SIViP
DOI 10.1007/s11760-007-0013-8

ORIGINAL PAPER

Satellite imagery based adaptive background models and shadow suppression

Anup Doshi · Mohan Manubhai Trivedi

Received: 26 October 2006 / Revised: 26 March 2007 / Accepted: 30 March 2007

Abstract Accurate segmentation of foreground objects in video scenes is critical for assuring reliable performance of vision systems for object tracking and situational awareness in outdoor scenes. Most existing techniques for background modeling and shadow suppression require that a number of parameters be "hand-tuned" based on environmental conditions. This paper presents two contributions to overcome such limitations. First, we develop and demonstrate a satellite imagery based approach for selecting appropriate background and shadow models. It is shown that the illumination conditions (i.e. cloud cover) of a scene can be reliably inferred from visible satellite images in the local region of the camera. The second contribution presented in the paper is introduction and evaluation of a Hybrid Cone-Cylinder Codebook (HC3) model which combines an adaptive efficient background model with HSV-color space shadow suppression into a single coherent framework. The structure of the HC3 model allows for seamless fusion of the satellite data. We are thereby able to exploit the fact that, for example, shadows are more pronounced on sunny days than cloudy days, allowing for more sensitive detection. The paper presents a set of experiments using day long sequences of videos from an operational surveillance system testbed. Results of these experimental analyses quantitatively illustrate the benefits of using satellite imagery to inform and adaptively adjust background and shadow modeling.

A. Doshi (✉) · M. M. Trivedi
Computer Vision and Robotics Research Laboratory,
University of California at San Diego,
9500 Gilman Drive, La Jolla, CA 92093-0434, USA
e-mail: andoshi@ucsd.edu

M. M. Trivedi
e-mail: mtrivedi@ucsd.edu

Keywords Foreground–background segmentation · Shadow suppression Satellite imagery · Cloud radiance · Hybrid Cone-Cylinder Codebook (HC3)

1 Introduction and motivation

Most applications in the realm of the outdoor visual surveillance all begin at a common root. The primary step of these vision algorithms is a segmentation of the foreground from the background. The end applications may be vastly different, from event capture and situational awareness [34] to vehicle counting to face identification, however an unreliable foreground estimate can wreak havoc on the results.

One potentially significant difficulty involved in this step is the presence of shadows cast by foreground objects. While shadows are usually not part of the foreground of interest, they inevitably show up in the foreground anyway. Algorithms that do not deal with shadows may have trouble segmenting two distinct objects or accurately determining the shape of an object. Thus much work has been done in the automatic detection and removal of shadows. Fortunately shadows are somewhat structured in their relationship to the background, and with that knowledge it is possible to remove the shadows.

Both the background removal and shadow suppression steps can be difficult to work with over a longer time period however. This is simply because for each scene and each lighting condition, usually several parameters must be manually set to achieve desirable results. Non-parametric algorithms exist but may suffer from lack of sensitivity as compared to the parametric ones [27]. Thus having an estimate of the environmental conditions would help greatly,

since the correct parameter set could then be automatically chosen.

With the advent of widely available weather data via the Internet, useful information could conveniently be gleamed from most any location (Fig. 1). Specifically, the most prevalent data comes in the form of satellite imagery. Geostationary satellites which gather reflectance data cover most of the globe, and their data is publicly available in near real-time, updated every 15 min. Given this data, which is available at resolutions of 1 km, the cloudiness and sunlight intensity can be inferred for any specific location.

In order to use the satellite data to its fullest capabilities, in terms of adjusting both the background and shadow parameters, a single coherent foreground extraction model is proposed. The Hybrid Cone-Cylinder Codebook (HC3) model incorporates the highly efficient Codebook background model with a shadow suppression algorithm in HSV color space [7]. The proposed model is designed on a fundamental level to attempt to provide coherence, faster performance, and more accurate pixel clustering reflecting basic properties of illumination changes in HSV space. Additionally, the HC3 model has several key parameters directly related to sunlight intensity, describing the variability of the background and the strength of the shadows. By automatically adjusting these parameters according to the satellite information, foreground segmentation becomes more sensitive and accurate.

The paper is organized as follows. Section 2 investigates previous work in the field, followed in Sect. 3 by a discussion of the acquisition and processing of satellite imagery. Section 4 then introduces the HC3 model and incorporates the satellite data. Evaluations are explained in Sect. 5, followed by Sect. 6.

2 Related research studies

Foreground–background segmentation is a relatively extensive field. In search of algorithms that would perform well in light of all the problems posed above, several main contenders appeared. Typically models for backgrounds are based on single Gaussians [9,10,16] or a Mixture of Gaussians (MOG) [1,13,30]. While single Gaussians fail to model complex backgrounds, MOGs have trouble with sensitive detections and fast variations [12,16]. Also, while MOGs may be able to converge to a complex background with enough components, the computation required may be quite complex [25].

Another method for modeling backgrounds involves nonparametric kernel density estimation [8]. This method can in many cases be prohibitively memory intensive [12,25]. The codebook method [12], described in detail in Sect. 4.2.1, uses a non-statistical clustering technique to achieve a fast and efficient model for the background while allowing for a moving background. There are many other techniques (Wallflower [32], Pfinder [35], and more [23]) based on texture, color spaces, and pixel-, region-, or frame-based approaches.

Shadow suppression has also been quite extensively studied. A common assumption is that shadows decrease the luminance of an image, while the chrominance stays relatively unchanged [2,6,9,10,18,25,29]. However there are many other methods for removing shadows, many of which are reviewed by Prati et al. [26,27]. Accordingly, the algorithm based on the HSV space [6] was cited as the most generally capable and robust shadow detection and suppression algorithm [27]. This algorithm is also described in further detail in Sect. 4.2.2. It is worth noting that not many shadow detection schemes treat "highlights," or lighting increases/specularities, as well. These can be interpreted as the inverse of shadows, and much of the same reasoning can be used for both. Therefore the work in this paper focuses on shadows and highlights both.

It is also possible to attempt to detect environmental changes in the scene based solely on the data from the camera itself [31]. Such a method requires considerable prior knowledge of the scene itself, whereas for the proposed method it is sufficient to know the scene is outdoors and during the daytime. Cavallaro et al. also detects changes in the background noise model and additionally considers various color spaces in order to remove shadows and segment the foreground [3]. Yet with the simple addition of the satellite sensor to infer weather, detection of scene changes becomes much more straightforward.

However we found a lack of work on using measured cloud reflectance from satellites for the purpose of aiding ground-based visual object detection. In the field of remote sensing and climatology, much has been done to characterize the properties of clouds and what can be seen from satellites [17,15,14,11,28,36], with fundamental work in radiative transfer going back to Chandrasekhar [4]. This work will be applied below to show the usefulness of the satellite imagery.

3 Principles and processing of satellite imagery

Satellite imagery is virtually an unknown player in the study of object detection and tracking. In the following sections an explanation is put forth as to why and how it is useful (Sects. 3.1.1, 3.1.2), along with how the data is acquired (Sect. 3.1.3) and how it is processed (Sect. 3.2).

3.1 Motivation

3.1.1 Ground illumination

In the interest of detecting objects using the visual spectrum, it is useful to study the luminance properties of a given scene. Generally the daytime irradiance on a ground object

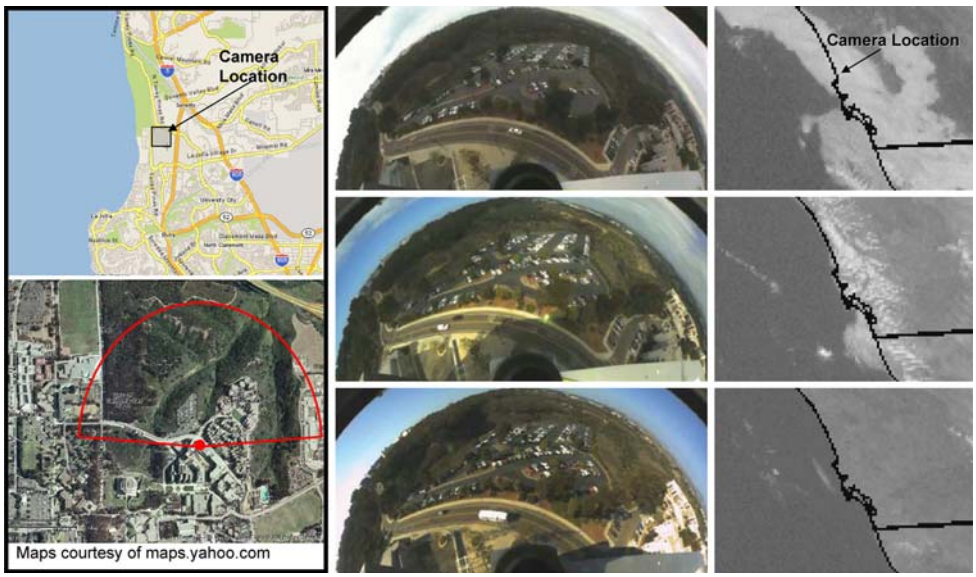

Fig. 1 Three examples of satellite images and corresponding scenes. *Top* Overcast day, *Middle* partly cloudy, *Bottom* sunny. Note the lack of shadows in the overcast scene, as compared with the dark shadows in the sunny scene. The location and field of view of the camera can be seen on the left

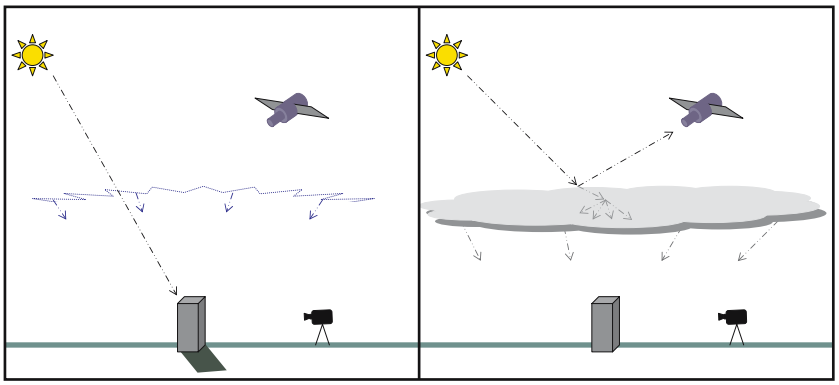

Fig. 2 An scene under direct sunlight and under cloudy conditions will be viewed differently. The cloudiness of the scene can be inferred from the reflected light by the satellite

(see Fig. 3) can be written as $I = I_{\text{sun}} + I_{\text{sky}}$, in the absence of artificial light [33]. When there are no clouds, I_{sun} dominates I_{sky} as there will be direct sunlight on the object. This will result in the appearance of shadows; further, the greater the intensity of I_{sun}, the more pronounced the shadow.

When the scene is overcast, the case is just the opposite. Direct sunlight no longer reaches the object, but instead diffuses through the clouds. I_{sky} then becomes diffused light of approximately equal strength from all directions, and the shadows disappear. As depicted in Fig. 2, a viewer of the same scene under such different conditions would have to adjust to each environment.

3.1.2 Using satellites to determine ground illumination

In order to characterize the I_{sky} parameter on cloudy days, it is necessary to measure the amount of light traveling through the cloud. There are potentially many alternatives to calculate this transmitted solar flux, including sensors on the ground measuring intensity at the visible wavelengths. However for every given scene, this would require additional equipment and labor costs. On the other hand, there are already sensors on satellites that cover most of the world. Since their data is publicly available, the usage becomes extremely scalable and quite cheap.

It is not immediately obvious why satellite sensors will help though. Here we try to show why the ground illumination properties can be deduced from measurements taken on a satellite.

Clouds tend to obscure any electromagnetic signal between wavelengths of 300 and 3,000 nm, including the $400 - 750$ nm visible spectrum [5]. The extent to which they do so can be analyzed by studying the solar radiative flux through the clouds. Specifically, any incident solar flux on the top of a cloud either reflects, gets absorbed, or transmits through the cloud. Thus we can write:

$$R + T + A = 1$$

where R, T, and A are the reflectance, transmissivity, and absorptivity of a given cloud [11]. This is assuming the Earth's surface albedo is very small, which is a common modeling technique. Typically clouds reflect much more radiation

back to space as compared to the surface without clouds [19]. In fact, according to [15] the effects of surface reflectance amount to less than 2% difference in the upward flux of the clouds. Thus for our purposes the surface albedo can be ignored in examining the interaction of solar radiation and clouds.

The energy transmitted through the cloud is then clearly directly proportional to I_{sky}, which is what we need to estimate (i.e., $I_{\text{sky}} \propto T$). The satellite, on the other hand, will measure energy that is reflected off the cloud. If there were a relationship between R and T, we would be able to exploit that to find I_{sky} given satellite measurements. Because of the absorption A, however, the relationship between reflected and transmitted energy is not immediately clear.

Radiative properties of clouds have been studied extensively [17,15,14,11,28,36], as they play an important role in climate predictions. Many models have been developed for the effect of clouds; among the most common is a class of models known as two-stream approximations to radiative transfer. These closed-form results are useful for being easy to interpret and accurately representing important parameters in the radiative transfer process (a general and thorough treatment of the analysis can be found in [17] and [14]). The equations of interest follow as

$$R = \frac{\omega_0 \tau'}{\mu_0} \beta_0$$

$$T = 1 - \frac{\tau'}{\mu_0}(1 - \omega_0(1 + \beta_0))$$

where ω_0 is the scattering coefficient divided by the sum of the scattering and absorption coefficients of the material, μ_0 is a function of the incident ray angle, β_0 is a quantity linked to the scattering law of single particles, and τ' is a measure of the optical depth, which is related to the amount of water in a vertical column of the cloud, or the "thickness" of the cloud [17]. It is rather straightforward then to see how thickness, absorption, and solar angle affect R and T. In essence as a function of these parameters, whenever reflectance increases, transmittance decreases. Thus we can conclude that the behavior of transmittance T can be inferred given reflectance data R.

3.1.3 Data acquisition

The National Weather Service (NWS) publicly distributes visible satellite imagery covering the entire United States [22], specifically in the area of La Jolla, CA where our cameras are situated. The data of interest is measured by the GOES-10 satellite, one of a series of Geostationary Operational Environmental Satellites, operated by the National Oceanic and Atmospheric Administration (NOAA). The satellite contains a myriad of sensors including an Imager that captures energy from two sources: (1) radiant energy produced by the Earth, and (2) reflected solar energy. The Imager itself incorporates five sensors; one in the visible wavelengths and the rest in infrared wavelengths [20]. The visible wavelength sensor detects energy solely from the reflection of solar energy, as the radiated energy from Earth occurs only in the infrared wavelengths [5].

The resolution of the visible-spectrum sensor is listed as approximately 1 km^2 [20]. This kind of resolution provides reflectance data on a given location with relatively impressive precision. Whereas a weather website or forecaster may list the conditions of a general area around a city, for example, the satellite image would pinpoint the cloudiness within a few city blocks. Moreover, the image is updated every fifteen minutes, providing near real-time reports of conditions.

3.2 Processing

For the purposes of this paper the satellite data was collected from [21] after every update. The location of the ground scene of interest was found on the satellite image, and 5-pixel-width box around that point was averaged to measure the reflectance of the clouds. This averaging was done to account for any noise in the raw data.

The raw data is quite affected by the angle of the sun, accounting for flux changes due to the time of day. These effects can be seen in Fig. 3a. In order to account for this factor, the measured reflectance was first subtracted from some reference curve, something akin to a "normal day". The "normal" curve was extracted as a smoothed average over several days of collected reflectance data. The relative sunlight intensity (RSI) is then defined as *RSI = normal reflectance – measured reflectance*, and makes it easy to analyze the data and predict the environmental conditions of the scene (Fig. 3b).

The cloudiness state parameter is then predicted from the RSI using MAP estimation, as follows:

$$c = \text{argmax}_c \{P(R|c)P(c)\}, \quad c \in \psi$$

where $\psi = \{\text{overcast, partly cloudy, clear}\}$ is the set of cloudiness classifications and R is the relative sunlight intensity. The prior $P(c)$ is assumed to be uniform, implying Maximum Likelihood estimation. For a specific geographical location, however, the prior could be adjusted based on weather history: for example, in San Diego it may always be sunny, and in Pittsburgh it may often be cloudy. The likelihoods $P(R|c)$ are modeled by Gaussian distributions, whose means and variances are learned from manually segmented sample data as seen in Fig. 3c. In this paper three classifications are used for ψ, representing the major cloudiness conditions. To include more cloudiness states would amount to collecting more training data to learn their likelihoods.

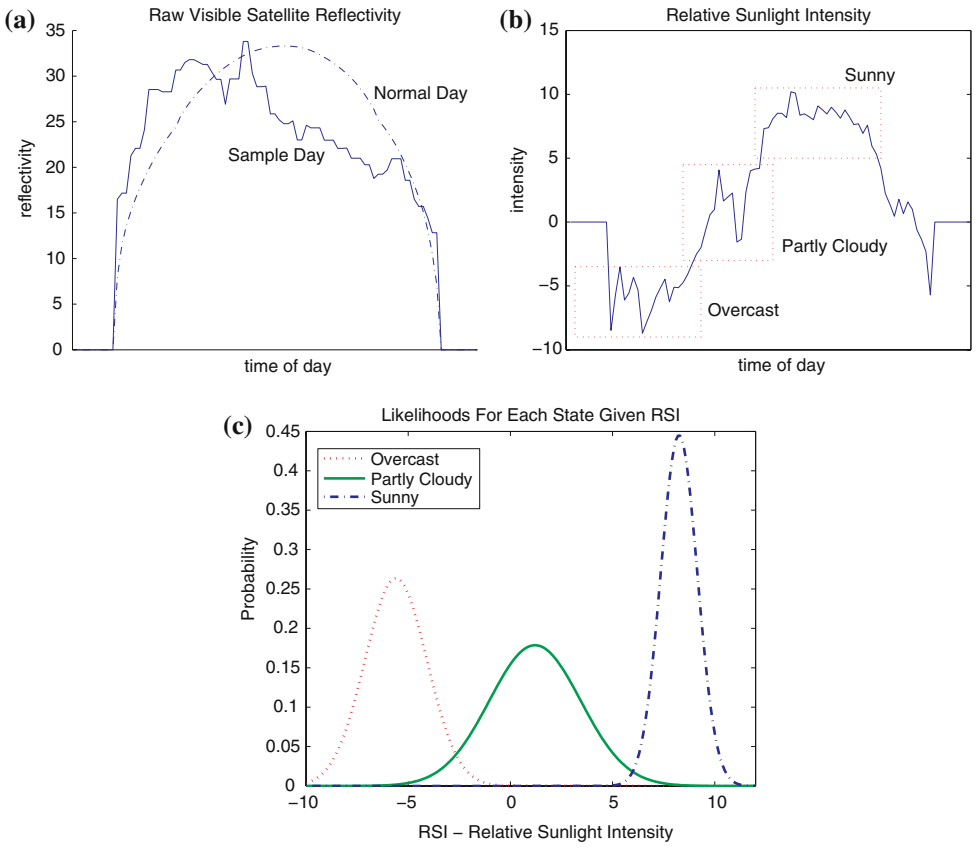

Fig. 3 Example day-long satellite data analysis. Samples of these three classifications can be seen in Fig. 1, respectively. **a** Reflectance is measured and compared against a "normal" day's curve. **b** The difference is plotted and segmented as overcast, partly cloudy, and sunny. **c** The data is used to train Gaussian Likelihood functions for each state, given the relative sunlight intensity

4 Hybrid Cone-Cylinder (HC3) model

A robust background model is now proposed which has been developed independently of the multi-sensory satellite data. The model is formulated, however, with the goal in mind of incorporating a daytime cloudiness parameter which was estimated above.

4.1 Motivation: background subtraction and shadow suppression

Outdoor surveillance and security settings are likely one of the most common uses of video cameras. Yet there are still many problems to be solved to make a stable and reliable computer vision system useful for those applications. The primary problem for any system is segmenting foreground objects from the background, in the face of various environmental challenges.

These challenges include global illumination changes; moving cast shadows during the daytime; specularities, highlights, and shadows due to artificial light sources at night; and moving backgrounds such as flags or trees. Global illumination changes can occur quickly, as in a cloud suddenly blocking out the sunlight, or relatively slowly, as the sun takes a while to set. Either way, the perceived background will change, so a model which removes that should adaptively update. Moving cast shadows are those shadows which move along with a foreground object, but are not a part of the object itself [6]. These shadows can cause objects to merge, distort their shape, or occlude other objects, so they can be very problematic.

An environmental challenge which is quite common to many outdoor scenes is of moving backgrounds. A common example is of trees or grass swaying in the wind. These movements may occur frequently and with an unknown periodicity (depending on the wind), so an algorithm to adaptively learn such motions would be useful.

Section 4.2 describes the original models chosen here for background subtraction and shadow segmentation in order to tackle the above problems. Following that Sect. 4.3 proposes a new model which fuses the original two, with several key modifications designed to make the model fundamentally coherent. Satellite data is incorporated into the model in Sect. 4.4.

4.2 Description of original models

4.2.1 Codebook model for background–foreground segmentation [12]

The codebook model for foreground extraction is quite simply a non-statistical clustering scheme with several important

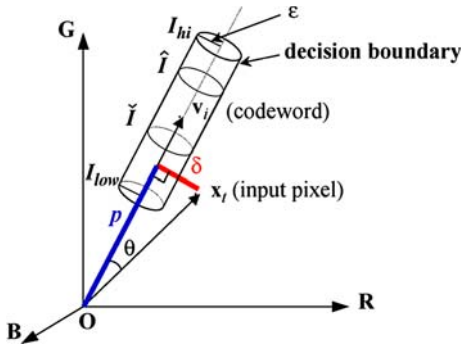

Fig. 4 Codebook 'codeword' model [12]

additional elements to make it robust against moving background. The motivation for having such a model is that it will be fast, because it is deterministic; efficient (requires little memory); adaptive; and able to handle complex backgrounds with sensitivity. A detailed description of the model can be found in [12].

In the codebook model, each pixel can have a variable number of codewords representing the background. The codeword is comprised of an RGB vector, and several auxiliary components which are used in the test data comparison. A test pixel is classified as a member of the codeword's set if it satisfies two conditions: (1) Brightness constraint: the intensity (norm[RGB]) should be within some range of the lowest intensity pixel in that codeword's set, and the highest intensity pixel in the codeword's set. (2) Color distance: the color or chromaticity (function of the angle between test vector and codeword rgb) should be within some constant. If both conditions are satisfied, the test pixel is added to the codeword set. Fig. 4 shows the codeword model in RGB space.

Several other auxiliary components are kept as a part of the codeword, and assist in training and pruning the codebook. Namely, these values are the first and last accesses of that codeword, along with the frequency of accesses and maximum length of time between consecutive accesses (or maximum negative run length, MNRL). These values are used to determine whether a codeword should remain in the background or not. Generally, a moving background such as a tree will have a smaller MNRL since the two components of the background (tree and non-tree) will get accessed relatively frequently.

During the training procedure, each pixel maintains a group of codewords which are updated with every incoming pixel. If an incoming pixel happens to be a member of a previously created codeword's set, that codeword is updated accordingly; if the incoming pixel has no codeword, a new one is created with that pixel as its "mean". At the end of the training, if the MNRL of a codeword is too lengthy, or its access frequency too small, that codeword is removed. This pruning step leaves a stable group of codewords for each pixel that define the background.

As testing proceeds, adaptive updating occurs whenever a test pixel falls into the cluster of a codeword. Additionally, when a pixel is classified as foreground, it is added as a codeword to a cache. In turn, if a pixel is repeatedly accessed from the cache, it moves into a layer of the background. These two together provide a basis for robustness under global illumination changes; however, there are additional measures in [12] which can be taken to gain more adaptability. A more detailed description of the training and testing process can be found in [12].

4.2.2 Shadow detection in HSV space [6]

The algorithm presented in [6] is a deterministic non-model based solution for finding moving cast shadows. It is based on the simple idea that shadows change the brightness of the background, but do not really affect the color values. For this reason, the HSV space is chosen to distinguish luminance (V) from chrominance (H and S). Using the logic presented above, a shadow classifier for a given pixel can simply boil down to the following expression:

$$SP_k(x,y) = \begin{cases} 1 & if\ \alpha \le \frac{I_k^V(x,y)}{B_k^V(x,y)} \le \beta \\ & \wedge (I_k^S(x,y) - B_k^S(x,y)) \le \tau_S \\ & \wedge \left| I_k^H(x,y) - B_k^H(x,y) \right| \le \tau_H \\ 0 & otherwise \end{cases} \quad (1)$$

where I_k and B_k are the input and background images, respectively. α, β, τ_S, τ_H are all parameters to be chosen.

Parameter selection is one critical issue here and will be discussed further in Sect. 4.4. However an analysis of the parameter choices and implications is provided in [6], and the choices of τ_S and τ_H in this paper are based upon those findings. It is also fairly straightforward to extend this algorithm to highlight detections due to sunlight. A highlight is defined as an increased luminance as opposed to a decreased luminance, so the equations above can be appropriately modified to find highlights as well.

4.3 Toward generalization of the codebook model

The two algorithms above were chosen for their various properties to implement a robust foreground detection algorithm. However two pressing issues, which are presented below, arose in the combination stage, suggesting a modified approach to these algorithms. First is a discussion about the choice to use the HSV color space for the entire algorithm, which simplifies the number of parameters and lessens the time of calculation. Then the issue of the codeword cluster

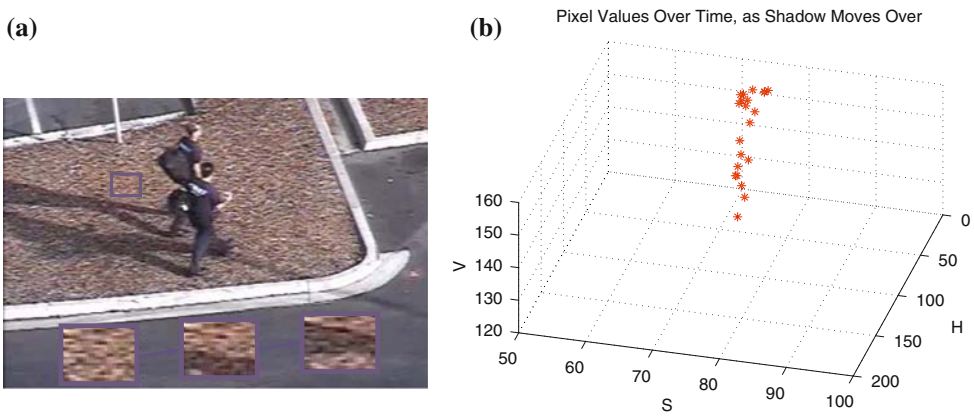

Fig. 5 Evidence for the conical nature of illumination changes in HSV space. As the intensity (V) increases, the variance in the H and S components increases. **a** Sample group of pixels as a shadow moves over. **b** Average value of the selected pixels over time

volumes and shadow and highlight volumes are considered and re-engineered to provide a more cohesive framework.

4.3.1 Color space modification

The original codebook model [12] uses the following method, as mentioned above, to distinguish intensity and chrominance. An intensity value is simply the L2-norm of the RGB components, and the chrominance is measured as a function of the angle between the input and reference values in RGB space. These calculations are only one way of estimating intensity, since the perceived brightness of each of the primary colors is actually quite different—green (0,255,0) actually appears much brighter than blue (0,0,255).

In HSV space, which is used in [6], intensity can be approximated as the V (value) component, where the chrominance encompasses the Hue and Saturation components. It is assumed that this V will be roughly equivalent to the total power of the spectrum; hence it is a good candidate to use for intensity. Upon closer inspection, though, it is apparent that the V component is a function only of the largest of the RGB components, and so it does not encode the entire spectrum.

However the main advantage of the HSV space is that the H and S components, which represent chromaticity, are independent of the intensity approximation V. This implies that the chrominance calculations can be done separately from the luminance calculations Also, according to [8], the HSV space more closely models human vision, where the V can be understood as roughly close to how humans perceive luminance. This implies that it is a more representative approximation of the true illumination intensity than the L2-norm. Finally according to [6], RGB space has been demonstrated to be less accurate than HSV space for detecting shadows.

Therefore to combine the codebook model and shadow suppression techniques, and in the interest of simplicity, both methods are converted into HSV space. This representation eases calculations in both cases, and allows for a smaller parameter set as well. The transformation into HSV space and corresponding distance calculations are shown to be effective and extremely fast.

4.3.2 Introducing the hybrid cone-cylinder volume

The main modification to the models involves the shape of the test volume around a background pixel, whose space defines the "cluster" associated with that pixel. In other words, the test pixels which fall inside this volume are associated with the corresponding background pixel, and labeled as background.

The advantages of having such a fixed volume are discussed in [12], but in essence boil down to simplicity and effectiveness. To develop a more dynamic model would involve laborious statistical calculations and potentially show little, if any, improvements.

However the choice of the volume in the original codebook model [12] does not account for certain cases. As shown in Fig. 4, a cylinder represents the cluster space around the label. Unfortunately it is apparent that at lower intensities, more chromaticity values would have the chance of lying within the background pixel's cylinder. This is evident when examining a background pixel of very low intensity: almost every low-intensity test pixel will be assigned to its cylinder. An unintended consequence of this is that as two similarly grouped pixels increase intensity, they have less chance of being in the same cluster, even though their respective chrominance remains the same.

As a more effective and intuitive choice of volumes, a cone corrects these problems and more precisely covers the colorspace. Therefore the codebook model here is modified to use cones instead of cylinders. Empirically, the conical volume works extremely well. Evidence of the conical nature of a pixel value over time in HSV space can be seen in Fig. 5.

The cone has already been used in shadow suppression techniques [25,24]. As mentioned in the shadow detection algorithm description, a shadow or highlight is simply the same chromaticity value as the original background pixel,

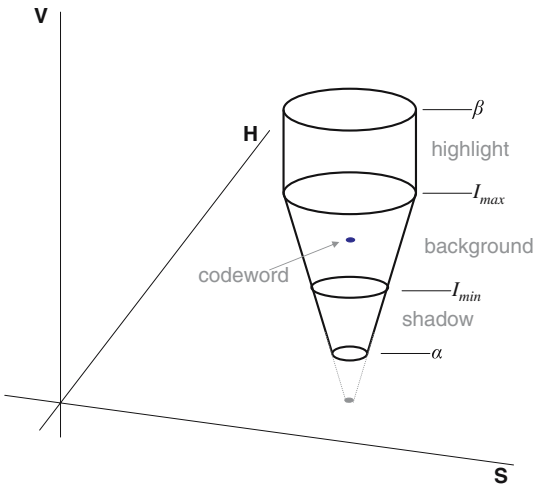

Fig. 6 Hybrid Cone-Cylinder model

with a lower or higher luminance. Thus the shadow and highlight "cones" would lie beyond the range of the codebook cone, adjacent on either side. However after testing, the pure conical highlight detector grabbed too many pixel values within its space and thus pushed up the false negative rate. With a desire for sensitive detections in mind, the highlight volume was limited to a cylinder. Therefore the final hybrid cone-cylinder volume used in the HC3 algorithm is shown in Fig. 6.

4.4 Incorporation of satellite data

The one clear drawback of the HC3 is its large parameter space, however that problem is precisely what is addressed by the additional sensor data. As can be seen in Fig. 6, there are four parameters directly related to the illumination conditions of the scene. The length of the conical portion is bounded on the bottom by the shadow parameter α, where the top of the shadow region is $I_{\min}$, also corresponding to the bottom of the codebook cone. Similar descriptions fit for $I_{\max}$ and β above that as well.

As the amount of direct sunlight increases, shadows become more pronounced, and so α should be lowered. This will cause more pixels to be classified as shadow. However if there is higher variability in the scene illumination, corresponding to a partly cloudy day, the difference $I_{\max} - I_{\min}$ should be increased, so that a greater range is classified as foreground. A greater $I_{\max}$ and β may also be required as direct sunlight intensely falls upon a region that was just behind a small cloud. Finally, if the scene illumination is due to overcast clouds, it will stay relatively constant, and so all the values in the parameter set can be brought much closer to their mean.

The intuitive nature of this model and its parameters make it easy to hand-pick models which will work well under various conditions. However this model-fitting stage could be formulated as a classification problem and trained automatically. That is, we would like to find a certain set of parameters which will work well on sunny days, and another set will work well on cloudy days, etc. However to save the burden of manually labeling so much data, we instead pose this model-fitting as a clustering problem, where we cluster the outputs of many models on a series of diverse inputs. This leads to a small set of representative models each of which have unique characteristics making them good in different situations, as is verified by experimental analysis. The algorithm developed for automatic discovery of model parameters is described in Algorithm 1, and carried out in Sect. 5.2.1.

Algorithm 1 Find HC3 parameters for different illumination conditions

1: Generate N parameter sets $\chi_i = [\alpha_i, I_{\min_i}, I_{\max_i}, \beta_i], i \in N$.
2: Choose input test sequence $\mathcal{U}$ of length L frames which includes different illumination conditions such as {overcast, partly cloudy, clear, ...}.
3: **for** $i = 1$ to N **do**
4: Evaluate HC3 background subtraction with parameters χ_i on sequence $\mathcal{U}$.
5: Put *foreground pixel counts* for each frame into a size $L \times 1$ vector T_i.
6: **end for**
7: Perform principal component analysis on $[T_1, T_2, \ldots, T_N]$ to obtain vectors $[t_1, t_2, \ldots, t_N]$, length$(t_i) = l << \text{length}(T_i) = L$.
8: Group vectors $[t_1, t_2, \ldots, t_N]$ using the k-means clustering algorithm, into $n < N$ clusters.
9: **for** $k = 1$ to n **do**
10: Let $\mathcal{V} = \{t_1, \ldots, t_j\}$ be the set of vectors belonging to cluster k.
11: Find the closest vector to the mean, $t_k^* = \arg\min_{\mathcal{V}} \|t_j - \text{mean}(\mathcal{V})\|_2$.
12: Let χ_k^* be the set of model parameters corresponding to t_k^*.
13: Find corresponding label$_k \in$ {overcast, partly cloudy, clear, ...} based on performance of model χ_k^* on $\mathcal{U}$.[1]
14: **end for**
15: Given illumination condition corresponding to label$_k$ from satellite data: choose HC3 model parameters χ_k^*.

We evaluate the performance of each of the N parameter sets in a range of conditions from sunny to overcast, and we compare these models based on the number of foreground pixels per frame. In other words, each of the N parameter sets generates a vector of pixel counts in response to the input test frames $\mathcal{U}$. Many models should perform similarly to others, however this case could be made without thousands of measurements. Thus a Principal Component Analysis is performed on the output data, leaving each model with a low-dimensional feature vector.

The models are then clustered by using k-means on the feature vectors. The model closest to the mean of each cluster is then chosen as the representative model for that cluster.

Finally, those models are designated labels according to their performance in each environment.[1]

5 Experiments and evaluation

The evaluation of this model was broken up into two sections. Section 5.1 provides an experimental validation of the HC3 model by itself, which is shown to be generally useful for many applications. Section 5.2 is an analysis of the fusion of the satellite data into this model.

5.1 Experimental validation of HC3 model without satellite data

Several experiments were performed to determine the potential effectiveness of the HC3 model by itself on data in various outdoor environments, including daytime and nighttime. First, two scenes were manually marked up as ground truth data, and those were used to quantitatively compare the HC3 model and original codebook model. Additionally, results are presented below on several other difficult environmental contexts.

The first two experiments were conducted on two sequences, each 100 frames long (3.5 s), which were selected for their complexity and appearance of cast shadows. These frames were manually marked up to represent ground truth. For each experiment, both the original codebook model [12] and the HC3 model were trained on a series of 500 frames and tested on the subsequent 100 frames. The training and testing procedures, which are the same for the codebook and HC3 models, are described in more detail in [12]. It is worth noting that the morphological open and close operators were used similarly in both models to remove noise and fill in holes, slightly reducing the sensitivity. For each frame, the errors were detected and a corresponding detection versus false alarm rate was plotted on an ROC graph. The true positive rate and false positive rate are defined as follows: $\text{TPR} = \frac{\text{true positives}}{\text{true positives+false negatives}}$ and $\text{FPR} = \frac{\text{false positives}}{\text{false positives+true negatives}}$.

Figure 7 shows results from a simple scene where two people are walking along the road with very long cast shadows. As is evident, the HC3 model does well in classifying shadows as background. Figure 8 displays a more complex scene, with moving background in the waving grass and dust storms, as well as cast shadows and dust kicked up by the car. The HC3 model correctly classifies most of the shadows as background, and the dust is classified as highlights and thereby background as well.

Evidently the ROC plots show the effectiveness of this algorithm. The errors in the HC3 model are generally well separated from the original model; this is mostly due to the classification of shadows and highlights as background. Table 1 compares the average error rates over all 100 frames for each sequence. While the detection rate of true positives remains quite similar for both algorithms, the false positive rate drops relatively significantly on average using HC3.

Table 1 Average FPR and TPR values. This demonstrates the performance improvements of the proposed HC3 model over the original codebook model, as well as a mixture of Gaussians with HSV shadow suppression

	People walking (%)		Car on dirt road (%)	
	FPR	TPR	FPR	TPR
HC3 model	1.1	77.9	0.2	79.4
Original model	3.7	68.7	2.4	83.4
MoG+HSV model	2.6	71.1	0.3	67.1

For an additional comparison measure, the Mixture of Gaussians background model [30] was implemented along with the HSV shadow suppression scheme in [6]. The two sequences were tested with this model for the background, and the average error rates are shown in Table 1. The HC3 model performs with a significantly higher sensitivity as evidenced by the higher True Positive Rates, while maintaining lower False Positive errors, as compared to the "MoG+HSV" model.

In Fig. 9 is an example of a sensitive people-detector using an omnidirectional camera. Even though the background is somewhat stable, the foreground objects are quite small and could easily be confused with noise. However in each case the people are detected quite accurately.

In terms of performance, the HC3 background segmentation phase runs at approximately 40 frames per second on videos of size 320×240. Thus it is quite clearly viable for real-time applications. Due to pruning and adaptive updating, the sizes of the codebook and cache stay relatively constant. In fact for most pixels 1–2 codewords are sufficient on average. A select few pixels need more codewords to deal with fluctuating background scenes.

5.2 Evaluation of HC3 with satellite data

Here the validation of the incorporated satellite data occurs. As demonstrated in the applications in Sect. 5.1, the HC3 model is generally useful, but only if the correct parameters are known. For daytime data, the cloudiness data from the satellite provide a direct means to estimate the correct

[1] Note that the mapping in step 13 of Algorithm 1, namely of which χ_k^* would work well in which condition, could be found automatically if there were a sufficient set of training data. However since each model actually has an intuitive link to the intensity measure, in this paper the correspondences between each satellite-measured parameter and codebook model are manually labeled based on theoretical performance, as explained above and in Table 2.

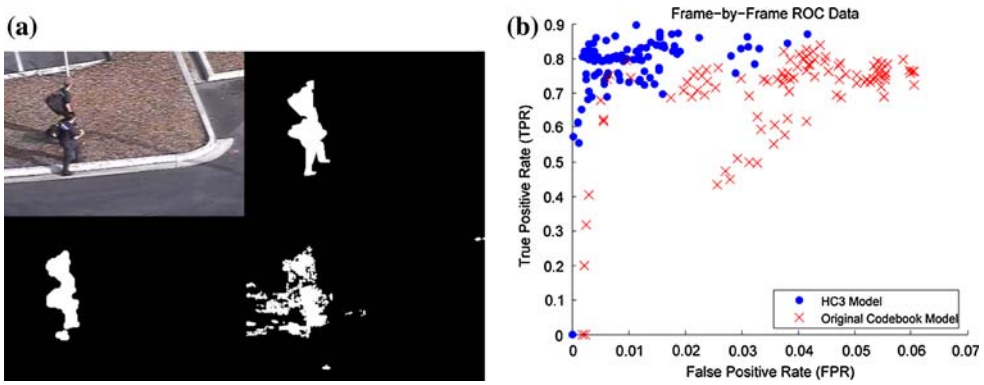

Fig. 7 Quantitative comparison of HC3 and original codebook algorithms—walking people with shadows. **a** Input frame (*top left*), ground truth (*top right*), HC3 Model (*bottom left*), Original Codebook Model (*bottom right*). **b** Errors for each of 100 frames in the sequence

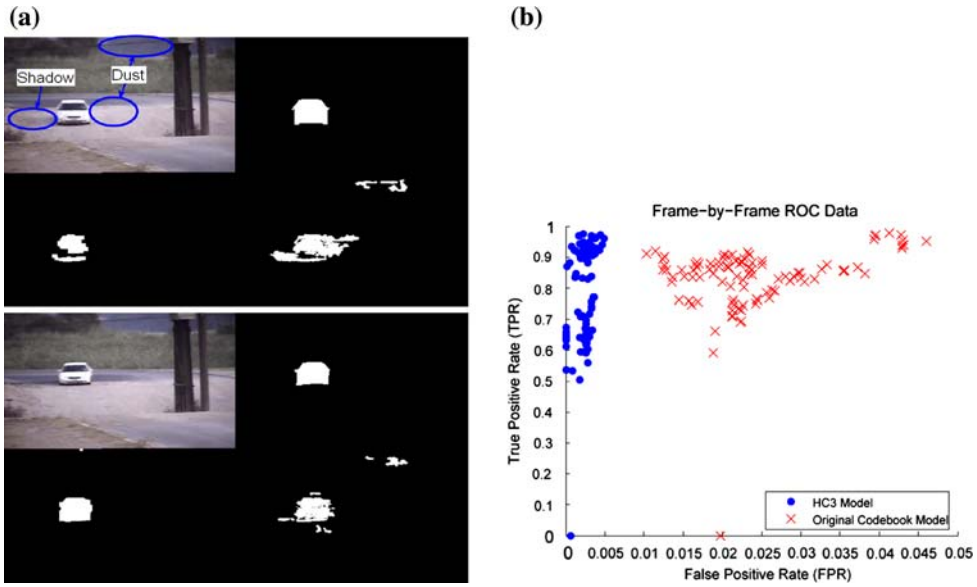

Fig. 8 Quantitative comparison of HC3 and original codebook algorithms—Car on dirt road. **a** Input frame (*top left*), ground truth (*top right*), HC3 model (*bottom left*), original codebook model (*bottom right*). **b** Errors for each of 100 frames in the sequence

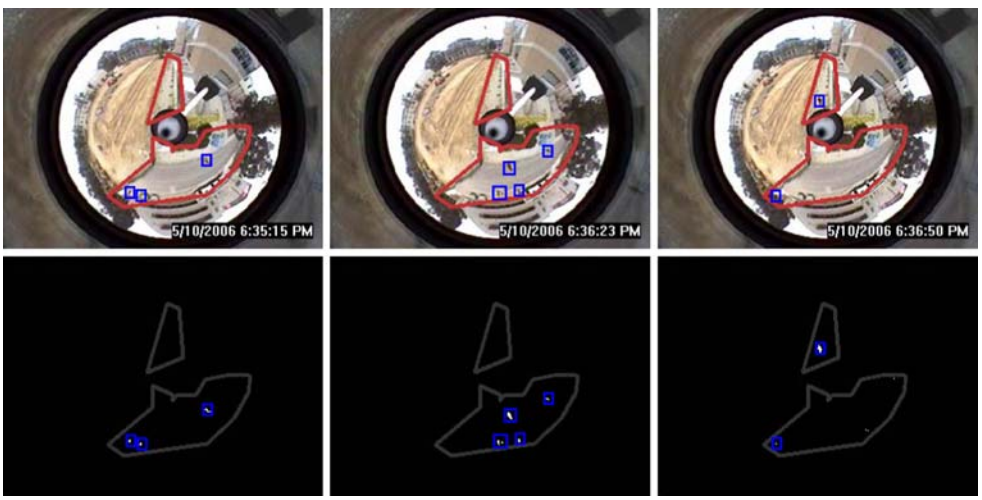

Fig. 9 Example of sensitive detection of people with an omni-directional camera. Input with a zone of interest (*top*), foreground (*bottom*). The HC3 model is able to detect very small objects with little noise

Table 2 Automatically discovered parameters. The first model, with very tight parameters, would work well on a cloudy day with no shadows. Models 2 and 3 are similar in that both have very small lower bounds for the shadow detector, implying they would remove shadows well on a sunny day. However one of the models has a larger $I_{\max}$, which would allow for more intense highlights from sunlight on a brighter day, for example

Model	α	β	$I_{\min}$	$I_{\max}$	Intuitive label
1	0.9	1.1	0.8	1.1	'overcast'
2	0.1	1.3	0.7	1.1	'partly cloudy'
3	0.1	1.3	0.7	1.3	'sunny'

parameters. The following set of experiments were performed to validate the use of the satellite data.

5.2.1 Automatic discovery of "Good" HC3 parameters

A primary goal given the satellite intensity measure is to determine what kind of model works well on any given condition; the process for doing so is described in Sect. 4.4, Algorithm 1. The model in this case, is defined as the choice of parameters $\chi_i = [\alpha, I_{\min}, I_{\max}, \beta]$ for the HC3 model. Following Algorithm 1, a range of $N = 90$ parameter vectors were generated and the corresponding background models were trained and tested on a representative sequence of $L = 5,000$ frames. This sequence included a range of conditions from sunny to overcast.

Following Algorithm 1, it was found by PCA that 99.5% of the information was stored in just $l = 5$ components. Thereby each of the 90 models was left with a vector of 5 representative feature points, which were then clustered using a k-means algorithm. It was found experimentally that $n = 3$ clusters were satisfactory to separate the data well. The vector closest to the mean of each of these clusters was then automatically chosen as a representative model, as defined in Step 11 of Algorithm 1. These three chosen sets of values are shown in Table 2. These parameter sets, or models, can be seen intuitively to work well on overcast scenes (Model 1), partly cloudy scenes (Model 2), and sunny scenes (Model 3). This mapping is shown to be valid in the next experiments.

5.2.2 Day-long experiment

The effectiveness of the information obtained from the satellite-based cloudiness parameter is now examined. In the following experiment we use the three models obtained above as corresponding to their intuitive usage—overcast, partly cloudy, and sunny respectively. In other words, we assume that the cloudiness parameter has told us to choose a certain model for that given time. Then we show the results of such adaptive model switching in real-time.

Over an entire 12-h day of data, we examine a sequence of 5,000 frames for each hour. For each sequence, the three background models whose parameters were automatically discovered above were trained on the first 500 frames and tested on the rest. As a measure of sensitivity, the number of detected foreground pixels was accumulated for each frame. The mean of this foreground count over the sequence was used as a measure of the output of the model. This metric was used as opposed to ROC metrics due to the lack of ground truthing for so many frames (however, a select few frames were marked up and used to analyze performance in Sect. 5.2.3).

Since the HC3 background model adaptively updates to deal with global illumination changes, in a similar manner to the original codebook model [12], it is a valid approximation to retrain the background at the beginning of each new sequence. However this adaptive updating or retraining step of each individual model does not greatly affect the performance differences between models. The significance of the proposed algorithm is that the cloudiness parameter motivates adaptively switching between models, to find the most appropriate one. This is the source of the most performance improvement, as can be seen in Fig. 10. Sample outputs from each time period and corresponding models can also be seen in Fig. 11.

Note that the satellite-based parameter informed the algorithm to switch models when the cloudiness conditions changed. The result of these switches was that the foreground was much more sensitive during overcast conditions, since there was less illumination variability and no shadows. When the scene became partly cloudy, suddenly the lighting became much more variable, and shadows started appearing. By switching to Model 2 and then Model 3, the algorithm was able to detect and discard the shadows and illumination changes as part of the background. Some of the sensitivity of the algorithm was lost, but the satellite-based model did outperform any of the other three models taken alone, as can be seen in Fig. 11 and Table 3. It is also worth mentioning that the local effects of the immediate switching are relatively minor, as the changes occur to improve performance immediately.

Examples of the scene and corresponding satellite image during each period can be seen in Fig. 1, and output results can be seen in Fig. 11. These outputs clearly show the advantages of being able to automatically switch between models when different conditions arise.

5.2.3 Ground-truth verification

To quantitatively verify the improvements in performance, a number of scenes were manually marked up to serve as ground truth. These included 20 difficult scenes from cloudy, partly cloudy, and sunny conditions; with and without

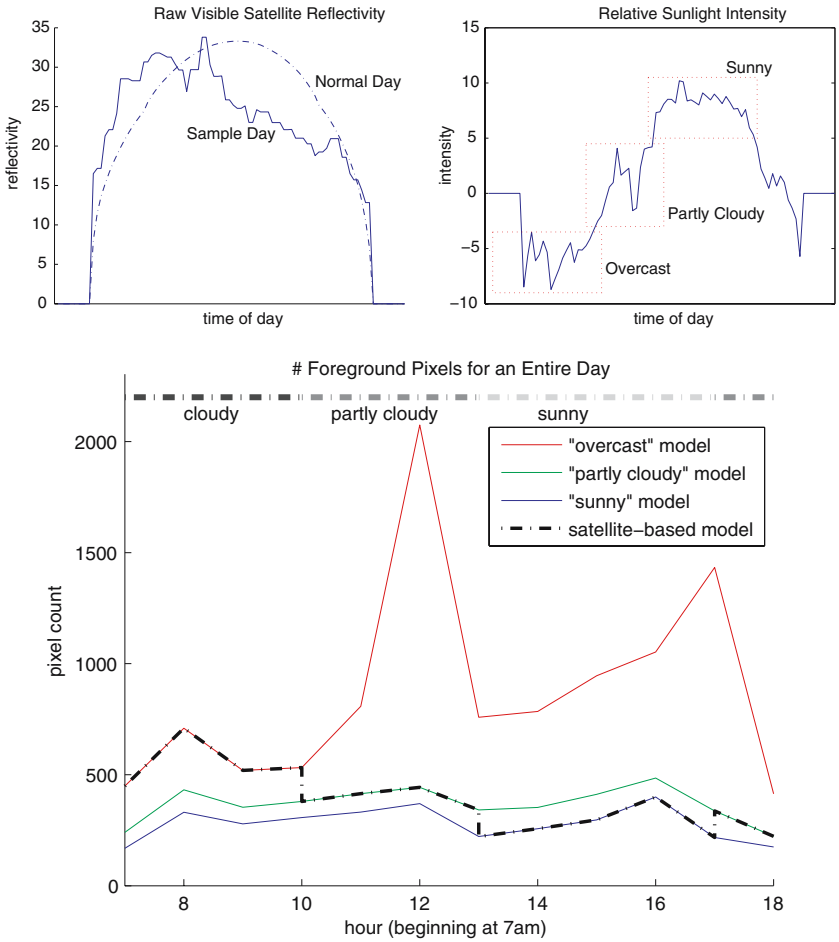

Fig. 10 Day-long experiment. Three models were trained and tested over an entire day-long sequence, whose satellite intensity measurements shown above (same as Fig. 3). The measurements were used to switch between models, giving the final satellite-based model shown above. Model 1 turns out to be more sensitive during cloudy conditions, but Models 2 and 3 perform better with illumination changes and shadows during sunny scenes. Sample outputs can be seen in Fig. 11

Table 3 Ground truth verification. True positive rates/false positive rates for each of the models, over several different difficult conditions and overall. Also, the overall raw errors(= FP + FN) are shown. The "Switching" model maintains the better overall performance by always keeping low FPRs and high TPRs, while the other models have poorer worst-case performances

	Ovc HC3 TPR/FPR	Pcl HC3 TPR/FPR	Sun HC3 TPR/FPR	**Switching** TPR/FPR	MoG+HSV TPR/FPR
Ovc conditions	55.84/0.028	40.82/0.010	30.53/0.004	**55.84/0.028**	56.39/0.327
Pcl conditions	59.21/0.649	44.37/0.037	35.35/0.013	**44.37/0.037**	57.94/1.859
Sun conditions	83.78/0.292	60.90/0.074	52.86/0.058	**52.86/0.058**	73.55/0.203
Overall	64.54/0.274	47.39/0.035	38.07/0.022	**51.83/0.039**	61.59/0.718
Raw errors	443.39	225.61	242.94	**223.11**	947.39

shadows, highlights, and during global illumination changes. For each of these environments, 5 models were tested: the three HC3 models that were automatically discovered in Sect. 5.2.1, the proposed satellite-based "Switching" HC3 model, and a reference "Mixture of Gaussians" model with HSV shadow suppression as described in Sect. 5.1.

Results can be seen in Table 3. Displayed are TPR and FPR for each environment as well as overall. Additionally, average raw pixel errors (defined as false positive count + false negative count) are shown.

Both the "overcast HC3" model and the "MoG+HSV" models have very good true positive rates overall; however, their false positive rates suffer under harsher conditions. The "sunny HC3" model has relatively low FPRs, but clearly suffers in detecting true positives. The "partly cloudy HC3" model could be a fair compromise; however there is still a

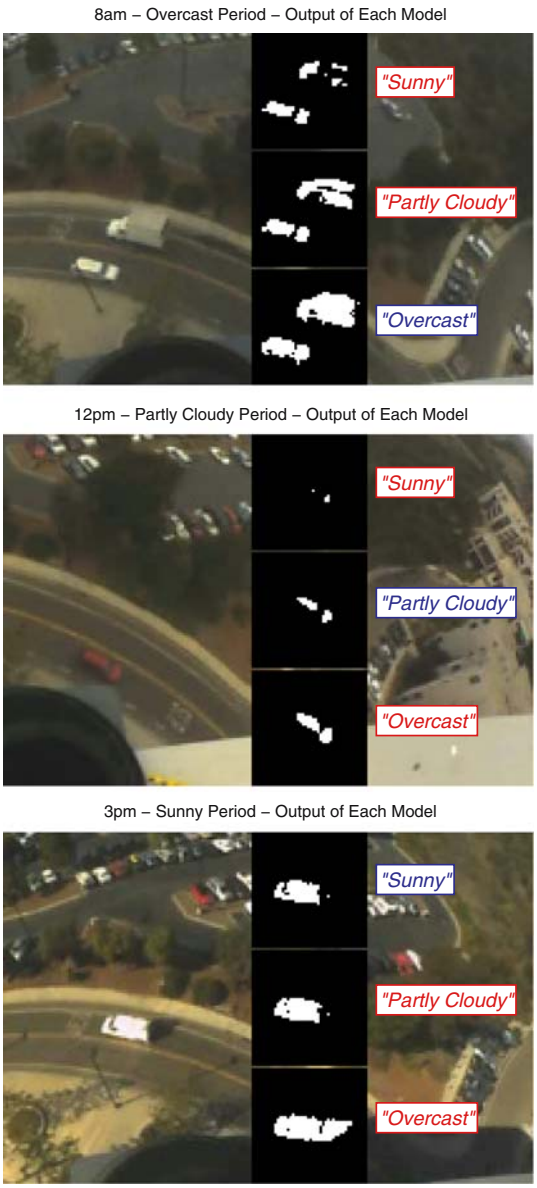

Fig. 11 Day-long experiment. Three models were trained and tested over an entire day-long sequence, as seen in Fig. 10. The "overcast" model is seen to be most sensitive in all cases, however this is a drawback when dealing with shadows in the bottom scene. The "partly cloudy" model is sensitive enough to detect objects in overcast conditions, yet still remove some shadows during sunny conditions. The "sunny" model loses traction during overcast periods, however produces more accurate foreground outputs during sunny conditions. Thus a clear improvement can be seen by using satellite data to switch between the models. The satellite-based choice of models is highlighted in blue for each case

significant variance in its performance over different conditions. By switching between models, we can avoid worst-case scenarios of each individual model, and *consistently* output good results. Sample outputs displaying these effects can be seen in Fig. 11. Thus an overall improvement in performance can be achieved with the addition of the proposed satellite-based cloudiness parameter.

6 Conclusions and future work

This paper focused on the introduction of the satellite as an additional sensor, which is uncommon in the modern computer vision field, especially with respect to object detection and tracking. A cohesive background and shadow model which could fully utilize the satellite-based cloudiness estimate was developed and shown to be quantitatively better, independently of the satellite data, in Sect. 5.1. Then in Sect. 5.2, the satellite-based adaptive HC3 model was shown to improve sensitivity and accuracy of detection, with comparatively fewer errors than reference models.

The Hybrid Cone-Cylinder Codebook (HC3) background modeling algorithm was designed to be real-time and robust against shadows, highlights, and moving backgrounds. Additionally, it is an adaptive model and thus capable of handling global illumination changes. These properties make the algorithm a good candidate to run on very long video sequences, in applications such as long-term surveillance and vehicle counting.

When the environmental context of the scene changes, however, satellite data was shown to be useful. The cloudiness parameter extracted from the satellite image was able to predict the amount of shadows as well as the variability or noise level of the scene. This information allowed the background model to change parameters in real time to adapt to the environment. Additionally, the satellite imagery was shown to be straightforward to integrate into the HC3 model, at a very minimal cost in terms of setting up additional sensors.

It would be interesting to explore the interplay between the satellite-based cloudiness parameter and other background models such as mixture of Gaussians. Doing so may prove difficult since it might not be immediately clear which parameters to adjust. As a next step, however, it will certainly be necessary to comprehensively compare the satellite-based adaptive HC3 algorithm, with other background subtraction methods and on longer data sets. Such analysis will provide motivation for re-tuning and adjusting the current model to apply in more general situations.

Another future direction involves the fact that even though HSV space empirically works quite well, other color spaces may also be appropriate. These include spaces such as CIE L*a*b*, which encode the estimated luminance in one component and chrominance in two other independent components. It may be the case that such a space is slightly more precise; however, careful analysis will be necessary and parameters will need to be re-tuned.

Finally, we have provided motivation and analysis for the use of satellite imagery during the background removal phase of object detection. It is a cheap, ubiquitous, and extremely informative source of information, taking into account that existing satellites already cover most of the globe and their data is publicly accessible. The use of satellite data for

estimation of environmental conditions is advantageous in that it also requires minimal prior knowledge of the scene. We have developed an adaptive background model that incorporates the satellite-based cloudiness estimate, and we have shown that model to have comparatively low error rates, with sensitive performance and effective shadow suppression over various environmental conditions.

Acknowledgments The authors are grateful for the support of the Technical Support Working Group of the US Department of Defense. They are most appreciative of the contributions and assistance of their colleagues from the CVRR laboratory, who made the experimental phase of the research possible. They would like to thank the reviewers for their careful reading and insightful comments.

References

1. Amnuaykanchanasin, P., Thongkamwitoon, T., Srisawaiwilai, N., Aramvith, S., Chalidabhongse, T.H.: Adaptive parametric statistical background subtraction for video segmentation. In:Proceedings of the Third ACM Int'l Workshop on Video Surveillance & Sensor Networks, pp.63–66, New York. ACM Press (2005)
2. Baba, M., Asada, N.: Shadow removal from a real picture. In: SIGGRAPH '03: ACM SIGGRAPH 2003 Sketches & Applications, pp.1–1, New York. ACM Press (2003)
3. Cavallaro, A., Salvador, E., Ebrahimi, T.: Shadow-aware object-based video process. IEE Visi. Image Signal Process. **152**(4), 14–22 (2005)
4. Chandrasekhar, S.: Radiative Transfer., 2nd edn. Dover Publications (1960)
5. Colwell, R. (ed.) : Manual of Remote Sensing, chap. 5, pp.165–230. American Society of Photogrammetry, Falls Church 2nd edn, (1983)
6. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving shadow suppression in moving object detection with HSV color information. In: IEEE Int'l Conf. Intelligent Transportation Systems, pp.334–339 (2001)
7. Doshi, A., Trivedi, M.M.: Hybrid cone-cylinder codebook model for foreground detection with shadow and highlight suppression. In: IEE Int'l Conf. on Advanced Video and Signal based Surveillance. Australia (2006)
8. Elgammal, A.M., Harwood, D., Davis, L.S.: Non-parametric model for background subtraction. In: European Conf. on Computer Vision, pp.751–767 (2000)
9. Han, H., Wang, Z., Liu, J., Li, Z., Li, B., Han, Z.: Adaptive background modeling with shadow suppression. In: Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE, vol. 1, pp.720–724 (2003)
10. Horprasert, T., Harwood, D., Davis, L.S.: A statistical approach for realtime robust background subtraction and shadow detection. In: Proceedings of IEEE Frame Rate Workshop, pp.1–19 (1999)
11. Joseph, J.H., Wiscombe, W., Weinman, J.: The delta-eddington approximation for radiative flux transfer. J. Atmos. Sci. **33**(12), 2452–2459 (1976)
12. Kim, K., Chalidabhonse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. Real-Time Imaging **11**(3), 167–256 (2005)
13. Lee, D.-S., Hull, J.J., Erol, B.: A bayesian framework for gaussian mixture background modeling. IEEE Int. Conf. Image Process. **3**, 973–976 (2003)
14. Liou K.-N.: Radiation and Cloud Processes in the Atmosphere: Theory, Observation, and Modeling. Oxford University Press (1992)
15. Liou, K.-N., Wittman, G.D.: Parameterization of the radiative properties of clouds. J. Atmos. Sci. **36**(7), 1261–1273 (1979)
16. Lluis, J., Miralles, X., Bastidas, O.: Reliable real-time foreground detection for video surveillance applications. In: Proceedings of the Third ACM Int'l Workshop on Video Surveillance & Sensor Networks, pp.59–62, New York. ACM Press (2005)
17. Meador, W., Weaver, W.: Two-stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement. J. Atmos. Sci. **37**(3), 630–643 (1980)
18. Mikic, I., Kogut, G., Cosman, P., Trivedi, M.M.: Moving shadow and object detection in traffic scenes. Int. Conf. on Pattern Recogn. (2000)
19. NASA: Clouds and radiation. Earth observatory: http://www.earthobservatory.nasa.gov/Library/Clouds/
20. GOES I-M Databook rev. 1, Space Systems-Loral. NASA, http://www.rsd.gsfc.nasa.gov/goes/text/goes.databook.html. Chapter on Imager, pp. 20–38 (2001)
21. NOAA: Current GOES-10 1 km VIS Satellite Imagery Centered on San Diego. http://www.sat.wrh.noaa.gov/satellite/1km/Sandiego/VIS1SAN.GIF
22. NOAA: Geostationary Satellite Server. http://www.goes.noaa.gov/
23. Piccardi, M.: Background subtraction techniques: a review. IEEE Int. Conf. Systems Man Cybern. **4**, 3099–3104 (2004)
24. Porikli, F., Thornton, J.: Shadow flow: A recursive method to learn moving cast shadows. In: IEEE International Conference on Computer Vision (ICCV) (2005)
25. Porikli, F., Tuzel O.: Bayesian background modeling for foreground detection. In:Proceedings of the Third ACM Int'l Workshop on Video Surveillance & Sensor Networks, pp.55–58, New York. ACM Press (2005)
26. Prati, A., Mikic, I., Cucchiara, R., Trivedi, M. M.: Analysis and detetcion of shadows in video streams: a comparative evaluation. In: IEEE CVPR Workshop on Empirical Evaluation Methods in Computer Vision (2001)
27. Prati, A., Mikic, I., Trivedi, M., Cucchiara, R.: Detecting moving shadows: Algorithms and evaluation. IEEE Trans. Pattern Anal. Mach. Intelli. **25**, 918–923 (2003)
28. Rossow, W.B.: Measuring cloud properties from space: a review. J. Climate **2**(3), 201–213 (1989)
29. Stauder, J., Mech, R., Ostermann, J.: Detection of moving cast shadows for object segmentation. IEEE Trans. Multimedia **1**(1), 65–76 (1999)
30. Stauffer, C., Grimson, W.L.: Adaptive background mixture models for real-time tracking. CVPR99 **II**, 246–252 (1999)
31. Tattersall, S., Dawson-Howee, K.: Adaptive shadow identification through automatic parameter estimation in video sequences. In: Irish Machine Vision and Image Processing Conference (IMVIP 2003), pp.57–64 (2003)
32. Toyama, K., Krumm, J., Brumitt, B. Meyers, B.: Wallflower: principles and practice of background maintenance. In:ICCV (1), pp.255–261 (1999)
33. Trivedi, M.M.: Analysis of high-resolution aerial images. In: Kasturi, R., Trivedi, M.M. (eds.) Image Analysis Applications, pp.281–305. Marcel Dekker (1990)
34. Trivedi, M.M., Gandhi, T.L., Huang, K.S.: Distributed interactive video arrays for event capture and enhanced situational awareness. IEEE Intell. Systems **20**(5), 58–66 (2005)
35. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: Real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 780–785 (1997)
36. Zhu, X., Arking, A.: Comparison of daily averaged reflection, transmission, and absorption for selected radiative flux transfer approximations. J. Atmos. Sci. **51**(24), 3580–3592 (1994)