

University of Freiburg at ImageCLEF06 - Radiograph Annotation Using Local Relational Features

Lokesh Setia, Alexandra Teynor, Alaa Halawani and Hans Burkhardt
Chair of Pattern Recognition and Image Processing (LMB),
Albert-Ludwigs-University Freiburg, Germany
{setia, teynor, halawani, burkhardt} @informatik.uni-freiburg.de

Abstract

This paper provides details of the experiments performed by the LMB group at the University of Freiburg, Germany, for the medical automatic annotation task in the ImageCLEF 2006. We use local features calculated around interest points, which have recently received excellent results for various image recognition and classification tasks. We propose the use of relational features, which are highly robust to illumination changes, and thus quite suitable for X-Ray images. Results with various feature and classifier settings are reported. A significant improvement in results is seen when the relative positions of the interest points are also taken into account during matching. For the given test set, our best run had a classification error rate of 16.7 %, just 0.5 % higher than the best overall submission, and therewith was ranked second in the medical automatic annotation task at the ImageCLEF 2006.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; I.5 [Pattern Recognition]: I.5.3 Clustering; I.5.4 Applications;

General Terms

Algorithms, Experimentation

Keywords

Classification, Radiograph, Image Annotation

1 Introduction

Growing image collections of various kinds have led to the need for sophisticated image analysis algorithms to aid in classification and recognition. This is especially true for medical image collections, where the cost for manual classification of images collected through daily clinical routine can be overwhelming. Automatic classification of these images would spare the skilled human labour from this repetitive, error-prone task. As an example, a study carried out by the University Hospital in Aachen [1] revealed over 15% human errors in assignment of a specific tag, for X-ray images taken during normal clinical routine. Unless corrected, these images cannot be



Figure 1: Intra-class variability within the class annotated as “*x-ray, plain radiography, coronal, upper extremity (arm), hand, musculoskeletal system*”.

found again with keyword search alone. Our group (LMB) participated in the ImageCLEF 2006 medical automatic annotation task. In this report, we describe our experiments and submitted runs using a local version of relational features which are very robust to illumination changes, and with different kinds of accumulators to aid in the matching process.

2 Database

The database for the task was made available by the IRMA group from the University Hospital, Aachen, Germany. It consists of 10,000 fully classified radiographs taken randomly from medical routine. 1,000 radiographs for which classification labels were not available to the participants had to be classified. The aim is to find out how well current techniques can identify image modality, body orientation, body region, and biological system examined based on the images. The results of the classification step can be used for multilingual image annotations as well as for DICOM header corrections. The images are annotated with the complete IRMA code, a multi-axial code for image annotation. However, for the task, we used only a flat classification scheme with a total of 116 classes. Figure 1 shows that a great deal of intra-class variability exists, mostly in illumination changes, small amounts of position and scale differences, and noise. On the other hand, Figure 2 shows that the inter-class distance can be quite tight, as the four frontal chest categories are at least pairwise very similar to each other. The confusion matrix between these four classes indicates that more than 80 % of the images were correctly classified.

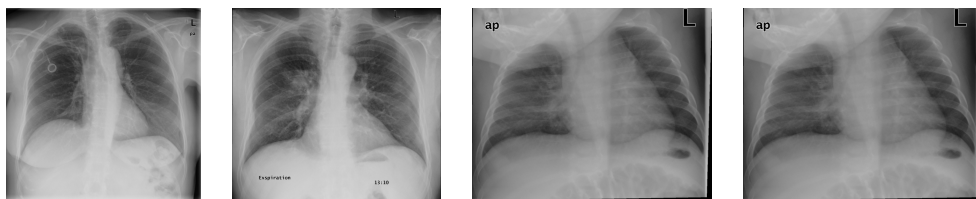


Figure 2: One example image each from class number 108, 109, 110 and 111 respectively. Some classes are hard to distinguish by the untrained eye, but surprisingly perform quite well by automatic methods.

3 Features

3.1 Local Features Generation

We use features inspired from invariant features based on group integration. Let G be the transformation group for the transformations against which the invariance is desired. For the case of radiograph images this would be the group of translations. For rotation, only partial invariance is desired as most radiographs are scanned upright. If an element $g \in G$ acting on the images, the transformed image is denoted by $g\mathbf{M}$. An invariant feature must satisfy $F(g\mathbf{M}) = F(\mathbf{M}), \forall g \in G$. Such invariant features can be constructed by integrating $f(g\mathbf{M})$ over the transformation group G [2].

$$I(\mathbf{M}) = 1/|G| \int_G f(g\mathbf{M})dg$$

Although the above features are invariant, they are not particularly discriminative, due to the fact that the integration runs over the whole image, which contains large insignificant regions. An intuitive extension is to use local features, which have received good results in several image processing tasks. The idea is to first use a so-called interest point detector, to select pixels with high information content in their neighbourhood. The meaning of “high information” differs from application to application. There are, for example, detectors which select points of high gradient magnitude, or corner points (points with multiple gradient orientations). A desirable characteristic expected from an interest-point detector is that it is robust towards the above mentioned transformations. In this work, we use the wavelet-based salient point detector proposed by Loupiaz and Sebe [3].

For each salient point $\mathbf{s} = (x, y)$, we extract relational feature subvector around it, given as:

$$R_k = [\mathfrak{R}(x, y, r_1, r_2, \phi, n)]_k, \quad k = 1, \dots, n$$

Then,

$$R_k = \text{rel}(\mathbf{I}(x_2, y_2) - \mathbf{I}(x_1, y_1)),$$

where

$$(x_1, y_1) = (x + r_1 \cos(k \cdot 2\pi/n), y + r_1 \sin(k \cdot 2\pi/n)),$$

and

$$(x_2, y_2) = (x + r_2 \cos(k \cdot 2\pi/n + \phi), y + r_2 \sin(k \cdot 2\pi/n + \phi))$$

with the `rel` operator defined as:

$$\text{rel}(x) = \begin{cases} 1 & \text{if } x < -\epsilon \\ \frac{\epsilon-x}{2\epsilon} & \text{if } -\epsilon \leq x \leq \epsilon \\ 0 & \text{if } \epsilon < x \end{cases}, \quad (1)$$

The relational functions were first introduced for texture analysis by Schael [4], motivated by the local binary patterns [5]. For each (r_1, r_2, ϕ) combination a local subvector is generated. In this work, we use 3 sets of parameters, $(0, 5, 0)$, $(3, 6, \pi/2)$ and $(2, 3, \pi)$, each with $n = 12$ to capture local information at different scales and orientations. The subvectors are concatenated to give a local feature vector for each salient point. The local feature vectors of all images are taken together and clustered using the k-means clustering algorithm. The number of clusters would be denoted here by N_c , and is determined empirically.

3.2 Building Global Feature Vector

An equally important step is that of estimating similarity between two image objects, once local features have been calculated for both of them. For the case that one expects to find for a given test image, a training image of basically the same object, correspondence-based matching algorithms are called for, i.e. for each test local feature vector, the best training vector and its location is recalled (e.g. through nearest neighbour). The location information can be put to use by doing a consistency check if the constellation of salient points build a valid model for the given class.

However, the large intra-class variability displayed by the radiograph images suggest that a global feature based matching would perform better, as small number of non-correspondencies would not penalize the matching process heavily. We build the following accumulator features from the cluster index image $I(\mathbf{s})$, i.e. an image which contains for each salient point, its assigned cluster index only.

3.2.1 All-Invariant Accumulator

The simplest histogram that can be built from a cluster index image is a 1-D accumulator counting the number of times a cluster occurred for a given image. All spatial information regarding the interest point is lost in the process. The feature can be mathematically written as:

$$\mathbf{F}(c) = \#\{\mathbf{s} \mid (I(\mathbf{s}) = c)\}, \quad c = 1, \dots, N_c$$

In our experiments, the crossvalidation results were never better than 60 %. It is clear that critical spatial information is being lost in the process of histogram construction.

3.2.2 Rotation Invariant Accumulator

We first incorporate spatial information, by counting pairs of salient points lying within a certain distance range, and possessing particular cluster indices, i.e. a 3-D accumulator defined by:

$$\mathbf{F}(c_1, c_2, d) = \#\{(\mathbf{s}_1, \mathbf{s}_2) \mid (I(\mathbf{s}_1) = c_1) \wedge (I(\mathbf{s}_2) = c_2) \\ \wedge (D_d < \|\mathbf{s}_1 - \mathbf{s}_2\|_2 < D_{d+1})\}$$

where the extra dimension d runs from 1 to N_d (number of distance bins).

The crossvalidation results improve to about 68 %, but it should be noted that this accumulator is rotation invariant (depends only on distance between salient points), while the images are upright. Incorporating unnecessary invariance leads to a loss of discriminative performance. Especially in this task of radiograph classification it can be seen that mirroring the position of interest points leads to a completely another class, for example, a left hand becomes a right hand, and so on. Thus we incorporate orientation information in the next section.

3.2.3 Orientation Variant Accumulator

We discussed this method in detail in [6]. The following cooccurrence matrix captures the statistical properties of the joint distribution of cluster membership indices which were derived from local features.

$$\mathbf{F}(c_1, c_2, d, a) = \#\{(\mathbf{s}_1, \mathbf{s}_2) \mid (I(\mathbf{s}_1) = c_1) \wedge (I(\mathbf{s}_2) = c_2) \\ \wedge (D_d < \|\mathbf{s}_1 - \mathbf{s}_2\|_2 < D_{d+1}) \\ \wedge (A_a < \angle(\mathbf{s}_1, \mathbf{s}_2) < A_{a+1})\}$$

with the new dimension a running from 1 to N_a (number of angle bins).

The size of the matrix is $N_c \times N_c \times N_d \times N_a$. For 20 clusters, 10 distance bins and 4 angle bins, this leads to a feature vector of size 16000. The classification is done with the help of a multi-class

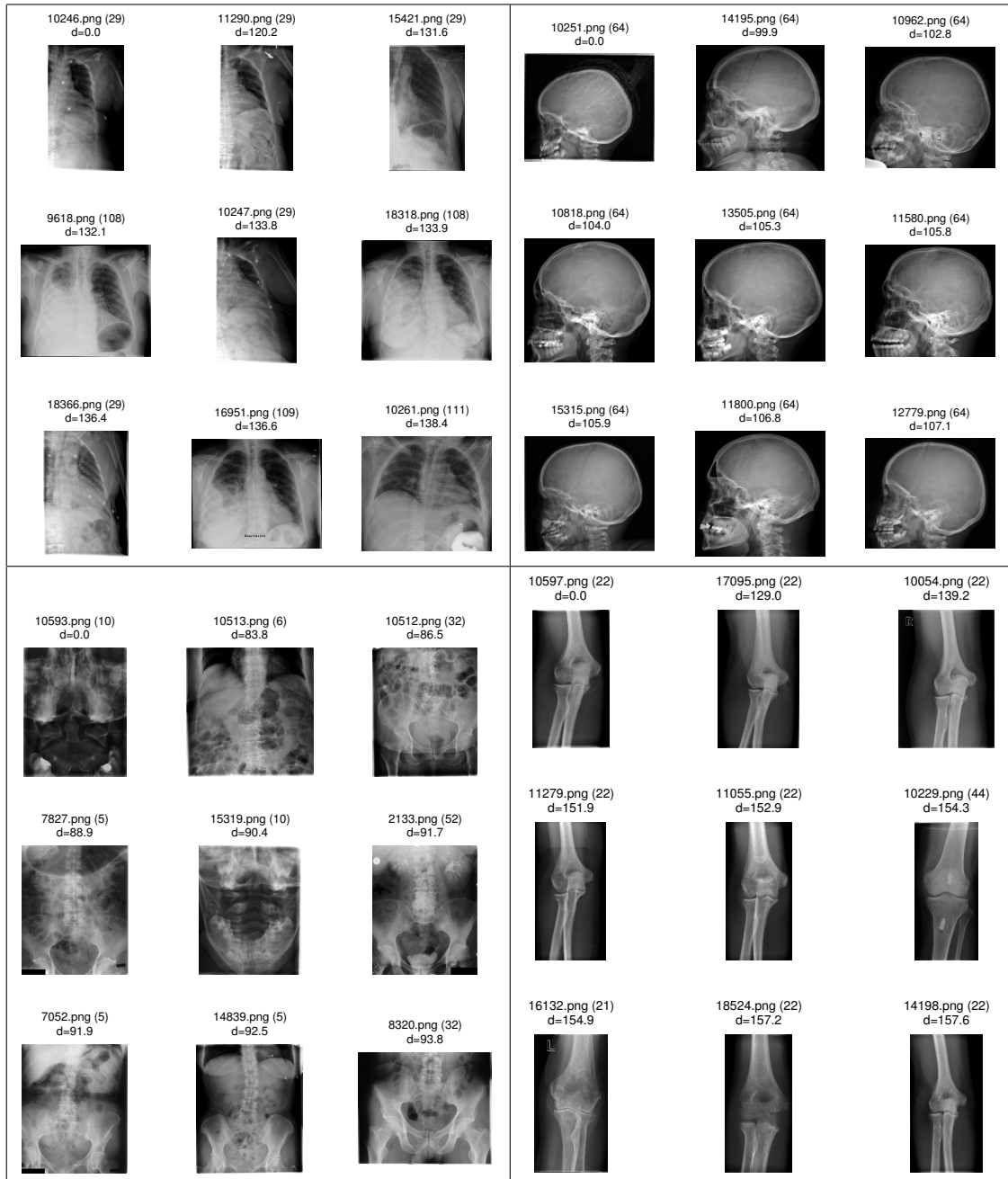


Figure 3: Top 8 nearest neighbour results for different query radiographs. The top left image in each box is the query image, results follow from left to right, then top to bottom. The image caption contains the image name and label in (brackets). Finally the distance according to the L_1 norm is displayed underneath.

SVM, running in a `one-vs-rest` mode, and using a histogram-intersection kernel. The libSVMTL implementation¹ was used for the experiments. Despite the large dimensionality, the speed should be acceptable for most cases. Training the SVM classifier with 9000 examples takes about 2 hours of computation time, the final classification of 1000 examples just under 10 minutes.

3.3 Results and Conclusion

We submitted two runs for the ImageCLEF medical annotation task. Both runs used the accumulator defined in Section 3.2.3 and were equally sized, with $N_c = 20$, $N_d = 10$ and $N_a = 4$. The only difference was in the number of salient points, N_s , selected per image. The first run used $N_s = 1000$, while the second used $N_s = 800$. For the given test set, the runs achieved an error rate of 16.7 % and 17.9 % respectively, which means that our best run was ranked second overall for the task. There might have been some room left for better parameter tuning, but it was not performed due to lack of resources, and because there are quite a few parameters to tune. Thus, instead of a joint optimization, each parameter was only individually selected through empirical methods.

Figure 3 shows the nearest neighbour results (L_1 norm) for four sample query images. It can be seen that mostly the features are very suited for the task. Local features are in general robust against occlusion and partial matching, which is in general advantageous for a recognition task. This is also true for the radiograph annotation task described in this paper, but there are exceptions. For example, the top left box contains a chest radiograph, and the radiographs with complete frontal chest, and partial frontal chest apparently belong to different classes (class number 108 and 29 respectively). It might be beneficial to use region specific extra features, once it is determined that the radiograph is e.g. one that of chest.

References

- [1] M. O. Gueld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann. Quality of DICOM header information for image categorization. In *Proc. SPIE Vol. 4685, p. 280-287, Medical Imaging 2002:*, pages 280–287, May 2002.
- [2] H. Schulz-Mirbach. Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, editors, *17. DAGM - Symposium "Mustererkennung"*, pages 1-14, Bielefeld, 1995. Reihe Informatik aktuell, Springer.
- [3] E. Louprias and N. Sebe. Wavelet-based salient points for image retrieval, 1999.
- [4] M. Schael. *Methoden zur Konstruktion invarianter Merkmale für die Texturanalyse*. PhD thesis, Albert-Ludwigs-Universität, Freiburg, June 2005.
- [5] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. In *Proc. Sixth European Conference on Computer Vision*, pages 404–420, Dublin, Ireland, 2000.
- [6] L. Setia, A. Teynor, A. Halawani and H. Burkhardt. Image Classification using Cluster-Cooccurrence Matrices of Local Relational Features In *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval (MIR 2006)*, Santa Barbara, CA, USA, October 26-27, 2006

¹<http://lmb.informatik.uni-freiburg.de/lmbsoft/libsvmtl/>