Polyps Recognition Using Fuzzy Trees

Orlando Chuquimia, Andrea Pinna, Xavier Dray, Bertrand Granado

▶ **To cite this version:**

Orlando Chuquimia, Andrea Pinna, Xavier Dray, Bertrand Granado. Polyps Recognition Using Fuzzy Trees. BHI-2017 International Conference on Biomedical and Health Informatics, Feb 2017, Orlando, FL, United States. pp.9-12, 10.1109/BHI.2017.7897192 . hal-01896834

HAL Id: hal-01896834
https://hal.science/hal-01896834

Submitted on 16 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

POLYPS RECOGNITION USING FUZZY TREES

CHUQUIMIA Orlando[1], PINNA Andrea[1], DRAY Xavier[2], BERTRAND Granado[1]

***Abstract*— In this article, we present our work on classifier to realize a Wireless Capsule Endoscopy (WCE) including a Smart Vision Chip (SVC). Our classifier is based on fuzzy tree and forest of fuzzy trees. We obtain a sensitivity of 92.80% and a specificity of 91.26% with a false detection rate of 8.74% on a large database, that we have constructed, composed of 18910 images containing 3895 polyps from 20 different video-colonoscopies.**

I. INTRODUCTION

Colorectal cancer is the second more common type of cancer and the second death cause by cancer around the world with a death rate of 51.03%, with more incidence in developed countries (60.80%) [5]. It is a real public health problem with 694000 deaths in 2012 [5], [3]. For this type of cancer, an early testing policy with a high degree of compliance is a big part of the solution, it is treatable in 90% of the cases if it is detected earlier before polyp become adenocarcinomas [1].

To realize the detection of polyps, colonoscopy is the gold standard but it is an invasive method where an anesthesia, a specialist and a controlled environment are necessary. It has shortcoming and risks as perforations, infections and not allow the visibility of all the regions near the colon.

To overcome these shortcomings, WCE was introduced by Paul Swain and developed since 2000 [7]. This autonomous medical device has opened up a new world of diagnosis for the gastroenterologists. It is a simple pill that the patient swallows and that transmits images of the gastrointestinal tract via a Radio Frequency communication through the body. Characteristics of existing WCE can be read in table I.

TABLE I
EXAMPLES OF WCE FOR SMALL BOWEL, ŒSOPHAGEUS AND COLON

Manufacturer	MedTronics	MedTronics	MedTronics
Operative Region	SB	Œsophageus	Colon
Battery (h)	8	0.5	10
Field of view (°)	156	172	172
Depth of field (mm)	0-30	0-30	0-30
Image Resolution	320x320	256x256	N/A
Image Sampling rate (fps)	2-6	35	4 -35
Image Transmitted (frame)	57600-172800	63000	144000-1260000

As we can notice in table I, WCE presents some limitations:

- a low image resolution compared to standard colonoscopy, image size in WCE is 320x320 compared to those of colonoscope that are higher than 1920x1080;
- a large amount of unecessary transmitted images, WCE transmits more than 50000 images with a majority without pathologic pattern;
- a limited energy budget that limits examination around 8 hours only, that is not sufficient to inspect the total intestinal tract.

To reduce the time to analyse the images produced by WCE, several automatic recognition algorithms to detect polyps were introduced [6]. These algorithms are running on an external computer and contribute to help the physician to realize his diagnosis. We can see the performances of the state of the art of these algorithms in table II. We can notice that the most performant results were obtained with small amounts of images in the database. We can also argue, as it is highlighted in [6], that the true sensitibity is difficult to determine due to the lack of an adequate gold standard.

Although these algorithms help the physician and can limit the errors in human interpretation, they are not sufficient to increase the autonomy of the WCE nor the resolution of the images. To do so, we propose a new paradigm of WCE. Its originality is to integrate a SVC inside the WCE to give it the capability of recognizing a polyp in situ. With this quality, a WCE can transmit only suspicious images in order to reduce their number, increasing its autonomy and providing opportunities to increase the image resolution. In this article, we present our work on the classifier to realize this WCE. It is based on fuzzy trees and allows the recognition of a polyp. We choose it for its high degree of integration in a system on chip [12]. We have evaluated the performance of stand alone fuzzy tree, but also of forest of fuzzy trees and we present the obtained results in this article.

II. POLYPS RECOGNITION

A. *Feature extraction*

The first step to recognize a polyp in an image is to extract the features of this image. We used a method based on co-occurence matrix described in [15]. In this method the classification is done on Region Of Interest (ROI) that correspond to a zone where there is a high suspicion of a presence of a polyp. We have delimited the ROI by hand based on the knowledge of a phyisician. We have extracted 2D texture and luminosity features of each ROI using the Co-occurrence Matrix. We extract 26 descriptors that become our attributes and we use them to train our fuzzy trees.

[1] LIP6, CNRS UMR 7606, Université Pierre et Marie Curie, FRANCE.
[2]APHP, Hôpital Saint-Antoine, Université Pierre et Marie Curie, 75012 Paris, France.

TABLE II
STATE OF THE ART RECOGNITION ALGORITHMS.

Authors	No. Images (polyps)	Source of images	Features 2D	Sensitivity	Specificity	F-measure
Kodogiannis, 2007[10]	140 / (70)	WCE	Shape	97	94	95.48
Karargyris, 2009[8]	50 / (10)	WCE	Shape	100	67.5	80.60
Karargyris & Bourbakis, 2011[9]	100	colonoscopy	Colour, texture, shape	96.2	70.2	81.17
Li & Meng, 2011[11]	1200	colonoscopy	Colour, texture	91.6	NA	NA
Bernal, 2012[2]	300 / (300)	colonoscopy	Shape	89.0	98.0	93.28
Romain et al., 2013[15]	1500 / (300)	colonoscopy	Shape, Texture	91.0	95.2	93.05
David et al., 2013[4]	30540	colonoscopy	Colour, shape	80.0	65.0	71.72
Mamonov & Figueiredo, 2014[13]	18900	colonoscopy	Colour, shape	81.0	90.0	85.26

B. *Fuzzy tree*

Fuzzy tree classifier allows us to manage imprecise data improving the detection robustness and takes in to account lack of knowledge of the attributes or descriptors [14].

This inductive recognition algortihm consists of two parts: learning phase and classification phase. In learning phase fuzzy trees, composed of membership functions, nodes, arcs and leaves, are constructed from a training dataset. In classification phase fuzzy trees classify the ROI belongs to the class which has the highest degree of membership, there is two classes: "1" stand for polyp and "0" stand for not polyp. We use three different methods, described in detail in [14], to realize the classification. These methods are: Classical Modus Ponem, Zadeh operators and Lukasiewicz.

C. *Creation of dataset*

We used a set of 20 video-colonoscopies, 10 of them contain polyps[16]. The dataset $\varepsilon\{A_1, A_2..., A_{26}, C_{0,1}\}$ is a table of 27 columns where each row correspond to a ROI ε_i with 26 attributes ranging from A_1 to A_{26} (Autocorrelation, Contrast, Matlab Correlation, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Matlab Energy, Entropy, Matlab Homogeneity, Homogeneity, Maximum probability, Sum of squares: Variance, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Inf measure of correlation 1, Inf measure of correlation 2, Inv difference normalized, Inv difference moment, mean, Variance, kurtosis and Skewness) and a class which it belongs "0" or "1". To construct the dataset and attach a class to each ROI, we use a ground-truth validated by a gastroenterologist. For the class "1", the ROI is extracted from the mask that was annotated by the gastroenterologist on an image. For the class "0" a randomly ROI selection is made. ROI's dimensions may have diferentes sizes depending on the polyps dimension. The dataset construction is ilustred in figure 1.

From the 20 videos colonscopies we have obtained a dataset composed by 131038 ROI (3856 ROI of class "1"). After we have built 50 learning datasets composed by 4628 dataset's ROI where 50% belong to class "1". All the learning datasets are built with the same set of class "1" (corresponding to 60% of class "1"), for the class "0" a random selection is made from the dataset (corresponding to 60% of class "0"). A test dataset of 6082 dataset's ROI was built and it is composed by 1542 ROI of class "1" (the remaining 40% of class "1").

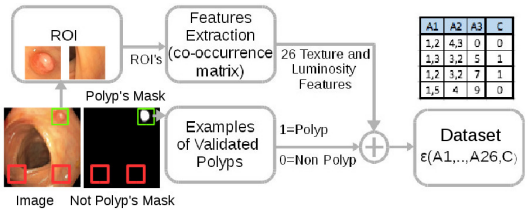

Fig. 1. Proposed diagram for the dataset creation.

D. *Learning phase*

To construct a fuzzy tree $\Phi(\varepsilon_{\{A_i,C_k\}}, H, P, T)$ we use for each training dataset $\varepsilon\{A_1, A_2, ..., A_{26}, C_{0,1}\}$ a discrimination method H to discriminate which attributes are more important. We use a partition strategy P that gives the notion of how to divide the dataset of examples into two groups, and a stop criterion T that indicates when the training dataset is in its minimum expression indicating that the construction is completed. The construction is described in [14] and it is composed by the following steps:

1) selection of the most important attribute with a measure of discrimination H. Measures used for imprecise data, or fuzzy data, are classified into three groups:
   - measures using Shannon entropy as reference,
   - measures using entropy star as reference,
   - measures using another calculation that entropy as a reference.
2) with the criterion P, we made the partition of the training dataset into two sub-sets. Before, we need to aggregate and discretize each attribute, that means calculating the membership functions and a break point of each attribute A_i in each class C_k for all the samples.
3) with each sub-set, we choose again the most important attribute using H and P until each sub-set training are in the form described for the stop criterion T.

In figure 2 this algorithm is illustrated. To construct a fuzzy tree from a training dataset we use the following criterions:

- a star entropy and Shannon entropy H modeled with entropy model proposed in [14].
- a fuzzy attributes aggregation called Generalized Fuzzy Partition using Mathematical Morphology (FPMM*), which is based on mathematical morphology and create

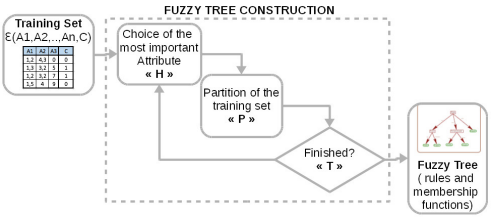

Fig. 2. Fuzzy trees algorithm construction.

erosion and dilation transductors.

- a classic partition P alpha-couple ($\alpha = 1/2$) dividing the learning dataset into two sub-sets.
- a stop criterion T when there are n or less examples in a sub-set, in our case $n = 5$.

At the end of the learning phase, we obtain a fuzzy tree composed by J membership functions $\mu_{j(x)}$ for each attribut and M rules: nodes corresponding to J attributs $A_{m(j)}$ for each rule, arcs corresponding to J break points $v_{m(j)}$ for each attribut and leaves corresponding to membership degrees μ_{mC_0} and μ_{mC_1} of each class of the rule.

E. *Classification phase*

To use fuzzy trees Φ to classify a observed objet $\varepsilon_i\{w_1, w_2, ..., w_{26}\}$ we first use the method of generalized Modus Ponem described below, and showed in figure 3.

1) We calculate a similarity degree $Deg(w_{m(j)}, v_{m(j)})$ between the observed value $w_{m(j)}$ and the break point $v_{m(j)}$ of each attribute j of the rule m using the triangular norm $\top$.

$$\forall_{j=1,2...,J} Deg(w_{m(j)}, v_{m(j)}) = \frac{\top(\mu_{j(w_{m(j)})}, \mu_{j(v_{m(j)})})}{\mu_{j(w_{m(j)})}} \quad (1)$$

In our case we use a triangular norm $\top$ equal to the minimum between $\mu_{j(w_{m(j)})}$ and $\mu_{j(v_{m(j)})}$.

2) We calculate a satisfiability degree $Fded_{m(c_k)}$ using all the similarity degrees $Deg(w_{m(j)}, v_{m(j)})$ of the J attributes of the rule m.

$$Fded_{m(1)} = \top_{j=1,..J} Deg(w_{m(j)}, v_{m(j)}) * \mu_{mC_1} \quad (2)$$

$$Fded_{m(0)} = \top_{j=1,..J} Deg(w_{m(j)}, v_{m(j)}) * \mu_{mC_0} \quad (3)$$

In our case the triangular norm $\top$ is equal to the multiplication between all $Deg(w_{m(j)}, v_{m(j)})$.

3) Finaly, we calculate a new membership degree μ_{c_k} using all the satisfiability degrees of the M rules.

$$\mu_{polyp} = \bot_{m=1,2...,M} Fded_{m(1)} \quad (4)$$

$$\mu_{n_polyp} = \bot_{m=1,2...,M} Fded_{m(0)} \quad (5)$$

In our case we use a conorm $\bot$ is equal to the maximum between all $Fded_{m(C_k)}$.

We use also Zadeh that propose conditional probability $P^*(c_k/(v_{m(1)}, v_{m(2)}.., v_{m(j)}))$ to recalculate each satisfiability degree $Fded_{m(c_k)}$.The final method is Lukasiewicz that uses a triangular norm $\top$ equal to the minimum between all $Deg(w_{m(j)}, v_{m(j)})$ to calculate each satisfiability degree $Fded_{m(c_k)}$. In the case of Classical Modus Ponem the similarity degree is only a comparison:

$$Deg(w_{m(j)}, v_{m(j)}) = sign(w_{m(j)} - v_{m(j)}) = \{0, 1\} \quad (6)$$

To mesure the classification rate we calculate a sensitivity

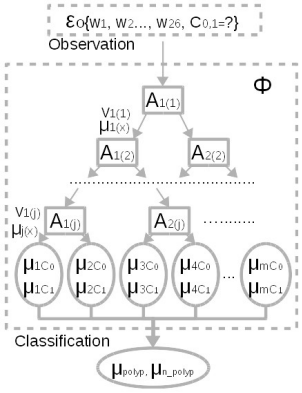

Fig. 3. Proposed diagram for the classification with fuzzy trees.

and specificity:

$$Sensitivity = \frac{true\ positives}{true\ positives + false\ negatives} \quad (7)$$

$$Specificity = \frac{true\ negatives}{true\ negatives + false\ positives} \quad (8)$$

Sensitivity measures the ability to detect polyps and specificity measures the ability to detect the absence of polyp. We test our 50 fuzzy trees constructed on the test dataset for each classification method: Classical Modus Ponem, Zadeh operators and Lukasiewicz. In the figure 4 we show the average and standard deviation of all the sensitivity and specificity results for each classification method. We

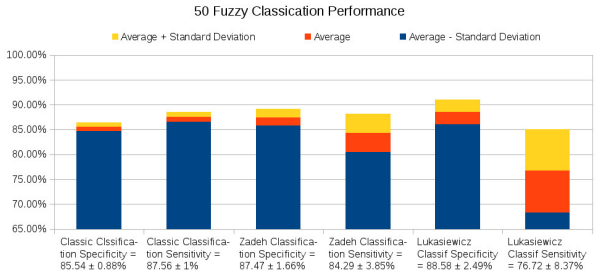

Fig. 4. Fuzzy Trees classification performance.

can notice, based on the analysis of figure 4, that Modus Ponem demonstrated higher sensitivity performance and a stable behavior, Zadeh and Lukasiewicz demonstrated higher specificity performance without a stable behavior.

III. FUZZY FOREST $\Psi\{\Phi_1, \Phi_2..., \Phi_n\}$

In order to enhance the classification rate we construct a fuzzy forest as a combination of n fuzzy trees. We

calculate a new degree of membership using a criterion with the degrees of membership of the n fuzzy trees $\forall_{i=1,2...,n}\{\mu_{polyp(\Phi_i)}, \mu_{n_polyp(\Phi_i)}\}$, this criterion is called conorm "$\perp$":

$$\Upsilon_{polyp} = \perp_{i=1,2...,n}\{\mu_{polyp(\Phi_i)}\} \quad (9)$$

$$\Upsilon_{n_polyp} = \perp_{i=1,2...,n}\{\mu_{n_polyp(\Phi_i)}\} \quad (10)$$

Also we propose to use a weighting factor correspond to the sensitivity and specificity classification of each fuzzy tree:

$$\Upsilon_{polyp} = \perp_{i=1,2...,n}\{\mu_{polyp(\Phi_i)} * Sensitivity_{(\Phi_i)}\} \quad (11)$$

$$\Upsilon_{n_polyp} = \perp_{i=1,2...,n}\{\mu_{n_polyp(\Phi_i)} * Specificity_{(\Phi_i)}\} \quad (12)$$

In our case, for the Classic Modus Ponem we propose a majority vote how conorm criterions, for Zadeh and Lukasiewicz we propose the following criterions of conorm weighted and not weighted: maximum, average and median.

We use the n fuzzy trees created to perform the test with different methods of conorm as it is showed in figure 5.

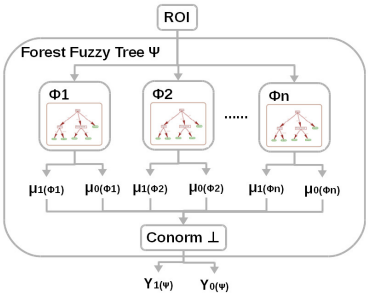

Fig. 5. Fuzzy Forest Diagram.

In the figure 6 we show the results of sensitivity and specificity classification of the fuzzy forest constructed from the 50 fuzzy trees testing it on the test dataset.

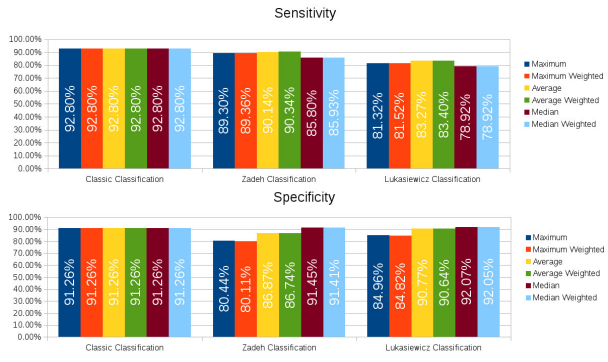

Fig. 6. Fuzzy forest sensitivity and specificity classification.

We can notice, based on the analysis of figure 6, that the sensitivity, specificity and stability performance are increased and for Modus Ponem we obtain a recognition rate higher than 90%.

IV. CONCLUSIONS

In this paper we have presented the fuzzy trees performance for the classification of polyps inside a WCE using texture and luminosity features. The experimentation has shown a sensitivity of 88.56%, specificity of 86.42% and a F-measure of 87.48 with fuzzy trees and a sensitivity up to 92.80%, specificity up to 91.26% and a F-measure of 92,02 with a fuzzy forest. This is at the state of the art compared to other methods presented in table II, and validated in a large database of 18910 images.

REFERENCES

[1] AARONSON, N. K., AHMEDZAI, S., BERGMAN, B., BULLINGER, M., CULL, A., DUEZ, N. J., FILIBERTI, A., FLECHTNER, H., FLEISHMAN, S. B., DE HAES, J. C., ET AL. The european organization for research and treatment of cancer qlq-c30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the national cancer institute 85*, 5 (1993), 365–376.

[2] BERNAL, J., SÁNCHEZ, J., AND VILARINO, F. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition 45*, 9 (2012), 3166–3182.

[3] CORREA, P. Human gastric carcinogenesis: a multistep and multifactorial processfirst american cancer society award lecture on cancer epidemiology and prevention. *Cancer research 52*, 24 (1992), 6735–6740.

[4] DAVID, E., BOIA, R., MALAESCU, A., AND CARNU, M. Automatic colon polyp detection in endoscopic capsule images. In *International Symposium on Signals, Circuits and Systems ISSCS2013* (2013).

[5] FERLAY, J., SOERJOMATARAM, I., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D. M., FORMAN, D., AND BRAY, F. Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International journal of cancer 136*, 5 (2015), E359–E386.

[6] IAKOVIDIS, D. K., AND KOULAOUZIDIS, A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nature Reviews Gastroenterology & Hepatology 12*, 3 (2015), 172–186.

[7] IDDAN, G., MERON, G., GLUKHOVSKY, A., AND SWAIN, P. Wireless capsule endoscopy. *Nature 405* (2000), 417.

[8] KARARGYRIS, A., AND BOURBAKIS, N. Identification of polyps in wireless capsule endoscopy videos using log gabor filters. In *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH* (2009), IEEE, pp. 143–147.

[9] KARARGYRIS, A., AND BOURBAKIS, N. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. *IEEE transactions on Biomedical Engineering 58*, 10 (2011), 2777–2786.

[10] KODOGIANNIS, V., AND BOULOUGOURA, M. An adaptive neurofuzzy approach for the diagnosis in wireless capsule endoscopy imaging. *International Journal of Information Technology 13*, 1 (2007), 46–56.

[11] LI, B., AND MENG, M. Q.-H. Automatic polyp detection for wireless capsule endoscopy images. *Expert Systems with Applications 39*, 12 (2012), 10952–10958.

[12] LIU, B.-D., AND HUANG, C.-Y. Design and implementation of the tree-based fuzzy logic controller. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 27*, 3 (1997), 475–487.

[13] MAMONOV, A. V., FIGUEIREDO, I. N., FIGUEIREDO, P. N., AND TSAI, Y.-H. R. Automated polyp detection in colon capsule endoscopy. *IEEE transactions on medical imaging 33*, 7 (2014), 1488–1502.

[14] MARSALA, C. *Apprentissage inductif en présence de données imprécises: Construction et utilisation d'arbres de décision flous*. PhD thesis, 1998.

[15] ROMAIN, O., HISTACE, A., SILVA, J., AYOUB, J., GRANADO, B., PINNA, A., DRAY, X., AND MARTEAU, P. Towards a multimodal wireless video capsule for detection of colonic polyps as prevention of colorectal cancer. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on* (2013), IEEE, pp. 1–6.

[16] TAJBAKHSH, N., GURUDU, S. R., AND LIANG, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE transactions on medical imaging 35*, 2 (2016), 630–644.