Large Scale Performance Measurement of Content-Based Automated Image-Orientation Detection

Shumeet Baluja
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA. 94043
shumeet@google.com

Henry A. Rowley
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA. 94043
har@google.com

***Abstract* – With the proliferation of digital cameras and self-publishing of photos, automatic detection of image orientation will become an important part of photo management systems. In this paper, we perform a large scale empirical test to determine whether the common techniques to automatically determine a photo's orientation are robust enough to handle the breadth of real-world images. We use a wide variety of features and color-spaces to address this problem. We use test photos gathered from the web and photo collections, including photos that are in color and black and white, realistic and abstract, and outdoor and indoor. Results show that current methods give satisfactory results on only a small subset of these images.**

I. INTRODUCTION

The wide-spread adoption of digital cameras and camera phones has lead to an explosion in the number of personal photos. Unfortunately, many cameras do not tag photographs with the orientation of the camera. Therefore, at some point in the photo management process, every user must manually ensure that each photo is in its correct orientation. We are interested in streamlining photo management tasks as much as possible; in this paper, we investigate the task of automatically determining the orientation of an image. Examples of this task are given in Figure 1.

This task is made complicated by the wide variety of photographs that are taken. Many of the more typical vacation images (such as sunsets, beaches, etc) have an easily recognizable pattern of light and dark that can be exploited to yield good results for this task. Indoor scenes are more difficult due to the variations in lighting sources, and finally abstracts or macro shots (close-ups) provide some of the greatest challenges, since there may be no clear anchor points or lighting sources in the image.

Recently, there have been a variety of approaches proposed to solve this problem [1-4]. All of these approaches use low-level features of the image, such as spatial color histograms and edges, which are fed into statistical classifiers. Results have been reported with accuracies as high as 90-97% of the images on small test sets. In this paper, we follow a similar, but simplified, methodology with a large set of features to determine which features are most effective. We also test the resultant classifiers on a very large set of images drawn from the web as well as professional photographs. Our goal is to determine if these approaches are ready for mass deployment in photo management applications.

Figure 1: Row 1: An easy example - lighting and sky-patch are giveaways. Row 2: Difficult close-up. Row 3: Difficult example; original image has no color information (sepia toned), contains reflections in water, and contains few markers. Row 4: If deeper information (such as face detection) was used, this image would be much easier to classify.

II. FEATURES AND ALGORITHMS

Following previous work [1-4], we first extract a large number of simple features from each image. The full set of features that we consider are listed here. From the original image, 15 simple transformed single channel images are computed:

- 1-3. R, G, B Channels
- 4-6. Y, I, Q Channels
- 7-9. Normalized versions of R,G,B; linearly scaled to span 0-255.

10-12. Normalized versions of Y,I,Q; linearly scaled to span 0-255.
13. Intensity (computed as average of R, G, B)
14. Horizontal edges computed from intensity
15. Vertical edges computed from intensity

For each of these transformed images, we compute the mean and variance of the entire image. In addition, we compute mean and variance of square subregions of the image. The subregions are squares that cover (1/2)x(1/2) to (1/6)x(1/6) of the image (there are a total of 91=1+4+9+16+25+36 squares). We also compute the mean and variance of rows and columns that cover 1/2 to 1/6 of the image (there are a total of 20=2+3+4+5+6 rows, and 20 columns). There are 1965 features representing averages (15*(91+20+20)) and 1965 features representing variances, for a total of 3930 features.

For each image to be classified, these features are computed for each of the possible 4 upright orientations. We use a Support Vector Machine (SVM) [5] to perform the statistical classification[1]. In the experiments with the full set of features, the SVM is trained to take the 3930 features as input and output positive value if the image is in the upright orientation, and a negative value otherwise.

III. EXPERIMENTAL SETUP

Training the SVMs to correctly identify upright images requires a large set of labeled training images. We gathered the training images from two sources: Corel 1,300,000 image library, disk #7 [6] (using mainly scenic shots, tourist type shots and people shots) and from images from Google's Image Search by searching for images of people by using query terms of "mother", "father", "grandmother", "grandfather", etc. A total of 6,009 unique images were used for training. For input into the SVM, the original upright image as well as three rotations of the image (in 90° increments) were converted into the features described in the previous section and used as inputs with positive and negative values as target outputs.

SVMs are sensitive to parameters settings. To ensure that we did not bias our results due to improper training of the SVMs, we tried a variety of settings. In total, we trained 120 SVMs using SVM-Lite [5]. The SVMs were trained with RBF kernels with the gamma parameter (controlling the RBF kernel) varying between 0.1 and 0.00001 and the C parameter (controlling the tradeoff between training error and margin) between 5 and 500,000. We used a set of 1192 images from the Corel disks (disk #8) [6] (again using scenic, tourist, and people shots) and a set of 2500 randomly selected images from the web as validation sets. The performance on the validation sets determines which of the 120 SVMs parameter settings to use in practice.

---

[1] Note that any classifier could be used here (SVM, K-NN, AdaBoost, Decision Tree, etc.) We use an SVM as it consistently provided either the best or near-best performance across a variety of previous studies.

A. Results with and without feature reduction

In addition to training the SVMs with all 3930 features for each image, we also tried selectively removing some of the 15 simple transformations described in the previous section. For each of the variations tried, we retrained the SVM with the parameter settings that were determined to be the best using the full-set of features.

Table 1 shows the results of using all features and also with various subsets of features. Note that the accuracy % indicates the percentage of images for which the classifier worked correctly – it gave the maximal response when the image was rotated into an upright position. For each image, all four 90° orientations were tried.

Table 1: Results with feature selection on validation sets. Accuracy is the % of images for which upright rotation had maximal output.

Image Transformations	# Feat	Corel-Val(%)	Web-Val(%)
Full Set	3930	78.7	52.1
Only Edge Features	524	74.4	42.0
Only Intensity Features	262	68.7	43.7
Only YIQ & Normalized YIQ	1572	77.0	52.5
Only RGB & Normalized RGB	1572	74.9	49.4
No Variance (only Mean)	1965	76.7	48.0
No Normalized RGB or YIQ	2358	79.0	51.4
No Edge Features	3406	76.2	50.6
No Sub-Squares (only row & col)	1230	79.1	48.0
Subregions (1,2) (not 3,4,5,6)	270	75.8	47.3
Subregions (1,2,3,4,5) (not 6)	2490	78.0	53.1

It is interesting to note that the difference in performance between using all and various subsets of the features was quite small. Interestingly, only using intensity features (ie no color information) did not degrade performance drastically, although it did have the largest impact in this study. Removing the edge features from the full set had little impact on the accuracy; interestingly, using *only* edge features worked almost as well. With this limited sampling of results, it is difficult to find general trends. Nonetheless, the relatively small range of performance achieved with different feature sets indicates that much of the important information is contained in many of the features. For the remainder of this paper, our studies will use the entire feature set. For efficiency in a deployed system, however, a smaller subset may be employed.

IV. RESULTS ON INDEPENDENT TEST SETS

In this section, we perform large scale tests on two independent test sets. We also show how the system may be used in practice, describe easy rejection schemes for images that cannot be classified, and show where we expect good performance and where we expect poor performance.

When the system is used in the real world, it is likely that it will be commonly used for photographs that are taken with a digital camera. For this usage case, a photographer may hold the camera rotated ± 90°, but is unlikely to hold the camera

rotated 180°. Therefore, we can constrain our search to three orientations, 0 and ± 90°.

If this system is deployed in a photo management system, it should reject samples that it is unsure about rather than make incorrect rotations. A simple method to do this is to reject any images for which there is not exactly one orientation that triggers a response by the SVM > 0.0 (remember the SVM is trained to output positive for upright, and negative for other orientations). This provides a simple rejection scheme that requires no training.

A. Web Images

For this test set, we use 7,500 images that were randomly selected from the web that were not in the training or validation sets. They are divided into 3 groups, with decreasing entropy (as measured in the RGB space). The accuracies are shown in Table 2. Note that we examine the results with and without rejection schemes, and with examining 3 or 4 orientations.

Table 2: Accuracies on 7500 Web Images.

Decreasing Entropy Sets (2500 images each)	4 Orientations			3 Orientations		
	No Reject	With Reject		No Reject	With Reject	
		% Correct	Reject Rate		% Correct	Reject Rate
1	55.4	65.2	40.0	67.1	79.1	45.3
2	49.1	58.1	41.2	61.3	72.8	48.8
3	36.8	44.1	57.7	48.1	60.2	64.0

For general web images, the results are quite poor, ranging from 37-55% when four orientations are considered and 48-67% when 3 orientations are considered. Note that the performance decreases as the entropy in the images decreases; this is expected as very low entropy images will have few distinguishing markers. Further, on the web, low entropy images are often cartoons, drawing, charts, presentations screenshots, etc – all images that have no lighting information or cues for determining orientation other than deep understanding of image content. Using the simple rejection heuristic increases the correctly classified rate significantly; however, in this case the number of images that are rejected are quite large (40-60%). It should be noted, however, that this test set is difficult; these images are often not photographs, and even the photos are often difficult to recognize without captions.

B. Full Photo Set from Corel Disk #6

In this test, we examine all of the photographs contained on Corel's disk #6 [6]. There are a total of 15,888 images in 182 categories. We did not pre-select images for those that we felt would be most likely to be taken by amateur photographers nor by those that we felt would be most amenable to these techniques. One of the goals of this paper is to provide a concrete, comparable baseline to which future work can be measured. To this end, we have included results on all 182 categories in Table 3. Table 3 divides the 182 categories into

5 buckets, depending on accuracy of our algorithm. Examples of the images in each of the category are shown in Figure 2.

When processing this entire set with 4 orientations, the SVM correctly identifies the correct orientation as upright in 61.9% of the images. The simple rejection scheme described earlier rejects 35.6% of the images as not-classified, and yields a 76.1% correct rate on the remaining images. When examining only 3 orientations, the accuracy is 69.7%; the rejection scheme rejects 41.6% of the images with 85.6% accuracy on the remainder.
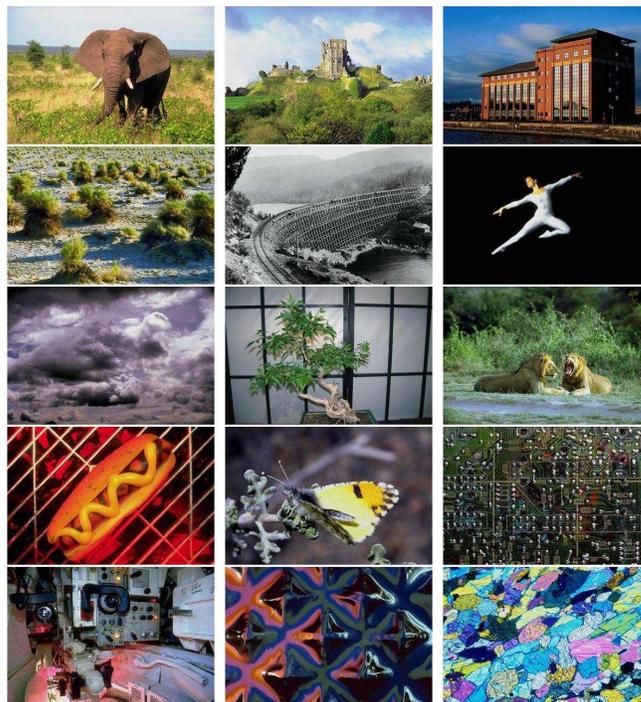
Figure 2: Three typical images from each bucket of performance shown in Table 3. Top row is highest performance, bottom row is lowest. Only horizontal images are shown here for layout purposes, the test set contains both orientations.

The best performers (accuracy=90-100% of the images detected in the upright orientation, measured without rejection with 4 orientations) are largely outdoor images taken of typical travel scenes, such as castles, etc. Most are fairly standard photos with traditional composition. Automated systems are expected to perform well here.

The second bucket (accuracy=70-90%) are categories very similar to the best performers. However, in these categories, there are more cloudy shots and shots where the ground is closer in color to the sky, for example in pictures of snow. Interestingly, there are also black and white photographs ("Canadian Historic Railroads") that fall into this performance category. This indicates that the detection is not reliant on color.

The third bucket (accuracy=50-70%) includes images that are indoors, contain close-ups of animals, pictures of people, birds in flight (with no other markers in the image), and

illustrations. These images are subject to errors because we do not do any deep image analysis.

The fourth and fifth buckets (accuracies < 50%) are the worst performing. Here, the images contain abstract patterns which have no *a priori* upright orientation, close ups, backgrounds and textures. Also included in these buckets are pictures of doors, which, not surprisingly, are not amenable to this type of approach. Here, the performance often degrades to random guessing (25%).

V. CONCLUSIONS

This paper has presented the results of a large scale test of automatic image orientation detection. The performance is acceptable for many outdoor images and more standard composition pictures that contain strong lighting and texture cues. For these classes of images, we match the previously reported accuracy rates. However, close-ups, illustrations, and abstract images are significant challenges to these approaches; unfortunately, these comprise a large portion of the images that need to be classified. The next step towards automating this procedure is employing approaches which attempt deeper understanding of the images, such as object detection. Detecting cars, upright faces and trees will provide a significant source of information for these automated systems.

REFERENCES
[1] Wang, Y. & Zhang, H. (2001), "Content-Based Image Orientation Detection with Support Vector Machines" in *IEEE Workshop on Content-based Access of Image and Video Libraries*, pp 17-23.
[2] Zhang, L, Li, M., Zhang, H (2002) "Boosting Image Orientation Detection with Indoor vs. Outdoor Classification", *Workshop on Applications of Computer Vision 2002*.
[3] Vailaya, A., Zhang, H., Yang, C., Liu, F., Jain, A. (2002) "Automatic Image Orientation Detection", *IEEE Transactions on Image Processing., 11,7.*
[4] Wang, Y., & Zhang, H. (2004) "Detecing Image Orientation based on low level visual content*" Comp. Vision and Image Understanding*, 2004.
[5] T. Joachims (1999) Making large-Scale SVM Learning Practical. *Advances in Kernel Methods - Support Vector Learning*, B. Schölkopf and C. Burges and A. Smola (ed).
[6] Corel-Gallery 1,300,000 (1999) Image Gallery – 16 Compact Disk Set – JB #40629.

Table 3: Performance on 182 Categories. Grouped by performance on 4-Orientation discrimination task with no rejection. Performance with Rejection and 3-Orientation task also given. Each category is shown as well as images correct / total images in each category.

% Correct (average)	Directory Name (# correct with 4-orientations/total images in directory)	4 orients With Reject		3 orients	3 orients With Reject	
		% Correct	Reject Rate %	No Reject	% Correct	Reject rate %
100-90 (94%)	Costumes 2/2; City_Life 5/5; Children_II 6/6; Boats_II 1/1; COASTAL 542/563; Architecture_V 82/86; Castles_of_Great_Britain 18/19; Castles 52/55; Castles_of_Europe_I 78/83; Churches_and_Cathedrals 49/53; Architecture_III 61/66; Architecture_VIII 82/89; Architecture_I 56/61; Austria 55/60; Canals_and_Waterways 21/23; Architecture_VII 73/80; coastal2 361/397; Eastern_Europe 30/33	97	8.6	95	98	9.6
90-70 (81%)	British_Royalty 8/9; Architecture_IX 70/79; ASIANARC 377/427; Architecture_VI 72/82; Contemporary_Buildings 79/90; Canadian_Rockies 86/98; Canada_An_Aerial_View 84/96; China_and_Tibet 39/45; Bridges_II 78/91; Architecture_X 64/75; Architecture_IV 58/68; Alaska 69/81; Ancient_Architecture_III 74/87; Elephants 85/100; Dawn_and_Dusk 85/100; Cities_of_Europe_I 56/66; Beaches 27/32; Arizona 58/69; Alien_Landscapes 84/100; Architecture_II 61/73; Ancient_Architecture_II 71/86; Copenhagen_Denmark 32/39; Bridges_III 54/66; Death_Valley 76/93; American_National_Parks 76/93; Ancient_Architecture 67/82; ARCHITEC 240/295; Bali_Indonesia 43/53; Devon_England 59/73; American_Wilderness 79/98; Egypt 57/71; Air_Travel 4/5; California_Coasts 75/94; Canada 43/54; Canadian_Farming 62/78; Animals 68/86; Denmark 33/42; Desert_Scenes 64/82; Czech_Republic 39/50; Coast_of_Norway 28/36; Antiquity 7/9; Agriculture_II 49/63; Alien_Landscapes_II 57/74; African_Specialty_Animals 77/100; Australia 60/78; Coastal_Landscapes 56/73; Croatia 48/63; African_Wildlife 76/100; Boston 28/37; Acadian_Nova_Scotia 34/45; Canadian_Historic_Railways_National_Archives_of_Canada 75/100; Autumn_in_Maine 54/72; Clouds_I 70/94; California_Parks 74/100; Belgium_and_Luxembourg 37/50; Arizona_Desert 71/100; Architectural_Details 51/69; Berlin 25/34; Colorado_Plateau 71/97; Circus_Fairs_and_Amusement_Parks 19/26; Cheetahs_Leopards_and_Jaguars 73/100; Canada_East_Coast 39/55; Aboriginal_People_-_National_Archives_of_Canada 70/100	89	20	86	94	24
70-50 (59%)	African_Antelope 69/99; Children_of_the_World 9/13; Cities_of_Italy 37/54; Beautiful_Bali 26/38; Bhutan 35/52; British_Columbia 39/58; Canadian_National_Parks 48/72; Backyard_Wildlife 66/100; China 25/38; Agriculture 47/72; Construction 15/23; Dolphins_and_Whales 65/100; Endangered_Species 64/100; Decorative_Hand-painted_Scenes 64/100; Clouds 64/100; Caribbean 30/47; Costa_Rica 47/74; Canoeing_Adventure 19/30; Chicago 36/57; Ancient_Carvings_and_Design 55/89; Couples_II 14/23; animals2 360/595; Dogs 59/99; Birds_IV 58/98; Bonsai_and_Penjing 59/100; Autumn 59/100; Alaskan_Wildlife 59/100; Creatures_III 53/90; Bears 57/100; Birds_V 54/95; Architectural_Details_II 38/67; Dinosaur_Illustrations 56/100; Birds_III 56/100; Bird_Illustrations 56/100; Creatures_I 54/98; Cougars 55/100; Creatures_II 50/91; Beverages 54/100; Creatures_V 46/86; Art_Crafts_and_Design_I 10/19; Creatures_IV 51/98; African_Birds 51/100; Dance 4/8; Cowboys 7/14; Couples 10/20; Brazil 19/38; Air_Travel_III 1/2	70	37	69	83	45
50-30 (40%)	closeup2 225/452; Commercial_Construction 25/51; Dog_Sledding 24/49; Christmas 33/68; Children 8/17; Alligators_Crocodiles_and_Reptiles 47/100; Doors_of_San_Francisco 46/99; Decorated_Pumpkins 30/68; Animals_Closeup 43/100; Communication_and_Technology 21/50; Christmas_Celebration 31/75; Desserts 41/100; Abstracts_and_Patterns 41/100; Army 2/5; Annuals_for_American_Gardens 39/100; Cuisine 38/100; Butterflies 38/100; Backgrounds_II 36/96; Apes 36/100; Backgrounds_and_Textures_V 34/96; Color_I 29/82; Doors_of_Paris 35/100; Caves 7/21; Ballet 3/9; Caverns 32/100; Arthropods 31/100; Abstract_Textures 31/100; Color_Backgrounds 30/100; Cactus_Flowers 30/100	48	53	52	64	63
< 30 (25%)	Color_Backgrounds_II 28/96; Backgrounds_and_Textures_III 29/100; Backgrounds_and_Textures_I 29/100; Agates_Crystals_and_Jaspers 29/100; Backgrounds_and_Textures_IV 28/99; Contemporary_Fabric 28/100; Agates 28/100; Butterflies_II 27/100; Beautiful_Roses 26/100; EMS_Rescue 1/4; Artist_Textures 25/100; Creative_Textures 24/100; Backgrounds_and_Textures_II 24/100; Barbecue_and_Salads 23/98; Crystallography 23/100; Creative_Crystals 23/100; Backgrounds_I 12/53; Abstract_Color 21/93; Cards 22/100; Bark_Textures 22/100; Abstract_Designs 22/100; Colors_and_Textures 21/100; Beads 21/100; Air_Travel_IV 0/1	27	69	34	37	76