IMAGE ORIENTATION DETECTION WITH INTEGRATED HUMAN PERCEPTION CUES (OR WHICH WAY IS UP)

Lei Wang[†], *Xu Liu*[†], *Lirong Xia*[†], *Guangyou Xu*[†], *Alfred Bruckstein*[‡]

[†]: Computer Department, Tsinghua University, Beijing, 100084, P.R.China
[‡]:Computer Science Department, Technion, Haifa 32000, Israel

ABSTRACT

In this paper, we propose a set of human perceptual cues used jointly to automatically detect image orientation. The cues used are: orientation of faces, position of the sky, brighter regions, and textured objects, and symmetry. We combine these cues in a Bayesian framework, and the photo acquiring model has been considered carefully as the prior knowledge of the image orientation. Results on more than a thousand different images provide a compelling argument that our approach is a viable one.

1. INTRODUCTION

With the increasing popularity of personal imaging devices such as cameras, scanners, digital cameras, and more recently, mobile devices with imaging capabilities, managing captured images is fast becoming an important issue that needs to be addressed. Because images can be captured at any orientation, having an automatic orientation detector would be a very useful tool in any image management system. Any image shown can have the option of being prewarped to an upright orientation for viewing. In addition to image management systems, the technique for automatic orientation detection would also be very useful for other applications that involve object detection and content-based image retrieval.

Image orientation adjustment is usually done manually. Most of the related prior work describe algorithms that detect the orientation of documents [1] and medical images [2]. An algorithm for automatic image orientation detection using a Bayesian learning framework was presented in [3, 4]. This algorithm is based on a training set with 1,995 image samples, and the features for classification are the spatial color moments. SVM [5] and boosting [6] based algorithms were proposed as well. All these algorithms were based on a learning and clustering framework, and assumed that the image orientation can only be 0°, 90°, 180°, or 270°.

Our approach does not impose such angular restrictions. We approach the orientation detection problem by considering human perceptual processes, rather than learning features and classifications independently of such considerations. Humans can easily perceive the orientation of an image [7, 8]. A human tends to find prior knowledge as cues to determine image orientation, such as orientation of a human face, depth cues, textures, and relative positions of sky and ground.

Although these cues are typically not sufficient in isolation (and the fact that some of the cues may be missing), the integration of all these cues can provide considerably more accurate orientation for a lot of images with a full range of rotation from 0° to 360°. While the integration could be a simple voting for orientation, we use Bayesian inference here to combine the outputs of different human perceptual cues. The photo acquiring model, which is the human's behaviour for taking or scanning photos, has been considered carefully in this framework as the prior knowledge of the image orientation.

The rest of the paper is organized as follows. In Section 2, we discuss the various visual perceptual cues used in our formulation. They are: orientation of faces, position of the sky, brighter regions, textured objects, and symmetry. This is followed by a description of how these cues are integrated in a Bayesian framework (Section 3). We show results of experiments in Section 4, discuss relevant issues in Section 5, and provide concluding remarks in Section 6.

2. HUMAN PERCEPTION CUES

Given an image, a human tends to extract relevant perceptual cues to infer the orientation of the image. These cues can be typically classified into three categories:

1. Objects with distinguishable orientation, such as human faces, human bodies, trees, animals and some written or printed characters;
2. Objects with a usually fixed position, for example, the sky is usually on the top, while the ground is usually at the bottom;
3. Low level features, such as light, texture, symmetry, edges, and segments. As an example, brighter things are considered to be up, while a larger segment is assumed to be below a smaller one for structural stability.

The first two categories are called high-level cues, since they are related to high-level semantics.

In this section, we investigate some of the cues mentioned above, which cover all three categories. For a given image, we use Θ to denote its orientation. The clockwise orientation for the correct "up" image is 0°. Assuming we have n cues, if Θ is given, we can evaluate the likelihood of the image (observation) for cue C_i under the given orientation $\Theta = \theta$. We denote it with $P(C_i|\Theta), (i = 1, \cdots, n)$.

2.1. Human Faces

Human faces have a very distinguishable orientation, as well as detectable. Many methods have been proposed to detect faces in an image. Here, an algorithm based on both template matching and support vector machine is used for face detection [9]. Any reliable face detection algorithm can be used; even if it is not rotation

This research is supported by Chinese NSF grant No. 60273005.

invariant, a good algorithm is usually tolerant of some slop. We rotate the image to various hypothesized orientations and then search only for upright faces. For each orientation, the output of the algorithm is the maximum reliability of the faces detected. This distribution is regarded as the likelihood of the image orientation wrt human faces. The reliability is between 0 and 1. The distribution of reliability is normalized to represent $P(C_1|\Theta)$, as shown in Figure 1.

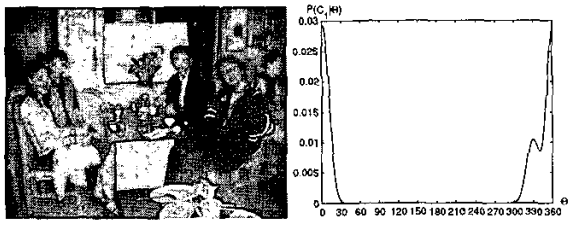

Fig. 1. Sample for face: Left: Original image; Right: Distribution for $P(C_1|\Theta)$.

2.2. Sky

In images that involve the sky, the sky is usually on top. As a result, the location of sky can be used to determine the orientation of an image. Vailaya [10] used color, texture and position to train a classifier for the sky. In our case, since position is highly related with the image orientation, we only use the color and texture. Two additional assumptions are made to decrease the false detection: 1. Sky regions are connected; 2. Sky regions must extend to some edge of the image, i.e., sky regions should not be surrounded by non-sky regions. These assumptions are reasonable for most images with sky.

After the sky area is detected, a given image orientation is validated by projecting sky and non-sky area points to the orientation, and evaluating how much the sky area points is on top. This is done by calculating

$$E_{sky} = \sum_{allpoints} \frac{S_p * H_p}{1 + \sigma_{Sp}},$$

where $S_p = 1$ if the point is belong to sky, and $S_p = 0$ otherwise. H_p is the height of the point, which is the point's projection on the orientation. σ_{Sp} is the variance for S_ps within the same height. As a result, E_{sky} will be maximum when all the sky points are on the top without non-sky points at the same height. When $E_{sky} \leq 0$, it is set to a very small ϵ. $P(C_2|\Theta) = E_{sky}$ after normalization. An example is shown in Figure 2.

2.3. Light

Humans tend to perceive brighter parts of the image to be on top. This is easy to understand if we notice that naturally most light sources come from above, such as sunlight, moonlight and light fixtures. This motivates us to evaluate the likelihood of orientation by examining how much the brighter parts are on the top. This is our third cue: $P(C_3|\Theta)$. Similar to the sky cue, $P(C_3|\Theta)$ is calculated by normalizing

$$E_{light} = \sum_{allpoints} \frac{I_p * H_p}{1 + \sigma_{Ip}}$$

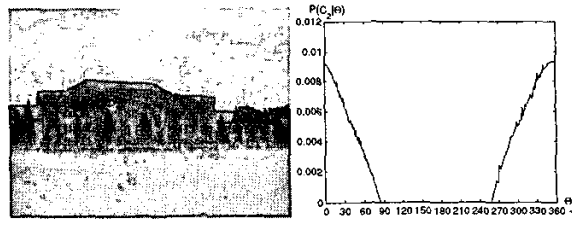

Fig. 2. Image with sky example. Left: Original image; Right: Distribution for $P(C_2|\Theta)$.

where I_p is the intensity of the point represented by its gray scale, H_p is the height of the point represented by its projection on the orientation, σ_{Ip} is the variance of the I_ps within the same height. Figure 3 shows an example where this cue predominates.

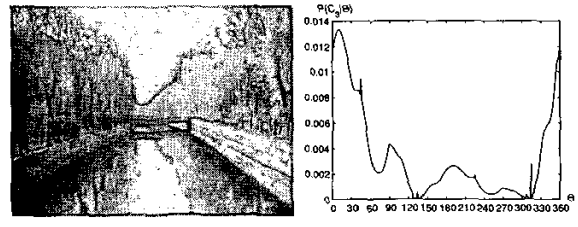

Fig. 3. Sample for dark and bright: Left : Original image; Right : Distribution for $P(C_3|\Theta)$.

2.4. Texture

Another characteristic of human perception is that textured areas are usually regarded to be at the lower part of the image. Just like sky for outdoor images and ceiling for indoor images, this behavior is expected because of gravity. As it is much easier to stay on the ground than to float in the sky, the part near the ground, usually corresponding to the lower part of the image, tend to have more variety of objects. So the texture is another cue for inferring the image orientation such as sky and light. Here the classic co-occurrence matrices [11] are used to describe the textured level of a point. The texture cue $P(C_4|\Theta)$ is then computed by normalizing

$$E_{texture} = \sum_{allpoints} \frac{T_p * H_p}{1 + \sigma_{Tp}}$$

where T_p is the texture measurement for the point, H_p is the height of the point represented by its projection on the orientation, σ_{Tp} is the variance of the T_ps within the same height. T_p is larger for textureless area. An example for texture cue is shown in Figure 4.

The evaluations with sky, light and texture are very similar. In fact, they have a uniform physical explanation. When we hang up an object, the object rotates automatically to the orientation determined by the hanging point and the gravity center of the object. Here we take different cues as the weight, and the calculation is just similar to computing the gravity center wrt the cue. In this way we can tell the right orientation using the gravity center.

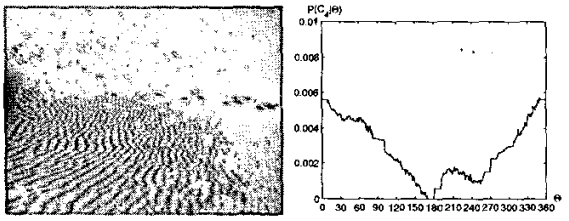

Fig. 4. Sample for texture: Left : Original image; Right : Distribution for $P(C_4|\Theta)$.

2.5. Symmetry

Many natural and man-made things are symmetric. Although some of them have multiple axes of symmetry, objects (such as human faces, human bodies, houses, and trees) which often appear in an image usually have only one symmetry axes along their upright orientation. In addition, people also inherently tend to interpret the image symmetrically. This cue can also be used to infer the orientation of an image. For each assumed orientation, we scan the symmetry axis along the orientation, and for each axis calculate the "symmetric distance." The "symmetric distance" is defined as the sum of squared intensity distance for points symmetrically neighbouring the axis within a window. All the symmetric distances are then added up and normalized to evaluate the likelihood between the image symmetry and the orientation. This is $P(C_5|\Theta)$, with an example shown in Figure 5. Obviously, this cue can not tell the difference between $0°$ and $180°$; this ambiguity can be resolved using other cues.

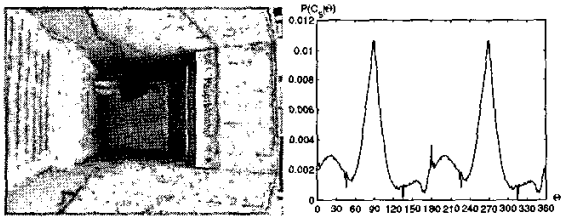

Fig. 5. Sample for symmetry: Left : Original image; Right : Distribution for $P(C_5|\Theta)$.

All the cues above are obviously not absolutely right, nor exhaustive. Natural images are so varied that each of the cues may fail in isolation. We integrate the outputs of all these cues because it is much less likely that all these cues will fail simultaneously, thus providing a degree of robustness against noise and image variation.

3. BAYESIAN INTEGRATION

When multiple cues are applied, the inference problem is to find the best orientation that have maximum likelihood with the cues. Assuming conditional independency, we have

$$\begin{aligned}\Theta &= \arg\max_{\Theta} P(\Theta|C_1, C_2, \cdots, C_5) \\ &= \arg\max_{\Theta} P(C_1|\Theta)P(C_2|\Theta)\cdots P(C_5|\Theta)P(\Theta)\end{aligned}$$

It is interesting to note that the role of $P(\Theta)$ in the equation. It can be regarded as the cue from camera model, which is the human's behavior for taking or scanning photos. Usually, photos are taken in an upright orientation, and sometimes the orientations are $90°$ or $270°$. The orientation may be off by a few degrees from these three typical directions. The possibilities for the other degrees are quite small. Like production model has been widely used in video analysis, photo acquiring model is also quite useful for image orientation detection. The distribution is given empirically, as shown in Figure 6.

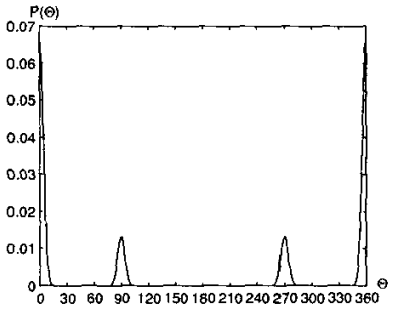

Fig. 6. Prior distribution for image orientations (e.g., $P(\Theta)$).

4. EXPERIMENTS

1287 images were used to evaluate the system. These images are collected from personal albums taken by digital cameras or scanned from photos. As a result, these images are highly varied.

Table 1 shows the results using different cues and the Bayesian integration. The estimated orientation is regarded to be correct for the $360°$ case if it is within $3°$ of the ground truth. For the four directions case, estimated orientation is approximated to its nearest neighbor in $\{0°, 90°, 180°, 270°\}$. It is clear that high-level cues such as faces and the sky are more accurate and stable, but they are limited to images that include them. Figure 7 shows some of the successful examples, while some of the images we failed are shown in Figure 8. Note that all the results are obtained without any complicated training.

Cue	360 degrees (%)	4 directions (%)
Face	92.9	94.5
Sky	90.1	92.2
Light	82.1	84.9
Texture	81.3	84.0
Symmetry	87.5	89.3
Integrated	92.6	94.1

Table 1. Detection rate with different cues and Bayesian integration. The ground truth of right orientation is given by human. Note that the detection rates for face and sky are calculated according to images with faces and sky separately.

From the results above we can see that the idea of using human perceptual cues to detect the image orientation is a powerful one.

Fig. 7. Images whose orientation is correctly detected. Detected orientation: Left Top: 90°, Right Top: 0°, Left Bottom: 356°, Right Bottom: 0°.

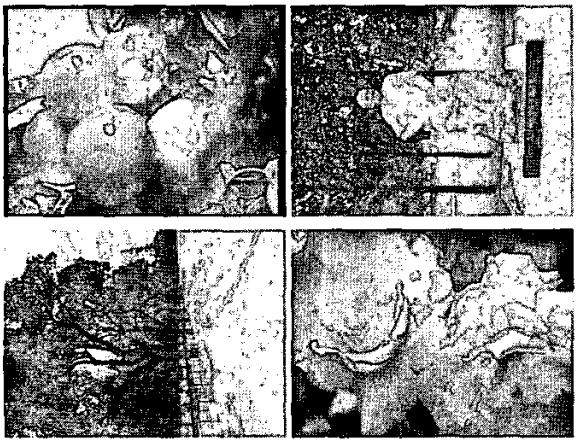

Fig. 8. Images whose orientation is incorrectly detected. Detected orientation: Left Top: 260°, Right Top: 90°, Left Bottom: 0°, Right Bottom: 109°.

5. DISCUSSION

Future improvement is obviously necessary, including developing new cues as well as improving the detection rate for existing cues. We are trying to make some comparison with existing learning based image orientation detection algorithms, as well as to incorporate some learning factors to improve the performance of our algorithm. We may also extend the idea used here to a solid human perception theory.

Each cue has its own limitation. When an image is about a close-up of a flower or some simple object, or it is something artistic, all the cues above may fail. In fact it is hard even for a human to tell the correct orientation. Here we mainly focus on images and photos taken in daily life. This is encouraged by Chang's [12] comments about multimedia retrieval, i.e., research work should be done for data sets with large quantities but low unit price. That is why we design our system to work for images taken in everyday life.

6. CONCLUSIONS

Image orientation detection is useful for object detection, content based retrieval, and image database management. Cues motivated by human perceptions are proposed and integrated by Bayesian inference to detect the orientation of an image. Without any complicated training, the proposed method has achieved a high orientation detection rate for a large number of real images. This is a compelling evidence that our technique is a good one.

7. ACKNOWLEDGMENT

The authors would like to thank Haizhou Ai and Bo Wu for providing the face detection program. Thanks go to Sing Bing Kang, Lei Zhang, Steve Lin, Feng Jing, Chang Yuan, Xiaoliang Wei as well for valuable proofreading and discussions.

8. REFERENCES

[1] R.S. Caprari, "Algorithm for text page up/down orientation determination," *Pattern Recognition Letters*, vol. 22, no. 4, pp. 311–317, 2000.

[2] M.G. Evanoff and K.M. McNeill, "Computer recognition of chest image orientation," in *11th IEEE Symposium on Computer-Based Medical Systems*, 1998, pp. 275–279.

[3] A. Vailaya, H.-J. Zhang, and A. Jain, "Automatic image orientation detection," in *ICIP'99*, Kobe, Japan, October 24-28 1999, vol. 2, pp. 600–604.

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

[5] Y.M. Wang and H.-J. Zhang, "Content-based image orientation detection with support vector machines," in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Hawaii, December 2001, pp. 17–23.

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

[7] I.P. Howard and W.B. Templeton, *Human Spatial Orientation*, Wiley, New York, 1966.

[8] I. Rock, *Orientation and Form*, Academic Press, New York, 1974.

[9] H.Z. Ai, L.H. Liang, and G.Y. Xu, "Face detection based on template matching and support vector machines," in *ICIP'01*, Thessaloniki, Greece, October 2001, pp. 1006–1009.

[10] A. Vailaya and A. Jain, "Detecting sky and vegetation in outdoor images," in *SPIE: Storage and Retrieval for Image and Video Databases VIII*, San Jose, CA, January 2000, vol. 3972.

[11] L.S. Davis, M. Clearman, and J.K. Aggarwal, "An empirical evaluation of generalized cooccurrence matrices," *PAMI*, vol. 3, no. 2, pp. 214–221, March 1981.

[12] S.F. Chang, "The holy grail of content-based media analysis," *IEEE Multimedia*, vol. 9, no. 2, pp. 6–10, April/June 2002.