Testing a Method for Statistical Image Classification in Image Retrieval

Christoph Rasche, Constantin Vertan
Laboratorul de Analiza si Prelucrarea Imaginilor
Universitatea Politehnica din Bucuresti
Bucuresti 061071, RO
rasche15@gmail.com, cvertan@alpha.imag.pub.ro

Abstract

We continued to test our image classification methodology in the photo-annotation task of the ImageCLEF competition [Nowak et al., 2011] using a visual-only approach performing automated labeling - however with little algorithmic improvement as compared to last year. Our labeling process consisted of three phases: 1) feature extraction using color and structural description; 2) classification using Linear Discriminant (LD), which provided the confidence (scalar) values; 3) post-processing by eliminating labels (setting binary values to 0) on the testing set thereby exploiting the calculated joint-probabilities for pairs of concepts from the training set. Our conclusions remain the same as last year: our approach provides reasonable, fast image classification.

1 Introduction

Our approach this year is essentially the same as last year and we participated to fine-tune some minor algorithmic details. We employ a massive structural descriptor extraction process, which is based on partitioning contour segments followed by relating the segments to form more complex descriptors. Although this structural description is very promising in essence, we have used it only in a statistical manner so far (because we have not developed the appropriate learning algorithm for the structural classification yet). Even though our statistical classification performance ranks mid-range only, there are some advantages to it over the other, more successful classification approaches. The majority of the other (last year's) approaches use SIFT features for representation and those features appear to be useful for texture identification (see also Torralba et al. 2005 for arguments); matching is done feature-by-feature and is relatively costly. In our approach in contrast, matching is done between image vectors, and is thus resonably fast; but our approach is costly at the feature extraction stage.

The main novelty of the presented approach is the use of a decomposition of structure as introduced in Rasche [Rasche, 2010]. The decomposition output is particularly suited to represent the geometry of contours and the geometry of their relations (pairs or clusters of contours), but it is applied here only in a statstical form for reason of simplicity, together with a color histogramming approach as described in Vertan et al. [Vertan and Boujemaa, 2000b]. This statistical classification has already been shown to be useful for video indexing [Ionescu et al., 2010].

Looking at the provided photo annotations we realized that the spatial size of the annotated object or scene can vary substantially in reference to the image size: an annotation can describe either the image content as a whole and is thus suitable for (semantic) image classification, or it can describe a part of a scene (e.g isolated objects) and is thus rather suited for object-detection systems. A clear distinction between whole or part annotation is difficult of course. A typical recognition system is specialized for one process, either image classification or object detection. Our methodology is geared toward image classification and therefore is limitedly useful for 'part' annotations.

2 Method

2.1 Feature Extraction

Color and Texture Characterization The classical histogram image content description approach was further refined by the classification of the image pixels in several classes, according to a local attribute (such as the edge strength). We can easily imagine a classification in three classes, consisting of pixels characterized by a small, medium and high edge strength. The number of classes is thus related to the number of quantization level of the pixel attribute's. At the limit, since every pixel has acquired a supplementary, highly relevant characteristic, we can easily imagine a one pixel per class approach, which will certainly provide a very accurate description of the image, but will require a very important size.

In order to keep the balance between the histogram size and the discrimination between pixels we propose to adaptively weight the contribution of each pixel of the image into the color distribution [Vertan and Boujemaa, 2000b]. This individual weighting allows a finer distinction between pixels having the same color and the construction of a weighted histogram that accounts both color distribution and statistical non-uniformity measures. Thus, we used a modified histogram as explained in [Vertan and Boujemaa, 2000b,Vertan and Boujemaa, 2000a]. In short: colors are uniformly quantized with 6 bins per RGB color component, yielding a 216 components feature vector per image.

Structure Characterization Images were downsampled to a maximum size of 300 pixels for any side length (width or height) to decrease computation time. The structural processing started with contour extraction ([Canny, 1986]) at 4 different scales (sigma=1,2,3 and 5). Contours were then partitioned and represented as described in (Rasche 2010) leading to 7 geometric and 5 appearance parameters for each contour segment (arc, 'wiggliness', curvature, circularity, edginess, symmetry, contrast, 'fuzziness'). Contour segments are then paired and clustered leading to another 58 parameters describing various distance measurements (between segments end and center points) and structural biases (degree of parallelism, T feature, L feature,...), see [Rasche, 2010] for details. For each parameter a 10-bin histogram is generated; the histograms are then concatenated to form a single vector of 700 dimensions. The average processing time for structural processing is ca. 40 seconds per image on a 2.6 GHz machine.

Integration of Color and Structure The color and structural parameters are then concatenated to a single image vector with ca. 916 dimensions (ca. 700 structural and 216 color parameters).

2.2 Classification

- LDA: A Linear Discriminant Analysis was applied to train a one-versus-all classifier for each of the 93 concepts (on the 8000 training images). This resulted in an average number of 19.3 labels per photo (average number of labels per training image: 11.9). The posterior values of the classifier are provided as confidence values.

Last year we also used a weighted average retrieval rank (ARR), but skipped it this year for lack of time unfortunately, although it was the better performing classifier in our runs last year.

2.3 Postprocessing - Label Elimination

Because the LDA method (see above) returned a much larger proportion of labels for the testing set (19.3 labels/image) than for the training set (11.9), we attempted to reduce the number of labels by eliminating unlikely labels based on the joint-probabilities observed in the training set. Within the training set, we determined which pairs appeared as mutual exclusive (joint probability equal 0). If a testing image contained a pair of labels that are mutual exclusive in the training set, then the one label (of the pair) was eliminated that showed a lower posterior value (obtained from the LDA classifier) in reference to the entire distribution of posterior values for each concept.

2.4 Runs

Both of our submitted runs had exactly the same preprocessing. The classification differed only by its postprocessing: one run had no post-processing, the other run was with postprocessing (with the label-elimination as explained above).

3 Results

The postprocessing did surprisingly not produce different results (or are the performance measures based on the [provided] scalar values only? That would explain the lack in difference).

Overall, the performance was worse in comparison to other approaches than last year, most likely because we did not exploit our other better-performing classifier (based on ranking).

4 Discussion

Although we applied our structural decomposition in a statistical manner only and on down-scaled image resolutions, it achieved already a performance comparable to other approaches. But in the future we should rather participate in the competition with a 'part-based' approach.

References

[Canny, 1986] Canny, J. (1986). A computational approach to edge-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.

[Ionescu et al., 2010] Ionescu, B., Rasche, C., Vertan, C., and Lambert, P. (2010). A contour-color-action approach to automatic classification of several common video genres. In *AMR 8th International Workshop on Adaptive Multimedia Retrieval. Linz, Austria*.

[Nowak et al., 2011] Nowak, S., Nagel, K., and Liebetrau, J. (2011). The clef 2011 photo annotation and concept-based retrieval tasks. In *CLEF 2011 working notes*, Amsterdam, NL.

[Rasche, 2010] Rasche, C. (2010). An approach to the parameterization of structure for fast categorization. *International Journal of Computer Vision*, 87:337–356.

[Torralba et al., 2008] Torralba, A., Fergus, R., and Freeman, W. T. (2008). 80 million tiny images: a large dataset for non-parametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 2008*, 30(11):1958–1970.

[Vertan and Boujemaa, 2000a] Vertan, C. and Boujemaa, N. (2000a). Spatially constrained color distributions for image indexing. In *Proc. of CGIP 2000*, pages 261–265, Saint Etienne, France.

[Vertan and Boujemaa, 2000b] Vertan, C. and Boujemaa, N. (2000b). Upgrading color distributions for image retrieval: Can we do better ? In Laurini, R., editor, *Advances in Visual Information Systems*, volume 1929 of *Lectures Notes in Computer Science LNCS*, chapter , pages 178–188. Springer Verlag, Berlin, Germany.