Profil Entropic visual Features for Visual Concept Detection in CLEF 2008 campaign

Herve GLOTIN & Zhongqiu ZHAO

Laboratoire des sciences de l'information et des systemes

UMR CNRS & Universite' Sud Toulon-Var France

glotin@univ-tln.fr, zhongqiuzhao@gmail.com

Abstract

In this task, we used only visual information to implement the VCDT task. We defined and compared two simple projection operators : the harmonic and arithmetic means. We proposed a new kind of compact features based on the entropy of pixels projection. These features, called Profil Entropy Features (PEF), were added to usual color means and variances, and then were fed to SVM classifiers for the detection of 17 visual concepts on the IARPR images during the CLEF 2008 campaign. The simple arithmetic mean projection is at the 4th best rank at the official test over 53 runs of around 20 laboratories. We show that the harmonic projection gives complementary information, and that its simple early fusion with arithmetic PEF yields to the third best rank system. As the runs of the other teams used state of the art SIFT an color histogram visual features, it could be concluded that PEF are efficient. Moreover, PEF are fast with around 10 images computed per second on usual pentium.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [Database Management]: Languages–Query Languages

General Terms

Measurement, Performance, Experimentation

Keywords

Rank Fusion, Image Retrieval

1   Introduction to Visual Concept Detection Clef2008 Task

ImageCLEF (http://www.imageclef.org/ImageCLEF2008) is the cross-language image retrieval track run as part of the Cross Language Evaluation Forum (CLEF) campaign. This track evaluates retrieval of images described by text captions based on queries in a different language; both text and image matching techniques are potentially exploitable. The visual concept detection task (VCDT) is one of five tasks of ImageCLEF 2008, whose goal is to identify visual concepts that would help in solving the photographic retrieval task in ImageCLEF 2008.
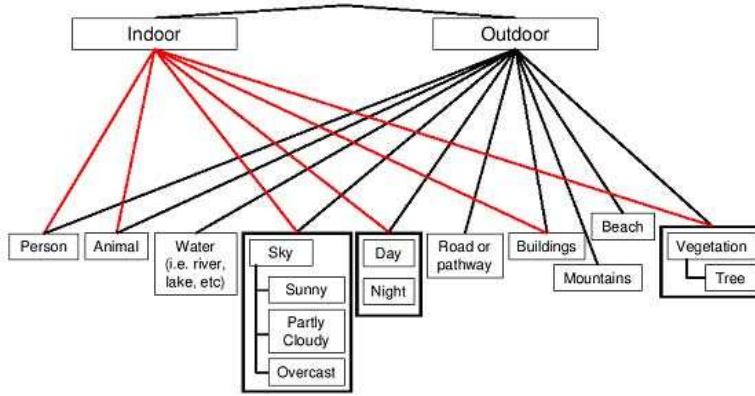
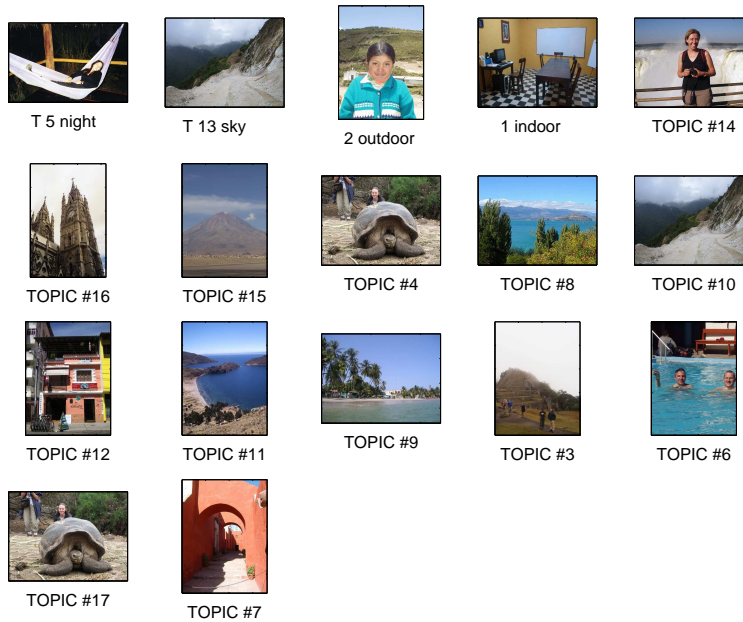Figure 1. The class hierarchy of the VCDT task of ImageCLEF 2008

Figure 2: Example of the 17 topic to find in the IAPR images database. The topics are sorted from the easier to the most difficult topic to classify according to our results.

Therefore, the VCDT task provides a training database of approximately 1,800 images which are classified according to the concept hierarchy described in Figure 1 along with their classification. Only these data may be used to train retrieval models. Figure 2 shows the examples for the 17 topics to find in the IAPR images database. So the retrieval task contains totally 17 topics. The test database consists of 1,000 images, for each of which participating groups are required to determine the presence/absence of the concepts.

In this task, we use the LS support vector machine (LS-SVM) to implement image retrieval, which will be detailed in Section 4.

2 Feature Extraction

An important step in content-based image retrieval (CBIR) system is the extraction of discriminant visual feature that are fast to compute. Information theory and Cognitive sciences can provide some inspiration for developping such feature.

Among the many visual features that have been studied, the distribution of color pixels in an image is the most common visual feature studied. The standard representation of color for content-based indexing in image databases is the color histogram. A different color representation is based on the information theoretic concept of entropy. Such entropic feature can simply equal the entropy of the pixel distribution of the image, as proposed in [1]. A more theoretical presentation of this kind of image entropy feature, accompanied by a practical description of its merits and limitations compared to color histograms, has been given in [2].

We propose in [3] a new feature equal to the pixel 'profil' entropy. A pixel profil can be a simple arithmetic mean in horizontal (or vertical) direction. The advantage of such feature is to combine raw shape and texture representations in a low cpu cost feature. These feature, associated to mean and color std, reached the second best rank in the official ImagEval 2006 campaing (see www.imageval.org and [4]).

In this paper we extend these features using another projection to get the pixel profil. We then propose also to use the harmonic mean of the pixel of each ligne or column. The idea is that the object or pixel region distribution, which is lost in arithmetic mean projection, could be partly catch by the harmonic mean. These two projections are then expected to give complementary and/or concept dependant informations. We detail below the extraction algorithm of these Profil Entropy Feature (PEF).

2.1 PEF Algorithm

Let I be an image, or any rectangular subpart of an image.

For each normalized color ($L = R + G + B$, $r = R/L$, $and g = G/L$), we first calculate two orthogonal profils by the projections of the pixels of I. We consider two simple orthogonal projection axes : the horizontal axis X (noted Π_X), versus the vertical one Y (noted Π_Y). The projection operator is either the arithmetic mean (noted 'Ar', then the projection is noted Π_X^{Ar}), as illustrated in Figure 3, or the harmonic mean of the pixels on each column or each ligne of I (noted 'Ha', then we have Π_X^{Ha}).

Then, we estimate the probability distribution function (pdf) of each profil according to [5]. Considering that the sources are ergodic, we finaly calculate each PEF equal to the normalized entropy ($H(pdf)/log(\#bins(pdf))$). We detail below each steps of the PEF extraction.

Let be op the selected projection,
for each color of I of $L(I)$ lignes and $C(I)$ columns :

$\Phi_X^{op}(I) = \hat{pdf}(\Pi_X^{op}(I))$, over $nbin_X(I) = round(\sqrt{C(I)})$ bins,
where Π_X^{op} is the vertical projection with operator op,
$PEF_X(I) = H(\Phi_X^{op}(I))/log(nbin_X(I))$.

$\Phi_Y^{op}(I) = \hat{pdf}(\Pi_Y^{op}(I))$, over $nbin_Y(I) = round(\sqrt{L(I)})$ bins,
$PEF_Y(I) = H(\Phi_Y^{op}(I))/log(nbin_Y(I))$.

We add to these PEF_a the usual entropic feature :
$\hat{pdf}(I) = $ pdf of all the pixels of I over $nbin_{XY}(I) = nbin_X(I) * nbin_Y(I)$ bins,
$PEF_.(I) = H(\hat{pdf}(I))/log(nbin_{XY}(I))$.

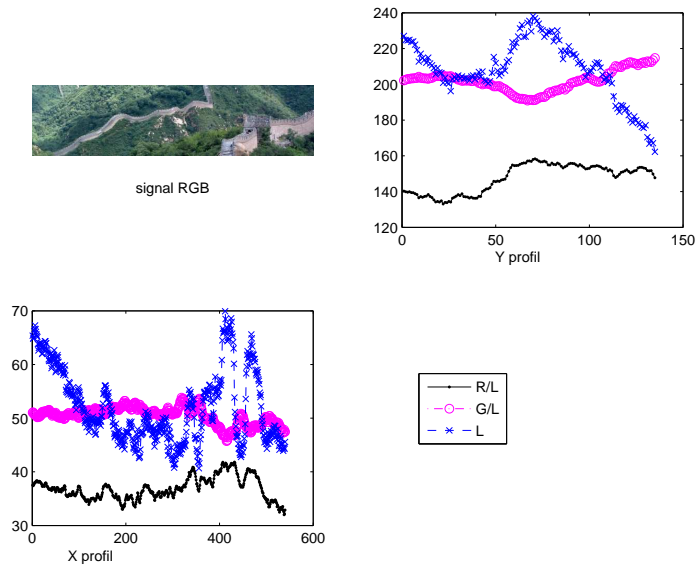And we finaly complete the PEF features by the usual mean and standard deviation of each normalized color of I.

Figure 3: Illustration of the horizontal and vertical profils using simple arithmetic projection (or sum) of each normalized color $r = R/L, g = G/L, L = R + G + B$.

2.2 PEF concatenations

We can calculate the PEF into three horizontal subimages as illustrated in Figure 4. We note such PEF '='. We also calculate the PEF in three vertical subimages, we note these PEF '‖‖'.

For each, we have 3 bands and 3 different PEF for each of the 3 colors, plus their mean and variance, thus we have $3 * 3 * 3 + 3 * 3 * 2 = 45$ dimensions for '=' or for '‖‖' features. We note '+' the feature concatenation of '=' and '‖‖' features, which has then 90 dimensions. Considering the two mean type, the PEF concatenation without repetition of the mean and std color are quite compact with a total of 126 dimensions (= 2 (subimages type '=' or '‖‖') * 3 (bands by subimages type) * 3 (rgL) * 4 (=4 Π types = (X or Y) * (Ar or Ha)) + 1 (=$H(I)$) + 2 (= mean and std))).
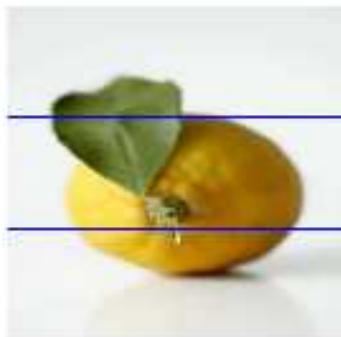

Figure 4: Illustration of the 3 subimages of type '=' (horizontal). The type '‖‖' is simple its transpose.
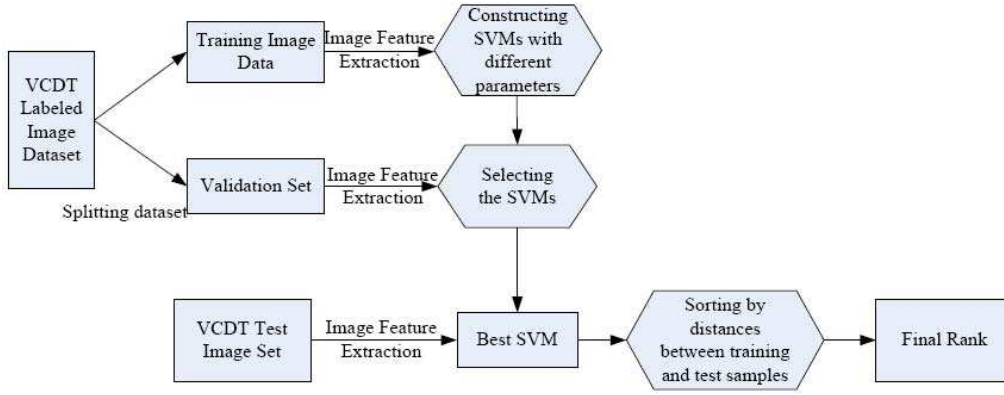
Figure 5. The framework for image retrieval of each topic

3   Support Vector Machines

The support vector machine (SVM) [6,7] first maps the data into a higher dimensional input space by some kernel functions to learn a separating hyperspace to maximize the margin. Currently, because of its good generalization capability, this technique has been widely applied in many areas such as face detection, image retrieval, and so on [8,9]. The SVM is typically based on an ε-insensitive cost function, meaning that approximation errors smaller than will not increase the cost function value. This results in a quadratic convex optimization problem. So instead of using an ε-insensitive cost function, a quadratic cost function can be used. The least squares support vector machines (LS-SVM) [10] are reformulations to the standard SVMs which lead to solving linear KKT systems instead. It is computationally attractive.

In our experiments, the RBF kernel

$$K(x_1 - x_2) = exp(-|x_1 - x_2|^2/\sigma^2)$$

is selected as the kernel function of our LS-SVM. So there is a corresponding parameter, σ , to be tuned. A large value of σ^2 indicates a stronger smoothing. Moreover, there is another parameter, γ, needing tuning to find the tradeoff between to stress minimizing of the complexity of the model and to stress good fitting of the training data points.

We train a total of 100 SVMs with different parameter values for each topic, and then we selected the best SVM using the validation set. In the experiments, we used the LS-SVMlab1.5 toolbox, which can be downloaded from http://www.esat.kuleuven.ac.be/sista/lssvmlab/.

4   Experimental Results

The process we adopt to implement the image retrieval in VCDT task is shown in Figure 5. It can also be depicted as the following steps:

Step 1) Split the VCDT labeled image dataset into 2 sets, namely training image dataset and validation set.

Step 2) Extract the visual features from the training image data using our extraction method; train and generate lots of SVM (or in the original run Kernel Discrimant Analysis or MlP) with different parameters.

Step 3) Use the validation set to select the best model

Step 4) Extract the visual features from the VCDT test image database using our extraction method; and then use the best model to find the best discriminant feature.

Step 5) Sort the test images by the distances from the positive training images and produce the final rank result.
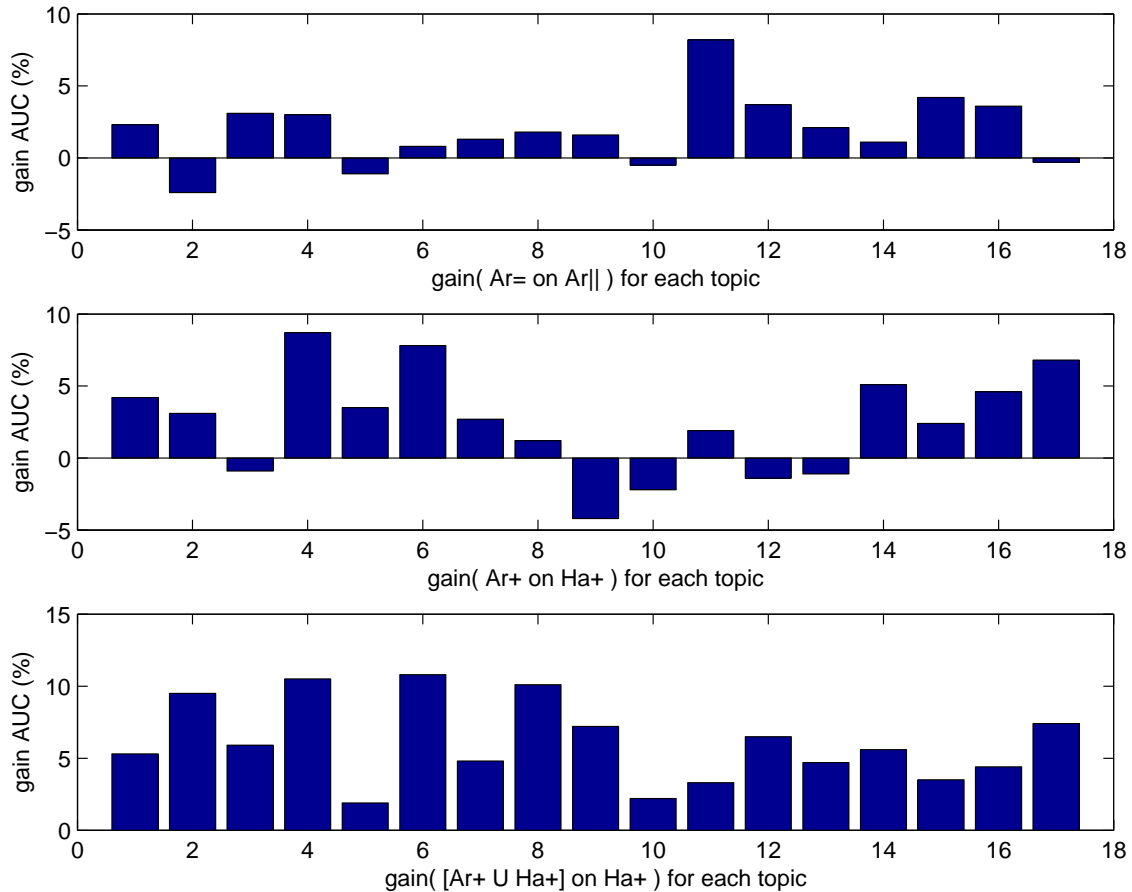
Figure 6: The relative gains for the different PEF

We show in Figure 6 the Area Under the curve (AUC) gains, calculated using the CLEF evaluation tool, for the differents PEF.

In the top of Figure 6 : gain (in %) of AUC using Arithmetic projection PEF from the = subimages compared to the AUC using PEF from the ∥∥ subimages. This shows that the choice of subimages type (horizontal or vertical subimages) could be optimized according to the topics.

Middle of Figure 6 : gain (in %) of AUC using concatenated Arithmetic PEF ('Ar+') compared to concatenated Harmonic PEF ('Ha+'). This shows that Ar PEF are more discriminant for most of the considered topics, except for 'tree' and 'mountains', where 'Ha' PEF are better.

Bottom of Figure 6 : gain (in %) of AUC using concatenated PEF (Ar+ and Ha+) against 'Ha+' alone. The gain are all positive and significant, showing the completion between Ar and Ha projection operators.

5 Discussion and conclusion

We finaly compared PEF scores to the 4 best team which participated to Clef VCDT 2008. Figure 7 gives for each topic the classification error (= the complementary of the usual area under the curve = 1 - Area Under the Curve). In average the results of our 126 PEF features (denoted by 'LSIS') are at the third rank into the initial official campaing (the average of the 17 topic errors is the given at index 18 in fig. 7). Xerox system is the best, certainly including SIFT features
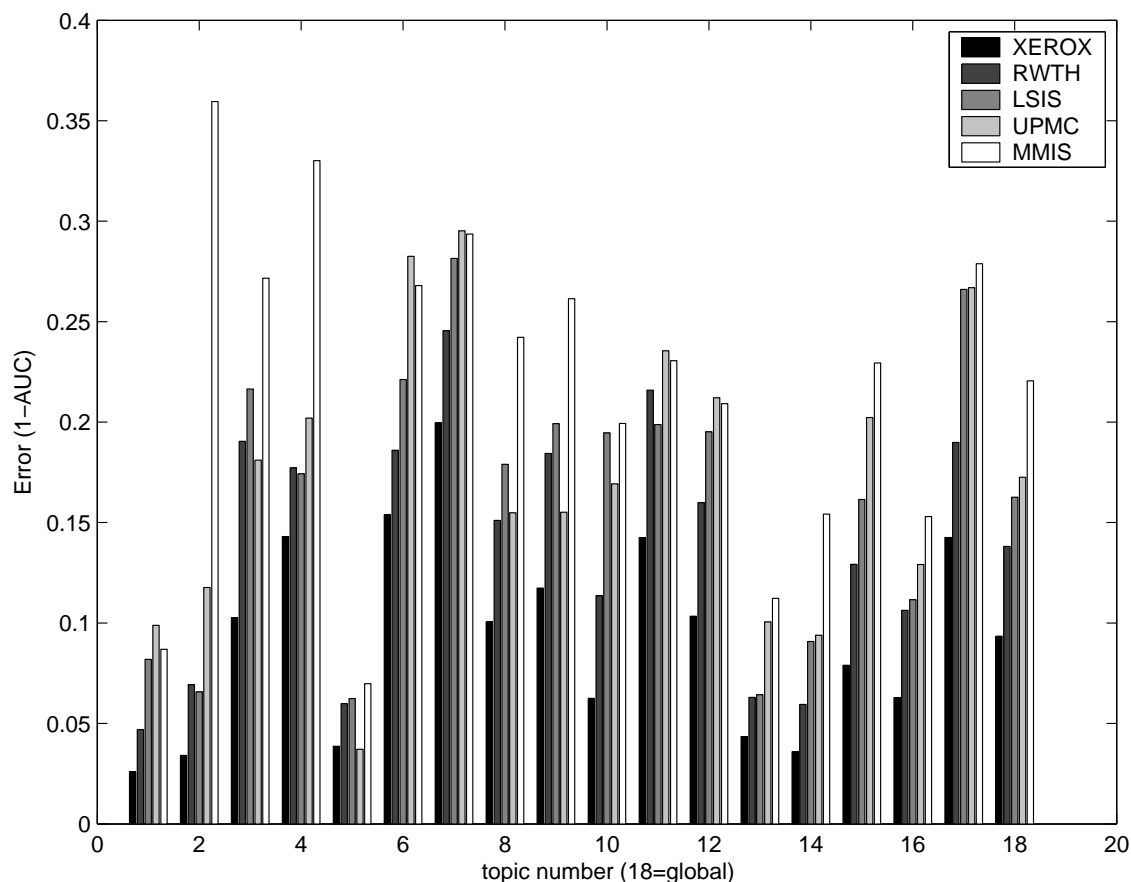
Figure 7: The different scores (=1-Area Under the Curve) for the 17 topics and the 4 best teams at the Clef VCDT task. The new fusion of the PEF features (LSIS) after the official campaign are at the third rank into the initial CLEF results. Xerox (see XEROX notes in this workshop, maybe using SIFT features) is the best. Usual color histograms features (around 200 dimensions, see this workshop note) used by UPMC seem similar or little less discriminant than PEF.

and large reference images database (see Xerox paper in this workshop). The usual perceptual color histograms features, of around 200 dimensions, that has been partly used by UPMC (see workshop note) seem similar or little less discriminant than PEF.

Acknowledgment

References

[1] M. Jagersand, Saliency maps and attention selection in scale and spatial coordinates: An information theoretic approach, in Proc. of 5th International Conference on Computer Vision, 1995.

[2] Iyengar J. Zachary, S.S and Barhen J., Content based image retrieval and information theory: A generalized approach, in Special Topic Is- sue on Visual Based Retrieval Systems and Web Mining, Journal of the American Society for Information Science and Technology, 2001, pp. 841-853.

[3] H. Glotin, "Robust Information Retrieval and perception for a scaled Lego-Audio-Video multi-structuration", Thesis of habilitation for research direction, University Sud Toulon-Var, 2007.

[4] S. Tollari and H. Glotin, Web image retrieval on imageval: Evidences on visualness and textualness concept dependency in fusion model, in ACM Int Conf on Image Video Retrieval, 2007.

[5] R. Moddemeijer, On estimation of entropy and mutual information of continuous distributions, Signal Processing, vol.16, no.3, pp. 233-246, 1989.

[6] Vapnik, V. 1995 The nature of statistical learning theory. Springer-Verlag, New York.

[7] Vapnik, V. 1998 Statistical learning theory. John Wiley, New York.

[8] Waring, C.A. and Liu, X. 2005. Face detection using spectral histograms and SVMs. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 35, 3 (June 2005), 467 - 476.

[9] Tong S., Edward, and Chang 2001. Support vector machine active learning for image retrieval. In Proceedings of the ninth ACM international conference on Multimedia Ottawa, (Canada, 2001), 107 - 118.

[10] Suykens, J.A.K. and Vandewalle, J. 1999. Least Squares Support Vector Machine Classifiers Neural Processing Letters, 9 (1999), 293-300.