# SUNSET SCENE CLASSIFICATION USING SIMULATED IMAGE RECOMPOSITION

*Matthew Boutell*

Department of Computer Science
University of Rochester
boutell@cs.rochester.edu

*Jiebo Luo, Robert T. Gray*

Electronic Imaging Products, R & D
Eastman Kodak Company
{luo, gray}@image.kodak.com

## ABSTRACT

Knowledge of the semantic classification of an image can be used to improve the accuracy of queries in content-based image organization and retrieval and to provide customized image enhancement. We developed an exemplar-based system for classifying sunset scenes. However, the performance of such a system depends largely on the size and quality of the set of training exemplars, which can be limited in practice. In addition, variations in scene content, as well as distracting regions, may exist in many testing images to prohibit good matches with the exemplars. We propose using simulated spatial and temporal image recomposition to address such issues. The recomposition schemes boost the recall of sunset images from a reasonably large data set by 10%, while holding the false positive rate constant.

## 1. INTRODUCTION

Automatically determining the semantic classification (e.g., sunset, picnic, beach) of an arbitrary image is a difficult problem. Much research has been done recently, and a variety of classifiers and feature sets have been proposed. The most common design for such systems has been to use low-level features (e.g., color, texture) and statistical pattern recognition techniques. Such systems are exemplar-based, relying on learning patterns from a training set [1, 2].

Semantic scene classification can improve the performance of content-based image organization and retrieval (CBIR). Many current CBIR systems allow a user to specify an image and search for images similar to it, where similarity is often defined only by color or texture properties. This so-called "query by example" has often proved to be inadequate [3]. Knowing the category of a scene *a priori* helps narrow the search space dramatically, reducing the search time, increasing the hit rate, and lowering the false-alarm rate [4].

Scene classification can also find application in image enhancement. Rather than applying generic color and exposure adjustment to all scenes, we can customize them to the scene, e.g., retaining or boosting brilliant colors in sunset images while removing warm-colored cast from tungsten-illuminated indoor images.

In the literature, sunset classification is limited to small, constrained data sets. Vailaya *et al.* reported 94.3% accuracy on the constrained problem of separating 155 sunset images from 373 forest and mountain scenes, which were chosen because they were unambiguous in their scene content compared to sunset scenes [2]. Belongie *et al.* [1] reported 89% accuracy on a limited data set (approximately 60 training images and 30 testing images from each of twelve classes). Smith and Li [4] reported 86.1% accuracy on a small dataset of 91 training images (including 10 sunsets) and 266 testing images (including 36 sunsets). It is unclear how well any of the above systems would generalize for a large data set. We are aware of no results on larger data sets in the literature (except within a CBIR framework, which does not translate directly into classification accuracy).

What are the obstacles to accurate scene classification and sunset detection in particular? The primary obstacle appears to be the tremendous variety of images found within most semantic classes. Exemplar-based systems must account for such variation in their training sets. Even hundreds of exemplars do not necessarily capture all of the variability inherent in some classes, like that of sunset images. Sunset images captured at various stages of the sunset can vary greatly in *color*, as the colors tend to become more brilliant as the sun approaches the horizon, and fade as time progresses further. The *composition* can also vary due in part to the camera's field of view: Does it encompass the horizon or the sky only? Where is the sun relative to the horizon? Is the sun centered or offset to one side?

A second reason for limited success in exemplar-based classification is that images often contain excessive or distracting foreground regions (Fig. 1), which cause the scene to look less prototypical and thus not match any of the training exemplars well. This is especially true in consumer images, where the typical snapshooter pays less attention to composition and lighting than would a professional photographer. Therefore, consumer images contain greater variability, causing the high performance (on professionally taken stock photo libraries such as the Corel database) of many existing systems to decline when used in this domain.

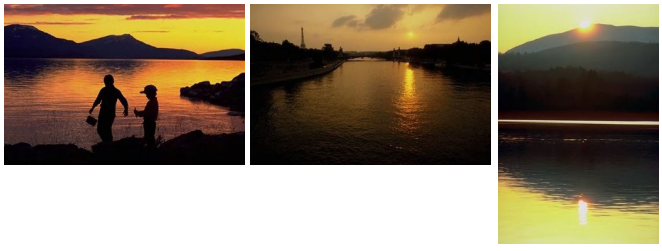

**Fig. 1**. Sunset images with distracting foreground regions.

We address each of these issues by introducing the concept of *spatial image recomposition*; designed to minimize the impact of undesirable composition (e.g., foreground objects), and of *simulated temporal image recomposition*, designed to minimize the effect of color changes occurring over time.

## 2. METHODOLOGY

We now present the features and classifier used in our baseline system, as well as the concept and use of image recomposition for boosting the system performance.

### 2.1. Baseline system

In the hierarchical image classification scheme in [2], sunsets were easily separated from mountain and forest scenes. Color was found to be more salient for the problem than other features, such as edge direction alone, confirming our intuition that sunsets are recognizable primarily by their brilliant, warm colors. Furthermore, we believe spatial information should be incorporated to distinguish sunsets from other scenes containing warm colors, such as those of desert rock formations and fireworks. Therefore, we used spatial color moments, dividing the image into 49 regions using a 7 x 7 grid and computing the mean and variance of each band of an Luv-transformed image. This yields 49 x 2 x 3 = 294 features.

We used a Support Vector Machine (SVM) as the classifier because SVMs have been shown to give higher performance than other classifiers, such as Learning Vector Quantizers (LVQ) on similar problems [5]. We used a Gaussian kernel, creating an RBF-style classifier. SVMs are designed for two-class problems, such as ours, and output a real number for each testing image. The sign is the classification and the magnitude can be used as a loose measure of the confidence.

### 2.2. Image recomposition

One of the largest obstacles to high performance in semantic scene classification is the *immense* variety, both in terms of color and composition, of images in each class. Because manually collecting large numbers of prototypical images to capture this variability is time consuming and still may not be comprehensive, it is critical to make efficient use of all available training data.

Furthermore, the best match of a testing image with the set of training exemplars occurs when the image matches an exemplar of its class, e.g., both in its color and its composition. However, the test image may contain variations not present in the training set. For instance, the degree of match is affected by the photographer's choice of *what* to capture in the image (affecting its composition) and *when* to capture the image (potentially affecting its color because of changes in the scene illuminant over time). If it were possible to "relive" the scene, one could attempt to derive images with more prototypical color and composition for the class (Fig. 2). How can one “relive” the scene? In other words, how can one transform a testing image into one that will match a prototypical exemplar better or produce more prototypical exemplars for training?

We propose a concept called *simulated spatial and temporal recomposition* to address the above issues. The different types and uses of spatial recomposition (mirroring and cropping images) and simulated temporal recomposition (shifting the color of images) will be elaborated in detail. While these types of image recomposition are by no means exhaustive, we found them to work well for sunset detection. We now detail the schemes we used to boost sunset classification performance.

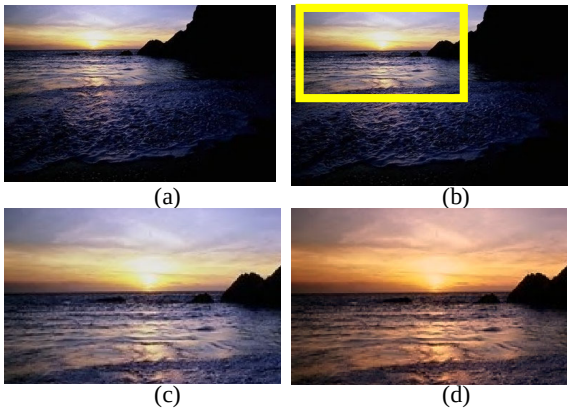


**Fig. 2.** “Reliving the scene.” The original scene (a) contains a salient subregion (b), which is cropped and resized (c). Finally, an illuminant shift (d) is applied, simulating a sunset occurring later in time.

### 2.3. Boosting performance using recomposition schemes

Using recomposition on a limited-size set of training data can yield a richer, more diverse set of exemplars. Our goal is to obtain these exemplars in an unsupervised manner, i.e., without having to inspect each image. One technique is to flip each image horizontally, assuming the image is in its correct orientation, thereby doubling the number of exemplars. Clearly, the classification of the new image is unchanged, e.g., flipping a sunset image with the sun on the left side of the image moves the sun to the right side, but it is still just as valid a sunset (Fig. 3).

Another technique is to crop the edges of an image. The assumption is that the salient portion of an image is in the center, and imperfect composition is caused by distractions in the periphery. Cropping from each side of the image, in turn, produces four new images of the same classification. Of course, one does not want to lose a salient part of the image (such as the sun or the horizon line in a sunset). If we conservatively crop a small amount, e.g., 10%, the *semantic* classification of a scene is highly unlikely to change, although the classification by an algorithm may change.

**Fig. 3.** Spatial recomposition. The original image (center) can be transformed by horizontal mirroring (left) or cropping (20% from bottom shown on right).

If still more training data is needed, one may use more aggressive recompositions, such as larger amounts of cropping or significant color shifts, however, care must be taken to screen out those images in which the recomposition has caused the image to lose salient scene content.

While recomposing the training set yields more exemplars, recomposing a test image, and classifying each new, recomposed image yields multiple classifications of the original image. In terms

of spatial recomposition, the edges of the image can be cropped in an attempt to better match the features of a test image against the exemplars. We may need to crop more aggressively (as in Fig. 2) to obtain such a match. However, if the classifier has been trained using mirrored images, there is no need to mirror the test image because symmetry is already built into the classifier.

Some classes of images, like sunsets, contain a large variation in their *global* color distribution. Shifting the overall color of the test image appropriately can yield a better match with a training exemplar. For example, an early and a late sunset may have the same *spatial* distribution of color (bright sky over dark foreground), but the overall appearance of the early sunset is much cooler, as a result of the color change in the scene illuminant. By artificially changing the color along the illuminant (red-blue) axis [6] toward the warmer side, we can simulate the appearance of capturing the image later in time. We refer to this illuminant shift as *simulated* temporal recomposition (Fig. 4). Likewise, variation in the *magnitude* of illuminant in the scene can be handled by a shift along the luminance axis. Color shifts along other axes may be applicable in other problem domains.

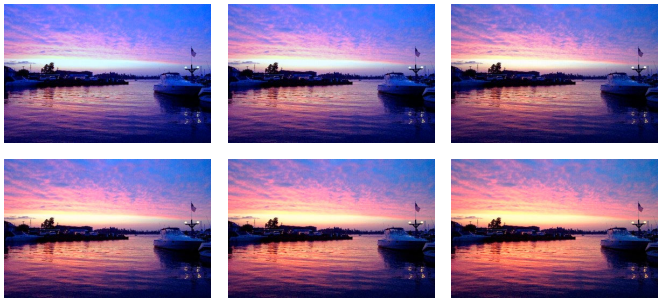

**Fig. 4.** Temporal recomposition. A series of illuminant shifts (along red-blue axis) in 1.5-stop increments, starting from –3 stops and ending at +4.5 stops, is shown. A *stop* is a photographic term equal to a 2X change in exposure or magnitude [6].

Whether using spatial or temporal recomposition, the classifier may or may not label a new, recomposed image with the same class as the original image. When the classifications of the recomposed images differ, we need to adjudicate using a meta-classifier. In a two-class problem, we found that one interesting meta-classifier of $r$ recompositions is to use the $m^{th}$ order statistic, e.g., the maximum ($m = 1$), the second largest ($m = 2$), or the median ($m = r/2$). Varying the parameter $m$ moves the classifier's position on the operating curve. Small $m$ classifies images positively in an aggressive manner, giving greater recall at the expense of more false positives. The choice of $m$ will clearly depend on the application.

Recomposition in testing is useful even when it has been used in training, because there is no guarantee that recomposition in training exemplars has exhaustively created all possible variations and completely filled the image space.

## 3. RESULTS

We experimented with 3 databases, as shown in Table 1. Database D1 was used as a benchmark for our baseline system: sunset and non-sunset images similar to those in [2] (in number and content) were chosen. We achieved comparable results to ensure that our baseline classifier is as good.

**Table 1.** Sunset Training and Testing Sets
(S: sunset, NS: non-sunset, C: confusing)

| Database | Size | S:NS | S:C | Source |
|---|---|---|---|---|
| D1 | 577 | 1:2 | 1:0 | Corel |
| D2 | 1342 | 1:5 | 2:1 | Corel + personal |
| D3 | 2085 | 1:9 | 1:0 | Consumer |
| TRAIN | 1766 | 1:5 | 2:1 | Corel + personal |

Database D2, consisting of a mix of consumer and stock photos, is a more challenging set to classify. We included 200 images with content chosen *purposefully* to push the limits of the system (such as fire, moon, fireworks, tungsten lighting, and red/orange flowers), giving a ratio of sunset to confusing images of 2:1. We also included a wide variety of non-sunset images, in an attempt to more realistically model the space of consumer and professional photos, giving an overall 1:5 ratio of sunset to non-sunset images.

The training set of 1766 images (*independently* derived from the same sources as D2) was enriched by spatial recomposition. For each image, we trained on the original image, four images generated by cropping 10% off each side of the image in turn, and the mirror images of each of the five, yielding ten times as many training images.

When testing with images from D2, we used spatial recomposition as well, classifying the original image and four each of 10%-cropped and 20%-cropped images. We combined them using the second highest score of the nine generated. It is less prone to outliers than the maximum score, which gives highest recall and highest false positive rates.

In an effort to correctly classify the subtle sunsets, we also incorporated simulated temporal recomposition in testing. We repeated the combination of nine spatially recomposed images twice, once on the original image, and once on the illuminant-shifted image (simulating an illuminant shift of +3 stops on the original image). The maximum of the two scores was used.

Performance is evaluated using three standard measures. *Accuracy* is defined as ratio of correct classifications to total images. *True positive rate* (TPR) is defined as the ratio of sunsets correctly classified (hits) to the total number of positive images (sunsets). *False positive rate* (FPR) is defined as the ratio of false positives to the number of negative (non-sunset) images. Table 2 shows how different recomposition schemes affected each of these measures.

**Table 2.** Effects of Image Recomposition.

| Dataset | Recomposition | TPR | FPR | ACC |
|---|---|---|---|---|
| D2 | None (baseline) | 74.8% | 3.8% | 92.7% |
| D2 | Training only | 74.8% | 2.0% | 94.2% |
| D2 | Testing only | 81.1% | 5.5% | 92.3% |
| D2 | Both | 85.1% | 3.8% | 94.3% |
| D2 | Both + temporal | 89.6% | 8.8% | 90.9% |
| D3 | Both | 71.3% | 0.5% | 96.4% |
| D3 | Both + temporal | 83.8% | 7.2% | 91.3% |

Using recomposition in the training set increased accuracy significantly, presumably because the set was much richer. This overcomes some of the effects of having a limited training set. Using recomposition in the testing set increased both the number of hits and the number of false positives. Finally, using

recomposition in both training and testing gave the best results overall. The best scenario gave an increase in a true positive rate of 10%, while holding the false positive rate constant

Using spatial recomposition on the testing images enabled us to meet our goal of correctly classifying sunset images with large distracting foreground regions: the images presented in Figs. 1 and 2 were all classified incorrectly by the baseline system, but were classified correctly when recomposition was used (gained by recomposition). The image shown in the middle of Fig. 1 is a good example of how recomposition by cropping can help. Cropping the large, dark, water region in the foreground from the image increases the SVM score substantially. The other images fared similarly, e.g., cropping the bottom 20% from the image on the right eliminated the confusing reflection in the water.

However, the number of false positives also increased, partially offsetting the gain in recall. Typical false positives induced by recomposition are shown in Fig. 5. Each image contains patterns not typical of sunsets (e.g., the multiple bright regions in the night images or the sky in the desert image), which, when cropped out, make the image appear much more sunset-like.

**Fig. 5**. False positives induced by spatial recomposition.

Some sunset images have prototypical composition, but weak colors, corresponding to early or late sunsets. Shifting the scene illuminant "warms up" these images, causing them to be classified correctly, but also introduces many false positives, both of which are shown in Fig. 6.

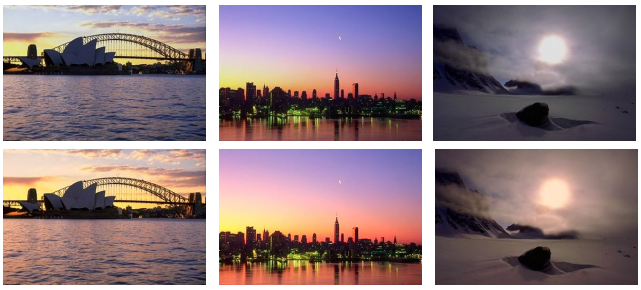

**Fig. 6.** Sunsets and false positives induced when temporal recomposition are used. The original (top) and illuminant-shifted (+3 stops) images (bottom) are shown. Note that the image on the far right is one of the purposefully confusing images: a winter scene with the sun low in the horizon, but not setting.

Sample images that were misclassified are presented in Fig. 7. The missed sunset images are understandable, given the features used. Some images contain distractions either in the middle or on both sides of the image, while others contain weak or non-typical color. Among the false positives, one can see a number of the purposefully chosen, confusing, non-sunset images. Some could possibly be avoided by incorporating other features, such as edge direction, but others, such as the fire and ballet scenes, may be nearly impossible to disambiguate.

The accuracy on D3 was excellent because we correctly rejected most of the non-sunset images. Our TPR was somewhat lower because of many subtle sunsets in the collection. Temporal recomposition helped gain many of the early sunsets.

## 4. CONCLUSIONS

With an exemplar-based sunset scene classifier as the baseline system, we are able to boost its performance by using image recomposition in training, which increases the diversity of exemplars. Furthermore, using recomposition in the testing phase allows us to tailor our operating point to the need of an application. Finally, the concepts used in this study should apply to any suitable baseline sunset scene classifier and generalize to other outdoor scene classes.

**Fig. 7.** Samples of incorrectly classified images from D2.

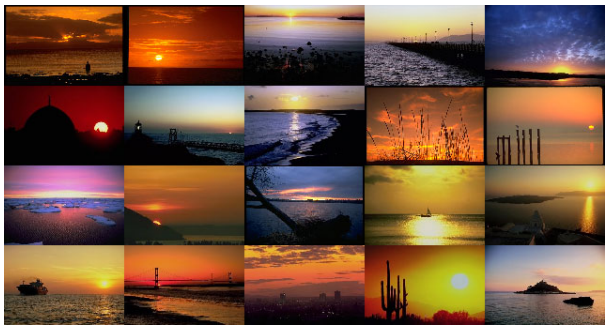

**Fig. 8.** Samples of correctly classified sunsets.

## 5. REFERENCES


[1] S. Belongie, C. Carson, H. Greenspan, and J. Malik, "Recognition of images in large databases using a learning framework," TR 97-939, U.C. Berkeley, 1997.

[2] A. Vailaya, M. Figueiredo, A. Jain, and H.J. Zhang, "Content-based hierarchical classification of vacation images," *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, 1999.

[3] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval: the end of the early years," *IEEE Trans. PAMI*, vol. 22, no. 12, pp. 1349–1380, 2000.

[4] J. R. Smith and C.-S. Li, "Image classification and querying using composite region templates," *Computer Vision Image Understanding*, vol. 75, no. 1/2, pp. 165–174, July/August 1999.

[5] Y. Wang and H. Zhang, "Content-based image orientation detection with support vector machines," *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, 2001.

[6] R.W.G. Hunt, The Reproduction of Colour, Fountain Press, England, 1995.