What do color changes reveal about an outdoor scene?

Kalyan Sunkavalli[1] Fabiano Romeiro[1] Wojciech Matusik[2] Todd Zickler[1] Hanspeter Pfister[1]

[1]Harvard University [2]Adobe Systems

Abstract

In an extended image sequence of an outdoor scene, one observes changes in color induced by variations in the spectral composition of daylight. This paper proposes a model for these temporal color changes and explores its use for the analysis of outdoor scenes from time-lapse video data. We show that the time-varying changes in direct sunlight and ambient skylight can be recovered with this model, and that an image sequence can be decomposed into two corresponding components. The decomposition provides access to both radiometric and geometric information about a scene, and we demonstrate how this can be exploited for a variety of visual tasks, including color-constancy, background subtraction, shadow detection, scene reconstruction, and camera geo-location.

1. Introduction

The importance of video-based scene analysis is growing rapidly in response to the proliferation of webcams and surveillance cameras being shared world-wide. Most of these cameras remain static with respect to the scene they observe, and when this is the case, their acquired videos contain tremendous temporal structure that can be used for many visual tasks. Compression, video summarization, background subtraction, camera geo-location, and video editing are but a few applications that have recently prospered from this type of analysis.

While temporal patterns in 'webcam data' have received significant attention, the same cannot be said of color patterns. Many webcams observe outdoor scenes, and as a result, the sequences they acquire are directly affected by changes in the spectral content of daylight. Variations in daylight induce color changes in video data, and these changes are correlated with the time of day, atmospheric conditions, weather, and camera geo-location and geo-orientation. Thus, one would expect the colorimetric patterns of outdoor webcam data to be an important source of scene information.

In this paper, we present a model for outdoor image sequences that accounts for this time-varying color information, and exploits the spectral structure of daylight. We explicitly represent the distinct time-varying colors of ambient daylight and direct sunlight, and in doing so, we show how an image sequence can be decomposed into two corresponding components. The decomposition provides access to a wealth of scene information, which can be divided into two categories:

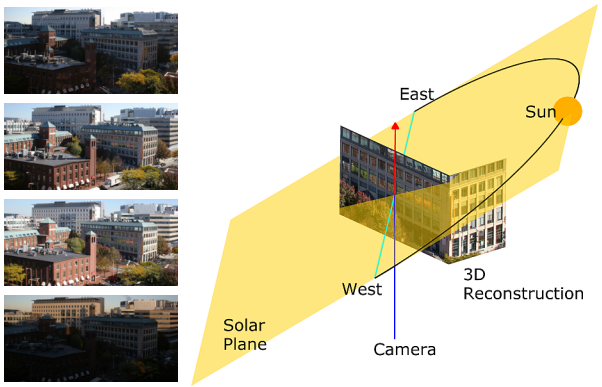

Figure 1. *Partial scene reconstruction and camera geo-location obtained with our model for time-varying color variation in an outdoor scene. By fitting our model to the input image sequence (left), we recover among other things the orientation of the solar plane relative to the local horizon (right). When combined with time stamps, this determines the latitude and longitude of the camera as well as its orientation in an astronomical coordinate system.*

1. *Per-pixel illuminant color and material properties*. Temporal variations in illuminant color are recovered separately at each scene point along with a color albedo. This provides a time-varying background model that handles cast shadows in a natural way. It also provides a trivial method for obtaining color-constant measurements of foreground objects, which is a hard problem otherwise.

2. *Scene and camera geometry*. The model provides partial information regarding the orientation of scene surfaces relative to the moving sun. By combining this information with standard geometric constraints we can predict shadow directions, recover scene geome-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

try, and locate and orient the camera in a celestial coordinate system (see Fig. 1).

2. Background and Related Work

There is a diversity of applications for our model, and in this section we discuss each separately.

Color Constancy. The goal of a computational color constancy algorithm is to extract an illuminant-invariant representation of an observed surface. Given a trichromatic (RGB) observation $\mathbf{I}^E$ acquired under unknown illuminant E, the aim is to predict the observation $\mathbf{I}^{E_o}$ that would occur under a canonical illuminant E_o. One can distinguish most color constancy algorithms along three different lines: the type of transform used for illuminant changes; the method used to estimate the transform for a given image; and whether the illuminant is homogeneous or varies throughout the scene.

Almost all existing methods model illuminant changes using 3×3 linear transforms $\mathbf{I}^{E_o} = \mathbf{M}^{E \to E_o}\mathbf{I}^E$ that are restricted to being diagonal or 'generalized diagonal' [7]. This restriction is important because it reduces the estimation problem from finding nine parameters of a general linear transform to finding only three diagonal entries. Restricting transforms to be diagonal or generalized diagonal (or even linear in the first place), implies joint restrictions on the sensors being employed, and the sets of illuminants and materials being observed [4]. General linear transforms are the least restrictive—and hence the most accurate—of the three. They are rarely used in practice, however, because robust methods for estimating nine parameters from an image do not yet exist. One of the contributions of our work is to show that by exploiting the colorimetric structure of outdoor images we can overcome this limitation and achieve reliable color constancy with general linear transforms.

Most color constancy algorithms also restrict their attention to scenes with a single illuminant (two notable exceptions are [1, 6]). In our context, however, outdoor images are captured under a mixture of two different light sources: direct sunlight and ambient skylight. Moreover, both the spectral content and the intensities of these two light sources change over the course of the day. Nonetheless, we show that we can recover the normalizing (general linear) transform parameters for any mixture of these two illuminants, and that we can do so independently for each pixel in each frame of an image sequence (see Fig. 4).

Intrinsic images. The notion of intrinsic images was introduced by Barrow and Tenenbaum [2], who studied the decomposition of an image according to the intrinsic characteristics of a scene, including illumination, reflectance, and surface geometry. Since that time, a number of related decompositions have been proposed. One such decomposition involves separating a single grayscale image into separate components for shading (relative source position) and surface reflectance (albedo) [15]. Finlayson et al. [9, 8] propose an alternative, color-based decomposition that recovers a reflectance component that is not just independent of shading but is independent of source color as well.

The problem of deriving intrinsic images can be simplified by using multiple images of a single scene under varying illumination. Weiss [18] uses a maximum-likelihood framework to estimate a single reflectance image and multiple illumination images from grayscale time-lapse video. This is further extended by Matsushita et al. [12] to derive time-varying reflectance and illumination images from similar data sets. More related to our work is that of Sunkavalli et al. [14], who propose a method for decomposing a color outdoor image sequence into components due to skylight illumination and sunlight illumination. Each of these two components is further factored into components due to reflectance and illumination that are optimized for compression and intuitive video editing. While this is related to our work, our motivation is quite different, and hence so is our model. We employ a more physically accurate model that uses general linear color transforms as opposed to diagonal transforms (which is what their model reduces to since they treat each color independently), and we make explicit assumptions about scene reflectance. This allows us to handle more general weather conditions ([14] is designed for cloudless scenes) and to recover explicit scene information such as illuminant colors, sun direction, camera position, etc.

Camera location and orientation. Estimating the geographic location and orientation of a camera from a time-stamped image sequence has rarely been considered. Cozman and Krotkov [5] extract sun altitudes from images and use them to estimate camera latitude and longitude (geo-location), and Trebi-Ollennu et al. [17] describe a system for planetary rovers that estimates camera orientation in a celestial coordinate system (geo-orientation). Both systems assume that the sun is visible in the images. Recently, Jacobs et al. [11] presented a method for geo-location based on correlating images with satellite data, but geo-orientation was not considered. In our work, we recover the position of the sun indirectly by observing its photometric effect on the scene. This provides both geo-location and geo-orientation without the need for satellite data and without requiring the sun to be in the camera's field of view (see Fig. 1).

Background subtraction/foreground detection. The core of most methods for background subtraction is the maintenance of a time-varying probability model for the intensity at each pixel. Foreground objects are then detected as low-probability observations (e.g., [16]). These methods can be difficult to apply to time-lapse data, where the time between

captured frames is on the order of minutes or more. In these cases, the 'background' can change dramatically between frames as clouds pass overhead and shadows change, and these intensity variations are difficult to distinguish from those caused by foreground objects. Our work suggests that the structure of daylight can be exploited to overcome this problem and obtain a reliable background model from time lapse data. By modeling the colors and intensities of both direct sunlight and ambient skylight over time, we can effectively predict how each scene point would appear under any mixture of these two illuminants in any given frame. Not only does this provide a means to detect foreground objects, but it also ensures that we do not return false-positive detections on the shadows that they cast (see Fig. 5).

3. A color model for outdoor image sequences

Since it is the most important natural source of radiant energy, the spectral content of daylight has received significant attention [19]. A variety of studies have shown that daylight spectra—including those of direct sunlight, ambient skylight, and combinations of the two—form a one-dimensional sub-manifold of spectral densities. When represented in chromaticity coordinates they form a 'daylight locus' that lies slightly offset from the Planckian locus of blackbody radiators. It is common to parameterize the daylight locus in terms of correlated color temperature. The correlated color temperatures of ambient skylight and direct sunlight are generally distinct, and each varies with weather, location, time of day, and time of year [19]. From a computational standpoint, it is often more convenient to represent daylight spectra in terms of a linear subspace and studies suggest that subspaces of two (or perhaps three) dimensions are sufficient.

As the spectral content of illumination changes, so does the color of an observed surface point. Restricting our attention to Lambertian surfaces and linear sensors, the trichromatic observation of any surface point under illuminant $E(\lambda)$ can be written as

$$I_k = \sigma \int C_k(\lambda)\rho(\lambda)E(\lambda)d\lambda, \tag{1}$$

where $C_k(\lambda)$ and $\rho(\lambda)$ are the sensor and spectral reflectance terms, respectively, and σ is a geometric scale factor that accounts for the angular distribution of incident radiant flux relative to the orientation of the observed surface patch.

We will use the notation $\mathbf{I}(x,t)$ for a trichromatic (RGB) image sequence parameterized by (linearized) pixel location x and time t. We choose linear transforms as our model for the effects of illuminant changes, and informed by the discussion above, we assume that the subspace containing daylight spectra is two-dimensional[1]. According to this assumption, the observation of any given material under any daylight spectral density (i.e., at any time of day and under any weather conditions) can be written as [7, 4]

$$\mathbf{I}(x,t) = \left(\sum_{i=1}^{2} c_i(t)\mathbf{M}_i\right)\boldsymbol{\rho}(x), \tag{2}$$

where $\boldsymbol{\rho}(x)$ is an illumination-independent material descriptor, $\mathbf{M}_i$ are fixed 3×3 invertible matrices that span the allowable transforms (and more), and c_i are the coordinates of a particular color transform in this basis. In the next section, we combine this color model with geometry terms to produce a complete model for outdoor image sequences.

3.1. Incorporating shading

We assume that the sequence is captured by a fixed camera in an outdoor environment. For the moment, we also assume that the scene is static, that reflectance at scene points is Lambertian, and that the irradiance incident at any scene point is entirely due to light from the sky and the sun (i.e., mutual illumination of scene points is negligible.) Under these assumptions, the sequence can be written as

$$\begin{aligned}\mathbf{I}(x,t) = \alpha(x,t)\left(\sum_{i=1}^{2} e_i^{\text{sky}}(t)\mathbf{M}_i\right)\boldsymbol{\rho}(x) \\ +\beta(x,t)\left(\sum_{i=1}^{2} e_i^{\text{sun}}(t)\mathbf{M}_i\right)\boldsymbol{\rho}(x),\end{aligned} \tag{3}$$

where $\boldsymbol{\rho}(x)$ is the material descriptor of each surface point (assumed to be of unit norm), the terms in parentheses model the effects of time-varying spectra of ambient skylight and direct sunlight, and $\alpha(x,t)$ and $\beta(x,t)$ encode the effects of albedo and scene geometry. Since the sun is a directional light source, we can write

$$\beta(x,t) = V(x,t)a(x)\cos(\omega^{sun}t + \phi(x)), \tag{4}$$

where $a(x)$ is the albedo intensity, ω^{sun} is the angular velocity of the sun, $\phi(x)$ is the projection of the surface normal at a scene point onto the plane spanned by the sun directions (the *solar plane*), and $V(x,t) \in [0,1]$ models cast shadows. This last function will be binary-valued on a cloudless day, but it will be real-valued under partly cloudy conditions.

Similarly, the α term represents the surface reflectance integrated against the ambient sky illumination. Analytical forms for this are very difficult to estimate but for our datasets we have found that a low-frequency cosine works well. Therefore, we write this term as

$$\alpha(x,t) = b(x)\cos(\omega^{sky}t), \tag{5}$$

[1] While some studies suggest that three dimensions are required, we have found that two are sufficient for our datasets.

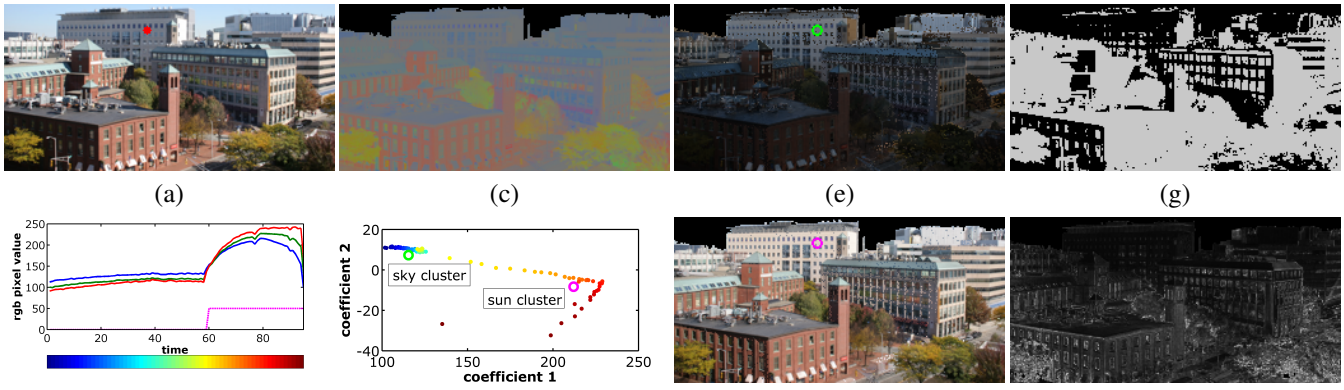

Figure 2. *Our color and shadow initialization procedure. (a) Frame 50 from the original time-lapse sequence. (b) RGB pixel values over time for the pixel indicated above. Since daylight spectra is low-dimensional, the time-varying color at each pixel lies in a plane in the color cube. A principal component analysis at each pixel allows us to recover each plane as well as a per-pixel normalized albedo (c). Projecting each pixel onto its dominant plane yields coefficients (d), shown with time coded using color from the colorbar in (b). These coefficients form two clusters that correspond to illumination by direct sunlight (f) or only ambient skylight (e), and based on these clusters we can estimate a binary shadow function (g). (Also shown for a single pixel as the magenta curve in (b)). (h) The ratio of the 3^{rd} to 2^{nd} eigenvalues at each pixel (scaled by 200). This is largest in regions of noise due to motion, foreground clutter etc., where the assumption of two-dimensional color variation for each pixels is violated.*

where $b(x)$ combines the intensity $a(x)$ and the *ambient occlusion* which represents the fraction of the hemispherical sky that is visible to each point.

3.2. Model fitting

While the model in Eq. 3 is non-linear and has a large number of parameters, these parameters are overconstrained by the input data. For a time-lapse image sequence with P pixels and F frames, we have $3PF$ observations but only $PF+5P+4F$ degrees of freedom. In order to fit the model to an input sequence, we begin by recovering the color parameters (M_1, M_2, and $\boldsymbol{\rho}(x)$) independent of intensity. This enables an initial decomposition into sun and sky components, which is then refined through a global optimization over the remaining parameters.

Material colors and a transform basis. From Eq. 2 it follows that the trichromatic observations $\mathbf{I}(x,\cdot)$ of a single pixel over the course of time will lie in a plane spanned by $\mathbf{M}_1\boldsymbol{\rho}(x)$ and $\mathbf{M}_2\boldsymbol{\rho}(x)$. A good estimate of this plane is found through a principal component analysis (PCA) of $\mathbf{I}(x,\cdot)$. The PCA yields color basis vectors $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$ corresponding to the three eigenvalues $\sigma_1 \geq \sigma_2 \geq \sigma_3$. The plane we seek has $\mathbf{u}_3$ as its normal vector. Doing this separately at each pixel yields a set of F planes, which induce constraints on the materials and transform basis matrices

$$\mathbf{u}_3(x)^\top(\mathbf{M}_1\rho(x)) = 0, \quad \mathbf{u}_3(x)^\top(\mathbf{M}_2\rho(x)) = 0. \qquad (6)$$

These constraints do not uniquely determine the unknown parameters. Arbitrary invertible linear transformations can be inserted between $\mathbf{M}_i$ and $\boldsymbol{\rho}(x)$, for example, and these correspond to changes of bases for the illuminant spectra and material spectral reflectance functions. These changes of bases are of no theoretical importance, but they do have practical implications. In particular, parameter choices for which the angle between $\mathbf{M}_1\boldsymbol{\rho}(x)$ and $\mathbf{M}_2\boldsymbol{\rho}(x)$ is small (for any scene point x) are poor because they will lead to numerical instabilities. A convenient method for choosing 'good' parameters is to find those that minimize the objective function

$$\mathcal{O}(\mathbf{M}_i,\boldsymbol{\rho}(x)) = \sum_{i=1}^{2}\sum_{x} ||\mathbf{M}_i\boldsymbol{\rho}(x) - \mathbf{u}_i(x)||^2 \qquad (7)$$

subject to the constraints in Eq. 6. Since $\mathbf{u}_1(x)$ and $\mathbf{u}_2(x)$ are orthonormal for all x, this ensures numerical stability in the subsequent analysis, and since $\mathbf{u}_1(x)$ is the dominant color direction at each scene point, it effectively chooses bases for the space of illuminants and spectral reflectances such that $M_1\boldsymbol{\rho}(x)$ is close to the mean color of the sequence.

When the scene contains foreground objects, interreflections, and non-Lambertian surfaces, estimates of the color plane for each pixel (i.e., the normals $\mathbf{u}_3(x)$) can be corrupted by outliers. In these cases, we have found that enforcing Eq. 6 as hard constraints yields poor results. A better approach is to perform an unconstrained minimization of the objective function in Eq. 7, which already has a soft version of the constraints 'built in'.

The shadow function. Central to the decomposition into sun and sky components is the estimation of the shadow function $V(x,t)$, which indicates whether the sun is visible to a scene point in a given frame. This function can be recovered by simultaneously exploiting the differences be-

Figure 3. *This figure shows (top row) frames 1, 30, 60 and 95 from the original video, (middle row) the reconstruction from our model, and (bottom row) the absolute error multiplied by 3.*

tween the color and intensity of sunlight and ambient daylight. For the moment, we assume that $V(x,t)$ is a binary function.

The material vectors $\boldsymbol{\rho}(x)$ and the transform basis $\{\mathbf{M}_1, \mathbf{M}_2\}$ define a color plane for each pixel, and by projecting the observations $\mathbf{I}(x,t)$ onto these planes we obtain the coefficients $\mathbf{c}(x,t) = (c_1(x,t), c_2(x,t))$ of Eq. 2. For a given pixel, the coefficients $\mathbf{c}(x,\cdot)$ provide a description of that pixel's color and intensity over time. Due to the differences between sunlight and skylight, the coefficients $\mathbf{c}(x,\cdot)$ will generally form two separate clusters corresponding to the times that the scene point is lit by the sun and those when it is not (Fig. 2(d)). We observe that the clusters differ in both intensity (distance from the origin) and color (polar angle). Using the cluster centers $\mathbf{c}^{sky}(x)$ and $\mathbf{c}^{sun}(x)$, we label a pixel as 'in shadow' or 'lit by the sun' on the basis of the distances $d^{sky}_{x,t} = ||\mathbf{c}(x,t) - \mathbf{c}^{sky}(x)||$ and $d^{sun}_{x,t} = ||\mathbf{c}(x,t) - \mathbf{c}^{sun}(x)||$. By applying a two-cluster k-means algorithm we can define a decision boundary $\mathcal{B}(x)$ for whether the sun is visible to a scene point or not.

While we could use these per-pixel decision boundaries to recover the binary shadow function $V(x,t)$, the results can be significantly improved by exploiting temporal and spatial coherence. To do this, we construct a graph in which each space-time point (x,t) is a node and each node is connected to its six nearest space-time neighbors. Using a standard graph cuts algorithm [3], we determine $V(x,t)$ as the binary labeling that minimizes (globally) an energy function. The unary (data) terms in the energy function measure the position of the coefficients $\mathbf{c}(x,t)$ relative to the decision boundaries $\mathcal{B}(x)$, and for the labels of sky and sun, are given by

$$D^{sun}_{x,t} = \frac{1}{1 + e^{-(3d^{sky}_{x,t} - d^{sun}_{x,t})}}, D^{sky}_{x,t} = 1 - D^{sun}_{x,t}. \quad (8)$$

The pairwise (smoothness) terms are based on Pott's model. The recovered binary shadow function $V(x,t)$ can be refined, for example, by updating the per-pixel cluster centers according to this labeling and repeating the graph-cuts procedure. In practice we have found this not to be necessary.

Remaining parameters. Points that are known to be in shadow determine the angular sky parameter ω^{sky} in Eq. 5. It can be estimated robustly using a non-linear least-squares optimization. By subtracting the ambient component from the input sequence, we obtain an approximate 'direct sun' sequence that can be used to estimate the angular sun velocity ω^{sun} in a similar fashion. Note that we need to consider only a small number of spatial points to recover these parameters.

Referring to Eq. 3, the remaining parameters to be estimated are the transform coefficients $e^{\text{sky}}_i(t)$ and $e^{\text{sun}}_i(t)$ and the surface albedos $\rho(x)$ and normal angles $\phi(x)$. Similar to [14] we randomly initialize these parameters and then iteratively update each in order to minimize the RMS difference between the model and the input sequence. The coefficients $e^{\text{sky}}_i(t)$ and $e^{\text{sun}}_i(t)$ are updated using linear least squares, and the normal angles $\phi(x)$ are updated using a one-dimensional exhaustive search for each pixel.

As a final step, the binary shadow function is relaxed by finding the real-valued function $V(x,t) \in [0,1]$ that minimizes the RMS reconstruction error. This is an important step for any scene captured under partly cloudy conditions.

Experimental Results. Table 1 and Fig. 3 show results for two sequences obtained from Sunkavalli et al. [14]. These sequences consist of roughly the same scene in two different weather conditions (sunny and partly cloudy), and each sequence was captured over the course of one day with approximately 250 seconds between frames. The accompa-

Figure 4. *Color Constancy. The top row shows the original frames 1, 35, 94 and 120 and the bottom row shows the corresponding images reconstructed with the sun and sky illuminant colors fixed to those of frame 35.*

Table 1. *RMS reconstruction errors.*

Data	# Imgs	Resolution	RMS Error
Sunny square	95	130×260	6.42%
Cloudy square	120	240×360	7.36%

nying video shows these sequences in their entirety. It is important to note that the visible portions of the sky in our sequences were not considered in the decomposition; for all the results shown in the paper and the video, they have been copied from the original data to avoid distracting the reader.

In our results, errors are caused by foreground objects, smoke, interreflections from windows, and saturation of the camera. Another significant source of error is deviation from the assumed Lambertian reflectance model. From examining our data, it seems as though a rough-diffuse model [13] would be more appropriate.

4. Implications for machine vision

The appearance of a scene depends on shape and reflectance, the scene illumination (both color and angular distribution), as well as the observer's viewpoint. Any visual task that requires some of this information seeks to recover it in a manner that is insensitive to changes in the others. By explicitly isolating many of these scene factors, our model enables novel approaches to some visual tasks and improves the performance of a number of others. Here we provide examples that relate to both color and geometry.

Color Constancy. As mentioned in Sect. 2, most (single image) color constancy algorithms restrict their attention to diagonal or generalized diagonal transforms when representing changes in illumination. Even with this restricted model, estimating the transform parameters in uncontrolled environments is hard to do reliably. In contrast, once our model is fit to an image sequence, the task of color constancy becomes trivial. Since we obtain illuminant transform parameters separately for each frame and sun/sky mixing coefficients independently for each pixel, we can obtain illuminant-invariant descriptions everywhere simply by manipulating these parameters. Fig. 4 shows an example in which the color in each frame of the sequence is corrected so that the effective sky and sunlight colors are constant over the course of the day. (They are held fixed to the colors observed in frame 35 of the sequence). Clear differences are visible between this and the original sequence, especially near dawn and dusk.

We emphasize that the color corrections are applied to the entire sequence, including the foreground objects. As a result, if one applies a color-based recognition algorithm to the color-corrected sequence instead of the original sequence, one can effectively obtain color-constant recognition with very little computational overhead. In addition, our use of general linear transforms can be expected to provide increased accuracy over what could be obtained using common diagonal or generalized diagonal transforms [4].

Background subtraction. Most background subtraction methods perform poorly when the illumination changes rapidly, for example, on a partly cloudy day. This problem is exacerbated in time-lapse data, where the time between frames is on the order of minutes, and the temporal coherence of foreground objects cannot be exploited. By modeling the entire scene over time, our model provides the means to handle these effects quite naturally. In particular, it immediately suggests two strategies for foreground detection. As noted earlier, the trichromatic observations $\mathbf{I}(x, \cdot)$ lie in the plane spanned by vectors $\mathbf{M}_1\boldsymbol{\rho}(x)$ and $\mathbf{M}_2\boldsymbol{\rho}(x)$. Thus, one approach to foreground detection is simply to measure the distance between an observation $\mathbf{I}(x, t)$ and its corresponding spanning plane. This approach has the advantage of ignoring shadows that are cast by foreground objects, since cast shadow induce variations *within* the spanning planes. A second approach is to use the complete time-varying reconstruction as a background model

Figure 5. *Simple foreground detection using per-pixel thresholds in color space. Frames 3, 41, 72, 88 from the original sequence with detected foreground pixels marked in red. Shadows cast by foreground objects are correctly ignored. Violations of our model (interreflections, saturated regions, etc.) trigger false positive responses.*

and to use simple background subtraction for each frame. Fig. 5 shows the result of a combination of these two approaches, and shows how one can identify cars on the street without false positive responses to the shadows they cast or to the shadows cast by moving clouds. We do see detection errors in some areas, however, and these correspond to saturated image points, dark foreground objects with low SNR, and inter-reflections from other buildings. Nonetheless, the detection results presented here suggest that our model will provide a useful input to a more sophisticated detection algorithm.

Scene geometry and camera geo-location. Our model provides direct access to the angular velocity of the sun ω^{sun} as well as the angles $\phi(x)$ in Eq. 4, which are one component of the surface normal at each scene point that corresponds to its projection onto the solar plane. This partial scene information can be combined with time-stamp information and common geometric constraints to recover scene geometry as well as the geo-location and geo-orientation of the camera.

Given three scene points x_i that are known to lie on three mutually orthogonal planes (two sides of a building and the ground plane for example), we can represent the normals $n_i = (\theta_i, \phi_i)$ in terms of spherical coordinates in a solar coordinate system (Z-axis is the normal to the solar plane and East is the X-axis). The azimuthal angles ϕ_i are equal to the corresponding $\phi(x_i)$ from our model up to a unknown, global additive constant. If each normal has a unique azimuthal component, our model gives two constraints on n_i in the form of the azimuthal differences $(\phi_{x_1} - \phi_{x_2})$ and $(\phi_{x_2} - \phi_{x_3})$. Combining these with mutual orthogonality constraints, the three normals are determined relative to the solar plane. (The same can be achieved from two orthogonal planes with the same albedo.)

If one of the recovered normals is the ground plane, the angle of the solar plane, and therefore the peak (or meridian) altitude of the sun is uniquely determined. In addition, the projection of the ground plane normal onto the solar plane provides the azimuthal angle ϕ_{peak} of the sun's peak passage and East corresponds to the direction in the solar plane with azimuthal angle $\phi_{peak} - \pi/2$. Thus, by observing orthogonal planes over the course of a day, we can achieve the functionality of a combined compass and sextant.

Given the date and UTC time-stamps for each frame, we know the UTC time of the sun's peak passage (i.e. its meridian altitude). This time uniquely determines both the latitude and longitude of the observed scene (e.g., using a nautical almanac such as http://aa.usno.navy.mil/data/docs/AltAz.php). Likewise, if we know the latitude and longitude of the camera (and the season and year) we can reverse this process and compute the date and a UTC time stamp for the peak frame and propagate time stamps to all frames in the sequence using the time interval. Results of these analyses for one of our sequences is shown in Table 2.

Table 2. Camera geo-location results.

True value	**Estimate**
$42^o21'57''$N $71^o05'33''$W	$42^o12'58''$N $70^o05'47''$W
t_{peak} = 16:25 UTC date = 10/27/06	t_{peak} = 16:30 UTC date = 10/28/06

The meridian altitude of the sun was found to be 34.3 degrees. Using the UTC time-stamps from the image sequence, this predicts a latitude and longitude that is only 83.7km from the ground truth position. Alternatively, had we known the true geo-location of the camera, as well as the year and season of acquisition, we would have estimated a UTC time that differs from the true value by only 5 minutes and a date that deviates from the actual one by a day.

Finally, if we know the vanishing lines corresponding to the three scene planes, the camera can be calibrated [10]. This yields the orientation of its optical axis relative to the solar plane, and in a celestial coordinate system. This achieves the functionality of a combined compass and inclinometer. A reconstruction of the scene is shown in Fig. 1. This includes the recovered solar plane, the orientation of the camera, and two reconstructed planes that are texture-mapped with the input frame that corresponds to the indicated sun direction.

Shadow Prediction. Once the solar plane is known, we can determine the sun direction within that plane for each frame of a sequence. This can be used, for example, to predict a time varying vanishing point on the image plane

Figure 6. *Estimated shadow direction in two different frames.*

that corresponds to the direction in which vertical objects will cast shadows onto the ground plane. If a vertical object (e.g., a person) is known to touch the ground plane at pixel x in a given frame, its shadow will lie on the line segment connecting x to the vanishing point of the shadow direction for that frame. This is demonstrated in Fig. 6, which shows predicted shadows vectors for some vertical objects that can be used for improved background subtraction.

5. Discussion

This paper presents a model for exploiting the colorimetric structure of extended outdoor image sequences. The model explicitly represents the time-varying spectral characteristics of direct sunlight and ambient skylight, and it allows an image sequence to be decomposed into distinct components that can be interpreted physically. The model can be reliably fit to time-lapse sequences in which the sun is visible for at least a fraction of a day; and once it is fit, it can be used for a variety of visual tasks. Examples in this paper include color constancy, background subtraction, scene reconstruction and camera geo-location.

The model could be improved by incorporating robust estimators into the fitting process, by using a more flexible reflectance model, and by making use of temporal patterns to appropriately handle 'time-varying textures' such as moving water and swaying trees.

There are a number of additional applications to be explored. By segmenting the scene according to albedo $\boldsymbol{\rho}(x)$ and surface normal angle $\phi(x)$, one may be able to use orthogonality constraints to produce a coarse reconstruction of the scene. This type of scene geometry has proven to be a useful source of context information for object recognition. Also, since there is a one-to-one correspondence between coordinates in our illuminant transform space (e_i^{sky}, e_i^{sun}) and complete spectral densities in the daylight locus, it may be possible to use our model to infer information about air quality and other atmospheric conditions.

Acknowledgements

Fabiano Romeiro and Todd Zickler were supported by NSF CAREER Award IIS-0546408.

References

[1] K. Barnard, G. D. Finlayson, and B. V. Funt. Color constancy for scenes with varying illumination. *Computer Vision and Image Understanding*, 65(2):311–321, March 1997.

[2] H. Barrow and J. Tenenbaum. *Recovering intrinsic scene characteristics from images*. Academic Press, 1978.

[3] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision. 26(9):1124–1137, 2004.

[4] H. Y. Chong, S. J. Gortler, and T. Zickler. The von Kries hypothesis and a basis for color constancy. In *Proc. IEEE Int. Conf. Computer Vision*, 2007.

[5] F. Cozman and E. Krotkov. Robot localization using a computer vision sextant. In *Proc. IEEE Conf. Robotics and Automation*, pages 106–111, 1995.

[6] M. Ebner. Color constancy using local color shifts. In *Proc. European Conf. Computer Vision*, pages 276–287, May 2004.

[7] G. Finlayson, M. Drew, and B. Funt. Color constancy: generalized diagonal transforms suffice. *J. Optical Society of America A*, 11(11):3011–3019, 1994.

[8] G. D. Finlayson, M. S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *8th European Conference on Computer Vision (ECCV 2004)*, pages 582–595, May 2004.

[9] G. D. Finlayson, S. D. Hordley, and M. S. Drew. Removing shadows from images. In *Proc. European Conf. Computer Vision*, pages 823–836, May 2002.

[10] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[11] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *Proc. IEEE Int. Conf. Computer Vision*, 2007.

[12] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3–10, 2003.

[13] M. Oren and S. Nayar. Generalization of the Lambertian model and implications for machine vision. *Int. Journal of Computer Vision*, 14:227–251, 1996.

[14] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz. Factored time-lapse video. *ACM Transactions on Graphics*, 26(3):101:1–101:10, July 2007.

[15] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. In *Neural Information Processing Systems*, pages 1343–1350, December 2002.

[16] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proc. IEEE Int. Conf. Computer Vision*, pages 255–261, 1999.

[17] A. Trebi-Ollennu, T. Huntsberger, Y. Cheng, E. T. Baumgartner, B. Kennedy, and P. Schenker. Design and analysis of a sun sensor for planetary rover absolute heading detection. *IEEE Trans. on Robotics and Automation*, 17(6), 2001.

[18] Y. Weiss. Deriving Intrinsic Images from Image Sequences. In *Proc. IEEE Int. Conf. Computer Vision*, pages II: 68–75, 2001.

[19] G. Wyszecki and W. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley New York, second edition, 2000.