

Automatic Key Frames Detection in Intravascular Ultrasound Sequences

Francesco Ciompi^{1,2}, Oriol Pujol^{1,2}, Simone Balocco^{1,2}, Xavier Carrillo³, Josepa Mauri³ and Petia Radeva^{1,2}

¹Dep. of Applied Mathematics and Analysis, University of Barcelona, Spain

²Computer Vision Center, Campus UAB, Bellaterra, Barcelona, Spain

³University Hospital “Germans Trias i Pujol”, Badalona, Spain

fciompi@maia.ub.es

Abstract. We present a method for the automatic detection of key frames in Intravascular Ultrasound (IVUS) sequences. The key frames are markers delimiting morphological changes along the vessel. The aim of defining key frames is two-fold: (1) they allow to summarize the content of the pullback into few representative frames; (2) they represent the basis for the automatic detection of clinical events in IVUS. The proposed approach achieved a compression ratio of 0.016 with respect to the original sequence and an average inter-frame distance of 61.76 frame, minimizing the number of missed clinical events.

1 Introduction

Intravascular Ultrasound (IVUS) is a catheter-based imaging technique generally used during percutaneous interventions. In clinical practice an IVUS acquisition consists in a set of frames obtained while the probe, inside the vessel, is pulled-back (pullback) at constant velocity (0.5 - 1.0mm/sec). As a result, a sequence of thousands of frames reproducing the internal vascular morphology is obtained (see Fig. 1). Despite the large amount of data in a pullback, physicians mainly focus on regions of the vessel characterized by clinically relevant observations (“*clinical events*”) such as presence of *stenosis*, characterized by *calcifications* or *lipid* pools, indicating potential threats for the patient health. An IVUS sequence can be considered as a connected series of clinical events (see Fig. 1). Since the presence of interesting clinical events always implies a morphological change, the quick identification of these regions is a fundamental task for the diagnosis of the patient clinical conditions; this is necessary for guiding physicians in choosing the most suited treatment for Percutaneous Coronary Intervention.

Current IVUS equipments provide two representations of the vessel: the *short-axis* view, representing the morphology at a fixed position of the vessel, and the *longitudinal* view, representing the morphology of the vessel at a fixed angular position. By using a cursor, the user can navigate the pullback in the two representations. Three main limitations affect the effectiveness of this procedure: (1) the number of redundant frames that need to be identified before finding the optimal visualization is huge; (2) the visualization of only one frame in the short-axis view hinders the possibility of comprehensively analyzing the vessel properties along the whole pullback; (3) in the longitudinal view, severe lesions can result invisible due to an inappropriate choice of the angle. The analysis of the vessel condition is thus reduced to a time consuming coarse-to-fine

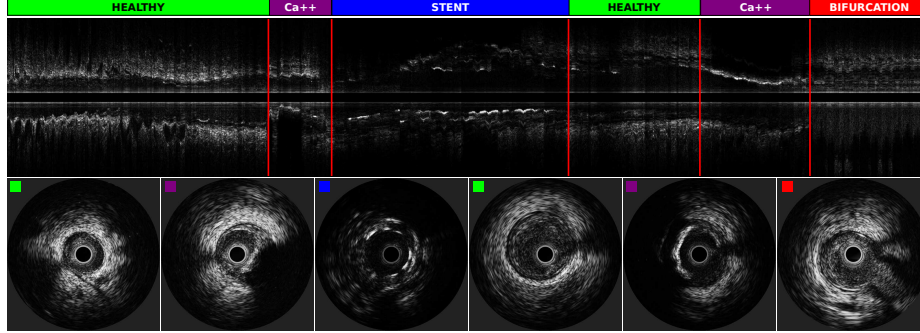


Fig. 1. Example of case navigation/visualization by means of key frames. Main clinical events (top) and their respective end points are marked in the longitudinal view of the pullback (center) as well as the corresponding key frames (bottom).

manual search by alternately changing both the angular and the longitudinal position of the cursor.

In this paper we present a technique for the automatic detection of key frames in IVUS sequences. Although some works on video summarization in medical imaging have been presented [1–3], to the best of our knowledge, this is the first time a video summarization method is proposed in IVUS. A key frame is commonly defined in literature as a frame in a sequence that marks the position where a significative change in the content occurs. In the context of an IVUS sequence, we assume that a key frame marks the separation between two consecutive morphological conditions of the vessel in the pullback. The benefit achieved with the detection of key frames is two-fold: (1) since the number of morphological events in a pullback is much smaller than the number of frames, the inspection by key frames allows a quick overview of the vessel clinical condition “*at a glance*” (see Fig. 1); (2) using key frames as markers for delimiting the clinical events in the pullback, the physicians can focus on the diseased part of the vessel, avoiding the navigation of the whole pullback.

State-of-the-art techniques for the detection of key frames are often based on the detection of changes in color [4–7], brightness [8], scene content and motion [9, 10] or on entropy measures [11] along the sequence. The straight application of state-of-the-art methods to the detection of key frames in IVUS sequences is limited by a series of factors: a continuous rotation effect in the images along the sequence, the repetition of frames due to the heart twisting, the uniform sequence brightness, the movement of the catheter, the vessel pulsation, the presence of speckle noise and the smoothed definition of anatomical regions. All these phenomena hinder the applicability of standard key frames extraction methods. On the other hand, the information related to the vessel morphology, can be considered as intrinsically robust with respect to rigid transformations and also to the presence of artifacts and noise. This motivated us to base the methodology for the detection of keyframes on a semantic description of the vessel rather than on IVUS images intensity analysis. For this purpose, a set of signals providing information on the morphological content of the vessel are used

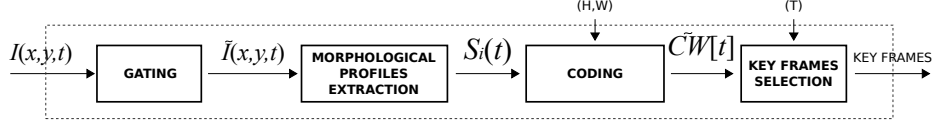


Fig. 2. Schematic representation of the key frames detection technique.

as input for the key frames detection technique, with the aim of detecting the positions in the pullback where a significative change in the morphology occurs. Examples of these signals are the *lumen area*, the *vessel area*, the amount and the composition of *atherosclerotic plaques* along the pullback.

2 Methodology

The flowchart of the proposed technique for the automatic detection of key frames is depicted in Figure 2. The method requires a sequence of IVUS images as well as a set of parameters. The output is the set of key frames, representatives of the main morphological changes of the pullback. In next sections, each block of the flowchart is described in details.

Gating It is known that the acquisition of an IVUS sequence is affected by catheter longitudinal swinging, induced by the cyclic heart twisting [12]. As a consequence, each pair of frames along the sequence suffers a roto-translation; furthermore, the same position of the vessel is imaged several times. In order to extract key frames that correspond to real morphological changes in the vessel, these phenomena must be avoided. As a first step of the keyframes extraction it is then necessary to apply a *gating* procedure to the input sequence. For this purpose, the recently presented image-gating technique based on motion blur analysis is used [12]. The gating process selects a set of positions $\mathcal{T} = \{\tau_1, \dots, \tau_N\}$ of the pullback where the motion blur is minimum, defined as *gated frames*. Given the input sequence $I(x, y, t)$, the gated sequence can be expressed as $\tilde{I}(x, y, t) = I(x, y, t)|_{t \in \mathcal{T}}$. Thus, the detection of the key frames is based on the analysis of the frames at positions $\tau_i \in \mathcal{T}$. As a consequence, the set of selected key frames represents a subset of the gated frames. By basing the analysis of keyframes on gated frames, the following advantages are achieved: (1) the presence of repetitive frames in the sequence is avoided. Furthermore, (2) since in the gated positions the amount of artifacts due to motion is minimum, a more reliable analysis of both RF data and texture is achieved [13]. Finally (3) the selected frames belong to the same cardiac phase leading to a consistent representation of the vessel morphology.

Morphological profiles extraction The detection of key frames is based on the analysis of vessel morphological profiles. Each profile consists in a signal that describes the value of a morphological measurement along the pullback. It is worth to note that the proposed framework is completely independent from the technique used to extract the morphological profiles, which can be obtained by means of any automatic or semi-automatic method, or even manually. In order to provide a basic description of the vessel morphology, we consider the following measurements $\tilde{S}_i[t]$, extracted from the sequence $\tilde{I}(x, y, t)$: *lumen area*, *vessel*

area, *fibrotic*, *lipidic* and *calcified* plaque area (see Fig. 3(c)). Then, with the aim to relate $\hat{S}_i[t]$ with the real input sequence $I(x, y, t)$, the full profile $S_i(t)$, at the same instant of the cardiac cycle, is reconstructed by interpolating the gated samples with a *spline*. A number of samples equal to the number of frames of the pullback is then obtained.

Coding Each morphological profile is then normalized and discretized. For this purpose, the Symbolic Aggregate approXimation (SAX) is used [14]. SAX is a technique that quantize a signal with null mean value and unitary standard deviation into a code of length H , obtained by windowing the input signal. The parameter H represents the resolution of the entire framework: for this reason, its value is set to the mean inter-frames distance of the gated frames. Consequently, we are implicitly assuming that no significative morphological changes in the vessel can occur in less than 1 *mm*: from the clinical point of view, this assumption is correct. Each element of the code can assume a value between zero and $W - 1$, where W is an input parameter for the SAX algorithm, representing the number of values used in the coding process. SAX assures that each value used in the signal coding is equiprobable. Information of both the global statistic of the signal and the local statistic of the windowed samples is considered in order to assign the proper value to the code. This guarantees that a morphological change is considered relevant only if it is significative with respect to the global signal trend. For each input profile $S_i(t)$, a code $C_i[t]$ is obtained. By joining the five computed profiles, a set of H *codewords* $CW[t]$ composed by words of five *letters* is obtained (see Figure 3(b)).

Keyframes selection The key frames are finally detected by processing the code-words $CW[t]$ obtained by gating $CW[t]$ (see Figure 3). Each word of $CW[t]$ is considered as a *descriptor* of the artery, since it encodes the morphological content for a particular position of the vessel. A distance measure can be applied to the discretized profile with the aim of detecting significative changes in the original profile. The set of key frames is defined as follows: the first frame f_0 of the sequence is considered as the first key frame and its word $CW[\tau_0]$ is stored, where $\tau_0 = 0$ represent the position of the first frame of the sequence. For all the successive words of $CW[t]$, the Euclidean distance with respect to the last key frame is computed as $dist(\widetilde{CW}_{f_n}, \widetilde{CW}_{f_{n-1}})$. If the distance is higher than a threshold T , the frame is selected as key frame and added to the set $Y = \{y_1, \dots, y_K\}$ of key frames positions, otherwise it is discarded. The described methodology requires three input parameters (H, W, T) representing the window size and the number of coding values in SAX, and the threshold for the selection of key frames, respectively. Their values are tuned by minimizing the *morphological event detection error* in a validation dataset manually labeled by experts. In the next section the technique for the estimation of the parameters is described and the method is validated.

3 Validation

Materials A set of 16 in-vivo pullbacks have been acquired by using Galaxy II IVUS equipment (Boston Scientific) and a 40 MHz catheter Atlantis SR Pro

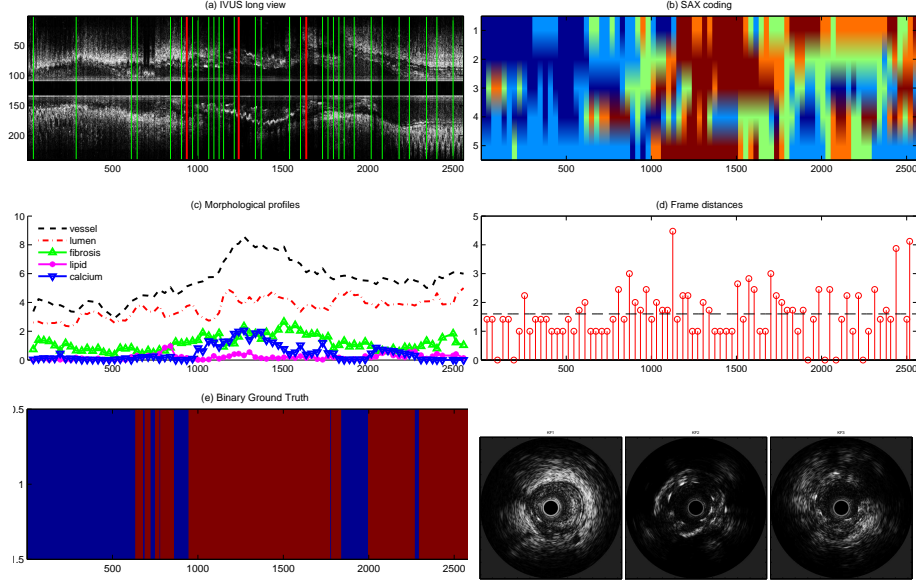


Fig. 3. The visual representation of the framework steps is illustrated. A longitudinal view of the IVUS pullback is provided (a), together with the corresponding morphological profiles (c). The discrete codeword obtained by coding the morphological profiles are represented (b) as well as the distances between consecutive codewords (d). The ground truth (e) represents the healthy (blue) and diseased (red) regions of the vessel. Finally, three examples of detected key frames (KF1-KF3) belonging to the same clinical event are shown.

(Boston Scientific). Radiofrequency data (RF) have been sampled with a 12 bit acquisition card at a sampling rate of 200MHz. The set \mathcal{T}_p of gated position is computed for each pullback by applying the technique described in [12].

For each frame, the five measurement detailed in section 2 have been performed. The lumen area and the vessel area are computed automatically by means of the methods proposed in [13, 15]. The two methods demonstrated to accurately detect lumen and media-adventitia border with a mean error of $0.17 (\pm 0.08) \text{ mm}$ and $0.21 (\pm 0.13) \text{ mm}$, respectively. The lumen and vessel borders define the atherosclerotic plaque area. The plaque characterization presented in [16], based on RF data processing, has been applied to the plaque region, thus computing the amount of fibrotic, lipidic and calcified plaques.

In order to validate the proposed methodology, for each pullback, the initial and the final frame for the most relevant clinical events have been manually labeled by two experts. The events considered are: *stenosis*, presence of *calcifications*, *stent*, presence of *soft plaques*. The intersection between the annotations of the two experts has been used as ground truth. In order to validate the key frames detection, only the position of the beginning/end of each event is required. For this purpose, given the set of labeled events, a binary signal that indicates the presence or absence of an event has been created by grouping all the labeled events as *diseased* regions of the vessel: the remaining parts of the vessel have

been considered as the *healthy* event (see Fig.3(e)). The position in the pullback corresponding to a separation between consecutive regions is defined as “*trigger*” and belongs to the set $X = \{x_1, \dots, x_L\}$. The set of triggers corresponds to manual annotation and is solely required to validate the methodology, since it is not part of the proposed method.

Parameters tuning The estimation of the method parameters (W, T, H) is obtained by optimizing a functional Φ that depends on the key frames detection quality. In order to define the most suited functional Φ for the specific problem, we assume as a necessary condition the presence of at least one key frame in correspondence of a small neighborhood Δ of each trigger position. This condition assures that the proposed methodology is able to detect at least all the clinical event transitions by accepting a distance Δ from the event transition. If fulfilled, the optimal vector parameters $\mathcal{W} = [W_{opt}, T_{opt}, H_{opt}]$ is computed by minimizing the functional $\Phi(\mathcal{W}) = \frac{|Y|}{|X|}$, subject to $|Y \cap X_\Delta| = |X|$ (\mathcal{A}), where $X_\Delta = \{x_1 - \Delta \leq x \leq x_1 + \Delta, \dots, x_L - \Delta \leq x \leq x_L + \Delta\}$ is the set of neighborhood of length 2Δ around each trigger position. During the tuning process, H has been set to the value proposed in section 2 ($H = 30$). The Δ value has been also considered in the tuning. Hence, $\mathcal{W} = [W, T, \Delta]$.

The tuning of the parameters has been performed according to the Leave-One-Patient-Out (LOPO) technique. Given the number of pullbacks, the p^{th} patient is discarded and the parameters \mathcal{W}_{opt}^p are computed. This process is repeated for all the patients. The solution of the whole optimization process is the one that contemporaneously accomplish the condition (\mathcal{A}) and that minimizes the average Φ value according to the LOPO technique. The LOPO scheme assures that the parameters computed at each fold act on the p^{th} pullback as it was a completely unknown case. By taking the intersection of the optimal parameters over all the folds, we provide *nominal characteristics* assuming the framework to work on new unknown cases with performances comparable to the ones obtained during the validation process.

The set of parameters have been estimated by performing an exhaustive search: Δ ranges from 0 to 5, which corresponds to the minimum event duration in the ground truth (5 mm); W ranges from 3 to 15 and T ranges from 1 to d_{max} , where d_{max} is the distance between a pair of codeword which values are set to W_{min} and W_{max} respectively. The Euclidean distance is used as metric between codewords, hence the value $T > 1$ is set. In this way we avoid the detection of key frames where only one profile exhibits a significative change. The values of the optimal parameters obtained by the tuning process is $\mathcal{W}_{opt} = [W_{opt}, T_{opt}, \Delta_{opt}] = [5, 1.6, 4]$.

Performance parameters In order to asses the quality of the proposed key frames detection method, the following performance parameters are considered: *Compression Ratio* (CR), defined as the ratio between the number of selected key frames and the number of frames in the pullback $|I|$: $CR = |X|/|I|$, where the operator $|\cdot|$ indicates the cardinality of a set; *Inter Frames Distance* (IFD), defined as the mean distance between detected key frames: $IFD = \frac{1}{|Y|-1} \sum_{i=2}^{|Y|} |y_i - y_{i-1}|$;

	<i>kFrames</i>	<i>Gating</i>	<i>Ground Truth</i>
<i>Compression Ratio (CR)</i>	0.016 ± 0.004	0.045 ± 0.022	0.004 ± 0.001
<i>Inter key frames distance (IFD)</i>	61.762 ± 48.460	22.0634 ± 13.525	215.484 ± 270.664
<i>Redundancy (Φ)</i>	4.890 ± 2.750	12.603 ± 6.393	1

Table 1. Quantitative results (MEAN \pm STD) for the defined scores computed with the key frames computed with the presented methodology (*kFrames*), gated frames (*Gating*) and triggers (*Ground Truth*).

Redundancy (Φ), defined as the ratio between the number of selected key frames and the number of triggers: $R_d = \frac{|Y|}{|X|}$. Both CR and IFD express the capability of the framework to compress the amount of input information.

The automatic key frames extraction technique has been applied to each pullback obtaining three sets of frame indexes: *kFrames* (*Y*), *Gating* (*T*) and *Ground Truth* (*I*). Table 1 reports the scores values computed by using the detected key frames against the set of gated frames and the set of triggers, respectively.

4 Discussion

The CR achieved with *kFrames* is closer to the Ground Truth than the gated set, showing a high summarization capability of our approach. The same trend is exhibited by the IFD. However, it is worth to note that both a low CR and a high IFD do not necessarily correspond to an optimal key frames detection since they do not depend on the definition of clinical events.

On the other hand, the redundancy score completes the analysis by indicating how the few selected key frames are representative of each clinical event. It is worth to note that the obtained value for redundancy, learned by cross-validation, as described in section 3, assures that no clinical events are lost by accepting a maximum distance Δ between the trigger and the key frame.

To corroborate this assumption, the distance between each detected key frame and the corresponding trigger is computed for all the database: the maximum experimental values is $\Delta = 4$, corresponding to the optimal value obtained by the tuning process. This fact demonstrates that the examples used in the tuning process, through the LOPO technique, are highly representative of the cases variety, and that the tuning process is effective.

Figure 3 illustrates three key frames (KF1-KF3) belonging to the same clinical event. It is interesting to note that, although they belong to the same clinical event, the appearance of the frames is substantially different: a significative change in the content has been correctly detected by the algorithm. This discrepancy between the labels and the detection can be explained as follows: (1) physicians tend to unify local morphological changes under the same semantic event; (2) it is difficult to define a standard *atlas* for the definition and the properties of clinical events in IVUS; (3) several morphological changes can belong to the same clinical event.

5 Conclusions

At the best of our knowledge, this is the first time an automatic method for the detection of key frames in IVUS has been proposed. For this reason, we consider the contribution of this paper as a first original contribution in the field of *video summarization* in IVUS. The presented technique provides a small set of key

frames that are highly representative of the vessel clinical condition. For this reason, they allow the quick and effective inspection of the morphology of the coronary artery during intervention. With respect to the morphological profiles, the proposed methodology is completely modular. Indeed, additional profiles (for example the presence of *bifurcations* or *stent*) can enrich the available set of signals used for the key frame extraction, thus providing a more complete representation of the vessel morphology. Finally, a further supervised analysis on the regions of the pullback delimited by key frames can lead to the automatic detection of the main clinical phenomena in the vessel.

References

1. Wu, S.E., Chang, S.F., H.: Echocardiogram videos: summarization, temporal segmentation and browsing. In: ICIP. (2002)
2. A. Roy and S. Sural and J. Mukherjee and A. K. Majumdar: State-based modeling and object extraction from echocardiogram video. TITB **12**(3) (2008) 366 – 376
3. Syeda-Mahmood, T., Beymer, D., Wang, F., Mahmood, A., Lundstrom, R., Shafee, N., Holve, T.: Automatic Selection of Keyframes from Angiogram Videos. In: ICPR, Ieee (August 2010) 4008–4011
4. H.J.Zhang, J.Y.A.Wang, Y.Altunbasak: Content-based video retrieval and compression. ICIP (1997) 13 – 16
5. B.Gunsel, A.M.Tekalp: Content-based video abstraction. ICIP (1998) 128 – 132
6. Y.Zhuang, Y.Rui, T.S.Huang, S.Mehrotra: Adaptive key-frame extraction using unsupervised clustering. ICIP (1998) 866 – 870
7. A. Hanjalic and H.J. Zhang: An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis. TCSVT (1999) 1280 – 1289
8. A. Hanjalic and R. L. Lagendijk and J. Biemond: A new key-frame allocation method for representing stored video-stream. First International Workshop on Image Databases and Multi Media Search (1996)
9. W.Wolf: Key frame selection by motion analysis. ICASSP (1996) 1228 – 1231
10. P.O.Gresle, T.S.Huang: Gisting of video documents: a key frames selection algorithm using relative activity measures. The 2nd Int. Conf. On Visual Information Systems (1997)
11. Mentzelopoulos, M., Psarrou, A.: Key-Frame Extraction Algorithm using Entropy Difference. Proceedings of the 6th ACM SIGMM (2004) 39 – 45
12. C. Gatta and S. Balocco and F. Ciompi and R. Hemetsberger and O. Rodriguez-Leor and P. Radeva: Real-time gating of IVUS sequences based on motion blur analysis: Method and quantitative validation. In: MICCAI 2010, LNCS 6362/2010. (2010) 59–67
13. S. Balocco and C. Gatta and F. Ciompi and O. Pujol and X. Carrillo and J. Mauri and P. Radeva: Combining Growcut And Temporal Correlation For IVUS Lumen Segmentation. In: IbPRIA, LNCS 6669/2011. (2011) 556–563
14. Lonardi, J.L., Keogh, E., Wei, L., Stefano: Experiencing SAX: a Novel Symbolic Representation of Time Series. Data Mining and Knowledge Discovery **15**(2) (2007) 107 – 144
15. F.Ciompi, O.Pujol, C.Gatta, X.Carrillo, J.Mauri, P.Radeva: A holistic approach for the detection of media-adventitia border in IVUS. MICCAI 2011, LNCS, Volume 6893/2011 **LNCS 6893**/ (2011) 401–408
16. Ciompi, F., Pujol, O., Gatta, C., Leor, O.R., Ferre, J.M., Radeva, P.: Fusing in-vitro and in-vivo intravascular ultrasound data for plaque characterization. IJCI **26** (2010) 763–779