclusion

The proposed SDM approach allows a dense and detailed representation of characteristic ONH variations. We expect that these models provide new insights to the glaucoma disease and can be potentially applied for automated glaucoma detection.

Acknowledgment

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

References

1. P.G. Sanfilippo, A. Cardini, A.W. Hewitt, J.G. Crowston, D.A. Mackey. Optic disc morphology - Rethinking shape. Progress in Retinal and Eye Research, 28(4), 227-248, 2009.
2. R. Bock, J. Meier, L.G. Nyúl, J. Hornegger, G. Michelson. Glaucoma risk index: Automated glaucoma detection from color fundus images. Med Image Anal, 14(3), 471-481, 2010.
3. J. Modersitzki, Numerical methods for image registration, Oxford University Press, USA, 2004.

A Framework for Voxel Based Analysis of the Optic Radiation Using DTI in Glaucoma

A. El-Rafei[1,4], T. Engelhorn[2], S. Wärntges[3], A. Dörfler[2], J. Hornegger[1,4], and G. Michelson[3,4,5]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Department of Neuroradiology
[3] Department of Opthalmology
[4] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
[5] Interdisciplinary Center of Ophthalmic Preventive Medicine and Imaging (IZPI), University Erlangen-Nuremberg, Germany

Abstract. Glaucoma is a neurodegenerative disease. The mechanism of glaucoma is not yet fully understood. In this work, we aim to investigate the effect of glaucoma on the optic radiation. A framework for the identification and analysis of the optic radiation using diffusion tensor imaging (DTI) is proposed. The results indicate that the proposed framework is able to efficiently localize the effect of glaucoma on the optic radiation.

1 Introduction

Diffusion tensor imaging (DTI) provides information about the white matter fiber orientation. Thus, it enables fiber tracking and reconstruction of white matter fibers. Moreover, parameters are derived from the diffusion tensor describing the degree of deviation from the isotropic diffusion such as fractional anisotropy (FA) which is used as an indicator of the fiber integrity.

2 Methods

The proposed system consists of the following steps: The optic radiation is automatically segmented using the segmentation algorithm developed by the authors [1]. Manual post-processing is performed to restrict the segmentation to the main fiber bundle of the optic radiation and to correct the segmentation errors. This results in a high degree of shape similarity of the optic radiation. A reference optic radiation is selected as that belonging to the normal subject with the largest optic radiation size. Optic radiations from all patients are non-rigidly registered to the reference [2]. Using the transformation fields from the registration step, tensor-derived parameters are as well transformed to the unified space of the reference. Finally, voxelwise statistical analysis using Mann Whitney U test is performed to determine the significant voxels.

3 Results

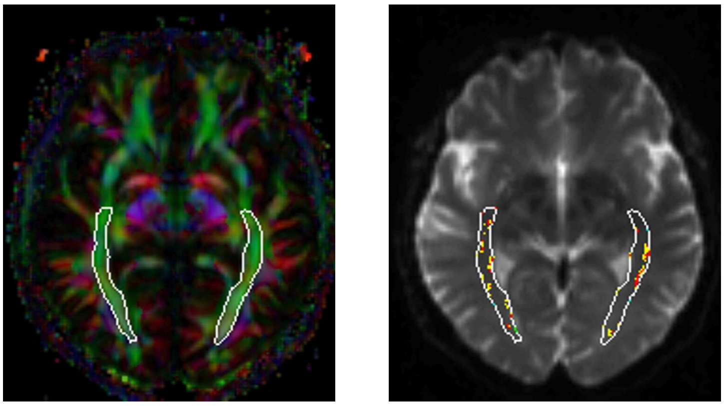
The system is applied to 23 subjects categorized into two age matched groups. A normal group of 10 subjects and a glaucoma group with 13 subjects. The voxels are considered significant if the p-value is less than 0.1. Figure 1 shows the significant regions detected by the system using FA analysis.

4 Conclusion

The framework is able to efficiently localize the effect of glaucoma on the optic radiation. Moreover, it avoids the dependence on the tensor-derived parameters affected by glaucoma.

Acknowledgements

The authors gratefully acknowledge funding of German academic exchange service (DAAD) and the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.



(a) (b)

Fig. 1. Reference optic radiation (a) and the significant optic radiation voxels based on the FA analysis (b).

References

1. A. El-Rafei et al.: Automatic Segmentation of the Optic Radiation using DTI in Glaucoma Patients. In: Computational Vision and Medical Image Processing - VipIMAGE 2009 (International Conference VipIMAGE 2009 - II ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing) Portugal : Taylor and Francis, pp. 293-298, 2009.
2. D. Rueckert et al.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18(8), 712 -721, 1999.

Multiple Frame OCT Data Merging: A Wavelet Denoising Approach

M. A. Mayer[1,2], R. P. Tornow[3], and J. Hornegger[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany,
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
[3] Department of Ophthalmology, University Clinic Erlangen,
Friedrich–Alexander University Erlangen-Nuremberg, Germany

Abstract. We propose a novel multiple frame optical coherence tomography (OCT) data merging approach that is based on wavelet denoising.

1 Introduction

Speckle noise suppression on OCT images is currently performed by averaging multiple frames [1]. In contrast to this common approach we propose a novel method originally developed for the denoising of Computed Tomography data [2]. It uses the structural properties of the actual image content to better differentiate between speckle noise and relevant tissue information.

2 Method

Each of the recorded single B-Scans is decomposed by a wavelet analysis. The wavelet coefficients representing a local frequency analysis of the images are then adapted by weights computed out of a statistical analysis. For each coefficient on every frame individual weights are calculated. Two weights are proposed: (i) A significance weight computed out of the local intensity differences that provides a local noise estimation. (ii) A correlation weight calculated from correlations in small neighborhoods that estimates whether an edge is present or not. The two weights are combined and applied on the detail coefficients. Finally the modified single frames are wavelet reconstructed and averaged.

To test the algorithm 455 linear B-scans were acquired from a pig's eye ex vivo with a Spectralis HRA+OCT, Heidelberg Engineering. Correlated noise was avoided by slightly moving the eye every 13 frames. All images are rigidly registered and averaged to form a noise suppressed gold standard. The signal-to-noise ratio (SNR) as well as sharpness reduction at selected borders using Full-Width-Half-Maximum is measured.

3 Results

Using the proposed method with only 8 frames we achieve a similar result compared to common averaging of 35 frames. This improvement is verified by an evaluation on multiple randomly selected sets. The SNR is increased by 104.0% while sharpness is only slightly reduced by 9.7%.

Fig. 1. Left: Average of 8 recorded OCT frames (Crop out of an OCT image). Middle: Standard wavelet thresholding. Right: Multiple frame wavelet denoising.

4 Conclusions

The proposed method makes use of the spatially varying noise statistics of multiple frame OCT-data of the same region. The evaluation showed that the SNR of OCT scans can be significantly improved while edge sharpness is preserved. The algorithm is applicable in 3D and can therefore be used to acquire high-quality OCT volume data in a shorter amount of time.

Acknowledgment

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

References

1. B. Sander, M. Larsen, L. Thrane, J. L. Hougaard, T. M. Jørgensen. Enhanced optical coherence tomography imaging by multiple scan averaging. Br J Ophthalmol, 89(2), 207212, 2005.
2. A. Borsdorf, R. Raupach, T. Flohr, J. Hornegger. Wavelet based Noise Reduction in CT-Images Using Correlation Analysis. IEEE Transactions on Medical Imaging, 27(12), 1685-1703, 2008.

Optical Density of Macular Pigment

A. Budai[1,4], J. Hornegger[1,4], G. Michelson[2,3,4], and R. P. Tornow[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
[3] Department of Ophthalmology
[4] Interdisciplinary Center of Ophthalmic Preventive Medicine and Imaging (IZPI),
Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Abstract. The receptors of the eye are protected from destructive radiation from the sun by the macular pigment. The loss of macular pigment can lead to eye damage or blindness. We develop an algorithm intended for clinical application to measure the macular density using fundus images as input. The method is based on the comparison of the absorption in the visible green and blue light domains. The measurements are evaluated by comparing multiple measurements from different illumination settings. The calculated pigment profiles show a correlation over 0.995.

1 Introduction

Most of the receptors in the eye are accumulated in the macula. These receptors are sensitive to the destructive radiation of the sun, and they are protected from it by the macular pigment. This pigment absorbs not only the destructive radiation, but absorbs a high amount of light in the visible blue light domain. In the clinical applications the density of the macular pigment is usually measured by a modified Heidelberg Retina Angiography (HRA) or multi-spectral fundus image acquisitions[3]. Here we present a method to measure the macular pigment density by comparing the blue and the green channels of color fundus images.

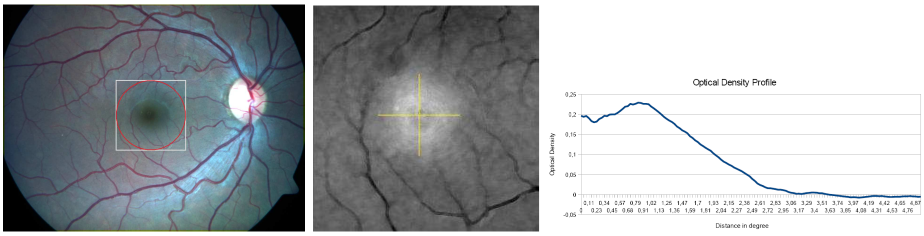
2 Methods

Retinal fundus images are acquired with a common fundus camera (FF 450 Carl Zeiss Meditec) with a special 3-band transmission filter[1,2] in the illuminating beam path. The images show no under- or over-saturation in any of the color channels[1]. A registration method is applied on the green and blue channels to correct the wavelength dependent aberration of the light. The only user interaction is a manual ROI selection around the macula region. The vessels of the images are excluded from the whole examination process. An optical pigment density image is generated by the following formula:

$$ODMP(x,y) = -\log\left(N * \frac{I_{blue}(x,y)}{I_{green}(x,y)}\right) \tag{1}$$

$$with \qquad N = \frac{I_{green}(reference)}{I_{blue}(reference)} \tag{2}$$

where N is a normalization factor, I_{blue} and I_{green} are the pixel intensities of the blue and green color channels, and the reference is an average of non-vessel pixels in 6° (visual angle) distance from the center of the selected macula region. From the density image a density profile is calculated to visualize the density changes as a function of distance from the center (example: Fig. 1).



(a) (b) (c)

Fig. 1. A filtered fundus image (a), the density image and the detected peak center (b), and the density profile showing the density as a function of distance from the center (c)

3 Evaluation

The algorithm was tested on multiple images of 2 subjects. The robustness was tested by comparing measurements from 9 images of the same eye. These images were taken using significantly changing illumination intensities from under-illuminated to almost over-saturated. All calculated density profiles showed similar curves. Their correlation was over 0.995 in each case.

4 Outlook

An evaluation of the measurements using multi-spectral fundus images or an HRA device is essential for further development. The algorithm has to be tested on high quality fundus images without the 3-band-filter. A fully automatic ROI selection algorithm has to be developed for the initial ROI selection.

References

1. R. P. Tornow *et al.*: Invest. Ophthalmol. Vis. Sci. 2007, Vol. 48, ARVO E-abstract: Program No. 1206
2. R. P. Tornow *et al.*: Patent application: Vorrichtung zur Optimierung von Farbaufnahmen DE102004010725A1 Deutsche Patent und Markenamt, 27.02.2004.
3. Lo J. bour *et al.*: Fundus Photography Measurement of Macular Pigment Density Distribution in Children, in: Investigate Ophthalmology & Visual Science, May 2002, Vol 43, No. 5, pages 1450-1455.

Implementation of an Intelligent Stimulator of Cyclopean Vision for Olympic Fencers

J. Paulus[1,2], J. Hornegger[1,2], and G. Michelson[2,3,4]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
[3] Department of Ophthalmology
[4] Interdisciplinary Center of Ophthalmic Preventive Medicine and Imaging (IZPI)
Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Abstract. Visual performance is not only dependent on the eye as the perception is processed in the brain. Perceptual learning using specific stimulators can improve this cognitive task allowing higher performance for athletes. This project's main goal is to develop a stimulator based on virtual reality to train and improve 3D vision of Olympic fencers.

The performance of human vision is an important issue in every day life. It is not only limited by properties of the eye as the visual system extends into the brain. Vision perception is processed in different retinal and cortical channels including spatial contrast, spatial orientation, texture, motion, cyclopean (3D) vision, and color. It was shown previously [1,2] that these channels can be trained by specific stimulation improving the visual performance. The goal of this project is to implement a visual stimulator for Olympic fencers based on virtual reality to train and improve their 3D vision to allow higher performance.

(i) The first main task will be to render a 3D object (e.g. sphere) that will fly around in front of the fencer. It will be presented in 3D by a stereo setup of two projectors using polarized filters.

(ii) The stereo perspective for the rendered object has to be adjusted according to the observer's viewing direction and position. Therefore, tracking of the 3D glasses is required. In virtual sports [3,4] this is mainly done via IR-cameras and passive markers.

(iii) Tracking of a real fencing weapon faces several problems (see iv). Therefore, in this step it is replaced by a special colored, stiff surrogate weapon showing a color not present in the scene. Simple background subtraction [5] in combination with a color based approach [6] are a possibility for tracking.

(iv) For a realistic analysis of fencers' 3D vision real fencing weapons with their specific behavior are required. They cause several problems: (a) Their specular surface will complicate the use of markers for IR-cameras. (b) Used markers in general have to be light weighted and small to avoid modifications of the balance and weight of the weapon. (c) For an accurate tracking high speed cameras are required, since the thin and flexible blade with its high velocity results in a smearing for normal cameras. One possibility for finding fencing blades is background subtraction combined with a Hough transform [7].

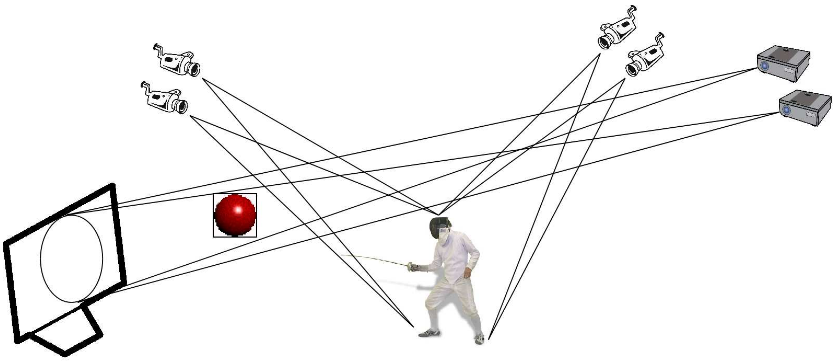


Fig. 1. Schematic setup: A 3D object is presented to a fencer via a stereo setup of two projectors. The weapon is tracked by four cameras to detect collisions with the object.

Acknowledgment

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German Research Foundation (DFG) in the framework of the German excellence initiative.

References

1. Polat, U., Ma-Naim, T., Belkin, M., Sagi, D.: Improving vision in adult amblyopia by perceptual learning. Proceedings of the National Academy of Sciences of the United States of America 101(17), National Academy of Sciences, USA (April 2004), 6692-6697
2. Polat, U.: Making perceptual learning practical to improve visual function. Vision Research 49(21), Elsevier B.V., USA (October 2009), 2566-2573
3. Rusdorf, S., Brunnett, G., Lorenz, M., Winkler, T.: Real-Time interaction with a humanoid avatar in an immersive table tennis simulation. IEEE Transactions on Visualization and Computer Graphics 13(1), IEEE Computer Society, USA (January/February 2007), 15–25
4. Bideau, B., Multon, F., Kulpa, R., Fradet, L., Arnaldi, B., Delamarche, P.: Using virtual reality to analyze sports performance. IEEE Computer Graphics and Applications 30(2), IEEE Computer Society, USA (November 2010), 14-21
5. Piccardi, M.: Background subtraction techniques: a review, IEEE International Conference on Systems, Man and Cybernetics, 4, IEEE Computer Society, USA (October 2004), 3099–3104
6. Javed, O., Shafique, K., Shah, M.: A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information. Workshop on Motion and Video Computing (MOTION'02) 0, IEEE Computer Society, USA (December 2002), 22–27
7. Mitchell, A., Mills, S.: Finding fencing foils using the Hough transform. Proceedings of Visual Information Engineering (VIE 2005), Institution of Electrical Engineers, United Kingdom (April 2005), 59-64

Sorting of Histological Image Sequences

S. Gaffling[1,2] and J. Hornegger[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany

Abstract. Methods for three-dimensional (3-D) reconstruction of histological image sequences depend on the correct order of images. Due to manual slice preparation, this might not always be given. We propose an algorithm that performs a consistency check and restores the correct image sequence based on a distance measure between slices.

1 Introduction

Methods for 3-D reconstruction of histological image sequences rely on the consistency of the data sets, especially on artifact-free images and their proper order. However, preparation of the data sets is done mostly manually, in particular slicing and digitizing. This commonly leads to discontinuities in the final image sequences, which prevents a proper reconstruction result. Therefore, we suggest that the image sequences should be checked for a consistent order, and the correct sequence should be restored prior to the reconstruction.

2 Methods

Since it is not possible to easily define a ">" or "<" relation on images, standard sorting algorithms are not applicable to solve this problem. Instead, we use the assumption that neighboring or close images should be more similar to each other than more distant ones.

Given an image sequence, the distance between one slice to the next is determined using rigid registration and calculation of the Mean Squared Distance (MSD),

$$\mathrm{MSD}(T, R) = \frac{1}{P} \sum_{i=1}^{P} (T_i - R_i)^2 \tag{1}$$

where T and R denote template and reference image, i is the i-th pixel in an image, and P is the number of considered pixels.

Then, a threshold is derived using the mean and standard deviation of the distances, and outliers with respect to this threshold are detected. The threshold is defined as

$$
\begin{aligned}
\Theta = \quad & \frac{1}{N-1} \cdot \sum_{i=1}^{N-1} \mathrm{MSD}\left(S_{i+1}, S_i\right) \\
& + \tau \cdot \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} \left[\mathrm{MSD}\big(S_{i+1}, S_i\big) - \mu\big(\mathrm{MSD}(S)\big)\right]^2} \\
= \quad & \mu\big(\mathrm{MSD}(S)\big) \quad + \quad \tau \cdot \sigma\big(\mathrm{MSD}(S)\big)
\end{aligned}
\tag{2}
$$

where N is the number of slices in the sequence, S_i is the i-th image of the sequence, and $\mu\big(\mathrm{MSD}(S)\big)$ and $\sigma\big(\mathrm{MSD}(S)\big)$ denote the mean resp. the standard deviation of the MSD values of the whole sequence S. τ is a user-defined weight for the standard deviation of the MSD values of the sequence.

The data set is divided into consistent blocks, where each block consists of a sequence of images with an MSD value smaller than the threshold.

Afterwards, the similarity between all block borders is calculated, and the blocks are matched and rearranged using the best found similarities. The reordered sequence is then used as input for the next iteration. The algorithm stops after a maximum allowed number of iterations, or earlier if no more changes are found.

Single images that are incorrectly positioned in the sequence, however, have to be treated separately. Despite missing at the original, true position, this location cannot be identified using the abovementioned threshold. So after iterating several times until no changes are found, single remaining outliers are detected and reintroduced into the sequence by matching the image against all others and choosing the position which fits best.

3 Results and Outlook

The approach was tested on a histological data set of a mouse brain, which was specially prepared for reconstruction, and can be considered mostly artifact free [1]. The image sequence was first intentionally manipulated, and then used as input for our sorting algorithm.

Figure 1 shows the reconstruction results of the unsorted sequence and after the sequence was rearranged using our algorithm.

As evaluation strategy, we propose to calculate the Accumulated Sequence Mismatch (ASM), i.e., the sum of the sequence position discrepancies of each image to its true successor:

$$
\mathrm{ASM}(S) = \sum_{i=1}^{N-1} \left[\, |P(S_{i+1}) - i)| - 1\right] \tag{3}
$$

with $P(S_{i+1})$ being the position index of the true succsessor of the i-th image. Subtraction of one prevents adding a penalty for correct neighbors. Since the suggested evaluation measure is not yet implemented, the evaluation was done using visual inspection.

In general, the method is able to find discontinuities in image sequences and to restore the correct sequence in many cases. The runtime depends mostly on the speed of the rigid registration method and the degree of parallelization. Further improvement will be a more flexible choice of the threshold, which demands further investigation and refinement.

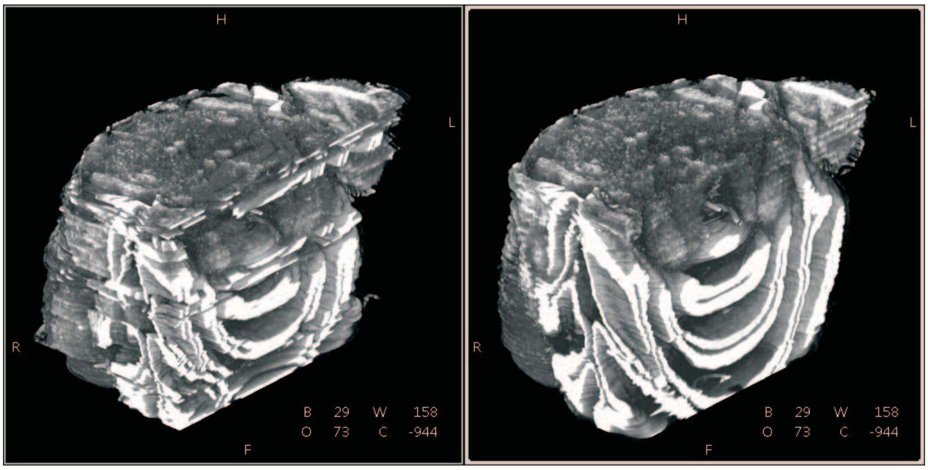


Fig. 1. 3-D reconstruction of a mouse brain. Left: Unsorted image sequence leads to discontinuities in the structure. Right: After sorting, the reconstruction appears much smoother.

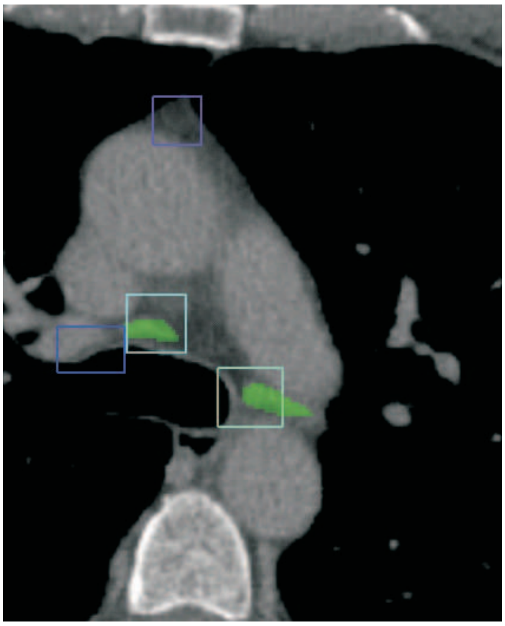
4 Acknowledgements

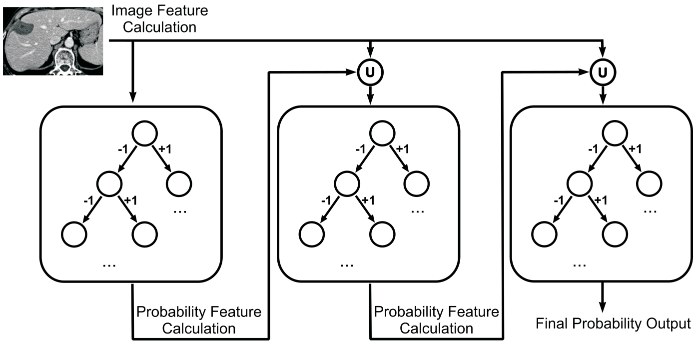
The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

References

1. T. Ju *et al.*: Building a 3D Atlas of the Mouse Brain. PhD thesis, Rice University, April 2005. Data set for download at http://www.geneatlas.org/gene/data/histology.zip.

Segmentation





Vessel Segmentation

M. Spiegel[1,3], D. Schuldhaus[1], J. Hornegger[1,3], and A. Dörfler[2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany,
[2] Department of Neuroradiology, Friedrich-Alexander-University
Erlangen-Nuremberg, Germany
[3] Erlangen Graduate School in Advanced Optical Technology (SAOT), Germany

Abstract. 3-D rotational angiography (3-DRA) denotes a well established approach at modern neuroradiology departments to examine cerebrovascular disease i.e. aneurysms, stenosis or fistule. 3-DRA images enable physicians to come up with a detailed analysis of vessel diameters, size and location of aneurysms. Nevertheless, 2-D digital substraction angiography (DSA) image series are still heavily used to support and guide this analysis. Thus, vessel segmentation in 2-D and 3-D supports this analysis to come up with a clinical descision. In this work, we present a 2-D and 3-D vessel segmentation method. The 3-D approach makes use of ellipsoids to approximate the tubular structure of the vessel and the 2-D DSA segmentation is a seed point driven graph-based technique.

1 3-D Segmentation

An interleaved ellipsoid model was used to approximate the contour of vessels [1]. A sphere, centered at the origin and aligned with its coordinate axis, can be defined by the following function

$$E(\mathbf{x}) = x^2 + y^2 + z^2 \tag{1}$$

where $\mathbf{x}$ denotes a point in $\mathbb{R}^3$. Considering a set of points $\mathbf{X} = \{\mathbf{x}|E(\mathbf{x}) = 1\}$ lying on the surface of the sphere, a coordinate transformation $\mathbf{T}(\mathbf{x})$ generalizes the sphere to an ellipsoid in terms of rotation, scaling and translation i.e. pose:

$$\mathbf{T}(\mathbf{x}) = \mathbf{R}(\mathbf{r})\mathbf{S}(\mathbf{s})\mathbf{x} + \mathbf{t} \tag{2}$$

$$\mathbf{a} = \left(\mathbf{r}^T, \mathbf{s}^T, \mathbf{t}^T\right)^T \tag{3}$$

where $\mathbf{a}$ is a vector in $\mathbb{R}^9$ holding the pose parameters. $\mathbf{R}$ and $\mathbf{S}$ represent rotation and scaling matrices respectively. This coordinate transformation yields a new set of ellipsoids $\mathbf{X}^{'} = \left\{\mathbf{x}|E(\mathbf{T}^{-1}(\mathbf{x})) = 1\right\}$. The parameter estimation is performed using a maximum likelihood estimator with the corresponding log-likelihood function:

$$F(\mathbf{a}, I_F, I_B) = \int_{\Re(\mathbf{a})} \log f(I(\mathbf{x}) - I_F))d\mathbf{x} + \int_{\Omega\setminus\Re(\mathbf{a})} \log f(I(\mathbf{x}) - I_B))d\mathbf{x} \tag{4}$$

where I_F and I_B denote foreground and background intensities respectively. $I(\mathbf{x})$ is the voxel intensity at position $\mathbf{x}$. $\Re(\mathbf{a})$ is a region defined by $E(\mathbf{T}^{-1}(\mathbf{x})) \leq 1$ and f determines the additive noise within the image. Figure 1 (a) shows a segmentation result.

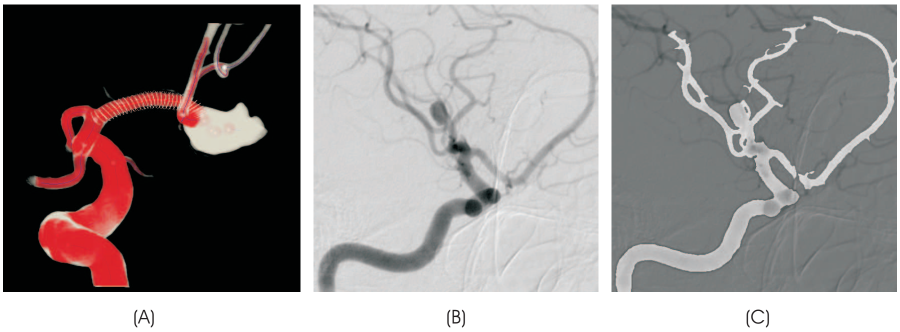


Fig. 1. Segmentation results: (a) 3-D segmentation (b) and (c) show the original 2-D DSA image and the 2-D segmentation result respectively.

2 2-D Segmentation

Seed points are provided by the user to initialize the 2-D segmentation method. A semi-automatic segmentation method is preferred in this case because the physician is often only interested in those vessel branches that lie in the vicinity of the pathology i.e. aneurysm or stenosis. A graph-based algorithm [2] is applied to compute the vessel centerline between the seed points and a local box enhancement method is used to segment the 2-D vessel branches. Figure 1 (b) and (c) illustrates this approach.

3 Outlook

In future, it is planned to combine the 2-D segmentation with the 3-D segmentation in order to incorporate additional information and to improve the 3-D segmentation.test

4 Acknowledgements

The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

References

1. Tyrrell, J., di Tomaso, E., Fuja, D., Tong, R., Kozak, K., Jain, R., Roysam, B.: Robust 3-d modeling of vasculature imagery using superellipsoids. IEEE Transactions on Medical Imaging 26(2) (February 2007) 223-237
2. Dijkstra, E.W.: A note on two problems in connections with graphs. Numerische Mathematic 1, 269-271 (1959)

Automatic Segmentation of Liver Lesions in CT Images by Means of Cascaded Boosting Trees

A. Militzer[1,2]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Siemens AG Healthcare, Computed Tomography, Siemensstr. 1, 91301 Forchheim,
Germany

Abstract. Planning liver interventions for tumor treatment requires precise knowledge about the location, size and shape of the target lesions. The same is true for follow-up examinations in image based therapy monitoring. Manual segmentation of these lesions, however, is tedious and too time consuming for clinical routine. Therefore, in this paper a novel method is presented that fully automatically performs simultaneous detection and segmentation of hypodense liver lesions in CT images. This is achieved by classifying points in the liver as being part of a lesion or of the background by means of a Probabilistic Boosting Tree. In a comprehensive evaluation of the system, detection rates of 79% could be achieved, with a segmentation sensitivity of 96% and 92% specificity at the same settings.

1 Introduction

Computed tomography (CT) images, acquired after intravenous injection of a contrast agent, are widely used by clinicians for diagnosis, treatment planning and monitoring of liver tumors. Up to six images are acquired, showing different phases of contrast enhancement in the liver. For the segmentation method presented here, images with venous contrast enhancement were used, as these provide the best contrast between lesion and parenchyma for many lesion types. A learning based segmentation approach was chosen in order to be able to capture the great variability in lesion appearance.

2 Classification Based Segmentation

During preprocessing, the liver is segmented first in order to reduce the search space and the complexity of the classification task. Next, intensities inside the liver are standardized using the method by Jäger et al. [1]. This step is crucial, as the overall image intensity in the contrasted liver heavily depends on factors like acquisition timing and individual perfusion.

For the pointwise classification a Probabilistic Boosting Tree (PBT) [2] is utilized. Due to its hierarchical nature, it is very well suited for problems with high intra-class variance. It assigns a lesion probability value to each point in the liver. In order to further stabilize and smooth the classification results, an iterative scheme is employed (see Fig. 1), in which the output image of each PBT is used to calculate additional features for the points, which are then classified again by another PBT [3]. Finally, the output probability map of the last PBT is transformed into a lesion mask.

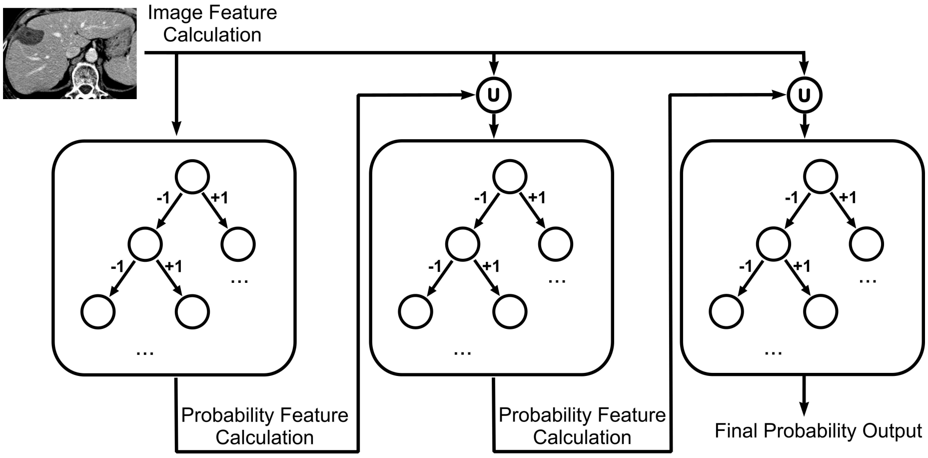


Fig. 1. Iterative classification scheme. For each point, image features are calculated and used for classification in the first PBT. From its output, further (probability) features are calculated. Together with the original image features, these are used for classification in the next classifier.

3 Experiments

The presented algorithm was evaluated on 15 CT datasets in a five-fold cross-validation, yielding a lesion detection rate of 79% at 4 false positive detections per patient. At the same settings, a segmentation sensitivity of 96% was achieved at 92% specificity. The maximum dice similarity coefficient was 1.48. Moreover, three iterative classification steps turned out to be sufficient for convergence.

References

1. Jäger, F.; Hornegger, J.: Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. IEEE Transactions on Medical Imaging 28 (1), pp. 137–150, Jan. 2009.
2. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. Proc. ICCV 2, pp. 1589–1596, Beijing, China, Oct. 2005.
3. Morra, J. H.; Tu, Z.; Apostolova, L. G.; Green, A.E.; Toga, A. W.; Thompson, P. M.: Automatic Subcortical Segmentation Using a Contextual Model. Proc. MICCAI, pp. 194–201, New York, USA, Sept. 2008.

Lymph Node Detection in 3-D Chest CT

J. Feulner[1], K. Zhou[2], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Siemens Corporate Research, Princeton, NJ, USA

Abstract. This work presents a method for fully automatic detection of mediastinal lymph nodes, which is a challenging problem due to low contrast to surrounding structures and clutter. In contrast to prior work, we heavily rely on prior anatomical knowledge, which is modelled as a spatial prior probability and learned from annotated data.

1 Introduction

Lymph nodes play an important role in clinical practice, especially in the mediastinal area. They routinely need to be considered during oncological examination related to all kinds of cancer [1,2], for instance lung cancer, where metastases settle in lymph nodes, but also lymphoma, which is a cancer of the lymphatic system itself. Furthermore, they are also relevant in case of inflammation in general.

Manually counting and measuring lymph nodes in the images is not only cumbersome but also error prone because annotations from different human observers and even from the same human observer vary significantly. However, automatic detection of lymph nodes is challenging as they are hard to see due to low contrast and irregular shape.

2 Method

We propose a method for fully automatic mediastinal lymph node detection in 3-D computed tomography (CT) images of the chest area. Discriminative learning [3,4] is used to detect lymph nodes based on their appearance. Because lymph nodes can easily be confused with other structures, it is vital to incorporate as much anatomical knowledge as possible to achieve good detection rates. Here, a learned prior of the spatial distribution is proposed to model this knowledge. As atlas matching is generally inaccurate in the chest area because of anatomical variations, this prior is not learned in the space of a single atlas, but in the space of multiple ones that are attached to anatomical structures. During test, the priors are weighted and merged according to spatial distances.

Fig. 1 shows some detection examples. In general, the detection score, which is color coded, reflects well the quality of the detection. False positives are mostly structures that look similar to lymph nodes like vessels, or sometimes they are at locations where lymph nodes are common. One reason for this behavior is the spatial prior, and another reason is that the detectors also learned common surrounding structures of lymph nodes.

3 Results

Cross-validation on 54 CT datasets showed that the prior based detector yields a true positive rate of 52.3% for seven false positives per volume image, which is about two times better than without a spatial prior.

4 Acknowledgements

This work was in parts supported by the German Federal Ministry of Economics and Technology.

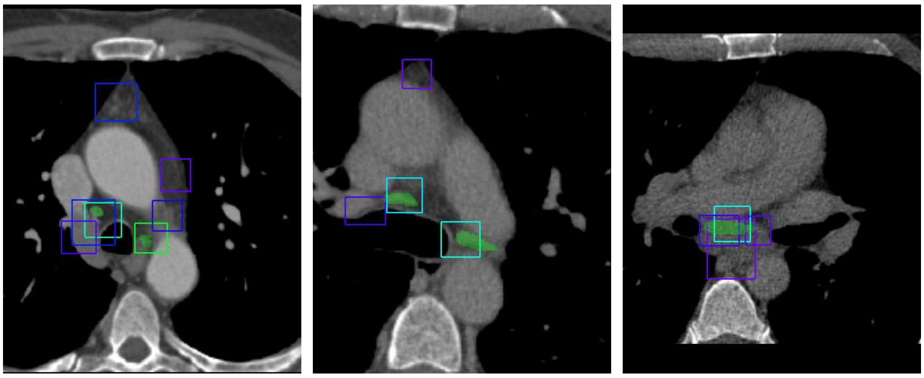


Fig. 1. Examples of detection results on unseen data. The boxes are the automatic detections, the manual ground truth segmentations are shown in green. The detection score in color coded: Violet means lowest, turquoise is medium and green is high confidence.

References

1. A. J. de Langen, P. Raijmakers, I. Riphagen, M. A. Paul, and O. S. Hoekstra. The size of mediastinal lymph nodes and its relation with metastatic involvement: a meta-analysis. Eur J Cardiothorac Surg, 29(1):26-29, 2006.
2. B. V. Duwe, D. H. Sterman, and A. I. Musani. Tumors of the mediastinum. Chest, 128(4):2893-2909, 2005
3. Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features. ICCV, pages 1-8, Rio de Janeiro, Oct. 2007
4. Z. Tu. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. ICCV, Volume 2, pages 1589-1596, Beijing, Oct. 2005

Automatic Determination of Symmetry Plane / 4-D Brain PET Tumour Classification

P. Ritt[1,2], T. Kuwert[1], and J. Hornegger[2]

[1] Clinic for Nuclear Medicine, University Hospital, Krankenhausstr. 12, 91054 Erlangen, Germany
[2] Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany

Abstract. In medical imaging, decisions based on symmetry of anatomical/functional structures are quite common. Registration techniques can be used to automate the procedure of determining the plane of symmetry.
Dynamic (4-D) PET images of the human brain provide additional information for a differentiation of tissue or tumour classes. An automated classification of brain tumours is described in the following.

1 Automatic Determination of Symmetry Plane

Approximate plane symmetry is often encountered in medical imaging. Among various examples like the skeletal system, blood vessels, lungs, kidneys and many more, the human brain shows a high degree of symmetry corresponding to the mid-sagittal plane. Many medical decisions on this field are influenced by physicians taking the known symmetry into account. Using symmetry information for automatic classification therefore seems to be a intuitive and natural way. In order to ensure full-automatic classification, an automated determination of the symmetry plane is necessary. This can be achieved by the use of registration techniques: In image registration, a similarity measure between two fused medical images is calculated. The similarity measure is considered to have a maximum when the two images are aligned in an optimal way relative to each other. The same similarity measure can be used to find an optimal reflection plane. In our case, a new dataset which is the original dataset reflected with respect to an arbitrary plane, is created. It can be shown [1] that with the help of a subsequent translation and rotation the image transformation is equal to a reflection with respect to any desired plane. In case of a translation and rotation that maximizes the similarity measure (rigid registration) between reflected and original dataset the result is equal to a reflection with respect to the symmetry plane. We use the mutual information based on image intensities as similarity measure. The visual results seem to be accurate (Fig. 1) and were confirmed by physicians, the technique can be applied to a variety of medical images (e.g. CT, SPECT, PET, MRI, ...).

2 4-D Brain PET Tumour Classification

It is possible to obtain 3-D PET images [2] dynamically (4-D), which means a 3-D image is acquired at multiple points in time. Dynamic images are used to monitor the

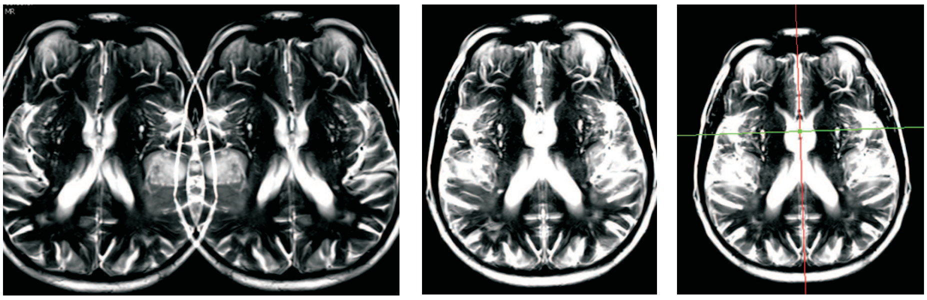


Fig. 1. T2 weighted Brain MRI. Left: Unregistered images; Middle: Registered images; Right: Calculated plane of symmetry

kinetics (speed/way of uptake) of a radio-pharmaceutical in a certain volume (e.g. of an amino acid in the brain). It can be helpful for enhanced diagnostics of brain tumours to evaluate various parameters (features) for a specific region of interest (ROI), e.g. the shape of the time activity curve (TAC), mean uptake, ratio to reference region, standardized uptake value. The manual pixel by pixel classification by physicians is the current standard in clinics and is rather time consuming. Therefore an automatic classification of the tumours according to the aforementioned features seems feasible. We developed a solution which automatically extracts the desired features from medical brain PET images. The solution is integrated into a state-of-the-art 3-D volume renderer (InSpace)[3]. Additionally the automatic classification using Support-Vector-Machines (SVM) was evaluated. Based on datasets from 79 patients out of the clinical routine, classification rates for the 2 class problem (tumour/non-tumour) are good (97%). However classification rates for the tumour type (4 classes: WHO1+2 Tumours, WHO3 Tumours, WHO4 Tumours, Oligodendrogliomal Tumours) are rather low (54%). An inaccurate ground-truth and overlapping classes in the feature space seem to be limiting factors. Possible solutions could be a more carefull selection of the patient collective and additional usage of features derived from other modalities (e.g. MRI, CT).

References

1. Tuzikov, A. ; Colliot, O. ; Bloch, I.: Evaluation of the symmetry plane in 3D MR brain images. *Pattern Recognition Letters* 2003; 24:2219–2233
2. Kuwert, T. ; Grünwald, F. ; Haberkorn, U. ; Krause, T.: *Nuklearmedizin* 2008, 27-30, Georg Thieme Verlag, Stuttgart
3. *http://www.insideinspace.com*

TOF-Applications

Rectification of Endoscopic Images by Gravity

K. Höller[1,2,3], S. Hempel[2,3], A. Schneider[3], J. Jahn[4], J. Guttierrez[4], T. Wittenberg[4],
H. Feußner[3], and J. Hornegger[1,2]

[1] Central Institute of Healthcare Engineering (ZIMT),
Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
[2] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[3] Workgroup for Minimal Invasive Surgery MITI, Klinikum r.d. Isar, Technical University Munich, Germany
[4] Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany

Abstract. Image rotation correction even in non-rigid endoscopic surgery (particularly NOTES) can be realized with a tiny MEMS tri-axial inertial sensor placed on the tip of an endoscope by measuring the impact of gravity on each of the three orthogonal axes. After an initial calibration the rotation angle can be estimated with some calculations out of these three values. Achievable repetition rate is above the usual endoscopic video frame rate of $30Hz$, accuracy is about one degree. The image rotation was performed by rotating just digitally a capture of the endoscopic analogue video signal which can be realized in real-time. Improvements and benefits have been evaluated in animal studies: Coordination of different instruments and estimation of tissue behaviour regarding gravity related deformation and movement was considered to be much more intuitive having a stable horizon on endoscopic images.

1 Method

One of the biggest problems with flexible endoscopy in Natural Orifice Translumenal Endoscopic Surgery (NOTES) [1] is the missing information about image orientation [2]. Thus, tip retro-flexion of a non-rigid endoscope causes image rotation angles till ±180 degrees.
Our approach to measure this orientation angle is to integrate a Micro Electro-Mechanical System (MEMS) based inertial sensor device in the endoscope's tip, which measures influencing forces in three orthogonal directions [3,4]. If the endoscope is not moving, only the acceleration of gravity has an effect on the three axes. Using an arc tangent function the rotation angle Φ can be computed out of acceleration values F_y and F_z on the two axes y and z orthogonal to the endoscopic line of view in x-direction:

$$\Phi = \arctan2(F_y, F_z) \tag{1}$$

First, a preceded 3×3 calibration matrix, which incorporates misalignment and scaling errors [5], has to be retrieved by initial measurements. Moreover a peak elimination is the result of clever down sampling the measuring frequency, which is considerably higher than the image frame rate, using a weighted sum algorithm. To avoid bouncing or jittering images as a result of the angle correction, additional filtering is necessary.

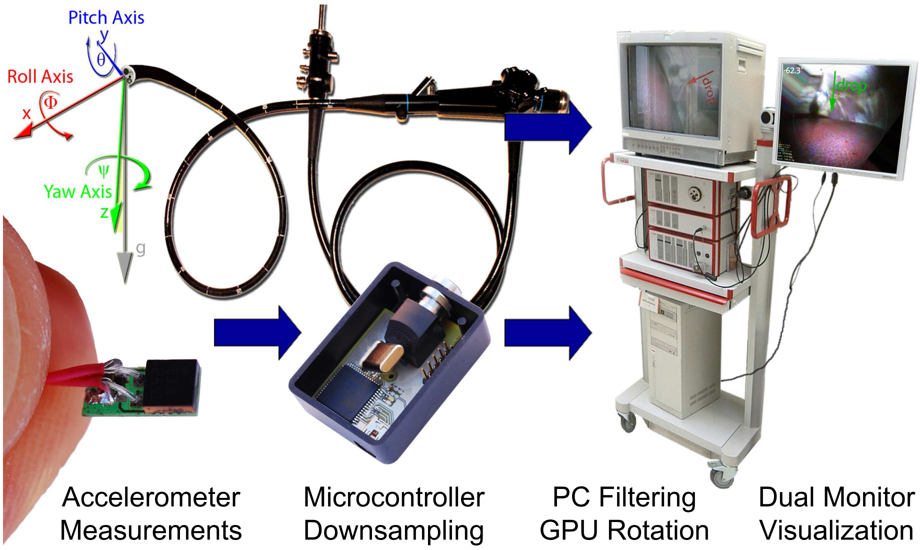


Fig. 1. EndoSens Hardware with ENDOrientation Algorithm for Endoscopic Image Rectification

Hence, prior to angle calculation, the measures of each axis are filtered with a Hann filter to smooth angle changes and with a minimum variation threshold to suppress dithering.
With the employed sensor there is a uniform quantization of 8 bit for a range of $\pm 2.3g$ for each axis. This implies a quantization accuracy of $0.018g$ per step or 110 steps for the focused range of $\pm g$. This is high enough to achieve a durable accuracy even to a degree within relatively calm movements. This is possible as roll angle Φ is calculated out of inverse trigonometric values of two orthogonal axes. Single extraordinary disturbed MEMS values or superposed accelerations with the same order of magnitude as gravity are suppressed by the ENDOrientation algorithm [6].

2 Results

During a porcine animal study, the navigation complexity of a hybrid endoscopic instrument during a NOTES peritoneoscopy with the well established trans-sigmoidal access [7] was compared with and without automated image rotation. To evaluate the benefit of automated real-time MEMS based image rectification, four different needle markers were inserted through the abdominal wall to the upper right, lower right, lower left and upper left quadrants. These four needle markers had to be grasped with a trans-abdominal introduced endoscopic needle holder under standardized conditions. The minimum time, the mean time and the maximum time have been lower with image rectification for every position [8]. All participating surgeons considered the complexity lower using our ENDOrientation technique.

References

1. Rattner, D., Kalloo, A.: ASGE/SAGES working group on Natural Orifice Translumenal Endoscopic Surgery: White Paper October 2005. Surg Endosc vol 20 (2006) 329-333.
2. Höller, K., Petrunina, M., Penne, J., Schneider, A., Wilhelm, D., Feussner, H., Hornegger, J.: Taking endoscopy to a higher dimension: Computer Aided 3-D NOTES. In: Proc. 4th Russian-Bavarian Con-ference on Bio-Medical Engineering, Zelenograd, Moscow, MIET (July 2008) 33-37.
3. Höller, K., Penne, J., Schneider, A. et al.: Endoscopic orientation correction. In: Medical Image Computing and Computer Assisted Intervention, 12th International Conference Proceedings MICCAI'09 London, UK (September 2009) 459-466.
4. Höller, K., Schneider, A., Jahn, J. et al.: Orientation of Endoscopic Images: Rectification by Gravity. In: J Biomed Eng, August 2010, in Press.
5. Höller, K.: Characterization and modeling of an inertial sensor for navigation of autonomous systems. Friedrich-Alexander-Universitt Erlangen-Nrnberg (October 2005).
6. Höller, K., Penne, J., Schneider, A. et al.: Suppression of shock based errors with gravity related endoscopic image rectifcation. In: Proc. 5th Russian-Bavarian Conference on Bio-Medical Engineering Munich, Germany (July 2009) 43-47.
7. Wilhelm, D., Meining, A., von Delius, S., et al.: An innovative, safe and sterile sigmoid access (ISSA) for NOTES. Endoscopy vol 39 (2007) 401-406.
8. Höller, K., Schneider, A., Jahn, J. et al.: Clinical evaluation of ENDOrientation: Gravity related rectifcation for endoscopic images. In: Proceedings of the ISPA, Salzburg, Austria (September 2009) 713-717.

Time-of-Flight Surface Imaging: Applications in Radiotherapy

S. Bauer, J. Wasza, and J. Hornegger

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. This project investigates the integration of time-of-flight surface imaging into the clinical workflow of fractionated radiotherapy. The major goal is to further improve the safety, quality and efficiency of dose delivery. Promising applications within the radiotherapy workflow are an optimized patient positioning and real-time motion management.

1 Time-of-Flight Surface Imaging

Time-of-flight (ToF) imaging provides a direct way to acquire metric surface data with a single sensor in real-time [1]. Active light sources attached to the camera emit an incoherent cosine-modulated optical signal in the non-visible spectrum of the infrared range. The light is reflected by the scene and enters the monocular camera where each sensor element performs a correlation of the local optical signal with the electrical reference of the emitted signal. The measured phase shift ϕ between both signals represents the propagation delay and the distance d from the sensor, $d = \frac{c}{2 \cdot f_{mod}} \cdot \frac{\phi}{2\pi}$, where f_{mod} denotes the modulation frequency, c the speed of light.

2 Applications in Radiotherapy

A variety of medical applications is expected to benefit from ToF imaging in the future [2,3,4]. The focus of this research project is on the development of methods that improve the accuracy, safety and workflow efficiency of fractionated radiotherapy (RT). As the name implies, the irradiation of the tumor is performed in multiple sessions according to a treatment plan that is generated from an initial computed tomography (CT) scan. In order to ensure an accurate adjustment of the external beam, the

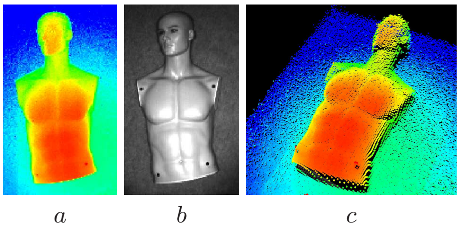


Fig. 1. ToF measurement data, delivered by the sensor. *a*: Distance. *b*: Amplitude (measure of confidence). *c*: 3D visualization of the point cloud.

oncological patient must be aligned to the reference position of the planning CT before each treatment session. In this project, we develop a surface-based method for patient positioning that relies on ToF data registration. In comparison to existing alignment strategies, our approach is marker-less, non-invasive, contactless and does avoid additional radiation exposure.

Real-time motion management is another promising application of ToF imaging for RT. We develop a motion management system for acquisition and classification of a multi-dimensional volumetric respiratory signal. The system monitors specific anatomical regions of the upper part of the body and determines the instantaneous phase within the respiration cycle.

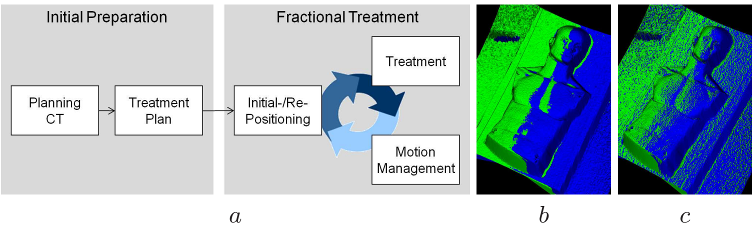


Fig. 2. *a*: Flowchart of the clinical workflow in fractionated radiotherapy. *b*, *c*: Two surface acquisitions, labeled in green and blue, respectively, before and after registration.

Acknowledgements

The authors acknowledge the support of the Bayerisches Staatsministerium für Wirtschaft, Verkehr und Technologie, Siemens AG, Healthcare Sector, Erlangen, Germany, and Softgate GmbH, Erlangen, Germany.

References

1. A. Kolb, E. Barth, R. Koch, and R. Larsen, “Time-of-flight sensors in computer graphics,” in *Eurographics State of the Art Reports*, pp. 119–134, Munich, Germany, Mar 2009.
2. C. Schaller, J. Penne, and J. Hornegger, “Time-of-flight sensor for respiratory motion gating,” *Medical Physics*, vol. 35, no. 7, pp. 3090–3093, Jun 2008.
3. C. Schaller, C. Rohkohl, J. Penne, M. Stürmer, and J. Hornegger, “Inverse C-arm positioning for interventional procedures using real-time body part detection,” in *MICCAI*, vol. 5761 of *LNCS*, Part I, pp. 549–556, London, UK, Oct 2009.
4. J. Penne, K. Höller, M. Stürmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feußner, B. Schmauss, and J. Hornegger, “Time-of-flight 3-D endoscopy,” in *MICCAI*, vol. 5761 of *LNCS*, Part I, pp. 467–474, London, UK, Oct 2009.

Patient Positioning with a Time-of-Flight Camera: Feature Based Approach

S. Placht, C. Schaller, M. Balda, and J. Hornegger

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Patient positioning is very important for an accurate dose delivery in radiotherapy. We present an automatic system, which computes the rigid body transform between two surfaces acquired by a ToF camera.

1 Introduction

In radiation therapy, tumors inside the human body are located and a treatment plan is created. An optimal alignment of the treatment beam to the cancerous tissue requires a very accurate positioning of the patient. To cope with that problem, there are on the one hand X-ray image based systems. However, the additional radiation exposure is harmful for the patient. On the other hand optical systems can be used to acquire the body surface. Commercially distributed systems are still very expensive. A cost-effective modality for the acquisition of objects is the Time-of-Flight (ToF) camera, which provides 3-D information for a whole scene in realtime. In the presented system, a ToF camera is used for patient positioning.

2 Methods

First of all, the camera is calibrated in order to reduce systematic distance errors. The coordinate system of the linear accelerator is detected by means of a chessboard-calibration-pattern. Temporal camera noise is suppressed with a Kalman filter [1]. Afterwards, the patient's surface is extracted from the ToF dataset and rigidly registered with the surface of a reference dataset. For the registration a two stage algorithm is implemented: The first step introduces a novel algorithm for the computation of surface features. Local surface patches are extracted and analyzed in the frequency domain. By means of the feature vectors corresponding points are detected and a first estimation for the rigid body transform is computed. The second stage consists of a modified version of the Iterative Closest Point (ICP) algorithm [2]. It refines the registration result and provides the relative transformation parameters for patient positioning.

3 Results

Using a body phantom, both the accuracy and the stability of the results is evaluated on a linear accelerator. Practical tests demonstrate a mean accuracy of 1.62 mm for the translational and 0.07° for the rotational component. The registration result of two not completely overlapping surfaces is visualized in Fig. 1.

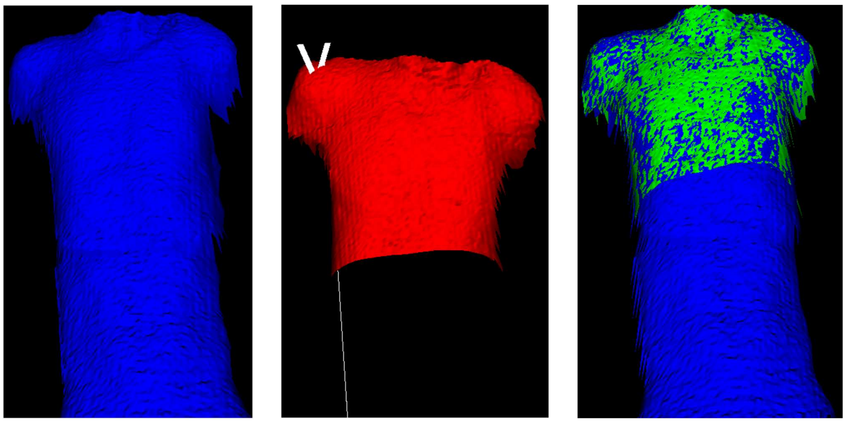

Complete surface Transformed and cut surface Registration result

Fig. 1. Registration result after pre-registration and ICP.

References

1. Welch, G.; Bishop, G.: An Introduction to the Kalman Filter, Chapel Hill, NC, USA, July 2006.
2. Rusinkiewicz, S.; Levoy, M.: Efficient Variants of the ICP Algorithm, in Proceedings of the Third Intl. Conf. on 3D Digital Imaging and Modeling, 2001, pp. 145-152.

There is no Milka Cow!

C. Schaller^{1,2}, M. Balda^{1,2}, and S. Placht^{1,2}

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
Metrilus GmbH, Martensstr.3, 91058 Erlangen, Germany

Abstract. Time-of-Flight (ToF) cameras [2] provide a cost-effective alternative to camera-based milking robots. Due to the fast distance measurement of ToF cameras one can determine the position of cow teats. Using this position a robot can precisely dock to the detected teats.

1 Introduction

A major pitfall for many livestock farms is the relatively high purchase price (> 100.000 Euro) of current automated milking robot systems. Furthermore, there is also a limited functionality of current systems. Often, ”‘inappropriate”’ cows cannot be processed by existing systems on the market. ToF-cameras offer a cost-effective and better alternative to systems based on structured light or laser scanners. These methods cannot capture fast movements of cows, and thus the time used for the docking process is increased. In addition, a ToF-based solution can be realized much cheaper. Thus a ToF-based milking robot is not only of higher quality, but also reduces the milking time. We are currently developing algorithms for automatic detection of the position of the teats. This information is used to control milking robots.

2 Methods

We us a standard CamCube ToF camera from PMD Technologies, Siegen. An amplitude image of 204×204 pixels and a corresponding distance map is acquired by the camera. The camera, is attached to a robot and pointing in direction of the cow udder. Based on the amplitude image, we segmented about 1500 cow teats and flipped the images on the y-axis. Therefore, we used about 3000 positive samples for training a classifier based on haar features. Furthermore, about 17000 negatives samples are used. To eliminate multiple and false detections, we introduced special constraints using the corresponding distance map. These constraints contain the computation of the overlap of detections and the correct window size in metric coordinates.

3 Results

Figure 1 shows the detection result of the just described system. The detected positions of the system are transfered to the robot via a serial protocol.

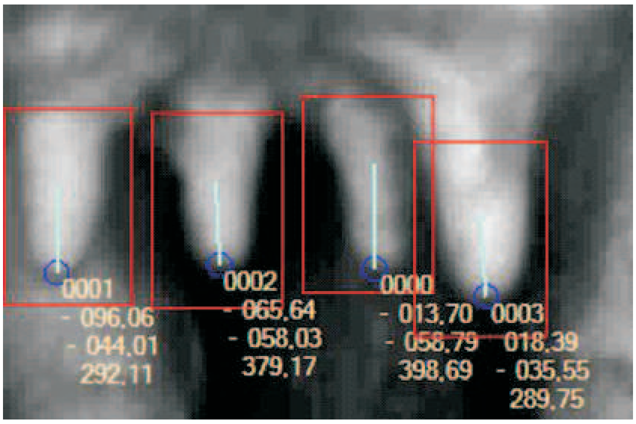


Fig. 1. Cow udder showing detected teats including the 3-D position.

References

1. Viola, P.; Jones, M.: Robust Real-time Object Detection, International Journal of Computer Vision, 2001
2. Ringbeck, T.; Profittlich, M.; Schaller, C.: Kameras für die dritte Dimension, In: Optik und Photonik Oktober (2009) No. 3 pp. 30-33

ToF Imaging of Foot Soles

M. Stürmer[1,2], G. Becker[2], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Lion Systems S.A., ecostart 2, rue du commerce, L-3895 Foetz, Luxembourg

Abstract. Measuring devices for fast, accurate and markerless 3D-acquisition of human feet are of large interest for a wide range of applications. From the fact that every foot is different arises the need for an easy to use and inexpensive measurement system that is able to image feet during gait. Time-of-Flight (ToF) imaging is a very promising technology to solve this task, as it measures 3D-images in realtime without the need for markers and does not illuminate the scene with any visible light.

1 Introduction

Both in the field of sports as well as in medicine and science the interest for fast 3D measuring devices remains high. In particular the deformation of feet during gait cannot be measured by static acquisition devices that are available nowadays. Even though there are several attempts to do high accuracy measurements of footsoles at high framerates [1] [2], no system could establish itself so far.

Our proposed system consists of a ToF-camera which is facing a glass-plate from below. The foot to measured is walking atop the glass and the camera records the sole. The main goals of this system are portability, easy calibration routines, high frame rate and the best possible measurement accuracy.

2 Calibration

The setup of the scanner device introduces two main measurement errors to the data: on one hand the glass plate with a thickness of 26 mm distorts the image and on the other hand inside of the closed housing internal reflections occur. Alltogether this leads to raw data which is not usable for suffcient accurate measuring, as deviations from a reference plane P_c reach values of up to 20 mm (fig. 1). Both effects are static. For correction a three-dimensional offset O with individual values for each pixel in the sense of a fixed pattern noise (fpn) is computed:

$$P' = \hat{T}(P) + O \ . \tag{1}$$

$$\hat{T} = \underset{T}{\operatorname{argmin}} \sum_i (P_c^i - T(P^i))^2 \ . \tag{2}$$

P denotes the raw measured 3D-data, T is a six-parameter rigid transformation (rotation and translation) to align the glass plate with the XY-plane in the global

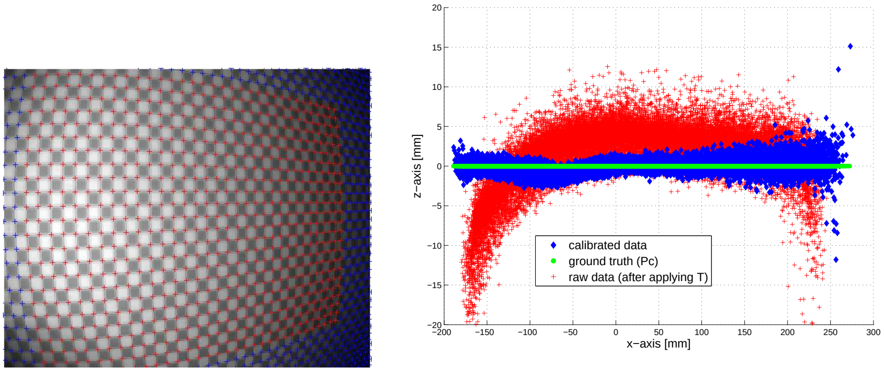


Fig. 1. Calibration pattern with 10 mm features for computing fpn O (left). Comparison of raw data, calibrated data and ground truth plane (right).

coordinate system and i is the index of a given pixel. O corrects the distortion of P. In (tab. 1) results for the measurement accuracy are presented. Mean height-values and standard deviation is shown in colums 2/6. To reduce noise in the raw data a flat plane with homogeneous IR-reflectivity was used. In column 3/7 and 4/8 the measurement accuracy for feature distances within a plane is given. The mean distance and standard deviation between cross sections with 20 $[mm]$ spacing is evaluted.

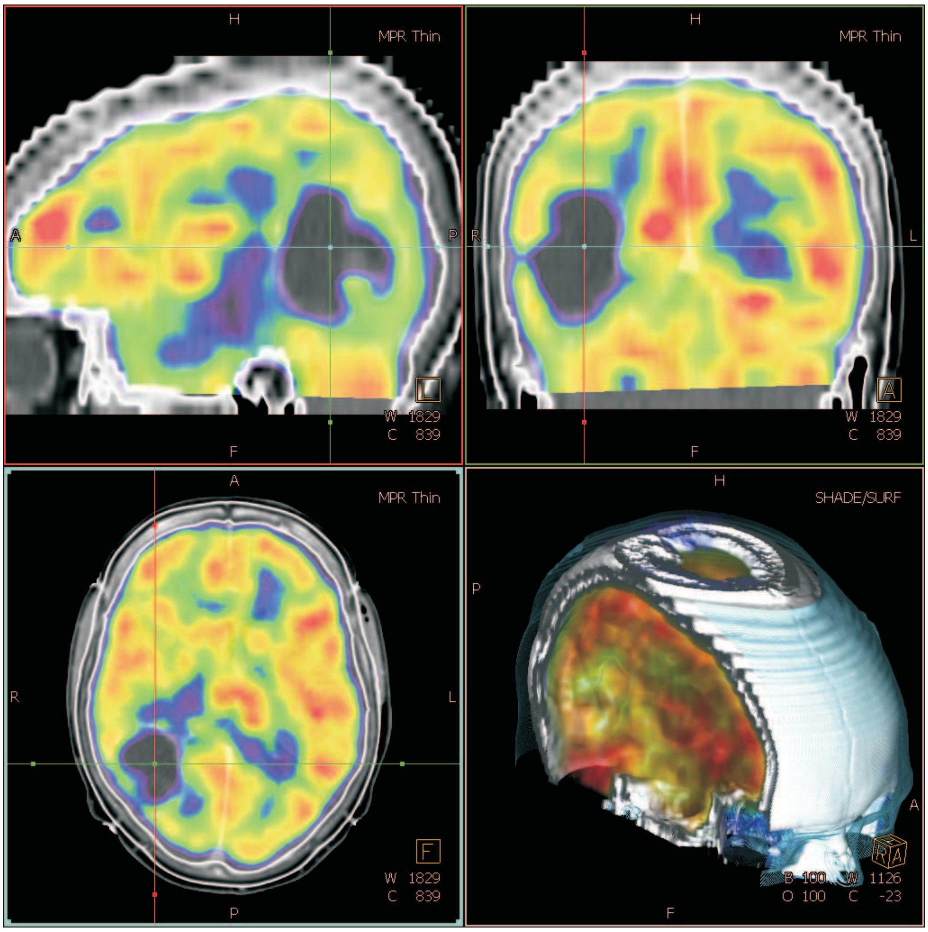
Table 1. Accuracy and noise of measurements. Values given in [mm]. HR: real height, HM: measured height, IP $[x|y]$: in-plane distances in $[x|y]$-direction.

HR	HM	IP x	IP y	HR	HM	IP x	IP y
23.8	25.5 ± 0.94	20.18 ± 0.49	20.15 ± 0.15	100.2	101 ± 0.83	19.99 ± 0.44	19.98 ± 0.16
42.9	44.3 ± 0.89	20.12 ± 0.51	20.10 ± 0.14	119.3	120 ± 0.79	19.96 ± 0.46	19.94 ± 0.17
62.0	63.2 ± 0.84	20.09 ± 0.52	20.05 ± 0.15	138.4	140 ± 0.77	19.96 ± 0.51	19.91 ± 0.17
81.1	82.4 ± 0.84	20.06 ± 0.49	20.00 ± 0.15	157.5	159 ± 0.89	19.92 ± 0.51	19.89 ± 0.19

References

1. T. Schmeltzpfenning, C. Plank *Dynamic foot scanning: A new approach for measurement of the human foot shape while walking.* Footwear Science, Vol. 1, June 2009, pp. 28-30
2. M. Kimura, M. Mochimaru, T. Kanade *Measurement of 3D foot shape deformation in motion.* Proceedings of the 5th ACM/IEEE International Workshop on Projector camera systems, 2008

Registration



Random Walks with Probabilistic Labels

V. Daum[1,2], D. Hahn[2], and J. Hornegger[2]

[1] Department of Nuclear Medicine, University Hospital,
Friedrich-Alexander-University Erlangen-Nuremberg (FAU), Krankenhausstr. 12,
91054 Erlangen, Germany
[2] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Random walk algorithms allow robust and fast image segmentation from initially specified labels on a number of pixels in the image. We propose an extension of the algorithm that allows an initialization with probabilistic labels. This modification makes the random walk algorithm much more suitable for processing the output of automatic classification algorithms, as it allows to reflect uncertainty in the labeling and it can overrule erroneous (outlier) labels.

1 Introduction

Automatic image segmentation approaches often consist of an initial classification of the desired image regions and a subsequent spatial treatment of the classification output. This aims to make the result spatially more consistent, i.e. close holes, remove stray misclassified pixels and adapt the classification result to visible boundaries in the data. The Random Walk algorithm [2] in its original form is designed for semi-automatic image segmentation: The user specifies two or more sets of labeled image pixels and the algorithm determines for every pixel in the image to which of those labels it most likely belongs. To make the algorithm better usable in a setting where labels are generated automatically, we propose the introduction of probabilistic labels, in a way similar to [1]. As a result, it is possible to directly feed the output of probabilistic classification algorithms into the fuzzy random walk.

2 Random Walks for Image Segmentation

Random walks describe the stochastic process of random movements on a graph. We will introduce the random walk algorithm in terms of its namesake random walker: Given an undirected graph and two or more sets of labeled vertices on the graph, we want to determine for each pixel, which set a random walker moving around on the graph will reach first, when started from that pixel. Formally we define the graph G to consist of a set of vertices $V = \{v_1, \ldots, v_n\}$ and a set of edges E. In the context of image segmentation each pixel constitute a vertex v_i and the edge set E contains all edges e_{ij} connecting directly adjacent vertices/pixels v_i and v_j in the image. Additionally, we define edge weights $w_{ij} = w(e_{ij})$, associated with the edges, that indicate how difficult it is for the random walker to cross that particular edge. The weights are usually derived from the gray values at the vertices v_i and v_j. The function w that maps from the edge e_{ij} to the according weight w_{ij} weights can, for instance be expressed in terms of the image intensities x_i at the vertex v_i.

$$w_{ij} = w(e_{ij}) = \begin{cases} e^{-\beta|x_i - x_j|} & \text{if } e_{ij} \in E \\ 0 & \text{else} \end{cases} \tag{1}$$

where β acts as a scaling parameter for the image intensity differences. At each vertex v_i the random walker moves to the neighboring vertex v_j with the transition probability $\pi_{i,j}$. These probabilities are defined in terms of the edge weights.

$$\pi_{i,j} = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}} \tag{2}$$

For interactive use $m \geq 2$ disjoint sets of labeled vertices $L_1, \ldots, L_m$ are defined manually. Additionally the set L_0 contains all the remaining unlabeled vertices $L_0 = V \setminus \bigcup_{k=1}^{m} L_k$. We will use the name of a set of labeled vertices synonymously for the corresponding label. It is then determined for every vertex which set of labeled vertices the random walker is likely to reach first. The probability for the event that the random walker started from vertex v_i reaches a vertex in L_k with $k > 0$ before any other labeled vertex, is denoted as $P(L_k|v_i)$.

$$P(L_k|v_i) = \chi(v_i \in L_k) + \chi(v_i \in L_0) \sum_{\substack{j=1 \\ j \neq i}}^{n} \pi_{i,j} P(L_k|v_j), \tag{3}$$

where χ denotes the indicator function.

3 Probabilistic Labels

In the random walk algorithm, as described above, the labels specified for a vertex are absolute and treated as ground truth. In the case of automatically generated initial labels it is beneficial if they are specified in a way that is less absolute. This is done by replacing the indicator function χ by a priori class probabilities.

$$P(L_k|v_i) = P(v_i \in L_k) + P(v_i \in L_0) \sum_{\substack{j=1 \\ j \neq i}}^{n} \pi_{i,j} P(L_k|v_j). \tag{4}$$

The formulation is, therefore, a generalization of the original equation. Usually, the probabilities provided as input to the random walk by an automatic classification algorithm will not contain an ,,unknown label" L_0. We will call these input class probabilities $P_{\text{input}}(v_i \in L_k)$. This is rectified by replacing them with derived initial label probabilities that have artificially introduced probabilities for the label L_0. They are defined by considering the input probabilities as conditional probabilities with respect to the label L_0.

$$\begin{aligned} P(v_i \in L_k | v_i \notin L_0) &:= P_{\text{input}}(v_i \in L_k) \\ P(v_i \in L_k) &= (1 - P(v_i \in L_0)) P_{\text{input}}(v_i \in L_k) \\ \text{where } 1 \leq k \leq m & \end{aligned} \tag{5}$$

Although it is certainly possible to define a value for $P(v_i \in L_0)$ for each vertex separately, we will use one global value u for all vertices in our experiments.

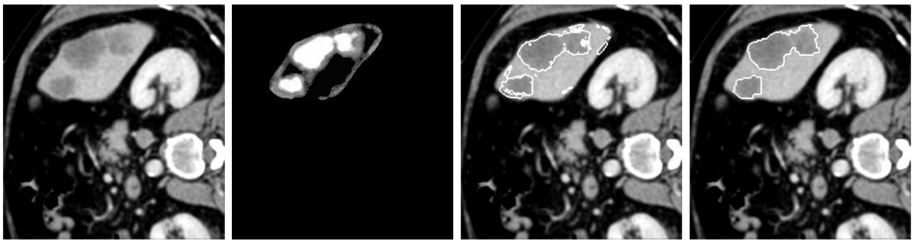


image $P_{\text{input}}(v_i \in L_1)$ $P_{\text{input}}(v_i \in L_1) > 0.5$ $P(L_1|v_i) > 0.5$

Fig. 1. From left to right: image, input label probabilities, resulting segmentation from thresholding $P_{\text{input}}(v_i \in L_1)$ probabilities at 0.5, segmentation after random walk; parameters used are $\beta = 0.02$ and $u = 0.99$.

4 Results

The algorithm, as introduced in this work, has been implemented for 2-D images in Octave/Matlab. The computed tomography (CT) images used in the experiments are preprocessed with minor Gaussian smoothing ($\sigma = 2$) to reduce image noise. As an application example the algorithm is applied in liver tumor segmentation. An initial classification is computed using probabilistic boosting trees [3] adapted for liver segmentation. The results of the initial classification and the refinement by the fuzzy random walk are shown in Fig. 1. Holes in the segmentation are nicely closed and even some very uncertain areas are retained.

Acknowledgment

The authors would like to thank Arne Militzer and Siemens Healthcare for providing the data and preliminary results of their current research on using probabilistic boosting trees for liver tumor segmentation. We also gratefully acknowledge the financial support by the BMBF grant "Medical Imaging", Network project: "Angiogenesis - Targeting for diagnosis and therapy", 01EZ0808, Principal Investigator, Dr. O. Prante, University Erlangen-Nürnberg and by the International Max Planck Research School for Optics and Imaging (IMPRS-OI).

References

1. Leo Grady. Multilabel random walker image segmentation using prior models. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 763–770, Washington, DC, USA, June 2005. IEEE Computer Society.
2. Leo Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, November 2006.
3. Zhuowen Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1589–1596, Washington, DC, USA, 2005. IEEE Computer Society.

Introduction to the Logarithm of Odds

D. Hahn and J. Hornegger

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Statistical shape information is needed for many medical image applications. Quite often, the single shape segmentations are encoded by signed distance maps (SDMs) and a principal component analysis (PCA) is performed. A statistical model is then fitted to the PCA components and high-dimensional SDMs can be projected onto lower dimensional representations. However, it is not obvious how to impose a vector space on SDMs and, in general, the addition of two SDMs is not an SDM anymore. Difficulties arise also when uncertainty about the shape boundaries have to be incorporated into the model. The following article presents an introduction to the concepts of the logarithm of odds (LogOdds) that is applied to shape model generation and that can be used to induce a vector space for probability atlases.

1 Introduction

In medical imaging, many segmentation tasks rely on prior information about the shape in order to improve the robustness of the approach. The training data for the shape information is usually provided by manual segmentation encoded as signed distance maps (SDMs), where the zero level set implicitly describes the boundary of the object. Values smaller than zero indicate the region outside and values larger than zero the inner region of the shape. In comparison to explicit formulations as parametric surfaces, the implicit SDM representation has some advantages: this representation covers topological changes, multiple anatomical shapes can be represented within the same model and the correspondence problem between the training data can be addressed using image registration techniques. In order to analyze the shape variation throughout the training data, a principal component analysis (PCA) is often used and a statistical model is fitted to the PCA components. During the segmentation, the high dimensional SDMs are projected onto the lower dimensional manifold spanned by the PCA components.

The major drawbacks of using SDMs directly for these computations is that SDMs do not form a vector space. The addition of two SDMs, for instance, is generally not a valid SDM anymore. Therefore, a new sample has to be projected back into the manifold of valid SMDs using the PCA components. It is also not clear how to incorporate uncertainty about the validity of the shape boundary into the resulting shape model.

2 Logarithm of Odds

In 2007, Pohl et al. [1] have presented an approach using the logarithm of odds (LogOdds) formulation to define a vector space on probabilistic atlases (PAs). In summary, the LogOdds represents a class of functions that maps the space of M discrete distributions $\mathbb{P}_M$ to the Euclidean space:

$$\begin{aligned}\mathbb{P}_M &= \left\{ \boldsymbol{p} | \boldsymbol{p} = (p_1, \ldots, p_M) \in (0,1)^M \wedge \sum_{i=1}^{M} p_i = 1 \right\} \\ &= \left\{ \boldsymbol{p} | \boldsymbol{p} = \left(p_1, \ldots, p_{M-1}, 1 - \sum_{i=1}^{M-1} p_i \right) \in (0,1)^M \right\}\end{aligned}$$

The multinomial LogOdds function is defined by the logit function that takes the vector of discrete probability distributions as input and maps it to the real numbers:

$$\begin{aligned}\operatorname{logit}(\boldsymbol{p}) &: \mathbb{P}_M \mapsto \mathbb{R}^{M-1} \ , \ \boldsymbol{p} \in \mathbb{P}_M \\ [\operatorname{logit}(\boldsymbol{p})]_i &= \log\left(\frac{p_i}{p_M}\right) \ , \ i \in \{1, \ldots, M-1\}\end{aligned}$$

Its inverse function is the generalized logistic function:

$$[\mathcal{P}(\boldsymbol{t})]_i = \begin{cases} \frac{e^{t_i}}{1+\sum_{j=1}^{M-1} e^{t_j}} \ , \ \text{for } i \in \{1, \ldots, M-1\} \\ \\ \frac{1}{1+\sum_{j=1}^{M-1} e^{t_j}} \ , \ \text{for } i = M \end{cases}$$

The space of discrete probability distributions can, therefore, be associated with the LogOdds space by the logit transform: $\mathbb{L}_{M-1} = \{\operatorname{logit}(\boldsymbol{p}) | \boldsymbol{p} \in \mathbb{P}_M\} \equiv \mathbb{R}^{M-1}$.

In the special case of a Bernoulli distribution for just a two-class problem - many segmentation problems can be reduced to this representation - the formulations can be written in simplified forms in terms of a binomial probability distribution (with $M = 2$ and $p \in \mathbb{R}$) and the sigmoid function:

$$\begin{aligned}\operatorname{logit}(p) &= \log\left(\frac{p}{1-p}\right) \\ \mathcal{P}(t) &= \frac{1}{1+e^{-t}}\end{aligned}$$

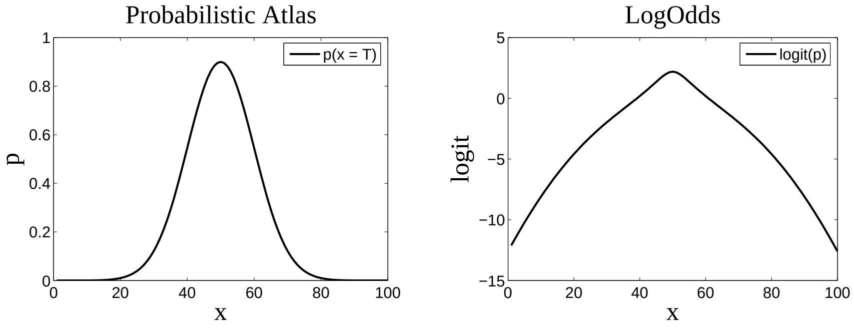
Fig.1 shows a one-dimensional example for the LogOdds representation of a given PA. The location of the object boundary is encoded as the zero level set of the LogOdds function, which corresponds to values of the PA of 0.5.

Assuming the independence of neighboring voxels, the LogOdds allows for a one-to-one mapping between the space of discrete distributions and an Euclidean space. This relations can, therefore, be used to define a probabilistic vector space $(\mathbb{P}, \oplus, \otimes) \leftrightarrow (\mathbb{L}, +, \cdot)$:

$$\begin{aligned}p_1 \oplus p_2 &= \mathcal{P}(\operatorname{logit}(p_1) + \operatorname{logit}(p_2)) = \frac{p_1 \cdot p_2}{p_1 \cdot p_2 + (1-p_1)(1-p_2)} \\ \alpha \otimes p &= \mathcal{P}(\alpha \cdot \operatorname{logit}(p)) = \frac{p^\alpha}{p^\alpha + (1-p)^\alpha} \ .\end{aligned}$$

Interestingly, the probabilistic addition of a normalized class dependent probability and the class prior yields the posterior probability for this class. Given two cases for either encountering class A or not $\overline{A}$ and a feature B, the posterior can be computed by the probabilistic addition operator induced by the LogOdds space:

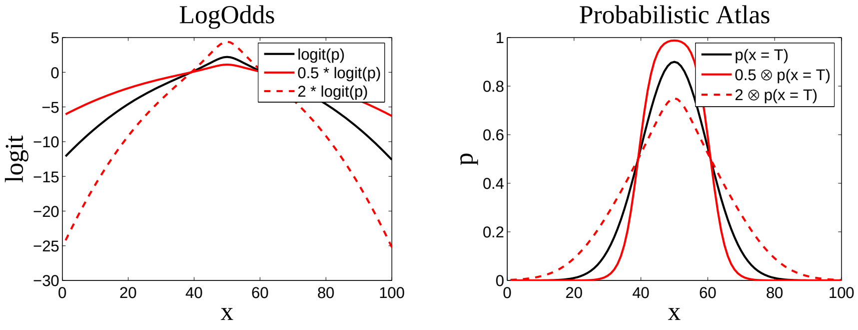
$$\frac{p(B|A)}{p(B|A) + p(B|\overline{A})} \oplus p(A) = p(A|B)$$



(a) (b)

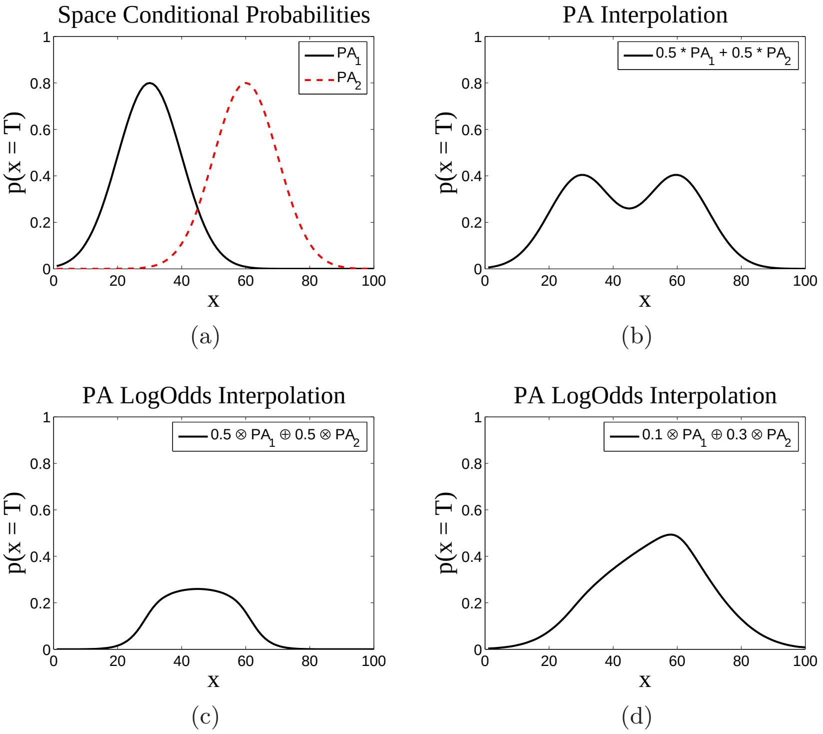
Fig. 1. Probabilistic atlas (spatial probability) (a) with its corresponding LogOdds function (b).

If used for image segmentation and shape model generation, the concept of LogOdds has some nice properties. It embeds SDMs in a vector space, as every SDM is part of the LogOdds space. Like SDMs, LogOdds encodes the boundary of the object implicitly via the zero level set. Moreover, the certainty of the boundary location can be controlled via the probabilistic scaling operator, which is demonstrated in Fig.2. Having defined a vector space on PAs through the LogOdds formulation, combinations between two PAs are no longer restricted to convex combinations and the result may be more intuitive. The following example (Fig.3) shows a combination of two Gaussian-like space conditional probabilities that may, for instance, represent two PAs at different states. A direct and convex combination yields a multi-modal distribution, although both PAs are mono-modal. Using the LogOdds, the result is again mono-modal, a result one would usually expect.



(a) (b)

Fig. 2. A scaling within LogOdds relates to a modified certainty of the boundary location in the PA space.



(a) (b)

(c) (d)

Fig. 3. Interpolation between two spatial probabilities (also known as space conditional probabilities) (a). The direct convex combination yields a multi-modal distribution (b), whereas the LogOdds interpolations (c), (d) result in more intuitive distributions.

3 Summary

The LogOdds concept can be applied in image segmentation methods to incorporate shape knowledge based on PAs. By interpreting the SDMs as part of the LogOdds space, the PA can be defined in terms of the generalized logistic function. The mapping between PAs and the LogOdds also induces a vector space for discrete distributions, which allows for more flexible weightings of the PCA components and also the incorporation of the uncertainty of the boundary location. Existing applications based on the LogOdds formulation are targetting brain segmentation and bias field estimation or also voxel-based classification.

References

1. Pohl, K. M.; Fisher, J.; Bouix, S.; Shenton, M.; McCarley, R. W.; Grimson, W. E. L.; Kikinis, R.; Wells, W. M.: Using the logarithm of odds to define a vector space on probabilistic atlases, Medical Image Analysis, 2007, 11(6), 465-477

2D/3D Registration using CUDA

J. Wasza, D. Hahn, M. Prümmer, and V. Daum

Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany

Abstract. Rigid registration and fusion of pre-operative 3D volumetric data with intra-operative 2D X-Ray images is an important task for many clinical applications, for instance, in non-invasive cancer treatment using radiation therapy. In this work, an intensity based approach is presented which establishes the proper alignment between the data by a gradient descent strategy based on analytically formulated derivatives of similarity measures.

1 Registration method

Intensity based methods address the 2D/3D registration problem by computing *digitally reconstructed radiographs* (DRRs) which are compared to real X-Ray images by means of similarity measures. Common metrics described in literature are the *sum of squared differences*, *pattern intensity*, *gradient difference* and *mutual information* [1]. The registration problem is formulated as a cost function $\mathcal{J}$ resulting in a non-linear optimization problem which can be solved by employing a gradient descent strategy. Using the ray-casting algorithm to generate a DRR, the gradient can be found in the image domain Ω as

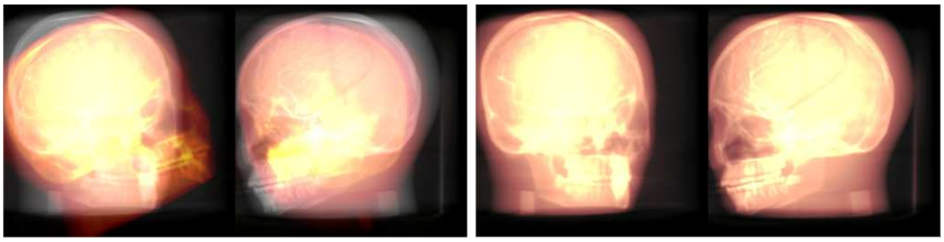
$$\nabla \mathcal{J} = \sum_{i,j \in \Omega} \alpha_{i,j} \int_{Ray} J_{\Phi} \left(\boldsymbol{c} + \lambda \boldsymbol{n}_{i,j} \right) \cdot \nabla \mathcal{V} \left(\Phi \left(\boldsymbol{c} + \lambda \boldsymbol{n}_{i,j} \right) \right) d\lambda \tag{1}$$

α is a weighting factor depending on the actual similarity measure, Φ denotes the rigid body transformation (three rotations, three translations) that aligns the volumetric data $\mathcal{V}$, and J_{Φ} is the Jacobian matrix of the transformation Φ. The optical center $\boldsymbol{c}$ and the direction vector $\boldsymbol{n}$ are defined by the perspective projection properties of the X-Ray device.

As the gradient $\nabla \mathcal{J}$ defined in equation 1 is composed of single gradients at discrete image positions $(i, j) \in \Omega$, the performance of the presented registration method can benefit from an implementation of gradient computation on general purpose graphics processing units (e.g. using Nvidia's CUDA technology [2]).

2 Results

Figure 1 shows the overlay of 3D data for the initial misalignment (a) and after registration (b). For this bi-plane setup (512 × 512 pixels, each) and a volume size of 256 × 256 × 256 the alignment was established in less than 4 s on an AMD Phenom II X4 920 and a Nvidia GeForce GTX 285. The target registration error was 0.4 mm. Experiments on different data sets showed similar results for small to medium initial misalignments (up to 20° rotation, translations depending on the actual problem). However, for large displacements the registration may fail due to the optimization process getting trapped in a local minimum. Regarding registration times, the size of the X-Ray images is the dominating factor.



(a) Initial misalignment (b) After registration

Fig. 1. Overlay of 3D data.

References

1. Penney, G.P. and Weese, J. and Little, J.A. and Desmedt, P. and Hill, D.L.G. and Hawkes, D.J.: A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE Transactions on Medical Imaging Vol.17 Number 4 (Aug. 1998) 586–595
2. Wasza, J.: 2D/3D Registration using CUDA. Diploma thesis at the Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Oct. 2009

An Evaluation Framework for a Fluoroscopy Study of Foot Kinematics

M. Prümmer[1], E. Graf[2], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Human Performance Laboratory, University of Calgary, Faculty of Kinesiology,
Canada

Abstract. In this work foot kinematics are derived. Bones shown in a CT Volume are registered to a temporal sequence of fluoroscopic images. A temporal sequence is acquired using two high-speed cameras mounted each on an image intensifier. The motion of the bones is computed performing a global rigid 3-D/2-D registration of the CT data to frames of a fluoroscopic image sequence. In a further refinement step a local neighborhood of a bone is defined by a segmentation mask such that each individual bone can be aligned in a local coordinate system. The focus of the presented work lies on the registration framework, especially on the optimization. First pre-liminary results are presented. They show that a fast registration can be achieved if most of the optimization work is done in the parameter space of the detector plane, in 2-D. In conclusion we can say that a fast intra-detector plane optimization, combined with rotation degrees of freedom (DOF) in 3-D, provides a fast and robust optimization strategy that outperforms standard 3-D optimization techniques.

1 Introduction

To evaluate the foot kinematics, especially of the bones, a fluoroscopy study with two high-speed image intensifiers is performed. The final goal of this work is to derive a motion model of the foot skeleton during running of a person. A 3-D volume from the foot of a person and a temporal sequence of fluoroscopic images (an example image is shown in Fig. 1) are acquired. This allows to register - using a local bone mask - a single bone from a CT volume to a sequence of fluoroscopic images. An evaluation framework is developed that allows to semi-automatically segment a single bone using the RandomWalk segmentation algorithm [3]. First, a global rigid alignment of a CT volume to each of the frames of the fluoroscopic images is performed and second, the segmentation acts as a local mask such that a local refinement can be performed. This provides a local registration of each single bone.

The problem statement of 3-D/2-D registration has already been addressed intensively in the work of [1]. However, the 3-D/2-D registration problem is still not satisfyingly solved. One challenging problem is due to the fact that the DOF of translation and rotation have a totally different impact to the energy function to be minimized. The energy function is most frequently defined between the forward projection of a CT volume and the acquired fluoroscopic reference images. Only few degrees of rotation can cause a rapid change inside the forward projection and thus the energy function becomes very jagged and the optimization becomes strongly local. Additionally, even in case of a perfect match, the content of the CT projection and the fluoroscopic image can differ among other things by an intensity scale and anatomical structure due to the limited field of view of the reconstructed CT volume. In conclusion of this observation, the choice of a proper similarity measure is important. A comparison between similarity measures and feature based registration is presented in [2].

2 Calibration Issues

A further issue for the kinematic study is the view direction of the two high-speed cameras. An optimal view setup for the registration problem is determined by an empirical study and via simulations using the registration software. The data acquisition environment is experimental and the calibration of a 3×4 projection matrix for each camera is computed based on a calibration phantom as shown in Fig. 1 (left).

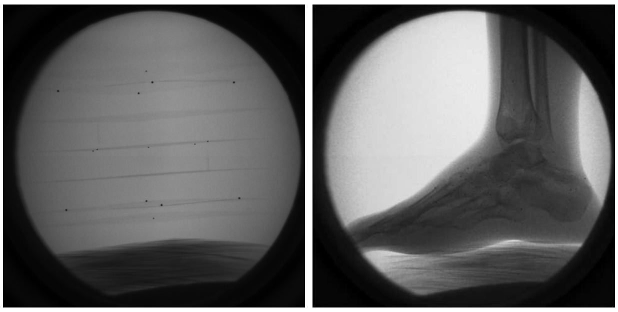


Fig. 1. Calibration phantom (left) and acquired fluoroscopic image (right).

3 Rigid 3-D/2-D Registration

A running trial of fluoroscopic images is not pre-aligned to the CT volume. Thus an initial registration is performed. For the rigid registration a transformation with six DOF has to be estimated. Three for translation and three for rotation. Many standard approaches optimize all six DOFs using a local optimization method like Powell's method. In case of larger rotations the optimization can become quite local and instead of continuing the update of the rotation, the translation according to the source-detector distance of the volume is updated. Such an example is shown in Fig. 2 where the registration failed. Our approach minimizes intra-detector motions in 2-D combined with the resulting DOFs in 3-D. This allows a fast intra-detector rotation and translation and makes the registration more robust.

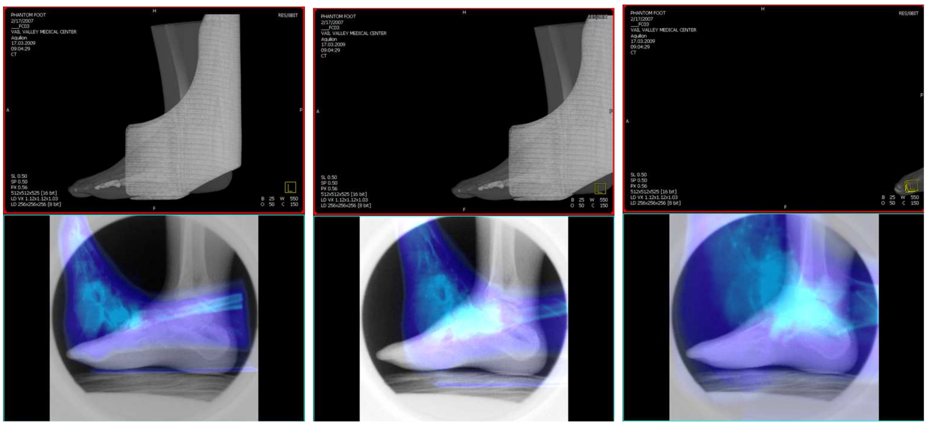


Fig. 2. Standard 3D-2D registration. Left: initial position; middle: after few iterations; right: failed result due to wrong depth translation.

4 Pre-liminary Results

A 3-D CT volume as shown in Fig. 2 (top row) is aligned to one fluoroscopic image. The registration setup is shown in Fig. 3. The initial position of the volume is shown in the left projection image. The middle image shows the reference and acquired fluoroscopic image. The right image shows the registration result using the intra-detector optimization technique.

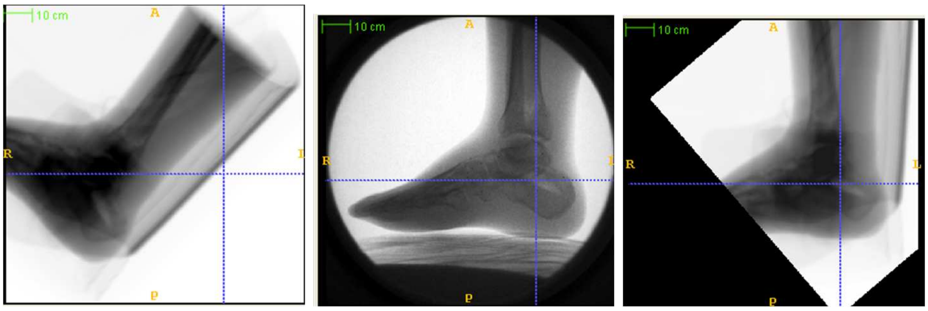


Fig. 3. Pre-liminare result of intra-detector registration.

5 Summary and Conclusions

An evaluation framework for a fluoroscopy study of foot kinematics is under development. A new intra-detector optimization technique is discussed and first pre-liminary results are presented in Fig. 3. The registration software allows to align single bones inside a CT volume to a sequence of fluoroscopic images manually and with an automatic registration method. In conclusion we can say that the optimization strategy of a 3-D/2-D registration is crucial. An intra-detector optimization strategy combined with 3-D DOFs can improve robustness and speed of the registration. The derivation of the foot kinematics is still in progress.

Acknowledgments This work is founded by Adidas, Herzogenaurach, Germany. Furthermore we thank Michael R. Torry at Biomechanics Research Laboratory, Steadman Hawkins Research Foundation, Vail, CO for the great support in hardware and data.

References

1. Lilla Zollei, 2D-3D rigid-body registration of X-ray Fluoroscopy and CT images, M.Sc. Thesis, Massachusetts Institute of Technology, 2001.
2. Robert A. McLaughlin, John Hipwell, David J. Hawkes, Ailson Noble, James V. Byrne, and Tim C. Cox. A Comparison of a Similarity-Based and a Feature-Based 2-D3-D Registration Method for Neurointerventional Use. IEEE Transactions on Medical Imaging, Vol. 24, Issue 8, p. 1058 1066, ISSN: 0278-0062, 2005.
3. Leo Grady and Gareth Funka-lea, Multi-Label Image Segmentation for Medical Applications Based on Graph-Theoretic Electrical Potentials, issn 0302-9743, vol. 3117, ECCV, p. 230–245, Springer, 2004.

Motion Compensation by Registration-Based Catheter Tracking

A. Brost[1], R. Liao[2], N. Strobel[3], and J. Hornegger[1]

[1] Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen,
Germany
[2] Siemens Corporate Research, Princeton, NJ, USA
[3] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. Atrial fibrillation is the most common sustained arrhythmia. Treatment is performed under fluoroscopic image guidance. Overlay images from pre-operative 3-D volumetric data can be used to add anatomical detail otherwise not visible under X-ray. Unfortunately, current fluoroscopic overlay images are static. We developed a system for image-based 2-D motion estimation and compensation as a solution to this problem. Respiratory and cardiac motion at the site of ablation is estimated by tracking a catheter device in fluoroscopic images. In our experiments, the circumferential catheter was successfully tracked in 13 sequences with an average tracking error of 0.6 mm.

1 Motivation

Recent research in the area of X-ray guidance for electrophysiology procedures found that augmented fluoroscopy using overlay images rendered from 3-D images facilitates more precise catheter navigation and a reduction in fluoroscopy time. Unfortunately, current image overlay approaches still lack the possibility for motion compensation. A first approach for 3-D respiratory motion compensation based on 3-D device tracking was proposed by the authors in [1,2]. In this work, we describe a method that requires only mono-plane fluoroscopy and therefore works in 2-D.

2 Motion Compensation by Catheter Tracking

The circumferential mapping catheter on the imaging plane is extracted by manual clicking followed by fast marching in the first frame of the fluoroscopy sequence. The 2-D catheter model is then calculated as an ellipse from the obtained points. After the catheter model has been generated from the first frame of the fluoroscopic sequence, it is tracked throughout the remainder of the sequence. To minimize the influence of peripheral structures, the region of interest is restricted around the center of the tracked catheter in the previous frame. Histogram equalization is applied to improve contrast. A vessel enhancement filter is used to improve line-like structures. This image is binarized using Otsu's method. Further, we apply a skeletonization algorithm. Finally, a distance map is calculated. These steps are summarized in Fig. 1. Catheter model tracking is achieved by performing model-based registration. The estimated 2-D translation can be directly applied to the 2-D overlay to move it in sync with the tracked device.

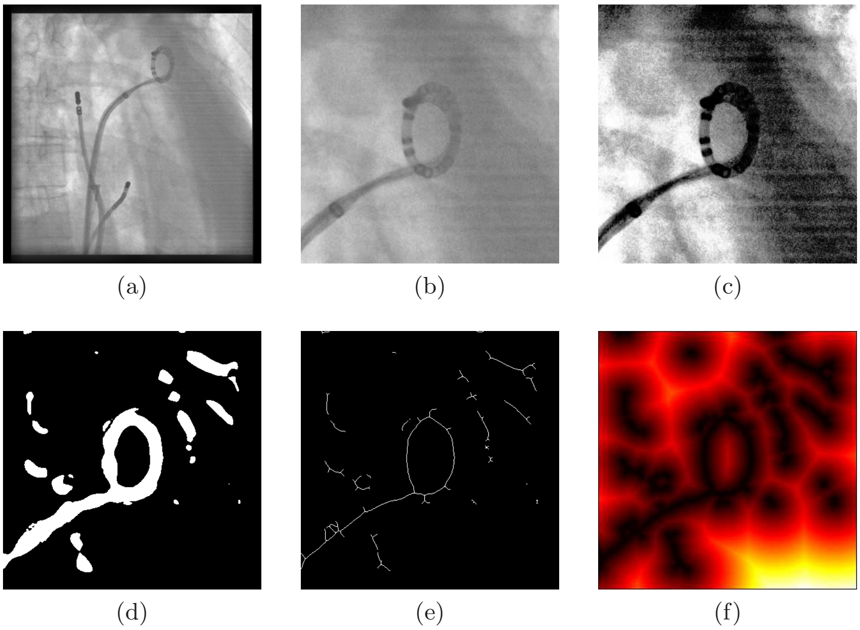


Fig. 1. Image processing steps to enhance the elliptical structure of the catheter.

3 Results and Conclusions

Catheter tracking is based on a model-based registration framework. We start by estimating a catheter model from the first frame of a fluoroscopic sequence. This model is then tracked throughout the remainder of the sequence using model-based registration. A clinical data set of 13 sequences was used for testing. The catheter was successfully tracked with an average tracking error of 0.6 mm.

Acknowledgements

The authors would like to thank Dr. med. Johannes Rieber for the gold-standard segmentation of the catheter and Teri Moore for collecting the clinical data. This work was supported by Siemens AG, Healthcare Sector, Forchheim, Germany.

References

1. Brost, A.; Liao, R.; Hornegger, Joachim; Strobel, N.: 3-D Respiratory Motion Compensation during EP Procedures by Image-Based 3-D Lasso Catheter Model Generation and Tracking. In Yang, G.Z., Hawkes, D.J., Rueckert, D., Noble, J.A., Taylor, C.J., eds.: Lect. Notes Comput. Sci., MICCAI '09: Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention, Volume 5761., London, UK (Sept. 2009) 394-401
2. Brost, A., Liao, R., Hornegger, J., Strobel, N.: 3D model-based catheter tracking for motion compensation in EP procedures. SPIE Medical Imaging, Volume 7625., San Diego, CA, USA (Feb. 2010) 762507

Author Index

.

www5.informatik.uni-erlangen.de

www.hss.de

Hanns
Seidel
Stiftung

.

ISBN 978-3-00-031916-7