Using Cloud Shadows to Infer Scene Structure and Camera Calibration

Nathan Jacobs, Brian Bies, Robert Pless
Computer Science and Engineering, Washington University in St. Louis, MO, USA
{jacobsn, bwb4, pless}@cse.wustl.edu

Abstract

We explore the use of clouds as a form of structured lighting to capture the 3D structure of outdoor scenes observed over time from a static camera. We derive two cues that relate 3D distances to changes in pixel intensity due to clouds shadows. The first cue is primarily spatial, works with low frame-rate time lapses, and supports estimating focal length and scene structure, up to a scale ambiguity. The second cue depends on cloud motion and has a more complex, but still linear, ambiguity. We describe a method that uses the spatial cue to estimate a depth map and a method that combines both cues. Results on time lapses of several outdoor scenes show that these cues enable estimating scene geometry and camera focal length.

1. Introduction

Although clouds are among the dominant features of outdoor scenes, with few exceptions visual inference algorithms treat their effects on the scene as noise. However, the shadows they cast on the ground over time give novel cues for inferring 3D scene models. Clouds are one instantiation of the first law of geography, due to Waldo Tobler: *"Everything is related to everything else, but near things are more related than distant things."* In a sense, we are applying this law to the problem of estimating a depth map from time-lapse imagery. The basic insight is that there is a relationship between the time series of intensity at two pixels and the distance between the imaged scene points. We describe two cues, one spatial and one temporal, that further refine this relationship. We also present algorithms that use these cues to estimate a depth map.

The first cue is purely spatial; it ignores the temporal ordering of the imagery and does not require a consistent wind velocity. We begin by considering that if the relationship between pixel time-series correlation and 3D distance is known, then there is the simple problem: Given an image and the 3D distance between every pair of scene points, find the 3D model of the scene that is consistent with the camera geometry and the distance constraints. However, the relationship between correlation and distance is unknown and depends on the scene and the type of clouds in the scene. We present a method that simultaneously solves for the relationship between distance and correlation and for a corresponding 3D scene model.

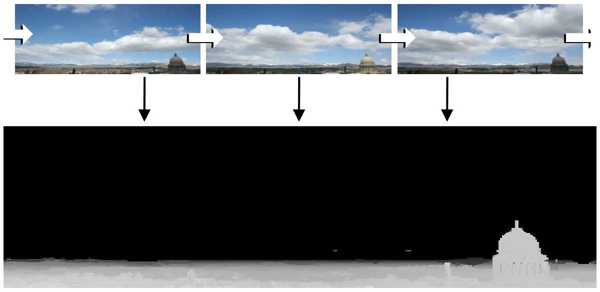


Figure 1: Clouds and cloud shadows are significant sources of appearance variation in outdoor scenes. In this work, we explore using cloud shadows as a cue to infer scene structure and camera focal length. The depth map was created using the method described in Section 3.1

The second cue requires higher frame-rate video and the ability to estimate the temporal offset between a pair of pixel-intensity time series. This temporal delay, coupled with knowledge of the wind velocity, allows us to define a set of linear constraints on the scene geometry. With these constraints there is a very clean geometric problem: Given an image and the distance between every pair of pixels projected onto the wind direction, solve for a 3D scene structure that is consistent with the projected distances and the camera geometry.

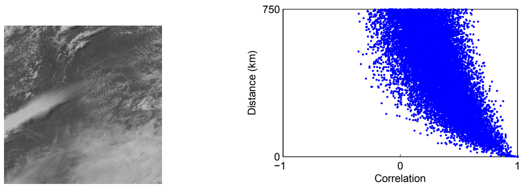
Our work falls into the broad body of work that aims to use natural variations as calibration cues and each of these methods makes certain assumptions. We, for example, require weather conditions in which it is possible to, mostly, isolate the intensity variations due to clouds from other sources of change. The methods we describe are a valuable addition to the emerging toolbox of automated outdoor-camera calibration techniques.

1.1. Related Work

Stochastic Models of Cloud Shapes The structure of clouds has been investigated both as an example of natural images that follow the power law and within the atmospheric sciences community. Natural images of clouds often exhibit structure where the expected correlation between two pixels is a function of the inverse of their distance [3], and furthermore, there is a scale invariance that may be characterized by a power law (with the ensemble spatial frequency amplitude spectra ranging from $f^{-0.9}$ to f^{-2} [1]). These trends have been validated for cloud cover patterns, with empirical studies demonstrating that the 2D auto-correlation is typically isotropic [15], but that the relationship of spatial-correlation to distance varies for different types of clouds (e.g. cumulus vs. cirrus clouds) [18]. This motivates our use of a non-parametric representation of the correlation-to-distance function.

Shadows in Video Surveillance For video surveillance applications, clouds are considered an unwanted source of image appearance variation. Background models explicitly designed to capture variation due to clouds include the classical adaptive mixture model [14] and subspace methods [12]. Farther removed from our application, object detection/recognition is disturbed by cast shadows because they can change the apparent shape and cause nearby objects to be merged. Several algorithms seek to minimize these effects, using a variety of approaches [13], including separating brightness and color changes [5].

Geometry and Location Estimation Using Natural Variations Within the field of remote sensing, shadows have long been used to estimate the height of ground structures from aerial or satellite imagery [4]. Recent work in analysis of time-lapse video from a fixed location have used changing lighting directions to cluster points with similar surface normals [9]. Other work has used known changes in the sun illumination direction to extract surface normal of scene patches [16], define constraints on radiometric camera calibration [8, 17], and estimate camera geo-location [17]. Work on the AMOS (Archive of Many Outdoor Scenes) dataset of time-lapse imagery demonstrates consistent diurnal variations across most outdoor cameras and simple methods for automated classification of images as *cloudy* or *sunny* [6]. This supports methods that estimate the geolocation of a camera, either by finding the maximally correlated location (through time) in a satellite view, or interpolating with respect to a set of cameras with known positions [7]. The recently created database of "webcam clip-art" includes camera calibration parameters to facilitate applications to illumination and appearance transfer across scenes [11].



(a) (b)

Figure 2: The correlation-to-distance relationship between sample points has a similar form at many scales. Here we show long-range correlations due to clouds in a sequence of satellite images. (a) A single 1 km-scale image from the visible-light band of the GOES-12 satellite. (b) The relationship between correlation and distance for a set of 90 satellite images captured during the summer months of 2008. Notice that the expected value of distance is a monotonically decreasing function of correlation and that the variance in the conditional distribution is much lower at closer distances.

2. Structural Cues Created by Cloud Shadows

The image of cloud shadows passing through a scene depends upon the camera and scene geometry. Here we describe two properties of outdoor-scene time lapses that depend on cloud shadows, are easy to measure, and, as we show in Section 3, can be used to infer camera and scene geometry.

2.1. Geographic Location Similarity

The closer two points are in the world the more likely they are to be covered by the shadow of the same cloud. Thus, for a static outdoor camera, the time series of pixel intensities are usually more similar for scene points that are close than for those that are far.

We begin by considering the correlations that arise between pixels in satellite imagery. The statistical properties of this approximately orthographic view are similar to the spatial properties of the cloud shadows cast onto the ground. We empirically show the relationship between correlation and distance for a small dataset of visible-light satellite images (all captured at noon on different days during the summer of 2008). The scatter plot in Figure 2, in which each point represents a pair of pixels, shows that the correlation of the pixel intensities is clearly related to the distance between the pixels. Additionally, It shows that the expected value of distance is a monotonically decreasing function of correlation.

This relationship also holds at a much finer scale. To show this, we compute correlation between pixels in a timelapse video captured by a static outdoor camera on a partly cloud day. Since we do not know the actual 3D distances between points we cannot generate a scatter plot as in the satellite example. Instead, Figure 3 shows examples of correlation maps generated by selecting one landmark pixel and comparing it to all others. The false-color images, colored by the correlation between a pair of pixels, clearly show that correlation is related to distance.

We note that different similarity measures between pairs of pixels could be used (and, in some cases, would likely work much better). We choose correlation because it is simple to compute online and works well in many scenes. Our work does not preclude the use of more sophisticated similarity metrics that explicitly reason about the presence of shadows using, for example, color cues. In Section 3, we show how to infer the focal length of the camera and a distance map of the scene using correlation maps as input.

2.2. Temporal Delay Due to Cloud Motion

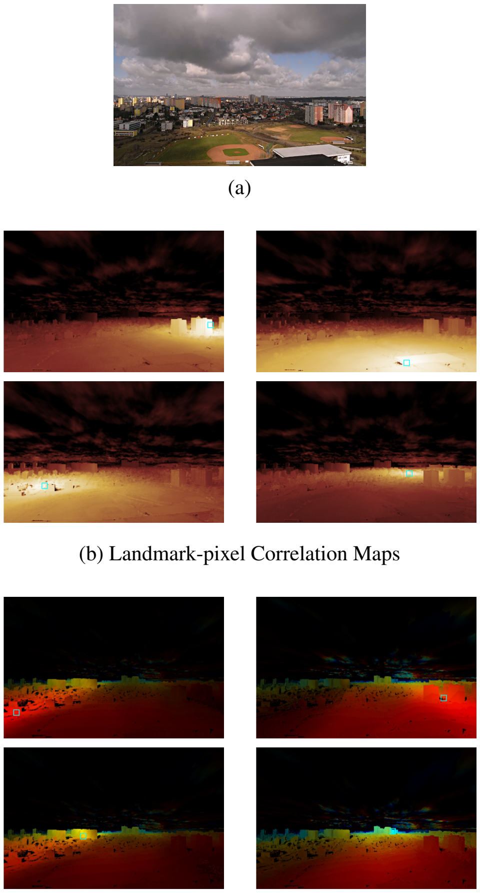
As clouds pass over a scene, each scene point exhibits a sequence of light and shadow. In the direction of the wind these time series are very similar but temporally offset relative to the geographic distance between the points (see Figure 4). Also, for short distances perpendicular to the wind direction we expect to see zero temporal delay. As in the previous cue, we expect correlation, after accounting for delay, to decrease with distance due to changing cloud shapes or, different clouds altogether if we move far enough perpendicular to the wind direction.

Our method for estimating the temporal offset between the time series of a pair of pixels consists of two phases. First we use cross-correlation to select the integral offset that gives the maximum correlation. Then we obtain a final estimate by finding the maxima of a quadratic model fit to the correlation values around the integer offset. We use the correlation of the temporally aligned signals as a confidence measure, *e.g*. low correlation means low confidence in the temporal offset estimate.

Figure 3 shows examples of false-color images constructed by combining the estimated delay and the temporally aligned correlation for every pixel, relative to a single landmark pixel. The motion of the clouds in this scene is nearly parallel with the optical axis, so the temporal delays are roughly equal horizontally across the image (i.e., perpendicular to the wind direction) but the correlations quickly decrease as distance from the pixel increases (i.e., different clouds are passing over those points). Orthogonally, the correlations are relatively higher in the direction of the wind but the delay changes rapidly.

3. Using Clouds to Infer Scene Structure

The dependence of correlation upon distance and the temporal delay induced by cloud shadow motion are both strong cues to the geometric structure of outdoor scenes. In this section, we describe several methods that use these cues to infer a depth map and simplified camera geometry.


(a)

(b) Landmark-pixel Correlation Maps

(c) Landmark-pixel Delay Maps

Figure 3: (a) A frame from a time-lapse video of an outdoor scene captured on a partly cloud day. (b) False-color images colored by the correlation of the time series of the highlighted landmark pixel with all other pixels in the image. (c) False-color images with colors based on the temporal delay between a landmark pixel and all other pixels in the scene. The hue of each pixel is determined by the delay and the value is determined by the confidence in the delay (low intensity regions are low confidence).

We assume a simplified pinhole camera model. Assuming a focal length, f, a point, $R_i = (X, Y, Z)$, in the world projects to an image location, expressed in normalized homogeneous coordinates as $r_i = (\frac{Xf}{Z}, \frac{Yf}{Z}, 1)$. For each pixel, i, the imaged 3D point, R_i, can be expressed as $R_i = \alpha_i r_i$ with depth, α_i. We define the 3D distance between two points as $d_{ij} = ||R_i - R_j||$. Note that the use of 3D distances is not technically correct; it should take into account the location of the sun. Consider, for example, that any two scene points in-line with the sun vector see the same cloud shadows and will therefore have similar time

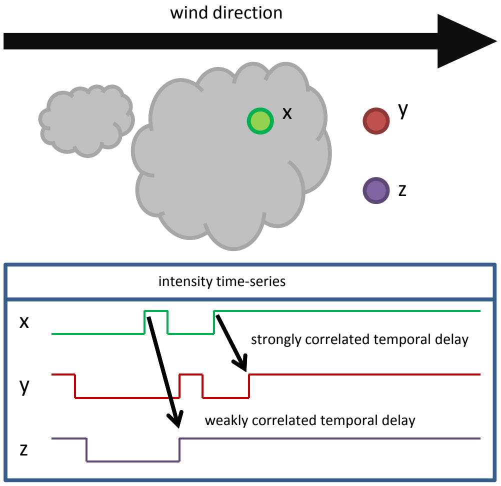

wind direction
x
y
z
intensity time-series
x
y
z
strongly correlated temporal delay
weakly correlated temporal delay

Figure 4: The time series of variation in light intensity induced by cloud shadows is dependent on the direction of motion of the wind, the shape of the clouds, and the geographic position of the scene points.

series. In our experiments, we handle this by modifying d_{ij} by projecting the points, along the sun direction vector, to the ground plane prior to computing the distance (if the sun vector is unknown we project points straight down). This gives distances that are meaningful with respect to time-series similarities induced by cloud shadows. Note that this creates a point ambiguity where the depth of a pixel ray that is parallel to the sun vector is unconstrained.

3.1. Estimating Scene Structure Using Pairwise Correlation

In outdoor scenes there is a strong relationship between correlation, ρ_{ij}, and 3D distance, d_{ij}, between the imaged scene points. In this section, we show how to estimate a depth map, $\mathbf{a} = \alpha_1, \ldots, \alpha_n$, for an outdoor scene using this relationship. The challenge with estimating the scene structure given the pairwise correlations is the unknown conditional relationship between correlation and distance between scene points, $E(d_{ij}|\rho_{ij})$. In other words, we do not know what the distance between a pair of points should be for a given value of correlation; this mapping depends on, among other factors, the type of clouds passing overhead.

We assume that the geographic correlation function (GCF) with respect to a single scene point is geographically isotropic (i.e., if you could view the correlation map of the scene from zenith, the iso-contours would be circular and the expected value of correlation would monotonically decrease with distance from the landmark pixel). This implies that for the correct scene geometry the distribution of correlation at a given distance is relatively low variance. We use this to define an error function for evaluating possible depth maps. More formally, for a good depth map, $\mathbf{a}$, we expect that the following value to be small:

$$V(\mathbf{a}) = \int \mathrm{Var}(d(\mathbf{a})|\rho)d\rho. \quad (1)$$

In a real scene, if the clouds have an isotropic GCF then the shadows cast by the clouds will likely have an anisotropic GCF unless the sun is directly overhead. Consider, for example, the elliptical shadow cast by a sphere onto the ground plane. In this work, we ignore this effect and expect the shadows to have an isotropic GCF regardless of the sun position. This is equivalent to modeling the cloud layer as having height zero.

3.1.1 Overview

We use Non-metric Multidimensional Scaling (NMDS) [10, 2] to simultaneously solve for $E(d_{ij}|\rho_{ij})$ and the depth map, $\mathbf{a}$. Like classical Multidimensional Scaling (MDS), NMDS solves for point locations given pairwise relationships between points. Unlike MDS, NMDS does not expect the input relationships to correspond to distances, instead the input is only required to have a monotonic relationship to distance. Since we assume that distance is a monotonically decreasing function of correlation, we can use the NMDS framework to solve for this mapping.

In our application NMDS works, from a high-level, as follows. First we initialize a planar depth map (see Section 3.1.2). Then we iterate through the following steps:

- determine d_{ij} for the current depth map,
- estimate the mapping from distance to correlation, $E(d_{ij}|\rho_{ij})$ (see Section 3.1.3),
- use the pairwise correlation, ρ_{ij}, and $E(d_{ij}|\rho_{ij})$ to compute a pairwise distance estimate,
- update the depth map to better fit the estimated distances (see Section 3.1.4).

We now describe the three components of this procedure in greater detail.

3.1.2 Initialization

Here we describe a method for initializing a depth map that makes the assumption that the scene is planar. We solve for the camera focal length, f, and external orientation parameters, θ_x and θ_z, that minimize the variance of the correlation-to-distance mapping. More formally, we choose parameters that minimize Equation 1,

$$\min_{f,\theta_x,\theta_z} V(f, \theta_x, \theta_z). \quad (2)$$

Note that in this case the 3D distances, d_{ij}, are a function of the three parameters, which together with a ground plane assumption imply a depth map. We exhaustively search over

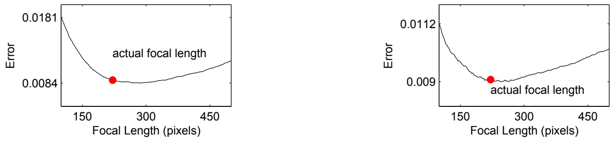


Figure 5: The error in different focal length values in our ground-plane based initialization method for two scenes. The red points correspond to the value of focal length provided by the camera in the image EXIF tags.

a reasonable range of parameters and choose the setting that minimizes the objective function.

Figure 5 shows the value of the error function defined above w.r.t. focal length for two scenes. We find that the estimated focal length is close to the ground truth value in both cases. We use the planar depth map to provide an initial estimate for the mapping from correlation to distance (the distance from the camera to the ground plane along each pixel ray). See Figure 6 for two examples of initial depth maps discovered using this method. This initial depth map is used to initialize the correlation-to-distance mapping, $E(d_{ij}|\rho_{ij})$.

3.1.3 Estimating Pairwise Distance Given Correlation

This section describes our model of the monotonic mapping from correlation to distance, $E(d_{ij}|\rho_{ij})$. Many simple parametric models could be used to fill this requirement but they impose restrictions on the mapping which can lead to substantial artifacts in the depth map. Instead we choose a non-parametric model that makes the following minimal assumptions on the form of the mapping:

- $\mathrm{E}(d_{ij}|\rho_{ij} = 1) = 0$, when the correlation is one the expected distance is zero,
- $\mathrm{E}(d_{ij}|\rho) \geq E(d_{ij}|\rho + \epsilon)$, expected distance is a monotonically decreasing function of correlation

These assumptions follow naturally from empirical studies on the spatial statistics of real clouds [15]. While these statistics are not present in all time-lapse videos, we leave for future work the task of determining which videos have the appropriate statistics.

We use the non-parametric regression method known as monotonic regression [10], to solve for a piecewise linear mapping from correlation to distance while respecting the constraints described above. The first step is choosing an optimal set of expected pairwise distances, $\hat{\mathbf{d}}$, for a fixed set control points uniformly sampled along the correlation axis (we use 100 control points). We choose values for $\hat{\mathbf{d}}$ that minimize $\sum \left|\hat{\mathbf{d}}_{\mathrm{Bin}(\rho_{ij})} - d_{ij}\right|$ relative to the distances, d_{ij}, implied by current scene model (initially a plane). Given the control point locations and corresponding optimal distance values we use linear interpolation to estimate the expected value of distance for a given correlation.

Examples of the correlation-to-distance mapping, $E(d_{ij}|\rho_{ij})$, are shown in Figure 6. Note that the expected values are reasonable when compared to the sample points and that they would be difficult to model with a single, well-justified parametric model. We use this regression model to define the expected distance between a pair of points, and we use this expected distance as input into the depth map improvement step described in the following section.

3.1.4 Translating Pairwise Distances Into Depths

We use $E(d_{ij}|\rho_{ij})$, defined in the previous section, to estimate a distance matrix. We pass this distance matrix as input to a nonlinear optimization-based Multidimensional Scaling (MDS) [2] procedure to translate estimated distances into 3D point locations. We augment MDS to respect the constraint that the 3D point locations must lie along rays defined by the camera geometry. We fix the focal length to the value estimated in the initialization step.

The error (*stress*) function for MDS is as follows:

$$S(\mathbf{a}) = \sum_{i,j} w_{ij}(d_{ij} - E(d|\rho_{ij}))^2 \tag{3}$$

where the weights, w_{ij}, are an increasing function of the correlation, ρ_{ij}. In other words, we expect the distance estimates from high-correlation pairs to be more accurate than those of lower-correlation pairs. In this work, we use $w_{ij} = \rho_{ij}^2$ for $0 \leq \rho_{ij}$ and $w_{ij} = 0$ for $\rho_{ij} < 0$. Recall that the 3D distance, d_{ij}, between imaged scene points is a function of the depths, $\mathbf{a}$, along pixel rays. We minimize the *stress* function with respect to the depths using the trust region method, constrained so that $\mathbf{a} \geq 0$. We use a straightforward application of the chain rule to compute the gradient and to form a diagonal approximation of the Hessian. We perform several descent iterations for a given distance matrix before re-estimating the correlation-to-distance mapping, $E(d_{ij}|\rho_{ij})$, using the updated point locations. Additionally, we constrain the average of the estimated pairwise distances to remain constant to avoid the trivial, zero-depth solution.

Ideally we would use all pairs of pixels when minimizing the *stress* function. We find that using a much smaller number yields excellent results and is substantially less resource intensive (we typically use around 100 randomly selected landmark pixels for a 320×240 image). In our Matlab implementation the complete depth estimation procedure, including the ground-plane based initialization, typically requires several minutes to complete.

This algorithm is essentially a projectively constrained variant of the Non-metric Multidimensional Scaling (NMDS) [10] algorithm. It is well known that NMDS is subject to local minima which can lead to suboptimal depth maps. This has not been a significant problem for depth estimation, but understanding this is an interesting area for future work. The majority of errors we see in the final depth maps are caused by erroneous, high correlations for distant pixel pairs. Frequent causes of this problem are insufficient imagery for estimating the correlation, large sun motions which cause higher correlations between surfaces with similar normals, and automatic camera exposure correction which causes shadowed pixels to be highly correlated across the image.

3.2. Estimating Scene Structure Using Temporal Delay in Cloudiness Signal

The motion of clouds due to wind causes nearby pixels to have similar but temporally offset intensity time series. Together these temporal offsets, $\Delta_{t(i,j)}$, give constraints on scene geometry. Section 2.2 shows examples of these temporal offsets.

Let W be a 3D wind vector which we assume it is fixed for the duration of the video. A pair of points in the world, R_i, R_j, that are in-line with the wind satisfy the linear constraint $R_i - R_j = W\Delta_{t(i,j)}$ where $\Delta t(i,j)$ is the time is takes for the wind (and therefore the clouds) to travel from point R_j to point R_i. However, the algorithm in Section 2.2 can often compute the temporal offset between pixels not exactly in-line with the wind. We generalize the constraint to account for this by projecting the displacement of the 3D points onto the wind direction, $\hat{W} = W/||W||$:

$$\hat{W}^\top(R_i - R_j) = \hat{W}^\top W\Delta_{t(i,j)}. \tag{4}$$

Based on the simplified camera imaging model, each pixel corresponds to a known direction, so the 3D point position, R_i, can be written as a depth, α_i, along the ray, r_i. Explicitly showing this constraint in terms of the unknown depths we find:

$$\begin{aligned} \hat{W}^\top(\alpha_i r_i - \alpha_j r_j) &= \hat{W}^\top W\Delta_{t(i,j)}, & (5)\\ \alpha_i\hat{W}^\top r_i - \alpha_j\hat{W}^\top r_j &= \hat{W}^\top W\Delta_{t(i,j)} & (6)\end{aligned}$$

This set of constraint defines a linear system,

$$\mathbf{Ma} = \mathbf{\Delta}, \tag{7}$$

where $\mathbf{a}$ is a vector of the (unknown) depth values, α_i, for each pixel, the rows of $\mathbf{M}$ contain two non-zero entries of the form $(\hat{W}^\top r_i, -\hat{W}^\top r_j)$, and $\mathbf{\Delta}$ contains the scaled temporal delays between pixels.

The constraint on depth due to temporal delay has an ambiguity. In all cases, the matrix $\mathbf{M}$ has a null space of dimension at least one. This is visible from the structure of $\mathbf{M}$, adding any multiple of $\alpha' = (\frac{1}{\hat{W}^\top r_1}, \frac{1}{\hat{W}^\top r_2}, \ldots)$ to the depth map, $\mathbf{a}$, does not change the left hand side of Equation 6. The next section describes how we overcome this ambiguity.

3.3. Combining Temporal Delay and Spatial Correlation

The two cues we describe have ambiguities, the scale ambiguity for the spatial cue and the null space ambiguity for the temporal cue, that prevent metric interpretation of the generated depth maps. Combining the two cues allows us to simultaneously remove both ambiguities and makes possible future work on metric scene estimation. We propose the following simple method.

Starting with the constraints defined by the temporal cue, we solve for a feasible depth map, $\mathbf{a}$, using a standard non-negative least squares solver. We then consider the set of solutions of the form $\mathbf{a} + k\alpha'$, and search over values of k to find a *good* depth map. While many criteria exist for evaluating a depth map we focus on combining the two cues we have described to remove this ambiguity. As with the spatial cue, we make the assumption that correlation is geographically isotropic. This motivates us to use the error function defined in Equation 2 to evaluate the different depth maps. The only difference is that we now search over the null space as opposed to the focal length and orientation parameters. In Section 4.2, we show results that demonstrate that depth maps with low error function values are more plausible than those with error function values.

4. Results

We demonstrate depth estimation on several outdoor scenes. In all examples, we resize the original images to be 320 pixels wide and assume that the sky has been manually masked off. In some cases, shadow regions are masked using automatic filtering methods based on thresholding the variance of the individual pixel time series.

4.1. Depth from Correlation

We show depth maps generated using the method described in Section 3.1. As input, we provide correlations between one hundred randomly selected pixels and all other pixels in the scene, in both cases we omit sky pixels. Examples of these correlation maps can be seen in Figure 3.

The first time lapse was captured over three hours with pictures captured every five seconds. Naïvely computing correlation on the entire video sequence yields a low quality correlation map due to long term and spatially broad changes caused by the sun motion and melting snow on fields in the near ground. Computing correlations over short temporal windows and then averaging these correlations removed these artifacts. Figure 6 shows the depth map estimated from this scene and the correlation-to-distance mapping we estimate as part of the optimization.

The second time lapse consists of 600 images captured over 50 minutes. Figure 6 shows the depth map estimated from this scene and the correlation-to-distance mapping we estimate as part of the optimization. The river and the sky were manually masked and the shadow regions were automatically masked by removing low-variance pixels.

A final example of using the spatial cue to estimate a depth map is shown in Figure 1. This time lapse demonstrates that NMDS is able to recover from significant errors in the initial depth map, for example the initial depth estimate of the rotunda was incorrect by several kilometers.

We emphasize that in these examples we perform no post-processing to improve the appearance of the generated depth maps. The optimization is based solely on geometric constraints on the camera geometry and the expectation that the correlation-to-distance mapping is geographically isotropic.

4.2. Depth from Combining Temporal Delay and Spatial Correlation

Figure 7 shows the depth map generated by the method described in Section 3.3. Note that to reduce memory usage we discard constraints for pixel pairs, ij, whose temporally aligned correlation is less than a threshold (we use threshold 0.85). The top row of the figure show results on a previously described scene. This result demonstrates that higher values of the error function lead to lower quality depth maps. For the second scene two-hundred frames of a time lapse (captured one frame every five seconds) was used to estimate a delay map. This delay map is translated into a depth map using the combined inference procedure.

5. Conclusion

We presented two novel cues, both due to cloud shadows, that are useful for estimating scene and camera geometry. The first cue, based on spatial correlation, leads to a natural formulation as a Non-metric Multidimensional scaling problem. The second cue, based on temporal delay in cloud signals, defines a set of linear constraints on scene depth that may enable metric depth estimates. These cues are unique in that they can work when other methods of inferring scene structure and camera geometry have difficulties. They require no camera motion, no haze or fog, no sun motion, and no moving people or cars. We also demonstrated how to combine these cues to obtain improved results. This work adds to the growing literature on using natural scene variations to calibrate cameras and extract scene information.

Acknowledgment

We gratefully acknowledge the support of NSF CAREER grant (IIS-0546383) which partially supported this work. We also thank Martin Setvak for the time lapses used to create Figures 3, 6, and 7.

References

[1] V. A. Billock. Neural acclimation to 1/f spatial frequency spectra in natural images transduced by the human visual system. *Phys. D*, 137(3-4):379–391, 2000.

[2] I. Borg and P. J. F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2nd ed. edition, September 2005.

[3] G. J. Burton and I. R. Moorhead. Color and spatial structure in natural scenes. *Applied Optics*, 26(1):157–170, 1987.

[4] P. M. Dare. Shadow analysis in high-resolution satellite imagery of urban areas. *Photogrammetric Engineering and Remote Sensing*, 71(2):169–177, 2005.

[5] T. Horprasert, D. Harwood, and L. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *IEEE ICCV FRAME-RATE Workshop*, 1999.

[6] N. Jacobs, N. Roman, and R. Pless. Consistent temporal variations in many outdoor scenes. In *CVPR*, June 2007.

[7] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *ICCV*, Oct. 2007.

[8] S. J. Kim, J.-M. Frahm, and M. Pollefeys. Radiometric calibration with illumination change for outdoor scene analysis. *CVPR*, pages 1–8, June 2008.

[9] S. J. Koppal and S. G. Narasimhan. Appearance derivatives for isonormal clustering of scenes. *IEEE PAMI*, 31(8):1375–1385, 2009.

[10] J. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2), June 1964.

[11] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Webcam clip art: Appearance and illuminant transfer from time-lapse sequences. *ACM Transactions on Graphics (SIGGRAPH Asia 2009)*, 28(5), December 2009.

[12] A. Mittal, A. Monnet, and N. Paragios. Scene modeling and change detection in dynamic scenes: A subspace approach. *Computer Vision and Image Understanding*, 113(1):63 – 79, 2009.

[13] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE PAMI*, 25(7):918–923, 2003.

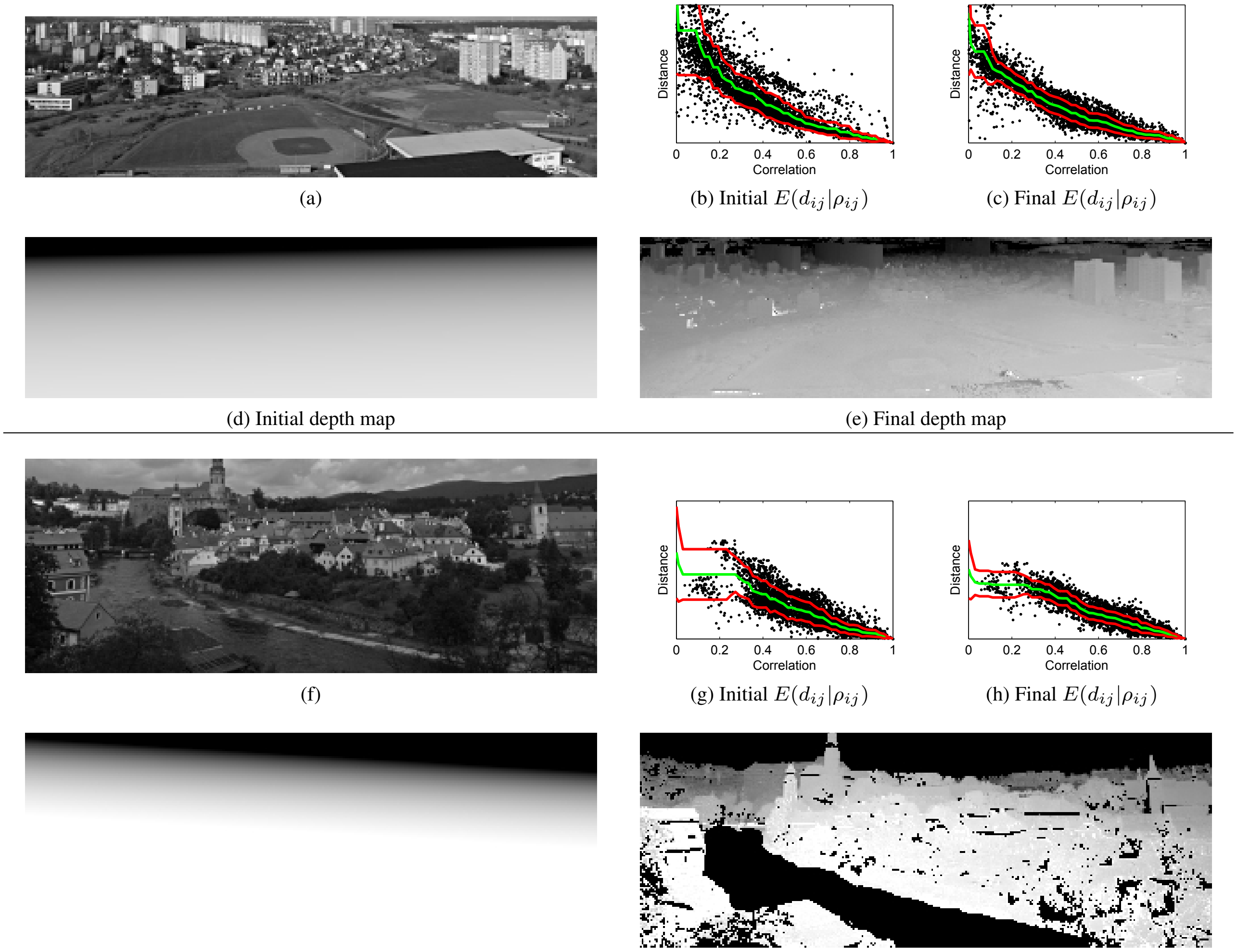
[14] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*, pages 2246–2252, 1999.

[15] C.-H. Sun and L. R. Thorne. Inferring spatial cloud statistics from limited field-of-view, zenith observations. In *Proceedings of the Fifth Atmospheric Radiation Measurements (ARM) Science Team Meeting*, pages 331–334. U.S. Department of Energy, 2000.

[16] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz. Factored time-lapse video. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3), Aug. 2007.

[17] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? *CVPR*, pages 1–8, June 2008.

[18] L. R. Thorne, K. Buch, C.-H. Sun, and C. Diegert. Data and image fusion for geometrical cloud characterization. Technical Report SAND97-9252, Sandia National Laboratories, 1997.



(a)

(b) Initial $E(d_{ij}|\rho_{ij})$

(c) Final $E(d_{ij}|\rho_{ij})$

(d) Initial depth map

(e) Final depth map

(f)

(g) Initial $E(d_{ij}|\rho_{ij})$

(h) Final $E(d_{ij}|\rho_{ij})$

(i) Initial depth map

(j) Final depth map

Figure 6: Examples of depth maps estimated using our NMDS-based method using correlations between pairs of pixels. The correlation-to-distance mappings at the optimal solution are clearly lower variance than those of the initial planar depth map.

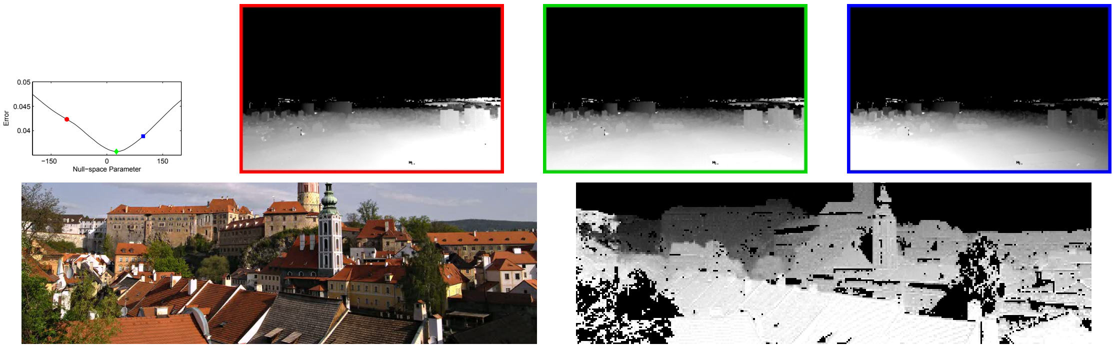


Figure 7: (top) A plot of the error function for differing depth maps created using the procedure described in Section 3.3. The plot highlights the smooth nature of this objective function for depth maps generated for different values of the null space parameter. The depth map generated at the optimal null space parameter is significantly more plausible than the others. (bottom) A cropped frame and a corresponding depth map generated for another scene using the same procedure.