ANNOTATION OF MEDICAL IMAGES USING THE SURF DESCRIPTOR

Anna Wojnar *

Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology

António M. G. Pinheiro

Remote Sensing Unit
Universidade da Beira Interior

ABSTRACT

This paper describes a method for medical images annotation based on the SURF descriptor and the SVM classifier. For the features extraction a Fast-Hessian detector was used. The feature matching was performed with a SVM with a quadratic kernel. The testing of the developed system was performed using a subset of the IRMA radiographic images. The results provided with the SURF descriptor are compared with the ones obtained using the SIFT descriptor with SVM classification. Applying the SURF descriptor resulted in improved classification of lung images with an accuracy over 96%. The research shows that the SURF is a potentially strong tool to be applied in the field of medical image annotation. Together with the SVM classification it may construct an efficient system for automatic medical image retrieval and annotation.

Index Terms— Image Annotation, Image Description, SURF, SVM

1. INTRODUCTION

Medical images are a crucial source of knowledge for therapeutic purposes. Due to the advances in digital technology there has been a large growth in the amount of digital images stored in recent years. Therefore it is crucial to develop efficient systems for medical image retrieval and annotation based on visual content. Some success of CBIR (Content-based Image Retrieval) in a narrow field of medical applications has been reported [1]. However, there has not still been any breakthrough with respect to large varied databases.

Despite the huge number of description techniques recently researched and published, only few represent reliable solutions, suitable for semantic annotation of images. Bay et al.[2] proposed a scale-invariant and resilient to image rotation Speeded-Up Robust Features (SURF) descriptor. It is based on the Fast-Hessian detector and makes efficient use of integral images. Lowe [3] proposed the Scale-Invariant Feature Transform (SIFT) descriptor invariant to image transformations based on the DoG detector.

*Work developed during a trainship at University of Beira Interior, Portugal.

In this paper, the potential application of the SURF descriptor for medical images annotation is tested. The images descriptors are classified using Support Vector Machines (SVM). SURF was reported for its successful application in video stabilisation with KD-tree algorithm for matching [4], as well as in an image search and object recognition used on mobile phone with the Nearest Neighbour matching method [5]. SVM has been successfully used in content based image retrieval. It was attempted to use multi-class SVM for automatic medical image annotation and retrieval based on global image features contents [6].

This paper is organized as follows. Feature detection with the SURF detector and descriptor is briefly described in Sec. 2. The pattern of descriptors selection for the classifier is presented in Sec. 3. In Sec. 4 experimental results are discussed. The contribution is summarized in Sec. 5. SIFT descriptor will not be described as it is a well-known and widely used tool [3].

2. THE SURF DETECTOR AND DESCRIPTOR

2.1. Detection

In order to detect interest points the Hessian matrix is approximated using a set of box-type filters (Fig. 1). These 9 x 9 box filters approximate second order Gaussian derivatives in y- and xy-direction with $\sigma = 1.2$ and represent the lowest scale for computing blob response maps. These derivatives are referred as D_{yy} and D_{xy} respectively. Due to the use of integral images [7] the derivatives can be evaluated at a very low computational cost.

For computational efficiency the weights applied to the rectangular regions are kept simple, but they are weight, yielding:

$$det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \qquad (1)$$

However, it was shown [2] that the weight does not have a significant impact on the results and suggested to keep it constant and equal to 0.9.

The output of the 9 x 9 box filters shown in Fig. 1 is the initial scale level of the scale-space. Subsequent layers are

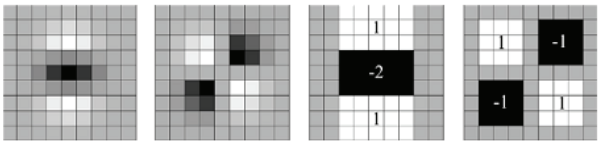


Fig. 1. Left to right: the Gaussian second order partial derivative in y- and xy-direction respectively; approximation of the second order Gaussian partial derivatives in y- and xy-directions [2].

obtained by upscaling the filters. The masks of the filters are increased gradually. The scale-space is divided into octaves. An octave represents a series of filter response maps obtained by convolving the same input image with a filter of increasing size. Each octave is subdivided into a constant number of scale levels. The octaves are overlapping in order to cover all possible scales seamlessly.

For localisation of interest points in the image and over scales 3D non-maximum suppression is applied, both spatially and over the neighbouring scales.

2.2. Description

The descriptor describes the distribution of the intensity content within the interest point neighbourhood. It is built on the distribution of first order Haar wavelet responses in x- and y-direction. Description consists of orientation assignment and extraction of the descriptor.

In the first step a reproducible orientation for the interest points is identified. For this purpose Haar wavelet responses in x- and y-direction within a circular neighbourhood around the interest point are calculated. The dominant orientation of the Haar wavelet responses is estimated by calculating the sum of all of the responses (vertical and horizontal ones) within a sliding orientation window of size $\pi/3$. The longest vector of the summed responses defines the orientation of the interest point.

For the extraction of the descriptor, a square region centred around the interest point and oriented along the selected orientation is created. This region is further split up into 16 square sub-regions for which values of sums of Haar wavelet responses $\sum d_x$ and $\sum d_y$ and their absolute values ($\sum |d_x|$ and $\sum |d_y|$) are computed. They form a feature vector of length 64 defined by eq. 2.

$$v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \tag{2}$$

3. ANNOTATION AND CLASSIFICATION

The developed system consists of two parts: the SURF descriptors extraction and building of the SVM classifier. Descriptors extraction was performed with the use of D.Kroon openSURF implementation based on Chris Evans OpenSurf c# [8].[1] Feature vectors of length 64 were extracted. The descriptors for the set of training images were selected and the classifier was built. Finally, for the purpose of comparison, the same procedure was performed with D. Lowe implementation of SIFT algorithm.[2] In case of SIFT descriptors feature vectors of length 128 were investigated.

The testing of the developed system was performed using the radiography images from IRMA database.[3] In this paper the annotation of images with lungs is reported. By the term lungs it is understood two-side lungs images that were taken either from front or back of the patient. The aim of the classification was to classify images into lungs and non-lungs. Images displaying one-side lung were classified separately. The image database used in the experiment contained randomly chosen 4241 test images (594 lungs and 3647 non-lungs images). Moreover, another 65 train images (21 lungs for positive training set and 44 non-lungs images for negative training set) were also selected from the database. Negative training set for classifier was based on 3 abdomin, 1 breast, 3 fingers, 5 hand, 2 patella, 2 shoulder, 4 skull, 6 spine, 3 lung images taken from the side and 15 unspecified images. Images of lungs taken from the side were considered a different feature than front and back lungs as it was impossible to find good matching points between the two.

Two classifiers were tested: k-Nearest Neighbours and SVM with various kernel functions, including linear, quadratic, Gaussian Radial Basis Function and the third order polynomial kernels. The SVM with the quadratic kernel function turned out to perform the best for this application. A crucial definition for the efficient work of the classifier was the appropriate selection of descriptors for the training set. For that, test images were matched with each other and concrete descriptors were chosen based on visual assessment and distance value calculated for the matched feature vectors. The distance between vectors was calculated according to eq. 3.

$$distance = \sum_i (D(i)_2 - D(i)_1)^2 \tag{3}$$

In order to obtain the positive training set, lung images were matched between themselves and the most representatives of the class were selected. In the case of the negative training set, images were matched with the other images displaying the same feature and also with lung images. If an image has good similarity results with its own cathegory and bad similarity results with lung images, then was selected as a good negative element. This allowed a good determination of the interest points that are representatives of the given features only. It was important not to choose descriptors that could be similar with the two features. It was determined that the unique interest points in lung images are located in the space

[1] www.mathworks.com/matlabcentral/fileexchange/28300
[2] www.cs.ubc.ca/ lowe/keypoints/
[3] www.ganymed.imib.rwth-aachen.de/irma/

between ribs, close to its cross. So, descriptors on these locations were chosen for the positive training set of the classifier. Images for the negative training set were chosen in order to include various body parts representatives. From those images the selected descriptors locations were characteristics of a given group that were not present in lung images. Finally, using the SURF descriptor a negative training set contained 215 descriptors derived from 44 images and a positive training set containing 117 descriptors chosen from 21 images was selected.

In each classification all descriptors for each image were classified. The decision whether an image belongs to lung class or non-lung class was determined by the amount of descriptors belonging to the class lung. The established threshold value for the SVM classifier was 0.2 (Fig. 2), which means that if images possess at least 20% of descriptors belonging to the class lung they are classified as lung images. For the kNN classifier the estimated threshold value turned out to be 0.3.

It was hard to decide how to classify images of one-side lungs. In order to determine if they possess the lung features or should rather be classified as some other feature, they were treated as a separate class of images. 22 one-side lung images were separated.

In case of the SIFT algorithm the same procedure was taken for the selection of the negative training set descriptors. However, it was impossible to find good matches between lungs images and therefore the creation of a good positive training set was unfeasible. Finally, a selection of 25 descriptors for the positive training set and 64 for the negative training set was made. Such small number of descriptors arises from the fact that the error of classification was incomparably larger than in case of the application of the SURF descriptor with even smaller number of descriptors.

4. RESULTS AND DISCUSSION

Table 1 shows the results of classification obtained for the classification of two-side lung images using SURF and SIFT descriptors.

Using the SURF description for the extraction of features from radiography images and quadratic SVM classification, an error of 3.4% was achieved. We were expecting that the SURF algorithm would have a good performance with this kind of images. However, with such a small training set, the results are surprisingly good. For the kNN classifier the overall error is almost the double, but the amount of false negative results is unacceptably high. The SIFT algorithm did not perform well. The overall classification error with this descriptor was over 50% with a huge amount of false positive results.

One-side lungs images classification performance showed a poor efficiency. About half of the images were misclassified. In one-side lung images there are well-defined ribs with their crossings, which is the desired feature. However, these

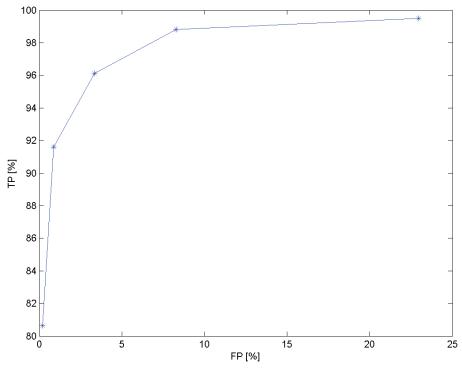


Fig. 2. ROC for the SVM classifier obtained for different threshold values (0.1, 0.15, 0.2, 0.25, 0.3).

	SURF		SIFT
	KNN	SVM	SVM
Positives	594	594	594
FN	107	23	69
% FN	18.0	3.9	11.6
Negatives	3647	3647	3647
FP	147	122	2199
% FP	4.0	3.3	60.3
all	4241	4241	4241
all misclassified	254	145	2268
% error	6.0	3.4	53.5

Table 1. Classification Results for kNN with k=5 and quadratic SVM

images also contain a very well-defined and detailed spine, which is usually not present in two-side lungs regardless the position the picture was taken. It is usually at least a little blurred (Fig.3). Hence, there is a need for a specific classifier trained only for this kind of images.

Most of the misclassified images were low quality images. In case of false positive results they were quite often lung images taken from the side or images with a small part of the lung only. For images containing small parts of lungs (Fig. 4) the situation might be similar to the one-side lung images. It is necessary to describe these features in a more complex way. Fig. 5 illustrates examples of some bad quality images classified either as false positives or false negatives.

5. CONCLUSIONS

It has been shown that the SURF descriptor is a tool with a huge potential in applications in the field of annotation of medical images. It was possible to achieve over 96% accuracy applying the SURF algorithm together with the

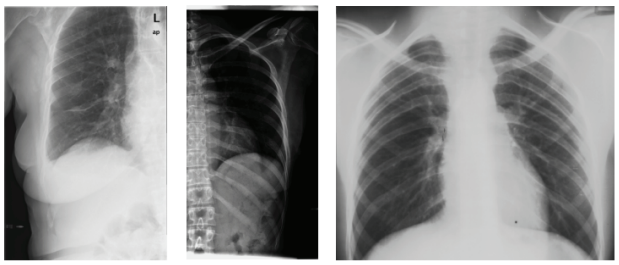


Fig. 3. Left to right: properly classified one-side lung, missclassified one-side lung with well-defined spine, two-side lungs.

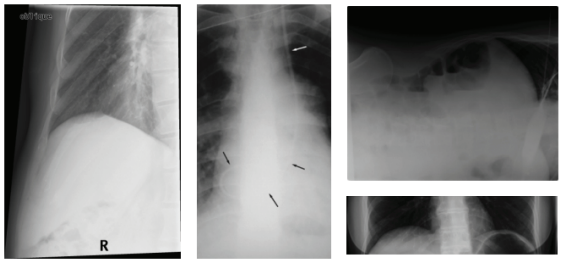


Fig. 4. Examples of images containing parts of lungs.

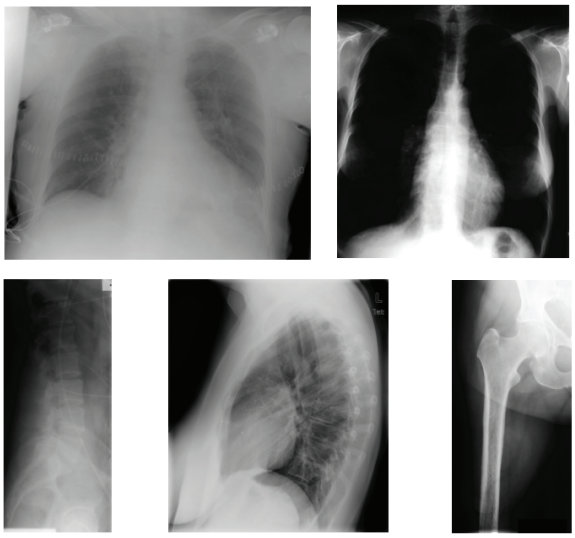


Fig. 5. Examples of misclassified images.

quadratic SVM classifier using a very small training set of images. Moreover, it was concluded that SIFT descriptor is not suitable for annotation of radiographic images.

Looking to the misclassified images examples it was concluded that images need to have a certain level of quality for a reliable annotation. They can not be obtained with overexposure to light or on the other hand, they can not be too dark. These kind of images result in a small amount of interest points, leading to a small classification rate. The definition of models for the image quality measurement in the sense required by SURF is important for any real system and can be a future research direction. Images that do not provide reliable annotation should be marked and manually annotated.

It is also essential to develop a strategy for dealing with images containing two features. As described in the end of Sec. 4 these kinds of images are also a source of classification errors. Moreover, implementing multiple classifiers would improve the annotation performance.

To summarize the SURF descriptor is a potentially strong tool to be applied in the field of medical image annotation. Together with a SVM classifier it may construct an efficient and reliable system for automatic medical image annotation and retrieval.

6. REFERENCES

[1] Henning Mller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications–clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1 – 23, 2004.

[2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, 110:346–359, June 2008.

[3] David G. Lowe. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the International Conference on Computer Vision - Volume 2*, ICCV '99, Washington, DC, USA, 1999.

[4] Keng-Yen Huang, Yi-Min Tsai, Chih-Chung Tsai, and Liang-Gee Chen. Video stabilization for vehicular applications using SURF-like descriptor and KD-tree. In *ICIP*, pages 3517–3520, 2010.

[5] Wei-Chao Chen, Yingen Xiong, Jiang Gao, Natasha Gelfand, and Radek Grzeszczuk. Efficient Extraction of Robust Image Features on Mobile Devices. In *ISMAR '07*, pages 1–2, Washington, DC, USA, 2007.

[6] Md. Mahmudur Rahman, Bipin C. Desai, and Prabir Bhattacharya. Supervised machine learning based medical image annotation and retrieval. In *In: Working Notes of the CLEF Workshop 2005*, 2005.

[7] Paul Viola and Michael Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition*, 1:511, 2001.

[8] Christopher Evans. Notes on the OpenSURF Library. Technical Report CSTR-09-001, University of Bristol, January 2009.