Joachim Hornegger, Kurt Höller,
Philipp Ritt, Anja Borsdorf,
Hans-Peter Niedermeier
(Eds.)

HSS-Cooperation Seminar

Pattern Recognition in Medical and Health Engineering

Proceedings of the
HSS-Cooperation Seminar
"Ingenieurwissenschaftliche Beiträge
für ein leistungsfähigeres Gesundheitssystem"
Wildbad Kreuth, Bavaria, July 22-25, 2008

Friedrich-Alexander-Universität
Erlangen-Nürnberg
TECHNISCHE FAKULTÄT

Volume Editors:

Joachim Hornegger
Kurt Höller
Philipp Ritt
Anja Borsdorf

Chair of Pattern Recognition (LME)
Erlangen, Germany

Hans-Peter Niedermeier

Hanns Seidel Foundation (HSS)
Munich, Germany

Technical Support:

Sebastian Korschofski
Martin Wegner

Erlangen, Germany

Printed in Erlangen, Germany

ISBN 3-921713-34-1

Preface

The research at the Chair of Pattern Recognition of the Friedrich-Alexander-University of Erlangen-Nuremberg concentrates on both the foundations and applications of pattern recognition in medical and health engineering. This rather diversified and interdisciplinary research focus is reflected by the lab structure and by a complex international network of industrial and scientific partnerships. The teams work on speech recognition, digital sports and sportronics, medical image processing and computer vision.
With more than 40 researchers and roughly 20 industrial and scientific partners, the success of the LME substantially relies on the intense and regular discussion of scientists of the various disciplines, the interdisciplinary interaction with medical experts, and the prototype system implementation and evaluation with industrial partners. With regular workshops like the Cooperation Seminar with the Hanns-Seidel-Foundation on "Engineering Science Contributions for a more Efficient Healthcare System" LME-researchers have a platform to present their recent achievements, discuss these with others, and consider the implications and impact of results to science, industry and society. Thus the main purpose of this workshop was to intensify the interdisciplinary dialog at the LME in an easy-going innovative environment that cannot be provided in the office.
The proceedings summarize the scientific contributions of the workshop and give an overview of the research topics. I know that the proceedings will not reflect the inspiring atmosphere of the workshop and the high motivation of its participants, but I guarantee that we had a great and productive time in Wildbad Kreuth. After three days of hard work, I drew the conclusion that the current research at the LME is at the cutting edge and that I work with a group of excellent, extremely smart and highly motivated young researchers.
The success of this workshop is the success of many people that contributed to the scientific part and helped us to organize the meeting. At this point I want to thank my academic teachers Professor Heinrich Niemann and Professor Volker Strehl. Both gave brilliant and entertaining invited talks. Professor Niemann was summarizing the first years of pattern recognition in Erlangen to the young researchers in the audience. Professor Strehl presented an overview of the life and achievements of the German mathematician August Ferdinand Moebius. Last but not least I want to thank the restless people who contributed to the organization of the workshop and the Hanns-Seidel-Foundation which was a perfect host.

July 2008

Joachim Hornegger

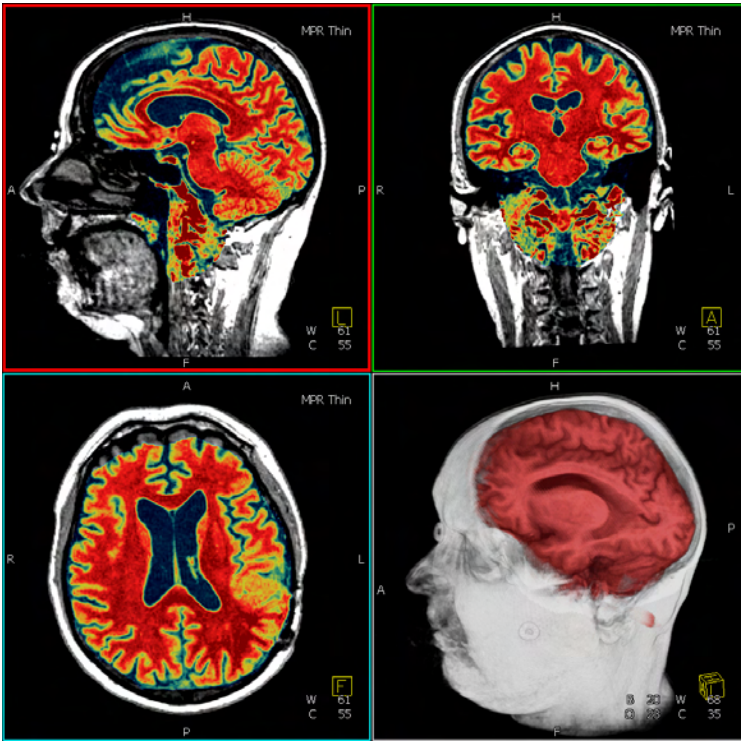
Head of the Chair of Computer Science 5 (Pattern Recognition)
of the Friedrich-Alexander University Erlangen-Nuremberg

Workshop Initiator

II

Table of Contents

Basics of Medical Image Processing



Dynamic CT Reconstruction: ART vs. Filtered Backprojection

M. Prümmer[1] and J. Hornegger[1]

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany.

Abstract. The reconstruction of moving objects like in cardiac applications requires reconstruction techniques that can incorporate a motion model. State-of-the-art reconstruction techniques like ART and FBP assume *well structured sampling* of the discretized reconstruction volume as given for parallel- and fan-beam geometry. For these cases shift-invariant point-spread functions (PSF) are observed and fast algorithms can be derived that allow in theory an exact reconstruction of the object. However, for a non-rigid motion the structured sampling might become individual for each lattice point in the image. In this work first considerations about the performance of ART and FBP according to the motion model dependent PSF are discussed.

1 Methods

In algebraic reconstruction the scan geometry and used pixel model defines the *system matrix* $\boldsymbol{A}$. It contains in each row the contribution (weights) of a specific ray that intersects a subset of pixels in the image $\boldsymbol{f}$ that is reconstructed. This leads to a linear system of equations:

$$\boldsymbol{A}\boldsymbol{f} = \boldsymbol{p}, \tag{1}$$

where the vector $\boldsymbol{p}$ contains the measured X-ray image data. Theoretically the system can also be solved via computing the *pseudo inverse*

$$\begin{aligned} \boldsymbol{f} &= \boldsymbol{A}^{\dagger}\boldsymbol{p} && (2)\\ &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{p}. && (3) \end{aligned}$$

In practice the system matrix is large and sparse and thus iterative solvers are applied to solve for the unknown image $\boldsymbol{f}$. In the following we use the formalism of the *pseudo inverse* as a link between ART and FBP. Generally spoken the matrix $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ defines the inverse of the PSF given by Grams matrix. If the PSF is shift-invariant, the reconstruction problem can be solved more efficiently since each image pixel can be reconstructed via an unfiltered backprojection $\boldsymbol{A}^T\boldsymbol{p}$, followed by filtering the blurred backprojected image using the inverse of the PSF. In case of a well structured sampling (fig. 1 left image), we observe a shift-invariant PSF and thus a fast shift-invariant filter can be applied. This reduces the complexity of the filtering significantly and thus results in a fast reconstruction algorithm. However for a non-rigid motion model the PSF can result in a shift-variant system (fig. 1 right image).

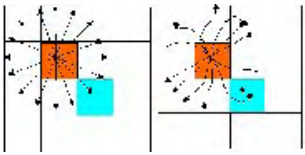


Fig. 1. Left image: regular structure for a parallel-beam; each pixel experiences the same sampling; there is always a ray that creates dependency between two pixel. Right image: may get pixel dependent sampling due to non-rigid motion; unequal dependency between pixel.

Dynamic ART For ART the motion model $\boldsymbol{u}$ can be introduced into the scan geometry (projector model) since the object and scan geometry are moving relative to each other. Therefore we adapt the system matrix according to the motion such that each single ray equation is modified to compensate the objects motion. Using a dynamic Kaczmarz approach we get

$$\boldsymbol{A}_{\boldsymbol{u}}\boldsymbol{f} = \boldsymbol{p} \tag{4}$$

$$f_j^{(n+1)} = f_j^{(n)} + \gamma \sum_{i=0}^{M-1} \frac{a_{i,j}^{u}}{||\boldsymbol{A}_{\boldsymbol{u}}^{i}||}(\boldsymbol{A}_{\boldsymbol{u}}^{i}\boldsymbol{f}^{(n)} - \boldsymbol{p}_i), \tag{5}$$

where f_j is the j-th image pixel, $\boldsymbol{A}_{\boldsymbol{u}}^{i}$ the i-th ray equation (modified according to the motion model $\boldsymbol{u}$), $\boldsymbol{p}_i$ the intensity observed at the i-th detector pixel and n denotes the iteration index. Neglecting the issue of data completeness to be able to reconstruct the object exactly, we are currently interested in a motion compensated least-square solution. The dynamic ART approach allows to reconstruct a moving object even for a highly non-rigid motion as shown in fig. 2. Probably one of the most important advantages of ART-solvers is the fact that no explicit filtering using the inverse PSF is required. This is done iteratively and in an implicit manner. However, the complexity of a dynamic forward and backward projector is much higher compared to a regular structured projector geometry.

Dynamic Filtered Backprojection For static objects the inverse PSF describes a shift-invariant filter system. Each row in $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ represents the same filter, but shifted according to the current pixel position that is filtered. The PSF-matrix $\boldsymbol{A}^T\boldsymbol{A}$ describes the point spread we get if the ideal image is forward projected and then afterwards again backprojected. Each image pixel $\boldsymbol{f}_j$ experiences the point-spread as given in the j-th row in the PSF-matrix. Such matrices are also called circulant *Toeplitz* matrices [1], since they describe a shift-invariant filter.

$$\boldsymbol{f} = \underbrace{(\boldsymbol{A}^T\boldsymbol{A})^{-1}}_{\text{inverse PSF}} \underbrace{\boldsymbol{A}^T\boldsymbol{p}}_{\text{unfiltered BP}} \tag{6}$$

The important property is that if $\boldsymbol{A}^T\boldsymbol{A}$ is Toeplitz, a fast FBP approach can be derived. Instead of performing an iterative dynamic ART approach a direct solution using a FBP

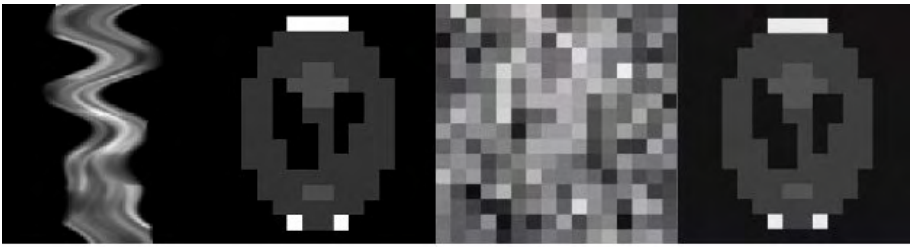


Fig. 2. Comparison between ART using a constant projection geometry and an according to the motion adapted dynamic geometry. From left to right: sinogram of a non-rigid moving shepp-logan phantom, ground-truth phantom, result of a non-motion corrected Kaczmarz approach, result of dynamic Kaczmarz.

approach can be derived. An example of a block Toeplitz matrix is presented in fig. 3. Here we observe a block-like structure, since the object size is bounded and therefore image pixels that lie close to the image boundary observe a sparser sampling compared to pixel that lie close to the rotation center of the scan geometry. For our considerations we assume that the object is not bounded and thus we observe in case of a static object a regular sampling of each image pixel as shown in fig. 1 (left image). However, if

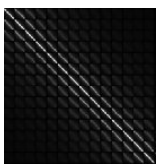


Fig. 3. Example of a PSF-matrix $\boldsymbol{A}^T\boldsymbol{A}$ for a static object (intensity encoded entries).

the object is moving and we adapt the system matrix to compensate for motion, the structure of the PSF-matrix changes. The structure of $\boldsymbol{A}^T\boldsymbol{A}$ can be used to investigate different motion models and conclude for each model if a resulting shift-invariant PSF is observed and thus a FBP approach can be derived. The PSF-matrix describes a shift-invariant filter if all diagonals are constant. Principally all pixel pairs $(\boldsymbol{f}_j, \boldsymbol{f}_l)$ that define a line with the angle θ contribute to the same diagonal in $\boldsymbol{A}^T\boldsymbol{A}$. The diagonal is specified by the angle θ. If all of these pixel pairs observe the same dependency - constant for each individual projection angle, but not necessarily same for different angles - according to all rays that connect a pixel-pair, the PSF describes a shift-invariant

system.

Conclusions The pseudo-inverse is a powerful theoretical framework to investigate both, ART and FBP for dynamic CT reconstruction. The filter complexity can be derived directly from the PSF-matrix. FBP approaches can be used for motions that can be modeled at least linear along each individual ray-equation.

References

1. Robert M. Gray: Toeplitz and circulant matrices: A review. Technical Report, Now Publishers Inc, 2002.

Optimization: The Father of all Problems in Registration

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

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] International Max-Planck Research School on Optics and Imaging

Abstract. Gradient descent based optimization, which is commonly used for registration techniques suffers greatly from different parameter scalings. Especially in rigid registration the rotational and translational parameters of the transformation are usually scaled very differently. Established, as well as new methods are discussed that help overcome this problem.

In the medical field registration or image fusion as it is often referenced as well is used for the alignment of images taken from different modalities or at different points in time (for an overview see [1]). The quality of the alignment is evaluated through an energy functional, the so called distance measure, that depends on a transformation $\Phi_{\mathbf{r},\mathbf{t}}$ and the two images R and T that have to be matched:

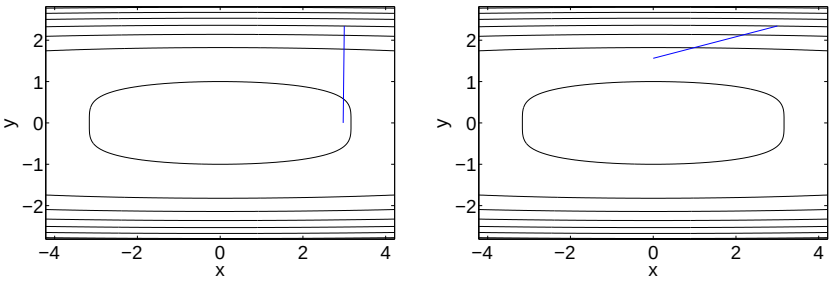
$$\min_{\mathbf{r},\mathbf{t}} \mathcal{D}(R, T \circ \Phi_{\mathbf{r},\mathbf{t}}) \tag{1}$$

For this work we will limit ourselves to rigid registration (i.e. only translation and rotation of the images). The transformation $\Phi_{\mathbf{r},\mathbf{t}}$ therefore only depends (in $3 - D$) on three rotational $\mathbf{r} = (r_1, r_2, r_3)$ (Euler angles, versor) and three translational parameters $\mathbf{t} = (t_1, t_2, t_3)$. Obviously a change in a rotational or a translational parameter can have quite different effects on the distance measure depending on

- image size
- center of rotation
- current quality of alignment

The optimization of this energy functional is commonly performed by gradient descent optimization. The problem with gradient descent based optimization is that the actual gradient yields a descent direction in each component of the gradient vector, but the overall direction is dominated by the parameters that vary the most (see figure 1(a)).

In order to compensate this, the parameter scaling can be estimated or guessed beforehand. For example the Insight Toolkit ITK [2] offers the user to set the according scaling parameters by hand. Another possibility is the application of Newton's method, which makes use of second order derivatives. The inverse Hessian used in Newton's method scales the components of the gradient vector back to the same order of magnitude. Locally, this works not only for quadratic functions but also for more general nonlinear functions (see figure 1(b)). But the use of Newton's requires the calculation of the second derivative of $\mathcal{D}$. This is not only very complicated but also error prone,



(a) Gradient descent is influenced by the different (linear) scalings of the nonlinear function.

(b) Newton's method has at least an estimate of the different scalings and comes much closer to the actual minimum.

Fig. 1. Example of a single step of gradient descent and Newton's method on the $2-D$ function $f(x,y) = x^2 + 10y^4$.

.

as this also requires the computation of spatial second order derivatives on the images, which is always very sensitive to noise.

The use of the Hessian in Newton's method actually does more than is really needed: It scales along the major axes / Eigenvectors of the variation of the gradient. In the special application of rigid registration it is known beforehand that the major scaling differences will be between the rotational and translational parameters. Therefore a diagonal approximation of the Hessian, with a constant entry for the rotational and the translational parameters each, should be sufficient to solve the scaling problem.

$$\mathcal{H}_{\mathcal{D}} = \begin{pmatrix} s_r & 0 & 0 & 0 & 0 & 0 \\ 0 & s_r & 0 & 0 & 0 & 0 \\ 0 & 0 & s_r & 0 & 0 & 0 \\ 0 & 0 & 0 & s_t & 0 & 0 \\ 0 & 0 & 0 & 0 & s_t & 0 \\ 0 & 0 & 0 & 0 & 0 & s_t \end{pmatrix} \tag{2}$$

These parameters can be estimated numerically by performing directional derivatives.

$$s_r = \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial \mathbf{v}_r} \tag{3}$$

$$s_t = \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial \mathbf{v}_t} \tag{4}$$

The directions along which the derivatives are performed can be chosen straightforward as

$$\mathbf{v}_r = (1,1,1,0,0,0) \tag{5}$$

$$\mathbf{v}_t = (0,0,0,1,1,1) \tag{6}$$

Another possible choice that performed really well in experiments is to use the actual descent direction in the rotational and translational parameter sets as the direction for the numerical derivatives.

$$\mathbf{v}_r = (\frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial r_1}, \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial r_2}, \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial r_3}, 0, 0, 0) \tag{7}$$

$$\mathbf{v}_t = (0, 0, 0, \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial t_1}, \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial t_2}, \frac{\partial \mathcal{D}'_{\mathbf{t},\mathbf{r}}}{\partial t_3}) \tag{8}$$

If the iteration is already close enough to the optimum it is also possible to keep s_r and s_t constant for several iterations (experiments showed that an update after 3 – 5 iterations is reasonable).

In summary it is therefore possible to estimate a good enough approximation of the Hessian by only two additional evaluations of the gradient of the distance measure every 3 to 5 iterations and thus overcome the scaling problems.

References

1. D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):R1–R45, 2001.
2. Luis Ibáñez, Will Schroeder, Lydia Ng, and Josh Cates. *The ITK Software Guide*. Kitware Inc., second edition, 2005.

Evaluation - The Ignored Stepchild of Image Registration

D. Hahn[1] and V. Daum[1]

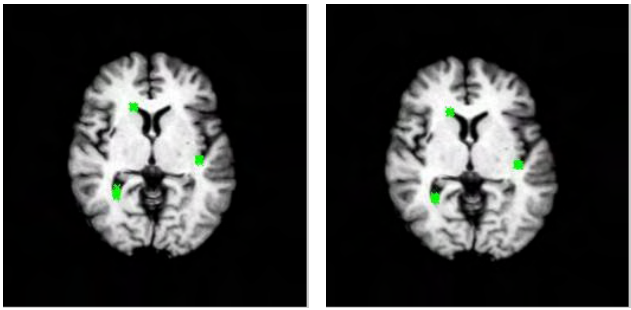
Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. Evaluation of image registration techniques, especially of non-rigid algorithms, is still an open issue. Usually, there exists no ground truth method that allows to quantify the results. Nonetheless, a profound evaluation is necessary to make objective statements about the accuracy and robustness of a registration algorithm.

An evaluation technique for image registration algorithms has to address several aspects. First of all, it has to support both rigid and non-rigid registrations. State-of-the-art evaluation techniques only quantify results of rigid methods. Non-rigid evaluations currently only consist in the generation of an artificially created ground truth deformation that is tried to recover during the registration. This clearly introduces a bias into the evaluation. A manual evaluation by medical experts is widely accepted, however, achieving the necessary measurements is very time consuming and more often than not, the results are only applicable for a single study, as the commercial software packages have no storage support for this kind of data. A quantitative evaluation has to address the accuracy and the robustness of an algorithm to draw objective conclusions about its quality. The accuracy has to be measured in terms of the exactness of a computed registration transform. The robustness criterion provides information about the robustness of the algorithm to achieve the same result from varying starting positions.

The goals of the proposed evaluation system can be briefly summarized as follows: an evaluation database has to be created for both rigid and non-rigid registration algorithms. Based on the knowledge of medical experts, the observer uncertainty has to be incorporated as one cannot expect the registration to deliver better results in regions of the images where even the experts are unsure. The entire system has to be reusable for multiple algorithms, comparisons between them and subsequent medical studies.

The proposed evaluation method consists of a combination of manual landmark localization with statistical information about the observer variability. A list of landmark positions is specified for the entire human body. The evaluation algorithm has to dynamically select the joint subset of the landmarks contained in both images for which the registration quality has to be stated. As every single landmark that is specified by the physician can be regarded as a point feature, the selection of the landmark is assumed to be normally distributed. From a practical point of view, this assumption is plausible as the process of selecting a specific position is a classical normal distributed process in nature. Nonetheless, the goodness of fit has to be varified using stochastic tests in order to undermine this statement. Every selected landmark is treated as a sample from a normal distribution that belongs to one observer. The shape of the distribution then provides the information about the observer uncertainty. With an increasing sharpness of



(a) Reference image. (b) Second mode of variation.

(c) Deformation field.

Fig. 1. Example deformation used for the computation of the measure (1).

the distribution, i.e. a smaller variance, the observer certainty increases and vice versa. With normally distributed landmark selections in two images, the accuracy can be measured by the application of the symmetric Kullback-Leibler divergence (KL) between the approximated normal distributions for the L landmark localizations:

$$\mathcal{M}(\Phi \circ R, T) = \sum_{i=1}^{L} \frac{\mathrm{KL}(\Phi \circ \mathcal{N}_{R,i} \| \mathcal{N}_{T,i}) + \mathrm{KL}(\mathcal{N}_{T,i} \| \Phi \circ \mathcal{N}_{R,i})}{2} , \tag{1}$$

where Φ is the estimated transform between the reference R and the template image T. $\mathcal{N}_{R,i}$ denotes the normal distribution for the i-th landmark location in the image R. The symmetric KL is used in order to achieve results independent of the order of the input images for the registration. In case of a rigid transform, the distance $\mathcal{M}$ can be analytically computed using the density transform theorem and the closed form solution for the KL between normal distributions [1]. In case of a non-rigid registration, the distributions are estimated from the transformed features, which again allows for the closed form solution of the measure. Below is an example of the measure for the following images with three landmark positions of different uncertainty (green points): The registration has been computed on these images with a weighting on the deformation of 0, 0.5 and 1 and the according difference images:

(a) $M = 91.92$, weighting of 0 (b) $M = 21.86$, weighting of 0.5

(c) $M = 0.97$, weighting of 1

Fig. 2. Difference images after application of the computed deformation with a global weighting factor. The rightmost image features the smallest distance between the images with only interpolation artifacts in the difference image.

References

1. S. Yoshizawa and K. Tanabe. Dual differential geometry associated with the Kullback-Leibler information on the Gaussian distributions and its 2-parameter deformations. *SUT Journal of Mathematics*, 35(1):113–137, 1999.

Computed Tomography

Quantitative Computed Tomography by Means of Local Spectral Weighting

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

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Siemens Healthcare, Forchheim, Germany

Abstract. The measured attenuation coefficient in conventional Computed Tomography (CT) is a system dependent quantity due to the polychromatic nature of the radiation produced by the X-ray tube. Local Spectral Weighting (LSW) models the relation between the energy-dependent physical attenuation coefficient and the attenuation values reconstructed from the CT measurements. This information can be used to correct for beam-hardening artifacts and improve image-based methods of quantitative CT like basis material decomposition (BMD) or Rho-Z-projection [1].

Reconstructed CT datasets usually consist of Hounsfield Units (HUs). Water gets the value 0 HU and air -1000 HU. These values are computed from the reconstructed attenuation values by a linear transform. The measured attenuation values are not a quantitative measure as they are system and object dependent. Quantitative CT (QCT) uses multiple measurements of the same object at different system weightings. The measurements are used to generate values that only depend on the material composition and density of the imaged object. The system weighting can, for instance, be changed by altering the acceleration voltage of the X-ray tube in the range of 80 to 140 kV (kVp switching, Dual Source CT).
In the case of monochromatic radiation at energy E_0, reconstructing the physical attenuation coefficient at location $\mathbf{r}$, $\mu(E_0, \mathbf{r})$, can be performed by inverting the Radon Transform. The incoming radiation I is measured on different ray paths $\mathbf{l}_{t,\theta}(\alpha)$ from source to detector channel t at rotation angle θ:

$$I = I_0 \exp\left(-\int_{-\infty}^{\infty} \mu(E_0, \mathbf{l}_{t,\theta}(\alpha))d\alpha\right) \Leftrightarrow P_\theta(t) = \ln\left(\frac{I_0}{I}\right) = \int_{-\infty}^{\infty} \mu(E_0, \mathbf{l}_{t,\theta}(\alpha)) \tag{1}$$

I_0 is the intensity of the incoming radiation if no attenuator is present.
For polychromatic radiation, however, the reconstructed system dependent attenuation value is dependent on the system weighting, which consists of the energy-dependent intensities of the tube spectrum and the energy-dependent detector responsivity. It is also influenced by the surrounding object, as the spectral composition of polychromatic radiation changes as it is passes through matter. This causes a shift in the effective energy of the radiation. Applying the reconstruction for monochromatic radiation as above leads to an inhomogeneous underestimation of the attenuation values. This artifact is

called *beam hardening*. It is especially visible in the vicinity of strong attenuators like bone or metal.
The link between the reconstructed attenuation values $\overline{\mu}(\mathbf{r})$ and the physical attenuation values can be expressed by a weighting function $\Omega(E, \mathbf{r})$, which is called *local weighting function*:

$$\overline{\mu}(\mathbf{r}) = \int_0^\infty \Omega(E, \mathbf{r})\mu(E, \mathbf{r})dE \tag{2}$$

For an infinitesimally small object without any surrounding matter, the local weighting is equal to the system weighting $w(E) = \frac{S(E)D(E)}{\int_0^\infty S(E')D(E')dE'}$. The system weighting depends on intensity of the incoming radiation $S(E)$ and the responsivity detector $D(E)$ [2] (see Fig. 1 for examples of $S(E)$ and $D(E)$).
Computing the local weighting within real objects requires to take the beam hardening caused by the surrounding object and the reconstruction process into account.
The computation of $\Omega(E, \mathbf{r})$ can be realized as an iterative process together with BMD. This local weighting can be used directly for image-based methods of QCT like BMD or Rho-Z-projection by replacing the global system weighting with the local spectral weighting. It can also be used to perform energy calibration of CT datasets and beam-hardening correction.

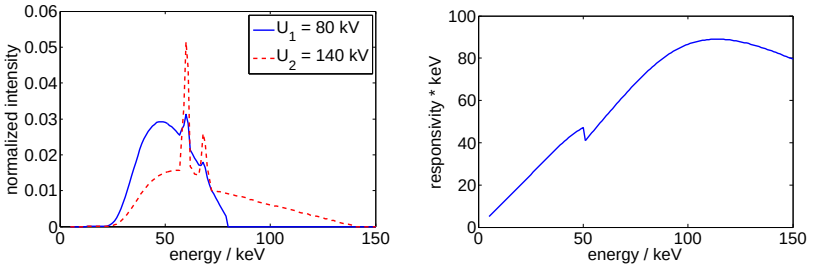


Fig. 1. Left: tube spectrum for $U_1 = 80\,\text{kV}$ and $U_2 = 140\,\text{kV}$ acceleration voltage; right: detector sensitivity for a typical scintillation detector.

References

1. B. J. Heismann, J. Leppert, K. Stierstorfer, *Density and atomic number measurements with spectral X-ray attenuation method*, Journal of Applied Physics, Vol. 94(3), pp. 2074-2079, Aug. 2003
2. S. Wirth, W. Metzger, K. Pham-Gia, B. J. Heismann, *Impact of Photon Transport Properties on the Detection Efficiency of Scintillator Arrays*, IEEE Nuclear Science Symposium Conference, M11-212, pp. 2602-2603, San Diego, Oct. 2006

Orientation Dependent Linearized Noise Propagation Through Indirect Fan-Beam FBP Reconstruction

A. Borsdorf[1,2], S. Kappler[2], R. Raupach[2], and J. Hornegger[1]

[1] Chair of Pattern Recognition (LME),
[2] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. An analytic, orientation dependent propagation of projection based noise variance estimates through indirect fan-beam FBP reconstruction is provided in order to adapt post-processing methods to the local image noise in CT.

Precise knowledge of the local image noise is important for the efficient application of post-processing methods such as bilateral filtering to computed tomography (CT) images. The non-stationary, object dependent noise in CT images directly results from the noise present in the projection data. Since quantum and electronics noise are the main noise sources, comparably simple physical models can be used for noise estimation in the individual projections.

We developed a fast approximate method for analytic propagation of these noise estimates through indirect fan-beam filtered backprojection (FBP) reconstruction [1]. Contrary to earlier publications in this field, we approximate the correlations within and between the parallel projections resulting from the rebinning and the convolution and take it into account for the further processing. The introduced approximations still lead to precise noise estimates below 3.6% relative error and allow an efficient implementation of the noise propagation method.

Furthermore, we compute orientation dependent noise estimates, e.g. for the horizontal and vertical directions [2]. This is done by projecting the noise estimates of the projections onto the horizontal and vertical directions and compute separate noise propagations according to our proposed method. The separation is constructed such that for every pixel in the image the overall noise variance is the sum of the horizontal and vertical noise variance. The resulting orientation dependent noise estimates are subsequently used for adapting bilateral filters. Both parts of the filter, the range and domain filter, are adapted to the non-isotropic and non-stationary nature of noise in CT images. A flowchart of the proposed methodology is presented in Fig. 1.

With this adaptation an improvement in SNR of more than 60% is achieved and a more homogeneous noise reduction results.

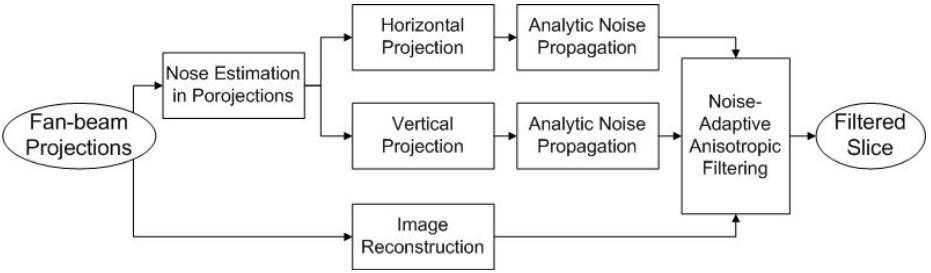


Fig. 1. Block diagram of our proposed method [2].

References

1. A. Borsdorf, S. Kappler al.: Analytic Noise Propagation in Indirect Fan-Beam FBP Reconstruction. Proceedings of the 30th Annual International IEEE EMBS Conference, Vancouver, British Columbia, Canada, 2701-2704, August 2008
2. A. Borsdorf, S. Kappler al.: Analytic Noise Propagation for Anisotropic Denoising of CT-Images. Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference (IEEE NSS/MIC), Dresden, Germany, October 2008 (to appear)

A Novel Method for 1D Atlas Matching of CT Volumes

Johannes Feulner[1] and S. Kevin Zhou[2]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Siemens Corporate Research, Inc., Princeton, NJ

Abstract. A similarity measure for axial CT slices is proposed and used for 1D atlas matching of CT volumes in axial direction. It is largely invariant to translation in sagittal and coronal direction. The restriction on one dimension allows exhaustive search and thus leads to robust results.

Being able to match a CT volume with an anatomical atlas offers important applications like initializing further image analysis steps. This problem is typically solved either by finding anatomical landmarks, which requires a large set of landmarks which can be detected robustly, or by intensity based registration, which is computationally expensive and prone to getting stuck in local optima. However, for many purposes an approximate 1D registration in axial direction is sufficient. Here a similarity measure for axial image slices is proposed which has its origin in scene classification [1,2]. Slices are described by a set of features, and the similarity of two slices is measured by comparing their feature histograms. Using this inter-slice-similarity measure, two volumes are rigidly registered in axial direction. Due to the one dimensionality of the problem and the similarity measure which can be computed efficiently, exhaustive search is feasible, which avoids the risk of local optima. Robust atlas matching is performed by using more than one atlas. After an outlier rejection step, the best registration result is taken as the final one. Figure 1 shows an overview of the registration process. Elevenfold cross-evaluation on 44 CT datasets proved the robustness and the performance of the proposed approach. For evaluation, the volumes were partitioned into pieces of five different sizes in order to measure the influence of the volume height. Especially for mid-size and large volumes, very robust results could be achieved at a reasonable time of 16s for a 100 slices volume.

References

1. Grauman, K.; Darrell, T.: The pyramid match kernel: discriminative classification with sets of image features Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 2005, 2, 1458-1465 Vol. 2
2. Lazebnik, S.; Schmid, C.; Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006, 2, 2169-2178

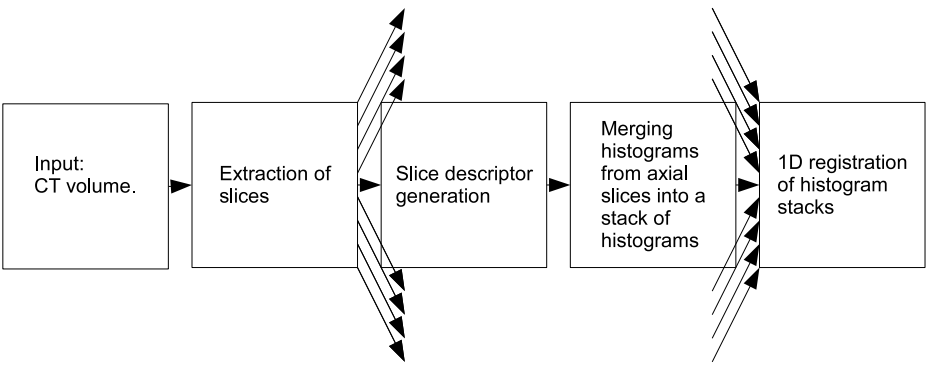


Fig. 1. Overview of the proposed system for 1D atlas matching. After slice extraction, axial slices are processed separately. The resulting slice descriptors are merged again, forming a stack. Resulting stacks are matched by 1D registration.

Shape-based Organ Segmentation

A. Wimmer[1,2], G. Soza[2], and J. Hornegger[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Computed Tomography, Siemens Healthcare Sector, Forchheim, Germany

Abstract. A method for the segmentation of organs using an implicit active shape model is described and evaluation results are presented for the segmentation of the liver from computed tomography scans.

Organ segmentation is frequently used in clinical routine, e.g. for intervention planning. Computed tomography (CT) acquisitions with limited soft tissue contrast and a high degree of noise lead to weak or incomplete organ boundaries, posing great challenge on segmentation methods without high-level knowledge.

The proposed method [1] incorporates high level knowledge in the form of an implicit active shape model [2], which consists of an active contour (resp. surface) whose deformation is constrained by a statistical shape model. Thereby, the formation of admissible shapes is encouraged. Both active contour and shape model are implemented within the level set framework, thus avoiding the need to establish correspondences amongst landmark points on reference shapes during training and the problem of self-intersection of the active contour during evolution.

The deformation of the active contour is guided by an appearance model [3]. A nearest neighbor classifier, consisting of intensity profiles sampled in surface normal direction on and off the true organ boundary during training is used to assign probabilities to image points according to their profiles. The active contour is then moved towards the most probable boundary points.

In order to constrain the deformation of the active contour, a statistical shape model is built by performing principal component analysis (PCA) on a set of reference segmentations, which were rigidly aligned and converted to signed distance maps beforehand. Shapes consistent with the training set are then expressed by the mean over all signed distance maps and a linear combination of the largest eigenmodes obtained from the PCA. The shape constraint is integrated through an energy functional, which quantifies the squared difference between active contour and shape model. The functional is optimized by evolving the active contour with a shape constraint term derived from its Euler-Lagrange equation, and by optimizing weights of the eigenmodes and parameters of an affine transformation aligning the shape model with the active contour.

The procedure of the described method is illustrated in Fig. 1, where both active contour and shape model are shown for different stages during the segmentation of a liver.

An evaluation was performed for the 2-D segmentation of the liver and results were compared to manual reference segmentations. Test cases comprised varying levels of contrast enhancement and both normal and pathologic livers with tumors. Over all 20

Fig. 1. Active contour (red) and shape model (white) for different stages of the segmentation process of a liver.

cases, the average contour difference was about 2 mm and the average area overlap was about 96%.

References

1. A. Wimmer, G. Soza, and J. Hornegger: Implicit Active Shape Model Employing Boundary Classifier. In ICPR 2008, to appear.
2. M. Rousson, N. Paragios, and R. Deriche: Implicit Active Shape Models for 3D Segmentation in MR Imaging. In MICCAI LNCS, volume 3216, pages 209–216, 2004.
3. B. van Ginnecken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever: Active shape model segmentation with optimal features. IEEE Transactions on Medical Imaging, 21(8):924–933, 2002.

Virtual Planning of Liver Interventions

A. Militzer[1 2], M. Sühling[2], and J. Hornegger[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Siemens AG Healthcare, Computed Tomography, Forchheim, Germany

Abstract. A novel liver intervention planning system is presented that will automatically segment liver lesions and provide all means necessary for effective planning of their treatment as well as monitoring of therapies.

Primary liver tumors are among the most frequent primary malignant tumors. Secondary liver tumors are even more common, amounting for about 90% of all liver lesions[1]. Especially colon cancer patients have a very high risk of developing liver metastases at some point [2]. However, so far computer assistance for physicians during diagnosis and surgical treatment of liver tumors is usually very basic.

The presented project aims at overcoming this deficit by providing an intervention planning system that, given a set of CT images, automatically detects and segments liver lesions, characterizes them, and analyzes their position relative to important anatomical structures like vessels or liver segments. This information can then not only be visualized in 3-D (fig. 1, right) for surgeons to get a better orientation. It is also used to simulate interventions including side effects such as liver regions being cut off their blood supply or drainage due to damaged vascular structures. Additionally, with the segmentation available, tumors that have not been resected can be monitored over a longer period of time in order to allow an assessment of tumor growth or shrinkage and thus the success of the treatment.

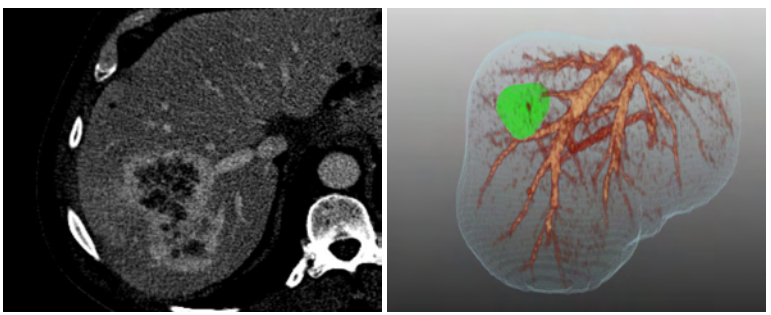


Fig. 1. Left: Venous CT image of the liver showing inhomogenous, border enhancing tumor. Right: 3-D visualization of liver with main venous vascular structures and tumor.

A major contribution of this work will be the automatic detection and segmentation of liver lesions. In particular, we want to focus on inhomogenous lesions and

lesions showing a strong enhancement along the tumor boundary (fig. 1, left). Current semi-automatic segmentation methods, which are often based on thresholding or region growing techniques, usually succeed for well-shaped, homogenous tumors, but fail for the mentioned more difficult cases. Therefore, machine learning techniques will be applied to robustly detect and segment these tumors in images. To achieve this robustness and reliability, information from various sources, such as CT images of different phases of contrast enhancement in the liver, will be incorporated into the process.

References

1. Layer, G; Gallkowski, U.: Lebertumoren. Chap.5 of *Radiologische Diagnostik in der Onkologie: Band 2 Gastrointestinum, Urogenitaltrakt, Retroperitoneum.* Springer Verlag Berlin Heidelberg, 2008, ISBN 978–3–540–29318–7
2. Layer, G.; Kopp, A. ; Müller–Schimpfle, M.: Kolorektales Karzinom. Chap.4 of *Radiologische Diagnostik in der Onkologie: Band 2 Gastrointestinum, Urogenitaltrakt, Retroperitoneum.* Springer Verlag Berlin Heidelberg, 2008, ISBN 978–3–540–29318–7

Magnetic Resonance Imaging

Normalization of Medical Images

F. Jäger

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Many imaging techniques lack a general intensity scale, like the Hounsfield units in Computed Tomography, due to artifacts induced by physical properties of the acquisition method (e.g., magnetic susceptibility effects) or human interaction (e.g. patient positioning). For many applications, however, normalized signal intensities are an unavoidable necessity or at least a possibility to significantly increase the quality as well as the performance of subsequent processing stages. Furthermore, the acquired images can be presented more easily without the need of the adaptation of imaging parameters by the physician. A very important issue in this context is the evaluation of normalization results. The work presented describes various different methods for the assessment of the correction quality in miscellaneous use cases. The method that was evaluated is described in [1]. It computes a non-parametric mapping between the images' intensity spaces, so that their probability density functions are as close as possible. Three different evaluation methods are presented. First, a certain tissue class was perturbed to show the superiority compared to single channel methods. Then, pathologies were enlarged (see Fig. 1) to demonstrate the robustness against severe anatomical changes. Finally, an intra volume slice normalization was evaluated using a distance measure basing on confidence maps.

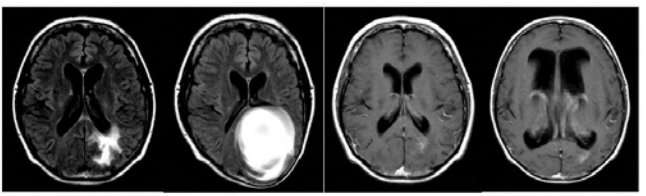


Fig. 1. Artificially enlarged pathologies in MRI head images: in the first two images an artificially evolving lesion is shown; the last two images show artificially enlarged ventricles.

References

1. Jäger, F. and Hornegger, J.: Non-Rigid Registration of Joint Histograms for Intensity Standardization in Magnetic Resonance Imaging, IEEE T-MI, *accepted for publication*

Discriminative Model-Constrained 3-D MR Image Segmentation

M. Wels[1,4], G. Carneiro[2], A. Aplas[3], M. Huber[4], J. Hornegger[a], and D. Comaniciu[2]

[1] Chair of Pattern Recognition, University Erlangen-Nuremberg, Germany,
[2] Siemens Corporate Research, IDS, Princeton, NJ, USA,
[3] Institute of Radiology, University Medical Center, Erlangen, Germany,
[4] Siemens, CT SE SCR 2, Erlangen, Germany

Abstract. We present recent achievements in ongoing projects on discriminative model-constrained 3-D image segmentation in brain MRI.

We report on a further development of our system for fully automated segmentation of pediatric brain tumors in multi-spectral 3-D magnetic resonance images (MRI). The system is based on the recently introduced discriminative model-constrained graph cuts segmentation approach [1] that combines probabilistic boosting trees (PBT) and lower-level segmentation via graph cuts in a unified Bayesian model based on the theory of Markov random fields (MRF). It is assembled by a strong discriminative observation model provided by the PBT algorithm for supervised learning of tumor appearance and a spatial prior that takes into account the pair-wise homogeneity of voxels in a 3-D neighborhood in terms of classification labels and multi-spectral intensities. The discriminative model relies not only on observed intensities at the voxel to be classified but also on the surrounding 3-D context for detecting candidate regions for pathology. Both approaches are integrated into a unified statistical framework. The method is applied to the challenging task of *pediatric* brain tumor segmentation, which is characterized by a high non-uniformity of both the pathology and the surrounding non-pathologic brain tissue. Quantitative evaluations show that our approach produces more accurate results in terms of overlap ratios than those reported for current state-of-the-art approaches to 3-D MR brain tumor segmentation both in adult as well as in pediatric patients. Figure 1 gives an visual impression of the segmentation results that could be achieved. Processing of one multi-spectral data set does not require any user interaction, and takes less time than previously proposed methods.

References

1. M. Wels, G. Carneiro, A. Aplas, M. Huber, J. Hornegger and D. Comaniciu: A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI, to appear in MICCAI 2008, Part I, LNCS 5241, pp. 67–75, Sept. 2008.

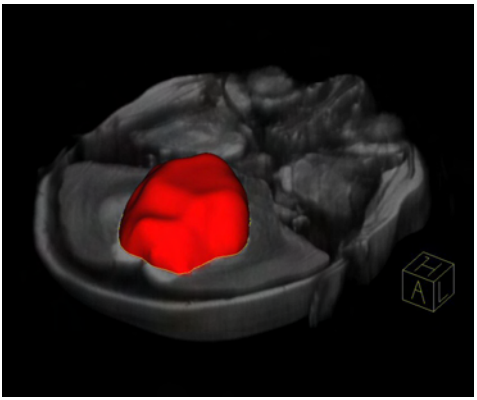


Fig. 1. Rendered result for pediatric brain tumor segmentation overlaid on the T2-weighted pulse sequence. Please view in color.

Interventional Angiography

CUDA based Simultaneous Algebraic Reconstruction Technique

B. Keck[1,2], H. Hofmann[2], H. Scherl[2], M. Kowarschik[2], and J. Hornegger[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Siemens Healthcare, Components, Medical Electronics & Imaging Solutions, Erlangen, Germany

Abstract. The Common Unified Device Architecture (CUDA) introduced in 2007 by NVIDIA is a new programming model making use of the unified shader design of the most recent graphics processing units (GPUs). We apply this novel technology to the Simultaneous Algebraic Reconstruction Technique (SART), which is an advanced iterative image reconstruction method for cone-beam CT.
We present an efficient implementation of the most time-consuming parts of the iterative reconstruction algorithm: forward- and back-projection. We also explain the required strategy to parallelize the algorithm for the CUDA architecture. Our implementation exhibits a speed-up of more than 32 compared to a state-of-the-art CPU using hardware-accelerated texture interpolation.

SART is a well-studied reconstruction method for cone-beam CT scanners. Due to its higher complexity, ART is rarely applied in most of today's medical CT systems. The typical medical environment requires fast reconstructions in order to save valuable time. We have already shown [1] that current GPUs offer massively parallel processing capability that can handle the computational complexity of three-dimensional cone-beam reconstruction.

We use our voxel-driven back-projection from [1] in order to distribute the corrective image onto the volume. Furthermore, a ray-caster based forward-projection [2] is employed to compute the updated projections to be compared to the acquired projections. Here we can make use of the hardware-accelerated interpolation of textures on graphics cards to compute the trilinearly interpolated samples along each projection ray.

Since CUDA 1.1 is not capable of 3-D textures, we implemented a method using a linear interpolation between two texture values that are both bilinearly interpolated by hardware. In our method the volume is represented as a stack of slices (S) organized as a texture. The volume-representing texture, therefore, has to be acquired and synchronized with the volume data (figure 1). Such a synchronization is referred to as a volume update. On the CPU side we used an existing multi-core based reconstruction framework, while using NVIDIAs QuadroFX 5600 on the GPU side. Our test data consists of simulated phantom projections. We used 228 projections representing a short-scan from a C-arm CT system to perform iterative reconstruction with a projection size of 256×128 pixels. The reconstruction yields a $512 \times 512 \times 350$ volume. In order to evaluate the performance of the GPU vs. the CPU we did a reconstruction with 20 iterations, while a volume update, in the GPU case, is performed each tenth projection. Therefore we could achieve a speedup of more than factor 32 compared to a 2GHz Intel Core2Duo processor.

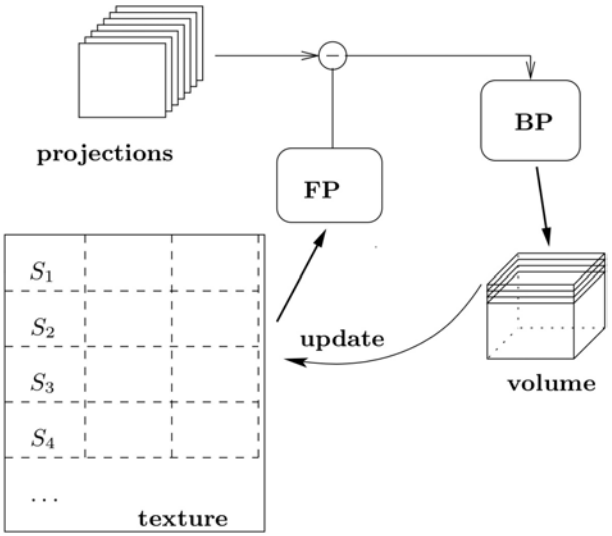


Fig. 1. GPU implementation principle: Volume represented in a texture by slices S_i is forward-projected (FP). After computing the corrective image, the back-projection (BP) distributes the result onto the volume. After performing an update the texture representation of the volume is equal to the volume.

References

1. Scherl H., Keck B., Kowarschik M., and Hornegger J.: Fast GPU-Based CT Reconstruction using the Common United Device Architecture (CUDA). In Eric C. Frey, editor, Nuclear Science Symposium, Medical Imaging Conference 2007, 4464–4466, 2007, Honolulu, Hawaii (USA).
2. Engel, K. et. al.: Real-time volume graphics. A K Peters, 2006.
3. Xu F., and Mueller K.: A comparative study of popular interpolation and integration methods for use in computed tomography. Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on, 1252–1255, April 2006.

Computational Fluid Dynamics - Influence of Segmentation on Resulting Flow Pattern

M. Spiegel[1], T. Redel[2], J. Hornegger[1], and A. Dörfler[3]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Siemens AG Healthcare Sector, Germany,
[3] Department of Neuroradiology, University Erlangen-Nuremberg, Germany

Abstract. The objective of this project is to develop a workflow to perform blood flow simulations by Computational Fluid Dynamics (CFD) of cerebral pathologies i.e. aneurysms. The focus lies on the segmentation of the complex vessel structures and on the examination of flow pattern given various vessel geometries.

Stroke is the third most common cause of death in the western world with 10-15% caused by intracerebral hemorrhage, e.g. rupture of an aneurysm. Today most of aneurysms are detected by accident and the reason of their appearance is not clear. Blood flow simulations will help physicians to provide a better diagnosis with respect to the risk of rupture by getting a closer look on occurring flow pattern. For instance, Castro et al. [1] analyzed the correlation between hemodynamic patterns and aneurysmal rupture. This work concentrates on the segmentation of cerebral vessel structures. For CFD the vessel geometry is the most prominent boundary condition. Therefore an evaluation of the influence of different segmentation results on resulting flow pattern will be included. Figure 1 (b) and (c) depicts one example of flow pattern where the blood inflow gets divided into two branches within the aneurysm.

State-of-the-art vessel segmentation techniques are statistical-based methods [2] and level-set formulations [3]. The basic concept of this work is to develop a segmentation algorithm combining 2-D and 3-D information in order to improve segmentation results. The first step will be to introduce a level-set based segmentation method whose level-set formulation is regularized in a way that tubular structures like vessels are easily segmented and non-tubular structures get penalized. Figure 1 (a) shows an initial result of a segmented vessel structure.

Phantom experiments, contrast agent flow experiments and sensitivity analysis of the segmentation results denote future work in the context of this project. The phantom experiments together with the sensitivity analysis will build a base to get a profound knowledge how different segmentation results will influence flow pattern resulting from CFD simulation. Moreover, a time parameterization of the pixel intensities within 2-D digital substracted angiograms (DSA) seems to be a promising way to gain more information about flow pattern within real vessel pathologies. These can be drawn in comparison to ”virtual angiograms” representing the flow pattern computed from the CFD simulation.

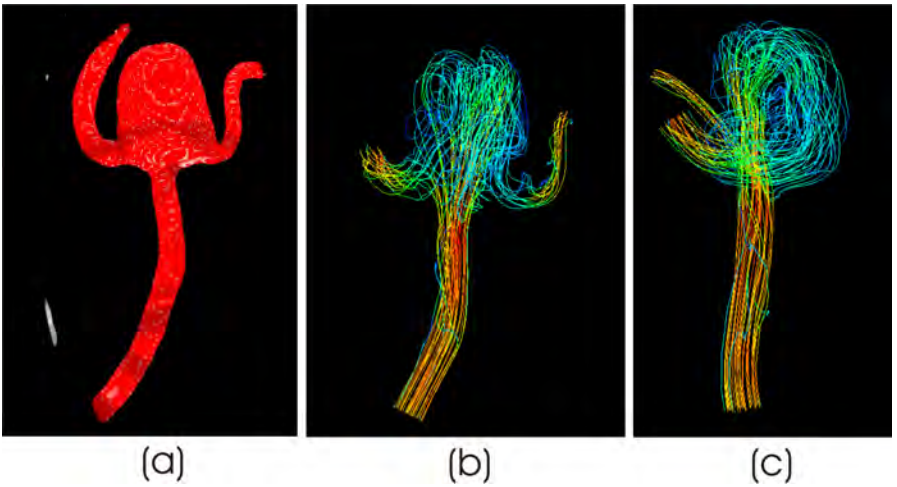


Fig. 1. Figure (a) illustrates a segmented vessel part with an aneurysm. The figures (b) and (c) show one instance of flow pattern within an aneurysm depicted with pathlines colored by velocity magnitude. Blue indicates low velocity values and red high velocity.

References

1. Castro, Marcelo; Putman, Christopher; Radaelli, Allesandro; Frangi, Alejandro; Cebral, Juan: Image-Based Investigation of Hemodynamics and Rupture of Cerebral Aneurysms of a Single Morphological Type: Terminal Aneurysms, Proc. of SPIE Medical Imaging 2008: Physiology, Function and Structure from Medical Images, San Diego, USA
2. Gan, Rui; Chung, A.C.S.; Wong, W.C.K.; Yu, S.C.H: Vascular segmentation in three-dimensional rotational angiography based on maximum intensity projections, IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004, vol. 1, pp.133-136
3. Deschamps, Thomas; Cohen, Laurent D.: Fast Extraction of Tubular and Tree 3-D Surfaces with front propagation methods, In Proc. of 16th ICPR, 1:731 -734, 2002

Tissue Perfusion Quantification With C-arm CT

A. Fieselmann[1,2], M. Zellerhoff[2], A. Ganguly[3], R. Fahrig[3], and J. Hornegger[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Siemens AG, Healthcare Sector, Forchheim, Germany
[3] Department of Radiology, Stanford University, CA, USA

Abstract. The aim of this project is to quantify the regional cerebral blood flow with C-arm CT - a modality that is suitable for 3-D imaging in the interventional suite.

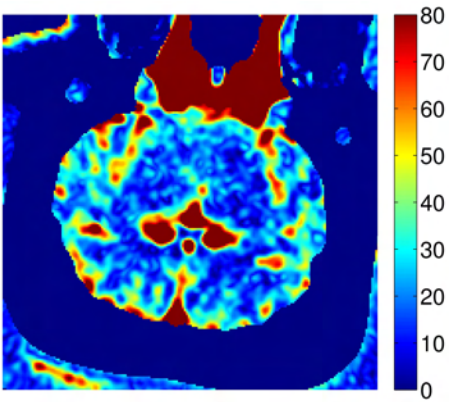


Fig. 1. Map of the regional cerebral blood flow (rCBF) obtained from a perfusion CT scan of a healthy pig (units: $[\mathrm{V_{blood}}/\mathrm{m_{tissue}}/\mathrm{t}]$ = ml/100g/min)

Perfusion CT measures the regional blood flow inside a tissue of interest. It can be used to diagnose stroke, the third leading cause of death in Western countries. Fig. 1 shows a map of the regional cerebral blood flow (rCBF) of a healthy pig that was computed from a perfusion CT scan using an in-house software for perfusion analysis.

A C-arm CT [1] can acquire 3-D images in the interventional suite. Perfusion imaging with a C-arm CT would have strong potentials because less resources would be blocked and the time between the imaging and the intervention could be reduced significantly.

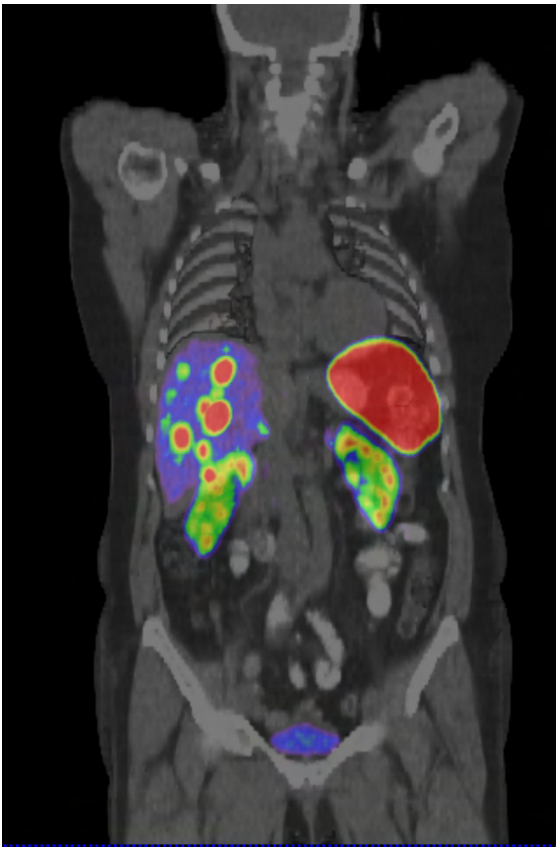
Perfusion C-arm CT is similar to perfusion CT, both modalities study the temporal evolution of a previously injected bolus of contrast agent by acquiring a sequence of scans of the tissue of interest. But in contrast to CT, the C-arm CT image reconstruction algorithms must take into account the slower rotational speed of the C-arm.

This projects focuses on the development of dynamic image reconstruction algorithms for perfusion C-arm CT. Furthermore it deals with the optimization of algorithms for perfusion analysis using scans acquired with a C-arm CT.

References

1. M. Zellerhoff, B. Scholz, E.-P. Rührnschopf, and T. Brunner. Low contrast 3-D reconstruction from C-arm data. In *Proceedings of SPIE Medical Imaging*, volume 5745, pages 646-655. SPIE Publications, 2005.

Molecular Imaging



Towards Quantitative SPECT: Error Estimation of SPECT OSEM with 3-D Resolution Recovery, Attenuation Correction and Scatter Correction

J. Zeintl[1], A. H. Vija[2], A. Yahil[2], X. Ding[2], E. G. Hawman[2], J. Hornegger[1], and T. Kuwert[3]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Siemens Medical Solutions USA, Inc., Molecular Imaging
[3] Clinic of Nuclear Medicine, University Erlangen-Nuremberg, Germany

Abstract. In this work we investigate errors in absolute SPECT quantitation introduced by maximum likelihood based iterative reconstruction such as ordered subset expectation maximization (OSEM-3D) with incorporated corrections for detector response, photon attenuation, and photon scattering.

We systematically investigate biases relevant to quantitative SPECT if OSEM with isotropic (3-D) depth dependent resolution recovery (OSEM-3D), attenuation and scatter correction is used. We focus on the dependencies of activity estimation errors on the projection operator, structure size, pixel size, and count density. We use 99mTc to establish a base line. We also investigate the non-stationary response of OSEM-3D and how statistical based global stopping criteria affect local bias. Four Siemens low energy collimators (LEUHR, LEHR, LEAP, LEHS) with geometric resolution between 4.4 mm and 13.1 mm at 10 cm distance and sensitivity between 100 cpm/μCi and 1020 cpm/μCi are tested with simulations and actual measurements of standard quality control phantoms. Images are reconstructed with OSEM-3D with attenuation and scatter correction. Global stopping criteria based on image features and statistical criteria are tested. Global and local contrast, resolution and count recovery are quantitatively measured for structure sizes between 4.8 mm and 31.8 mm and pixel dimensions between 0.6 mm and 4.8 mm. We also investigate noise correlation properties. Results show high fluctuations of OSEM operation points, the choice of iteration number and number of subsets, across structure sizes and collimator resolutions. In addition, inconsistencies between statistical and image feature based stopping criteria are observed. Contrast and emission recovery errors are comparable for LEUHR, LEHR, and LEAP collimators beyond 9.5 mm structure sizes (see Fig. 1). LEHS collimation in combination with OSEM-3D results in the highest errors in terms of contrast and quantitative emission recovery.

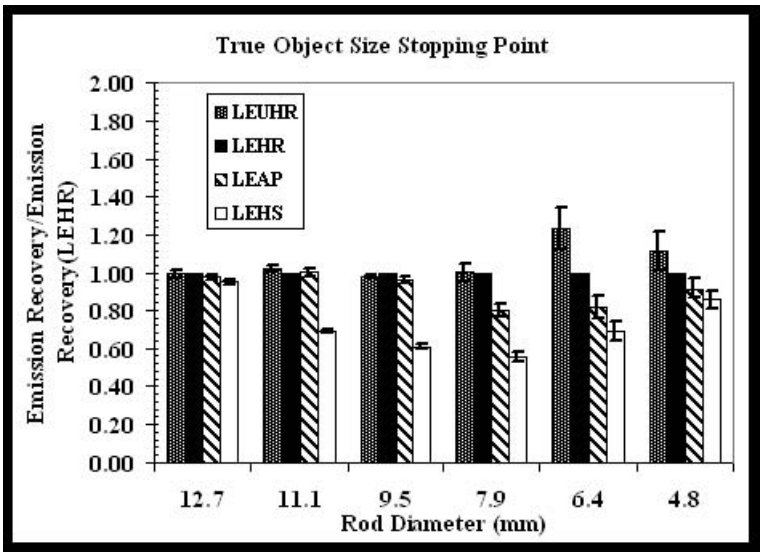


Fig. 1. Comparison of four low energy collimators in terms of emission recovery when OSEM reconstruction was stopped at the true object size values respectively. 3-D simulation of the Hot Rod Phantom with 10% background activity, 2.4mm pixel and $30x10^6$ counts. Emission Recovery values are normalized to the LEHR results. Error bars shown are standard deviations (n=4)

Methods of Correlative Imaging: PET-MR

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

[1] Clinic of Nuclear Medicine, University Hospital Erlangen, Germany
[2] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. Correlative imaging is the combination of imaging modalities which use different physical effects for generating medical images. E.g. Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Single Photon Emission Tomography (SPECT) or Ultrasonics (US). The combination and correlation of information gathered from different modalities provides gained diagnostic abilities than the standalone application of one modality.

1 PET-MR

Multi-modality image acquisition devices like SPECT-CT or PET-CT are already well established in nuclear medicine. CT images supplement the SPECT and PET images with additional anatomic information, whereas standalone emission images usually only supply functional parameters. In contrast to MR, CT suffers from weak soft tissue contrast, so differentiation of e.g. white and grey matter in brain images hardly possible. To outcome this limitation the idea of a new multimodality imaging device, PET-MR [2], was considered (see figure 1 for an PET-MR image). So far PET-MR only exists as

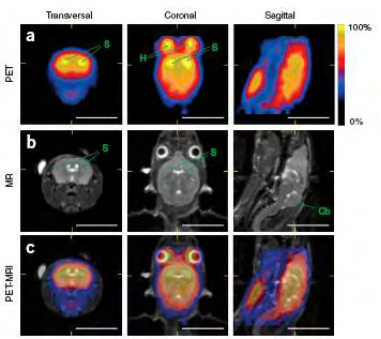


Fig. 1. Image from [2]. (a) shows the PET image of a mouse head. (b) the simultaneously aquired MR image. (c) the fusion of both images.

prototypes in facilities in Jülich and Tübingen (Germany). For PET-MR a yet unsolved

point is attenuation correction. Attenuation correction in medical imaging is necessary for quantitative and reliable diagnostics in every photon emission imaging device (SPECT, PET). So far attenuation correction is achieved with the help of transmission scans [1] [4] (e.g. CT) which is not possible in PET-MR. The transmission scan images are transformed to linear attenuation coefficients (LAC) appropriate to the photon energy (e.g. 140 keV for Tc-99m) of the radioactive pharmaceuticals. A new approach also suitable for PET-MR generates the LAC-maps out of classified (usually 3 classes), automatic segmented MR images. Segmentation techniques include machine learning [3] with the help of pre-labeled MR images, but also unlearned methods like graph-cut segmentation. Another promising procedure is the nonrigid registration of a MR image with an already existing CT from earlier examinations or with an atlas image, which is kind of an average calculated out of CT images for a certain patient collective. The atlas image itself can be converted to an LAC-map and therefore the PET aquisitions can be corrected. For the segmentation, special MR sequences, e.g. ultrashort time-to-echo (UTE) [5] sequences seem to compromise higher classification rates.

References

1. Zaidi, H. ; Hasegawa, B.: Determination of the Attenuation Map in Emission Tomography. Special Contribution to the Society of Nuclear Medicine 2002.
2. Judenhofer, M. ; Wehrl, H. ; Pichler, B.: Simultaneous PET-MRI: a new approach for functional and morphological imaging. *Nature Medicine* 2008; 14:459–465.
3. Zaidi, H.: Is MR-guided Attenuation Correction a Viable Option for Dual-Modality PET/MR Imaging? *Radiology* 2007; 244:639–642.
4. Patton, J. ; Turkington, T.G.: SPECT/CT Physical Principles and Attenuation Correction. *Journal of Nuclear Medicine Technology* 2008; 36:1–10.
5. Tyler, D. ; Robson, D. ; Henkelmann, M. ; Young, I. ; Bydder, G.: Magnetic Resonance Imaging With Ultrashort TE (UTE) PULSE Sequences: Technical Considerations. *Journal of Magnetic Resonance Imaging* 2007; 25:279–289.

4-D SPECT acquisition simulation using GATE and the NCAT phantom

H. Hofmann[1], J. Zeintl[1], T. Kuwert[2], and J. Hornegger[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Clinic of Nuclear Medicine, University Erlangen-Nuremberg, Germany

Abstract. In this work realistic SPECT simulations were conducted on a compute cluster using a computerized phantom. Furthermore, the influence of attenuation correction (AC) with different X-Ray CT acquisition schemes on the estimation of the ejection fraction (EF) was evaluated using real phantom data.

GATE, an emission tomography simulation framework, was installed on a compute cluster. It extends the underlying high energy physics simulation software, Geant4, with assets and operations that ease the modeling of emission tomography acquisitions. A sample acquisition setup is shown in Figure 1.

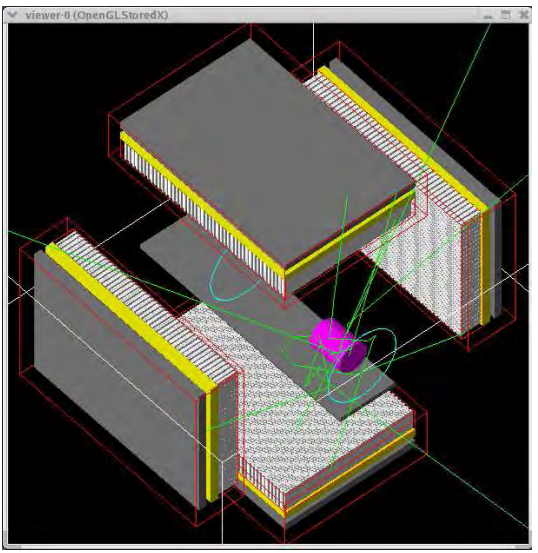


Fig. 1. Visualization of primitive simulated SPECT acquisition.

Different experiments were conducted to evaluate the correctness of the simulation results. First, collimators have been designed based upon the specification of real

Siemens Symbia™ devices [1]. The sensitivity and geometric resolution of the simulations showed good agreement with the specification. Then, simulations were executed with a computerized torso phantom (NCAT [2]). This phantom had been designed by Segars [3]. We developed several tools that were necessary to use the output volumes from NCAT in GATE.

Although utilizing 200 CPU cores on average, simulations with reasonable high activity took about six days to complete. We encountered some problems with such huge simulations. The simulation was split into 9600 jobs, which the cluster queuing system was not prepared for at first. Then, the output comprised about 50,000 files which might be too much for some file systems. And finally, there were 32 GB of output data, waiting to be processed. Nevertheless, the simulated projections and their reconstruction look realistic and promising.

In addition to the simulations, we acquired projections of the Dynamic Cardiac Phantom (DCP) from DataSpectrum with a real device (Siemens Symbia™ T6). This phantom has a pump attached to it, so it can simulate the blood flow in a beating heart. The pump can also be moved to eight fixed positions that are distributed over the heart cycle.

Nine X-Ray CT scans were taken. One with the pump of the DCP continuously moving, and one for each of the eight stop-positions. Also, a high count gated SPECT tomo with eight time slots, synced to the phantom's ECG signal, was acquired.

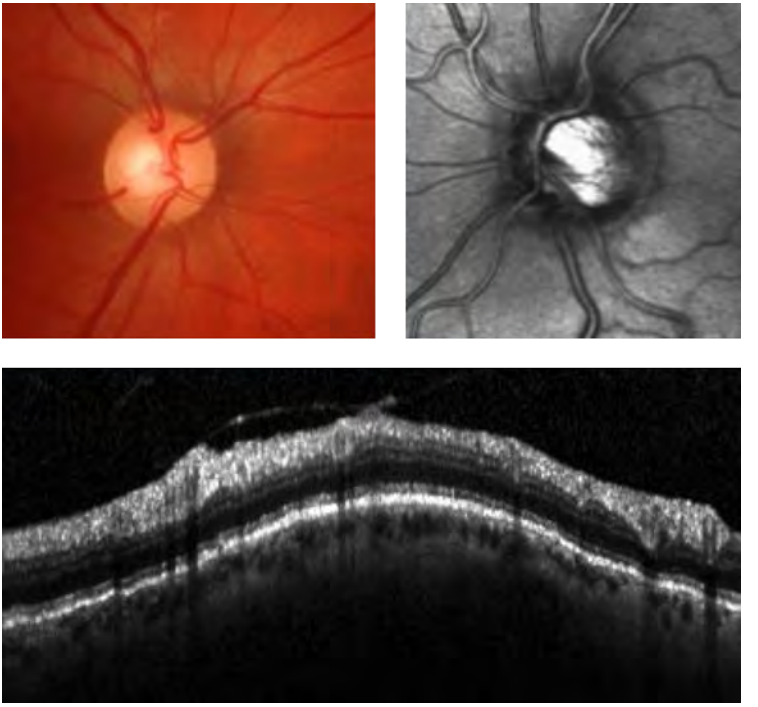
The gated SPECT tomo dataset was then split into eight static tomos representing one time slot each. These were reconstructed using the different X-Ray CT scans of the stopped phantom for AC to simulate AC with a gated X-Ray CT. The reconstructed volumes were then merged into a recon gated SPECT. As reference, this was also done using the continuous CT for AC of all time slots.

Finally, the EF of the reconstructions was measured using 4-D MSPECT and compared to the reference reconstruction. The measurements showed only insignificant difference of maximum 0.9%.

References

1. Siemens Medical Solutions USA, Inc.: Symbia™ TruePoint SPECT·CT. System Specifications, USA (2005)
2. Segars, W. P.: The 4D NCAT Phantom. Website, http://dmip.rad.jhmi.edu/people/faculty/Paul/Segars_research.htm#NCAT, accessed Aug 2008
3. Segars, W. P.: Development and Application of the new Dynamic NURBS-Based Cardiac-Torso (NCAT) Phantom. Ph.D. Dissertation, University of North Carolina at Chapel Hill, NC (2001)

Computer automated Detection of Eye Diseases



Glaucoma Risk Index: A Comparison to State of the Art Techniques

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

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] International Max Planck Research School Optics and Imaging
[3] Department of Ophthalmology, University Erlangen-Nuremberg, Germany

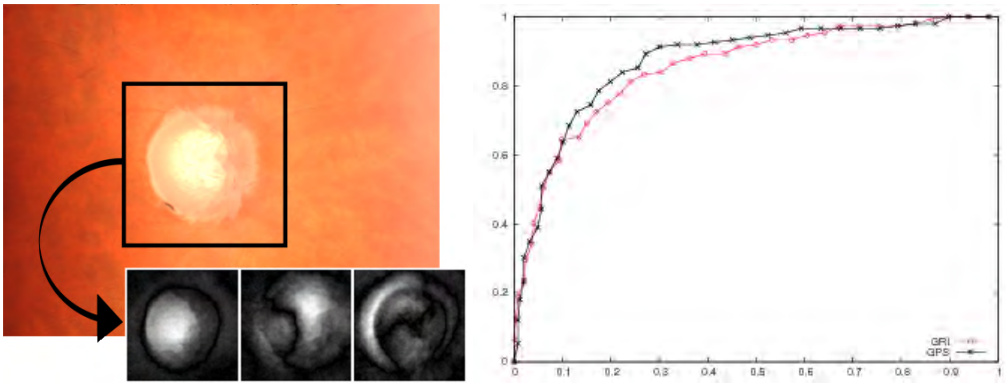
Abstract. The Glaucoma Risk Index (GRI) from fundus photographs provides a competitive performance of 78% accuracy in glaucoma detection. Compared to other methods the GRI derives the information from a cheap and easy to acquire image modality.

The presented Glaucoma Risk Index (GRI) classifies color fundus photos of the eye to detect the glaucoma disease. The approach reaches a classification accuracy of 78% in a two class problem (healthy vs. glaucomatous eyes) taking a gold standard diagnosis of ophthalmologists as a basis.

In contrast to previous classification methods of retina images we do not use segmentation measurements as features. Instead we use preprocessed and normalized fundus photos [1] and apply a pure data driven appearance based approach for feature calculation [2] (Fig. 1a). The 2-stage classification scheme helps to combine classifiers of different image inputs. The evaluation shows that the second stage improves the separability of clear cases and it makes the final decision more robust.

The area under the ROC of the system is 0.85 with a sensitivity of 78% in detecting the glaucoma samples (specificity 76%). The results in our evaluation with 575 images show comparable performance to a commercial Glaucoma Probability Score (GPS) of the HRT III Lasertomograph on the same patients (Fig. 1b). It can be concluded from the sample comparison between the GRI and the GPS that both indices capture different information. It is promising to combine both modalities or even both indices to a common glaucoma marker in the future.

Our method shows that the cheapest, easiest and most widespread retina image modality can be used to automatically estimate a glaucoma risk. This GRI could be a helpful indicator for experts in the clinics as well as for ophthalmologists in medical practice.



(a) (b)

Fig. 1. (a) Glaucoma Risk Index (GRI): Processed fundus image (vessels removed) and model generation by PCA (first three eigenimages of a training set with 180 samples). (b) ROC of GRI compared to established GPS index from HRT images.

References

1. Meier, J., Bock, R., Michelson, G., Nyúl, L.G., Hornegger, J.: Effects of Preprocessing Eye Fundus Images on Appearance Based Glaucoma Classification. In: 12th International Conference on Computer Analysis of Images and Patterns, CAIP. Lecture Notes in Computer Science 4673, Berlin (2007) 165–173
2. Bock, R., Meier, J., Michelson, G., Nyúl, L.G., Hornegger, J.: Classifying Glaucoma with Image-Based Features from Fundus Photographs. In: 9th Annual Symposium of the German Association for Pattern Recognition, DAGM. Lecture Notes in Computer Science 4713, Berlin (2007) 355–365

Glaucoma Risk Index: Towards Integration of Knowledge from Temporal Changes

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

[1] Chair of Pattern Recognition (LME)
[2] School of Advanced Optical Technologies (SAOT)
[3] Department of Ophthalmology
University of Erlangen-Nuremberg, Germany

Abstract. Temporal changes of papilla region are important in diagnosing glaucoma. The incorporation of these changes to image based glaucoma risk index might improve its performance.

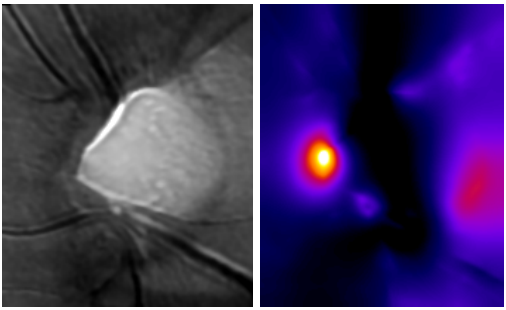
Glaucoma is a chronic eye disease that is characterized by a successive and irreversible loss of optic nerve fibers. As early detection is essential to prevent glaucoma, our developed automated glaucoma risk index (GRI) [1] is very useful for screening applications. The fundus based GRI linearly models structural changes among different patients. But, the temporal development of glaucoma remains unconsidered and should be included from longitudinal retinal image data.

We use topographic images of HRT from a longitudinal study of several patients that were tracked over ten to 15 years. The acquired images are compared to one baseline examination and capture deformations caused by (i) image acquisition and (ii) structural glaucomatous changes. The undesired variations due to acquisition are compensated using a normalization of the retinal surface by a parabolic surface fitting and a non-rigid registration with a high stiffness value. The remaining glaucomatous changes were detected by the magnitude of non-rigid deformations as well as mean and Gaussian curvatures. The latter non-rigid registration was performed with low stiffness to cover small structural changes. These different methods capture similar changing regions while non-rigid registration is most specific (Fig. 1). As there is no ground truth available it is still not clear how to differentiate between structural and acquisition caused variations.

The further usage of the topographic shape information to compensate acquisition based variations seems to be an adequate method to incorporate the pure glaucomatous changes to the glaucoma risk index.

References

1. Bock, R., Meier, J., Michelson, G., Nyúl, L.G., Hornegger, J.: Classifying Glaucoma with Image-Based Features from Fundus Photographs. In Hamprecht, F.A., Schnörr, C., Jähne, B., eds.: 9th Annual Symposium of the German Association for Pattern Recognition, DAGM. Lecture Notes in Computer Science (LNCS) 4713. Volume 4713/2007., Berlin (2007) 355–365



(a) (b)

Fig. 1. (a) Topography image of Heidelberg Retina Tomograph (HRT) used as baseline examination (bright regions denote deeper areas). (b) Glaucomatous changes over a period of seven years detected by non-rigid registration (strong deformations are marked by bright colors).

Analysis of the Visual System in Glaucoma Patients Using Diffusion Tensor Imaging

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

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Department of Opthalmology, University Erlangen-Nuremberg, Germany,
[3] Department of Neuroradiology, University Erlangen-Nuremberg, Germany,
[4] Siemens AG.

Abstract. In this work, a correlation between the glaucomatous signs of optic nerve and the optic radiation atrophy is being investigated. Diffusion tensor imaging is used to quantitatively segment and analyze the human visual system.

Millions of people in the world suffer from glaucoma, the second leading cause of vision loss in the world. Fewer than 50% of people with glaucoma are aware of their disease. The blindness caused by glaucoma is irreversible. Therefore, improved methods of screening and therapy for glaucoma are urgently needed.

Different imaging modalities exist for acquiring images of the eye. Most of these modalities focus on imaging the retina, but the human visual system does not only consist of the eye but it is extended into the brain through the optic nerve till the visual cortex.

Based on primary investigations, a correlation was found between the glaucomatous signs of optic nerve and the optic radiation atrophy, but the obtained results were limited by the qualitative analysis of the optic radiation.

Diffusion tensor imaging is based on diffusion weighted magnetic resonance imaging [1][2], where water diffusion is used to weight the magnetic resonance images. Therefore, diffusion weighted images can be used to detect the microscopic structure of the brain. The diffusion tensor imaging modality is the only modality that allows tracking the nerve fibers in the brain non-invasively.

The aim of this work is the volumetric quantification of the visual system in general and the optic radiation in particular by segmenting it using diffusion tensor imaging. A quantitative analysis of the segmented visual system will be performed and compared to an analysis of glaucoma using eye imaging modalities to investigate the correlation between the optic radiation atrophy and the glaucomatous signs of the optic nerve quantitatively.

The results of this research could illustrate the relation between the damage done in the various parts of the visual system due to glaucoma, and provide a new technique for glaucoma screening using diffusion tensor imaging.

References

1. S. Mori and P.B. Barker: Diffusion Magnetic Resonance Imaging: Its Principles and Applications. The Anatomical Record (new anat.), 257:102-109, 1999.
2. D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat: Diffusion tensor imaging: concepts and applications. Journal of Magnetic Resonance Imaging, 13:534-546, 2001.

Fast and Robust Nerve Fiber Layer Segmentation on OCT Scans

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

[1] Chair of Pattern Recognition (LME),
[2] School of Advanced Optical Technologies (SAOT),
[3] Department of Ophthalmology, University Erlangen-Nuremberg, Germany

Abstract. A nerve fiber layer segmentation method on retinal OCT scans is developed to support ophthalmologists in diagnosing the glaucoma disease.

Over the last few years optical coherence (OCT) tomography [1] has been introduced to clinical workflows in ophthalmology. OCT systems deliver depth scans of the retinal background. Today the modality is capable of providing images of sufficient resolution for a high diagnostic value to the physician [2]. Unlike other medical imaging fields like radiology, supporting image processing algorithms are just getting into the focus of the ophthalmic imaging research community.

One of the most severe eye diseases is glaucoma. It is the second leading cause of blindness in the world. OCT imaging allows a direct visualisation of the nerve fiber layer (NFL). The thickness of the NFL is an important indicator for glaucoma, because glaucoma can be described as the loss of nerve fibers and supporting tissue in the eye. The goal of this work is to develop a new NFL segmentation algorithm to ease the interpretation of OCT images for the ophtalmologists.

Our currently proposed method for segmenting the NFL on circular B-Scans around the papilla is based on the following principles: The outer nuclear layer (ONL) is easily detected on the images, so the search space can be restricted depending on the sought-after layer. The upper NFL boundary (UNFL) and the retinal pigment epithelium (RPE) can be found by extrema detection in the gradient image along the A-Scans. The RPE is used as an even line for geometry correction. This geometry correction leads to a better comparability between subjects and examinations. The lower boundary of the NFL (LNFL) is detected by a combination of feature based approach as well as a gradient based technique. Taking the highest intensity drop-of along the A-Scan between the UNFL and the ONL delivers exact results in blood vessel regions. In all other parts of the image features as for example the mean neighbouring intensity value are calculated for every extremum of the gradient image in the designated region. These features are then clustered by fuzzy c-Means clustering and thus splitted into a NFL-class and remainders. The most distant members from the UNFL in the NFL-class define the LNFL. The algorithm can be applied to both normal and glaucoma patients. An example result of a glaucomatous eye is shown in Fig. 1.

This work will contribute to an extended unterstanding of the development and progression of the glaucoma disease by being part of an automated detection system.

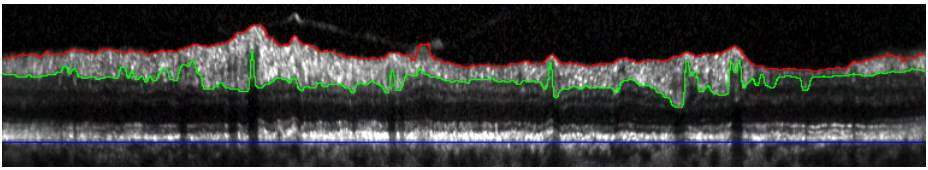


Fig. 1. Result of the segmentation algorithm: Geometry corrected OCT image with segmented NFL. Upper nerve fiber layer (red), lower nerve fiber layer (green) and lower retinal pigment epithelium boundary (blue).

References

1. D. Huang, E.A. Swanson et al.: Optical Coherence Tomography. Science, Nov. 1991: Vol.254(5035), 1178-81, 1991
2. M. Wojtkowski, R. Leitgeb et al.: In vivo human retinal imaging by Fourier domain optical coherence tomography. Journal of Biomedical Optics, Vol 7(3), 457-463, July 2002

Non-rigid Registration for Interpolation of Defect Slices in Sequences of Microscopic Photographs

Simone Gaffling[1], J. Hornegger[1], and E. Luetjen-Drecoll[2]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Institute of Anatomy II, University Erlangen-Nuremberg, Germany

Abstract. The work implements and compares four different non-rigid registration methods, which are used for interpolation of defect slices of a sequence of histological images, with the aim to be able to reconstruct the 3D structure of the tissue of interest.

1 Content

Glaucoma disease is one of the most common eye diseases. One of the most important risk factors is elevated intraocular pressure. This pressure can be reduced by applying the drug Bimatoprost®. Side effects of this treatment are elongation and thickening of eye lashes. To investigate the pathophysiology of these side effects, mice were treated with Bimatoprost, and the morphological changes were analyzed in 1 μm thick serial sections through the lids. The purpose of this study was the 3-D reconstruction of the lashes and the related hair follicles.

Unfortunately, the creation process of the histological slices introduces errors to the final images, since mechanical stress is implied on the tissue samples, making them unusable for reconstruction purposes. Fig. 1 shows an example of such an image.

MicroSlice Visualization is a software tool that was developed at the Chair of Pattern Recognition, University Erlangen-Nuremberg, to deal with the problems associated with this process, and to finally reconstruct the 3D structure. Consecutive steps include the identification of corrupt slices, rigid registration of neighboring slices to match their orientation, the interpolation of defect slices, an optional segmentation of structures of interest and finally the generation of the volume and its rendering.

The basic idea for the interpolation of images is to use the adjacent slices. First, a deformation field is calculated, that gives for each pixel in one image the translation necessary to transform it into the other neighboring image. The interpolated image or image sequence is created by partial application of the deformation field on the preceding image.

Four methods representing different approaches to non-rigid image registration were

implemented and compared. Three of them are standard techniques available in the Insight Toolkit ITK, namely B-Spline registration, which deforms the image by translating the control points of a B-Spline grid, Demons registration, which computes deforming forces out of the intensities, and Curvature registration, a variational formulation of the problem which restricts the degree of deformation using a regularizer. The fourth approach was implemented at the Chair of Pattern Recognition and is similar to the curvature registration, but differs in implementation details, especially the optimization strategy. The interpolation results show a smooth transition from the template image to the reference image, which was found satisfactory by experts in this field.

Further work will primarily extend the preprocessing of the images, with main focus on the correction of artifacts stemming from the creation of the slices, like blurred regions and tissue warping, but will also include color normalization and elimination of post-mortal artifacts.

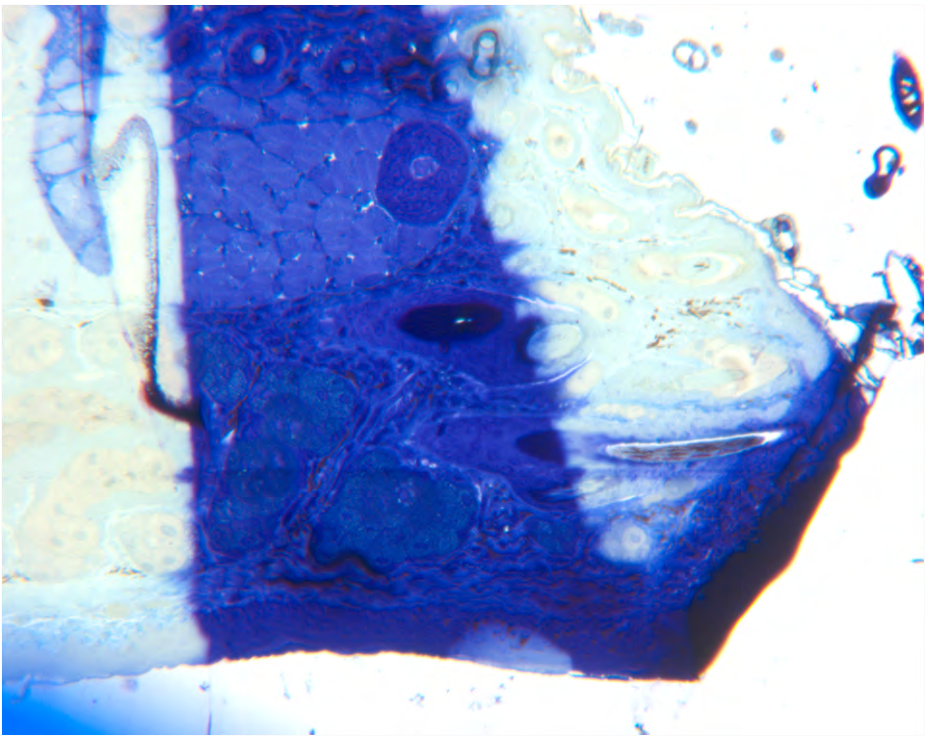
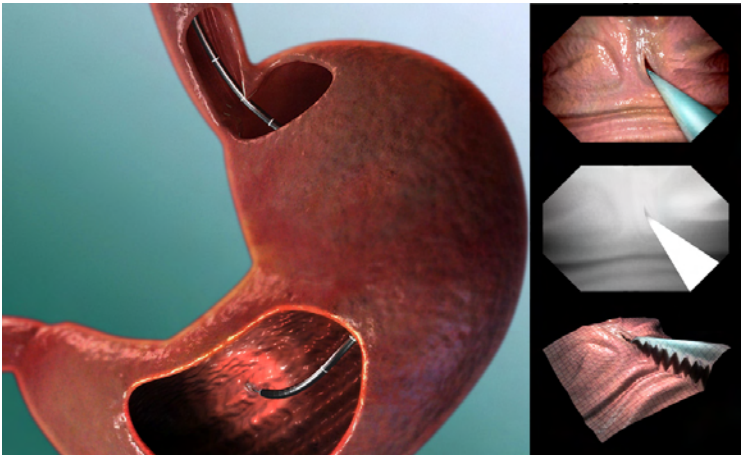


Fig. 1. Sample of an image which is unsuitable for reconstrution.

Medical TOF-Applications & Endoscopy



Enhanced Endoscopic Engineering

K. Höller[1], A. Schneider[2], J. Gutierrez[3], T. Wittenberg[3], J. Hornegger[1], and H. Feussner[2]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Workgroup for Minimal Invasive Surgery MITI, Klinikum r.d.Isar, Technical University, Munich, Germany,
[3] Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany

Abstract. With a tiny MEMS tri-axial inertial sensor, image orientation correction in flexible endoscopic surgery (NOTES) can be realized by measuring gravity.

One of the biggest problems with flexible endoscopy in Natural Orifice Translumenal Endoscopic Surgery (NOTES) is the missing information about image orientation [1]. Thus, tip retroflexion of a non-rigid endoscope causes image rotation angles till 180 degree. One possibility to measure this orientation is to integrate a Micro Electro-Mechanical System inertial sensor device in the endoscope tip, which measures influencing forces in three orthogonal directions (fig. 1). If the endoscope is not moving, only the acceleration of gravity has an effect on the three axes.

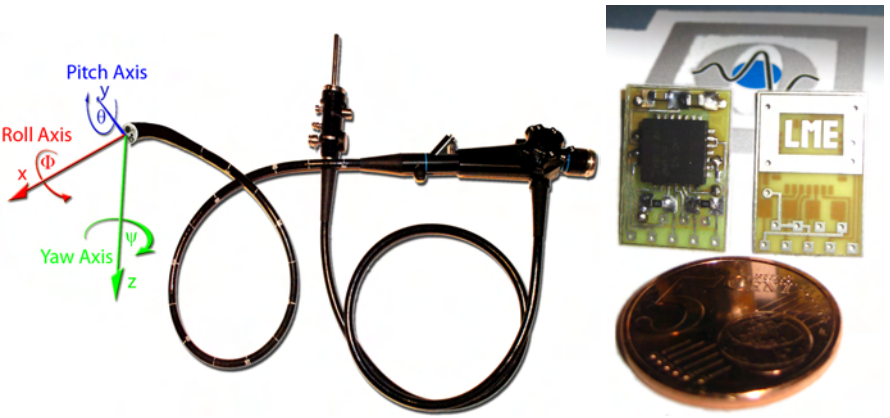


Fig. 1. Orientation of a flexible endoscope and implemented inertial measurement unit

To describe the orientation of the endoscope in relation to the gravity, an Cartesian "endoscope board navigation system" (according to the DIN 9300 aeronautical standard) with axes $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ is used [2]. The tip points in x-direction which is the boresight, image bottom is in z-direction and the y-axis is orthogonal to boths in horizontal image direction to the right. Rotations about these axes are called roll Φ (around x), pitch Θ (around y) and yaw Ψ (about z). Gravity $\boldsymbol{g}$ is considered as an external independent vector.

To correct the image orientation only the affect of gravity on each axis can be used, there are no explicit angle informations. Equation 1 expresses, how rotation parameters Φ, Θ and Ψ of the IMU (Inertial Measurement Unit) have to be chosen to get back to a spatial orientation with $\boldsymbol{z} \parallel \boldsymbol{g}$:

$$\begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\Phi) & sin(\Phi) \\ 0 & -sin(\Phi) & cos(\Phi) \end{bmatrix} \cdot \begin{bmatrix} cos(\Theta) & 0 & -sin(\Theta) \\ 0 & 1 & 0 \\ sin(\Theta) & 0 & cos(\Theta) \end{bmatrix} \cdot \\ \cdot \begin{bmatrix} cos(\Psi) & sin(\Psi) & 0 \\ -sin(\Psi) & cos(\Psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} = \begin{bmatrix} -sin(\Theta) \\ sin(\Phi)cos(\Theta) \\ cos(\Phi)cos(\Theta) \end{bmatrix} \cdot |\boldsymbol{g}| \tag{1}$$

with Φ: roll, Θ: pitch, Ψ: yaw
and $F_{x,y,z}$: measured acceleration, g: gravity

Using the two-argument function atan2 to handle the ambiguity
one finally can compute roll Φ for $F_x \neq \pm g$ and pitch Θ for all values:

$$\Phi = atan2(F_y, F_z) \tag{2}$$

$$\Theta = \arcsin\left(\frac{-F_x}{g}\right) \tag{3}$$

Since $\boldsymbol{z} \parallel \boldsymbol{g}$ is the only proposition, yaw Ψ around z cannot be computed with this approach. If $F_x = \pm g$ ($\rightarrow F_y = F_z = 0$), roll Φ is not determinable either. As for rotation correction only results for roll Φ are needed, these problems do not need to be discussed here. To avoid movement influence, orientation correction is only applied if there is no other acceleration and gravity g is the only affecting force.

$$F_x^2 + F_y^2 + F_z^2 = g^2 \tag{4}$$

Accelerometer Errors consist of several effects: Constant deterministic offset, drift (= variable random offset), scaling error (= incorrect factor), misalignment (= imprecisely mounted), nonlinearity and quantization error [3]. Misalignment, scaling error and constant offset can be analyzed and permanent corrected with a calibration matrix [4]:

$$\begin{bmatrix} F_x \\ F_y \\ F_z \\ 1 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} & T_x \\ M_{21} & M_{22} & M_{23} & T_y \\ M_{31} & M_{32} & M_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} F_{x0} \\ F_{y0} \\ F_{z0} \\ 1 \end{bmatrix} \tag{5}$$

A first prototype of a 12mm by 18mm circuit board with a 3-axis MEMS chip for acceleration measurement, SMD capacitors for power supply HF denoising and a two-wire I^2C communication interface was already build with this approach (fig. 1).

References

1. K. Höller, M. Petrunina, J. Penne, A. Schneider, D. Wilhelm, H. Feußner and J. Hornegger: Taking endoscopy to a higher dimension: Computer Aided 3-D NOTES. In: Proc. 4th Russian-Bavarian Con-ference on Bio-Medical Engineering, RBC-BME 2008, MIET, Zelenograd, Moscow, Russia, 08.-09.07.2008
2. D. H. Titterton and J. L. Weston: Strapdown Inertial Navigation Technology, 2nd Edition, IEE Radar, Sonar, Navigation and Avionics Series 17, 2004
3. R. Dorobantu: Simulation des Verhaltens einer low-cost Strapdown IMU unter Laborbedingungen; Schriftenreihe des IAPG, 1999
4. K. Höller: Charakterisierung und Modellierung eines Inertialsensors zur Navigation autonomer Systeme; Diploma thesis, LIKE, Friedrich-Alexander-University Erlangen-Nuremberg, Germany 2005

Amplitude Value Normalization for Time-of-Flight Cameras

J. Penne[1], M. Stürmer[1], and J. Hornegger[1]

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. An approach for the normalization of amplitude values acquired by Time-of-Flight (ToF) cameras [1] is proposed which takes into account integration-time dependency, geometric attenuation and inhomogeneous active scene illumination.

It can be observed that the illumination of a scene by a ToF camera is inhomogeneous: Towards the border of the field of view, the illumination gets darker.
But this is not the only effect: As the actively emitted light gets attenuated with respect to the traveled distance (according to the inverse-square-law and the general attenuation of electromagnetic waves) the amplitude values for the same object at different distances differ. As the distance of an observed point is provided by the ToF camera, these effects can also be circumvented.
Last but not least, the integration time, which controls how long photons are collected in each pixel, leads to changing amplitude values even if the same object is observed at the same distance.
In the following sections it will be outlined how the instability of the amplitude data with respect to the radiation characteristics of the illumination unit, the distance of the observed objects and the integration time can be reduced. The result will be normalized amplitude values only depending on the material of the observed objects and the specularity of the reflection at the object yielding data which is more beneficial for segmentation or tracking tasks.

Fig. 1. Original (left) and normalized (right) amplitude data in comparison.

References

1. Lange, R.: 3-D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology. Ph.D. Dissertation, University of Siegen, Germany (2000) 26–29

Patient positioning using Time-of-Flight Cameras

C. Schaller[1], A. Adelt[1], J. Penne[1], and J. Hornegger[1]

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. An approach to use Time-of-Flight cameras for patient positioning in a radiotherapy workflow.

1 Method

Patient positioning is a crucial issue in the field of radiotherapy [2]. In a common workflow, a planning CT scan of the patient is acquired in order to plan the treatment process. Right before the treatment session starts, the patient has to be positioned accurately to ensure the treatment plan can be applied correctly. Recently, a new emerging imaging technology called Time-of-Flight (ToF) is available. This new technology uses a single sensor which is feasible to acquire 3-D data in real-time. We present a method to correct the position of a patient by acquiring body surfaces using a ToF sensor. The ToF sensor illuminates the patient's body with an optical reference signal and provides a 3-D point set of the observed scene. Given two 3-D point sets of the patient's body the patient's misalignment can be identified and corrected. The surface matching process is divided into a segmentation and registration process. Before the two point sets are registered the patient's body has to be separated from the background. After this preprocessing step an iterative closest point (ICP) algorithm [1] is applied to the datasets in order to determine translation and rotation parameters to correct the position of the patient. Current results show that for rigid phantoms it is possible to obtain a mean accuracy of 2.88 mm.

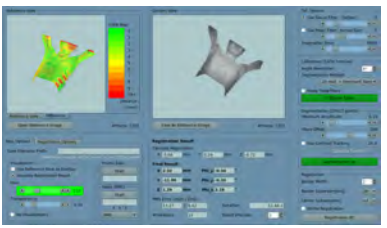


Fig. 1. Patient Positioning GUI

References

1. D. Chetverikov and D. Svirko and D. Stepanov. The trimmed iterative closest point algorithm. In International Conference on Pattern Recognition 2002, Quebec, August, pages 545–548
2. P. J. Schoeffel, W. Harms, G. Sroka-Perez, W. Schlegel, and C. P. Karger. Accuracy of a commercial optical 3d surface imaging system for realignment of patients for radiotherapy of the thorax. Phys.Med.Biol., Volume 52, Number 13, pages 3949–3963, 2007.

Medical TOF-Applications & Gesture Recognition

Survey about Gesture Recognition with a Time-Of-Flight Camera

E. N. K. Kollorz[1], J. Penne[1], J. Hornegger[1], and A. Barke[2]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Audi Electronics Venture GmbH, Gaimersheim, Germany

Abstract. Gesture recognition is performed with a time-of-flight camera in the automotive environment for regulating control elements. Goal of this survey "Operational concept next generation" is to measure the accuracy which can be achieved when a person points onto specific buttons on control elements. Dependencies between camera position, locations and styles of buttons on control elements were analyzed as well as the different user behaviors.

Gesture recognition is one possibility which can be applied to control buttons on control elements besides speech recognition. We use a time-of-flight camera for gesture recognition because in the acquired data the hand can be separated easily from the center console. The measured distance values of the time-of-flight camera can be used to distinguish between certain gestures and to compute the distance to the center console.
Certain aspects of earlier work were included into this study [1]. To distinguish between gestures a scan line is used. The scan line is calculated from the minimum and the center of gravity of the trailed user hand. Due to reflections of some shiny parts of the control elements, several pixels have to be masked out manually. The distance values of the time-of-flight camera are preprocessed with a median filter because of the large fluctuating distance range returned by the camera. A 3D plane is fitted through the preprocessed 3D points of the control elements to determine the distance of the finger tip to the control element.

For evaluation purposes 15 persons performed the survey. Unfortunately, the distances between the buttons on the control elements and the finger tip were unreliable. One major result is that the different coordinate systems (human eye, camera) as well as the size of the buttons of the control elements play an important role for the performance and accuracy of the gesture recognition system.

References

1. Kollorz, Eva Nicole Karin; Barke, Alexander; Hornegger, Joachim: Gesture recognition with a time-of-flight camera. Dynamic 3D Imaging (Workshop in Conjunction with DAGM'07), Heidelberg, Germany, (2007) 86–93

Application of Time-Of-Flight Technology in Tele-Physiotherapy

A. Mahmoud[1], A. Maier[2], J. Penne[2], and M. Stürmer[2]

[1] National School of Engineers, Tunis (ENIT), Tunisia
[2] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,

This proposal concerns a solution for regular physiotherapy sessions which are compulsory for many patients having chronic joint problems and not having enough mobility to go to the clinic. Since the expenses of physiotherapy sessions at home are quite high and since this kind of sessions only require observation and directions from the therapist, it is possible to carry it out online. In this paper, propose a novel system for Tele-physiotherapy based on the use of Time-Of-Flight (ToF) Camera. The ToF Camera acquires 3-D surface data by measuring the distance of equidistant points forming a pixel matrix. The Camera that we use (Mesa Swiss Ranger SR-3000) has a resolution of 176x144 pixels [1]. The distance acquired is coded in a 64-bit double making the size of each frame of raw data 198Kbyte (if we send only distances and we compute the 3-D coordinates in the client side). Hence, a solution for compressing data has to be developed to let the transmission of these data be possible in real time. After several studies of some possibilities, we agreed on compressing the raw data into an MPEG video which can be streamed in real time to the client using the ffmpeg API [2].

The Y value (of the YUV components of the MPEG stream) contains the down sampled distance to one byte: the minimum distance (0.5m) corresponds to 0 while the maximum distance (2.0m) corresponds to 255. On the client side, the visualization tool, based on the 'Sammon Viewer' [3], computes the 3-D coordinates and displays the points in a convivial way as shown in fig. 1.

The final user (the doctor) is able to interact with the tool by moving the 3-D video to observe the patient from different angles while he is performing the required movements.

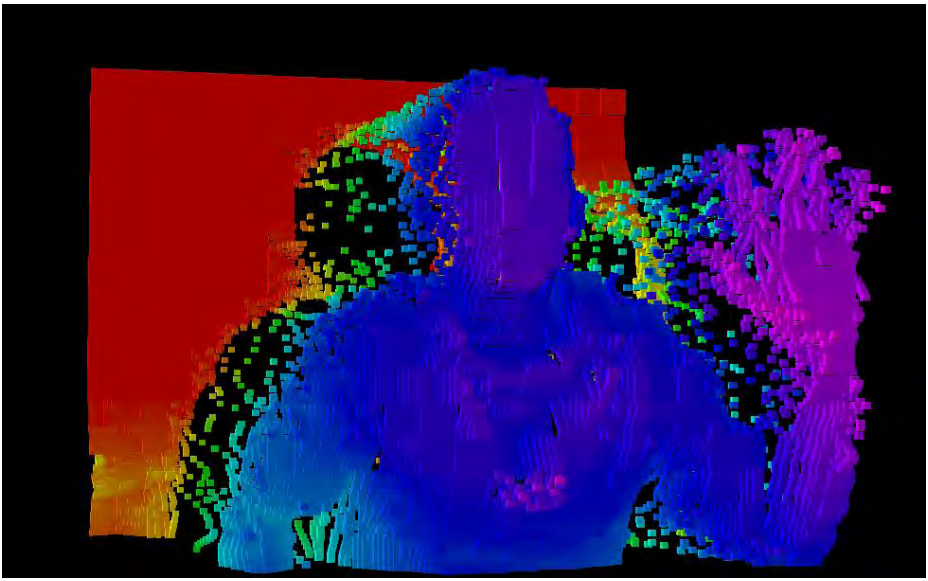


Fig. 1. Visualization tool of the 3-D surface data on the client side

References

1. T. Oggier, B. Büttgen, F. Lustenberger, G. Becker, B. Rüegg, A. Hodac. SwissRanger SR3000 and First Experiences based on Miniaturized 3D-TOF Cameras. RIM Days, ETH Zürich, Switzerland, (2005)
2. M. Stürmer, A. Maier, J. Penne, S. Soutschek, C. Schaller, R. Handschu, M. Scibor, E. Nöth. 3-D Tele-Medical Speech Therapy using Time-of-Flight Technology. European Biomedical Engineering Congress 2008, to appear.
3. A. Maier, J. Exner, S. Steidl, A. Batliner, T. Haderlein, E. Nöth. An Extension to the Sammon Mapping for the Robust Visualization of Speaker Dependencies.In: P. Sojka et al. (Eds.): TSD 2008, LNAI 5246, pp. 381–388, 2008

Time-of-Flight Distance Streaming using MPEG Compression

M. Stürmer[1], A. Maier[1], J. Penne[1], and J. Hornegger[1]

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. Video-streaming and -conferencing is widely used nowadays. As conventional cameras acquire simple 2-D images, only this information is send. ToF cameras push the dimension of the data to a higher level, so a intuitive 3-D visualization of ToF data which is streamed over the network is desirable. We propose a method for broadcasting distance data acquired by Time-of-Flight (ToF) cameras over ethernet by use of MPEG compression.

For streaming of the distance data from the ToF camera MPEG compression is used, meaning that the distances have to be scaled down to a 8bit representation. As the original data is provided as float, a quantization error is introduced which depends on the desired working range of the camera. By clipping the distance values below $500mm$ and above a maximum value r between $850mm$ and $1400mm$ a quantization error is introduced. Another error is added by the compression algorithm of the MPEG encoder, which is in average about five times larger than the quantization error. Nevertheless the quantization itself destroys details in the distance data when applying large r on structures with a high distance resolution (see fig. 1). The mean error of the reconstructed distances is between $1.6mm$ for $r = 850mm$ and $3.2mm$ for $r = 1400mm$. Standard deviations reach from $1.4mm$ to $3.1mm$ respectively.

On the receiver side the cartesian x, y, z-coordinates are reconstructed from the distances using the intrinsic parameters of the camera. The given point cloud is then visualized by a 3-D viewer which allows arbitrary views at the transmitted structure (see fig. 1).

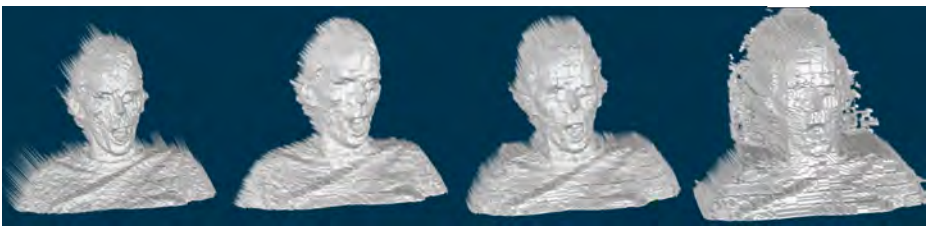
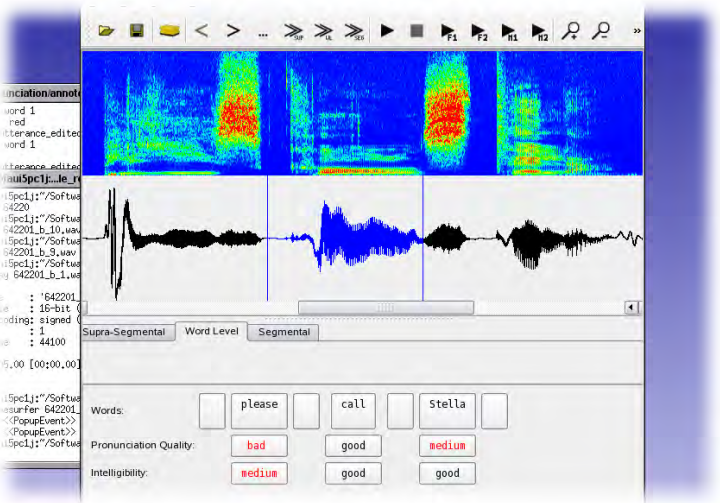


Fig. 1. Original and reconstructed distance data. From left to right: original data, $r = 85cm$, $r = 100cm$, $r = 140cm$.

Speech Evaluation



Influence of Reading Errors on the Text-Based Evaluation of Intelligibility of Pathologic Voices

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

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Dept. of Phoniatrics and Pedaudiology, University Erlangen-Nuremberg, Germany

Abstract. The influence of reading errors on the human-machine correlation of intelligibility evaluation based on a read-out text was examined. It was found not to be significant.

In speech therapy and rehabilitation, a patient's voice has to be evaluated. For this study, funded by the German Cancer Aid (Deutsche Krebshilfe; #107873), the test persons read a standard text which was processed by automatic speech recognition and prosodic analysis. We examined the influence of reading errors on the results of automatic evaluation and the correlation to human evaluation.

The test files were recorded from 85 patients suffering from cancer in different regions of the larynx. Each person read the text "Der Nordwind und die Sonne", a phonetically balanced text with 108 words (71 disjunctive). As a reference for automatic evaluation, five experienced phoniatricians and speech scientists evaluated each speaker's intelligibility according to a 5-point scale.

Due to reading errors, repetitions and remarks like "I don't have my glasses with me.", many samples contained out-of-text phenomena. In order to determine the influence of these phenomena on the evaluation results, a second version of the data set was created by removing additional words where possible.

The correlation between the human evaluation results and the average word accuracy (WA), the word recognition rate (WR) and the prosodic features, respectively, did not significantly change when the reading errors were removed from the audio files (see Table 1 and 2). Hence, the reading errors in text recordings do not have to be eliminated before automatic evaluation.

Table 1. Recognition results for the speech corpora with and without reading errors and correlation between automatic measure and human rating *(rightmost column)*

	measure	avg.	st. dev.	min.	max.	correl.
with errors	WA	48.0	17.2	3.4	81.3	–0.61
without errors	WA	49.3	17.0	10.1	81.3	–0.61
with errors	WR	53.2	15.3	9.1	82.2	–0.56
without errors	WR	54.1	15.4	9.1	82.2	–0.55

Table 2. Correlation r between selected prosodic features and human intelligibility ratings; presented are criteria with a correlation of $|r| \geq 0.5$

feature	correlation	
	with errors	without errors
ratio of duration of unvoiced segments and file length	+0.51	+0.53
duration of silent pause after current word	+0.54	+0.53
normalized energy of word-pause-word interval	+0.62	+0.59
normalized duration of word-pause-word interval	+0.63	+0.60

Automatic Pronunciation Assessment for Computer-Assisted Language Learning

F. Hönig[1], A. Batliner[1], A. Maier[1,2], K. Weilhammer[3], and E. Nöth[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Division of Phoniatrics and Pedaudiology, University Hospital Erlangen, Germany
[3] digital publishing AG, München

Abstract. The new research project C-AuDiT (Computer-assisted pronunciation and dialog training) is presented. Account is given on data collection, annotation and steps toward automatic pronunciation scoring.

Pronunciation training is vital for foreign language learning. In class however, time is usually short for the individual learner, and for computer-assisted language learning usually no or only crude pronunciation ratings are provided. C-AuDiT, a research project funded by the German Federal Ministry for Education and Research, brings together the expertise of digital publishing, one of the world leaders in language learning software, and the Chair of Pattern Recognition in order to improve current approaches to pronunciation scoring. The focus is on giving precise feedback where the error occured and on including prosodic scoring such as word stress. Another topic in C-AuDiT is natural dialog training.

For data collection, a comprehensive list of utterances has been compiled that contains two phonetically richt standard texts, shorter and longer sentences from different domains, tongue twisters and covers stress shifts in similar words (China/Chinese) or in noun/verb pairs (the project/to project). Per speaker, the material amounts to 30 minutes of speech. In total, 50 speakers are planned.

An annotation protocol has been defined that allows annotation of word insertions/deletions/substitutions, missing/superfluous word stresses and pronunciation quality and intelligibility (on a scale between 1 and 3) on word level. For prosodic phenomena, the general prosodic quality of an utterance is rated (1-3), the rhythmic structure (1-3), pitch contour (1-3) and whether the sentence mood is realized correctly. Furthermore, missing/superfluous phrase boundaries and phrase accents are annotated. To support the labellers, a GUI has been created that offers signal visualization and a framework to carry out the annotation.

Based on the prosody and pronfex [1] modules of the chair, an approach for word stress recognition has been developed.

References

1. Hacker, Christian; Maier, Andreas; Heßler, Andre; Guthunz, Ute; Nöth, Elmar: Caller: Computer Assisted Language Learning from Erlangen - Pronunciation Training and More. ICL 2007 – Proc. Int. Conf. Interactive Computer Aided Learning, September 26-29, 2007, Villach, Austria.

Automatic Detection of Speech Disorders in Children with Cleft Lip and Palate

A. Maier[1,2], F. Hönig[1], T. Haderlein[1,2], M. Schuster[2], and E. Nöth[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Division of Phoniatrics and Pedaudiology, University Hospital Erlangen, Germany

Abstract. Speech of children with cleft lip and palate (CLP) shows special characteristics such as hypernasality, backing, and weakening of plosives. In total five criteria were subjectively assessed by an experienced speech expert on the phone level and compared to an automatic evaluation system.

In clinical practice, articulation disorders are mainly evaluated by subjective tools. The simplest method is the auditive perception, mostly performed by a speech therapist. Previous studies have shown that experience is an important factor that influences the subjective estimation of speech disorders which leads to inaccurate evaluation by persons with only few years of experience as speech therapist. Therefore, an automatic and objective evaluation of articulation disorders is necessary.

In this paper, we present a new technical procedure for the measurement and evaluation of specific speech disorders and compare the results obtained with subjective ratings of an experienced speech therapist. This subjective evaluation was used as a gold standard to train a classification system. The automatic system achieves recognition results on frame, phone, and word level of up to 75.8 % CL. On speaker level significant and high correlations between the subjective evaluation and the automatic system of up to 0.89 are obtained. A more detailed description of the system is given in [1].

References

1. Maier, Andreas; Hönig, Florian; Hacker, Christian; Schuster, Maria; Nöth, Elmar: Automatic Evaluation of Characteristic Speech Disorders in Children with Cleft Lip and Palate Interspeech 2008 – Proc. Int. Conf. on Spoken Language Processing, 12th International Conference on Spoken Language Processing, September 22-26, 2008, Brisbane, Australia, Proceedings, to appear

Voice Quality Features for the Analysis of the Aging Voice

W. Spiegl and E. Nöth

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. In this study the changes in vocal aging are analyzed by examination of two voice quality features: jitter and shimmer.

During aging the voice changes. Among other reasons this is not only effected by changes in the place of articulation, but also by changes in the dynamics of the articulators. To cover these dynamics voice quality features are analyzed in this study. The applied features are jitter and shimmer: jitter describes the irregularities in pitch periods and shimmer the irregularities in the energy of pitch periods. The idea is with increasing age these irregularities increase as well. The calculation of the features follows the approach in [1].
In the experiment a longitudinal (same speaker, different ages) data set is analyzed: 30 recordings of Alistair Cooke (1947-2003, 10.6 h in total, 25 min each recording). The features are calculated for each recording/year on a voiced segment basis. Afterwards the values are averaged per year, which results in one feature value per year.
For the averaged jitter the results in this study show an age correlation with 0.66 on the Cooke data. In figure 1 the averaged jitter is plotted against the age. In contrast the shimmer feature does not show a trend (0.06 correlation). Hence, the averaged jitter appears as a good feature for regression/classification of Cooke's age.
In upcoming experiments the jitter and shimmer feature will be analyzed also on cross-sectional (different speakers, different ages) data sets to generalize the results.

References

1. Slyh, R.E.; Nelson, W.T.; Hansen, E.G.: Analysis of Mrate, Shimmer, Jitter, and F0 Contour Features across Stress and Speaking Style in the SUSAS Database, in ICASSP 1999, pp.2091-2094.

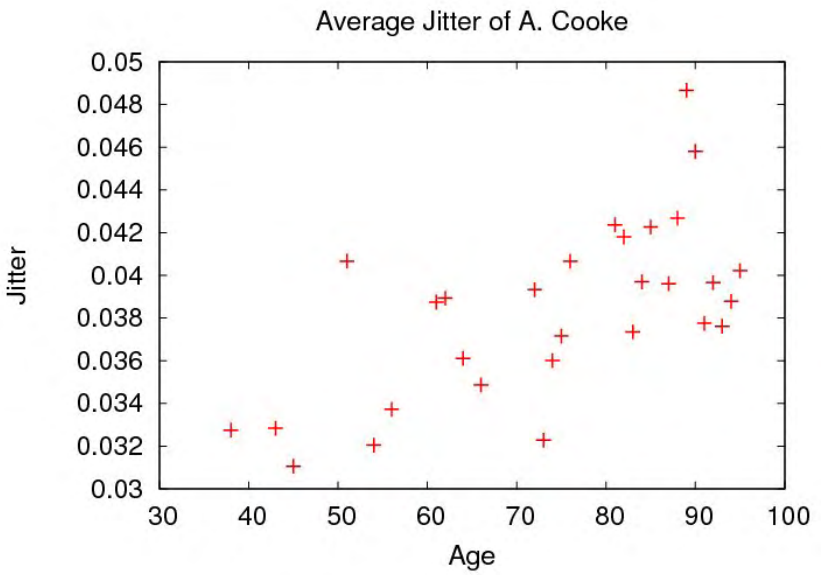
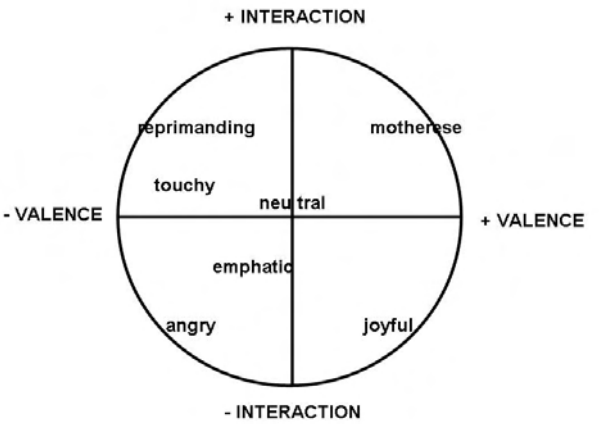


Fig. 1. The averaged jitter in the recordings of Alistair Cooke shows a trend with an age correlation of 0.66.

Emotion Recognition



Patterns, Prototypes, Performance

A. Batliner[1], D. Seppi[2], B. Schuller[3], S. Steidl[1], T. Vogt[4], J. Wagner[4], L. Devillers[5], L. Vidrascu[5], N. Amir[6], and V. Aharonson[7]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany
[2] Fondazione Bruno Kessler (FBK) – irst, Trento, Italy
[3] Institute for Human-Machine Communication, Technische Universität München (TUM), Germany
[4] Multimedia Concepts and their Applications, University of Augsburg (UA), Germany
[5] Spoken Language Processing Group (LIMSI-CNRS), Orsay Cedex, France
[6] Dep. of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University (TAU), Israel
[7] Tel Aviv academic college of engineering (AFEKA), Tel Aviv, Israel

Abstract. We address the impact of more or less clear, i.e. prototypical, cases on the classification performance in emotion recognition.

We report on classification results for emotional user states (German database of children interacting with a pet robot, recorded at our institute). We are modelling the following four classes: motherese, neutral, emphatic, angry. Starting with five emotion labels per word, we obtain chunks, i.e. meaningful units such as clauses or phrases, with different degrees of prototypicality: the more labels in a chunk belong to the class the chunk is attributed to, the higher its prototypicality. Six sites computed acoustic and linguistic features independently from each other, following in part different strategies. The initiative to co-operate was taken by us within the European Network of Excellence HUMAINE under the name CEICES (Combining Efforts for Improving automatic Classification of Emotional user States). A total of 4232 features were pooled together and grouped into 10 low level descriptor types: duration, energy, pitch, spectrum, cepstrum, voice quality, and wavelets as acoustic, and bag-of-words, part-of-speech, and semantic classes as linguistic feature types.

We apply Support Vector Machines using 150 features each with the highest individual Information Gain Ratio. Figure 1 shows the classification performance for chunks with different prototypicality, for acoustic and for linguistic features separately, and for both acoustic and linguistic features taken together. Note that above a threshold $thr = 80$ (80 percent or more of the labels belong to the class the whole chunk is attributed to) our 4-class problem is mutilated because of sparse data, thus realistic estimates can only be obtained for $thr \leq 80$.

Our data and our results can be seen as being typical for realistic databases: a tidy, balanced set of classes is not given, and can even less be maintained when going over to more prototypical constellations. However, we can demonstrate that the degree of prototypicality chosen clearly amounts to a marked difference in classification performance, e.g. for thr w/o (baseline) vs. 70 to 16.8 percent points (26.7% relative improvement). This difference is higher than the one normally obtained by optimizing feature sets or classifiers.

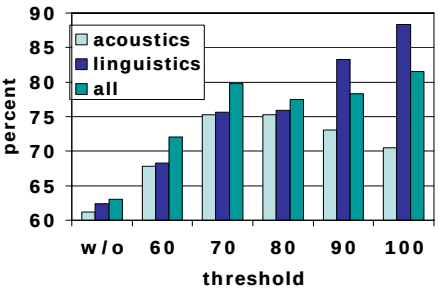


Fig. 1. Classification performance in percent on the y-axis for chunks with different prototypicality; threshold given on x-axis

If we are using only prototypes of different degrees of prototypicality for training, and are testing on the baseline thr w/o, classification performance goes down systematically. Thus prototypes cannot model variability in the data and, used for training, yield sub-optimal results. Even if our prototypes cannot simply be put on the same level as acted data, this result makes it less probable that using acted data for training is the solution for the sparse data problem.

Details and more references are given in [1].

References

1. Seppi, D., Batliner, A., Schuller, B., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., and Aharonson, V.: Patterns, Prototypes, Performance: Classifying Emotional User States. Proceedings of Interspeech, Brisbane. 2008. to appear

Emotion Recognition from Speech: Evaluation of Features based on the Teager Energy Profile

S. Steidl and E. Nöth

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. Features based on the Teager energy profile of 16 critical frequency bands, proposed by Zhou et al. for the detection of stress are evaluation for emotion recognition. The results are compared to the ones of other feature types.

Linear speech production models assume that the airflow propagates in the vocal tract as a plane wave and that this pulsatory flow is the source of the sound production. In contrast, studies by Teager [1] state that concomitant vortices are distributed throughout the vocal tract and that interactions between the vortex flows are the true source of sound production. The Teager energy operator (TEO) has been developed in order to reflect the instantaneous energy of these nonlinear interactions. The TEO is defined as follows:

$$\Psi(x_n) = (x_n)^2 - x_{n-1} \cdot x_{n+1} \approx A^2 \Omega^2 \ . \tag{1}$$

Its output is proportional to both the square of the amplitude and the square of the frequency of a harmonic oscillations $x_n = A\cos(\Omega n + \phi)$. Fig. 1 shows the Teager energy profile (TEP) for a bandpass filtered speech signal (Gabor filter centered at 1 kHz with a bandwidth of 160 Hz).

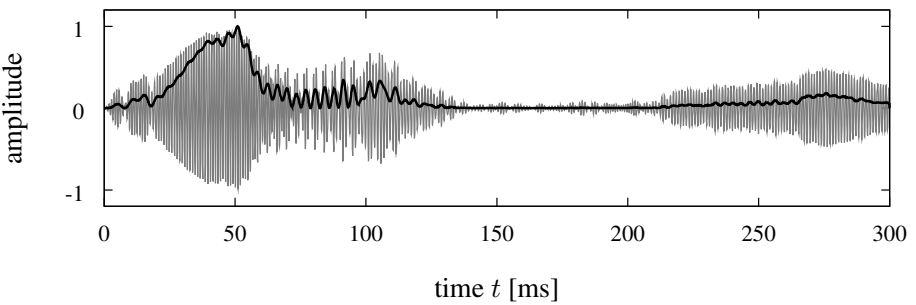


Fig. 1. Computation of the TEO-CB-Auto-Env features after [2]

In [2], features for stress detection are proposed that are based on this TEP. The most promising, the *critical band based TEO autocorrelation envelope area* (TEO-CB-Auto-Env) features have been reimplemented and evaluated for emotion recognition

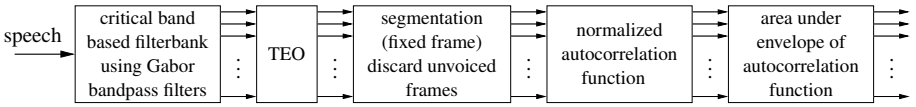


Fig. 2. Computation of the TEO-CB-Auto-Env features after [2]

on the FAU Aibo Emotion Corpus. The chain of computational steps is depicted in Fig. 2. The TEO-CB-Auto-Env features are focused on representing the variations of pitch harmonics. Experiments in [2] have shown that the fundamental frequency and consequently also the distribution pattern of pitch harmonics across critical frequency bands changes under stressful conditions compared to a neutral speaking style. The resulting TEO features are influenced both by the differences in the number of harmonic terms within each band as well as by differences in the regularity of each harmonic.

Classification results are reported for the 4-class problem ***A**nger* (subsuming the three states *angry*, *touchy/irritated*, and *reprimanding*), ***E**mphatic*, ***N**eutral*, and ***M**otherese* on the chunk level for the Aibo chunk set, which consists of 13,217 words in 4,543 chunks. The results are given in terms of the F measure, i. e. the harmonic mean of the absolute recognition rate and the unweighted average recall. If the framewise TEO-CB-Auto-Env features are classified directly on the frame level using GMMs and the a posteriori scores are multiplied for all frames within one chunk, 50.0 % F are achieved on the chunk level. The result is slightly better (52.1 % F) if the framewise features are described on the chunk level by their average, their standard deviation, and their extrema, and are then classified with artificial neural networks. Compared to other types of features, the TEO-CB-Auto-Env features are in the same range as features based on the first four formants and their bandwidths. Nevertheless, they perform worse than our set of prosodic features modeling F0, energy, and duration (59.8 % F) or our standard framewise MFCC features (61.1 % F). Unfortunately, the performance of the prosodic and the MFCC features could not be improved by adding the TEO-CB-Auto-Env features.

References

1. H. M. Teager, and S. M. Teager: Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract. NATO Advanced Study Institute on Speech Production and Speech Modelling, Vol. 55, pp. 241–261, 1990.
2. G. Zhou, J. H. L. Hansen, and J. F. Kaiser: Nonlinear Feature Based Classification of Speech Under Stress. IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 3, pp. 201–216, 2001.

Effects of Vocal Aging on Fundamental Frequency and Formants

S. Mwangi[1,2], A. Maier[1], T. Haderlein[1], F. Hönig[1], and E. Nöth[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Electrical and Electronics Department, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Abstract. The influence of vocal aging on the fundamental frequency and the formant frequencies was studied using recordings of Queen Elizabeth II that were made within a period of 50 years.

In this study we examine changes of the articulatory organs from young adulthood to old age. In particular, we focus on how the larynx and vocal folds age along with the rest of the body. The objective of the project is to automatically compute the effects of vocal aging on fundamental frequency and the formants.

The data consists of Queen Elizabeth's annual Christmas speech since the age of 26 to 76. The idea is to have the same speaker over a long period of time.

Voiced and unvoiced speech segments are discriminated by using thresholding parameters, short-time energy, zero-crossing rate and maximum amplitude. Short-time autocorrelation of a windowed speech segment is applied to reveal periodicity in a voiced speech segment. This property is used to extract the fundamental frequency of the Queen's speech. Linear predictive analysis is used to design a source-filter model. The formants are estimated based on this model by locating the acoustic resonance that makes up the filter.

The results indicate that vocal aging causes the decline of the fundamental and first formant frequencies while the second and third formants are not closely correlated to vocal aging. The correlation between the fundamental frequency and age is –0.90 while the average correlation for 1st formant and age is –0.81.

Figure 1 shows the change in the fundamental frequency in the test data.

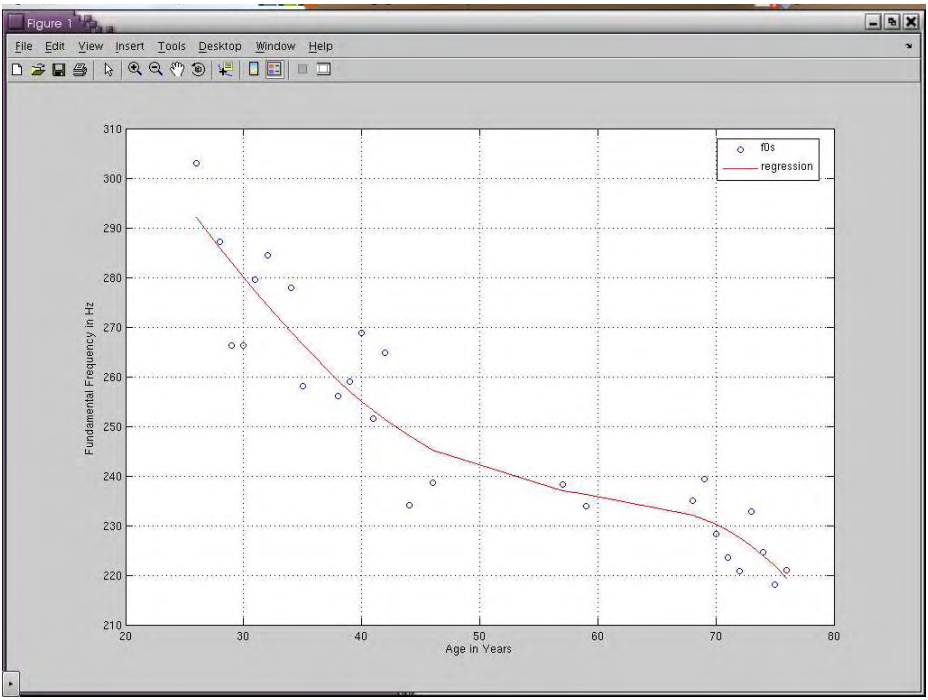


Fig. 1. Fundamental frequency in the Queen's speech between the age of 26 and 76; the regression is denoted by a fourth-order polynomial.

InformARTik: Technology, Art and Communication

S. Soutschek[1], J. Penne[2], H. Erzigkeit[1], and J. Kornhuber[1]

[1] Department of Psychiatry and Psychotherapy, University Erlangen-Nuremberg, Germany,
[2] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. The high standard of living in the first world and the continuous improvement of medical care will lead to an increase of the anticipated average live in the near future. For that reason, sustainable solutions to support elderly people in their everyday activities, to perform continuous screenings and to provide assistance in case of emergency need to be developed.[1]

It is important for the success of such technical systems that they achieve a high acceptance level primarily from the end-user, who normally is not familiar with technology. Within the project "informARTik" art objects are developed as attractive sensors, which attract attention and arouse interest to maximize the acceptance of such systems. The focus of the developed systems is to provide sensors for a continuous screening to support clinical intervention (e.g. the detection of early symptoms or estimation of risk factors). For a practical evaluation of the sensors and to show their applicability in a normal home environment, they are integrated into the "Intelligent Senior Adapted" (ISA)-house located at the Chair of Pattern Recognition.

To reduce compatibility problems when a new demonstrator system is integrated to the ISA-house, a modular environment, which can easily be adapted to new sensors or functionalities, has been integrated. Furthermore the complete implementation of the modules is done platform independent. To show the functionality and the idea behind informARTik, first demonstrators have already been implemented.

Example number one, a virtual fish tank, originally developed by Benjamin Deutsch, was re-implemented and integrated as a module. When a person enters the field of view of a Time of Flight camera, which is mounted next to the display, the fishes inside the fish tank will follow the person's head. This resulted in the observation that the direct interaction between the user and the system aroused high interest and the person started to play with the fishes. During this time, biosignals or emotions can be measured which enable the system to provide useful information concerning the physical or psychological state of the user.

Example number two, a demonstrator for the acquisition of clinical relevant data, is a fitness exercise (see figure 1), where all body parts are segmented and their position is analyzed in 3-D. With that information it is possible to gain information about the physical performance of the user.

Further demonstrators which arouse interest, increase acceptance and support clinical intervention will come up very soon.

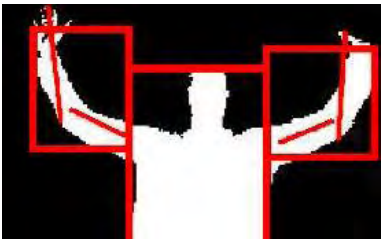


Fig. 1. Segmented bodyparts in a binarized 2 1/2-D image

References

1. Eisenmenger, M. et al.: Bevölkerung Deutschlands bis 2050. 11. koordinierte Bevölkerungsvorausberechnung. Wiesbaden: Statistisches Bundesamt 2006.

Biomedical Engineering

Analysis and Averaging of 3-D Foot Datasets

B. Eskofier[1], R. Grimmer[1], H. Schlarb[2], B. Wirth[3], J. Hornegger[1], and M. Rumpf[3]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] adidas innovation team (ait.), adidas AG, Scheinfeld, Germany,
[3] Institute for Numerical Simulation, University of Bonn, Germany

Abstract. In this work, we quantitatively examine the applicability of a Mumford-Shah based 3-D surface averaging method on laser-scanned human feet shapes.

Currently, only sparse or outdated information about the human body shape is available. This is true especially for the human foot. More information about its shape, and especially a good average shape representation, would be beneficial not only for sports manufacturers. We therefore examined the applicability of a Mumford-Shah based shape averaging technique developed by Wirth et al. [1].

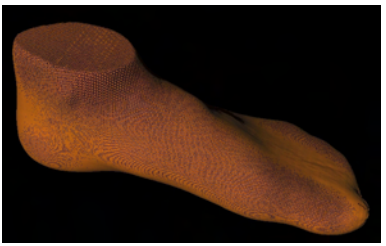


Fig. 1. Example foot.

For the quantitative analysis we used data from a worldwide study conducted by the adidas AG, who collected about 12000 3-D surface point datasets of human feet using a laser scanner made by Vorum Research, Inc., see Figure 1. Three subsets were taken from the study data:

- 24 datasets of German soccer players (GER), selection criterion was foot length $l = [256\,\text{mm},\ 270\,\text{mm}]$
- 22 datasets of Chinese women (CHN), selection criterion was foot length $l = [233\,\text{mm},\ 246\,\text{mm}]$
- 7 datasets from one individual (LME).

To average these subsets with n datasets, a shape v with minimal distance to the given shapes $v_i,\ i = 1 \ldots n$ is sought such that the accumulated energy of the deforma-

tions ϕ_i with $v_i = v \circ \phi_i$ together with a mismatch penalizing term becomes minimal:

$$E[v, v_1, \ldots, v_n, \phi_1, \ldots, \phi_n] = \sum_{i=1}^{n} (\hat{W}[\phi_i] + \gamma F[v, v_i, \phi_i]) \rightarrow \text{min!}$$

The hyperelastic deformation energy $\hat{W}[\phi_i]$ is borrowed from physics and prevents cracks and foldings of the image deformations, while the complementary energy term $F[v, v_i, \phi_i]$ penalizes a mismatch. To get to the necessary smooth representation of the surfaces, a Mumfurd-Shah approach [2] is used. It is a widely employed segmentation model for images, however, it is hard to calculate. Therefore, an Ambrosio-Tortorelli approximation [3] is used for the shape representation.

To calculate the mean shape, a multiscale method is used to pass over local minima. The images v, v_i and the deformations ϕ_i are discretized on a uniform rectangular grid via a finite element method. The discretized images V_i are obtained at each grid level separately by Ambrosio-Tortorelli segmentation, V is then updated by solving $\sum_{i=1}^{n} \langle \partial_V F|_{V,V_i,\Phi_i} \rangle = 0$ via conjugate gradient iteration. The Φ_i are then established by regularized gradient descent with Armijo line search.

The quantitative evaluation of the resulting average shapes was done using two methods developed by Grimmer et al. [4]. The first compares the volumetric distances of two datasets, while the second method computes 19 expert-defined features on the human foot such as the ball circumference or toe width.

We found that in general, the averaging method is suited for the application on human feet. All quantified foot features, as well as the volumetric discrepancies, were in the expected range. There is, however, still room for improvement. As the technique is global, small structures as for examples some bone protrusions get smoothed over. The circumvention of this problem as well as the application of the proposed method on other body surface data, e.g. of the elbow, will be the subject of further investigations.

References

1. Rumpf, M. ; Wirth, B.: Mumford-Shah-based averaging of shapes. In preparation, Germany (2008)
2. Mumford, D. ; Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics, 42(5), 1989, 577-685
3. Ambrosio, L. ; Tortorelli, V. M.: Approximation of functionals depending on jumps by elliptic functionals via Gamma-convergence. Communications on Pure and Applied Mathematics, 43(8), 1990, 999-1036
4. Grimmer, R. ; Eskofier, B. ; Schlarb, H. ; Hornegger, J.: Comparison and Classification of 3-D Objects Surface Point Clouds on the Example of Feet. Submitted to Journal of Machine Vision and Applications, Germany (2008)

Semi-Automatic Manufacturing of Medical Prostheses by using Expert Knowledge

K. Sickel[1], V. Bubnik[2], R. Melkisetoglu[2], S. Baloch[2], and S. Azernikov[2]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Siemens Corporate Research (SCR), Princeton, NJ, USA

Abstract. The current process of manufacturing medical prostheses consists of many human interaction steps, e.g. the operators have a 3-D model of the unfinished prosthesis and detail and model it using a CAD software. This has the advantage of incorporating the knowledge and experience of the operator. The linked disadvantage is that the result is not reproducible and consistent. To overcome these issues the CAD software was extended by incorporating expert knowledge into the system.

The knowledge for the manufacturing process is available as general work instructions in written form and in terms of experience of the expert users. This knowledge must be made available for the CAD software. To achieve this a script interpreter is implemented. The script interpreter is based on the GNU tools flex and bison [1].

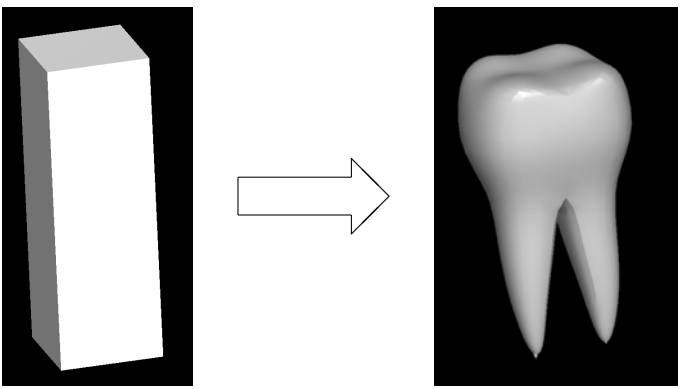


Fig. 1. Example models for a detailing and modeling process. On the left a basic 3-D model is shown. On the right is the result of a detailing and modeling process.

The associated script language is powerful enough to describe cutting and rounding operations by using special features. These features are detected on the fly on the 3-D model. The features consist of points, planes, curves and regions. The feature definition and selection is based on expert knowledge in the field and discussions with CAD

software users. The features mimic the features which the human user needs to make its detailing and modeling decisions. The challenges lie in the high shape variability. Hence robust feature detection and a script with fall back strategies are needed.

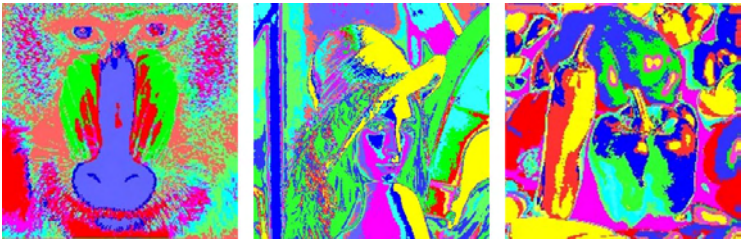
In the current testing phase the CAD software is now able to produce proposals to the user, e.g. show a cutting plane on the screen. The user can decide to accept, modify or reject the script proposal. The proposals are consistent and reproducible for every model. An expected side effect is a reduced working time per model. See fig. 1 for an example model.

This phase is characterized by its semi-automatic procedure. During a modeling session the differences between script proposal and user are stored for later analysis. In the next phase this information will be used to enhance the script performance as well as the feature recognition algorithms. The final target is to have a script based fully automatic system, which mimics the experienced user.

References

1. J. Levine, T. Mason and D, Brown, *Lex and Yacc: UNIX Programming Tools* O'Reilly Media, Sebastopol, CA, 2nd editon, 1992

Computer Vision



Albedo Estimation of Faces in 3-D

E. Eibenberger[1] and E. Angelopoulou[1]

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. Several applications, like face recognition or face relighting, require the knowledge of the skin albedo of faces. For this purpose an accurate yet relatively simple skin reflectance model describing the interaction between light and skin has to be built. The albedo estimation process itself can be expressed as an optimization problem.

Sinha et al. [3] have analyzed how the human visual system recognizes faces. Among others, they conclude that pigmentation cues of the face are at least as important as shape cues. However, automatic face recognition systems often ignore the coloring of the face, as it is affected by illumination. Highlights and shadows, for instance, change the appearance of faces. Thus, the pigmentation of the same face can significantly vary under different illumination conditions. As a result the recognition rate of face recognition algorithms is considerably reduced.

The goal of this research is the removal of illumination effects in face detection algorithms. For this purpose the albedo of the face has to be extracted. The albedo is defined as the ratio of diffusely reflected to incident electromagnetic radiation. This research project is undertaken in collaboration with Prof. Ioannis A. Kakadiaris (Computational Biomedicine Lab, University of Houston). The basis for this albedo estimation process was developed by Toderici et al. [4].

The underlying data consist of the geometry of the face and the corresponding texture, which is affected by the incident illumination conditions. An example of the input data is provided by Figure 1. For the albedo estimation process a skin reflectance model describing the interaction between light and skin has to be built up. It is essential that this model specifies both the diffuse and the specular reflectance of the skin in a physically accurate way. At the same time, the model should be simple enough, so as to be relatively computationally efficient. We are currently investigating improving the model used by Toderici et al. [4], as well as relaxing some of their assumptions.

A special case that we will investigate is regions of the skin where a very large part of the incident light is specularly reflected. In these areas no information about the skin albedo can be extracted. In order to detect highlights, a physically based method, computing the spectral gradient of the color texture, can be applied [1]. After the albedo estimation, the highlight regions will then be filled by image inpainting or a similar technique.

Another important aspect is that during the estimation process not only local illumination effects (specular and diffuse reflection), but also global illumination effects (shadows) have to be considered. All these factors, which affect skin reflectance, can make the albedo estimation process computationally expensive. Thus, we will also carefully analyze proper parameter initialization which can have a considerable impact.

Fig. 1. The input data for the albedo estimation are made up of the geometry and the texture of the face. This example is taken from the FERET database [2], in which the data were captured under uncontrolled illumination.

References

1. Angelopoulou, E.: Specular Highlight Detection Based on the Fresnel Reflection Coefficient. IEEE International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, Brazil, Oct. 2007, 1–8
2. Phillips, P.J.; Flynn, P.J; Scruggs, T.; Bowyer, K.W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W.: Overview of the Face Recognition Grand Challenge. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, USA, June 2005, 947–954
3. Sinha, P.; Balas, B.; Ostrovsky, Y.; Russel R.: Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About. Proceedings of the IEEE 94 (11) (2006), 1948–1962
4. Toderici, G.; Passalis, G.; Theoharis, T.; Kakadiaris, I.A.: An Automated Method for Human Face Modeling and Relighting with Application to Face Recognition. Proceedings of the First International Workshop on Photometric Analysis For Computer Vision (PACV 2007), Rio de Janeiro, Brazil, Oct. 2007, 1–8

On Fusion of Camera, Digital Maps and GPS Data for Lane Detection and Auto-Calibration Algorithms

A. Linarth[1 2], A. Döbert[1 2], and E. Angelopoulou[1]

[1] Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany,
[2] Elektrobit Automotive GmbH

Abstract. This work presents two studies on the application of sensor fusion methodologies for modeling a vehicle's environment. First the enhancements of a camera based lane detection system through the fusion of digital maps and GPS data is discussed. An issue on calibrating the devices is also presented, which arised from the application of the previous system into portable devices.

Regarding new technologies on automotive industry, the advanced driver assistance field is expected to grow more than 500% between 2006 and 2013 [1]. The idea behind it is, to either provide higher comfort, like easing a parking maneuver, or to enhance safety, by warning and acting on the vehicle dynamics, avoiding accidents. In such a dynamic environment challenges arise on the digital description of the vehicle's surroundings and on the identification of hazardous situations.

Within this context, we examined the potential of using fusion of different sensors' data as a means for achieving robust perception of the environment. By taking advantage of synchronized GPS data and digital maps for example, it is possible to impose constraints on a lane estimation task [2][3]. With the map information, features that better represent the lane can be selected (see Fig. 1) and the curvature of the model can be constrained during the lane estimation process. The result of this approach is mainly a better stability of the detection process. The overall computational speed is also improved, once the search space is reduced and the system is kept in a reliable tracking state for a longer time, avoiding undesired re-initialization phases.

In such systems (like an automotive setup), problems concerning final product integration must also be taken in account. If such a lane detection system is intended to run on a Portable Navigation Device (PND), usual assumptions such as fixed calibration must be relaxed. The use of standard approaches, as for example, through the acquisition of images of a known pattern, would be unacceptable for the end user, which would have to re-calibrate the system every time the device is attached back to its cradle. Such problems, though often ignored by researchers in the automotive field, can represent a challenging task, where assumptions like a static environment are far from being satisfied. A solution to this dynamic calibration problem can also take advantage of sensor data fusion, like, for example, the use of the car's velocity to estimate changes on the video sequence. Multiple view geometry [4] approaches represent possible methodologies to target such problem.

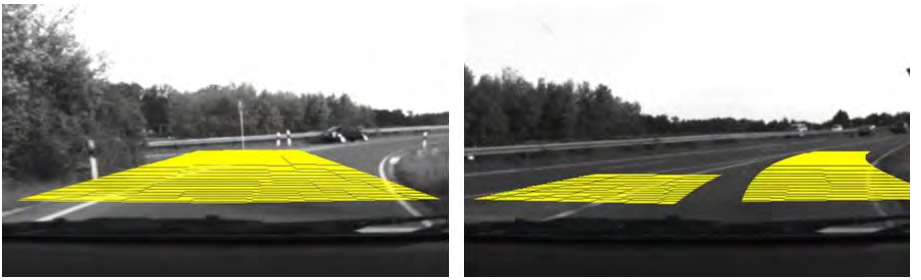


Fig. 1. Feature search region, without (left) and with (right) map and GPS data fusion

Sensing the environment around a vehicle is still an open research field. Several techniques have being studied for years but many algorithms are still far from an acceptable level required by the market. Our work has shown that sensor fusion approaches is an essential methodology, since they have the potential of enabling new solutions, improving robustness, and, in some cases, even reducing of the overall computational effort.

References

1. Mak, K.: Advanced Driver Assistance Systems: Assessing Opportunities and Challenges. Forecast, Strategy Analytics, March, 2007
2. Möhler, N.; John, D.; Voigtländer, M.: Lane Detection for a Situation Adaptive Lane Keeping Support System, the SAFELANE System. In proc. of Advanced Microsystems for Automotive Applications, Berlin, April, 2006, pages 485-500
3. Cramer, H.: Modelle zur Multisensoriellen Erfassung des Fahrzeugumfeldes mit Hilfe von Schätzverfahren. Ph.D. Dissertation, University of Technology, Chemnitz, Germany, 2006
4. Hartley, R.; Zisserman, A.: Multiple View Geometry in Computer Vision. 2nd Edition, Cambridge University Press, 2004

Towards a Feature Set for Sensor Sata Fusion

C. Riess, S. Fuchs, and E. Angelopoulou

Chair of Pattern Recognition (LME), University Erlangen-Nuremberg, Germany

Abstract. In order to successfully fuse sensor data on a level higher than the data level, it is first necessary to create the globally best possible single features with respect to the recognition result. Here, we discuss recent advances in feature extraction for quality assessment of printed products.

One of the current projects at LME deals with quality assessment in industrial applications (also known as optical inspection). Though this is a well-examined area (see for example [2]) we are investigating how the employment of diverse sensors and the fusion of their data can improve visual inspection. Our work in particular considers the validation of visible watermarks and a general model for the color distribution on printings. In the following, we briefly describe the state of our work and recent results:
Watermarks: On documents with embedded watermarks it is necessary to validate their existence and proper visual appearance. We work with a camera setup that looks at reflectance and transmittance images of the printings. In the transmittance image watermarks appear brighter than the rest of the printing. We computed 10 distance measures between a template watermark and the printing on all three RGB color channels, resulting in a vector containing 30 features for each of the different watermark regions on the printing. In the first level of processing, the features of a single region were combined and classified using decision trees and neural networks that were trained using reinforcement learning. Results from different regions were in a second level fused using again decision trees, neural networks, and also Adaboost. From three sets of printings, old and new watermarks could be discriminated with an accuracy of 80% using decision trees as the final classifier. To the best of our knowledge, no comparable types of watermark quality assessment do exist so far. Furthermore, the available training set was very small, consisting of only 13 new, 13 old and 9 samples that were altered with oil (because oil gives a similar visual impression similar to watermarks). Therefore we are convinced that our result can be further improved if additional data is used.
Color model: Aiming at more general evaluations on printings, we plan to exploit the fact that in mass production the desired output is known in advance. Therefore, we currently try to establish a more general model for describing printing quality based on color information. Currently, we consider color histograms on regions of the printing. We looked into segmentation methodologies based on unsupervised clustering (see i.e. [1]), that could help to reduce the image information. A particular problem therein is the formulation of an appropriate color distance measure. As an example, Fig. 40 shows two sample beer mats, one unchanged and one containing undesired stains, and the differences in their color histograms in RGB space. The unchanged beer mat is shown as blue dots, the stained beer mat as green crosses. In unaltered regions, the

density of the two histograms is about equal, while the stained region shows up as the dense green region in the middle right part of the shown histogram.

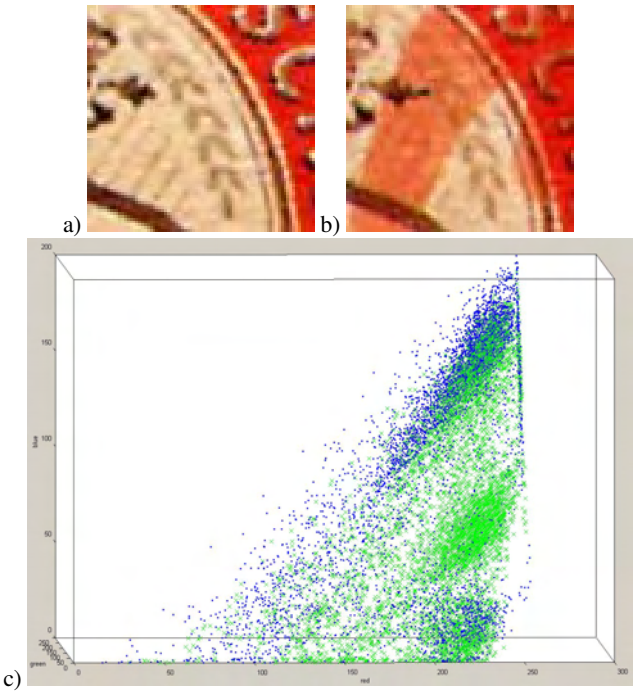


Fig. 1. (a) part of an unchanged beer mat, (b) the same part with stains on it. (c) differences in the color histogram, the unchanged beer mat appears blue, the stained green.

A future perspective is to look for possible application of these techniques in the domain of image forensics. This new research direction aims at detecting tampering of digital images.

References

1. Comaniciu, D. ; Meer, P.: Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), volume 24, number 5, May 2002, pp. 603-619
2. Malamas, E. ; Petrakis, E. ; Zervakis, M. ; Petit, L. ; Legat, J.: A survey on industrial vision systems, applications and tools. Image and Vision Computing, volume 21, number 2, February 2003, pp. 171-188

Author Index

www5.informatik.uni-erlangen.de

www.hss.de

.

ISBN 3-921713-34-1