Morphology, growth and patterning of the developing heart: methods and applications

de Boer, B.A.

Citation for published version (APA):
Towards the automatic registration of histological sections into a 3D reference model

BNAIC’07. Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence. 2007 41-48

deo Boer, B.A.
Ruijter, J.M.
Voorbaak, F.P.J.M.
Abstract
In research in cardiac development it is unavoidable to use sections of biological tissue, it is difficult to interpret these sections because of the complexity of the developing heart. 3D computer reconstructions can provide an anatomical context of such sections and make their interpretation easier. However, for practical reasons, researchers often do not stain a complete series of sections and, therefore, cannot make a 3D reconstruction. We are developing a program for tracing the anatomical context of individual tissue sections, by automatically fitting 2D sections into 3D reference reconstructions and thus enabling the retrieval of their right location and orientation. In this paper we show that a basic version of the program, using a primarily brute force pixel-based approach, already gives promising results. The performance of this basic program can substantially be improved if the program is extended with the use of relatively simple image features.
Introduction

In this paper we describe work in progress which is aimed at developing a system which automatically fits 2D images into 3D reference models for the specific application of fitting histological sections of embryonic (mouse) hearts into 3D computer reconstructions of such hearts. As a basis for this system, we first implemented a straightforward brute force pixel-based approach, where the (2D) input section is compared to a large number of virtual (2D) cross sections of the (3D) reference model. Section similarity is computed based on the distances of non-zero pixels of one image to the nearest non-zero pixels in the other image. The implementation of this computation makes use of the distance transform of the images.

The base system is being extended by including the use of some image features, and in this paper we show how the performance of the base system can already be improved substantially by using a few relatively simple features, such as density of non-zero pixels. Two performance characteristics are considered: computation time and the quality of the resulting fit. This fit quality is measured using a panel of experts.

The biomedical motivation for the system is that in research in cardiac development one cannot avoid using 2D sections. The reason for this is that the staining agents used for studying the location of gene products have a limited penetration depth. Sectioning a piece of tissue, containing the heart, leads to loss of the precise spatial orientation which hampers interpretation of the section. Having these 2D sections placed (at the proper location with the proper orientation) into a 3D reconstruction of a developing heart provides considerably more information than can be deduced from a section alone or even from a stack of sections.

The remainder of this paper is organized as follows. The background section contains some information both on biomedical issues and on the 3D computer reconstructions. In this section also related work is mentioned. Some details of the developed system (including the added features used in the extended system) are given in the next section. The results section describes the results of the performance measurements of the two versions of the developed system, and we conclude with discussion and conclusions.

Background

Biological background

In birds and mammals the mature heart consists of 4 chambers arranged in two parallel pairs: the right and left atrium are exclusively connected to the right and left ventricle, respectively. The right side of the heart serves the pulmonary circulation whereas the left side of the heart handles the systemic circulation (Moorman and Christoffels, 2003). Both heart halves beat synchronous with regular paced contractions where the contraction of the ventricles is perfectly adjusted to the contraction of the atria.

This complex dual circuited four-chambered heart is formed from a single-circuited heart tube during embryonic development. This primary heart tube pumps with a peristaltic wave. The heart of higher vertebrates must function from the moment it forms to supply the embryo with nutrients (Hoogaars et al., 2007). Therefore the transformation from a simple tube to a four-chambered heart with a sophisticated conduction system has to take place while the heart continues functioning (Moorman and Christoffels, 2003). Figure 1 shows a schematic illustration of the development of an embryonic heart of a higher vertebrate. This illustration is based on
a developmental series of a mouse heart from day 8.5 after fertilization (E8.5; embryonic day 8.5). Important to consider is the growth rate and size of the developing heart. For instance, a developing mouse heart grows from 250 µm at day E8.5 to 1.5 mm at day E14.5, i.e., a six times increase in diameter in six days (Soufan et al., 2003). The morphological changes are enormous which makes it difficult to analyse and understand cross sections through a developing heart.

3D computer reconstructions and TRACTS

In research in cardiac development it is unavoidable to use serial sectioned biological material for 3D computer reconstructions, because of the required level of detail and the limited penetration of staining agents into tissues (Ruijter et al., 2004). The problem of automatic 3D reconstruction using serial images has been studied by several authors, e.g. (Guest and Baldock, 1995; Verbeek et al., 1995). Soufan et al. (Soufan et al., 2003; Soufan et al., 2007) reconstructed a series of developing mouse hearts from sections of hearts in which in situ hybridization with markers for myocardium (heart muscle tissue) enabled the automatic selection of the myocardium. The 3D reconstructions show details and provide insights that are not clear from studying the series of sections themselves. Thus, the reconstructions are a useful tool to increase our understanding of the development of the myocardium, which is important because of the high incidence of congenital heart malformations (almost 1% of live births).

Although 3D reconstructions can provide additional insights, not all researchers choose to stain the required complete series of sections. In many cases, only a limited number of sections are stained for specific proteins or mRNAs. Moreover, these sections are often of unknown orientation and not exactly timed. Interpretation of such sections is then hampered by lack of information on their anatomical context.

To provide this information on the anatomical context, the departments of Anatomy & Embryology and Medical Informatics of the Academic Medical Center cooperate in the development of an application for TRacing the Anatomical Context of Tissue Sections (TRACTS). This is done by automatically fitting 2D sections of an embryonic mouse heart into 3D computer reconstructions of mouse hearts. The application is being developed with two main goals in mind. The first goal is to provide users with information about the orientation of their tissue sections. The second goal is to collect and present gene expression information within its anatomical context.

There are several research projects where spatial information, e.g. anatomical context, is added to gene expression databases (Sunkin, 2006). One of those, closely related to our work, is EMAGE (Baldock et al., 2003; Christiansen et al., 2006). There are two main differences between the
Towards Automatic Registration of Histological Sections

approaches of EMAGE and TRACTS. Firstly, EMAGE is embryo wide, while we primarily focus on the heart. Secondly, the spatial information in EMAGE is found manually with the MAPaint software (http://genex.hgu.mrc.ac.uk/Software/paint/), whereas with TRACTS we aim at developing a program for doing this automatically.

There is a lot of research done on (medical) image registration (for reviews see (Hill et al., 2001; Maintz and Viergever, 1998)). Most techniques described are developed for the registration of different 2D images, or of 2D projections to 3D spatial data, or of 3D data to 3D data. There are only a few publications dealing with the problem of fitting 2D slices into 3D volume data. One example of such work is Birkfellner et al. (2007), but the approach described in this work is not aimed at an application with high morphological variation. Another example is the matching of a complete stack of histological sections to a 3D MRI image (Jacobs et al., 1999). The use of a stack makes this approach essentially a 3D to 3D registration problem.

The aim of TRACTS is to fit histological sections to a developmental series of 3D heart models. These models, which will be referred to as reference models, are reconstructed from a stack of histological sections, which are stained using a mix of probes specifically chosen to visualize the myocardium. With a simple threshold the myocardium is extracted, resulting in binary images (Fig. 3). In the present version of the program only one reference model is used. During further development of the program, a complete series of reference models, covering all relevant stages of development, will be implemented. The reference model used is a reconstruction of an E12.5 mouse heart (Fig. 2a) (Soufan et al., 2003). The resolution of this reference model is 321x269x173 voxels with a voxel size of 5.155x5.155x7 µm.

For the performance measurements described below we used another 3D reconstruction of a heart of the same age as the reference model (Fig. 2b). This reconstruction that will be referred to as test model is used to create a test set of virtual sections. The resolution of the test model is 290x304x110 voxels with a voxel size of 5.155x5.155x10 µm.

The occurrence of biological and technical variation is obvious when observing the reference and test model (Fig. 2). Although both models are approximately the same age, they are quite different in appearance. The test model (Fig. 2b) clearly shows some artefacts, caused by deformations during the sectioning. These deformations made some neighbouring sections hard to align during the reconstruction procedure.

Figure 2a. Reference model

Figure 2b. Test model
The developed system

Basic program

As a basic version of the automatic fitting program, we first implemented a straightforward brute force pixel-based approach, where a 2D input section is compared to a large number of virtual 2D cross sections of the 3D reference model. To avoid repetitive computations, a database was created containing cross sections of the reference model. Only for the cross sections which are comparable in size, a similarity with the input section is computed.

The similarity measure used is based on the mean Euclidean distance of points on the contours of the heart in the input section to the contours in every cross-section through the reference model, and vice versa. The contour is considered to be a good representation of the shape of the heart. To compute this distance measure a distance transformed image is made from the input section and from the reference model sections, all resized to 128x128 pixels (Fig. 3). The pixel values in these distance transformed images represent the Euclidean distance to the heart contour. To find the distances between the contours, the distance transformed image of an input contour image is masked by the contour of the reference image, the remaining distance values are summed and divided by the number of pixels in the masking contour. This is done both ways and the two resulting average distance values are summed to supply the distance measure.

The database of cross sections of the reference model contains sections that were generated by taking a step size of 1 voxel over the three central axes. At every position 64 cross-sections are computed with a different combination of tilting angle and tilting direction. The tilting angle reached from 0 to 40 degrees (in steps of 10 degrees) and the number of tilting directions increases with each increment of the tilting angle. All model sections were normalized to a pixel size equal to the original x and y resolution of the model. Only those sections containing over 20 pixels of myocardium were used. Every section was cropped and the size of the bounding box of the heart was determined and stored. Then the model images were resized to a standard size of 128x128 pixels. The contours of the heart in those resized images were determined and their distance transformed images were computed. Both the contour and the distance transformed images are stored. This resulted in 44711 image pairs stored in the database. The positional and size information were stored in a lookup table.

Before an input section can be compared to the reference model some pre-processing has to take place. In the input section the myocardium needs to be separated from the background using a threshold. The resulting binary image is rotated in 32 steps and mirrored at each position. The resulting 64 images are processed like the model sections: determine size, resize to 128x128 pixels, determine contour and compute distance transform. The resulting images are used for comparison with the reference model.

The basic program uses essentially a brute force strategy. However, for computational reasons, not all sections from the reference model are compared to the input section: Only those that are similar in size are selected. The size difference between both sections may not exceed 20%, which is based on expert knowledge that embryonic hearts never differ over 20% in size at the same developmental stage. For every input and model section pair that meets the size criterion the distance measure is calculated. The lowest distance measure is considered to indicate the best match.
**Extended program**

The basic version of the program already uses a restricted brute force approach, since it selects only cross-sections from the reference model comparable in size to the input section. As we will show in the performance measurement, this basic version results in a considerable number of fits which are rated unsatisfactory by experts. To improve on this, we extended the basic program by adding the use of some features in an attempt to exclude the unsatisfactory fitting sections. Since this results in a lower number of distance transform based comparisons, it has the added benefit of making the program faster.

The decision which features to use is based on an analysis of the performance measurement of the basic program. Features that exclude the unsatisfactory rated fit results without affecting the good rated fits are preferred. Several potential features were tested, and three of these were selected based on their individual performance. Thresholds for these features where determined after some experimentation and in consultation with experts.

The first feature implemented was the tissue density. The basic program regularly fits sections with a large amount of tissue, for instance a section containing a thick ventricular wall, to sections with a lower number of tissue pixels, for instance a section containing only a thin atrial wall. Such a difference in thickness is reflected by the part of the section containing tissue. This tissue density is measured by counting the number of pixels containing tissue in the resized images.

With the previous feature it is still possible to fit atria on ventricles and ventricles on atria when the amount of tissue is very similar. Therefore, we introduce the centre of mass of the section as a second feature. This feature prevents that images with a very different tissue distribution are matched.

The last implemented feature uses a regional tissue density measure. For this feature the density in the four corners is measured of every section. To compensate for both technical and biological deformations of the tissue, the density measurement of the input sections is done by moving the region of interest (Fig. 4).

**Performance measurements**

To measure the performance of the fitting program two characteristics are considered: computation time and the quality of the resulting fit. The reported computation times are obtained using a 3 GHz Pentium IV processor, with 1024 Mb of RAM. An expert panel was asked to rate the fit results, since the actual spatial relation between the test model and the reference model is unknown and there is no gold standard fit algorithm available. Four experts were asked to rate the similarity of the anatomical context of the test section and the result section, on a scale from 0-10.

Before the performance of the basic and extended program was measured, we did some preliminary tests on the program algorithms for errors and computation time. The comparison with all the sections of the database takes on average 27.5 minutes per input section. This prompted our decision to use a restricted brute force approach in the basic program.

In one of these preliminary tests a sample of 100 sections from the reference database is fitted to the reference model. This test resulted in a perfect match for all sections. Another test was used to determine whether the resolution of the reference database is sufficient. To this end, a set of 100 sections was created from the reference model with random choices of the location on the three axes, the tilting angle and the tilting direction. All sections containing heart tissue were
fitted to database sections closely resembling the random chosen input images. From this we concluded that the resolution of the reference database is sufficient.

After the preliminary tests the program was ready to be tested in its basic version. A set of arbitrary cross sections was taken from the test model. Cross sections deemed irrelevant by experts, for example peripheral sections containing just small caps of indeterminable tissue, were removed from this set, which left a sample of 65 relevant cross-sections. This test set was fitted to the reference model. The mean computation time was just over 6 minutes per input section. The mean individual rating per fitted section is plotted in figure 5 (error bars indicate the standard deviation between experts). The overall mean rating of the basic program was 6.6.

In figure 5 also the mean individual rating per section for the extended program is plotted. These results were obtained using the same test sample as used with the basic program. In this case the overall mean rating is 7.5, and the computation time dropped from 6 to 1 minute per section, which we consider adequate for the intended use of the program. The right side of figure 5 shows that a considerable part of the previously unsatisfactory rated sections shows much higher ratings in the extended program.

In figure 6 the distributions of the mean rating for every section for both the basic (a) and the extended (b) is plotted. To make the interpretation easier these ratings are divided into three categories: Good when the mean rating is 8 or higher, reasonable when rated between 5.5 and 8, and unsatisfactory when rated below 5.5. The anatomical context of each reference section found by the basic program is rated good in 41% of the cases by the expert panel, whereas 37% is considered reasonable and 22% unsatisfactory. For the extended program, the fraction of good-rated sections increased to 60%, whereas the reasonable and unsatisfactory ones decreased to 26% and 14%, respectively.

**Conclusion and discussion**

As shown in the previous section, the program performance increases with the implementation of relatively
simple features. The fraction of fitted sections rated as good by the expert panel increased 1.5 times in the extended program compared to the basic program. At the same time the fraction of reasonably and unsatisfactory rated sections dropped by one third. This promising degree of performance improvement in fit quality is accompanied by a six-fold drop in computation time to approximately 1 minute per section. This computation time is adequate for the intended use of the program and needs therefore no further improvement.

In spite of the fit quality improvement, the performance in this area is not yet sufficient, and we expect it will be possible to obtain better results in the near future. However, a performance of 100% good fits might be an unrealistic goal, since due to biological variation not every input section may have a satisfactory counterpart in the reference model. Our aim is to have a good fit whenever possible.

In figure 5 the mean rate per section given by the expert panel is shown. This graph clearly shows that many of the sections that had an unsatisfactory fit in the case of the basic program fitted much better when the extended version was used. We tried to design the features in such a way
that none of the good fit results from the basic program was lost using the extended program. We succeeded in all cases except for one section (number 20), in which the mean rate dropped from 8.25 to 4.75. A detailed further analysis of this section is needed to evaluate whether it is possible to correct this problem without losing performance elsewhere.

There are two noteworthy remarks to be made on the test set. The first remark is that we used the test set for the performance measurement of the extended program, while the same test set was used for measuring the performance of the basic program and for selecting the features to implement in the extended version. Therefore the performance of the extended program might be a bit overestimated. The second remark is that the sections taken from the test model were more deformed than usual for a “real” histological section. This extra deformation is a result of the reconstruction process, in which not all sections could be perfectly aligned (as is clearly visible in Fig. 2b). As a result of this extra deformation, program performance might be underestimated. When we will test a (near) final version of the program we will take care to avoid both problems by using an independent test set based on a high quality reconstruction.

In further development the program will be expanded with the ability to fit heart sections of ages other than the E12.5 model used in the present version. Due to the enormous change in shape during embryonic development there is no guarantee that the presently implemented features will work for all ages, but given the simplicity of the features it is expected that they will be reasonably robust. Nevertheless, when additional features will be added to further improve performance, it is likely that these features will be more specific to a particular developmental stage of the heart.

Theoretically the currently used method should be applicable to a wide range of applications for medical imaging where 2D sections are obtained and a 3D reference model is available. To reach an optimal performance the thresholds used for the features have to be optimized.

Future work will also include a re-evaluation of the chosen resolution of the reference database and a study of other similarity measures than the presently used measure based on (Euclidian) distance of non-zero pixels to the nearest non-zero pixel in the other image. Other measures, for example based on mutual information or on Fourier Descriptors, could perhaps provide better performance when used instead of, or in addition to, the distance transform based measure. We would also like to be able to report not just the best fit, but also the reliability of the found fit.

Presently, expert knowledge is used to judge the fit quality and based on these judgements new features are proposed. In the future we will consider using features directly obtained from expert knowledge, such as the absence or presence of particular crucial structures. It is unlikely that such features could replace the use of a pixel-based comparison altogether, but they could be a useful addition to a pixel-based approach.

**Acknowledgements**

We thank Maurice van den Hoff, Mathilda Mommersteeg and Alexandre Soufan for their participation in the expert panel and helpful analyses. Antoon Moorman and Arie Hasman gave valuable suggestions on both program development and the use of features. We like to thank Jaco Hagoort for providing the 3D reconstructions and his technical support.
References


