

Automatic Image Orientation Determination with Natural Image Statistics

Siwei Lyu
Department of Computer Science
Dartmouth College
Hanover NH 03755

In this paper, we propose a new method for automatically determining image orientations. This method is based on a set of natural image statistics collected from a multi-scale multi-orientation image decomposition (e.g., wavelets). From these statistics, a two-stage hierarchal classification with multiple binary SVM classifiers is employed to determine image orientation. The proposed method is evaluated and compared to existing methods with experiments performed on 18040 natural images, where it showed promising performance.

1 Introduction

Displaying images in their correct orientations is one of the basic requirements in image processing. While manually adjusting orientations for several images is trivial, it is more efficient to be able to automate on several hundred digital photographs taken from a field trip or a vacation. One solution is to have the digital cameras record, at the time of capture, the orientation information in the image file (for instance, a user-defined tag in the JPEG header). However, there is yet not a widely accepted protocol for image processing softwares to take advantage of such information, and most legacy digital images were taken with cameras without such a feature. A more practical alternative then is to design systems that are able to determine image orientations with signal processing.

Technically, the goal of automatic image orientation determination is to classify an image to one of the four possible orientations, corresponding to rotation angles of 0° , 90° , 180° and 270° . Nevertheless, in practice, it is usually sufficient to determine if an image is landscape-oriented (0° or 180° orientation) or portrait-oriented (90° or 270° orientation), as it is rare to take a picture upside down. Existing automatic image orientation determination methods fall into two main categories. Top-down methods are based on high-level perception cues (e.g., the detection of faces, skies and walls [3]), or semantic relations in image contents (e.g., textured area in lower part [11]). Though a closer modeling of the human perception process, top-down methods suffer from the instabilities of current object detection and recognition algorithms, and are more likely to bias to a particular set of training images. On the other hand, bottom-up methods determine image orientations with low-level features, examples include the color moments [9] and the edge direction histograms [12, 13]. Compared to high-level cues, low-level features are more robust and reliable. Furthermore, psychophysical studies also confirmed that low-level features are critical for humans performance on determining image orientations [4].

In this paper, we propose a new low-level image feature for orientation determination, which consists of a set of natural image statistics collected from a multi-scale multi-orientation image decomposition (e.g., wavelets). Previously, we have shown that these image statistics are effective in detecting image steganography [5] and differentiating natural images from computer generated images [6], as they capture statistical correlations within natural images across different scales and color channels. In this work, these statistics are combined with a hierarchical two-stage classification with multiple binary SVM classifiers to determine image orientation. Experimental results on 18,040 natural images of the proposed method is reported and compared to existing methods.

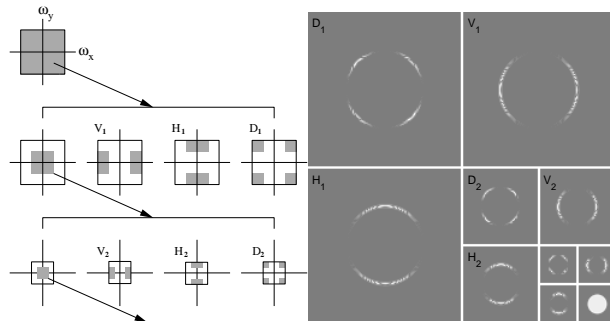


Figure 1: Shown on the left is an idealized multi-scale and multi-orientation decomposition of frequency space. Shown, from top to bottom, are levels 0, 1, and 2, and from left to right, are the low-pass, vertical, horizontal, and diagonal subbands. Shown on the right is the magnitude of a multi-scale and orientation decomposition of a “disc” image.

2 Natural Image Statistics

The image statistics are collected from a multi-scale multi-orientation¹ image decomposition based on separable quadrature mirror filters (QMFs) [8]. As shown in Figure 1, such a decomposition splits the frequency space into multiple scales and orientations (vertical, horizontal, and diagonal). For a color (RGB) image, the decomposition is applied independently to each color channel. The resulting vertical, horizontal, and diagonal subbands at scale i are denoted as $V_i^c(x, y)$, $H_i^c(x, y)$, and $D_i^c(x, y)$, where $c \in \{r, g, b\}$.

One important characteristics of natural images is that the coefficients in each oriented subband assume distributions characterized by a sharp peak at zero and large symmetric tails [1]. This is because natural images typically contain large smooth regions and abrupt transitions (e.g., edges). The smooth regions, though dominant, produce small coefficients near zero, while the transitions generate large coefficients. Instead of directly modeling these distributions, a set of statistics (mean, variance, skewness, and kurtosis) are collected to characterize them for simplicity.

There are also higher-order correlations within the decomposition among coefficients not captured by their marginal distributions [1]. Salient image features (e.g., edges) tend to orient spatially and extend across multiple scales and color channels. As a result, the coefficient magnitudes around such image features, which measure the localized energy at each spatial location, are correlated across space, orientation, scale and color channels. For example, a vertical edge creates coefficients with large magnitudes in the vertical subbands which are likely to have upper and lower spatial neighbors with large magnitudes. Similarly, if there is a coefficient with a large magnitude at scale i , it is also very likely that its “parent” at scale $i + 1$ will also have

¹This is the orientation of the 2D filters.

a large magnitude.

To this end, a second set of statistics are collected from the linear prediction errors of coefficient magnitudes [1]. Consider the vertical subband of the green channel at scale i , $V_i^g(x, y)$, a linear predictor for the magnitudes of these coefficients in a subset of all possible spatial, orientation, scale, and color neighbors² is formed as:

$$|V_i^g(x, y)| = w_1|V_i^g(x-1, y)| + w_2|V_i^g(x+1, y)| \\ + w_3|V_i^g(x, y-1)| + w_4|V_i^g(x, y+1)| \\ + w_5|V_{i+1}^g(x/2, y/2)| + w_6|D_{i+1}^g(x/2, y/2)| \\ + w_7|D_i^g(x, y)| + w_8|V_i^b(x, y)| + w_9|V_i^r(x, y)|, (1)$$

where $|\cdot|$ denotes magnitude and w_k are the scalar weights. Evaluating Eq. (1) across the whole subband yields:

$$\vec{v} = Q\vec{w}, \quad (2)$$

where \vec{v} is formed by all $|V_i^g(x, y)|$ strung out into a column vector (to reduce sensitivity to noise, only magnitudes greater than a pre-given threshold are considered), the columns of the matrix Q contain the neighboring coefficient magnitudes as specified in Eq. (1), and $\vec{w} = (w_1 \dots w_9)^T$. Eq. (2) is solved with the least squares as:

$$\vec{w} = (Q^T Q)^{-1} Q^T \vec{v}. \quad (3)$$

Similar linear predictors are constructed in all other subbands corresponding to different orientations, scales and color channels, with slightly different neighborhood settings.

With the linear predictors, the log errors between the actual and predicted coefficient magnitudes are computed as:

$$\vec{p} = \log(\vec{v}) - \log(|Q\vec{w}|), \quad (4)$$

where the $\log(\cdot)$ is computed point-wise on each vector component. Then the mean, variance, skewness, and kurtosis are collected to characterize the error distributions of each subband in the decomposition.

For a QMF decomposition with n scales, the total number of coefficient statistics is $36(n-1)$ (4 statistics for 3 oriented subbands and $(n-1)$ levels per color channel), and for similar reasons, the total number of error statistics is also $36(n-1)$, yielding a grand sum of $72(n-1)$ statistics. Specifically, for a decomposition of 4 levels, this setting yields 216 statistics, which are the features for determining image orientation.

3 Classification

Based on these image statistics, non-linear support vector machine (SVM) classifiers [10] are employed to determine

²The particular choice of neighbors was motivated by the observations of [1] and modified to include non-casual neighbors.

image orientation. Instead of treating the detection as a multi-class problem [12, 13], we adopt a hierarchal two-stage decision tree with SVM as base classifiers. In the first stage, one binary SVM classifier is used to differentiate images with landscape orientations (0° or 180°) from those with portrait orientations (90° or 270°). The second stage of classification takes landscape or portrait images and further determine their orientations with two more binary SVM classifiers, the $0/180$ classifier and the $90/270$ classifier. To give probabilistic meanings to the outputs of each binary SVM classifier, they are calibrated to the posterior probabilities of classification with a logistic function, whose parameters are estimated with a nonlinear least-squares [7]. Another important aspect in building the classifiers for image orientation determination is to use a proper rejection criterion. As pointed out by several authors [3, 12, 13], there are many images lacking clear notion of orientation, due to factors such as homogeneous textures, close-up views and nearly diagonal rotations. These images are inherently ambiguous and are subject to rejection by the classifiers. Specifically, images with a classifier output near 0.5 are thus rejected as being too ambiguous for classification (labeled as N/D). The number of images being rejected is controlled by a pre-given threshold t that defines the projection region as $[0.5-t, 0.5+t]$. The overall process of image orientation determination is shown in Figure 2.

Compared to the multi-class classification method, where multiple binary classifiers are combined in either the one-against-all or pairwise fashion, the two-stage framework is more tailored to the orientation determination problem and affords several advantages. First, only three binary classifiers are needed, whereas there are four binary classifiers in one-against-all and six in all-pair classifications. Using less classifiers simplifies the overall training process, which requires less training data. Secondly, the output of each binary classifier has specific meaning in the proposed two-stage framework, obviating merging outputs of the composing binary classifiers. Finally, as pointed out earlier, in many practical applications such as organizing personal photo albums, it is sufficient to determine the orientation of an image to the level of portrait/landscape. Thus the intermediate classifier outputs from the first stage can be reported directly without further processing.

4 Experiments

To empirically evaluate the proposed orientation determination method, we conducted a set of experiments on an image database of 18040 photographic images. Images in this database come from various sources ranging from professional image galleries to personal photo albums. these images span a range of contents (e.g., landscapes, city scenes and portraits) and imaging conditions (e.g., indoor and out-

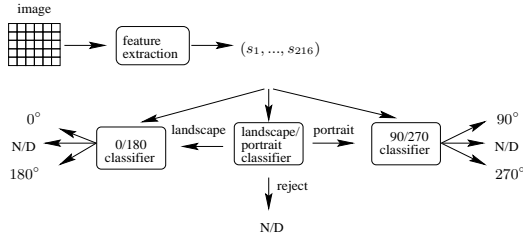


Figure 2: Overall process of image orientation determination. N/D stands for “not detectable”, corresponding to images being rejected by the classifiers.

door lighting, close up and far away views, etc.). Among all the 18040 images, 9177 are landscape-oriented and 8863 are portrait oriented. A further categorization in orientations shows that there are 8946 images with a 0° orientation angle, and 7992 images with a 90° orientation angle. There are relative fewer images of 270° and 180° in the database (871 and 231, respectively). As such image orientations are less frequently used in practice.

From this image database, 6000 landscape and 6000 portrait images were used to train the landscape/portrait classifier, while the rest formed the testing set. The training set of the 0/180 classifier is constructed from the 6000 labeled landscape images. Besides these images, it also includes their 180° rotated copies, to accommodate the relative small number of 180° oriented images. The training set of the 90/270 classifier is similarly formed from the 6000 portrait images.

From each image, training and testing alike, image statistics as described in Section 2 were extracted. To accommodate different image sizes, statistics were collected from the central 256×256 image region. For each image region, a four-level three-orientation QMF pyramid was constructed for each color channel, from which 216 coefficient and error statistics were collected to form a 216-D feature vector. For a basis of comparison, two other low-level image features used for orientation determination, color moments (CM) [9] and edge direction histograms (EDH) [12] were also collected on each image. The CM and EDH features were vectors of 288 and 945 dimensions, respectively. As a standard pre-processing step in SVM classification, each dimension in all type of features were normalized over training examples to the same scale. From the collected image feature vectors, the three binary nonlinear SVM classifiers with radial basis function (RBF) kernels were trained and tested³. The parameters of the SVMs, i.e., the regularization factor and the width of the RBF kernel were found by cross-validation.

³SVM algorithm in our experiments was implemented with package LIBSVM [2].

	rejection rate			
	0%	10%	20%	50%
CM	69.4	73.6	84.3	88.2
EDH	72.3	79.1	89.7	93.2
our work	78.9	81.3	90.4	95.1

Table 1: Classification accuracies of the landscape/portrait classifier, with different low-level features and different rejection rates

	rejection rate			
	0%	10%	20%	50%
CM	71.4	78.9	86.3	93.2
EDH	69.3	75.2	87.6	95.1
our work	67.2	73.1	84.9	91.6

Table 2: Classification accuracies of the 0/180 classifier, with different low-level features and different rejection rates

Listed in Table 1 are the classification accuracies of the landscape/portrait classifier on the testing set, with varying rejection rates. For comparison, performances of SVM classifiers with CM and EDH features are also shown. Note that the proposed low-level feature of natural image statistics achieved a better performance than both the CM and the EDH features, while having a relative lower dimensionality. Shown in Figure 3 are some images whose orientations are correctly determined by the classifier based on the proposed feature, and in Figure 4, examples of images whose orientations are incorrectly determined are shown. It seems that the proposed image statistics captures certain structural regularities in an image on which the classification is based. Shown in Figure 5 are some examples of images being rejected by the classifier, corresponding to a rejection rate of 10%. Many of the rejected images lack a definite orientation and can be plausibly explained as either landscape or portrait oriented.

Shown in Table 2 and 3 are the classification accuracies, with different rejection rates, of the 0/180 classifiers and 90/270 classifiers with different feature types. In these cases, however, the proposed image statistics features did not have an obvious advantage over the other feature types. One possible reason is that the statistics collected are more sensitive to a 90° rotation, a total reshuffle of components in the feature vector with all statistics of vertical subbands and horizontal subbands switching their positions, than a 180° rotation of an image. Nevertheless, by combining different features in different stages of classification, better performance is expected.

	rejection rate			
	0%	10%	20%	50%
CM	53.7	65.8	73.2	88.7
EDH	61.3	67.8	76.5	87.6
our work	59.8	65.7	77.9	89.3

Table 3: Classification accuracies of the 90/270 classifier, with different low-level features and different rejection rates

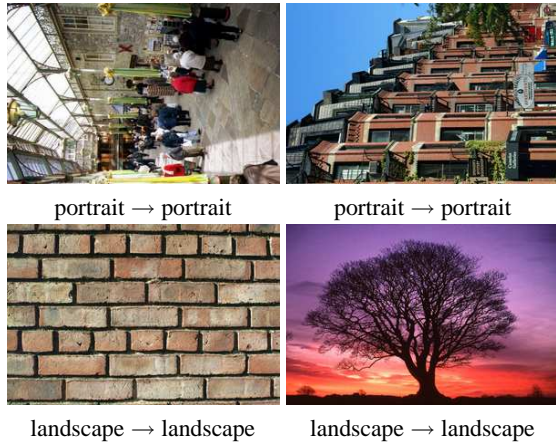


Figure 3: Examples of images whose orientations are misclassified by the landscape/portrait classifier using image statistics features.

5 Discussion

In this paper, we present a method for automatically detecting image orientations, based on a set of natural image statistics and SVM classification. The image statistics capture regularities in different oriented natural images and the nonlinear SVM classifier transform such difference into a computable procedure efficiently. Experimental results based on 18040 natural images seem to confirm the efficacy of the proposed method.

However, our work also indicates that there is no single low-level feature sufficient to reliably determine image orientation. One of our on-going work is to combine low-level features of different types to achieve the optimal performance.

Acknowledgment

The author would like to thank Hany Farid for helpful and inspiring comments about the paper. This work was supported by Hany Farid under an Alfred P. Sloan Fellowship, an NSF CAREER Award (IIS99-83806), an NSF Infrastructure Grant (EIA-98-02068), and under Award No. 2000-

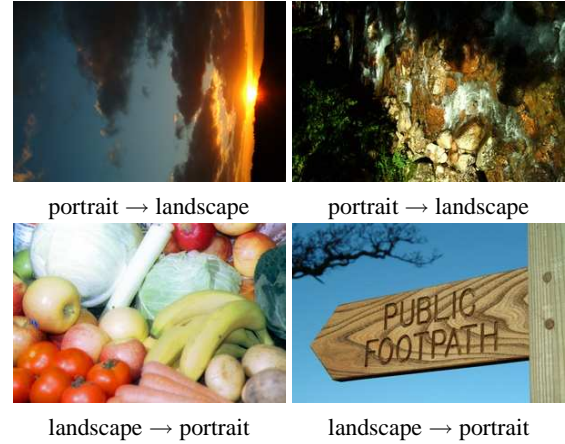


Figure 4: Examples of images whose orientations are misclassified by the landscape/portrait classifier using image statistics features.



Figure 5: Rejected images by the landscape/portrait classifier using image statistics features with a rejection rate of 10%.

DT-CX-K001 from the Office for Domestic Preparedness, U.S. Department of Homeland Security (points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Homeland Security).

References

- [1] R.W. Buccigrossi and E.P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8(12):1688–1701, 1999.
- [2] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [3] J. Luo and M. Boutell. A probabilistic approach to image orientation detection via confidence-based integration of low-

- level and semantic cues. In *International Workshop on Multimedia Data and Document Engineering*, 2004.
- [4] J. Luo, D. Crandall, A. Singhal, M. Boutell, and R. Gray. Psychophysical study of image orientation perception. *Spatial Vision*, 2003.
- [5] S. Lyu and H. Farid. Detecting hidden messages using higher-order statistics and support vector machines. In *5th International Workshop on Information Hiding*, Noordwijkerhout, The Netherlands, 2002.
- [6] S. Lyu and H. Farid. How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 53(2):845–850, 2005.
- [7] J. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In *Advances in Neural Information Processing Systems (NIPS)*, 1999.
- [8] E.P. Simoncelli and E.H. Adelson. *Subband image coding*, chapter Subband transforms, pages 143–192. Kluwer Academic, 1990.
- [9] A. Vailaya, H. Zhang, and A. Jain. Automatic image orientation detection. In *International Conference on Image Processing (ICIP)*, 1999.
- [10] V. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 1995.
- [11] 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 *International Conference on Image Processing (ICIP)*, 2003.
- [12] Y. Wang and H. Zhang. Content-based image orientation detection with support vector machines. In *IEEE Workshop on Content-base Access of Image and Video Libraries (CAIVL)*, 2001.
- [13] L. Zhang, M. Li, and H. Zhang. Boosting image orientation detection with indoor vs. outdoor classification. In *IEEE Workshop on Applications of Computer Vision (WACV)*, 2002.