PRE-CLASSIFICATION FOR AUTOMATIC IMAGE ORIENTATION

Hervé Le Borgne and Noel E. O'Connor

Centre for Digital Video Processing, Dublin City University, Dublin 9, Ireland
e-mail: {hlborgne, oconnorn}@eeng.dcu.ie

ABSTRACT

In this paper, we propose a novel method for automatic orientation of digital images. The approach is based on exploiting the properties of local statistics of natural scenes. In this way, we address some of the difficulties encountered in previous works in this area. The main contribution of this paper is to introduce a pre-classification step into carefully defined categories in order to simplify subsequent orientation detection. The proposed algorithm was tested on 9068 images and compared to existing state of the art in the area. Results show a significant improvement over previous works.

1. INTRODUCTION

Given the increasingly ubiquitous nature of image capture devices (e.g. digital cameras, camera phones, image scanners) it is now quite common for personal digital photo albums to contain hundreds or thousand images. As such, content management systems are required to allow users to efficiently organise their collections and this has been a key driver of research in content-based information retrieval in recent years. A key enabling technology for any photo management system is automatic image orientation detection and correction - photo collections of images from various sources typically contain many that are not oriented correctly from a viewer's perspective. Typically, it is desirable to determine the correct orientation of images (among the four "main" orientations *i.e* 0°, 90°, 180° and 270°,), not only so that they can be displayed to the user properly, but also because many other image processing algorithms assume correctly oriented images (e.g. face detection and recognition). As with many other image processing tasks, image orientation detection can be performed effortlessly by human perception but is an extremely challenging problem to compute.

In general, two different approaches to image orientation detection exist in the literature. The first, consists of learning a low-level description of images at correct and incorrect orientations and then using some classification scheme to determine the orientation. A study was conducted in [1] to determine the best feature to use for this, resulting in the choice of spatial color moments. Several classifiers were also compared, the best results being obtained with a Support Vector Machine (SVM) and the proposed Learning Vector Quantized (LVQ) based classifier. In [2], color moments and edge histograms were used as image signatures with one SVM trained for each feature and each orientation. The classification scheme employed was a "one-against-all" strategy consisting of classifying the image based on the classifier with the largest output. A rejection scheme was subsequently added, consisting of not classifying images with low confidence, thus automatically leading to improved results. The authors in [3] used the same features and architecture as in [2] but proposed adaBoost as a replacement for the SVM. The main difference is thus a rejection scheme, that extends the one in [2] to process indoor and outdoor images in different manner. This was motivated by the much lower accuracy of the method on indoor images.

In the second approach, originally proposed in [4] and subsequently in [5], image orientation detection is based on low-level feature extraction followed by detecting a small number of "human perception" or "semantic" cues. An arbitrary choice of specific features, such as "blue/cloudy sky", "grass" or "faces", are detected thanks to specific tools then integrated with low-level features using a bayesian framework. In [5] the choice of the additive features is justified by the output of a psychophysical study on image orientation perception. However, this study consisted of asking some users to cite the list of cues *they think they use* to detect orientation and thus may not necessarily correspond to the real (sub-conscious) criteria that users employ.

We agree with the idea that additional higher-level cues must be integrated into the classification scheme to obtain better results. However, our idea in this paper is to exploit the local statistics that are inherent to all natural images. As explained in Section 2.1 this approach can address some of the difficulties that have been encountered in previous work on image orientation. Our approach is to pre-classify a test image into a particular category in order to facilitate

Part of this work was supported by the European Commission under contract FP6-001765 aceMedia (URL: http://www.acemedia.org). Authors are also supported by Enterprise Ireland and Egide that fund the Ulysse research project ReSEND (FR/2005/56).

its subsequent processing (Section 2.2). We tested our algorithm on a large image database and compared to previous approaches (Section 3). The results obtained to date and directions for future research to further these results are discussed in Section 4.

2. PROPOSED APPROACH

2.1. Exploiting the local statistics of natural images

In [6], it was shown that their local and global statistics of natural scenes exhibit spatial differences depending on the depth perceived in the scene. For close-up scenes statistics tend to be quite stationary, whilst they are non-stationary for large viewed scenes (also known as *open* scenes). Interestingly, natural scenes preserve a right-left symmetry at each scale and their local statistics only differ along the bottom-up axis.

These observations are reflected in the limitations associated with previous work on image orientation. For example, Luo and Boutell assume the existence of some "easy" and "challenging" images in the orientation task [5], with easy images described as "long distance, outdoor scenes, and images with sky" but the author do not take particular advantage of this assumption in their algorithm. In a similar vein, [3] observed that indoor images are much more difficult to orientate than outdoor ones. To address this, they proposed a pre-classification step into indoor and outdoor images, each of which are processing differently in the rejection scheme. However, if no rejection step is employed, as in our paper, their method becomes equivalent to the one of [2] albeit using adaBoost instead of an SVM.

The "easy" group identified by [5] could be explained by the presence of sky. However, since the sky is far from being systematically blue in real consumer photo databases (see fig 2), it is more likely due to the non-stationary local statistics of these *open* scenes. Similarly, the difficulties met by [3] to orientate indoor images can be explained by the stationary nature of their local statistics. Considering the works made on natural scenes statistics and the difficulties met in previous work on image orientation, we propose in this paper to process the images according to the type of scene they belong.

2.2. Processing Steps

Our system framework, illustrated in fig 1, has three steps both for training and testing: feature extraction, category type classification and orientation detection. The features extracted in the first step correspond to edge direction histogram (EDH), color moments (CM) and scalable colour. The first two features are extracted for use in the image orientation step and these features are used so that we can compare with previous similar work. The third feature, scalable colour, is used in the classification step only. Since classification occurs prior to orientation detection, it is important to use a rotation invariant feature such as scalable colour [7]. The EDH was implemented as proposed in the MPEG-7 standard [7] to capture the luminance edge information. Edges are extracted onto a 4×4 grid and grouped into five categories ((0°, 45°, 90°, 135° and isotropic), resulting in a 80-dimensional vector. Previous work uses larger vectors due to finer grids (5×5 in [5, 2, 3]) and a larger number of directions (17 in [5], 37 in [2, 3]). The two first color moments were extracted on each layer in the LUV color space, again using a 4×4 grid ([1] used a 10×10 grid whilst [5, 2, 3] used a 5×5 grid), leading to a 96-dimensional vector for orientation detection that was centered (null mean) and reduced (unitary standard deviation).

The second step consists of classifying the image into pre-determined categories. These categories were chosen in such a way so as to be prototypical of natural scenes. In both *artificial*, i.e containing a majority of human-made structures, and *natural*, i.e without human-made elements, scenes, we selected both close-up and wide-view scenes. This corresponds to four categories, denoted $\mathcal{C}_j$ in the following. Unlike this approach, [5] pruned the close-up scenes from the training set in order to simplify SVM training. In our approach, two different classifiers were tested: a K-nearest-neighbour classifier [8] and a Support Vector Classifier [9] with polynomial kernel of third degree or a sigmoid one, and a one-against-one strategy to implement the multi-class classification [10].

The third step is the orientation detection itself. During the training phase, one SVM classifier [9] is trained for each direction θ_i using images from each prototypical class $\mathcal{C}_j$ only. Let $L_j = N_{train}/4$ training data (x_k, y_k) for category $\mathcal{C}_j$, where $x_k \in \mathbb{R}^{176}$ and $y_k = 1$ for direction θ_i and $y_k = 0$ for the three other directions (one-against-all implementation [10]). The corresponding classifier solves the following problem:

$$\begin{aligned} \min_{w^{i,j}, b^{i,j}, \xi^{i,j}} & \left\{ \frac{1}{2} (w^{i,j})^T w^{i,j} + C \sum_{L_j}^{k=1} \xi_k^{i,j} \right\} \\ (w^{i,j})^T \phi(x_k) + b^{i,j} & \geq 1 - \xi_k^{i,j}, \text{if } y_k = 1 \\ (w^{i,j})^T \phi(x_k) + b^{i,j} & \leq 1 - \xi_k^{i,j}, \text{if } y_k = 0 \\ \xi_k^{i,j} & \geq 0, k = 1, \ldots, L_j \end{aligned} \quad (1)$$

where the training data x_k is mapped to a higher dimensional space by the function ϕ, C is a penalty parameter and $\xi_k^{i,j}$ are a measure of the misclassification errors. The idea of SVM is to maximize the margin $(2/\|w^{i,j}\|)$ between images at direction θ_i and others (within the class $\mathcal{C}_j$). When data is not linearly separable, the penalty term reduces the number of training errors.

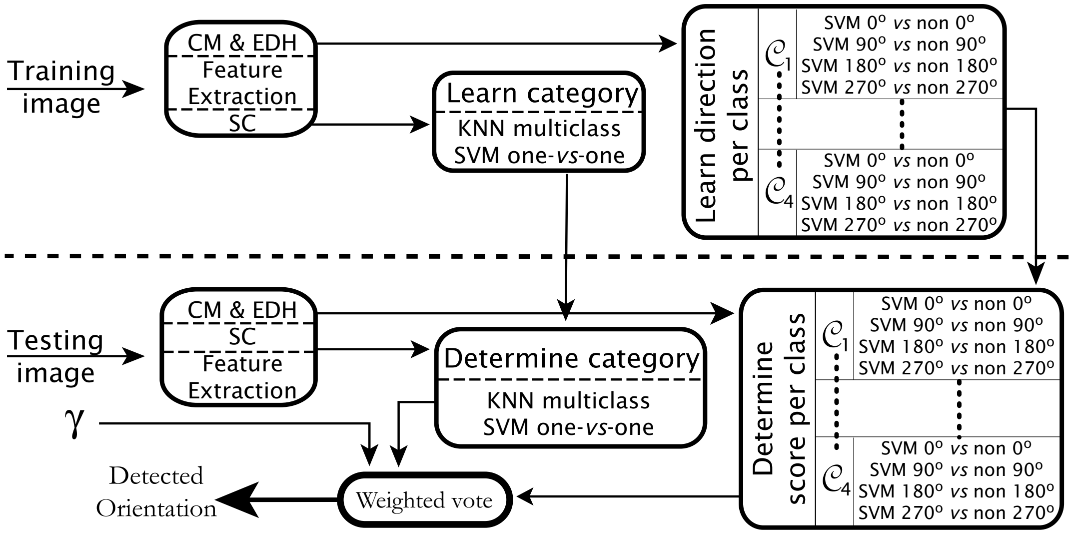

Fig. 1. Our system framework

After solving (1), there are 16 decision functions corresponding to the four direction on the four prototypical classes. Given the feature vector X_t of a test image, that has been pre-classified into the prototypical class $\mathcal{C}_{j_0}$, its orientation is calculated as:

$$arg\max_i \quad \Big\{ sign\Big\{ (w^{i,j_0})^T \phi(x_t) + b^{i,j_0} + \gamma\Big(\sum_{j \neq j_0} ((w^{i,j_0})^T \phi(x_t) + b^{i,j_0}) \Big) \Big\} \Big\} \quad (2)$$

where γ is a weighting coefficient giving the relative importance of the chosen pre-classified category.

3. EXPERIMENTAL EVALUATION

The image database consists of a mixture of professional images from Corel[1] and the goodshoot[2] stock photo library as well as personal pictures. We used $4 \times 4 \times 100 = 1600$ training images (100 samples per classes and per orientation) and $4 \times 2267 = 9068$ test images. Note that each image was used in all four directions. The size of our testing database is at least as large as those in previous works. However, unlike previous work in which the training set was typically twice the size of the testing set, the size of our training database is less than 20% of the test set. This attempts to reflect a more realistic application scenario, since in practice the full set of potential users will submit many more testing images to the system than any training set could contain. Two classifiers were tested for the second step.For all SVM classifiers, we used the LibSVM implementation [11], with a polynomial kernel of degree 3 for the third step (orientation detection).

We compared our proposed approach to our own implementations of three other approaches already reported. The *merged SVM* was proposed in [1]. It consists of merging CM and EDH then using one classifier to learn each orientation. On testing, the largest output among the four SVM gives the orientation. The *parallel SVM* method was proposed by [2]. CM and EDH were used as image signatures with one SVM trained for each feature and each orientation. The classification scheme employed was a "one-against-all" strategy. A second layer of four SVM (one per orientation) can be added to learn the output of both features instead of averaging. In both cases, normalisation as suggested in [2] was used.

Overall results for orientation detection are shown in Table 1. The best result for our approach is almost 8% better then the *parallel SVM* and 3% better than the *merged SVM*. The poor performance of *parallel SVM* is surprising considering the results reported in [2]. This may be due to the much smaller training set used in our experiments.

We tested different strategies for the pre-classification step (see Table 1 as well as different values of γ. The best result was obtained with a KNN classifier ($K = 5$) although the choice of the classifier did not make a significant difference. Rather, significant variation in the results according to γ can be observed with best results obtained at $\gamma = 1/2$. The particular case of $\gamma = 1$ for which all the classes have the same weight gives reasonable results of 76.8%. On

[1] http://www.corel.com
[2] http://www.goodshoot.com

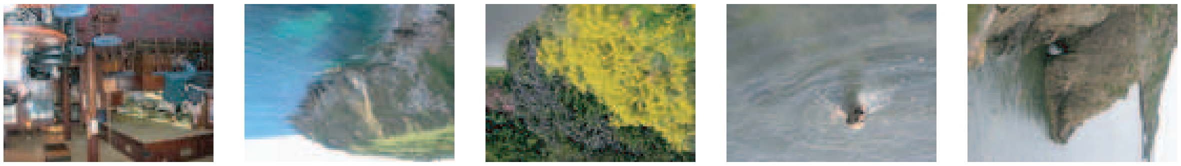

Fig. 2. Examples of failures of our method (KNN. K=5 - $\gamma = 1/2$), . Image were originally naturally oriented. They are shown in the detected orientation

the other hand, when the pre-classified class is used alone ($\gamma = 0$), the results drop to 73% with a 5-NN classifier. Some examples of where the approach fails are illustrated in Figure 2, and some typically correspond to images for which even humans would have difficulty in determining orientation.

Experiments	Accuracy
Parallel SVM 1 layer [2]	70.0%
Parallel SVM 2 layers [2]	68.9%
Merged SVM [1]	74.6%
KNN (K=8) - ($\gamma = 1/2$)	77.5%
KNN (K=8) - ($\gamma = 1/4$)	76.9%
KNN (K=8) - ($\gamma = 0$)	72.9%
KNN (K=5) - ($\gamma = 1/2$)	77.8%
KNN (K=5) - ($\gamma = 1/4$)	77.3%
KNN (K=5) - ($\gamma = 0$)	73.0%
SVM (poly) - ($\gamma = 1/2$)	77.1%
SVM (poly) - ($\gamma = 1/4$)	76.5%
SVM (poly) - ($\gamma = 0$)	72.5%
SVM (sigm) - ($\gamma = 1/2$)	77.3%
SVM (sigm) - ($\gamma = 1/4$)	76.4%
SVM (sigm) - ($\gamma = 0$)	69.4%
any (equal weights) - ($\gamma = 1$)	76.8%

Table 1. Accuracy of overall orientation detection considering different pre-classification strategies.

4. CONCLUSION

In this paper we proposed a new scheme to automatically detect image orientation in natural images, based on a pre-classification step that takes advantage of the local statistical structure of natural scenes. With only four prototypical categories, the proposed algorithm performed extremely well compared to other published approaches. Results reported are comparable to (or better than) those reported in [2] and [5]. Furthermore, it should be noted that our approach was trained using a significantly smaller training set than that used in previous approaches.

5. REFERENCES

[1] A. Vailaya, H.J. Zhang, C. Yang, F.-I. Liu, and A.K. Jain, "Automatic image orientation detection," *IEEE trans. Image Processing*, vol. 11, no. 7, pp. 746–755, 2002.

[2] Y.M. Wang and H. Zhang, "Detecting image orientation based on low-level visual content," *Computer Vision and Image Understanding*, vol. 93, pp. 328–346, 2004.

[3] L. Zhang, M. Li, and H.-J. Zhang, "Boosting image orientation detection with indoor vs. outdoor classification," in *WACV'02*, December 2002, Florida, USA.

[4] L. Wang, X. Liu, L. Xia, G. Xu, and A. Bruckstein, "Image orientation detection with integrated human perception cues (or which way is up)," in *IEEE international conference on Image Processing*, 2003, vol. 2, Barcelona, Spain.

[5] J. Luo and M. Boutell, "Automatic image orientation detection via confidence-based integration of low-level and semantic cues," *IEEE Trans PAMI*, vol. 27, no. 5, pp. 715–726, May 2005.

[6] A. Torralba and A. Oliva, "Statistics of natural images categories," *Network: Computation in Neural Systems*, vol. 14, pp. 391–412, 2003.

[7] B.S. Manjunath, J.-R Ohm, V.V Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE trans. CSVT*, vol. 11, no. 6, pp. 703–715, 2001.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience Publication, 2000.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*, NY:Springer-Verlag, 1995.

[10] C.-W. Hsu and C.-J. Lin, "A comparison on methods for multi-class support vector machines," *IEEE trans. on Neural Networks*, vol. 13, pp. 415–425, 2002.

[11] C.C. Chang and C.J. Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at `www.csie.ntu.edu.tw/~cjlin/libsvm`.