Robust Identification of Contrasted Frames in Fluoroscopic Images

Matthias Hoffmann[1], Simone Müller[1], Klaus Kurzidim[2], Norbert Strobel[3], Joachim Hornegger[1,4]

[1]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[2]Klinik für Herzrhythmusstörungen, Krankenhaus Barmherzige Brüder, Regensburg, Germany
[3]Siemens AG, Healthcare, Forchheim, Germany
[4]Erlangen Graduate School in Advanced Optical Technologies (SAOT), Erlangen, Germany
matthias.hoffmann@cs.fau.de

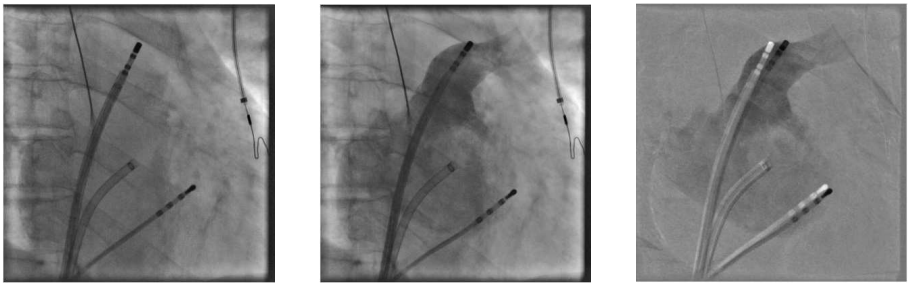
Abstract. For automatic registration of 3-D models of the left atrium to fluoroscopic images, a reliable classification of images containing contrast agent is necessary. Inspired by previous approaches on contrast agent detection, we propose a learning-based framework which is able to classify contrasted frames more robustly than previous methods. Furthermore, we performed a quantitative evaluation on a clinical data set consisting of 34 angiographies. Our learning-based approach reached a classification rate of 79.5%. The beginning of a contrast injection was detected correctly in 79.4%.

1 Introduction

Atrial fibrillation (AF) is a heart rhythm disorder characterized by fast and chaotic electrical excitation of the atria [1]. It is the most common heart disease with about two million of the US population affected. A common catheter based therapy of AF is the electrical isolation of the pulmonary veins performed in electro-physiology (EP) labs [2]. EP-labs are typically equipped with a C-arm X-ray system that is used to acquire fluoroscopic images of the left atrium (LA). To get a better 3-D impression in the projective X-ray images, biplane C-arm systems can be used, to acquire images from two different directions at the same time. Contrary to catheters, the LA is only visible in fluoroscopic images, if a contrast agent is injected. In order to provide the surgeon with a permanent overview of the heart's shape, a 3D-model may be superimposed on the fluoroscopy [3]. If such 3-D models are gained pre-operatively by CT or MRI, a registration to the C-arm coordinate system has to be performed.

Nowadays the registration is mostly performed manually [4], but recently, also automatic approaches based on contrast agent were proposed [5,6]. Both require an uncontrasted and a contrasted frame to obtain the contrast agent by means of a Digital Subtraction Angiography (DSA), see Fig. 1 for an example.

H. Handels et al. (Hrsg.), Bildverarbeitung für die Medizin 2015, Informatik aktuell, pages 23 - 28 DOI: 10.1007/978-3-662-46224-9_6,



(a) (b) (c)

Fig. 1. Example of a reference frame (a), a contrasted frame (b) and a computed DSA-image of one acquisition (c).

So far, only Zhao et al. [6] published an approach for automatic detection of contrasted frames for LA angiographies.

Approaches to contrast agent detection for different anatomical structures were proposed by Condurache et al. [7], Chen et al. [8] and Liao et al. [9]. Condurache et al. [7] used the 98-percentile of image intensities to classify contrasted images of coronary arteries, Chen et al. [8] applied for this task a learning-based framework. Liao et al. [9] published an approach for detecting contrast agent in images of the aortic root by calculating the similarity of histograms. However, this method was designed to decide if a sequence contains contrast agent at all and not to detect the first contrasted frame.

The detection of contrasted frames in LA angiographies is more complicated: As the LA is a large object, the density of contrast agent is lower. Additionally, the large movement of catheters in the LA may introduce artifacts in subtraction images and their histograms. We present a robust identification of contrasted frames in fluoroscopic images which extends the methods of Zhao et al. and Condurache et al. and combines them into a learning based framework using a support vector machine (SVM). As there exists no quantitative evaluation of the method by Zhao et al., we evaluate our novel approach as well as the approach by Zhao et al. and an adaption of the method by Condurache et al. using a set of 34 clinical fluoroscopic sequences.

2 Materials and Methods

All methods described here require a sequence of fluoroscopic images $\boldsymbol{I}_k$, $k = 1, \ldots, n$ of n containing contrasted and uncontrasted frames. They make all use of DSA: To obtain the injected contrast agent, an uncontrasted reference frame $\boldsymbol{I}_\mathrm{r}$ is subtracted from a contrasted frame k, $\boldsymbol{I}_{\mathrm{DSA},k} = \boldsymbol{I}_k - \boldsymbol{I}_\mathrm{r}$.

As our method takes up some ideas of the methods by Zhao et al. and Condurache et al., we first describe them briefly in Section 2.1. Then, we present in Section 2.2 how we integrate and extend them into a learning-based framework. Finally, in Section 2.3 we present our evaluation material and error measures.

2.1 State of the art

Approach according to Zhao et al. [6] First, a Difference Digital Subtraction Angiography (DDSA) $\boldsymbol{I}_{\mathrm{DDSA},k}$ is calculated by subtracting two neighboring fluoroscopic frames to estimate how the amount of contrast agent changes.

$$\boldsymbol{I}_{\mathrm{DDSA},k} = \boldsymbol{I}_{\mathrm{DSA},k} - \boldsymbol{I}_{\mathrm{DSA},k-1} = \boldsymbol{I}_k - \boldsymbol{I}_{k-1}. \tag{1}$$

A threshold T_{Z}, determined by e.g. a training step, is applied to $\boldsymbol{I}_{\mathrm{DDSA},k}$ to detect the region with newly contrasted pixels. The number of pixels $\boldsymbol{p}$, with $\boldsymbol{I}_{\mathrm{DDSA},k}(\boldsymbol{p}) < T_{\mathrm{Z}}$ is denoted by n_k. Given a second threshold value N_{pixel}, the first frame with $n_k > N_{\mathrm{pixel}}$ is taken as first contrasted frame. The value of N_{pixel} was empirically set as 1000 in a 256×256 image, i.e. 1.53% of the image size.

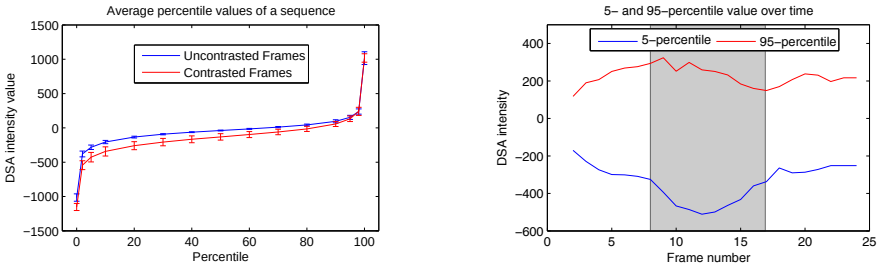
Approach according to Condurache et al. [7] First, a tophat-filter is applied to the images to highlight vessel-like structures. Then, the 98-percentile value $x(k)$ of the pixel intensities is computed over all frames k. To denoise the curve $x(k)$ it is low-pass filtered resulting in $y(k)$. The distribution of $y(k)$ in uncontrasted images in this sequence is modeled as a Gaussian $\mathcal{N}_0(\mu_0, \sigma_0^2)$. μ_0 and σ_0 are estimated using frames which are known to be uncontrasted. Based on the concept of a significance test, they compute a threshold $T_{\mathrm{C}} = \mu_0 + l \cdot \sigma_0$. A frame k is considered to be contrasted if $y(k) > T_{\mathrm{C}}$.

To adapt this method to angiographies of the LA, we replace the tophat-filter by a DSA as a tophat-filter is not suitable to enhance large contrasted areas which appear in LA angiographies. We estimate μ_0 and σ_0 using the first three frames which are always uncontrasted. l is estimated in a training step.

2.2 Learning-based Framework

We use a linear SVM [10] to combine several features which are partially inspired by the methods described in section 2.1. Before computing the features, the image borders which are covered by shutters are removed. The first part of the features is based on pixel intensity percentiles as proposed by Condurache et al. To enhance the contrasted area we used subtraction images $\boldsymbol{I}_{\mathrm{DSA},k}$ with $\boldsymbol{I}_0$ as reference frame instead of a tophat-filter. Furthermore, we used not only the 98-percentile but chose 15 percentiles, namely the 0-, 2-, 5-, 10-, 20-, 30-, 40-, 50-, 60-, 70-, 80-, 90-, 95-, 98-, 100-percentiles of DSA images. These features for frame k are denoted by f_1^k to f_{15}^k. Like Zhao et al. [6], we used features based on the computation of DDSA-images, see Equation 1. They are referred to as f_{16}^k to f_{20}^k. These features are the sums of pixels below a certain threshold. We use the values 0, -50, -100, -150, -200 as thresholds.

In most cases, the contrast state of a frame is the same as for its neighboring frames. So, the features f_1^{k-1} to f_{20}^{k-1} of the previous frame and the features f_1^{k+1} to f_{20}^{k+1} of the following frame are included. To set the feature values in relationship to the overall brightness which varies from patient to patient, the mean value of the 98-percentile of $\boldsymbol{I}_{\mathrm{DSA},2}$ and $\boldsymbol{I}_{\mathrm{DSA},3}$ is added as feature f_{21}



(a) (b)

Fig. 2. (a) Mean value and standard deviation of the percentiles for contrasted and uncontrasted frames. (b) Value of the 5- and 95-percentile. It becomes clear that the 5-percentile is better suited to distinguish contrasted frames (gray area) from uncontrasted frames.

which is constant for all frames of a sequence. The feature vector for frame k includes 61 features: $\boldsymbol{f}^k = (f_1^{k-1}, \dots f_{20}^{k-1}, f_1^k, \dots f_{20}^k, f_1^{k+1}, \dots f_{20}^{k+1}, f_{21})^T$. If a biplane system is used as in our case, the information of both views can be combined ending up in 122 features per frame. A feature selection is done by a three-fold crossvalidation on training data in order to get an insight into the importance of the different features.

2.3 Evaluation

We evaluated our SVM-based approach on 34 biplane angiographic sequences from 15 patients. The image series had a resolution of 1024×1024 pixels, 7 fps and different maximal frame numbers from 8 up to 57. Furthermore, an evaluation of the approach by Zhao et al. and Condurache et al. was performed. To allow a fair comparison, the parameters T_Z and l were trained in a leave-one-out manner as well as the SVM. So, the feature selection was performed as three-fold crossvalidation nested into the leave-one-out crossvalidation. The parameter C of the SVM was set to $\frac{n_c+n_u}{2n_c}$ and $\frac{n_c+n_u}{2n_u}$ for the training samples of n_c constrasted and n_u uncontrasted frames, respectively. The first frame was used as uncontrasted reference frame and was excluded from training and evaluation as well as the second frame which uses features from the first frame, too. Also frames where contrast agent flows out from the LA are not used for training as it is often difficult to decide whether they should be labeled contrasted or not. However, these ambiguous frames were used for evaluation but marked separately.

We performed two types of evaluation: For automatic contrast based registration [5,6] it is important to identify the first contrasted frame. Therefore we calculated the difference between the position of the first contrasted frame and the position of the first frame which was classified as contrasted. Sequences with a difference of -1, 0 or 1 were considered as detected correctly. Second, we measured the classification rate. This evaluation was not performed for the method by Zhao et al. [6] as it can only find the first contrasted frame.

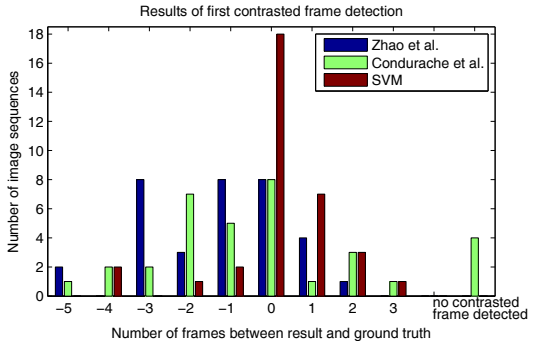


Fig. 3. Difference between the first contrasted frame and first frame classified as contrasted. A negative difference indicates a too early detected contrast injection.

Table 1. Confusion matrices for the results of our evaluation

		all frames				unambigous frames			
		Condurache		SVM-based		Condurache		SVM-based	
		contr.	uncontr.	contr.	uncontr.	contr.	uncontr.	contr.	uncontr.
class	contrasted	240	90	295	35	142	25	145	22
	uncontrasted	110	131	82	159	36	80	9	107

3 Results

Results for the first frame detection are provided in Fig. 3. The confusion matrix is given in Table 1 for the approach based on Condurache et al. and the SVM based algorithm. In 79.4% of all sequences, the SVM-based approach was able to find the first contrasted frame, the approach by Zhao et al. found in 58.8% the first frame and the adaption of Condurache's approach found in 41.2% the first frame correctly. The classification rate was 79.5% for the SVM-based approach and 65.0% for the adaption of Condurache's method. If only clearly contrasted and clearly uncontrasted frames were evaluated, the classification rate was 89.0% for the SVM-based approach and 78.4% for the approach according to Condurache et al.

4 Discussion

It turns out that the approach by Condurache et al. does not perform well, probably as the first three frames are not sufficient to estimate μ_0 and σ_0 reliably. Also the approach by Zhao et al. yields unsatisfying results as it compares each frame only to its previous frame and relies on a strong diffusion of contrast agent from one to the next frame. This is, however, not always given, especially for higher frame rates. The SVM-based approach can take much more information into account to distinguish contrasted from uncontrasted frames. Also additional

information, e.g. from previous frames or from a second image plane, could easily be integrated to achieve more reliable results. This results in a high classification rate and a more reliable detection of the first contrasted frame. The features selected most often in the feature selection were the 2-, 50- and 90-percentile values, the number of pixels in the DSA image below an intensity of -150 and the mean of the 98-percentile of the first two DSA frames.

Acknowledgments. This work was supported by the German Federal Ministry of Education and Research (BMBF) in the context of the initiative Spitzencluster Medical Valley - Europäische Metropolregion Nürnberg, project grant No. 12EX1012A. Additional funding was provided by Siemens AG, Healthcare.

Disclaimer. The concepts and information presented in this paper are based on research and are not commercially available.

References

1. Calkins H, Brugada J, Packer D, et al. HRS/EHRA/ECAS Expert Consensus Statement on Catheter and Surgical Ablation of Atrial Fibrillation: Recommendations for Personnel, Policy, Procedures and Follow-Up. Europace. 2007;9(6):335 – 379.
2. Dilling-Boer D, van der Merwe N, Adams J, et al. Ablation of Focally Induced Atrial Fibrillation:. J Cardiovasc Electrophysiol. 2004;15(2):200–205.
3. Brost A, Raab J, Kleinoeder A, et al. Medizinische Bildverarbeitung für die minimalinvasive Behandlung von Vorhofflimmern. Deutsche Zeitschrift für Klinische Forschung, Innovation, und Praxis (DZKF). 2013;17(6):36 – 41.
4. Bourier F, Vukajlovic D, Brost A, Hornegger J, Strobel N, Kurzidim K. Pulmonary vein isolation supported by MRI-derived 3D-augmented biplane fluoroscopy: A feasibility study and a quantitative analysis of the accuracy of the technique. J Cardiovasc Electrophysiol. 2007;115:3057–3063.
5. Thivierge-Gaulin D, Chou CR, Kiraly A, et al. 3D-2D registration based on mesh-derived image bisection. Biomedical Image Registration. 2012; p. 70–78.
6. Zhao X, Miao S, Du L, Liao R. Robust 2-D/3-D Registration of CT Volumes with Contrast-Enhanced X-ray Sequences in Electro-physiology Based on a Weighted Similarity Measure and Sequential Subspace Optimization. IEEE International Conference on Acoustics, Speech and Signal Processing. 2013; p. 934–938.
7. Condurache A, Aach T, Eck K, Bredno J. Fast detection and processing of arbitrary contrast agent injections in coronary angiopgraphy and fluoroscopy. Procs BVM. 2004; p. 5–9.
8. Chen T, Funka-Lea G, Comaniciu D. Robust and Fast Contrast Inflow Detection for 2D X-ray Fluoroscopy. Medical Image Computing and Computer-Assisted Intervention. 2011; p. 243–250.
9. Liao R, You W, Liu Y, Yan M, John M, Shea S. Integrated spatiotemporal analysis for automatic contrast agent inflow detection on angiography and fluoroscopy during transcatheter aortic valve implantation. Medical Physics. 2013;40(4).
10. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273–297.

Errata

Mistakes found after publication were corrrected in this version. The changes made are listed below:

- In section 2.2, the original publication states

  "the mean value of the 98-percentile of $\boldsymbol{I}_2$ and $\boldsymbol{I}_3$ is added as feature f_{21}".

  $\boldsymbol{I}_2$ and $\boldsymbol{I}_3$ were replaced by $\boldsymbol{I}_{\mathrm{DSA},2}$ and $\boldsymbol{I}_{\mathrm{DSA},3}$, respectively.

- In section 2.2, the original publication states

  "The feature vector for frame k includes 64 features:
  $\boldsymbol{f}^k = (f_1^{k-1}, \dots f_{21}^{k-1}, f_1^k, \dots f_{21}^k, f_1^{k+1}, \dots f_{21}^{k+1}, f_{21})^T$. If a biplane system is used as in our case, the information of both views can be combined ending up in 128 features per frame."

  Actually, f_{21} was included only once into the feature vector, ending up with 61 features per frame and 122 for a biplane system, respectively. This part was replaced by

  "The feature vector for frame k includes 61 features:
  $\boldsymbol{f}^k = (f_1^{k-1}, \dots f_{20}^{k-1}, f_1^k, \dots f_{20}^k, f_1^{k+1}, \dots f_{20}^{k+1}, f_{21})^T$. If a biplane system is used as in our case, the information of both views can be combined ending up in 122 features per frame."