2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Transforming Morning to Afternoon using Linear Regression Techniques

Stephanie M. Lowry, Michael J. Milford, *Member, IEEE*, and Gordon F. Wyeth, *Member, IEEE*

***Abstract*— Visual localization in outdoor environments is often hampered by the natural variation in appearance caused by such things as weather phenomena, diurnal fluctuations in lighting, and seasonal changes. Such changes are global across an environment and, in the case of global light changes and seasonal variation, the change in appearance occurs in a regular, cyclic manner. Visual localization could be greatly improved if it were possible to predict the appearance of a particular location at a particular time, based on the appearance of the location in the past and knowledge of the nature of appearance change over time.**

In this paper, we investigate whether global appearance changes in an environment can be learned sufficiently to improve visual localization performance. We use time of day as a test case, and generate transformations between morning and afternoon using sample images from a training set. We demonstrate the learned transformation can be generalized from training data and show the resulting visual localization on a test set is improved relative to raw image comparison. The improvement in localization remains when the area is revisited several weeks later.

I. Introduction

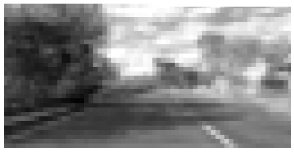
Visual localization can perform strongly on outdoor environments in situations where the environment does not undergo drastic perceptual change [1-3]. However, many systems do not perform well under changes such as that due to lighting changes, weather phenomena and seasonal changes [4].

This paper investigates whether global changes such as lighting variation in the environment can be *learned,* and subsequently *predicted*, using the lighting changes across time of day as a test case. Using sample images from a real environment, linear regression techniques are used to learn a transformation from one time of day to another (see Figure 1). The transformed images are then used to perform visual localization in the afternoon, based on images taken in the morning.

The paper proceeds as follows. Section II reviews the current approaches to appearance-based localization in changing environments. Section III describes our proposed approach to learning image transformations on an environment. Section IV describes the environment on which we perform our experiment. Section V presents results comparing the ability of transformed images to predict the future appearance of a location relative to using the untransformed image. The paper concludes in Section VI with discussion and future work.

S.M. Lowry, M.J. Milford and G.F. Wyeth are with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia. stephanie.lowry@student.qut.edu.au. This work was in part funded by the Australian Research Council through Discovery project DP1113006. S.M. Lowry is supported by an Australian Postgraduate Award and a QUT Vice-Chancellor's Scholarship.

(a) Actual morning image



(b) Predicted afternoon image

(c) Actual afternoon image

Figure 1. Linear regression techniques can be used to transform a low resolution morning image (a) to an afternoon image (b). The actual "ground truth" afternoon image is shown in (c).

II. Background

Appearance-based localization systems for robots that operate persistently over perceptually changing environments must either use comparison techniques that are robust to typical changes (lighting, weather conditions, seasonal variation), or else be able to *predict* or *learn* what those changes might be. In this section we summarize both approaches.

A. Localization in Changing Environments

Conventional image comparison methods such as SURF [5] features and SIFT [6] features do not perform well over environmental changes such as seasonal variation and changes in lighting [7]. A common method of improving performance is to include a geometric verification stage [4, 8, 9]. Image pre-processing such as patch normalization [10] or other filtering techniques can improve matching when lighting changes have occurred, as can applying the physics of light and illumination [11]. Another option is integration with other sensing modalities such as range sensors [12]. A novel solution to the problem of achieving lighting invariant visual localization is to use a scanning laser-rangefinder in place of a camera [13], and convert the resulting data into a *camera-like image* that is unaffected by lighting change.

One approach using a camera-only system is to generate a *plastic map* [14] that is built on robot experiences rather than physical locations – that is, a new experience is generated each time a robot visits a location that it does not recognize as the same place, and localization occurs within experiences rather than a physical or topological space. However, this approach loses potentially valuable connectivity information, and certainly the more frequently the robot is able to

correctly localize (and thus link experiences) the more useful the resulting map will be.

Localizing against image sequences [10, 15] rather than individual images can significantly improve visual localization ability when environments have experienced drastic change. A sequence-based approach does not require the image comparison step to achieve 100% correctness. However, as for any system, a better visual front-end can only improve overall performance.

B. Predicting Change for Localization

Predicting change can involve either a temporal generalization – predicting appearance at a previously unseen time – or a spatial generalization – predicting appearance at a previously unseen location. An example of temporal generalization is seen in [16], where feature co-occurrence maps generated during training runs at different times of day are interpolated between to achieve localization on the same route, at times between these two. However, the system requires that the appearance changes gradually between runs.

An example of spatial generalization is presented in [17]. In this system, training data across seasonal changes is used to generate a superpixel vocabulary and a "dictionary". This dictionary can be used to translate a visual word seen in one season to its matching appearance during another season.

III. Proposed Approach

In this paper we present a method of generalizing about temporal change in appearance over an environment. The method uses matched image pairs from morning to afternoon as training data to learn what temporal changes occur throughout different spatial locations. Learning is performed using linear regression techniques on the matched image pairs to find a transformation from one time of day to another. This transformation is then used to predict the change in appearance of a location at a different time of day.

We present results across two sets of test data, both of which are distinct from the training data. The first test set was captured at the same time as the training data, while the second was captured several weeks later. We show that localization ability can be improved by learning a transformation across time and using that transformation to predict the future appearance of an image. These transformations can either be performed globally across an environment, or be combined with a classification step to find transformations for distinct image clusters.

A. Global linear regression

In order to learn an appropriate transformation, we require training data in the form of two image sets $\{A_1, A_2, ..., A_n\}$ taken over time period t_1 (e.g. morning) and $\{B_1, B_2, ..., B_n\}$ taken over time period t_2 (e.g. afternoon) such that for each $i = 1, ..., n$, the image A_i and B_i come from the same location. If the images are of size $m \times p$ pixels, the transformation will be a matrix of size $(mp+1) \times mp$, where the 1 allows for an offset term.

For the global linear regression technique, the transformation T is calculated via a matrix inversion:

$$T = (A'A)^{-1} A'B. \quad (1)$$

Here A and B are two matrices where each row of A represents an image from $\{A_1, A_2, ..., A_n\}$ augmented by 1 to provide the offset and each row of B represents an image from $\{B_1, B_2, ..., B_n\}$. Once calculated, this transformation can be applied to any image captured in time period t_1 to generate a prediction of the appearance of that image in time period t_2.

B. Clustered Linear Regressions

The transformation created in the previous section was generated across all images from the training set. In this section, we classify the images into groups of similar images, and then learn a separate transformation for each image type.

To classify the images, we used the *k*-means clustering technique in image space; that is, the space where each pixel of the $m \times p$ images represented a separate dimension. The *k*-means approach generates *k* clusters from the data, and for each cluster determines a centroid point that represents the mean of all data points in each cluster. In our case, this centroid can be considered a "mean image" for each cluster and some examples of mean images are shown in Figure 2.

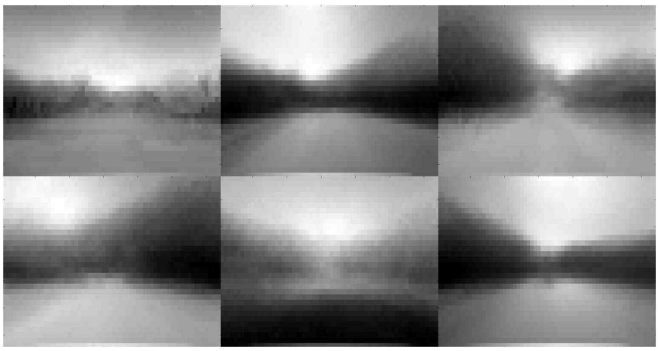


Figure 2. Mean images generated by *k*-means clustering using $k = 6$.

Following the clustering stage, the learning transformation described in Equation (1) was performed individually for each cluster. Test images were allocated to clusters using a nearest-neighbor approach – the distance from an image I to each cluster centroid was calculated and I was allocated to the cluster for which the distance was the least. Figure 3 shows examples of clustered images.

In order to use this transformation on an image I, the image is first classified according to the defined clustering using the same nearest-neighbor approach as for the training images. The predicted change in appearance of the image is then calculated using the relevant transformation.

IV. Experiment Setup

In the following experiments, we tested the effectiveness of both global and clustered linear regression in predicting the appearance of locations over time. A global linear regression and a clustered linear regression over image clusters (with $k = 6$ clusters) were generated over the training data.

Figure 3. Examples of image grouped using *k*-means clustering and nearest-neigbor allocation. Each of these correspond to the cluster with a centroid denoted by the matching "mean image" in Figure 2.

A. *Testing Environment*

The experimental environment was a large outdoor dataset first presented in [18]. This dataset provided visual and GPS data from a car driven around suburban streets for approximately 15 kilometers (see Figure 4). The route was repeated at a number of different times of day, and then the entire dataset was repeated about three weeks later. In this paper, we test on circuits from both sets of traversals (here denoted as the Week 1 and Week 2 sets). GPS data was logged at 1 Hz and used for ground truth.

Figure 4. Ground truth of St Lucia dataset. (Imagery ©2012 Cnes/Spot Image, DigitalGlobe, GeoEye, Sinclair Knight Merz & Fugro).

B. *Image Processing*

The images were captured by a forward-facing commercial web camera at a resolution of 640×480 pixels and an average frequency of 15 fps. For this experiment the images were down sampled to 32×64 pixels, resulting in a 2049×2048 transformation. As the dataset circuits were captured at different times ranging from early morning to late afternoon, the environment experienced moderate perceptual change, due largely to the effect of shadows (see Figure 5 for sample images).

C. *Training and Testing Environments*

Transformations were learned from a morning dataset captured at 8:45am to two afternoon datasets – 14:10pm and 15:45pm. Each of these datasets from the Week 1 set were split into two – the first half was used for training and the second half for testing. The complete 14:10pm and 15:45pm datasets from the Week 2 set were used as test sets to test if the learned transformations could still be applied several weeks later.

Figure 5. Sample images from dataset, showing the effect of light and shadow on the appearance of locations. The images on the left were taken in the morning; the images on the right were taken in the afternoon.

V. Results

In this section we present the results of transformations from images captured in the morning to images captured in the afternoon. The "predicted" images generated from the learned transformations were compared to actual "ground truth" images from the afternoon. The mean squared error between the predicted image and the actual images was calculated and compared to the mean squared error between the original morning image and the true afternoon image. If I_{am} is an image captured in the morning and I_{pm} is an image captured at the same location in the afternoon, then the pixel-wise mean squared error is given by:

$$MSE = \frac{1}{2048}\sum_{p=1}^{2048}\left(I_{am}(p) - I_{pm}(p)\right)^2 . \qquad (2)$$

This value is compared to the mean squared error of using the transformed image I_{am}^{trans} :

$$MSE^{trans} = \frac{1}{2048}\sum_{p=1}^{2048}\left(I_{am}^{trans}(p) - I_{pm}(p)\right)^2. \qquad (3)$$

Precision-recall curves were also generated to test the localization ability of the transformed images. Image comparison was performed using the sum of squared pixel differences (an equivalent calculation to the *MSE* calculation). Localization was determined by finding the image with the least MSE. Similar pixel comparison methods are employed as a visual front-end by SeqSLAM [19] and have been used to perform effective localization over changing environments.

The performance of the transformation was evaluated on images gathered at 08:45 (morning) when compared to images gathered at 14:10 (afternoon) in Figure 6. Figure 6 (a) shows the mean squared error on the training set for comparisons between the afternoon images and (i) the original morning images (left hand box); (ii) the globally transformed morning images (center box); and (iii) the clustered transformed images (right hand box). Figure 6 (b) shows the same calculations, but on the test set. Figure 6 (c) shows the precision and recall curve for appearance based localization achieved on the unseen Week 1 test data, demonstrating the ability of the transformation to improve localization performance. Figure 7 shows the same performance metrics for transformations from 08:45 to 15:45 – later in the afternoon.

The transformation was also tested on data captured three weeks later. Figures 8 and 9 show the improvements in the precision and recall curve for visual localization using the Week 2 test data.

The results displayed in Figures 6-9 show that it is possible to some degree to "learn" the transformation an environment undergoes over the course of a day, using simple linear techniques. There is a significant decrease in error shown in Figure 6 (a) and Figure 7 (a) when the predicted images are compared to the afternoon images in place of the original morning images. The decrease in error is smaller for the test images (Figures 6 (b) and 7 (b)). However, the improvement in similarity is sufficient to provide a localization improvement. The improvement in localization ability is also retained, when the area is revisited three weeks later, as shown in Figures 8 and 9.

The clustered transformation outperforms the global transformation at localization in each case, except for high precision cases in Figure 7 (c). Both outperform the untransformed images on every dataset, achieving between 3% and 7% recall at 100% precision, compared to the untransformed images whose performance varies between 0% and 2%. Furthermore, if the precision is dropped to 95% the clustered transformation method consistently achieves approximately 15% on all the tested datasets.

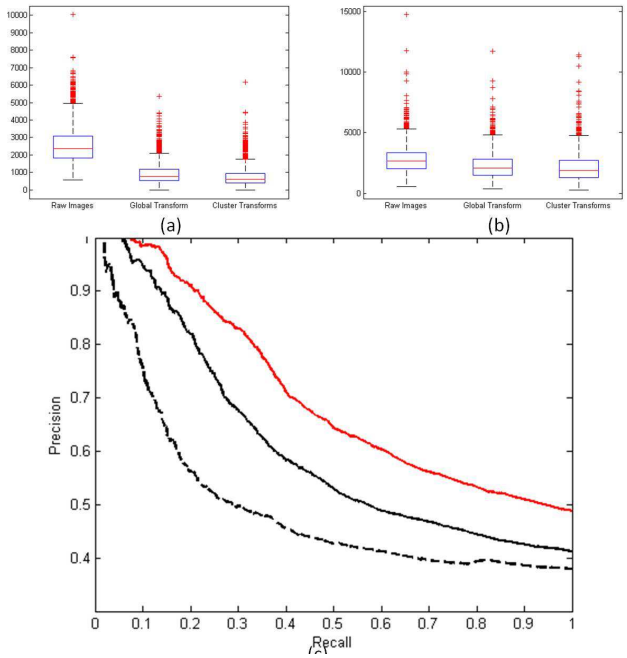


Figure 6. Results of transformation from 8:45 am to 14:10 pm datasets, including MSE on training set (a), MSE on test set (b), and precision-recall curve (c) on test set, localizing at 14:10pm from images captured at 8:45am. The dashed black dashed line shows localization using the original 8:45am images. The solid black line uses a global transformation to predict images at 14:10pm, and the solid red line uses clustered independent linear regressions across image clusters to predict images.

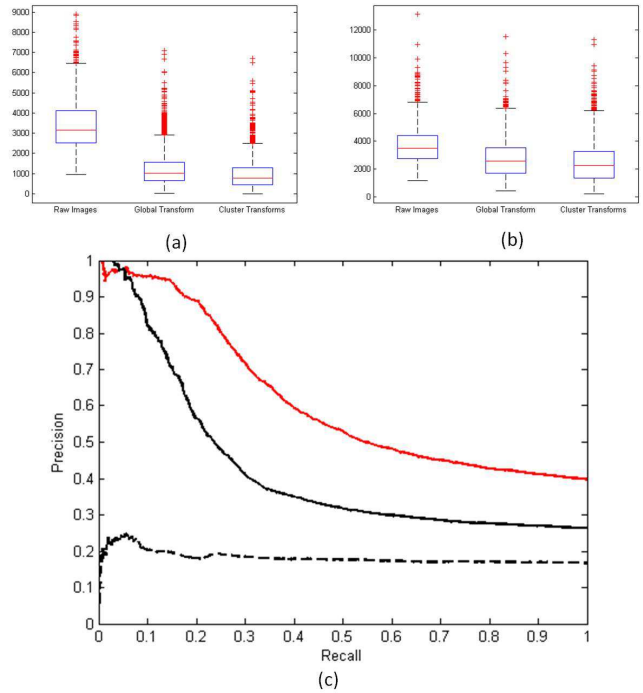


Figure 7. Results of transformation from 8:45 am to 15:45 pm datasets, including MSE on training set (a), MSE on test set (b), and precision-recall curve (c) on test set, localizing at 14:10pm from images captured at 8:45am. The dashed black dashed line shows localization using the original 8:45am images. The solid black line uses a global transformation to predict images at 15:45pm, and the solid red line uses clustered independent linear regressions across image clusters to predict images.

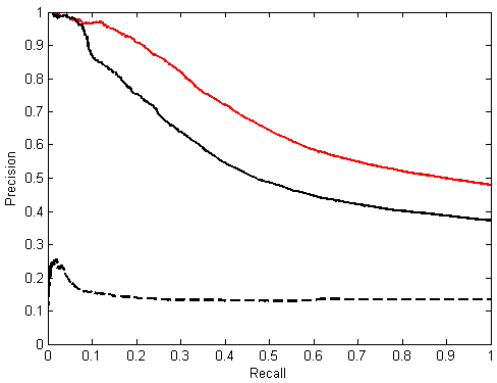


Figure 8. Precision-recall curve from 8:45 am to 15:45 pm datasets, across 3 week gap for untransformed images (dashed black line), global transformation process (solid black line) and cluster-based transformation (solid red line).

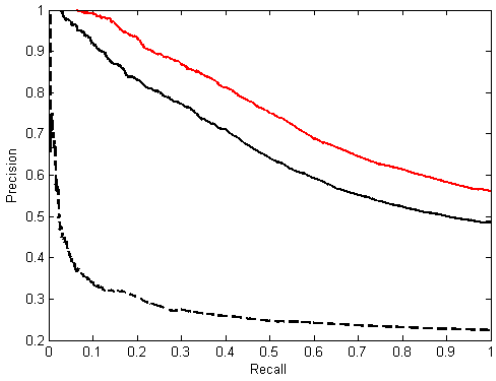


Figure 9. Precision-recall curve from 8:45 am to 14:10 pm datasets, across 3 week gap for untransformed images (dashed black line), global transformation process (solid black line) and cluster-based transformation (solid red line).

The examples of image transformations between morning and afternoon show the nature of the underlying transformation. Figure 10 shows the effect of a linear transformation and Figure 11 shows some outcomes of a clustered linear transformation on images from the morning dataset. The left-hand image in each row is the original morning image. The middle image is the "predicted" afternoon image, and the right-most image is a true afternoon image taken from the same location. The most notable effect is the change in shadows on the road in each case, and the change to the sky. The transformations are able to both remove and add shadows for the change in time of day.

Figure 10. Examples of successful image prediction from morning to afternoon using a global linear transformation. Each row shows (from left to right), original morning image, predicted afternoon image, actual afternoon image.

Figure 11. Examples of successful image prediction from morning to afternoon using a cluster-based linear transformation. Each row shows (from left to right), original morning image, predicted afternoon image, actual afternoon image. Note the change in the sky, and the appearance and disappearance of shadows from the road.

VI. Discussion and Conclusion

This paper presents a preliminary investigation into learning appearance transformations across environments that experience global lighting change. The results show linear regression techniques can be used to approximate the appearance change experienced by an outdoor environment over the course of a day, and the resulting predictive model can improve the ability to perform visual localization. The results are improved if the environment is clustered into a few coarse "types". As few as six clusters can improve precision and recall when localizing in the late afternoon based on early morning data.

Currently the transformation operates in a very high-dimension space. Each image pixel in the 32×64 sized image represents one out of a total of 2048 separate dimensions. The goal of this research was to determine whether transformations in image space could be successfully performed, but the high dimensionality of the dataset leads to over-fitting the training data. This problem limits the spatial generalizability of the transformation, and the algorithm only performs effectively when the test and training data come from areas that are nearby. Future work will explore forms of dimensionality reduction such as PCA, which has been used [20] on imagery from static webcams to determine spatial and temporal subsets, and partial least squares regression [21].

The system could benefit from more sophisticated learning techniques. An outlier removal process such as RANSAC [22] could be applied to the training set to remove any images that are too different – for example, when the camera's view is blocked by a bus passing by. We are currently investigating the use of Locally Weighted Regression [23] to replace the current *k*-means clustering process with nearest neighbor partitioning of the image data. Other non-linear regression techniques may also improve performance, as may the application of prediction techniques that learn affine transfers from similar scenes [24].

A different transformation must be generated between each time of day, with each configuration requiring its own training set. Future work includes developing transformations that are more general, and can be applied across different environmental configurations.

Acknowledgment

We thank Arren Glover for access to the St Lucia dataset. This dataset is available for download at https://wiki.qut.edu.au/display/cyphy/St+Lucia+Multiple+Times+of+Day).

References

[1] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *International Journal of Robotics Research,* vol. 27, pp. 647-665, Jun 2008.

[2] W. Maddern, M. Milford, and G. Wyeth, "CAT-SLAM: Probabilistic Localisation and Mapping using a Continuous Appearance-based Trajectory," *International Journal of Robotics Research,* vol. 31, pp. 429-451, 2012.

[3] A. Murillo, G. Singh, J. Kosecka, and J. Guerrero, "Localization in Urban Environments Using a Panoramic Gist Descriptor," *Ieee Transactions on Robotics,* pp. 1-15, 2013.

[4] C. Valgren and A. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems,* vol. 58, pp. 157-165, Feb 2010.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding,* vol. 110, pp. 346-359, Jun 2008.

[6] D. Lowe, "Object recognition from local scale-invariant features," in *IEEE International Conference on Computer Vision (ICCV)*, 1999, pp. 1150-1157.

[7] C. Valgren and A. Lilienthal, "SIFT, SURF and Seasons: Long-term Outdoor Localization Using Local Features," in *European Conference on Mobile Robotics (ECMR)*, 2007, pp. 1-6.

[8] C. Cadena and J. Neira, "A learning algorithm for place recognition," presented at the ICRA 2011 Workshop on Long-term Autonomy, Shanghai, China, 2011.

[9] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *International Journal of Robotics Research,* vol. 30, pp. 1100-1123, 2011.

[10] M. Milford and G. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1643-1649.

[11] P. Corke, R. Paul, W. Churchill, and P. Newman, "Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, 2013, pp. 2085-2092.

[12] H. Badino, D. Huber, and T. Kanade, "Real-time topometric localization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1635-1642.

[13] C. McManus, P. Furgale, and T. D. Barfoot, "Towards lighting-invariant visual navigation: An appearance-based approach using scanning laser-rangefinders," *Robotics and Autonomous Systems,* 2013.

[14] W. Churchill and P. Newman, "Practice makes perfect? Managing and leveraging visual experiences for lifelong navigation," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 2012, pp. 4525-4532.

[15] M. Milford, I. Turner, and P. Corke, "Long exposure localization in darkness using consumer cameras," in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, 2013.

[16] E. Johns and G.-Z. Yang, "Feature Co-occurrence Maps: Appearance-based Localisation Throughout the Day," in *Proc. ICRA*, 2013.

[17] N. Sünderhauf, P. Neubert, and P. Protzel, "Predicting the change– a step towards life-long operation in everyday environments," presented at the Robotics: Science and Systems (RSS) Robotics Challenges and Vision Workshop, Berlin, 2013.

[18] A. Glover, W. Maddern, M. Milford, and G. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 3507-3512.

[19] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, 2012, pp. 1643-1649.

[20] A. Abrams, E. Feder, and R. Pless, "Exploratory analysis of time-lapse imagery with fast subset PCA," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*, 2011, pp. 336-343.

[21] H. Wold, "Partial least squares," *Encyclopedia of Statistical Sciences,* 1985.

[22] M. A. Fischler and R. C. Bolles, "Random Sample Consensus - a Paradigm for Model-Fitting with Applications to Image-Analysis and Automated Cartography," *Communications of the ACM,* vol. 24, pp. 381-395, 1981.

[23] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association,* vol. 74, pp. 829-836, 1979.

[24] Y. Shih, S. Paris, F. Durand, and W. T. Freeman, "Data-driven hallucination of different times of day from a single outdoor photo," *ACM Transactions on Graphics (TOG),* vol. 32, p. 200, 2013.